(Sackler NAS Colloquium) Self-Perpetuating Structural States in Biology, Disease, and Genetics

Proceedings of the National Academy of Sciences

More information

Find similar titles

Share this PDF

Visit the National Academies Press online and register for...

✔ Instant access to free PDF downloads of titles from the

- ▪ NATIONAL ACADEMY OF SCIENCES

- ▪ NATIONAL ACADEMY OF ENGINEERING

- ▪ INSTITUTE OF MEDICINE

- ▪ NATIONAL RESEARCH COUNCIL

✔ 10% off print titles

✔ Custom notification of new releases in your field of interest

✔ Special offers and discounts

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Arthur M. Sackler
C O L L O Q U I A
OF THE NATIONAL ACADEMY OF SCIENCES

Self-Perpetuating Structural States in Biology, Disease, and Genetics

Edited by Susan Lindquist and Steve Henikoff

National Academy
of Sciences

Washington, DC

March 22–24, 2002

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the U.S. Congress in 1863, the Academy has a mandate that requires it to advise the Federal Government on scientific and technical matters.

Arthur M. Sackler
C O L L O Q U I A
OF THE NATIONAL ACADEMY OF SCIENCES

Self-Perpetuating Structural States in Biology, Disease, and Genetics

National Academy of Sciences
Washington, D.C.

Arthur M. Sackler, M.D.
1913–1987

Born in Brooklyn, New York, Arthur M. Sackler was educated in the arts, sciences, and humanities at New York University. These interests remained the focus of his life, as he became widely known as a scientist, art collector, and philanthropist, endowing institutions of learning and culture throughout the world.

He felt that his fundamental role was as a doctor, a vocation he decided upon at the age of four. After completing his internship and service as house physician at Lincoln Hospital in New York City, he became a resident in psychiatry at Creedmoor State Hospital. There, in the 1940s, he started research that resulted in more than 150 papers in neuroendocrinology, psychiatry, and experimental medicine. He considered his scientific research in the metabolic basis of schizophrenia his most significant contribution to science and served as editor of the *Journal of Clinical and Experimental Psychobiology* from 1950 to 1962. In 1960 he started publication of *Medical Tribune*, a weekly medical newspaper that reached over one million readers in 20 countries. He established the Laboratories for Therapeutic Research in 1938, a facility in New York for basic research that he directed until 1983.

As a generous benefactor to the causes of medicine and basic science, Arthur Sackler built and contributed to a wide range of scientific institutions: the Sackler School of Medicine established in 1972 at Tel Aviv University, Tel Aviv, Israel; the Sackler Institute of Graduate Biomedical Science at New York University, founded in 1980; the Arthur M. Sackler Science Center dedicated in 1985 at Clark University, Worcester, Massachusetts; and the Sackler School of Graduate Biomedical Sciences, established in 1980, and the Arthur M. Sackler Center for Health Communications, established in 1986, both at Tufts University, Boston, Massachusetts.

His pre-eminence in the art world is already legendary. According to his wife Jillian, one of his favorite relaxations was to visit museums and art galleries and pick out great pieces others had overlooked. His interest in art is reflected in his philanthropy; he endowed galleries at the Metropolitan Museum of Art and Princeton University, a museum at Harvard University, and the Arthur M. Sackler Gallery of Asian Art in Washington, DC. True to his oft-stated determination to create bridges between peoples, he offered to build a teaching museum in China, which Jillian made possible after his death, and in 1993 opened the Arthur M. Sackler Museum of Art and Archaeology at Peking University in Beijing.

In a world that often sees science and art as two separate cultures, Arthur Sackler saw them as inextricably related. In a speech given at the State University of New York at Stony Brook, *Some reflections on the arts, sciences and humanities,* a year before his death, he observed: "Communication is, for me, the *primum movens* of all culture. In the arts. . . I find the emotional component most moving. In science, it is the intellectual content. Both are deeply interlinked in the humanities." The Arthur M. Sackler Colloquia at the National Academy of Sciences pay tribute to this faith in communication as the prime mover of knowledge and culture.

PNAS

Proceedings of the National Academy of Sciences
of the United States of America

Contents

Colloquium

Self-perpetuating structural states in biology, disease, and genetics

Susan L. Lindquist*† and Steven Henikoff‡

*Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA 02142; and ‡The Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Over the past half-century, the central dogma, in which DNA makes RNA makes protein, has dominated thinking in biology, with continuing refinements in understanding of DNA inheritance, gene expression, and macromolecular interactions. However, we have also witnessed the elucidation of epigenetic phenomena that violate conventional notions of inheritance. Protein-only inheritance involves the transmission of phenotypes by self-perpetuating changes in protein conformation. Proteins that constitute chromatin can also transmit heritable information, for example, via posttranslational modifications of histones.

Both the transmission of phenotypes via the formation of protein conformations and the inheritance of chromatin states involve self-perpetuating assemblies of proteins, and there is evidence for some common structural features and conceptual frameworks between them. To foster interactions between researchers in these two fields, the National Academy of Sciences convened an Arthur M. Sackler Colloquium entitled "Self-Perpetuating Structural States in Biology, Disease, and Genetics" in Washington, DC, on March 22–24, 2002. Participants described new phenomenology and provided insights into fundamental mechanisms of protein and chromatin inheritance. Perhaps most surprising to attendees was emerging evidence that these unconventional modes of inheritance may be common.

First described in studies of scrapie and other transmissible encephalopathies in mammals, prions were later shown to cause some classical phenotypes in yeast. In each case, an alternative protein conformation leads to formation of structures resembling amyloid fibers seen in human disease. How these are seeded has been elucidated by *in vitro* studies, leading to a satisfying picture of prion-like protein propagation. Other cases of prion inheritance have been discovered in genetic screens, which suggests that we are seeing only the tip of the iceberg. Indeed, it now appears that amyloid fiber formation is the default state for misfolded proteins, and fibrillar aggregates found in amyloidoses result from defects in the cellular machinery that prevents protein misfolding.

Excitement also pervades the chromatin field, with new insights into how nucleosomes specify and maintain distinct chromatin states. Remarkably, a single modification of a histone tail residue underlies the distinction between euchromatin and heterochromatin, and even maintenance of DNA methylation can depend on histone tail modification. From insights such as these, we have begun to realize that the relationship between chromatin conformation and gene expression might have a simple basis. Genetic and biochemical approaches have begun to elucidate how histone-modifying enzymes and nonhistone structural proteins regulate chromatin inheritance. Although these alternate mechanisms of inheritance have shaken our blind faith in the central dogma, they whet our appetite for further revolutionary insights.

Colloquium

Transmission of prions

C. Weissmann*, M. Enari†, P.-C. Klöhn, D. Rossi‡, and E. Flechsig§

Medical Research Council Prion Unit, Department of Neurodegenerative Disease, Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom

The "protein only" hypothesis states that the infectious agent causing transmissible spongiform encephalopathies is a conformational isomer of PrP, a host protein predominantly expressed in brain, and is strongly supported by many lines of evidence. Prion diseases are so far unique among conformational diseases in that they are transmissible, not only experimentally but also by natural routes, mainly by ingestion. A striking feature of prions is their extraordinary resistance to conventional sterilization procedures, and their capacity to bind to surfaces of metal and plastic without losing infectivity. This property, first observed in a clinical setting, is now being investigated in experimental settings, both in animals and in cell culture.

Transmissible spongiform encephalopathies (TSEs), or prion diseases, are degenerative disorders of the central nervous system leading to motor dysfunction, dementia, and death. Prion diseases include scrapie of sheep, bovine spongiform encephalopathy (BSE) in cattle, and human diseases such as Creutzfeldt-Jakob disease (CJD), Gerstmann-Sträussler-Scheinker syndrome (GSS) and fatal familial insomnia (FFI). More recently, variant CJD (vCJD), ascribed to consumption of BSE-contaminated products (1), has claimed over 120 victims. Neither humoral nor cellular immunological responses have been detected in prion diseases.

Transmissibility of scrapie was first demonstrated in 1939 (2). The remarkable resistance of the causative agent, later designated prion, was revealed early on when 10% of a flock of Scottish sheep came down with scrapie after having been injected with a vaccine against looping ill prepared from formaldehyde-treated sheep brain extract (3). The agent's unusual resistance to UV irradiation suggested that it might be devoid of nucleic acid (4). The "protein only" hypothesis (5) in its updated version (6) proposes that the prion is a conformational isomer of the normal host protein PrP^C (7, 8), which is found predominantly on the outer surface of neurons, attached by a glycosylphosphatidylinositol (GPI) anchor. The abnormal conformer, when introduced into the organism, is thought to cause the conversion of PrP^C into a likeness of itself.

In prion disease, a largely protease-resistant, aggregated form of PrP designated PrP^{Sc}, accumulates, mainly in brain. It is believed to be the principal or only constituent of the prion (6). No differences in the primary structure of PrP^C and PrP^{Sc} were detected, suggesting that they differ in their conformation (9). The tertiary structure of PrP^C has been elucidated (10), whereas that of PrP^{Sc} has not; however, the β-sheet content of PrP^{Sc} was shown to be high whereas that of PrP^C is low (11, 12). The conclusion that some form of PrP is the essential, perhaps only, constituent of the infectious agent is based on compelling biochemical and genetic evidence (13, 14). The finding that PrP knockout ($Prnp^{o/o}$) mice are completely protected against scrapie disease and fail to propagate prions (15, 16) and that introduction of murine $Prnp$ transgenes into these mice restores susceptibility to prions (17) is one of the main supports for the "protein only" hypothesis.

Within the framework of the "protein only" hypothesis, the "refolding model" (Fig. 1A) proposes that PrP^C unfolds to some extent and refolds under the influence of a PrP^{Sc} molecule and that the two states are separated by an activation energy barrier (18). The "nucleation model" (Fig. 1B) postulates that PrP^C is in equilibrium with PrP^{Sc} (or a precursor thereof), that the equilibrium is largely in favor of PrP^C, and that PrP^{Sc} is stable only when it forms a multimer. Once such a multimer or seed is present, monomer addition ensues rapidly (19). "Breakage" of aggregates must be postulated to explain the exponential increase of PrP^{Sc} during infection (20). Conversion in vitro of PrP^C to a PrP^{Sc}-like product has been achieved by incubating ^{35}S-labeled PrP^C with PrP^{Sc} and demonstrating the appearance of a protease-resistant radioactive product with the mobility of protease-treated authentic PrP^{Sc} (21). This in vitro conversion exhibits the species specificity (22) and strain-specificity (23) observed in vivo. However, because the yield is less than stoichiometric with regard to the PrP^{Sc} used as seed, it has not been possible to determine whether or not there was an increase in infectivity. Perhaps the "cyclic amplification" procedure reported recently will lead to this goal (24). Although it has been possible to convert recombinant PrP^C into a β-sheet-rich, partially protease-resistant structure by physico-chemical procedures (25, 26), there have so far been no reports that such material gives rise to transmissible prion disease (27–29). Also, it has so far not been possible to renature completely denatured prion preparation to an infectious state (30, 31) although the infectivity of partially inactivated material can be increased by renaturation under certain conditions (32, 33). Prusiner and his colleagues have reported that intracerebral injection of a synthetic 55-residue peptide corresponding to region 89–143 of mouse PrP with a P101L substitution can induce neurological, prion-like disease, however this is achieved only in transgenic mice expressing PrP with the same mutation (29). The caveats here are that these transgenic mice show spontaneous disease even without inoculation, albeit only much later, and that transmissibility has yet to be demonstrated.

"Natural" Transmission of Prions

Although prion diseases are not contagious in the strict sense, i.e., by direct contact, they are transmissible perorally and parenterally. The BSE epidemic that emerged in the mid-eighties

A "Refolding" model

B "Seeding" model

Fig. 1. Models for the conformational conversion of PrP^C to PrP^{Sc}. (*A*) The "refolding" model. The conformational change is kinetically controlled, a high activation energy barrier preventing spontaneous conversion at detectable rates. Interaction with exogenously introduced PrP^{Sc} causes PrP^C to undergo an induced conformational change to yield PrP^{Sc}. This reaction could be facilitated by an enzyme or chaperone. In the case of certain mutations in PrP^C, spontaneous conversion to PrP^{Sc} may occur as a rare event, explaining why familial CJD or GSS arise spontaneously, albeit late in life. Sporadic CJD may come about when an extremely rare event (occurring in about one in a million individuals per year) leads to spontaneous conversion of PrP^C to PrP^{Sc}. (*B*) The "seeding" model. PrP^C and PrP^{Sc} (or a PrP^{Sc}-like molecule, light) are in equilibrium, with PrP^C strongly favored. PrP^{Sc} is stabilized only when it adds onto a crystal-like seed or aggregate of PrP^{Sc} (dark). Seed formation is rare; however, once a seed is present, monomer addition ensues rapidly. To explain exponential conversion rates, aggregates must be continuously fragmented, generating increasing surfaces for accretion.

and led to about 180,000 clinically diagnosed cases (and likely to many times more nondiagnosed ones) was fueled by the feeding of BSE-prion-contaminated bone-and-meat meal to cattle (34). The kuru epidemic that developed in the first half of the 20th century in Papua New Guinea was caused by ritualistic cannibalism (35) and is believed to have originated from a case of sporadic CJD. Variant CJD is thought to come about by ingestion of BSE-prion-contaminated foodstuff, and certainly mice (36), sheep (37), calves (38), and non-human primates (39, 40) can be experimentally infected with the BSE agent by the oral route. It appears quite likely that sheep scrapie spreads by ingestion of the infectious agent, although the source has not been established; infected placenta has been suggested (41), but scrapie-prion-contaminated feces are a likely possibility that merits investigation. Perhaps the appearance of vCJD in predominantly young individuals is due to infection by contaminated foodstuff through wounds resulting from teething and tooth loss between early infancy and adolescence. Experimental transmission by the dental route has been shown in hamster (42).

Somehow prions make their way from the digestive tract to the central nervous system (Fig. 2). The relative resistance of prion infectivity to protease digestion (43) probably allows a significant proportion of the infectious agent to survive passage through the digestive tract (36). It is not clear how prions pass through the
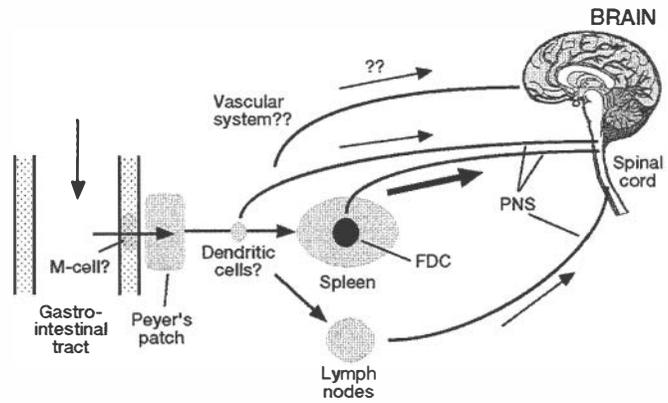
Fig. 2. Possible routes of propagation of ingested prions. After oral uptake, prions may penetrate the intestinal mucosa through M cells and reach Peyer's patches as well as the enteric nervous system. Depending on the host, prions may replicate and accumulate in spleen and lymph nodes. Myeloid dendritic cells are thought to mediate transport within the lymphoreticular system. From the lymphoreticular system and likely from other sites, prions proceed along the peripheral nervous system to finally reach the brain, either directly via the vagus nerve or via the spinal cord, under involvement of the sympathetic nervous system.

intestinal mucosa. M cells, which are portals for antigens and pathogens (44–46), are able to mediate transport of prions, at least in an experimental setting (47). Thus, after oral uptake, the infectious agent may penetrate the mucosa through M cells and reach Peyer's patches, where they are found early on (36) as well as the enteric nervous system (48). Depending on the host, other tissues of the lymphoreticular system (LRS), in particular the spleen but also lymph nodes (49), are sites in which prions replicate and accumulate; this result is the case in sheep scrapie, experimental BSE in sheep, vCJD in man, and experimental mouse scrapie, but not BSE in cattle (50). Recent reports suggest that myeloid dendritic cells mediate transport within the lymphoreticular system (51, 52). Interestingly, mature B cells (with or without PrP^C expression) are required for amplification of prions in spleen (53), not, however, because they themselves harbor or multiply prions (54), but because they are required for the maturation of follicular dendritic cells, the cells in which prion amplification and PrP^{Sc} accumulation occur (55, 56). Nonetheless, neuroinvasion is possible even in the absence of follicular dendritic cells, suggesting that other cell types in the periphery also can amplify prions (49, 57). From the LRS and likely from other sites, prions proceed along the peripheral nervous system to finally reach the brain, either directly via the vagus nerve (58) or via the spinal cord, under involvement of the sympathetic nervous system (59). If a sufficient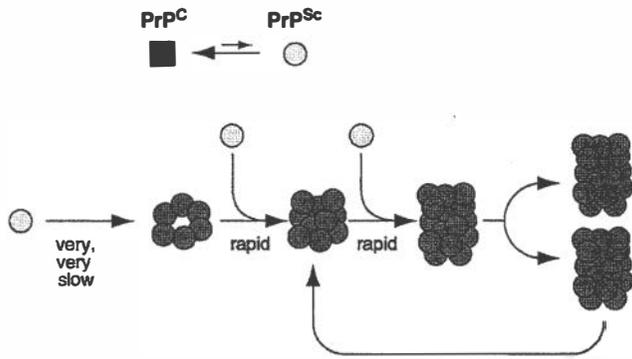ly high dose of prions is administered i.p., neuroinvasion can occur without participation of the LRS (60). Although prions have not been detected in muscle of scrapie or BSE-infected animals, infectivity has been found in some, but not all skeletal muscles of mice experimentally infected with ME7 or RML prions (61).

Not only the biosynthesis of prions, but also their spread depends on PrP-containing cells. This result was demonstrated by the finding that a PrP-expressing neuroectodermal graft in the brain of a *Prnp*$^{o/o}$ mouse could be infected by intracerebral injection of mouse prions but not by intraocular (62) or i.p. inoculation (63). Even after irradiation and reconstitution with a PrP-expressing lymphohemopoietic system, prions failed to reach the graft after i.p. or i.v. inoculation, showing that neuroinvasion, at least in the mouse, was not mediated by prion transport through the circulation (63) and underlining the requirement of an interposed PrP-expressing compartment, later shown to be the peripheral nervous system (60). In the case of

Fig. 3. Transmission of mouse scrapie prions by stainless steel wire. Stainless steel wires were inserted into the brains of scrapie-infected mice for 5, 30, or 120 min, washed exhaustively, and introduced permanently into brains of indicator mice. Five minutes of contact sufficed for the wire to acquire a maximum load of infectivity, equivalent to the injection of 30 μl of 1% homogenate of the same brain. Data from ref. 73.

Table 1. Transient insertion of infectious wires into brains of indicator mice

Inoculation	Sick/ total	Incubation time ± SD, days
Wires infected by exposure to scrapie brain		
Transient insertion into indicator mice		
30 min	4/4*	94 ± 10
120 min	2/2†	87 ± 113
Permanent insertion into indicator mice		
Wires not previously inserted	3/3	71 ± 2
Wires after transient insertion for:		
30 min	4/4	71 ± 3
120 min	5/5	68 ± 1
Controls		
Wires exposed to brain homogenate	6/6	76 ± 3
Brain homogenate (1%, 0.03 ml)	3/3	69 ± 3

Infectious wires were prepared by insertion for 5 min into scrapie-infected mouse brain. After washing, they were inserted into brains of six deeply anesthetised *Tga20* indicator mice for the times indicated. The recovered wires were washed and implanted into *Tga20* mice. As controls, wires incubated with 10% homogenate (6.8 log LD$_{50}$ units/ml) of the same brain and the homogenate itself were introduced into indicator mice. Modified from ref. 73.
*Two of six mice died on the day of the intervention.
†Four of six mice died within a day of the intervention.

had retained infectious prions over several years and despite repeated attempts at sterilization (70, 71).

Experimental Transmission of Surface-Bound Prions

The electrode described above had a complex structure: a steel shaft of about 6 mm diameter, with multiple silver contacts separated by rings of insulating plastic allowing for the existence of crevices into which infectious material might have penetrated. To clarify whether prions would bind to a homogeneous surface, we used fine stainless steel wires as model for a surgical instrument. In a first experiment, wires were incubated overnight with brain homogenate from a terminally sick, murine-scrapie-infected mouse, washed exhaustively with PBS, and permanently implanted into brains of indicator mice. This procedure resulted in scrapie disease withi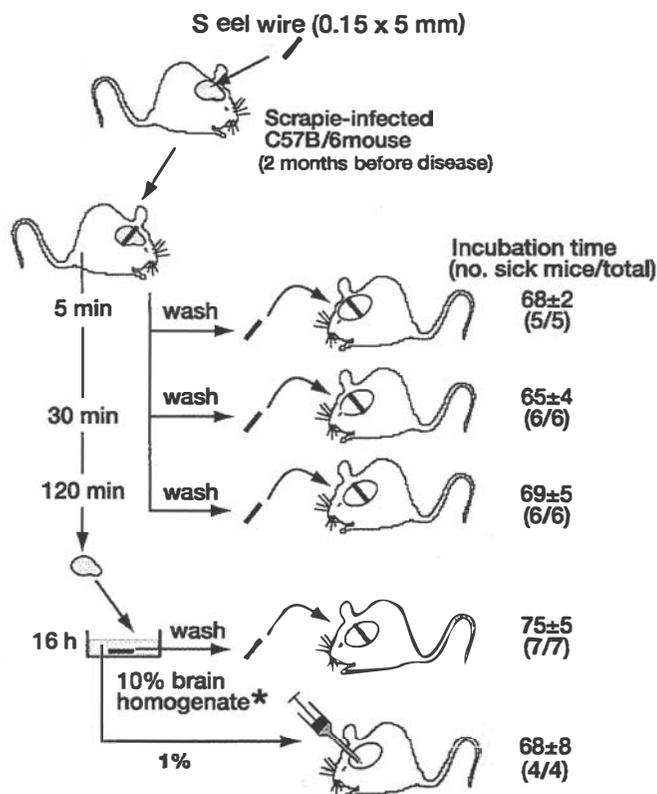n about 70 days, an incubation time only slightly longer than that obtained by injecting 30 μl of 1% brain homogenate (72). To mimic more closely real-life conditions, stainless steel wires were inserted directly into the brains of scrapie-infected, clinically still healthy mice for various periods of time, washed exhaustively, and assayed by permanent insertion into brains of indicator mice. Surprisingly, 5 min of contact sufficed for the wire to acquire a maximum load of infectivity, equivalent to the injection of 30 μl of 1% homogenate of the same brain (Fig. 3). A second important question regards the length of time an infectious wire must remain in contact with brain tissue to initiate disease. Rather than leave the infectious wires permanently in the indicator mouse, they were inserted transiently, for 30 or 120 min, to mimic the conditions that might obtain during a surgical operation. As shown in Table 1, a contact time of 30 min was sufficient to elicit disease, albeit with lower efficiency than was obtained after permanent insertion, as evidenced by the longer incubation time. The wires that had been inserted transiently into indicator mice remained fully infectious when introduced permanently into a further set of indicator mice (Table 1; ref. 73), reflecting the persistence of infectivity, as in the incident with the intracerebral electrode described above.

Why are wires exposed to infected brain or brain homogenates at least as infectious as injected homogenates, which contain far more protein than can be bound to a wire? The surfaces of steel and other metals tightly bind what appears to be a monolayer of protein (74–76). The unexpected high infectivity of steel wires

experimental mouse scrapie, prion infectivity could not be detected in leukocytes (64), nor was infectivity detected in the blood of BSE-infected cattle (50) or scrapie-infected sheep (65). However, a low but reproducible titer of prions was detected in blood of scrapie-infected hamsters (66). Also, 1 of 19 sheep transfused with blood from experimentally, orally BSE-infected sheep came down with prion disease (67). The level of prions in blood, which in all cases examined appears to be low or undetectable by the mostly not very sensitive methods used, may vary in different species and/or with different prion strains.

Iatrogenic Transmission of Prions

Almost 300 cases of involuntary transmission of CJD by medical interventions have been reported (68). Most cases are due to injection of cadaveric human growth hormone or transplantation of dura mater; however, a few incidents associated with cornea transplantation have been reported. Four instances of CJD after neurosurgical intervention have been attributed to surgical instruments that had previously been used on CJD patients (69); however, causality was proven only in one case. An electrode that had been inserted into the cortex of an unrecognized CJD patient was subjected to a decontamination procedure involving treatment with benzene, 70% ethanol, and formaldehyde vapor. It was then used in succession on two young patients and cleaned as above after each use. Within 2 yr, both patients came down with CJD. After these events, the tip of the electrode was implanted into the brain of a chimpanzee where it too caused lethal spongiform encephalopathy, proving that the electrode

Fig. 4. Infection of mouse neuroblastoma cells by plastic-bound prions. Wells of a polystyrene 96-well microtiter plate were exposed to various dilutions of a homogenate of scrapie-prion-infected mouse brain, washed exhaustively, and dried. Ten thousand N2a/Bos2 mouse neuroblastoma cells (77) were cultured in the wells for 3 days, then transferred to 24-well plates and cultured 4 wk, splitting 1:10 twice a week. The cells were then transferred to coverslips and assayed for the presence of PrP^{Sc} (78). Optimal infectivity resulted when plates were coated with 0.0125% homogenate. High concentrations of brain proteins bound to the plastic appear inhibitory for cell infection (P.-C.K. and C.W., unpublished results).

could be due to selective binding of infectious particles or a higher potency of surface bound infectivity. It has been shown that, despite the resistance of PrP^{Sc} and scrapie infectivity to treatment *in vitro* with proteinase K, prion titers in brain after intracerebral inoculation decrease below the level of detectability within 4 days or less (15). On the other hand, infectious wires left for 5 days in brain still retained infectivity (73). Perhaps metal-bound prions may be protected against rapid degradation in the brain, and their apparently high specific infectivity may therefore be due to the long persistence of relatively low levels of infectivity. It can be mentioned in passing that prion-coated gold wires exhibit similar infectivity intracerebrally as steel wires (73), and that plastic surfaces, such as polystyrene (Fig. 4), polypropylene, or polyethylene also tightly bind prions and transmit scrapie infectivity to adherent susceptible cultured cells (M.E., D.R., P.-C.K. and C.W., unpublished data).

We attempted to elute PrP from infectious steel wires with 2M NaOH, but failed to detect either protein (detection limit, 50 ng per wire) or PrP (detection limit 15 pg per wire). On the other hand, PrP immunoreactivity could be detected at the surface of prion-coated wires by chemiluminescence (73). This finding raises the question as to whether infection of brain tissue elicited by infected wires comes about by direct contact with irreversibly surface-bound prions or whether it is due to a slow, so far undetected release of prions. This question is difficult to answer experimentally; however, it would seem that intimate contact between the prion-loaded surface and target cells is a prerequisite for transmission of infectivity. Prion-coated wires were placed on monolayers of mouse neuroblastoma cells highly susceptible to mouse prions (77). After 1 to 14 days, the wires, to which some cells had adhered, were transferred onto coverslips in the wells of a tissue culture plate and incubated for 14 days, allowing the cells to migrate off the wire and multiply. Cells derived from both the residual monolayer and the wire were blotted onto nitrocellulose membranes and assayed for the presence of protease-resistant PrP, the surrogate marker of prion infection (78). Only the cells derived from the infected wire, but not from the residual monolayer, were PrP^{Sc} positive (Fig. 5) and contained infectivity (M.E., E.F. and C.W., unpublished data). This experiment shows that intimate contact between the prion-carrying surface and susceptible cells greatly promotes infection or is prerequisite. Similarly, cell-to-cell transmission of infectivity in cell culture is orders of magnitude more efficient than transmission by a prion preparation (79).
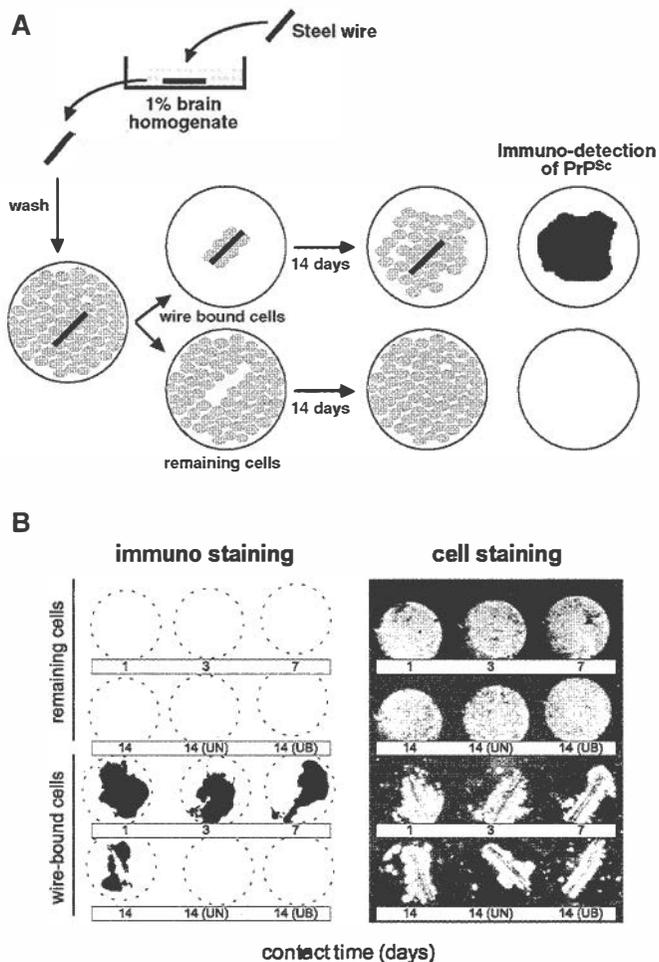
Fig. 5. Neuroblastoma cells are infected by contact with prion-coated stainless steel wires. (*A*) Stainless steel wires were exposed to scrapie-infected brain homogenates, washed, and placed on a confluent layer of neuroblastoma cells. After periods ranging from 1 to 14 days, the wire, to which a few cells had attached, were placed on a coverslip in a separate well and cultured for further 14 days. Both the cells remaining in the original dish ("remaining cells") and those derived from the cells clinging to the wire were assayed for PrP^{Sc} by the cell blot assay (78) and the mouse bioassay. (*B Left*) The cultures derived from wire-bound cells have been infected, as evidenced by the accumulation of PrP^{Sc}, whereas residual cells remain uninfected. (*Right*) The location of cells as revealed by ethidium bromide staining. UN, Blank wire; UB, wire treated with uninfected brain homogenate (M.E. and C.W., unpublished results).

The availability of prion-coated steel wires mimicking contaminated surgical instruments makes it possible to assess the efficacy of sterilization conditions on surface-bound prions. Preliminary results (Table 2) confirm that treatment with formaldehyde is insufficient to sterilize infectious wires, whereas treatment with sodium hydroxide, guanidinium thiocyanate (73), or autoclaving at 121°C for 20 min is efficacious (E.F. and C.W., unpublished results). It is, however, not appropriate to derive from these experiments recommendations for the sterilization of surgical instruments; it will first be necessary to validate the procedures scaling up the contact surface between metal and brain tissue and, importantly, using vCJD prions in a susceptible host, preferably a non-human primate.

Concluding Remarks

Twenty or more diseases of humans are associated with the deposition of β-sheet-rich protein aggregates, or amyloid (80,

Table 2. Effect of various treatments on the infectivity of wire-bound prions

	Sick/total	Incubation time ± SD, days
Uninfected wires		
Untreated	0/3	>260
Infected wires		
Untreated	6/6	76 ± 5
NaOH (1 M, 1 h, 25°C)	0/6	>260
Formaldehyde (10%, 1 h, 25°C)	6/6	92 ± 8
Guanidinium thiocyanate (4 M, 16 h, 25°C)	0/6	>260
Autoclaving (121°C, 20 min)	0/6	>170

Modified from ref. 73 and unpublished results (E.F. and C.W.).

81). They are frequently designated "conformational diseases" although it is not in all cases clear whether, or to what extent, the misfolded proteins are the cause of the disease rather than the consequence. Prion diseases are so far unique conformational diseases, because they are transmissible by misfolded protein, not only under experimental conditions but also naturally, predominantly by ingestion. Although in certain cases the inception of an experimental amyloidosis can be accelerated by the injection of amyloid into a predisposed host (82), prions are exceptional in that they are able to enter their hosts by natural portals and make their way from the gut to the brain, utilizing intermediate tissues for amplification. In the case of microbes and viruses, such sophisticated behavior is attributed to evolutionary processes, that is, genomic mutations and selection of mutants that most readily enter their host and find a suitable niche in which to replicate and/or perpetuate themselves; however, prion protein is encoded by the genome of its host. So what drives the prion to become more efficient in the destruction of its parent? We can only speculate. For example, the "misfolded" form of PrP may have originated as a "messenger" protein that on the one hand has or had a physiological function but on the other has a malignant potential that is rarely realized and was not selected against because evolutionary pressure does not operate efficiently at postreproductive age. It has been proposed that in yeast a "prion-like" phenomenon involving Sup35 may confer selective advantage on yeast growing under fluctuating environmental conditions (83). Another possibility is that PrP/PrP^{Sc} is derived from an ancient pathogen whose genetic material was integrated into the genome of its host and harnessed to fulfill a useful function while its pathogenic potential was minimized. More trivially, mammalian prion disease could be the result of the natural propensity of proteins to assume a β-sheet-rich conformation (84), a failure of the organism to prevent their formation and accumulation in some cases, and the coincidental ability of the conformational isomer to penetrate organisms and cells through natural portals.

1. Will, R. G., Ironside, J. W., Zeidler, M., Cousens, S. N., Estibeiro, K., Alperovitch, A., Poser, S., Pocchiari, M., Hofman, A. & Smith, P. G. (1996) *Lancet* **347**, 921–925.
2. Cuille, J. & Chelle, P. L. (1939) *C. R. Seances Acad. Sci.* **208**, 1058–1160.
3. Gordon, W. S. (1946) *Vet. Rec.* **58**, 516–520.
4. Alper, T., Cramp, W. A., Haig, D. A. & Clarke, M. C. (1967) *Nature (London)* **214**, 764–766.
5. Griffith, J. S. (1967) *Nature (London)* **215**, 1043–1044.
6. Prusiner, S. B. (1989) *Annu. Rev. Microbiol.* **43**, 345–374.
7. Basler, K., Oesch, B., Scott, M., Westaway, D., Walchli, M., Groth, D. F., McKinley, M. P., Prusiner, S. B. & Weissmann, C. (1986) *Cell* **46**, 417–428.
8. Oesch, B., Westaway, D., Wälchli, M., McKinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. B., Hood, L. E., Prusiner, S. B. & Weissmann, C. (1985) *Cell* **40**, 735–746.
9. Stahl, N., Baldwin, M. A., Teplow, D. B., Hood, L., Gibson, B. W., Burlingame, A. L. & Prusiner, S. B. (1993) *Biochemistry* **32**, 1991–2002.
10. Riek, R., Hornemann, S., Wider, G., Glockshuber, R. & Wüthrich, K. (1997) *FEBS Lett.* **413**, 282–288.
11. Pan, K. M., Baldwin, M., Nguyen, J., Gasset, M., Serban, A., Groth, D., Mehlhorn, I., Huang, Z., Fletterick, R. J., Cohen, F. E. & Prusiner, S. B. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10962–10966.
12. Caughey, B. W., Dong, A., Bhat, K. S., Ernst, D., Hayes, S. F. & Caughey, W. S. (1991) *Biochemistry* **30**, 7672–7680.
13. Weissmann, C. (1999) *J. Biol. Chem.* **274**, 3–6.
14. Prusiner, S. B. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13363–13383.
15. Büeler, H., Aguzzi, A., Sailer, A., Greiner, R. A., Autenried, P., Aguet, M. & Weissmann, C. (1993) *Cell* **73**, 1339–1347.
16. Sailer, A., Büeler, H., Fischer, M., Aguzzi, A. & Weissmann, C. (1994) *Cell* **77**, 967–968.
17. Fischer, M., Rülicke, T., Raeber, A., Sailer, A., Moser, M., Oesch, B., Brandner, S., Aguzzi, A. & Weissmann, C. (1996) *EMBO J.* **15**, 1255–1264.
18. Prusiner, S. B. (1991) *Science* **252**, 1515–1522.
19. Jarrett, J. T. & Lansbury, P. J. (1993) *Cell* **73**, 1055–1058.
20. Orgel, L. E. (1996) *Chem. Biol.* **3**, 413–414.
21. Kocisko, D. A., Come, J. H., Priola, S. A., Chesebro, B., Raymond, G. J., Lansbury, P. T. & Caughey, B. (1994) *Nature (London)* **370**, 471–474.
22. Raymond, G. J., Hope, J., Kocisko, D. A., Priola, S. A., Raymond, L. D., Bossers, A., Ironside, J., Will, R. G., Chen, S. G., Petersen, R. B., *et al.* (1997) *Nature (London)* **388**, 285–288.
23. Bessen, R. A., Kocisko, D. A., Raymond, G. J., Nandan, S., Lansbury, P. T. & Caughey, B. (1995) *Nature (London)* **375**, 698–700.
24. Saborio, G. P., Permanne, B. & Soto, C. (2001) *Nature (London)* **411**, 810–813.
25. Lu, B. Y. & Chang, J. Y. (2001) *Biochemistry* **40**, 13390–13396.
26. Jackson, G. S., Hosszu, L. L., Power, A., Hill, A. F., Kenney, J., Saibil, H., Craven, C. J., Waltho, J. P., Clarke, A. R. & Collinge, J. (1999) *Science* **283**, 1935–1937.
27. Hill, A. F., Antoniou, M. & Collinge, J. (1999) *J. Gen. Virol.* **80**, 11–14.
28. Shaked, G. M., Fridlander, G., Meiner, Z., Taraboulos, A. & Gabizon, R. (1999) *J. Biol. Chem.* **274**, 17981–17986.
29. Kaneko, K., Ball, H. L., Wille, H., Zhang, H., Groth, D., Torchia, M., Tremblay, P., Safar, J., Prusiner, S. B., DeArmond, S. J., *et al.* (2000) *J. Mol. Biol.* **295**, 997–1007.
30. Post, K., Brown, D. R., Groschup, M., Kretzschmar, H. A. & Riesner, D. (2000) *Arch. Virol.*, Suppl., 265–273.
31. Prusiner, S. B., Groth, D., Serban, A., Stahl, N. & Gabizon, R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2793–2797.
32. Shaked, G. M., Meiner, Z., Avraham, I., Taraboulos, A. & Gabizon, R. (2001) *J. Biol. Chem.* **276**, 14324–14328.
33. McKenzie, D., Bartz, J., Mirwald, J., Olander, D., Marsh, R. & Aiken, J. (1998) *J. Biol. Chem.* **273**, 25545–25547.
34. Kimberlin, R. H. & Wilesmith, J. W. (1994) *Ann. N. Y. Acad. Sci.* **724**, 210–220.
35. Alpers, M. (1979) in *Slow Transmissible Diseases of the Nervous System*, eds. Prusiner, S. B. & Hadlow, W. J. (Academic, New York), Vol. 1, p. 67.
36. Maignien, T., Lasmezas, C. I., Beringue, V., Dormont, D. & Deslys, J. P. (1999) *J. Gen. Virol.* **80**, 3035–3042.
37. Jeffrey, M., Ryder, S., Martin, S., Hawkins, S. A., Terry, L., Berthelin-Baker, C. & Bellworthy, S. J. (2001) *J. Comp. Pathol.* **124**, 280–289.
38. Bradley, R. (1996) in *Bovine Spongiform Encephalopathy: The BSE Dilemma*, ed. Gibbs, C. J., Jr. (Springer, New York), p. 11.
39. Bons, N., Mestre-Frances, N., Belli, P., Cathala, F., Gajdusek, D. C. & Brown, P. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4046–4051.
40. Ridley, R. M. & Baker, H. F. (1996) *Lancet* **348**, 1174 (lett.).
41. Race, R., Jenny, A. & Sutton, D. (1998) *J. Infect. Dis.* **178**, 949–953.
42. Ingrosso, L., Pisani, F. & Pocchiari, M. (1999) *J. Gen. Virol.* **80**, 3043–3047.
43. Bolton, D. C., McKinley, M. P. & Prusiner, S. B. (1982) *Science* **218**, 1309–1311.
44. Sansonetti, P. J. & Phalipon, A. (1999) *Semin. Immunol.* **11**, 193–203.
45. Kaiserlian, D. & Etchart, N. (1999) *Semin. Immunol.* **11**, 217–224.
46. Hathaway, L. J. & Kraehenbuhl, J. P. (2000) *Cell. Mol. Life Sci.* **57**, 323–332.
47. Heppner, F. L., Christ, A. D., Klein, M. A., Prinz, M., Fried, M., Kraehenbuhl, J. P. & Aguzzi, A. (2001) *Nat. Med.* **7**, 976–977.
48. Beekes, M. & McBride, P. A. (2000) *Neurosci. Lett.* **278**, 181–184.
49. Prinz, M., Montrasio, F., Klein, M. A., Schwarz, P., Priller, J., Odermatt, B., Pfeffer, K. & Aguzzi, A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 919–924.
50. Bradley, R. (1999) *Dev. Biol. Stand.* **99**, 35–40.
51. Aucouturier, P., Geissmann, F., Damotte, D., Saborio, G. P., Meeker, H. C., Kascsak, R., Carp, R. I. & Wisniewski, T. (2001) *J. Clin. Invest.* **108**, 703–708.
52. Huang, F. P., Farquhar, C. F., Mabbott, N. A., Bruce, M. E. & MacPherson, G. G. (2002) *J. Gen. Virol.* **83**, 267–271.
53. Klein, M. A., Frigg, R., Raeber, A. J., Flechsig, E., Hegyi, I., Zinkernagel, R. M., Weissmann, C. & Aguzzi, A. (1998) *Nat. Med.* **4**, 1429–1433.

54. Montrasio, F., Cozzio, A., Flechsig, E., Rossi, D., Klein, M. A., Rulicke, T., Raeber, A. J., Vosshenrich, C. A., Proft, J., Aguzzi, A. & Weissmann, C. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4034–4037.

55. Montrasio, F., Frigg, R., Glatzel, M., Klein, M. A., Mackay, F., Aguzzi, A. & Weissmann, C. (2000) *Science* **288,** 1257–1259.

56. Mabbott, N. A., Mackay, F., Minns, F. & Bruce, M. E. (2000) *Nat. Med.* **6,** 719–720.

57. Oldstone, M. B., Race, R., Thomas, D., Lewicki, H., Homann, D., Smelt, S., Holz, A., Koni, P., Lo, D., Chesebro, B. & Flavell, R. (2002) *J. Virol.* **76,** 4357–4363.

58. Beekes, M., McBride, P. A. & Baldauf, E. (1998) *J. Gen. Virol.* **79,** 601–607.

59. Bencsik, A., Lezmi, S., Hunsmann, G. & Baron, T. (2001) *Dev. Immunol.* **8,** 235–241.

60. Race, R., Oldstone, M. & Chesebro, B. (2000) *J. Virol* **74,** 828–833.

61. Bosque, P. J., Ryou, C., Telling, G., Peretz, D., Legname, G., DeArmond, S. J. & Prusiner, S. B. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 3812–3817.

62. Brandner, S., Raeber, A., Sailer, A., Blattler, T., Fischer, M., Weissmann, C. & Aguzzi, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13148–13151.

63. Blättler, T., Brandner, S., Raeber, A. J., Klein, M. A., Voigtlander, T., Weissmann, C. & Aguzzi, A. (1997) *Nature (London)* **389,** 69–73.

64. Raeber, A. J., Klein, M. A., Frigg, R., Flechsig, E., Aguzzi, A. & Weissmann, C. (1999) *EMBO J.* **18,** 2702–2706.

65. Hadlow, W. J., Kennedy, R. C. & Race, R. E. (1982) *J. Infect. Dis.* **146,** 657–664.

66. Holada, K., Vostal, J. G., Theisen, P. W., MacAuley, C., Gregori, L. & Rohwer, R. G. (2002) *J. Virol.* **76,** 4649–4650.

67. Houston, F., Foster, J. D., Chong, A., Hunter, N. & Bostock, C. J. (2000) *Lancet* **356,** 999–1000.

68. Brown, P., Preece, M., Brandel, J. P., Sato, T., McShane, L., Zerr, I., Fletcher, A., Will, R. G., Pocchiari, M., Cashman, N. R., *et al.* (2000) *Neurology* **55,** 1075–1081.

69. Brown, P. (1996) in *Prion Diseases*, eds. Baker, H. & Ridley, R. M. (Humana, Totowa, NJ), pp. 139–154.

70. Bernoulli, C., Siegfried, J., Baumgartner, G., Regli, F., Rabinowicz, T., Gajdusek, D. C. & Gibbs, C. J., Jr. (1977) *Lancet* **1,** 478–479.

71. Gibbs, C. J., Jr., Asher, D. M., Kobrine, A., Amyx, H. L., Sulima, M. P. & Gajdusek, D. C. (1994) *J. Neurol. Neurosurg. Psychiatry* **57,** 757–758.

72. Zobeley, E., Flechsig, E., Cozzio, A., Enari, M. & Weissmann, C. (1999) *Mol. Med.* **5,** 240–243.

73. Flechsig, E., Hegyi, I., Enari, M., Schwarz, P., Collinge, J. & Weissmann, C. (2001) *Mol. Med.* **7,** 679–684.

74. Williams, D. F., Askill, I. N. & Smith, R. (1985) *J. Biomed. Mater. Res.* **19,** 313–320.

75. Williams, R. L. & Williams, D. F. (1988) *Biomaterials* **9,** 206–212.

76. Eckert, R., Jeney, S. & Horber, J. K. (1997) *Cell Biol. Int.* **21,** 707–713.

77. Enari, M., Flechsig, E. & Weissmann, C. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 9295–9299.

78. Bosque, P. J. & Prusiner, S. B. (2000) *J. Virol.* **74,** 4377–4386.

79. Kanu, N., Imokawa, Y., Drechsel, D. N., Williamson, R. A., Birkett, C. R., Bostock, C. J. & Brockes, J. P. (2002) *Curr. Biol.* **12,** 523–530.

80. Carrell, R. W. & Lomas, D. A. (1997) *Lancet* **350,** 134–138.

81. Lansbury, P. T., Jr. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3342–3344.

82. Johan, K., Westermark, G., Engstrom, U., Gustavsson, A., Hultman, P. & Westermark, P. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2558–2563.

83. True, H. L. & Lindquist, S. L. (2000) *Nature (London)* **407,** 477–483.

84. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3590–3594.

Colloquium

Conservation of a portion of the *S. cerevisiae* Ure2p prion domain that interacts with the full-length protein

Herman K. Edskes and Reed B. Wickner*

Laboratory of Biochemistry and Genetics, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0830

The [URE3] prion of *Saccharomyces cerevisiae* is a self-propagating inactive amyloid form of the Ure2 protein. Ure2p residues 1–65 constitute the prion domain, and the remaining C-terminal portion regulates nitrogen catabolism. We have examined the *URE2* genes of wild-type isolates of *S. cerevisiae* and those of several pathogenic yeasts and a filamentous fungus. We find that the normal function of the *S. cerevisiae* Ure2p in nitrogen regulation is fully complemented by the Ure2p of *Candida albicans*, *Candida glabrata*, *Candida kefyr*, *Candida maltosa*, *Saccharomyces bayanus*, and *Saccharomyces paradoxus*, all of which have high homology in the C-terminal nitrogen regulation domain. However, there is considerable divergence of their N-terminal domains from that of Ure2p of *S. cerevisiae*. [URE3^{Sc}] showed efficient transmission into *S. cerevisiae* ure2Δ cells if expressing a Ure2p of species within *Saccharomyces*. However, [URE3^{Sc}] did not seed self-propagating inactivation of the Ure2p's from the other yeasts. When overexpressed as a fusion with green fluorescent protein, residues 5–47 of the *S. cerevisiae* prion domain are necessary for curing the [URE3] prion. Residues 11–39 are necessary for an inactivating interaction with the full-length Ure2p. A nearly identical region is highly conserved among many of the yeasts examined in this study, despite the wide divergence of sequences found in other parts of the N-terminal domains.

T he word "prion" means "infectious protein." Considerable evidence supports a prion basis for the transmissible spongiform encephalopathies (TSEs) of mammals, with an amyloid form of the PrP protein as the culprit (1). The gene for PrP controls the clinical and pathological features of the TSEs (2–7). The scrapie agent is far more radiation resistant than even small genome viruses (8), and purification of the infectious agent purifies an amyloid form of PrP (9). PrP is clearly necessary for and central to infectivity, but showing that it is sufficient has been difficult.

[URE3] (10) and [PSI+] (11) are nonchromosomal genes of *S. cerevisiae* whose molecular basis was long obscure. Genetic evidence first identified [URE3] and [PSI+] as prions of the yeast Ure2p and Sup35p, respectively (12). Three criteria distinguishing prions from viruses and plasmids are (*i*) after curing a prion, it can arise again *de novo* in the cured strain, (*ii*) overexpression of the protein increases the frequency with which the prion arises *de novo*, and (*iii*) the prion's propagation depends on the chromosomal gene encoding the protein, but the presence of the prion has a similar phenotype to recessive mutation of the chromosomal gene (12). Both [URE3] and [PSI+] satisfy all three criteria as prions of Ure2p and Sup35p, respectively (reviewed in refs. 13–16)

Ure2p is a regulator of nitrogen catabolism, acting by binding to the Gln-3 GATA transcription factor and thereby keeping the latter in the cytoplasm when the medium contains a rich nitrogen source such as NH_3 or glutamine (17–22). This prevents the transcription of genes, such as *DAL5*, encoding enzymes or transporters needed to use poor nitrogen sources (23–25). The N-terminal 65–90

residues of Ure2p largely determine the prion properties of the 354 residue protein (26, 27), whereas the C-terminal 261 residues are sufficient for nitrogen regulation (18, 26). Overexpression of the prion domain is sufficient to induce the *de novo* appearance of [URE3] at rates far higher than even the elevated rates observed on overexpression of the full-length Ure2p protein (26, 27). A similar phenomenon has been observed for the Sup35p prion domain and induction of [PSI+] appearance (28). Moreover, expression of just the prion domain is sufficient to maintain [URE3] (29).

The first biochemical evidence for the yeast prions and hint of their molecular basis was the observation that Ure2p is protease-resistant in extracts of [URE3] cells, but not in extracts of normal strains (26). The similarity of this finding to the protease resistance of PrP in scrapie brains (9) was, of course, highly suggestive. Moreover, it is the prion domain of Ure2p that forms the protease-resistant core of the prion form (26, 30, 31). Ure2p is aggregated *in vivo* specifically in [URE3] strains, and this aggregation requires the prion domain (32).

Ure2p purified from yeast is a stably soluble dimer (30, 33), but the synthetic prion domain, Ure2p1–65, spontaneously and rapidly forms amyloid fibers *in vitro* (30). Moreover, in the same way that expression of the prion domain induces the *de novo* appearance of the [URE3] prion *in vivo*, its presence *in vitro* induces the full-length Ure2p to form amyloid filaments containing both the prion domain fragment and the full-length molecule (30). The self-propagation of this reaction, the specificity for the Ure2p prion domain, and the similarity of the protease-resistance patterns of this *in vitro* amyloid to that of Ure2p in [URE3] cells strongly support the concept that [URE3] is a self-propagating inactive amyloid form of Ure2p.

Ure2p filaments have been directly observed *in vivo* specifically in [URE3] cells (31). These filaments were observed in large networks, localized to limited areas of the cytoplasm, with generally only a single network observed in a single cell section. Other areas of the cytoplasm were depleted of Ure2p (31).

Several lines of evidence suggest a structure for the Ure2p amyloid in which the prion domain forms a central β-sheet-rich core surrounded by the appended functional domain. Protease digestion of the 400 nm wide amyloid filaments formed by full-length Ure2p leaves narrow filaments morphologically similar to those formed by the prion domain alone and composed of the prion domain (30). Aggregated Ure2p from [URE3] cells shows little reaction with

antibody to the prion domain, but reacts well with anti-Ure2C (31). The same is true of Ure2p filaments seen by electron microscopy specifically in [URE3] cells (31). This observation suggests that, in this structure, the prion domain is inside and the C-terminal domain is outside.

The Ure2p prion domain was fused to the N terminus of several enzymes with small substrates (34), including glutatione *S*-transferase (GST), which is similar to the C-terminal domain of Ure2p (18, 35, 36). Amyloid formed in each case, and the enzymatic activity of the fusion proteins were essentially the same in the amyloid form as in the soluble form, when suitable correction was made for diffusion effects (34). These results suggested that amyloid formation does not inactivate Ure2p function by changing the conformation of the C-terminal domain, but rather by either sterically blocking interaction of Ure2p with Gln3p or by Ure2p being diffusion-limited in its filament form (34). Several of the fusion proteins formed monofilaments with a helical form. The helical repeat length was consistent within a given filament, but, remarkably, varied dramatically from one filament to another, even though the filaments were composed of the same fusion protein (34). This finding indicates that the geometry of Ure2p filament formation is determined by some stochastic process that occurs during filament initiation.

The epidemic of bovine spongiform encephalopathy in the United Kingdom, followed by over 100 human TSE cases caused by consumption of infected bovine material has highlighted the importance of cross-species transmission of the mammalian TSEs. A similar phenomenon has been demonstrated in variants of [PSI+] in which the region of Sup35 of various yeasts corresponding to the prion domain of Sup35Sc was fused to the *S. cerevisiae* C-terminal domain of Sup35 (37–39). These studies demonstrated that the N termini of the Sup35 proteins of *Pichia methanolica* and *Candida albicans* can act as prion domains. Moreover, the divergence of these prion domains correlated with a "species barrier," much like that long documented for the mammalian TSEs.

We previously showed that overexpression of fragments of Ure2p or fusions of such fragments with green fluorescent protein (GFP) could efficiently cure the [URE3] prion (32). Here we have defined the portions of Ure2N and Ure2C required for this curing activity. As one approach to the functional significance of the Ure2N curing region, we examined homologs of Ure2p and found strong conservation of this part of the otherwise rapidly evolving N-terminal domain.

Materials and Methods

Yeast Strains and Media. Media were as described by Sherman (40). The ureidosuccinate (USA) uptake phenotype of *ura2* strains was tested on synthetic dextrose plates to which was added the required amino acids and bases (except uracil) and 100 μg/ml of USA.

S. cerevisiae "wild-type" cultures were obtained from three sources. Some were locally purchased; these include cultures sold for making bread (SAF Perfect Rise yeast, Lesaffre, Bruxelles, Belgium; Fleischmann Active Dry yeast, www.breadworld.com; Peter McPhie's Sour Dough strain, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda) as well as cultures sold for home beer and wine making (Red Star Dry Wine Yeast, Premier cuvee; Wyeast catalog no. 1007, www.wyeastlab.com, German Ale; Wyeast catalog no. 2112xL, California Lager; White Labs catalog no. WLP002, www.whitelabs.com, English Ale yeast; Boots home beer making genuine brewers yeast). Yeast cultures from Centraalbureau voor Schimmelcultures (CBS, www.cbs.knaw.nl) in the Netherlands (CBS400, palm wine from *Elaies guineensis*, Ivory Coast; CBS405, bili wine from *Osbeckia grandiflora*, West Africa; CBS429, fermenting must of champagne grapes; CBS2087 flower of lychee, Tonkin, China; CBS2247, grape must, Cape Province, South Africa; CBS3093, alpechin, Spain; CBS4734, from juice of sugar cane; CBS5112, grape must, Spain; CBS5287, grapes, Russia; CBS6216,

tap water, Rotterdam, The Netherlands; CBS7957, cassava flour, Sao Paulo, Brazil). Clinical isolates of *S. cerevisiae* (41) were kindly provided by J. McCusker (Duke University Medical Center, Durham, NC). YJM145 is a segregant from YJM128 which was cultured from the lung of a man with immune deficiency syndrome. YJM413 is a segregant from clinical isolate YJM454. YJM280 is a segregant from YJM273, which was cultured from peritoneal fluid of a patient. YJM320 is a segregant from YJM309, which was cultured from the blood of a patient. YJM326 is a segregant from clinical isolate YJM310. YJM339 is a segregant from YJM311, which was cultured from the bile tube of a patient.

Saccharomyces bayanus (YJM562) and *Saccharomyces paradoxus* (YJM498) were kindly provided by J. McCusker. *Candida glabrata* (37A; ref. 42), *Candida kefyr* (telemorph is *Kluyveromyces marxianus*) (B4425) (43), *Candida maltosa* (B4430) (44), *C. albicans* (Darlington strain) (45), and *Candida lipolytica* (telemorph is *Yarrowia lipolytica*) (B3163) (46) were kindly provided by K. J. Kwon-Chung (National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda). *Ashbya gossypii* (47) was purchased from American Type Culture Collection (catalog no. 8717).

Plasmid Constructions. PCR used *Pfu* Turbo polymerase (Stratagene) unless otherwise stated.
Construction of yeast expression plasmids. pH7 (2μ *LEU2* P$_{ADH1}$) and pH317 (2μ*LEU2* P$_{GAL1}$) have been described (32, 48). pH199, a 2μ *LEU2* plasmid containing GFP under control of the *ADH1* promoter, has also been described (32).

To create pH722 (*LEU2 CEN* P$_{URE2}$), first the *NheI–Bam*HI bordered *ADH1* promoter of pH124 (32) was replaced by the similarly bordered *GAL1* promoter from pH250 (48), creating pH316 (*LEU2 CEN* P$_{GAL1}$). Then, a 413-bp *URE2* promoter fragment bordered by *NheI* and *Bam*HI sites was amplified by PCR from *S. cerevisisae* strain S288C using oligos HE194 (5′-AGAGCTAGCTTAGTAGAGCTGTGTAGAG-3′) and HE195b (5′-TTGGGATCCAACTTAATTTGCAGCTTAAAAC-3′) and cloned into the *Eco*RV site of pBC KS+ creating pH497. Replacing the *NheI–Bam*HI bordered *GAL1* promoter of pH316 with the similarly bordered *URE2* promoter from pH497 resulted in pH722. Likewise, replacing the *NheI–Bam*HI bordered *GAL1* promoter of pH317 with the *URE2* promoter from pH497 resulted in pH723 (*LEU2* 2μ P$_{URE2}$).

The *Hind*III and *Xba*I sites were removed from the *TRP1* gene in the 2μ *TRP1* vector pRS424 (49) by site-directed mutagenesis using oligos HE128 (5′-AAGAGAGCCCCGAAAGTTTA-CATTTTATGTTAGCTG-3′) and HE129 (5′-GGCCGCA-GAATGTGCTCTTGATTCCGATGCTGACTTG-3′), respectively, resulting in plasmid pH342. The *ADH1* promoter was amplified from pH7 by using oligos HE66 (5′-AGAGCTAGCAT-TACGCCAGCAACTTCT-3′) and HE67 (5′-ACAAGATCTTA-ATGCAGCCGGTAGAG-3′) and ligated into *Pvu*II digested pH342 creating pH401 (*TRP1* 2μ P$_{ADH1}$). The *TRP1* and *ADH1* promoters are facing each other in this construct.
Truncations of the Ure2p C terminus fused to GFP. pH328 contains the C-terminal part of *URE2* starting at Asp-66 (32). Further N-terminal deletions were made by amplifying *URE2*–GFP fusions from pH326 (32). PCR products were cloned into the *Eco*RV site of pBC KS+ (Stratagene), sequenced, and inserted as *Bam*HI–*Xho*I fragments into the *Bam*HI–*Xho*I window of pH7.

Deletions in the C-terminal fragment of the *URE2*–GFP fusion proteins were created through amplification of *URE2* fragments from pH13 (32), cloned into the *Eco*RV site of pBC KS+ and sequenced. The truncated fragments bordered by *Bam*HI and *Not*I sites were fused to GFP through exchange with the *Bam*HI–*Not*I fragment from pVTG4 (32).
Plasmids expressing truncations of the Ure2p N terminus fused to GFP. pVTG4 containing a part of *URE2* terminating at Arg-65 has been described (32). Further N-terminal deletions were made by ampli-

Table 1. Curing and complementation by Ure2C-GFP fusion proteins

Plasmid	URE2 part*	GFP signal	Curing,[†] USA−/Total	Complementation of ure2Δ
pH328	1–2, 66–354–GFP	+	40/40	+
pH409	1, 86–354–GFP	+	40/40	+
pH410	1, 106–354–GFP	+	40/40	−
pH550	1, 111–354–GFP	+	40/100	−
pH495	1, 116–354–GFP	+/−	0/100	−
pH411	1, 126–354–GFP	−	0/40	−
pH445	1–2, 66–293–GFP	+	0/40	−
pH444	1–2, 66–313–GFP	+	0/40	−
pH443	1–2, 66–333–GFP	+	0/40	−
pH494	1–2, 66–344–GFP	−	0/100	−
pH760	1–2, 66–347–GFP	+	98/100	−
pH549	1–2, 66–349–GFP	+	100/100	−
pH199	–GFP	+		−

*The portions of Ure2p are shown as residue numbers. After the Ure2p portion is the sequence GGR followed by GFP.

[†]Curing was assayed in strain YHE64 (*MATα ura2 leu2 trp1* [URE3]) as USA− transformants/total. USA complementation was assayed in strain YHE887 (*MATα leu2 ura2 ure2*).

fying *URE2*–GFP fusions from pVTG4. PCR products were cloned into the *Eco*RV site of pBC KS+, sequenced, and transferred as *Bam*HI–*Xho*I fragments into the *Bam*HI–*Xho*I window of pH7.

C-terminal truncations were made by amplifying the *ADH1* promoter and parts of *URE2* from pVTG4. PCR products were cloned into the *Eco*RV site of pBC KS+ and sequenced. The truncated *URE2* fragments bordered by *Bam*HI and *Not*I sites were fused to GFP through exchange with the *Bam*HI–*Not*I fragment from pVTG4. In pH767, S33 is encoded by AGC instead of AGT, and in pH548, T41 is encoded by ACT instead of ACA.

The $URE2N^{1-45-SGR}$-GFP fragment from pH547 was transferred as a *Bam*HI–*Xho*I fragment into the *Bam*HI–*Xho*I window of pH401 creating pH792.

Cloning *URE2* from Different *S. cerevisiae* Strains. Yeast strains were grown on yeast extract peptone adenine dextrose (YPAD) to single colonies. Colonies were resuspended in 50 μl H_2O containing 3 mg/ml zymolyase and incubated at 37°C for 30 min. One microliter of this suspension was used to amplify the *URE2* gene with oligos HE252 (5′-CTGCAAATTAAGTTGTACACC-3′) and HE253 (5′-TTCCTCCTTCTTCTTTCTTTC-3′). PCR products were cloned into *Eco*RV digested pBC KS+ and sequenced.

Cloning *URE2* Homologs from Different Fungi. All yeasts were grown in liquid YPAD, harvested, and genomic DNA was extracted as described (50). *A. gossypii* was grown on solid YPAD, and genomic DNA was extracted as described (51). Degenerate PCR primers were designed based on the alignment of *S. cerevisiae URE2* and *C. albicans URE2* (strain SC5314 of *C. albicans*; http://www-sequence.stanford.edu/group/candida/). Two degenerate sense strand primers, HE207 (5′-CCIAAYGGITTYAARGTIG-CIATH-3′; Y = C + T; R = A + G; H = A + C + T) and HE208 (5′-GGICAYGCICCIATGATHGGICAR-3′), and three degenerate antisense strand primers, HE209 (5′-RTAI-GCIGCIGCGTTYTCIGTRTC-3′), HE210 (5′-RTCIACIACRT-TRTTCCAIGGIAC-3′), and HE211 (5′-CATCATRTGYT-TIGTCCAYTTRTA-3′), were used to amplify *URE2* homologs by using PCR supermix (GIBCO/BRL). PCR products were cloned into pCR2.1/TOPO (Invitrogen) and sequenced. Based on these sequences, nested organism-specific inverse primer sets were designed. Genomic DNA was digested with one of a number of restriction endonucleases and the generated fragments were self ligated. The 5′ and 3′ regions of the *URE2* homologs were amplified by using PCR supermix and cloned into pCR2.1/TOPO. To ensure

that full-length *URE2* ORFs were identified, PCR primers were designed that hybridized to the 5′ and 3′ untranslated regions. After amplification of *URE2*, the PCR products were cloned into *Eco*RV-digested pBC KS+ and sequenced. No PCR product could be obtained that contained the 5′ untranslated region of the *URE2* gene of *S. paradoxus*. A PCR product could be obtained when genomic DNA of *S. paradoxus* was used in a PCR reaction with the *S. bayanus*-specific 5′ oligo. Finally, the *URE2* homologs were amplified by PCR using oligos that created a *Bam*HI site followed by the nucleotides CAA upstream of the start AUG and a *Xho*I site immediately downstream of the stop codon. *C. lipolytica* was amplified by using a PCR primer that created a *Hin*dIII site immediately downstream of the stop codon of *URE2* as it contains an internal *Xho*I. PCR products were cloned into the *Eco*RV site of pBC KS+ (Stratagene) and sequenced. The different *URE2* ORFs were cloned as *Bam*HI–*Xho*I fragments into the different expression vectors. Only *C. lipolytica URE2* was cloned as a *Bam*HI–*Hin*dIII fragment into the different expression vectors. If *URE2* was expressed under the control of the *S. cerevisiae URE2* promoter, the ORFs were first cloned into expression vectors containing the *GAL1* promoter. Subsequently, the *Nhe*I–*Bam*HI bordered *GAL1* promoter was replaced by the similarly bordered *URE2* promoter.

Results

Interaction Domains of Ure2p Based on Curing. We previously showed that overexpression of parts of Ure2p or of their fusions with GFP led to efficient curing of the [URE3] prion (32). Ure2C (residues 66–354) fused to GFP could both complement the nitrogen regulation function of Ure2p and cure. We now find that the N terminus of the curing region of Ure2C (fused to GFP) is between residues 111 and 116, whereas its C terminus is between residues 333 and 349 (Table 1). The amounts of fusion protein expressed from the various constructs was checked by the level of green fluorescence (Table 1). Comparable amounts were expressed from most constructs, but some constructs showed decreased expression, making the N terminus of the region needed for curing ambiguous. The minimal portion of the fusion proteins for complementing a ure2Δ mutation is residues 86–354.

Although overexpression of the N-terminal domain of Ure2p induces [URE3] prion formation, this same domain, when overexpressed in a [URE3] strain, cures cells of the prion (32). Curing also takes place when the Ure2p N terminus fused to GFP is overexpressed in a [URE3] strain (32). By making N- and C-terminal deletions in the Ure2 domain of these GFP fusion proteins (Table

Plasmid	URE2 part	GFP signal*	Curing,† USA−/total	Interference,‡ USA+/10⁶
pH199	−GFP	+ cytopl.	15/718	13
pVTG4	M1, **M2**–R65–GFP	+	439/440	68,000
pH545	M1, **N3**–R65–GFP	+	200/200	93,000
pH486	M1, **N4**–R65–GFP	+	100/100	73,000
pH762	M1, **N5**–R65–GFP	+	100/100	22,000
pH487	M1, **G6**–R65–GFP	+	9/300	11,000
pH408	M1, **N7**–R65–GFP	+	229/340	109,000
pH763	M1, **V9**–R65–GFP	+	197/200	35,000
pH764	M1, **N11**–R65–GFP	+	6/180	79,000
pH349	M1, **S13**–R65–GFP	+	0/118	71
pH350	M1, **R24**–R65–GFP	+ cytopl.	1/118	3
pH351	M1, **S34**–R65–GFP	+	0/118	8
pH352	M1, **N45**–R65–GFP	+ cytopl.	0/118	3
pH769	M1–**S33**, S63–R65–GFP	+ cytopl.	5/100	4
pH768	M1–**I35**, S63–R65–GFP	+ cytopl.	4/100	1
pH767	M1–**F37**, S63–R65–GFP	+ mainly cytopl.	4/100	4,600
pH442	M1–**F39**, S63–R65–GFP	+	0/240	130,000
pH484	M1–**V43**, S63–R65–GFP	−	0/200	66
pH548	M1–**N44**, S63–R65–GFP	+	129/300	60,000
pH547	M1–**N45**, S63–R65–GFP	+	101/200	57,000
pH546	M1–**N46**, S63–R65–GFP	+	80/200	61,000
pH485	M1–**N47**, S63–R65–GFP	+	295/300	39,000
pH766	M1–**N49**, S63–R65–GFP	+	86/100	70,000
pH441	M1–**N50**, S63–R65–GFP	+	139/140	63,000
pH765	M1–**S53**, S63–R65–GFP	+	95/100	53,000

*The GFP signal of most fusion constructs transformed into the [URE3] strain was aggregated. Others showed an even cytoplasmic distribution ('cytopl.').

†Curing was tested as in Table 1.

‡Interference was measured as USA+ cells per 10⁶ cells. All USA+ clones tested became USA− on loss of the plasmid, indicating that this is not [URE3] induction, but simply interference with Ure2p action.

2), we more accurately defined this [URE3] curing region. We find that residues 5–47 are needed for this curing activity. Comparable levels of protein were expressed for nearly all of the constructs as judged by the level of GFP fluorescence.

Unstable Inactivation of Ure2p by Overexpressed Ure2N–GFP. Overexpression of Ure2N–GFP from pVTG4 (CEN P_{ADH1} URE2N) results in the frequent appearance of USA+ colonies in either strain 3686 (MATα trp1 ura2 leu2) (Table 2) or YHE142 [3686 × 3385 (MATa kar1 ura2 leu2 his⁻)]. However, loss of the plasmid from 18 of these colonies uniformly (324 colonies tested) resulted in loss of the USA+ phenotype, whereas subclones that retained the plasmid remained USA+ (314 colonies tested). Moreover, this USA+ phenotype was not well transferred by cytoplasmic mixing, again indicating that it is not a stably self-propagating change. Cytoduction into 3385ρ° + pVTG12 (CEN LEU2 P_{URE2} URE2N–GFP) gave 23 cytoductants, of which 21 were USA− and two were weakly USA+. Cytoduction into 3385ρ° + pH312 (vector) produced only 25 USA− cytoductants. The failure of cytoduction and the instability of the USA+ phenotype shows that this is not caused by generation of the [URE3] prion.

Deletion mutants of the prion domain were tested for ability to inactivate Ure2p as above (Table 2). Most of the fusion proteins inactivating Ure2p appeared to be aggregated in vivo (Table 2), suggesting that Ure2p may have been recruited into aggregates, even though these would not stably propagate. In contrast to the aggregation of overproduced Ure2p–GFP fusion proteins, their production at normal levels generally leads to aggregation only in [URE3] cells (32). The part of Ure2N needed for this high

frequency conversion to USA+ extends from residue 11 to residue 39 (Table 2). We will see below that this region corresponds roughly to a conserved sequence in a variety of yeasts.

Natural Variants of Ure2. The N-terminal region of Ure2p contains a number of asparagine runs and is generally asparagine/serine rich. This finding could suggest a high degree of plasticity in this region. In searching for natural variation of Ure2p, we first examined various strains of S. cerevisiae including pathogenic isolates from immunocompromised patients, and strains used for brewing and baking from geographically widely scattered sources (Table 3). These strains contained 10 silent codon changes, all in the C-terminal domain. All but one were clustered in the region between A224 and A264. Only three changes that altered the amino acid sequence were found, but each, N23S, N70Y, and insertion of an N (AAT codon) between V43 and N44, was in the prion domain. Half of the URE2 sequences (13 of 26) of the strains examined were identical to that of strain S288C. Thus, although amino acid changes are found in the N terminus of Ure2p, the sequence of this region is completely conserved in many of the strains tested, so that [URE3] could occur in wild-type strains of S. cerevisiae.

Conservation of Ure2 N Terminus Among Different Fungi. URE2 homologs were isolated by using degenerate PCR from S. bayanus, S. paradoxus, C. glabrata, C. kefyr, (telomorph is K. marxianus), A. gossypii, C. maltosa, C. albicans, and C. lipolytica (telomorph is Y. lipolytica). Two URE2 alleles were identified in the analyzed strain of C. kefyr. All URE2 homologs have a highly conserved C-terminal domain that starts at Ser-100 of the S. cerevisiae sequence (see Table

Table 3. Amino acid changes in natural isolates of *S. cerevisiae*

Strain	Changes
Σ1278b; YJM145; YJM413; YJM280; YJM326; SAF; Fleischmann; Boots home; McPhie sourdough; CBS2087 Lychee, China; CBS4734 Sugar cane; CBS7957 Cassava flour Brazil; German Ale	None
YJM320	A224
CBS400 Palm wine–Ivory Coast; CBS405 Bili wine, West Africa	A264
CBS3093 Alpechin, Spain; CBS5112 Grape must Spain	L231
CBS5287 Grapes Russia	V245
English Ale; California lager; CBS6216 tap water Roterdam	N23S
CBS429 grapes, France; Red Star Wine; CBS2247 Grape must, S. Africa	N70Y V258
YJM339	N43b S10, E260, R344

N43b means insertion of an asparagine residue after amino acid 43. A224 means a change in codon 224 without changing the amino acid encoded.

4). The N-terminal regions show substantial divergence, although they remain asparagine/glutamine rich (Fig. 1). Interestingly, the N-terminal region identified by deletion analysis as needed for curing of [URE3] is conserved among *S. cerevisiae*, *S bayanus*, *S. paradoxus*, *C. glabrata*, *C. kefyr*, and *A. gossypii*. This region is missing from *C. maltosa*, *C. albicans*, and *C. lipolytica*. As *C. albicans* is an asexual organism, it was possible that this domain is missing in the particular strain we examined. However, a *C. stellatoidae* strain we examined (synonymous with *C. albicans*), and two additional *C. albicans* strains (present in the database) all had N-terminal domains identical to the first *C. albicans* strain, and lacked the conserved N-terminal region. The *URE2* sequences of *S. bayanus* and *S. paradoxus* are nearly identical, but differ in the asparagine-rich domain between Ser-40 and Leu-81 of the *S. cerevisiae* sequence, indicating a relative plasticity of this region.

Complementation of a *S. cerevisiae* *ure2* Deletion by *URE2* Homologs. The *URE2* ORFs of *S. cerevisiae*, *S. bayanus*, *S. paradoxus*, *C. glabrata*, *C. kefyr*, *A. gossypii*, *C. maltosa*, *C. albicans*, and *C. lipolytica* were placed under control of the *S. cerevisiae* URE2 promoter in pH722 (*LEU2 CEN* P$_{URE2}$). Complementation was assayed as the

Table 4. Conservation of *URE2* C-terminal domains

Species	Identity with *S. cerevisiae*, %		
	Amino acid	DNA	Complementation
S. Paradoxus	100	94	+
S. Bayanus	99	85	+
C. Glabrata	92	78	+
C. Kefyr-1	91	79	+
C. Kefyr-2	91	79	+
K. Lactis	91	77	+
A. gossypii	89	71	−/+*
C. Maltosa	82	73	+
C. Albicans	80	73	+
C. Lipolytica	78	66	−*

Comparison starts at Ser-100 of *S. cerevisiae* and equivalent positions in other organisms. The sequence from *K. lactis* was obtained from the GenBank database (gi: 14009513). Complementation is based on expression of the whole sequence with the *S. cerevisiae* URE2 promoter on a CEN plasmid in YHE888.
*Not improved if the *URE2* homolog is on a multicopy plasmid.

inability to grow on USA plates. *C. lipolytica URE2* did not complement, and *A. gossypii URE2* complemented weakly. For both the complementation was not improved significantly when expressed from the 2μ plasmid pH723 (also under control of the *S. cerevisiae* URE2 promoter). All of the other homologs complemented the *ure2* deletion in strain YHE888 (*MATα ura2 leu2::hisG trp1::hisG ure2::G418*), preventing USA uptake with ammonia, while allowing it with proline as a nitrogen source.

Induction of [URE3] by Overproduced *URE2* Homologs. The *URE2* homologs were placed under control of the *GAL1* promoter and overexpressed in a strain with an intact chromosomal *S. cerevisiae* URE2 gene. Induction of [URE3] was assayed as appearance of USA+ colonies (Table 5). Only *S. paradoxus* and, to a slight degree, *S. bayanus* were able to induce the appearance of [URE3] at higher than background rates. The lower than background rates for most of the other homologs probably reflects their masking the spontaneous [URE3] events by complementing the functional deficit of *S. cerevisiae* Ure2p and not being themselves inactivated by the *S. cerevisiae* [URE3] (see below).

Curing of [URE3] by Overexpression of Ure2p Homologs. Each of the Ure2p homologs was expressed from the *GAL1* promoter on a *LEU2 CEN* plasmid in strain YHE64 (*MATα ura2 leu2 trp1* [URE3]). Transformants were confirmed to still have [URE3] as judged by the USA+ phenotype (20 USA+ of 20 tested in each case). This finding shows that homolog production is efficiently repressed on glucose medium, a critical point for this experiment. Transformants were grown to single colonies on YPAGal2%Raf1% to overexpress the Ure2p homolog, and colonies were then replicaplated to leucine dropout plates containing dextrose. Leu+ colonies were spotted on a grid on dextrose media lacking leucine, and replicaplated twice to allow growth and dilution of any remaining heterologous Ure2p. Patches were then replicaplated to USA plates to score retention or loss of [URE3]. There was no curing by any of the *Saccharomyces* Ure2 or by the nonfunctional *C. lipolytica* Ure2p. However, each of the other Ure2s completely cured [URE3]. That this is curing, and not masking of the USA+ phenotype, is shown by the fact that the glucose-repressed transformants are USA+, and the cells are again repressed by glucose after the expression of homolog.

Can Overexpressed Ure2N$^{1-45}$-GFP Inactivate Ure2 Homologs? Strain YHE888 (*ure2*) containing plasmids expressing *URE2* homologs

```
                              |<---interference domain--->|
                              |<----------curing domain------------>|
                              1      10        20        30        40        50        60
         S. cerevisiae        MMNNNGNQVSNLSNALRQVNIGNRNSNTTTDQSNINFEFSTGV---NNNNNNNSSS-----NNNNVQN
         S. paradoxus         ......................................A..........S...NNNNN...A..
         S. bayanus           .....................................PS.................NS..S...
         C. glabrata          MGDSR.TGTI....S.........SGQ------K...Y...NGL------V.D.GNHNLVNTSED.----
         C. kefyr-1           MQQDMHNG-----GT..TI....S......L..............AID.NQQQLMEEV.Q.SMNAFNIQQQHQQQQE.
         C. kefyr-2           MQQDMHNG-----GT..TI....S....SL...............AID.NQQQLMEEV.Q.SMNAFNIQQQHQQQQE.
         K. lactis            MQQDMQNG-----GP..TI....S....N.L..............SID.NQQQLLEEA.QGSINAYNAQ---QQQQEH
         A. gossypii          MQQEIRNSNTPNTGS..P.....S.....HL..............SIDYTSRA------------------------
         C. lipolytica
         C. maltosa
         C. albicans                                                            MMSTDQHIQ..

                                                                      X-ray structure ->
                              70        80        90                                      |>
         S. cerevisiae        NNSGRNGSQNNDNENNIKNT-----LEQHRQQQQA-------------------------------FSDMSHVEY
         S. paradoxus         ..........S...G....D..............................................
         S. bayanus           S.N....T...N...S..D-----I.........................................
         C. glabrata          .KD.SINTNMMSRQVP.QH.HGSQL.Q.E.MNE.Q-------------------------------NP.....
         C. kefyr-1           VQKQQEQQ.QQLQQQQQQQQQQQQQQQ.QQ....QLQQQQQLQQHHHHQQRQQHPNNNVQAGTSQQQML.QGANSIDS
         C. kefyr-2           VQK-----.QQLQQQQQQQQQQQQQQQ.QQ....Q-QQQQQLQQHHHHQQRQQHANNNVQAGTSQQQML.QGANSIDS
         K. lactis            LQ--QQAQ.QQLHMQQLQQAQQQQAQQ.AH....QVHQ----VQHQHVQQ------DHMPIGQSQQQAMYQGPNPIDS
         A. gossypii          --------------------------P...PLDELSRGAAGGPAAGGGPEGSNMVGASSAAQVTA.GGP.V.DS
         C. lipolytica                                       MSGAHTISHLSAGLKSVNIGDQQQNEANLNLLQQQLENEAT.QTQQ
         C. maltosa           M.M.DQRIPQ.TGD.SNNSNSN-NNNNNNNNTHTISNLSAGLKSVSLTDQQQNEVNLNLLQQQLHRESSNQQQQ
         C. albicans          M.DNS.N.N.SN.N.TNN.NNNQSVNVNVNNTNNNTQTISNLSAGLKSVSLTDQQQNEVNLNLLQQQLHQEASTQQQQ

                              100       110       120       130       140       150       160       170
         S. cerevisiae        SRITKFFQEQPLEGYTLFSHRSAPNGFKVAIVLSELGFHYNTIFLDFNLGEHRAPEFVSVNPNARVPALIDHGMDNLS
         S. paradoxus         ............................................................................
         S. bayanus           .......................................................................N.....
         C. glabrata          ...S....N..M..............S......LQ...............................LE..A
         C. kefyr-1           ........N..M...............NM...............AI...........N....
         C. kefyr-2           ........N..M...............NM...............AI...........N....
         K. lactis            ........N..M...............NM...............AI..........N.E...
         A. gossypii          ........N..M.............M.LN...............AI...........SL....
         C. lipolytica        L..Q..AQ..SD.............I....ELPF........H..Q.....TI...........FN..T.
         C. maltosa           ....Q...N..A.............I....NLPF........N..Q.....TI...........FNE.T.
         C. albicans          ....Q...N..T..F..........I....NLPF..F.....N..Q.T....TI...........YN..T.
```

Fig. 1. Alignment of Ure2p homologs from yeasts and a filamentous fungus by using GCG pileup. The "." indicates identity, whereas "−" indicates a gap. S and T residues are red, and Q and N residues are green. The C-terminal (nitrogen regulation) domain begins at residue 100 in the *S. cerevisiae* sequence and continues to residue 354, though the figure only shows through residue 174. The "interference domain" is the portion of Ure2N which when overexpressed as a GFP fusion interferes with Ure2p activity. The "curing domain" is the portion that, fused to GFP, is needed for curing [URE3].

(*LEU2 CEN* $P_{URE2}URE2^{homolog}$) was transformed with either pH401 (*TRP1* 2μ P_{AHD1}) or pH792 (*TRP1* 2μ $P_{AHD1}URE2N^{1-45}$-GFP), and transformants were plated at serial dilutions on USA plates. Only the *S. cerevisiae* Ure2p was inactivated through overexpression of $URE2N^{1-45}$-GFP.

Transformants containing the *S. cerevisiae*, *C. glabrata*, or *C. kefyr URE2* were also plated quantitatively giving the same result as above. It is surprising that the nearly identical *S. bayanus* and *S. paradoxus* Ure2s were not inactivated. The fact that the source of Ure2p determined whether interference was observed indicates that Ure2N$^{1-45}$ was not interfering with the action of another component of the nitrogen control pathway.

Ability of Ure2 Homologs to Propagate [URE3]. *URE2* homologs were expressed in strain YHE888 (*ure2*) under control of the *S. cerevisiae* URE2 promoter from a centromeric *LEU2* plasmid. [URE3] was introduced into these strains by cytoduction from strain 4833–3B (*MATa ura2 arg1 kar1-1* [URE3-1]), and the cytoductants were examined for the USA phenotype. Only the *URE2* homologs from *S. bayanus* (11 USA+ of 11 cytoductants) and *S. paradoxus*

(6 USA+ of 11 cytoductants) were able to propagate [URE3]. These are the same two Ure2s that can induce *S. cerevisiae* [URE3] appearance when overexpressed. The Ure2s of *C. kefyr* (both genes), *C. albicans*, *C. maltosa*, and *C. glabrata* gave only USA− cytoductants. The *Ashbya* Ure2p could not be tested because of its incomplete complementation ability.

Discussion

Curing of [URE3]. Because propagation of [PSI+] (52) and [URE3] (53) require Hsp104, interference with its production can cure either prion. Guanidine at millimolar concentrations can also cure either [PSI+] (54) or [URE3] (12, 55) by a mechanism that appears to involve inactivation of Hsp-104 (56, 57). [URE3] is also cured by overexpression of parts of Ure2p, particularly when they are fused to GFP (32). This curing by the "hair of the dog" method has a potentially broad application, and is known to be effective in curing tissue culture cells of PrP (58). We proposed that the Ure2p fragments or fusion proteins join the growing filaments, but do not themselves provide a growing point, thus poisoning the linear amyloid crystal, but there are other possibilities.

Table 5. Ability of URE2 homologs to induce or cure [URE3] in *S. cerevisiae*

URE2 gene	Plasmid	Induction, USA+/10^6 cells	Plasmid	Curing, USA+/USA−
Vector	pH317	22	pH316	39/1
S. cerevisiae	pH739	7,700	pH740	40/0
S. bayanus	pH661	52	pH679	40/0
S. paradoxus	pH660	6,300	pH678	40/0
C. glabrata	pH659	<1	pH677	0/40
A. gossypii	pH656	1	pH674	0/40
C. kefyr-1	pH713	2	pH711	0/40
C. kefyr-2	pH714	2	pH712	0/40
C. albicans	pH563	6	pH672	0/40
C. maltosa	pH657	<0.1	pH675	0/40
C. lipolytica	pH658	22	pH676	40/0

Strain YHE711 (*MAT*α *ura2 leu2::hisG*) was transformed with 2μ plasmids carrying *URE2* homologs under control of the *GAL1* promoter. Individual transformants were grown to saturation in leucine dropout medium containing 2% galactose and 1% rafinose and plated in 10-fold dilutions onto USA plates to assay [URE3] induction. For curing, centromeric plasmids were transformed into YHE64. Transformants were confirmed to still carry [URE3], then grown to single colonies on YPAGal2%Raf2% to overexpress the Ure2p homolog. Leu+ colonies were grown as patches three times on dextrose, then tested for USA phenotype.

Here, we defined the parts of the C-terminal and N-terminal domains of Ure2p necessary for their curing activities. The borders of the part of Ure2C needed for curing [URE3], are approximately residues 111 and 347. The domain needed for complementation is larger, extending from approximately residue 86 to residue 354. Interaction of Ure2p with Gln3p requires at least Ure2p residues 151–330 (22), and dimer formation has been demonstrated for Ure2p residues 90–354, but not for Ure2p residues 111–354 (33). The C-terminal domain of Ure2p may cure [URE3] by forming heterodimers with the full-length Ure2p. This interaction might compete with its incorporation into the filaments. [URE3] prion stability is not affected by the nitrogen source (29), so it is unlikely that interactions with Gln3p or other factors involved in nitrogen regulation explain this curing phenomenon. Mks1p is necessary for [URE3] prion generation, not for propagation, indicating that possible interactions with this protein are not likely to be involved.

By using the yeast two hybrid method, evidence for an interaction of Ure2p residues 1–96 with Ure2p residues 152–354 has been obtained (59). However, our results show that a larger segment of Ure2C is needed for the curing, arguing against explaining the curing by this interaction of N terminus and C terminus competing for N terminus–N terminus interactions.

We find that the part of Ure2N (as a fusion with GFP) needed for curing of [URE3] is N5 to N47, a relatively short region. In this case, the crystal poisoning mechanism, in which "impurities" prevent crystal growth, is particularly attractive. A slightly smaller region, amino acid residues N11 to N44, is necessary for interfering with Ure2p activity on overexpression. All of the constructs able to interfere with Ure2p appear to be aggregated as judged by the nonhomogeneous distribution of GFP fluorescence. It is possible that the overexpressed Ure2–GFP fusion protein forms aggregates that sequester the full-length Ure2p, but that these aggregates are not self propagating, or at least do not initiate a self-propagating aggregation of the full-length Ure2p. Negative complementation of Sup35p by a fragment of its prion domain has also been observed, but whether or not this is associated with aggregation is not yet known (60).

Whatever the mechanism of inhibition, the N11 to N44 region probably interacts with full-length Ure2p, though we do not yet have direct evidence for this interaction. This region corresponds quite closely with the part of the N-terminal domain that is conserved among a group of Ure2p homologs (Fig. 1). The conservation of this region, despite wide divergence of sequence in the N-terminal part, suggests that this region is important for some function. In addition to the two-hybrid data (59), functional data suggests that Ure2N and Ure2C interact. Deletion of Ure2 residues 1–65 weak-

ens the ability of Ure2C to carry out its function in nitrogen regulation (29). Likewise, deletions of all or parts of Ure2C dramatically increase the frequency with which Ure2p changes to the prion form (26), suggesting that the C-terminal domain stabilizes the N-terminal prion domain, perhaps by an interaction.

Ure2 Homologs. In surveying clinical isolates, brewing strains, and baking strains of *S. cerevisiae* from a variety of sources, we find that the *URE2* sequence is well conserved, and the few amino acid changes observed are in the N-terminal prion domain. Similar results were obtained by Jensen *et al.* (61) studying the Sup35 protein. Examination of *URE2* genes from a series of yeasts and the fungus *A. gossypii* shows that the C-terminal part of Ure2p is highly conserved. Although Ure2p is homologous to the θ group of GSTs (18, 62), that similarity is only about 30%, whereas these proteins are 80–90% identical to each other in their C-terminal domains. For example, the "cap region" constitutes a loop with an α helix (36) of the *S. cerevisiae* Ure2p that is missing in most GSTs, but is present in all of the Ure2p homologs studied here as well as in that of *Kluyveromyces lactis* (GenBank accession no. AAK51642). Moreover, most of the yeast and fungal homologs fully complement the *S. cerevisiae* ure2Δ mutant, indicating that the nitrogen regulation function is conserved.

Unlike other θ class GSTs, the *S. cerevisiae* Ure2p has Ala-122 and His-187 instead of the consensus Ser and Tyr, respectively. In the sequences obtained here, the Ala for Ser substitution is maintained, but the *C. lipolitica*, *C. maltosa*, and *C. albicans* Ure2 proteins contain the consensus Tyr-187, and the consensus Gly-136 is replaced with Asp in *C. kefyr*, *K. lactis*, *C. maltosa*, and *C. albicans*, and with Glu in *C. lipolitica*. Thus, all of the Ure2p homologs diverge at critical residues from the GST consensus sequence.

All of the Ure2 homologs studied here have an asparagine/glutamine-rich N-terminal extension not found in the enzymatically active GSTs or in homologs from *Schizosaccharomyces pombe* or *Neurospora crassa* (ref. 63 and www-genome.wi.mit.edu). Interestingly, precisely those homologs with an N-terminal extension have the "cap region" insert in the C-terminal domain. Are they functionally related? The functional significance of the N-terminal extension of Ure2p remains a mystery. It is unlikely that prion formation helps cells regulate nitrogen catabolism, because it differs from the normal regulation mainly in lacking flexibility. Although Ure2C can regulate nitrogen catabolism without the prion domain, this regulation is less efficient than that carried out by the full-length Ure2p (26, 29). This helper activity of the N terminus may be sufficient to explain its retention in evolution.

A domain that is only present in yeasts closely related to *S.*

Edskes and Wickner

cerevisiae and the filamentous fungus *A. gossypii* (also closely related to *S. cerevisiae*) is the region between S10 and I35. The conservation of this region, despite the wide sequence divergence of the remainder of the N termini, suggests the presence of some functional constraint. There is no similarly conserved portion of the Sup35p prion domain (37–39).

The curing of [URE3] by homologs of Ure2p is striking in that all complementing Ure2s can cure except for the Ure2s of *Saccharomyces* species, which can participate in the [URE3] process. The results suggest that the C-terminal domain is doing the curing in these cases. However, it is not the complementation *per se* masking the [URE3] phenotype, because cells are assayed for [URE3] when the expression of the homolog is repressed on glucose.

Species Barrier for [URE3]. Our efforts to transmit [URE3] from *S. cerevisiae* Ure2p to Ure2p of other fungi was only successful for the Ure2s from other *Saccharomyces* species. These were the most closely related in N-terminal sequence, and would thus be expected to have the lowest barrier to transmission. The fact that transmission of [URE3] was not observed to the *C. glabrata* or *C. kefyr* Ure2

proteins from the *S. cerevisiae* Ure2p, despite all having the conserved sequence in the residue 10–40 region, indicates that this region is not sufficient to allow prion transmission.

Similar results were obtained on testing the Ure2p homologs for ability to induce the *de novo* appearance of *S. cerevisiae* [URE3]. Only *S. paradoxus*, and to a slight extent *S. bayanus*, were able to induce [URE3] appearance. Others behaved like Ure2C (26) in showing an apparent decrease in frequency of [URE3], indicative of efficient complementation and inability to be affected by [URE3]. *C. lipolytica* did not induce, but does not complement *ure2Δ* either, and showed a normal background frequency of [URE3].

Further work will be needed to determine whether the non-*Saccharomyces* Ure2ps can form prions on their own. It will also be important to determine whether the conserved S10 to I35 peptide participates in amyloid formation, and with what, if any, part of Ure2C it interacts.

1. Weissmann, C., Enari, M., Klöhn, P.-C., Rossi, D. & Flechsig, E. (2002) *Proc. Natl. Acad. Sci. USA* **99**, Suppl. 4, 16378–16383.
2. Dickinson, A. G., Meikle, V. M. H. & Fraser, H. (1968) *J. Comp. Pathol.* **78**, 293–299.
3. Oesch, B., Westaway, D., Walchli, M., McKinley, M. P., Kent, S. B., Aebersold, R., Barry, R. A., Tempst, P., Templow, D. B., Hood, L. E., *et al.* (1985) *Cell* **40**, 735–746.
4. Carlson, G. A., Kingsbury, D. T., Goodman, P. A., Coleman, S., Marshall, S. T., DeArmond, S., Westaway, D. & Prusiner, S. B. (1986) *Cell* **46**, 503–511.
5. Hsiao, K., Baker, H. F., Crow, T. J., Poulter, M., Owen, F., Terwilliger, J. D., Westaway, D., Ott, J. & Prusiner, S. B. (1989) *Nature (London)* **338**, 342–345.
6. Prusiner, S. B., Scott, M., Foster, D., Pan, K.-M., Groth, D., Mirenda, C., Torchia, M., Yang, S.-L., Serban, D., Carlson, G. A., *et al.* (1990) *Cell* **63**, 673–686.
7. Bueler, H., Aguzzi, A., Sailer, A., Greiner, R.-A., Autenried, P., Aguet, M. & Weissmann, C. (1993) *Cell* **73**, 1339–1347.
8. Alper, T., Cramp, W. A., Haig, D. A. & Clarke, M. C. (1967) *Nature (London)* **214**, 764–766.
9. Bolton, D. C., McKinley, M. P. & Prusiner, S. B. (1982) *Science* **218**, 1309–1311.
10. Lacroute, F. (1971) *J. Bacteriol.* **106**, 519–522.
11. Cox, B. S. (1965) *Heredity* **20**, 505–521.
12. Wickner, R. B. (1994) *Science* **264**, 566–569.
13. Wickner, R. B., Edskes, H. K., Roberts, B. T., Pierce, M. & Baxa, U. (2002) *Adv. Genet.* **46**, 485–525.
14. Bradley, M. E., Edskes, H. K., Hong, J. Y., Wickner, R. B. & Liebman, S. W. (2002) *Proc. Natl. Acad. Sci. USA* **99**, Suppl. 4, 16392–16399.
15. Wickner, R. B., Taylor, K. L., Edskes, H. K., Maddelein, M.-L., Moriyama, H. & Roberts, B. T. (2001) *Adv. Prot. Chem.* **57**, 313–334.
16. Wickner, R. B., Taylor, K. L., Edskes, H. K., Maddelein, M.-L., Moriyama, H. & Roberts, B. T. (1999) *Microbiol. Mol. Biol. Revs.* **63**, 844–861.
17. Drillien, R. & Lacroute, F. (1972) *J. Bacteriol.* **109**, 203–208.
18. Coschigano, P. W. & Magasanik, B. (1991) *Mol. Cell. Biol.* **11**, 822–832.
19. Beck, T. & Hall, M. N. (1999) *Nature (London)* **402**, 689–692.
20. Cardenas, M. E., Cutler, N. S., Lorenz, M. C., Di Como, C. J. & Heitman, J. (1999) *Genes Dev.* **13**, 3271–3279.
21. Hardwick, J. S., Kuruvilla, F. G., Tong, J. K., Shamji, A. F. & Schreiber, S. L. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14866–14870.
22. Kulkarni, A. A., Abul-Hamd, A. T., Rai, R., El Berry, H. & Cooper, T. G. (2001) *J. Biol. Chem.* **276**, 32136–32144.
23. Cooper, T. G. (1982) in *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, eds. Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 2, pp. 39–99.
24. Rai, R., Genbauffe, F., Lea, H. Z. & Cooper, T. G. (1987) *J. Bacteriol.* **169**, 3521–3524.
25. Magasanik, B. (1992) in *The Molecular and Cellular Biology of the Yeast Saccharomyces*, eds. Jones, E. W., Pringle, J. R. & Broach, J. R. (Cold Spring Harbor Lab. Press, Plainview, NY), Vol. 2, pp. 283–317.
26. Masison, D. C. & Wickner, R. B. (1995) *Science* **270**, 93–95.
27. Maddelein, M.-L. & Wickner, R. B. (1999) *Mol. Cell. Biol.* **19**, 4516–4524.
28. Kochneva-Pervukhova, N. V., Poznyakovski, A. I., Smirnov, V. N. & Ter-Avanesyan, M. D. (1998) *Curr. Genet.* **34**, 146–151.
29. Masison, D. C., Maddelein, M.-L. & Wickner, R. B. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 12503–12508.
30. Taylor, K. L., Cheng, N., Williams, R. W., Steven, A. C. & Wickner, R. B. (1999) *Science* **283**, 1339–1343.
31. Speransky, V., Taylor, K. L., Edskes, H. K., Wickner, R. B. & Steven, A. (2001) *J. Cell. Biol.* **153**, 1327–1335.
32. Edskes, H. K., Gray, V. T. & Wickner, R. B. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1498–1503.
33. Perrett, S., Freeman, S. J., Butler, P. J. G. & Fersht, A. R. (1999) *J. Mol. Biol.* **290**, 331–345.
34. Baxa, U., Speransky, V., Steven, A. C. & Wickner, R. B. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5253–5260.
35. Umland, T. C., Taylor, K. L., Rhee, S., Wickner, R. B. & Davies, D. R. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 1459–1464.
36. Bousset, L., Beirhali, H., Janin, J., Melki, R. & Morera, S. (2001) *Structure (London)* **9**, 39–46.
37. Chernoff, Y. O., Galkin, A. P., Lewitin, E., Chernova, T. A., Newnam, G. P. & Belenkly, S. M. (2000) *Mol. Microbiol.* **35**, 865–876.
38. Kushnirov, V. V., Kochneva-Pervukhova, N. V., Cechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. (2000) *EMBO J.* **19**, 324–331.
39. Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. (2000) *Cell* **100**, 277–288.
40. Sherman, F. (1991) *Methods Enzymol.* **194**, 3–21.
41. McCusker, J. H., Clemons, K. V., Stevens, D. A. & Davis, R. W. (1994) *Genetics* **136**, 1261–1269.
42. Miyazaki, H., Miyazaki, Y., Gerber, A., Parkinson, T., Hitchcock, C., Falconer, D. J., Ward, D. J., Marsden, K. & Bennett, J. E. (1998) *Antimicrob. Agents Chemother.* **42**, 1695–1701.
43. Collins, M. S. & Pappagianis, D. (1974) *Sabouraudia* **12**, 329–340.
44. Bassel, J., Phaff, H. J., Mortimer, R. K. & Miranda, M. (1978) *Intl. J. Syst. Bacteriol.* **28**, 427–432.
45. Kakeya, H., Miyazaki, Y., Miyazaki, H., Nyswaner, K., Grimberg, B. & Bennett, J. E. (2000) *Antimicrob. Agents Chemother.* **44**, 2985–2990.
46. Barth, G. & Gaillardin, C. (1997) *FEMS Microbiol. Rev.* **19**, 219–237.
47. Ashby, S. F. & Nowell, N. (1926) *Ann. Bot.* **40**, 69–83.
48. Edskes, H. K. & Wickner, R. B. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 6625–6629.
49. Christianson, T. W., Sikorski, R. S., Dante, M., Shero, J. H. & Hieter, P. (1992) *Gene* **110**, 119–122.
50. Pitkin, J. W., Panaccione, D. G. & Walton, J. D. (1996) *Microbiology* **142**, 1557–1565.
51. Timberlake, W. E. (1980) *Dev. Biol.* **78**, 497–510.
52. Chernoff, Y. O., Lindquist, S. L., Ono, B.-I., Inge-Vechtomov, S. G. & Liebman, S. W. (1995) *Science* **268**, 880–884.
53. Moriyama, H., Edskes, H. K. & Wickner, R. B. (2000) *Mol. Cell. Biol.* **20**, 8916–8922.
54. Tuite, M. F., Mundy, C. R. & Cox, B. S. (1981) *Genetics* **98**, 691–711.
55. Aigle, M. (1979) Ph.D. thesis (L'Universite Louis Pasteur de Strasbourg, Strasbourg, France), pp. 95.
56. Jung, G. & Masison, D. C. (2001) *Curr. Microbiol.* **43**, 7–10.
57. Ferreira, P. C., Ness, F., Edwards, S. R., Cox, B. S. & Tuite, M. F. (2001) *Mol. Microbiol.* **40**, 1357–1369.
58. Chabry, J., Caughey, B. & Chesebro, B. (1998) *J. Biol. Chem.* **273**, 13203–13207.
59. Fernandez-Bellot, E., Guillemet, E., Baudin-Baillieu, A., Gaumer, S., Komar, A. A. & Cullin, C. (1999) *Biochem. J.* **338**, 403–407.
60. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. (1996) *Genetics* **144**, 1375–1386.
61. Jensen, M. A., True, H. L., Chernoff, Y. O. & Lindquist, S. (2001) *Genetics* **159**, 527–535.
62. Rossjohn, J., Board, P. G., Parker, M. W. & Wilce, M. C. (1996) *Protein Eng.* **9**, 327–332.
63. Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., *et al.* (2002) *Nature (London)* **415**, 871–880.

Colloquium

Interactions among prions and prion "strains" in yeast

Michael E. Bradley*, Herman K. Edskes†, Joo Y. Hong*, Reed B. Wickner†, and Susan W. Liebman*‡

*Laboratory for Molecular Biology, Department of Biological Sciences, University of Illinois, 900 South Ashland Avenue, Chicago, IL 60607; and †Laboratory of Biochemistry and Genetics, National Institute of Diabetes, Digestive and Kidney Diseases, Building 8, Room 225, National Institutes of Health, 8 Center Drive MSC0830, Bethesda, MD 20892-0830

Prions are "infectious" proteins. When Sup35, a yeast translation termination factor, is aggregated in its [*PSI*+] prion form its function is compromised. When Rnq1 is aggregated in its [*PIN*+] prion form, it promotes the *de novo* appearance of [*PSI*+]. Heritable variants (strains) of [*PSI*+] with distinct phenotypes have been isolated and are analogous to mammalian prion strains with different pathologies. Here, we describe heritable variants of the [*PIN*+] prion that are distinguished by the efficiency with which they enhance the *de novo* appearance of [*PSI*+]. Unlike [*PSI*+] variants, where the strength of translation termination corresponds to the level of soluble Sup35, the phenotypes of these [*PIN*+] variants do not correspond to levels of soluble Rnq1. However, diploids and meiotic progeny from crosses between either different [*PSI*+], or different [*PIN*+] variants, always have the phenotype of the parental variant with the least soluble Sup35 or Rnq1, respectively. Apparently faster growing prion variants cure cells of slower growing or less stable variants of the same prion. We also find that *YDJ1* overexpression eliminates some but not other [*PIN*+] variants and that prions are destabilized by meiosis. Finally, we show that, like its affect on [*PSI*+] appearance, [*PIN*+] enhances the *de novo* appearance of [*URE3*]. Surprisingly, [*PSI*+] inhibited [*URE3*] appearance. These results reinforce earlier reports that heterologous prions interact, but suggest that such interactions can not only positively, but also negatively, influence the *de novo* generation of prions.

Prions are best known as the infectious agents proposed to be responsible for the mammalian transmissible spongiform encephalopathies including scrapie in sheep, mad cow disease in cattle, and Creutzfeldt-Jakob disease in humans (1). The prion form of the PrP protein is proposed to propagate its abnormal form to other "normal" PrP protein molecules with the same primary sequence (2, 3). Evidence suggesting that self-propagating prion proteins are not limited to PrP was presented in 1994 by Wickner (4). Three yeast proteins with self-perpetuating, alternate conformations have now been well described: [*PSI*+] (5), the prion form of the translational termination factor Sup35 (6–8); [*URE3*] (9), the prion form of the nitrogen catabolite repressor Ure2 (4, 8); and [*PIN*+] (10, 11), the prion form of Rnq1 (12, 13). In [*PSI*+] and [*URE3*] cells, Sup35 and Ure2 are respectively inactivated by aggregation, causing the same phenotypes as mutations in the *SUP35* and *URE2* genes. No phenotype has yet been associated with inactivation of *RNQ1*.

The *de novo* appearance of each of these yeast prions, [*PSI*+], [*URE3*], and [*PIN*+], is enhanced by overproducing the corresponding prion domains (12, 14, 15). The increased number of protein molecules presumably enhances the chance that a prion seed will form *de novo*. However, *de novo* appearance of some prions depends on the presence of other prions or prion-like aggregates (13, 16). We first described [*PIN*+] as a prion-like element having the phenotype of allowing overproduction of Sup35 to convert [*psi*−] cells to [*PSI*+] (10, 11), and later showed that [*PIN*+] is equivalent to the prion form of Rnq1 (13). The presence of [*URE3*]

(13) or the artificial fusion prion [*NU*+] (16) also permitted overexpression of Sup35 to induce the appearance of [*PSI*+].

The existence of different heritable forms or strains of prions is a fascinating chapter in the biology of prions. Prion diseases exhibit variable incubation times, neurodegenerative patterns, and PrP prion deposits, all of which remain distinct on transmission in inbred mammals (17). Recent evidence supports the idea that prion strain variation is a result of the PrP protein's ability to propagate in different heritable prion forms (18–20). Others, however, view the existence of prion strains as more compatible with a viral model for prion disease (21). The finding of [*PSI*+] strains (14), and more recently [*URE3*] strains (22), in yeast, where the viral hypothesis is unreasonable, supports the idea that prion strains result from multiple prion protein forms.

Distinct strains of [*PSI*+] have different mitotic stabilities (frequencies of [*PSI*+] loss), translational termination activities as measured by suppression of nonsense codons (14), and levels of nonaggregated Sup35 (23). Weak [*PSI*+] are less stable than strong [*PSI*+] in mitotic division (14), and the levels of nonaggregated Sup35 and accompanying translational termination are higher in weak [*PSI*+] cells than in strong [*PSI*+] cells (23). Strains of [*PSI*+] have also been distinguished by their differential responses to mutations in the *SUP35* gene (24, 25) and by their responses to chaperones (26). Several [*PSI*+] strains have been shown to be dominant, non-Mendelian traits when crossed with [*psi*−] (5, 27). In addition, the strong [*PSI*+] phenotype appears in diploids made from mating strong and weak [*PSI*+] cells (23, 24).

Several *in vitro* studies support the hypothesis that strains of [*PSI*+] result from distinct protein conformations of Sup35. Purified Sup35 prion domain (Sup35NM) forms fibers *in vitro* with either wavy or straight structures (28). Also, a purified chimeric Sup35NM was shown to form aggregates with distinct conformations and distinct *in vitro* seeding activities (29). Most convincingly, protein extracts from strong [*PSI*+] cells converted purified Sup35NM into fibers more efficiently than did protein extracts from weak [*PSI*+] cells (30, 31).

The continued propagation of prions depends on normal chaperone protein levels. The finding that either deleting or overexpressing the *HSP104* chaperone gene causes the elimination of [*PSI*+] (32) supported the prion model for [*PSI*+] because it suggested that [*PSI*+] formation involved a conformational change. Deleting, but not overexpressing, *HSP104* eliminates [*URE3*] (33) and [*PIN*+] (10, 34), and overexpressing *YDJ1*, which encodes an Hsp40 family chaperone, promotes the loss of [*URE3*] (33) and a weak strain of *Pichia methanolica* [*PSI*+] in *Saccharomyes cerevisiae*

---

(26). Hsp104 functions together with Hsp40 and Hsp70 (35) to promote the renaturation of denatured or aggregated proteins (36). The effect of Hsp104 on [*PSI*⁺] is modified by levels of, and alterations in, the Hsp70 family members Ssa and Ssb (26, 37–39). Maintaining the prion form of Rnq1 requires specific domains of the Hsp40 family member Sis1 (34). In addition, a specific mutation of *SIS1* caused an altered aggregation pattern of the Rnq1 prion that appeared to be heritable even in the absence of the *SIS1* mutation (34).

Here, we demonstrate the existence of different strains of [*PIN*⁺] with distinct phenotypes. We determine the relative competitiveness of these [*PIN*⁺] prion strains and of [*PSI*⁺] prion strains and find that one factor foretells the outcomes of competitions between two variants of the same prion. We also find that, whereas the [*PIN*⁺] prion enhances the *de novo* appearance of [*URE3*], the presence of the [*PSI*⁺] prion inhibits [*URE3*] appearance. The non-Mendelian segregation of prions has been reported to deviate from the classical 4:0 ratio (4, 5, 9, 27, 40). Here, we carefully document this effect for [*PSI*⁺] and [*URE3*], and show that it is due to meiosis and not the conditions used to stimulate sporulation. Finally, we show that overproducing the chaperone Ydj1 promotes the elimination of some [*PIN*⁺] strains.

Materials and Methods

Cultivation Procedures. Standard yeast media and cultivation procedures were used (41). Yeast extract/peptone/dextrose (YPD) with 5 mM guanidine·hydrochloride (YPD + GuHCl) was used to eliminate prions (42). YPD with 10 mg/liter cycloheximde (YPD + Cyh) was used to select for cycloheximide-resistant mutations (*cyh*^{R}) and for random spores in [*PSI*⁺] experiments. Synthetic medium with 3 mg/liter cycloheximide was used to select for cytoductants. Growth on YPD plates containing 40 μg/ml of ethidium bromide converted strains to [*rho*⁻] (43). Casamino acid (CA) medium contained 0.13% yeast nitrogen base, 0.5% ammonium sulfate, 1% casamino acids, and 2% glucose or 2% glycerol. Uracil, adenine, and tryptophan were added to CA when necessary. CA with 5 mg/liter cycloheximide (CA + Cyh) was used to select for random spores in [*PIN*⁺] experiments. Synthetic medium containing galactose and raffinose (SGal + Raf) was used to overexpress *YDJ1* or *URE2* from the galactose-inducible promoter *GAL1*. Copper sulfate (Cu; 50 μM) was added to medium to induce expression of *RNQ1* or *SUP35* under the control of the inducible *CUP1* promoter.

Strains. Opposite mating type yeast strains that are isogenic, sporulate efficiently when intercrossed, and contain the *ade1–14* allele, which permits weak and strong [*PSI*⁺] to be distinguished, were constructed by mating 74-D694 (*MATa ade1–14 trp1–289 his3-Δ200 leu2-3,112 ura3–52*) (44) with an efficiently sporulating strain, NKY292 (*MATα lys2 ura3 leu2::hisG ho::LYS2*) (45) (kindly supplied by D. Bishop, University of Chicago) and backcrossing *MATα* meiotic progeny to 74-D694 four times. Progeny from the final backcross were diploidized by transforming them with pGAL-HO (46) (kindly provided by R. Esposito, University of Chicago). The diploids were sporulated and dissected to obtain *MATa* and *MATα* segregants that otherwise had the same genotype as 74-D694. L1842 (*MATa*) and L1843 (*MATα*) are one pair of opposite mating type progeny from a single tetrad, as are L1844 (*MATa*) and L1845 (*MATα*). L2176 is a spontaneous *cyh*^{R} derivative of L1845 in which the *cyh*^{R} mutation was shown to be recessive.

To induce weak and strong [*PSI*⁺], L1842, L1843, L1844, and L1845 were transformed with pEMBL-SUP35. Transformants were grown on plasmid selective medium for ≈14 generations and subsequently on YPD for ≈14 generations to promote plasmid loss before plating the cultures for individual colonies on YPD. L2010 and L2012 are, respectively, weak and strong [*PSI*⁺] derivatives of L1844.

YHE711 (*MATα ura2 leu2* [*psi*⁻] [*ure-o*] [*PIN*⁺]) (47) was scored

as [*PIN*⁺] because expression of an RNQ1-GFP fusion in the strain gave rise to discrete aggregates characteristic of [*PIN*⁺] (12, 13). In addition, as expected in a [*PIN*⁺] background, overexpression of SUP35NM-GFP in YHE711 led to the appearance of ribbon and curve aggregates characteristic of newly induced [*PSI*⁺] (48). Derivatives of YHE711 grown in YPD + GuHCl failed to give rise to either of these types of aggregates and are therefore [*pin*⁻]. An *RNQ1* deletion derivative of the [*PIN*⁺] version of YHE711 was constructed by transformation with a PCR product of the *RNQ1*::kanMX4 insertion from yeast strain [American Type Culture Collection (ATCC) no. 4003435; ref. 49] amplified with DNA primers HE230 (*RNQ1* 5′ UTR, 5′-CACGTATTTCAGTTGTCC-3′) and HE231 (*RNQ1* 3′ UTR, 5′-CCACTCTTACATTGT-CATT-3′). Transformants were selected on YPD containing 300 μg/ml G418, after a recovery period in YPD. To confirm the disruption of *RNQ1*, candidate mutants were analyzed by PCR using primers HE265 (*RNQ1* 5′ UTR, 5′-GAATGATCCATCGT-TCTTAC-3′), HE266 (*RNQ1* 3′ UTR, 5′-GATGGCTTATATC-CTGCTC-3′), HE267 (kanMX4 pointing to 5′ end, 5′-CTGCAGC-GAGGAGCCGTAAT-3′), and HE268 (kanMX4 pointing to 3′ end, 5′-TGATTTTGATGACGAGCGTAAT-3′).

GuHCl-treated versions of yeast strains A3099 (*MATα ade2-1 SUQ5 lys1-1 his3–11,15 leu1 kar1-1 ura3::KanMX4* [*psi*⁻][*rho*⁻]) (12), c10B-H49 (*MATα ade2-1 SUQ5 lys1-1 his3–11,15 leu1 kar1-1 cyh*^{R} [*psi*⁻][*rho*⁻]) (50), BY4741 (*MATa his3-Δ1 leu2-Δ met15-Δ ura3-Δ* [*psi*⁻] [*PIN*⁺]) (from Research Genetics), and 3385 (*MATa ura2 leu2 kar1 his*⁻ [*psi*⁻] [*ure-o*][*PIN*⁺]) (4) were used in cytoduction experiments.

Plasmids. A 2μ plasmid (pEMBL-SUP35) with *URA3 leu2-d* markers and *SUP35* under its native promoter (51) was used to overproduce Sup35 at moderate levels on synthetic medium lacking uracil (−Ura) or high levels (on −Leu). The defective *LEU2* promoter present in the *leu2-d* allele on this plasmid selects for overamplification of the plasmid on −Leu. Moderate overexpression of *SUP35* was used to induce [*PSI*⁺]; high-level overexpression was used to distinguish different variants of [*PIN*⁺] on the basis of growth inhibition. *YDJ1* under the control of the *GAL1* promoter is present on a *CEN LEU2* plasmid (p901); the parent plasmid without *YDJ1* is pH316 (33). Plasmids pRNQ1-GFP and pSUP35NM-GFP, which respectively contain the fusions of either *RNQ1* or the NM domains of *SUP35* to green fluorescent protein (GFP) under the control of the *CUP1* promoter, were used to score for [*PIN*⁺] as described previously (13). Plasmids used for the [*URE3*] induction experiments were 2μ-based, with a *LEU2* marker and the *GAL1* promoter to express *URE2* (pH376), *URE2* (1–65) (pH382), *URE2Δ*$_{151–158}$ (pH377), or no expression of *URE2* as a negative control (pH317) (52).

Cytoduction. Cytoductions were performed by crossing [*RHO*⁺] donors to [*rho*⁻] recipients. Either the donor or the recipient carried the *kar1-1* allele, which inhibits nuclear fusion (53). When the recipient was cycloheximide-resistant (*cyh*^{R}), cytoductants were selected on synthetic glycerol medium containing cycloheximide. Otherwise, diploids and cytoductants were selected on synthetic glycerol medium deficient in a nutrient required by the donor strain for growth. Cytoductants were then identified by subcloning the population and screening colonies for the recipient mating type and auxotrophic markers.

Analyses of [*PSI*⁺] Variants. After inducing weak and strong [*PSI*⁺] in L1842, L1843, L1844, and L1845, the L1842 and L1843 derivatives (shown in Fig. 1) and the L1844 and L1845 derivatives were crossed in all possible combinations. We assayed diploids for [*PSI*⁺] strength by color on YPD and level of growth on synthetic medium lacking adenine (−Ade). The diploids were sporulated after propagating for approximately 42 generations. Meiotic progeny from
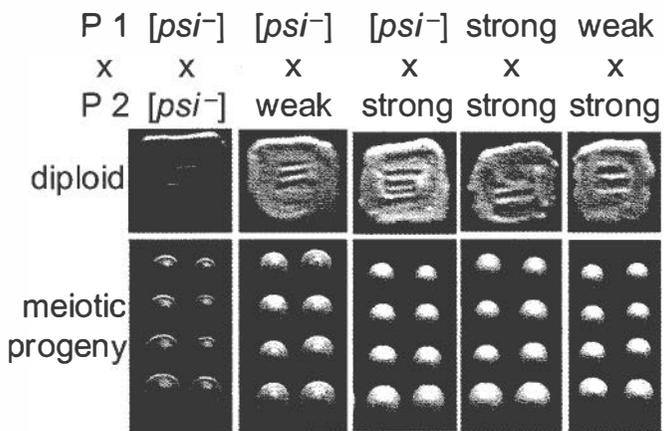
Fig. 1. Meiotic inheritance of [PSI⁺] variants. Isogenic haploid yeast, parent 1 (P1; L1842 derivatives) and parent 2 (P2; L1843 derivatives) carrying the indicated [PSI⁺] variants were mated. One representative diploid and two tetrads from each cross are shown.

each diploid were assayed for [PSI⁺] strength, mating type, and curability of [PSI⁺] by growth on YPD + GuHCl.

To perform the random spore analyses of [PSI⁺] loss frequencies, weak and strong [PSI⁺] derivatives of L1844 (L2010 and L2012, respectively) were mated to a GuHCl-treated cyh^R derivative of L1845 (L2274) to produce diploids SL-1142 and SL-1143, respectively. The resulting diploids were sporulated, and frequencies of [PSI⁺] loss were determined in at least three independent trials by counting red vs. total number of colonies after suspending cells in 100 μl of 10% gluculase, vortexing for 20 s, diluting 1 × 10⁻⁴ in water, and plating to YPD + Cyh.

Analyses of [PIN⁺] Variants. Phenotypes of [PIN⁺] variants in 74-D694, L1844, and L1845 were determined by overexpressing *SUP35* from the pEMBL-SUP35 plasmid. Transformants were patched to −Ura, where the plasmid is present in moderate copy number, and then spotted to −Ura, −Leu, and −Ade. The induction of [PSI⁺] was identified as GuHCl-curable nonsense suppression of *ade1–14* that was independent of the plasmid used to induce the appearance of [PSI⁺]. Phenotypes of [PIN⁺] variants in c10B-H49 were determined by overexpressing *SUP35* and spreading transformants to −Ade.

To determine the relative competitiveness of the [PIN⁺] variants, we first generated isogenic opposite mating type yeast carrying each of the [PIN⁺] variants. Independent [PIN⁺] derivatives of 74-D694 were cytoduced into a GuHCl-treated derivative of c10B-H49 and from there into a GuHCl-treated [pin⁻] derivative of L1844. Diploids made from crosses of the final cytoductants to a GuHCl-treated [pin⁻] cyh^R derivative of L1845 were transformed with pEMBL-SUP35 and sporulated. The [PIN⁺] phenotypes of cyh^R progeny that maintained pEMBL-SUP35 (obtained by plating sporulated cultures to CA–Ura + Cyh) were determined by growth on −Leu and −Ade. *MATα* progeny carrying the different [PIN⁺] variants were backcrossed to the original [PIN⁺] derivatives of 74-D694. Diploids were tested for the [PIN⁺] phenotypes and sporulated after propagating for approximately 28 generations. Random spores were selected on CA–Ura + Cyh, and their [PIN⁺] phenotypes were determined.

Comparison of Rnq1 Among [PIN⁺] Variants. Derivatives of 74-D694 were grown in liquid YPD to mid-log (OD$_{600}$ ≈ 1.0). Harvested cells were resuspended in lysis buffer [50 mM Tris, pH 8.0/150 mM NaCl/0.2% Triton X-100/1.9 μg/ml aprotinin/3.5 μg/ml E-64/5 μg/ml leupeptin/5 μg/ml pepstatin/400 μg/ml 1,10 phenanthroline/500 μg/ml PMSF/50 μg/ml N-(p-tosyl)lysine chloromethyl ketone (TLCK)], and mixed with 750 μl of glass beads/lysis buffer slurries. Total protein lysates were obtained by vortexing each tube eight times for 10 s, with intermittent incubations on ice, and removing cell debris at 10,000 × g for 10 min. The protein concentrations of the cleared lysates were measured (Bio-Rad Protein Assay), and lysis buffer was added to normalize the samples. Total protein lysate (1–2 mg) was fractionated at 280,000 × g for 30 min in a Sorvall TLA100.1 rotor. Pelleted proteins were resuspended in 200 μl of lysis buffer. Rnq1 was detected with a polyclonal antibody (kind gift from S. Lindquist, University of Chicago).

Influence of [PIN⁺] and [PSI⁺] on [URE3] Appearance. [PIN⁺] and [PSI⁺] derivatives from 74-D694 were cytoduced into a GuHCl-treated version of c10B-H49, and from there into a GuHCl-treated version of 3385, and finally from 3385 into a GuHCl-treated version of YHE711. The [URE3] prion induction assay was performed as described (52). Briefly, [ure-o] strains were transformed with pH317, pH376, or pH382 and transformant colonies were individually grown to saturation in SGal + Raf-Leu. Starting from 10⁷ cells per plate, serial dilutions were plated on synthetic, dextrose-based medium containing 100 μg/ml ureidosuccinate (USA). Colonies appearing after 5 days at 30°C were recorded.

Elimination of [PIN⁺] by YDJ1 Overexpression. Various [psi⁻] [PIN⁺] derivatives of 74-D694, as well as one [psi⁻] [pin⁻] control, were transformed with p901 (33) or the control plasmid pH316 (lacking YDJ1 but containing the GAL1 promoter). Transformants were grown on media containing galactose to induce YDJ1 expression. Two transformants for each strain and plasmid combination were subcloned on SGal + Raf-Leu 2–3 times consecutively by picking 3–4 same-sized colonies for each successive colony purification step. Colonies underwent an average of 20 cell doublings before being purified again or tested for [PIN⁺].

The elimination of [PIN⁺] was scored by checking for the loss of aggregated Rnq1. Colonies were patched to YPD and crossed to tester strains: GuHCl-treated versions of 64-D697 (MATα ade1–14 trp1–289 lys9-A21 leu2-3,112 ura3–52) or SL1010–1A (MATα ade1–14 met8-1 trp1-1 his5–2 leu2-1 ura3–52) transformed with pRNQ1-GFP (13). Diploids from these crosses were selected by complementation on medium (SC-His,Lys,Ura + Cu) that selects for maintenance of pRNQ1-GFP and contains 50 μM Cu to induce expression of the fusion protein. In this assay, [PIN⁺] colonies have bright green aggregates in the majority of cells, and [pin⁻] colonies have evenly distributed green fluorescence.

Results

Genetic Analysis of Strains of [PSI⁺]. To avoid confusion of yeast strains with prion strains we sometimes refer to the latter as variants. Yeast bearing strong or weak [PSI⁺] prion variants were previously distinguished (14) by the efficiency with which they could suppress the *ade1–14* nonsense allele, which contains a premature stop codon that prevents the protein from being completely translated. Yeast that are [psi⁻] do not grow on −Ade because translation termination at the premature *ade1–14* stop codon is efficient, and they are red on YPD due to the accumulation of a metabolic intermediate of the adenine biosynthesis pathway. Yeast strains with weak [PSI⁺] grow poorly on −Ade and are pink on YPD because they maintain lower levels of functional Sup35 compared with [psi⁻]; consequently, they partially suppress termination at the premature *ade1–14* stop codon and produce some functional Ade1. Yeast strains with strong [PSI⁺] grow well on −Ade and are white on YPD because they maintain even less functional Sup35 than weak [PSI⁺], and they efficiently suppress termination at the premature *ade1–14* stop codon (23, 30).

Diploids resulting from pairwise matings of isogenic yeast that were either [psi⁻], weak [PSI⁺], or strong [PSI⁺] displayed nonsense suppression levels equal to that of the [PSI⁺] parent with the "strongest" phenotype (Fig. 1). Tetrad analyses revealed that strong

Table 1. Tetrad analysis of [*PSI*⁺] variants

Parent 1	Parent 2	Tetrads dissected	Viable progeny	[*PSI*⁺] phenotypes of meiotic progeny		
				[*psi*⁻]	Weak	Strong
[*psi*⁻]	[*psi*⁻]	14	56	56	0	0
Weak	[*psi*⁻]	59	200	10	190	0
Weak	Weak	33	118	2	116	0
Strong	[*psi*⁻]	33	121	0	0	121
Strong	Strong	16	60	0	0	60
Weak	Strong	18	67	0	0	67
Strong	Weak	21	77	0	0	77

Each row represents the sum of progeny obtained from two to four independent diploids. At least one diploid was from a cross between derivatives of L1842 and L1843, and one was from a cross between derivatives of L1844 and L1845. Weak and strong refer to the [*PSI*⁺] variants.

[*PSI*⁺] diploids always segregated strong [*PSI*⁺] in a 4:0 ratio and weak [*PSI*⁺] diploids usually segregated weak [*PSI*⁺] in a 4:0 ratio (Fig. 1 and Table 1).

Meiosis Eliminates [*PSI*⁺] and [*URE3*] Prions. Approximately 2–5% of the spores from weak [*PSI*⁺] diploids were [*psi*⁻] (Table 1, rows 2 and 3), which is significantly more than the 0.7% loss of weak [*PSI*⁺] among the mitotic progeny from one of the parents of these diploids (average of ≈2,800 colonies from three independent, equally represented trials). This, together with an earlier finding that a weak [*PSI*⁺] (then called [*ETA*⁺]) was very unstable in meiosis (27), led us to investigate this phenomenon further.

Haploid cells did not exhibit an enhanced loss of weak [*PSI*⁺] when exposed to the same conditions that induce meiosis in an isogenic weak [*PSI*⁺] diploid. We incubated three yeast strains on sporulation medium: a weak [*PSI*⁺] haploid, L2010; a weak [*PSI*⁺] diploid isogenic to L2010 but heterozygous for *cyh^R*, SL-1142; and a weak [*PSI*⁺] *cyh^R* meiotic segregant from this diploid, SL1142–1A. Random spores were selected from the diploid culture on YPD + Cyh. SL1142–1A was also plated on this medium, and L2010 was plated on YPD lacking cycloheximide. The frequency of [*psi*⁻] among the random spore colonies was 5.7% (average of ≈2,400 colonies from three independent, equally represented trials). The frequency of [*psi*⁻] among mitotic colonies from either of the haploid controls was only 0.4% (7 [*psi*⁻] of 1,670 colonies from L2010, and 15 [*psi*⁻] of 3,530 colonies from SL1142–1A). Therefore, the observed effect was not a result of the conditions used to induce sporulation. The effect was also not due to heightened instability in the diploid phase because the frequency of weak [*PSI*⁺] loss from mitotic diploid progeny was only 0.07% (average of ≈3,450 colonies from three independent, equally represented trials). Thus, some aspect of meiosis interferes with the inheritance of [*PSI*⁺].

Likewise, although [*URE3*] is highly stable during mitotic growth, it is frequently lost in meiotic segregants (4, 9, 40). We examined the effect of meiosis-inducing conditions on the stability of [*URE3*]. [*URE3*] was stable in mitotic growth: it was efficiently cytoduced from strain 1735 to 1019 (16 of 16 cytoductants examined were USA⁺). Growing 1735 on sporulation medium did not decrease the stability of [*URE3*]. Furthermore, a diploid made by crossing 1735 with 1019 stably maintained [*URE3*] (100 of 100 colonies examined after growth on YPD were USA⁺). However, sporulation of this diploid produced mostly USA⁻ spores (39 of 48 USA⁻ spores). Thus, the process of meiosis causes loss of [*URE3*].

[*PIN*⁺] Variants. We previously described the isolation of spontaneously appearing [*PIN*⁺] colonies after prolonged incubation of a [*pin*⁻] [*psi*⁻] derivative of 74-D694 (11). Rare [*PIN*⁺] cells were detected by selecting for the appearance of [*PSI*⁺] after overex-
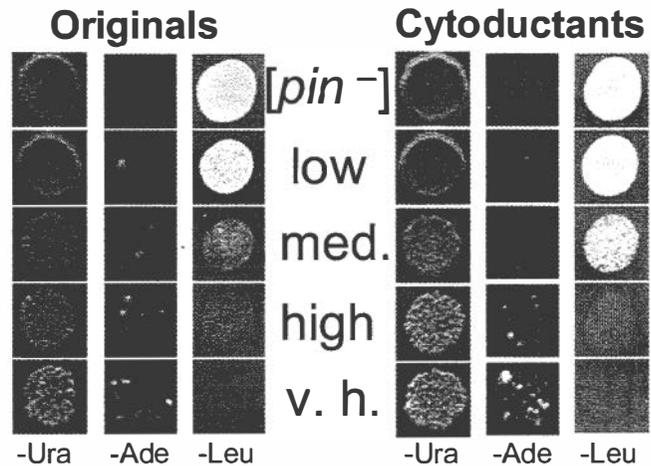

Fig. 2. Characterization of [*PIN*⁺] variants and their inheritance through cytoduction. Independent derivatives of the original 74-D694 are shown (Originals). These were each cytoduced into a [*pin*⁻] version of c10B-H49, a *kar1–1* yeast strain, and from there back into a [*pin*⁻] version of 74-D694 (Cytoductants). Both the originals and cytoductants carry the pEMBL-SUP35 plasmid. When the plasmid is maintained at moderate level on −Ura distinct levels of [*PSI*⁺] induction are observed on transfer to −Ade. When the plasmid is amplified on −Leu, different levels of growth inhibition are observed. The different [*PIN*⁺] variants are cytoplasmically inherited because cytoductants and donors exhibit the same [*PIN*⁺] phenotypes. Row 1 is the [*pin*⁻] control; rows 2, 3, and 5 are the low, medium (med.), and very high (v.h.) spontaneous [*PIN*⁺] variants obtained in [*pin*⁻] 74-D694; row four is high [*PIN*⁺] from the original 74-D694.

pression of *SUP35*. We eliminated [*PSI*⁺], but not [*PIN*⁺], by overexpressing *HSP104*. Later, the [*PIN*⁺] status of these isolates was shown to be a consequence of the prion form of Rnq1 because the loss of [*PIN*⁺] correlated with the loss of Rnq1 aggregates (13).

The phenotypes of [*PIN*⁺] were originally described as allowing moderate overproduction of Sup35 to convert [*psi*⁻] cells to [*PSI*⁺] and as inhibiting growth in the presence of extreme overproduction of Sup35 (10). We now distinguish [*PIN*⁺] variants with different levels (low, medium, high, and very high) of [*PSI*⁺] induction and growth inhibition (Fig. 2).

Because [*PIN*⁺] is cytoducible (12, 13), cytoduction should transfer the distinct phenotypes if they result from heritable [*PIN*⁺] variants, but not if they are the result of Mendelian modifier mutations. Derivatives of 74-D694 carrying the different [*PIN*⁺] isolates and a [*pin*⁻] control were cytoduced into a [*pin*⁻] derivative of c10B-H49, and from there back into a [*pin*⁻] derivative of 74-D694. The cytoductants displayed the donor's [*PIN*⁺] phenotypes (Fig. 2). This result proves that Mendelian mutations do not cause the variable phenotypes.

Genetic Analysis of Variants of [*PIN*⁺]. To determine the relative competitiveness of the different [*PIN*⁺] variants, we crossed pairs of opposite mating type yeast harboring each of the different [*PIN*⁺] variants in all possible combinations and determined the [*PIN*⁺] phenotypes of the resulting diploids by measuring [*PSI*⁺] induction and growth inhibition levels. High [*PIN*⁺] outcompetes low and medium [*PIN*⁺] (data not shown), and medium [*PIN*⁺] outcompetes low (Fig. 3); however, very high [*PIN*⁺] was outcompeted by high, medium, and low [*PIN*⁺] (Fig. 3 and data not shown). Therefore, the winner in these competitions is not always the one with the "strongest" [*PIN*⁺] phenotype. Meiotic progeny always exhibited the same phenotype as the diploid parent (Table 2). This result was true even when two different [*PIN*⁺] variants were crossed. For example, crosses between low [*PIN*⁺] and very high [*PIN*⁺] produced low [*PIN*⁺] diploids whose meiotic progeny always inherited low [*PIN*⁺].

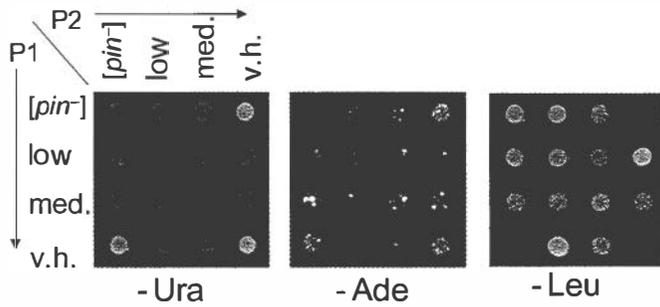Because very high [*PIN*⁺] was outcompeted by low, medium, and

Fig. 3. Genetic analysis of variants of [PIN+]. Independent diploids carrying the pEMBL-SUP35 plasmid reveal the outcome of crosses between haploid parents (P1 and P2) carrying the [pin−], low, medium (med.), or very high (v.h.) [PIN+] variants. The [PIN+] phenotypes are scored as in Fig. 2.
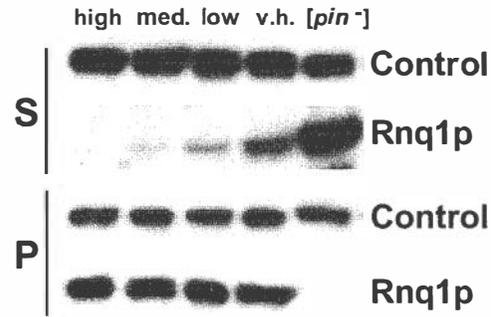

Fig. 4. Comparison of levels of soluble and aggregated Rnq1 among [PIN+] variants. Lysates were fractionated into soluble (S) and pellet (P) fractions by ultra-centrifugation. Rnq1 was detected by Western blotting with polyclonal Rnq1 antibody (kind gift from S. Lindquist). Stripped blots were then re-probed with monoclonal Sup35 antibody (Control) as a loading control. Note, the soluble Rnq1 in this figure were exposed twice as long as the Rnq1 in the pellet. The gradient depicted was generally reproducible; however, in three of eight independent protein isolations, the level of soluble Rnq1 in the v.h. and low variants appeared similar.

high [PIN+], it was possible that the latter phenotypes were caused by a combination of [PIN+] and an additional "modifier" prion distinct from [PIN+]. In this case, diploids formed from crosses of very high [PIN+] to, e.g., low [PIN+] would contain both the [PIN+] prion and the modifier prion, resulting in the low [PIN+] phenotype. To test this possibility, we cytoduced low and medium [PIN+] derivatives of c10B-H49 into a Δrnq1 derivative of BY4741 that, while unable to maintain [PIN+], should be able to maintain other cytoduced modifiers. The Δrnq1 recipient was then cytoduced into a very high [PIN+] derivative of c10B-H49 to test whether the hypothesized prion modifier would convert very high into low or medium [PIN+]. Because most cytoductants remained very high [PIN+] regardless of whether [pin−] control (11 of 12 independent cytoductants) or presumptive modifier-containing donor cytoplasm (11 of 12 independent cytoductants) was used, the data do not support the prion modifier hypothesis. In control experiments, we cytoduced weak or medium [PIN+] derivatives of BY4741 into a very high [PIN+] derivative of c10B-H49 and found that the majority of the recipients (10 of 15) were indeed converted into the phenotype of the donor (low or medium). Thus, it appears that the distinct phenotypes result from heritable differences in the [PIN+] aggregates themselves.

Comparison of Rnq1 Aggregation Among [PIN+] Variants. We compared the amounts of soluble and aggregated Rnq1 in the high, medium, low and very high [PIN+] variants. Each maintained indistinguishable amounts of aggregated Rnq1, but the levels of soluble Rnq1 showed strain-specific differences (Fig. 4). Strikingly, the hierarchy of [PIN+] variants determined by the competition

experiments described above exactly corresponded to the gradient, from least to most, of soluble Rnq1 exhibited by these strains. For example, the variant with the most soluble Rnq1, very high, was lost when crossed with each of the other [PIN+] variants, whereas the variant with the least soluble Rnq1, high, outcompeted all of the other [PIN+] variants.

Effect of [PIN+] and [PSI+] on [URE3] Appearance. Because [PIN+] facilitates the appearance of [PSI+] (10, 11), we asked whether it has a similar effect on the appearance of another prion, [URE3]. We compared the frequency with which overexpression of URE2 can induce the appearance of [URE3] in the [PIN+] yeast strain YHE711 and a [pin−] derivative of this strain obtained by growth on GuHCl. The original YHE711 strain had enhanced [URE3] induction relative to the [pin−] derivative, which had a drastically reduced frequency of [URE3] appearance (Table 3). Because deleting RNQ1 from the original [PIN+] strain also abolished the ability to induce [URE3], YHE711 does not harbor other elements in addition to [PIN+] that independently allow for [URE3] induction (Table 3). Clearly, [PIN+] facilitates [URE3] appearance. We further showed that [PIN+] did not cause greater overproduction of Ure2, nor did it stabilize newly appearing [URE3] (data not shown).

To determine whether the different variants of [PIN+] that were characterized above in terms of their distinct [PSI+] induction frequencies could also be distinguished on the basis of their effect on [URE3] induction, we cytoduced several different variants of [PIN+] into a GuHCl-treated derivative of YHE711. Surprisingly, the original YHE711 [PIN+] facilitated [URE3] appearance more efficiently than even the very high [PIN+] (Table 3), yet the original YHE711 [PIN+] was less efficient than very high [PIN+] in the Sup35-based [PSI+] induction phenotypes (data not shown). Whereas the presence of the low, medium, or very high [PIN+] elements enhanced the frequency of [URE3] induction relative to the [pin−] control, we could not consistently observe any differences in [URE3] induction levels among strains with the low, medium and very high [PIN+] variants (Table 3).

Although it generally appears that prions enhance the appearance of other prions (13), we find here that [PSI+] does not facilitate [URE3] induction, but rather inhibits its appearance. By overexpressing the URE2Δ151–158 allele, which efficiently induces [URE3] (15, 54), we found that a [PSI+] [pin−] derivative of YHE711 is less inducible to [URE3] compared with a [psi−] [pin−] derivative (averages of 346 ± 107 vs. 989 ± 394 [URE3]/10^6 cells from six independent experiments).

Table 2. Meiotic inheritance of [PIN+] variants

Parent 1	Parent 2	[PIN+] phenotypes of meiotic progeny				
		Total	Low	Med.	V.H.	[pin−]
[pin−]	[pin−]	26	0	0	0	26
[pin−]	Low	30	29	0	0	1
[pin−]	Med.	29	0	29	0	0
[pin−]	V.H.	30	0	0	30	0
Low	Low	30	29	0	0	1
Med.	Med.	30	0	30	0	0
V.H.	V.H.	30	0	0	30	1
Low	V.H.	62	60	0	0	2
Med.	V.H.	60	0	60	0	0
Med.	Low	60	0	58	0	2

[PIN+] phenotypes were scored using the SUP35 overexpression assays shown in Fig. 2. For each row, ≈10 progeny were obtained from 3 or 6 independent diploids. Low, medium (Med.), and very high (V.H.) variants of [PIN+] were used.

Table 3. Influence of [PIN+] on [URE3] appearance

| Experiment | [pin−] | Δrnq1 | Frequency of [URE3] as % of [pin−] control | | | |
| | | | [PIN+] Variants | | | |
			Original	Very high	Medium	Low
1	100 ± 59		4,600 ± 1,100			
2	100 ± 49		4,300 ± 2,400			
3	100 ± 50		37,000 ± 12,500	3,500 ± 1,200		
4	100 ± 39		9,200 ± 4,600	550 ± 80		
5*	100	19	23,000	3,500		
6*	100	23	1,100	1,100		
7*	100			830	650	720
8	100 ± 50			800 ± 360	2,100 ± 1,100	1,200 ± 650
9	100 ± 142		15,00 ± 6,400	2,800 ± 1,300	1,000 ± 1,100	
10	100 ± 143		39,000 ± 19,000	510 ± 230	5,700 ± 3,800	
Average	100	21	16,700	1,700	2,400	960

Full-length Ure2 was overproduced in derivatives of YHE711. The frequency of [URE3] colonies appearing in the [pin−] strain per 10^6 cells plated is normalized to 100, and other values were normalized accordingly. The YHE711 derivatives were: GuHCl-treated ([pin−]), RNQ1 deletion (Δrnq1), not GuHCl-treated (Original), GuHCl-treated and cytoduced with very high, medium, or low [PIN+]. Averages and standard deviations are shown when three or more transformants were assayed. Induction of [URE3] by overproducing just the Ure2 prion domain (1-65 aa) was also facilitated by the presence of [PIN+] (data not shown). Each experiment also included controls in which URE2 was not overexpressed, where the numbers of [URE3] colonies was very low (not shown).
*Indicates that only two measurements were performed, the average is shown.

Effect of YDJ1 Overexpression on [PIN+]. Various [psi−] [PIN+] derivatives of 74-D694, as well as a [psi−] [pin−] control, were transformed with p901, carrying YDJ1 under the control of the GAL1 promoter. Transformants were grown on media containing galactose to induce YDJ1 expression, after which Rnq1 aggregation was used to detect [PIN+]. Some [PIN+] variants were readily cured by overproducing Ydj1, whereas other [PIN+] variants were not cured (Table 4). [PIN+] was never lost in control experiments by using the pH316 empty vector (Table 4). The experiment gave similar results when repeated with two of the curable and three of the incurable [PIN+] variants (Table 4). Thus, we find that over-expressing YDJ1 promotes the loss of some [PIN+] variants.

Discussion

According to the protein-only prion model, two explanations of the prion strain phenomena are possible. Prion variants may result from inherent flexibility of the tertiary structure allowing one chain of amino acids to have two or more self-perpetuating conformations that are stably inherited. Alternatively, variants might result from a single tertiary conformation arranged into two or more different quaternary arrangements that are stable and self-perpetuating. Here, the existence of distinct heritable variants of the [PIN+] prion is described, and the phenotypic differences between them are shown not to be due to either nuclear or cytoplasmic modifiers. It is now clear that each of the well characterized yeast prions—[PSI+], [URE3], and [PIN+]—can exist as different distinct heritable variants.

Although different prions, e.g., [PSI+] and [PIN+], or [PSI+] and [URE3], can be maintained together in a single cell (11, 13, 16), such coexistence may not be possible for two variants of the same prion. Indeed, diploids formed by mating cells with different [PSI+] variants that could be distinguished from each other by Sup35-GFP staining retained only one of the [PSI+] variants (25). Furthermore, if prion variants could coexist, one might expect diploids formed from crosses between isogenic cells bearing weak and strong [PSI+] to exhibit a phenotype more extreme than strong [PSI+], and to occasionally segregate out both strong and weak [PSI+] in mitotic and meiotic growth. We show here that this is not the case. Rather, except for rare cases of loss of [PSI+], the diploids and all mitotic and meiotic progeny were indistinguishable from the strong [PSI+] parent. Furthermore, weak [PSI+] did not emerge from the diploids created by crossing weak and strong [PSI+] even after stimulation of [PSI+] loss by short-term growth in medium containing GuHCl (M.E.B. and S.W.L., unpublished work).

Although one could argue that strong and weak [PSI+] coexist, but that the phenotype of strong [PSI+] cannot be made any stronger, similar results obtained for crosses between cells bearing the different [PIN+] variants cannot easily be explained in this way. This result is because, unlike crosses between weak and strong [PSI+], where the variant with the strongest phenotype prevails, [PIN+] variants with less dramatic phenotypes (low, medium, and high [PIN+]) prevailed over very high [PIN+], which has the most dramatic phenotype. Thus, if the different [PIN+] variants coexisted, the very high [PIN+] phenotype could not reasonably be expected to be hidden by the [PIN+] variants with milder phenotypes.

Because two variants of the same prion compete for the same pool of newly synthesized protein to reproduce and be heritable, faster replicating variants should eventually outcompete slower or less stable variants by starving them for convertible protein (Fig. 5). Indeed we found that [PIN+] variants, which maintain little soluble Rnq1 but abundant aggregated Rnq1, indicating fast reproduction,

Table 4. YDJ1 cures some [PIN+] variants

| | Expt. 1 | | | | Expt. 2 | | | |
| | Control | | YDJ1 | | Control | | YDJ1 | |
Strain	[pin−]	[PIN+]	[pin−]	[PIN+]	[pin−]	[PIN+]	[pin−]	[PIN+]
[pin−]	24	0	24	0	35	0	48	0
V.H.	0	24	0	24				
Low	0	24	0	24				
Med.	0	24	8	16				
High	0	24	2	22	0	52	0	64
L1941	0	24	3	21				
L1952	0	24	14	10	0	61	29	32
L1947	0	24	22	2	0	66	11	56
L1949	0	24	0	24	0	49	4	62
L1956	0	24	0	24	0	46	0	66

Strains are derivatives of 74-D694. The very high (V.H.), low, medium (Med.), and high [PIN+] variants are indicated as such. Other independent [PIN+] isolates are also indicated (L1941–1956). Colonies checked from the control (pH316) and YDJ1 bearing (p901) transformants are shown.
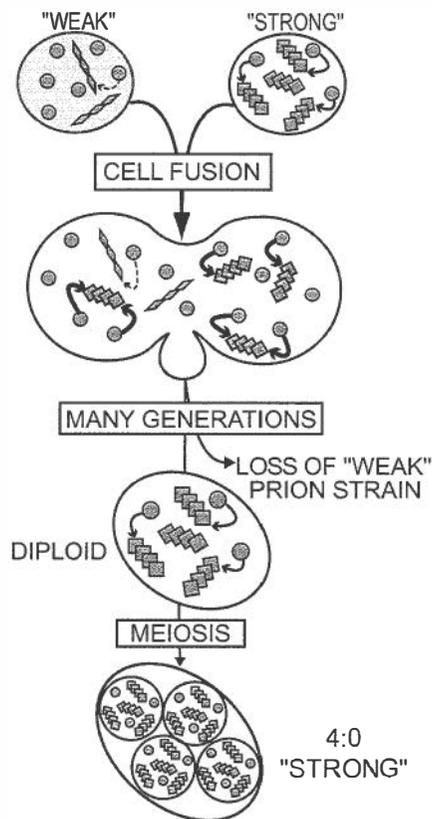
Fig. 5. Cartoon depicting a competition for soluble protein between variants of the same prion. Haploids carrying soluble protein (free circles) and aggregated-prion protein as either "weak" (rhomboids) or "strong" (squares) variants are mated. Mitotic growth of the diploid results in loss of the weak variant, presumably because it is starved for convertible soluble protein. When the diploid sporulates, each spore inherits the strong prion variant.

Fig. 6. Models illustrating the different seeding preferences proposed for two [PIN+] variants. (A) The very high [PIN+] variant inefficiently seeds Ure2 (red circles), but efficiently seeds Sup35 (green triangles) converting them into the [PSI+] shape. (B) The original YHE711 [PIN+] variant inefficiently seeds Sup35, but efficiently seeds Ure2, converting them into the [URE3] shape. Both [PIN+] variants propagate their forms by converting Rnq1 (blue rectangles) with the highest efficiency.

did take over when combined with [PIN+] variants that maintain more soluble Rnq1 and less abundant aggregated Rnq1. Previous observations indicate that [PSI+] variants show a similar pattern (23, 24).

Recent work has demonstrated that certain prions facilitate the appearance of other prions: [PIN+] and [URE3] (13), or the artificial fusion protein prion [NU+] (16), permit overexpression of SUP35 to induce the appearance of [PSI+]; and [PSI+] and [URE3] facilitate the appearance of [PIN+] (13). Here, we show that [PIN+] facilitates, but [PSI+] inhibits, the de novo appearance of [URE3].

Two mechanisms of prion-facilitated prion appearance have been proposed (13, 16). According to the seeding model, heterologous prions provide a template for initial cross-seeding of a de novo prion aggregate. The titration model hypothesizes that pre-existing heterologous prion aggregates sequester a protein that normally inhibits prion appearance, thereby allowing other prions to appear more easily. Our current finding of no correlation between the efficiencies with which the different [PIN+] variants promote the appearance of [PSI+] and the efficiencies with which they promote the appearance of [URE3] can most easily be explained by the cross-seeding model. We propose that some [PIN+] variants cross-seed Sup35 better than Ure2, whereas others exhibit the opposite preference (Fig. 6). More complicated scenarios involve combinations of the seeding and titration models, or multiple inhibitors with distinct binding properties.

The [PIN+] variants described here cannot easily be distinguished by the amounts of aggregated Rnq1 (Fig. 4). There is also no correlation between the levels of soluble Rnq1 and the phenotypes of the [PIN+] variants: the order of increasing soluble Rnq1 levels

is high, medium, low, then very high. One possibility to explain this conundrum is that different prion conformations of Rnq1 are better at influencing [PSI+] appearance, and these conformations are only coincidentally distinguishable by soluble Rnq1 levels. Another possibility is that accessory proteins, such as Sis1 (34), are associated in different amounts with each of the [PIN+] variants. The presence of such proteins may hinder the action of [PIN+] or, if these are chaperone proteins, they might be essential for creating the action of [PIN+]. It is also possible that other variants of [PIN+] that do not facilitate the induction of [PSI+] or [URE3] may exist.

We have unexpectedly found that the presence of one prion can inhibit the de novo appearance of another, because [PSI+] inhibited the appearance of [URE3]. Whereas the effect of [PSI+] on the induction of [URE3] is inhibitory, it still suggests that heterologous prions interact. [PSI+] may inhibit de novo [URE3] appearance by occasionally joining and "poisoning" [URE3] seeds thereby inhibiting [URE3] propagation as previously proposed to explain the [PSI+] curing effect of certain SUP35 mutants and the [URE3] curing effect of URE2-GFP (55–57). Alternatively, [PSI+] may sequester Ure2Δ151–158 thereby reducing the amount of protein available to form [URE3] seeds. That [PSI+] stimulates [PIN+] appearance but inhibits [URE3] appearance is inconsistent with the inhibitor model and more compatible with the idea that heterologous prions directly interact—sometimes causing cross-seeding and sometimes causing inhibition.

Prions are stable, heritable elements. Yeast cells should therefore acquire multiple different prions by mating with other cells—even though the prions may be disadvantageous under some circumstances. Our finding that prions are occasionally eliminated by meiosis may be an indication that yeast have evolved mechanisms of ridding themselves of prions. Several possibilities might explain how prions are eliminated by meiosis. Because overexpressing HSP104 is known to eliminate [PSI+] (32), the elevated HSP104 levels associated with sporulating cultures (58) might disrupt the inheritance of [PSI+]. Alternatively, because meiotic progeny re-

ceive less cytoplasm than mitotic daughter cells (59), they might lose [*PSI*⁺] more frequently as a result of inheriting fewer [*PSI*⁺] seeds. Because most Ure2 amyloid filaments are found in a single cytoplasmic network in [*URE3*] cells (60), it is not surprising that [*URE3*] is often lost in meiosis.

Overexpression of *YDJ1,* which interacts with Hsp104 and Hsp70 to rescue denatured proteins (35), was previously shown to cause the loss of a [*URE3*] variant (33). Here, we show that overexpressing *YDJ1* eliminates some, but not other, [*PIN*⁺] variants. [*URE3*] and [*PIN*⁺] are both eliminated by deletion, but not overexpression, of *HSP104* (10, 33, 34). The spontaneous [*PIN*⁺] variants described in this paper, including those that were not cured by Ydj1 overexpression, were cured by deleting *HSP104* (M.E.B. and S.W.L., unpublished work). Possibly, *YDJ1* may cure some [*PIN*⁺] variants by sequestering Hsp104 (33), but other [*PIN*⁺] variants may be less sensitive to the reduction of Hsp104. Alternatively, Ydj1 may stimulate protein refolding and therefore cure [*PIN*⁺] by disaggregating it. If this result were true, it could mean that some [*PIN*⁺] variants maintain too many seeds, or replicate too quickly to be eliminated by overexpressing just Ydj1. Possibly overexpressing Hsp104 or Hsp70 along with Ydj1 would eliminate some of these incurable [*PIN*⁺] variants. Indeed, overexpression of Ydj1 together with Ssa1 or Ssb1 did cause loss of a weak [*PSI*⁺] (26).

Our results show that [*PIN*⁺] prions exhibit strain variation, as do [*PSI*⁺] and [*URE3*] (14, 22), further supporting the hypothesis that strain variation is compatible with the protein-only model for prions. The recent suggestions that prions affect the appearance of other prions (13, 16) and that prion strain variation and species barrier are related phenomena (29) indicate that elucidating the molecular basis of [*PIN*⁺] strain variation may yield important clues about the transmissibility of prion diseases between different species.

1. Prusiner, S. B., Scott, M. R., DeArmond, S. J. & Cohen, F. E. (1998) *Cell* **93,** 337–348.
2. Griffith, J. S. (1967) *Nature (London)* **215,** 1043–1044.
3. Prusiner, S. B. (1982) *Science* **216,** 136–144.
4. Wickner, R. B. (1994) *Science* **264,** 566–569.
5. Cox, B. S. (1965) *Heredity* **20,** 505–521.
6. Liebman, S. W. & Derkatch, I. L. (1999) *J. Biol. Chem.* **274,** 1181–1184.
7. Serio, T. R. & Lindquist, S. L. (1999) *Annu. Rev. Cell. Dev. Biol.* **15,** 661–703.
8. Wickner, R. B., Taylor, K. L., Edskes, H. K., Maddelein, M. L., Moriyama, H. & Roberts, B. T. (2001) *Adv. Protein Chem.* **57,** 313–334.
9. Lacroute, F. (1971) *J. Bacteriol.* **106,** 519–522.
10. Derkatch, I. L., Bradley, M. E., Zhou, P., Chernoff, Y. O. & Liebman, S. W. (1997) *Genetics* **147,** 507–519.
11. Derkatch, I. L., Bradley, M. E., Masse, S. V., Zadorsky, S. P., Polozkov, G. V., Inge-Vechtomov, S. G. & Liebman, S. W. (2000) *EMBO J.* **19,** 1942–1952.
12. Sondheimer, N. & Lindquist, S. (2000) *Mol. Cell* **5,** 163–172.
13. Derkatch, I. L., Bradley, M. E., Hong, J. Y. & Liebman, S. W. (2001) *Cell* **106,** 171–182.
14. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. (1996) *Genetics* **144,** 1375–1386.
15. Masison, D. C. & Wickner, R. B. (1995) *Science* **270,** 93–95.
16. Osherovich, L. Z. & Weissman, J. S. (2001) *Cell* **106,** 183–194.
17. Mabbott, N. A. & Bruce, M. E. (2001) *J. Gen. Virol.* **82,** 2307–2318.
18. Bessen, R. A., Kocisko, D. A., Raymond, G. J., Nandan, S., Lansbury, P. T. & Caughey, B. (1995) *Nature (London)* **375,** 698–700.
19. Caughey, B., Kocisko, D. A., Raymond, G. J. & Lansbury, P. T. (1995) *Chem. Biol.* **2,** 807–817.
20. Safar, J., Wille, H., Itri, V., Groth, D., Serban, H., Torchia, M., Cohen, F. E. & Prusiner, S. B. (1998) *Nat. Med.* **4,** 1157–1165.
21. Mestel, R. (1996) *Science* **273,** 184–189.
22. Schlumpberger, M., Prusiner, S. B. & Herskowitz, I. (2001) *Mol. Cell. Biol.* **21,** 7035–7046.
23. Zhou, P., Derkatch, I. L., Uptain, S. M., Patino, M. M., Lindquist, S. & Liebman, S. W. (1999) *EMBO J.* **18,** 1182–1191.
24. Derkatch, I. L., Bradley, M. E., Zhou, P. & Liebman, S. W. (1999) *Curr. Genet.* **35,** 59–67.
25. King, C. Y. (2001) *J. Mol. Biol.* **307,** 1247–1260.
26. Kushnirov, V. V., Kryndushkin, D. S., Boguta, M., Smirnov, V. N. & Ter-Avanesyan, M. D. (2000) *Curr. Biol.* **10,** 1443–1446.
27. Liebman, S. W. & All-Robyn, J. A. (1984) *Curr. Genet.* **8,** 567–573.
28. Glover, J. R., Kowal, A. S., Schirmer, E. C., Patino, M. M., Liu, J. J. & Lindquist, S. (1997) *Cell* **89,** 811–819.
29. Chien, P. & Weissman, J. S. (2001) *Nature (London)* **410,** 223–227.
30. Uptain, S. M., Sawicki, G. J., Caughey, B. & Lindquist, S. (2001) *EMBO J.* **20,** 6236–6245.
31. Kochneva-Pervukhova, N. V., Chechenova, M. B., Valouev, I. A., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. (2001) *Yeast* **18,** 489–497.
32. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. (1995) *Science* **268,** 880–884.
33. Moriyama, H., Edskes, H. K. & Wickner, R. B. (2000) *Mol. Cell. Biol.* **20,** 8916–8922.
34. Sondheimer, N., Lopez, N., Craig, E. A. & Lindquist, S. (2001) *EMBO J.* **20,** 2435–2442.
35. Glover, J. R. & Lindquist, S. (1998) *Cell* **94,** 73–82.
36. Parsell, D. A., Kowal, A. S., Singer, M. A. & Lindquist, S. (1994) *Nature (London)* **372,** 475–478.
37. Chernoff, Y. O., Newnam, G. P., Kumar, J., Allen, K. & Zink, A. D. (1999) *Mol. Cell. Biol.* **19,** 8103–8112.
38. Newnam, G. P., Wegrzyn, R. D., Lindquist, S. L. & Chernoff, Y. O. (1999) *Mol. Cell. Biol.* **19,** 1325–1333.
39. Jung, G., Jones, G., Wegrzyn, R. D. & Masison, D. C. (2000) *Genetics* **156,** 559–570.
40. Aigle, M. & Lacroute, F. (1975) *Mol. Gen. Genet.* **136,** 327–335.
41. Sherman, F., Fink, G. R. & Hicks, J. B. (1986) *Methods in Yeast Genetics* (Cold Spring Harbor Lab. Press, Plainview, NY).
42. Tuite, M. F., Mundy, C. R. & Cox, B. S. (1981) *Genetics* **98,** 691–711.
43. Goldring, E. S., Grossman, L. I., Krupnick, D., Cryer, D. R. & Marmur, J. (1970) *J. Mol. Biol.* **52,** 323–335.
44. Chernoff, Y. O., Derkach, I. L. & Inge-Vechtomov, S. G. (1993) *Curr. Genet.* **24,** 268–270.
45. Kane, S. M. & Roth, R. (1974) *J. Bacteriol.* **118,** 8–14.
46. Herskowitz, I. & Jensen, R. E. (1991) *Methods Enzymol.* **194,** 132–146.
47. Edskes, H. K., Hanover, J. A. & Wickner, R. B. (1999) *Genetics* **153,** 585–594.
48. Zhou, P., Derkatch, I. L. & Liebman, S. W. (2001) *Mol. Microbiol.* **39,** 37–46.
49. Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J. D., Bussey, H., et al. (1999) *Science* **285,** 901–906.
50. Kochneva-Pervukhova, N. V., Poznyakovski, A. I., Smirnov, V. N. & Ter-Avanesyan, M. D. (1998) *Curr. Genet.* **34,** 146–151.
51. Ter-Avanesyan, M. D., Kushnirov, V. V., Dagkesamanskaya, A. R., Didichenko, S. A., Chernoff, Y. O., Inge-Vechtomov, S. G. & Smirnov, V. N. (1993) *Mol. Microbiol.* **7,** 683–692.
52. Edskes, H. K. & Wickner, R. B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 6625–6629.
53. Conde, J. & Fink, G. R. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 3651–3655.
54. Maddelein, M. L. & Wickner, R. B. (1999) *Mol. Cell. Biol.* **19,** 4516–4524.
55. Kochneva-Pervukhova, N. V., Paushkin, S. V., Kushnirov, V. V., Cox, B. S., Tuite, M. F. & Ter-Avanesyan, M. D. (1998) *EMBO J.* **17,** 5805–5810.
56. DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. (1998) *Cell* **93,** 1241–1252.
57. Edskes, H. K., Gray, V. T. & Wickner, R. B. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 1498–1503.
58. Sanchez, Y., Taulien, J., Borkovich, K. A. & Lindquist, S. (1992) *EMBO J.* **11,** 2357–2364.
59. Brewer, B. J. & Fangman, W. L. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5380–5384.
60. Speransky, V. V., Taylor, K. L., Edskes, H. K., Wickner, R. B. & Steven, A. C. (2001) *J. Cell. Biol.* **153,** 1327–1336.

Colloquium

Identification of benzothiazoles as potential polyglutamine aggregation inhibitors of Huntington's disease by using an automated filter retardation assay

Volker Heiser*†, Sabine Engemann*‡, Wolfgang Bröcker*†, Ilona Dunkel*, Annett Boeddrich*‡, Stephanie Waelter*‡, Eddi Nordhoff*†, Rudi Lurz*, Nancy Schugardt*‡, Susanne Rautenberg*‡, Christian Herhaus§, Gerhard Barnickel§, Henning Böttcher§, Hans Lehrach*, and Erich E. Wanker*‡¶

*Max-Planck-Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany; and §Merck KGaA, Frankfurter Strasse 250, D-64271 Darmstadt, Germany

Preventing the formation of insoluble polyglutamine containing protein aggregates in neurons may represent an attractive therapeutic strategy to ameliorate Huntington's disease (HD). Therefore, the ability to screen for small molecules that suppress the self-assembly of huntingtin would have potential clinical and significant research applications. We have developed an automated filter retardation assay for the rapid identification of chemical compounds that prevent HD exon 1 protein aggregation *in vitro*. Using this method, a total of 25 benzothiazole derivatives that inhibit huntingtin fibrillogenesis in a dose-dependent manner were discovered from a library of ≈184,000 small molecules. The results obtained by the filter assay were confirmed by immunoblotting, electron microscopy, and mass spectrometry. Furthermore, cell culture studies revealed that 2-amino-4,7-dimethyl-benzothiazol-6-ol, a chemical compound similar to riluzole, significantly inhibits HD exon 1 aggregation *in vivo*. These findings may provide the basis for a new therapeutic approach to prevent the accumulation of insoluble protein aggregates in Huntington's disease and related glutamine repeat disorders.

Huntington's disease (HD) is a progressive neurodegenerative disorder with no effective treatment (1). The disease is caused by an elongated CAG trinucleotide repeat expansion located within exon 1 of the *IT-15* gene encoding huntingtin, an ≈350-kDa protein of unknown function. The CAG repeat is translated into a polyglutamine (polyQ) stretch. In HD patients, huntingtin is expressed with 38–180 glutamine residues, whereas in healthy individuals the protein is synthesized with 8–37 glutamine residues (2, 3). Thus, the disease develops when a critical length of about 37 glutamine residues (pathological threshold) is exceeded, whereas a polyQ tract of fewer than 37 glutamine residues is tolerated by neuronal cells (4).

The accumulation of ubiquitinated polyQ-containing protein aggregates in neuronal inclusions is a pathological hallmark of HD and related glutamine repeat disorders (5). Whether the formation of huntingtin aggregates in brain is the cause, or merely the consequence, of disease, however, is still unclear (6). Within the last few years, aggregation of polyQ-containing proteins has been reproduced in various *in vitro* and *in vivo* model systems (7–11). Evidence has been presented that the process of aggregate formation is causally linked with disease progression. Recently, Yamamoto *et al.* (12) have shown that blockage of *HD exon 1* expression in symptomatic transgenic mice results in disappearance of insoluble protein aggregates, as well as motor dysfunction, suggesting that protein aggregation *in vivo* is associated with disease progression.

The formation of insoluble polyQ-containing protein aggregates *in vitro* and in cell culture model systems has been inhibited by specific antibodies (13), peptides (14), heat shock proteins (15, 16), and chemical compounds (17). Moreover, antibodies (18), β-sheet breaker peptides (19), and chemical compounds (20–24) that prevent protein aggregation in various models of Alzheimer's and prion diseases have been described. However, the effect of these and other potential therapeutic molecules on the progression of neurodegenerative disorders in humans needs to be evaluated.

In this study, we have developed an automated filter retardation assay for the identification of HD exon 1 aggregation inhibitors. In contrast to conventional methods, our assay is eminently suited for high-throughput detection of chemical compounds that block huntingtin fibrillogenesis because it allows the parallel, nonradioactive screening of up to 384 potential aggregation inhibitors on a single filter membrane. Here, we report the rapid screening and characterization of a previously uncharacterized class of polyglutamine aggregation inhibitors from large chemical compound libraries by using this approach.

Materials and Methods

Chemical Compounds, Enzymes, and Instrumentation. A library containing 184,880 chemical compounds was provided by Merck. All compounds were dissolved in 100% DMSO at a conc. of 3 mM. Riluzole was purchased from BioTrend (Cologne, Germany). Doxycycline and elastase were obtained from Sigma-Aldrich, and trypsin (modified version) was purchased from La Roche (Mannheim, Germany). The Cy3-labeled donkey anti-rabbit IgG was obtained from Jackson ImmunoResearch. Parallel liquid handling in 96-and 384-well format plates was performed with Tomtec (Orange, CT) Quadra 96SV and Quadra 384S robots.

Construction of Plasmids and Protein Purification. Standard protocols for DNA manipulations were followed. *Escherichia coli* SURE (stop unwanted rearrangement events; Stratagene) was used as host strain for plasmid construction and protein expression. Plasmids pCAG51 and pCAG51ΔP are described elsewhere (7, 25). *E. coli* SURE carrying the plasmids pCAG51 or pCAG51ΔP was used for the expression of GST-HD51 (GST-HD exon 1 fusion protein with 51 glutamines) and GST-HD51ΔP fusion proteins, respectively. Recombinant proteins were purified under native conditions by affinity chromatography on glutathione agarose as described (25), and aliquots were subjected to matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) MS for sequence confirmation. Recombinant proteins were stored at a conc. of 25 μM at $-80°$C.

High-Throughput Screening Procedure. In each well of six hundred 384-well microtiter plates, 15 μl of a 20 μM solution of the chemical compound to be tested (in 150 mM NaCl, 20 mM Tris·HCl (pH 8.0), 2 mM CaCl$_2$, and 6.7% DMSO) was mixed with 15 μl of predigested GST-HD51 fusion protein (1.25 μM) by using a pipetting robot. Total removal of the GST tag from the fusion protein GST-HD51 was achieved by elastase treatment (3 min at 37°C) before the addition of chemical compounds. Samples were incubated for 16 h at 37°C to allow aggregate formation. Reactions were stopped by addition of 30 μl of 4% SDS/100 mM DTT, followed by heating at 98°C for 10 min. Aliquots corresponding to 250 ng of GST-HD51 protein were filtered through a cellulose acetate membrane (0.2 μm, Schleicher & Schuell) by using a 384-well vacuum dot blot apparatus. Captured aggregates were detected by incubation with HD1 antibody (diluted 1:5,000), followed by incubation with alkaline phosphatase-conjugated anti-rabbit secondary antibody and the fluorescent substrate AttoPhos. Signals corresponding to SDS-insoluble aggregates were quantified by using AIDA 2.0 image analysis software (Raytest, Straubenhardt, Germany).

SDS/PAGE and Western Blotting. Proteins present in whole cell extracts or aggregation reactions were denatured, separated by SDS/PAGE (10 or 12%), and transferred to nitrocellulose. Membranes were blocked with 3% nonfat dry milk in Tris-buffered saline (TBS) containing 0.05% Tween 20 and incubated with the HD1 antibody (7). Secondary antibody was peroxidase-conjugated anti-rabbit IgG (Roche). Immunoreactive protein was detected by using enhanced chemiluminescence reagent (ECL, Amersham Pharmacia).

MS and Microscopic Analysis. The conditions for proteolytic digestion of GST-HD51ΔP with trypsin have been described (17). Mass maps of digested proteins were recorded on a Bruker Scout MTP Reflex III MALDI mass spectrometer (Bruker Daltonik, Germany) using the matrix α-cyano-4-hydroxycinamic acid. For electron microscopic analysis, the trypsin-digested GST-HD51 fusion protein was adjusted to a final conc. of 50 μg/ml in 40 mM Tris·HCl (pH 8.0) and 150 mM NaCl. Samples were negatively stained with 1% uranyl acetate and viewed in a Philips CM100 electron microscope (Philips Electron Optics, Eindhoven, The Netherlands). Electron micrographs of perinuclear inclusion bodies in 293 Tet-Off cells were generated as described (11).

Culturing of 293 Tet-Off Cells. Cells were grown in L-glutamine-free minimum essential medium with Earle's salts (GIBCO/BRL) supplemented with 10% FBS, 2 mM L-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycin, 100 μg/ml G418, 150 μg/ml hygromycin B, and 10 ng/ml doxycycline in poly-L-lysine-coated cell culture flasks. HDQ51 (Flag-tagged HD exon 1 protein with 51 glutamines) expression was induced by thoroughly washing the cells with PBS and adding fresh medium lacking doxycycline.

After incubation for 4 h in the absence of doxycycline, chemical compounds were added to the medium at the indicated concentrations. The compound-containing medium was changed daily, and, after 72–76 h of incubation, cells were harvested and protein extracts were prepared as described (26). Aliquots corresponding to 10–25 μg of protein were used for the cellulose acetate filter retardation assay (25). Protein concentration was determined by using the Bradford protein assay.

Indirect Immunofluorescence Microscopy. 293 Tet-Off cells were grown in Leighton tubes (Costar) in the presence or absence of PGL-135 (conc. 25 or 50 μM). Expression of HDQ51 was induced for 3 days by removal of doxycycline from the culture medium. Chemical compounds were added 4 h after incubation of cells in doxycycline-free medium. Indirect immunofluorescence microscopy was performed after 3 days as described (11).

Results

Inhibitor Screening. To identify polyQ aggregation inhibitors from large chemical libraries, a high-throughput *in vitro* screening assay was developed. The principle of this automated filter retardation assay is shown schematically in Fig. 1A. The assay is based on the finding that polyQ-containing protein aggregates are resistant to SDS and selectively retained on a cellulose acetate filter, whereas SDS-soluble protein under the same conditions is not. The aggregates retained on the filter membrane are then detected and quantified by immunoblot analysis using specific antibodies (25).

GST-HD51 was incubated for 3 min at 37°C with elastase, resulting in the complete removal of the GST tag from the fusion protein and the initiation of the aggregation process (data not shown). Then, immediately, chemical compounds were added to give a final conc. of 10 μM by using a pipetting robot. Incubation was continued for 16 h at 37°C to permit aggregate formation. Proteins were reduced and denatured by boiling in 2% SDS/50 mM DTT and filtered through a cellulose acetate membrane by using a 384-well dot blot apparatus. After washing of the membrane with 0.2% SDS, the amount of insoluble polyQ-containing protein aggregates retained on the surface was quantified by immunoblotting and subsequent image analysis. Compounds that reduced the signal intensity by more than 15% relative to the signal intensity of noninhibited HD51 aggregation reactions were identified as hits (Fig. 1B, F8). As positive controls, elastase-digested GST-HD51 protein treated with the known polyglutamine aggregation inhibitors Congo red and thioflavine S (17) (control I), as well as undigested GST-HD51 protein (control II), which does not aggregate (7), were applied to the filter membrane. Samples treated with the solvent DMSO and untreated GST-HD51 elastase cleavage products were used as negative controls (Fig. 1B).

High-throughput screening of an ≈184,000-chemical compound library by using the automated filter retardation assay resulted in the identification of about 300 chemical compounds that inhibited HD51 protein aggregation in a dose-dependent manner. Among these, we identified 25 benzothiazole derivatives by cluster analysis. Benzothiazole derivatives have been shown previously to be effective in treating neurodegenerative disorders such as amyotrophic lateral sclerosis (27). For example, riluzole (2-amino-6-trifluoromethoxybenzothiazole), a potent antagonist of glutamate release, has been reported to slow down disease progression in amyotrophic lateral sclerosis patients (28, 29). First clinical data on riluzole treatment of HD patients indicate a positive effect on hyperkinesia (30), and recently riluzole was reported to prolong the lifespan of a mouse model of HD (31). However, its mode of action is largely unknown. The identification of benzothiazoles as polyglutamine aggregation inhibitors is highly relevant because they are very promising candidates for future drug development. The structures of the
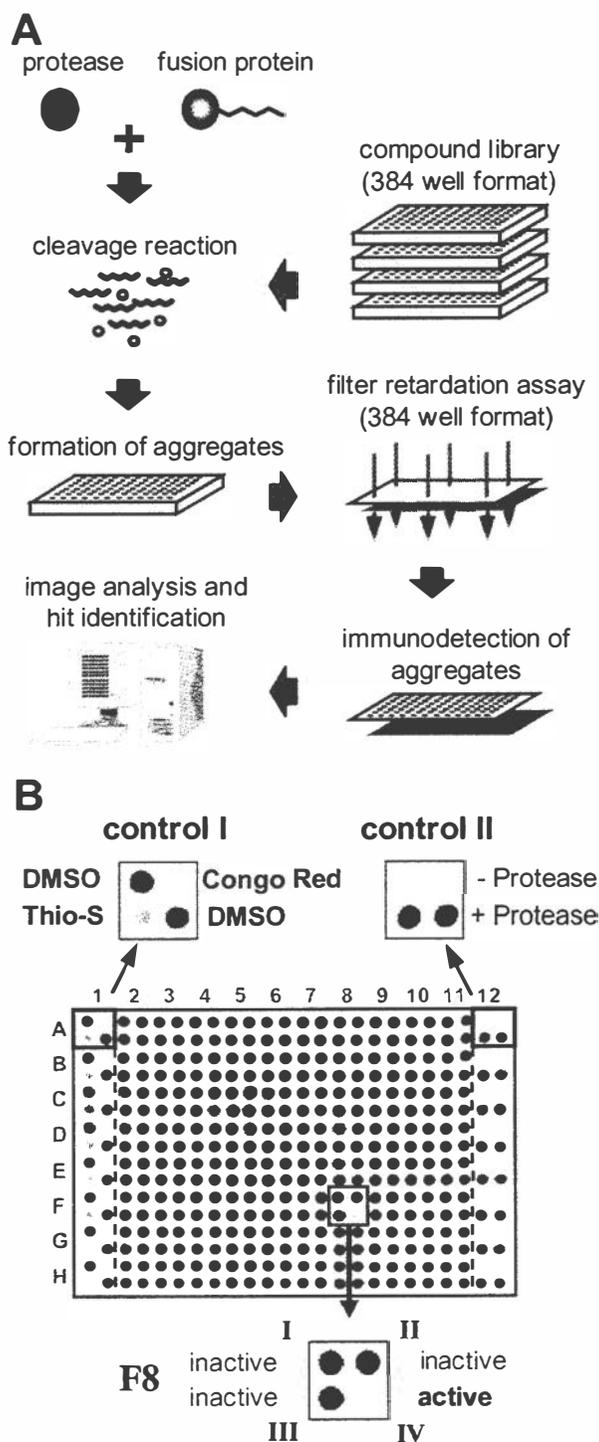
A

protease   fusion protein

+

cleavage reaction

compound library
(384 well format)

formation of aggregates

filter retardation assay
(384 well format)

image analysis and
hit identification

immunodetection of
aggregates

B

control I   **control II**

DMSO   Congo Red   - Protease
Thio-S   DMSO   + Protease

F8

	I	II	
	inactive	inactive	
	inactive	**active**	
	III	IV	

Fig. 1. Development of an automated filter retardation assay for high-throughput drug screening. (A) Flowchart of the automated filter retardation assay. (B) Effect of chemical compounds on HD exon 1 aggregation as monitored by the filter assay. GST-HD51 fusion protein at a conc. of 1.25 μM was predigested for 3 min at 37°C with elastase to remove totally the GST tag from the fusion protein and to initiate aggregation of the HDQ51 protein. Immediately after this step, various chemical compounds were added (final conc. 10 μM) and samples were incubated for an additional 16 h at 37°C. Then, aggregation reactions were filtered through a cellulose acetate membrane by using a 384-well dot blot apparatus. Captured aggregates were detected by immunoblotting using the HD1 antibody. On each filter membrane, 320 different chemical compounds were tested (squares A2–H11). Squares A1–H1 and A12–H12 were used for control samples. In square F8, an inhibitory compound was identified. ThioS, thioflavine S.

eight most effective compounds in the *in vitro* screen having IC$_{50}$ values in the range of 1–11 μM are shown in Table 1.

The effect of increasing concentrations of the compounds PGL-001, PGL-005, and PGL-034 on HD51 aggregation as monitored by the filter retardation assay is shown in Fig. 2A and B. PGL-001, PGL-005, and PGL-034 inhibited HD51 aggregation with IC$_{50}$ values of 1.2, 8.2, and 2.2 μM, respectively, whereas the solvent DMSO alone was ineffective. The inhibitory effect of the benzothiazole derivatives PGL-001 and PGL-034 on HD51 aggregation *in vitro* was also confirmed by electron microscopy (Fig. 3A). After an 18-h incubation of trypsin-digested GST-HD51 fusion protein (final conc. 100 ng/μl) without added chemical compound, numerous clusters of high-molecular-weight fibrils with a ribbon-like morphology (diameter, 20–40 nm) were detected. In strong contrast, treatment of the trypsin-digested GST-HD51 protein with PGL-001 or PGL-034 (each at 20 μM) resulted in the appearance of only a few fibrils with a diameter of 6–8 nm. In agreement with previous observations (7), electron microscopy of undigested GST-HD51 protein (negative control) predominantly showed spherical particles with a diameter of 6–7 nm.

Recently, we have shown that formation of insoluble HD exon 1 protein aggregates *in vitro* is a nucleation-dependent process (32) and that addition of chemical compounds such as Congo red or thioflavine S to the aggregation reaction significantly delays the assembly of monomeric HD exon 1 protein into high-molecular-weight fibrillar structures (17). To test whether the benzothiazole derivatives PGL-001 and PGL-034 have a similar effect on HD51 protein aggregation, a time course experiment was performed in which the decrease of soluble HD51 protein in the presence or absence of chemical compounds was monitored by SDS/PAGE and immunoblotting (Fig. 3B). After 24 h of incubation of trypsin-digested GST-HD51 protein, SDS-soluble HD51 protein was detected only in the presence of PGL-001 and PGL-034, indicating that these compounds have delayed the assembly of monomeric HD51 into insoluble protein aggregates. The soluble HD51 protein that resisted aggregate formation was readily accessible to proteolytic degradation. Thus, incubation for 30 min with proteinase K resulted in complete degradation of HD51 protein, whereas the high-molecular-weight polyQ-containing protein aggregates were largely resistant to proteinase K treatment (Fig. 3B). The results obtained by SDS/PAGE and immunoblotting were confirmed by MALDI-TOF MS analyses using a GST-HD exon 1 fusion protein, GST-HD51ΔP, that lacks the proline-rich region C-terminal to the polyQ tract (25). As shown in Fig. 3C, after incubation for 24 h, soluble HD51ΔP protein was detected only in samples that contained the compound PGL-001. In control samples lacking an added chemical compound, HD51ΔP assembled quantitatively into high-molecular-weight aggregates that are not accessible to MALDI-TOF MS analysis (17). Similar results were obtained with PGL-005 and PGL-034 (data not shown), indicating that benzothiazoles efficiently slow down the assembly of HD exon 1 protein into protease-resistant high-molecular-weight aggregates *in vitro*.

Inhibition of HD Exon 1 Protein Aggregation in a Cell Culture Model System of HD. To determine whether benzothiazoles are also capable of inhibiting HD exon 1 protein aggregation *in vivo*, a cell-based 96-well microtiter plate huntingtin aggregation assay was developed. In this assay, the tetracycline (tet)-regulated expression system (33) was used for the synthesis of HDQ51 in 293 Tet-Off cells (11). Cultivation of cells for 48–72 h in the absence of doxycycline induces the expression of HDQ51, resulting in the formation of large perinuclear inclusion bodies that mainly consist of fibrillar HDQ51 protein (11). The HDQ51 fibrils formed in 293 Tet-Off cells have a diameter of $\approx$10 nm (data not shown).

Table 1. Benzothiazoles inhibit HD exon 1 aggregation *in vitro*

HTS no.	Structure	Name	IC_{50}, μM
PGL-374		Benzothiazole-2,5,6-triamine	3.5 ± 1.2
PGL-034		[6,6']Bibenzothiazolyl-2,2'-diamine	2.2 ± 0.2
PGL-016		6-(2-Amino-benzothiazol-6-ylsulfanyl)-benzothiazol-2-ylamine	7.5 ± 2.0
PGL-317		2-[3-(6-Amino-benzothiazol-2-yl)-phenyl]-benzothiazol-6-ylamine	2.5 ± 1.6
PGL-005		3-Amino-N-[4-(6-methyl-benzothiazol-2-yl)-phenyl]-benzamide	8.2 ± 0.5
PGL-072		3-Methoxy-N-[4-(6-methyl-benzothiazol-2-yl)-phenyl]-benzamide	9.3 ± 7.2
PGL-077		4-[(4-Benzothiazol-2-yl-phenylimino)-methyl]-2,6-dibromo-benzene-1,3-diol	11.0 ± 4.5
PGL-001		N-(6-Phenylcarbamoyl-benzothiazol-2-yl)-terephthalamic acid methyl ester	1.2 ± 0.5
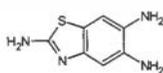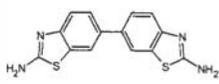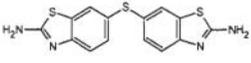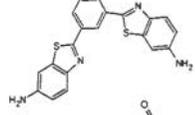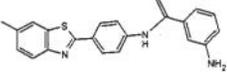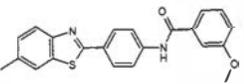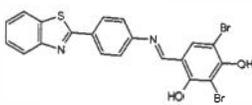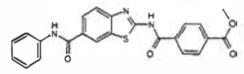
IC_{50} values shown are the average of four independent determinations (±SE).

Induced 293 Tet-Off cells expressing HDQ51 were grown for 72 h in the presence or absence of benzothiazole derivatives. Then, the cells were lysed, and the protein concentration in the cell extracts was determined. Because the protein concentration largely correlates with the cell density, a reduction of protein concentration by the addition of a given chemical compound indicates that the compound is toxic for 293 Tet-Off cells. For the detection of SDS-insoluble protein aggregates, protein extracts were filtered through a cellulose acetate membrane, and the amount of captured aggregates was quantified by immunoblotting and image analysis.

We tested more than 100 benzothiazole derivatives, including those identified by the high-throughput *in vitro* screen (Table 1), as well as several structural analogues obtained by computer analysis. Unexpectedly, we found that the most potent compounds in the cell-free assays—such as PGL-001, PGL-005, PGL-016, and PGL-034 (Figs. 2 and 3; Table 1)—were toxic for 293 Tet-Off cells (data not shown). Therefore, their activity on HDQ51 aggregation cannot be evaluated in this cell model system of HD. A similar result was also obtained, when the FDA-approved compound riluzole was tested in the cell-based assay (Fig. 4A). Increasing concentrations of riluzole markedly reduced the protein concentration, suggesting that the effect of this compound on HDQ51 aggregation is due to toxicity, but not due to an inhibition of this process (Fig. 4 *B* and *C*). In strong contrast, the benzothiazole derivatives PGL-135, PGL-137, and PGL-201 identified by computer analysis because of structural homology to the compounds shown in Table 1 were nontoxic for 293 Tet-Off cells (Fig. 4). We found that the compounds PGL-135 and PGL-137, but not PGL-201, inhibited HDQ51 aggregation in the cell-based assay. As shown in Fig. 4*B*, treatment of 293 Tet-Off cells with increasing concentrations of PGL-135 inhibited HDQ51 aggregation in a dose-dependent manner (IC_{50} value ≈40 μM). In comparison, PGL-137 suppressed HD exon 1 aggregation with an IC_{50} value of ≈100 μM
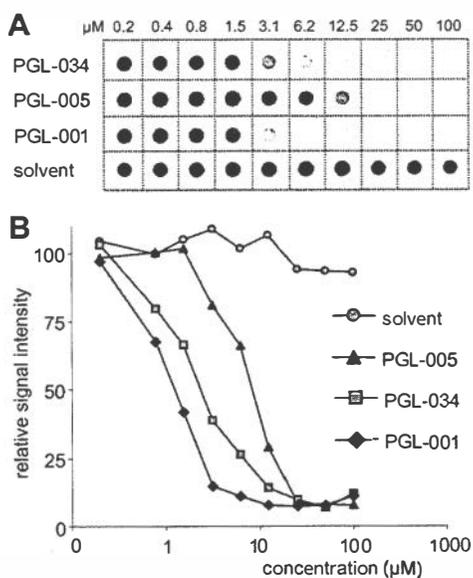

Fig. 2. *In vitro* inhibition of HD exon 1 aggregation by benzothiazoles. (*A*) Effect of the indicated chemical compounds on HD exon 1 aggregation as monitored by the filter retardation assay. (*B*) Quantification of the dot blot results shown in *A*. The signal intensity from the sample without added chemical compound (solvent) was arbitrarily set as 100. For structure of the chemical compounds, see Table 1.
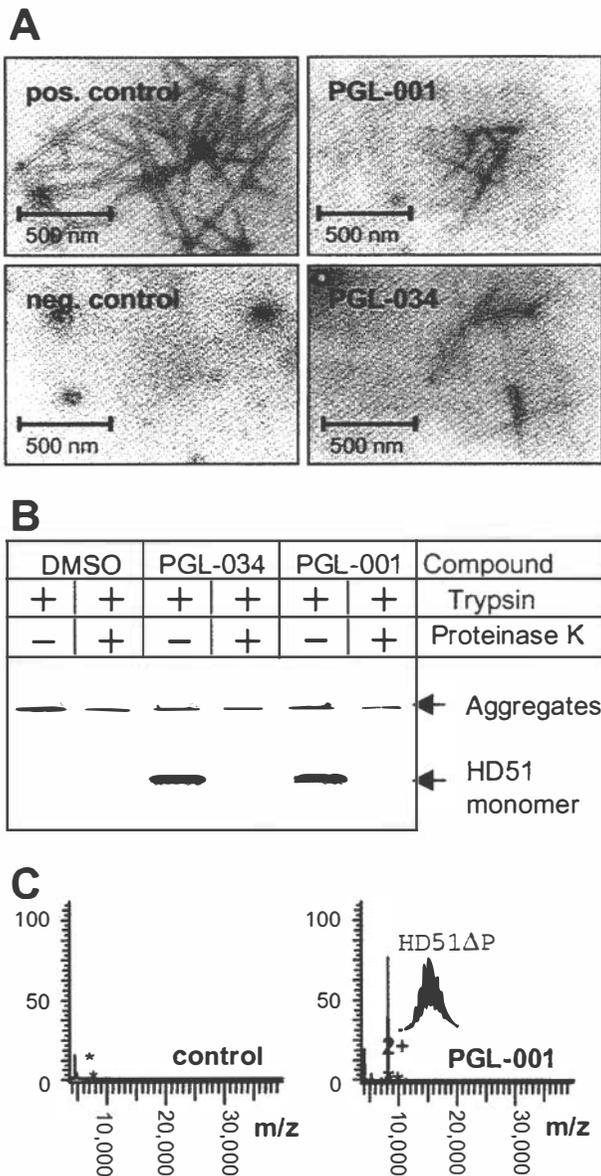
A

B

C

Fig. 3. Analysis of HD exon 1 aggregation *in vitro* by using secondary assays. (*A*) Electron micrographs of HD51 fibrils formed in the presence or absence of the indicated benzothiazoles. Trypsin-digested GST-HD51 protein at 1.5 μM was incubated for 16 h at 37°C either without chemical compound (pos. control) or with PGL-001 and PGL-034 (final conc. 20 μM). Representative examples of fibrillar structures are shown. (*B*) Western blot analysis of trypsin-digested GST-HD51 fusion protein. Aggregation reactions were performed for 16 h in the presence or absence of the indicated chemical compounds. Where indicated (+), aliquots were taken and incubated for additional 30 min with 10 ng/μl proteinase K. Samples corresponding to 200 ng of fusion protein were analyzed by SDS/PAGE and immunoblotting using the HD1 antibody. (*C*) Effect of the compound PGL-001 on HD exon 1 aggregation as monitored by MALDI-TOF MS. GST-HD51ΔP at 2.5 μM was incubated for 18 h at 37°C with trypsin in the absence (control) or presence of PGL-001 (final conc. 20 μM). Analysis by MALDI-TOF MS revealed that, only in the presence of PGL-001, monomeric HD51ΔP peptide (monoisotopic mass of 8,074.74 Da) is detectable; without added compound, the peptide has assembled into insoluble high-molecular-weight protein aggregates that cannot be detected by MALDI-TOF MS.

(data not shown). Together, these results indicate that the benzothiazole derivatives PGL-135 and PGL-137 are nontoxic for 293 Tet-Off cells and inhibit HDQ51 aggregation *in vivo*.

The effect of PGL-135 on the formation of HDQ51 aggregates in 293 Tet-Off cells was also examined by indirect immunofluorescence microscopy (Fig. 5*A*). Treatment of cells with PGL-135 significantly reduced the formation of large perinuclear inclusion bodies in a dose-dependent manner. Cultivation of the cells with 25 and 50 μM PGL-135 resulted in a 30 and 50% reduction of the average number of inclusion bodies per cell, respectively, consistent with the results obtained by the filter retardation assay (Fig. 4*B*).

Discussion

In this study, we have developed cell-free and cell-based assays for the identification of chemical compounds that prevent the formation of SDS-insoluble HD exon 1 aggregates. Using a high-throughput *in vitro* filter retardation assay, we have identified about 300 chemical compounds in a library of $\approx$184,000 small molecules that suppress the accumulation of polyQ-containing huntingtin aggregates in a dose-dependent manner. Among these compounds, five larger groups with 5–35 structurally related small molecules were found. In this study, we have concentrated on the analysis of one group of small molecules containing 25 different benzothiazole derivatives.

Benzothiazoles are highly interesting molecules for drug development, because they already have been shown to be useful for treating cerebrovascular and neurodegenerative disorders. For example, riluzole extends the survival of neurons in amyotrophic lateral sclerosis patients (28, 29). It also has been tested for therapy of HD patients, where treatment with riluzole has positive effects on choreatic hyperkinesia (30). Moreover, it has been reported that riluzole leads to an extended lifespan in a mouse model of HD (31). Immunohistochemistry revealed profound changes in ubiquitination of HD-characteristic neuronal intranuclear inclusions, but no significant size reduction of huntingtin-positive inclusions was observed, indicating that this molecule does not directly interfere with huntingtin aggregation (31).

Our data show that benzothiazole derivatives are potent inhibitors of HD exon 1 aggregation *in vitro* and in cell culture model systems of HD. However, the mechanism of action of these molecules on a molecular level is unclear. We have shown previously that polyQ-containing HD exon 1 aggregates are formed by a nucleation-dependent polymerization (32). Thus, aggregation does not proceed immediately but after a lag phase during which monomers slowly assemble into unstable oligomeric structures (nuclei). Once these structures have formed, the addition of further monomers becomes thermodynamically favorable, resulting in the rapid accumulation of large fibrils (4). Currently, the structure of the huntingtin oligomers and fibrils is unknown; however, it is likely that both structures consist of cross β-sheets (4, 32).

We propose that benzothiazoles slow down HD exon 1 aggregation because they bind to polyQ-containing β-sheet structures. Similar to Congo red, they may interfere with nucleus formation and/or fibril growth. However, detailed *in vitro* drug-protein binding studies will be necessary to address these questions in more detail. Currently, it is not known whether benzothiazole derivatives such as PGL-001 preferentially interact with monomers, oligomers, or fibrillar structures. We have found that PGL-001 inhibits the aggregation of HDQ51 protein (conc. 0.625 μM) with an IC_{50} of 1.2 μM, suggesting that a protein-inhibitor stoichiometry of 1:2 is necessary and sufficient to inhibit HD exon 1 aggregation *in vitro*.

The activity of benzothiazoles to prevent huntingtin aggregation was also assessed in a cell culture model of HD. We found that, at conc. of 10–150 μM, the most potent compounds in the cell-free assays—such as PGL-001 or PGL-034—as well as the FDA-approved drug riluzole (Fig. 4), are toxic for 293 Tet-Off cells. Thus, the impact of these compounds on HD exon 1 aggregation *in vivo* cannot be determined. In comparison, the
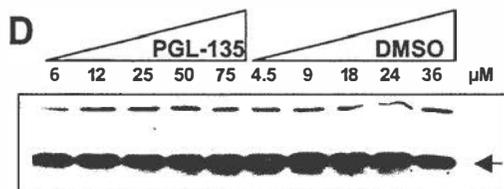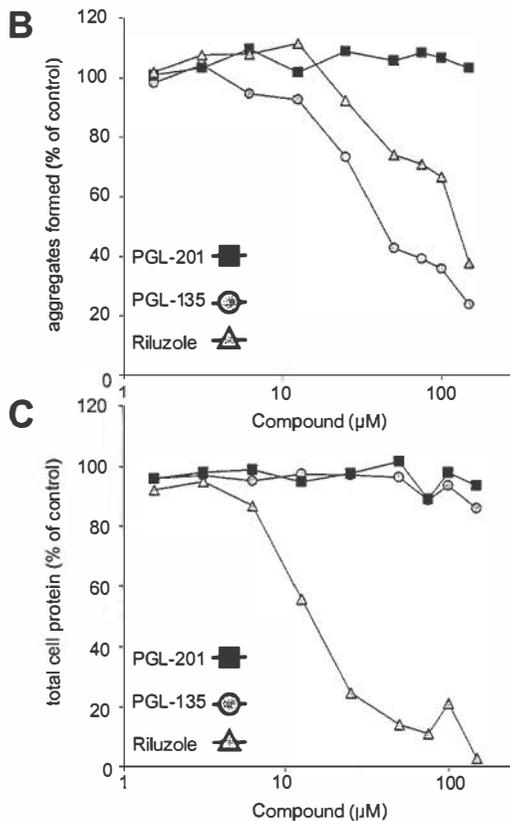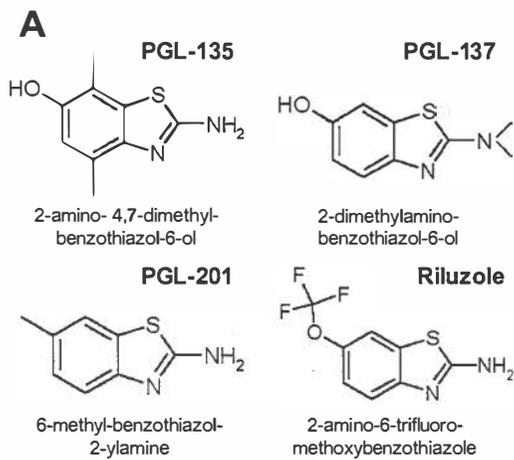
Fig. 4. Inhibition of HD exon 1 aggregation in 293 Tet-Off cells. (*A*) Structure of chemical compounds counteracting HDQ51 aggregation *in vivo*. (*B*) Quantification of filter retardation assay results. Cells were incubated for 72 h in the presence of various concentrations of the indicated chemical compounds. Protein extracts were prepared and filtered through a cellulose acetate membrane; captured SDS-insoluble protein aggregates were detected with the HD1 antibody. The dots corresponding to the control reactions without added chemical compound (not shown) were arbitrarily set as 100. (*C*) Relative protein concentrations of the cell extracts analyzed in *B*. (*D*) Western blot of cell extracts prepared from 293 Tet-Off cells after treatment with increasing concentrations of the chemical compound PGL-135 and the solvent DMSO. The arrow indicates the HDQ51 monomer.
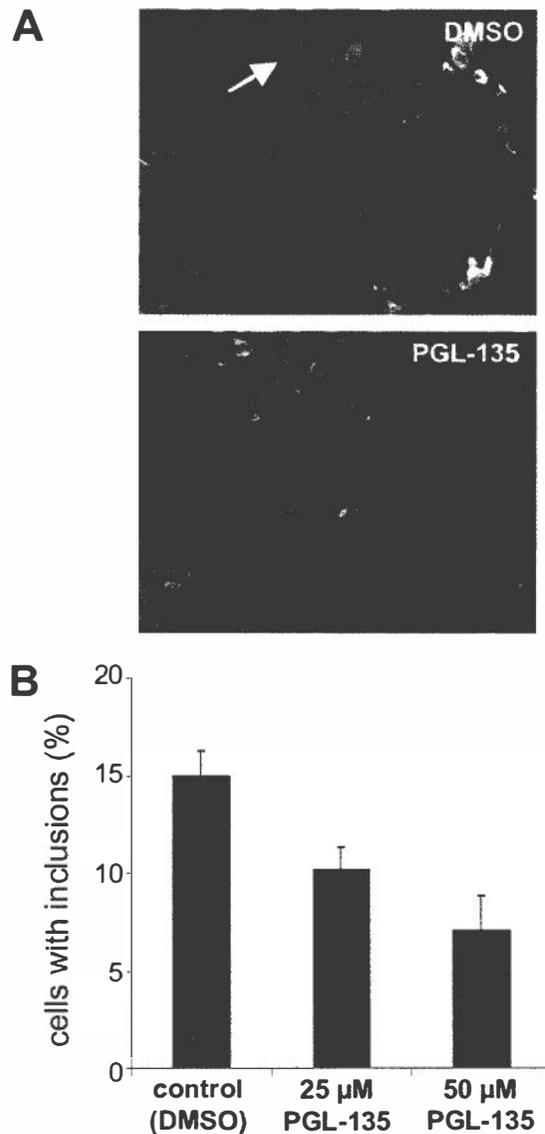
Fig. 5. Immunofluorescence microscopy analysis. (*A*) 293 Tet-Off cells expressing Flag-tagged HD exon 1 protein with 51 glutamines (HDQ51) were cultivated for 72 h in the presence or absence of PGL-135. Formation of inclusion bodies with aggregated HDQ51 protein was followed by indirect immunofluorescence microscopy using the anti-Flag antibody. Inclusion bodies are indicated as red dots by the arrow. Nuclei were counterstained with Hoechst. A total of about 5,000 PGL-135-treated and untreated cells were examined. (*B*) Quantification of cells with inclusion bodies.

structurally related benzothiazole derivatives PGL-135 and PGL-137, identified by computer analysis, were nontoxic in the cell-based assay under the same conditions. We suggest that the compounds PGL-135 and PGL-137, similar to the most active compounds in the cell-free assay, directly interact with mutant HD exon 1 protein and thereby slow down the formation of insoluble protein aggregates *in vivo*. However, at this stage, we cannot exclude an indirect mode of action in the cell culture model of HD. For example, misfolding and aggregation of polyQ-containing HD exon 1 protein could be prevented by stimulation of a heat shock response or activation of the ubiquitin/proteasome system (11, 15, 16). Experiments to address these questions are needed.

Our studies show that the automated filter retardation assay described above is suitable for high-throughput screening of

chemical compounds that prevent aggregate formation. Using this cell-free assay, very large compound libraries can be screened because the system is fast, robust, and relatively cheap. Furthermore, the amount needed of each substance for the aggregation reaction is small compared with other *in vitro* aggregation drug screening assays (34). Another strength of this assay is its ability to detect not only compounds that inhibit fibril growth (35), but also compounds that prevent nucleation.

A basic problem with drug screenings is that they tend to yield different results in cell-free, as opposed to cell-based, assays. Many compounds that show beneficial activities in the cell-free experiments are inactive or toxic in cell culture. We, e.g., found that benzothiazole derivatives that showed the greatest inhibitory effect *in vitro* were toxic in the cell assays. This finding means that the first step in screening can be to some extent misleading and that only the combination of cell-free and cell-based assays will result in the successful identification of small molecules that have the potential to be further developed into effective drugs. The challenge is to find chemical compounds that are nontoxic, that have a reasonable brain permeability, and that prevent the formation of huntingtin aggregates in the patient's neurons.

We have described a previously uncharacterized class of chemical compounds that prevent the formation of HD exon 1 aggregates *in vitro* and are biologically active in cell cultures. Now, the inhibitory activity of these molecules will be demonstrated in transgenic mouse models of HD and ultimately in humans.

1. Harper, P. S. (1991) *Huntington's Disease* (Saunders, London).
2. Rubinsztein, D. C., Leggo, J., Coles, R., Almqvist, E., Biancalana, V., Cassiman, J. J., Chotai, K., Connarty, M., Crauford, D., Curtis, A., *et al.* (1996) *Am. J. Hum. Genet.* **59,** 16–22.
3. Sathasivam, K., Amaechi, I., Mangiarini, L. & Bates, G. (1997) *Hum. Genet.* **99,** 692–695.
4. Wanker, E. E. (2000) *Biol. Chem.* **381,** 937–942.
5. Paulson, H. L. (1999) *Am. J. Hum. Genet.* **64,** 339–345.
6. Wanker, E. E. (2000) *Mol. Med. Today* **6,** 387–391.
7. Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H. & Wanker, E. E. (1997) *Cell* **90,** 549–558.
8. Krobitsch, S. & Lindquist, S. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 1589–1594.
9. Faber, P. W., Alter, J. R., MacDonald, M. E. & Hart, A. C. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 179–184.
10. Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L. & Bates, G. P. (1997) *Cell* **90,** 537–548.
11. Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H. & Wanker, E. E. (2001) *Mol. Biol. Cell* **12,** 1393–1407.
12. Yamamoto, A., Lucas, J. J. & Hen, R. (2000) *Cell* **101,** 57–66.
13. Lecerf, J. M., Shirley, T. L., Zhu, Q., Kazantsev, A., Amersdorfer, P., Housman, D. E., Messer, A. & Huston, J. S. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4764–4769.
14. Nagai, Y., Tucker, T., Ren, H. Z., Kenan, D. J., Henderson, B. S., Keene, J. D., Strittmatter, W. J. & Burke, J. R. (2000) *J. Biol. Chem.* **275,** 10437–10442.
15. Sittler, A., Lurz, R., Lueder, G., Priller, J., Hayer-Hartl, M. K., Hartl, F. U., Lehrach, H. & Wanker, E. E. (2001) *Hum. Mol. Genet.* **10,** 1307–1315.
16. Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K. & Hartl, F. U. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7841–7846.
17. Heiser, V., Scherzinger, E., Boeddrich, A., Nordhoff, E., Lurz, R., Schugardt, N., Lehrach, H. & Wanker, E. E. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 6739–6744.
18. Bard, F., Cannon, C., Barbour, R., Burke, R. L., Games, D., Grajeda, H., Guido, T., Hu, K., Huang, J., Johnson-Wood, K., *et al.* (2000) *Nat. Med.* **6,** 916–919.
19. Soto, C., Sigurdsson, E. M., Morelli, L., Kumar, R. A., Castano, E. M. & Frangione, B. (1998) *Nat. Med.* **4,** 822–826.
20. Korth, C., May, B. C., Cohen, F. E. & Prusiner, S. B. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 9836–9841.
21. Tatzelt, J., Prusiner, S. B. & Welch, W. J. (1996) *EMBO J.* **15,** 6363–6373.
22. Klunk, W. E., Debnath, M. L., Koros, A. M. & Pettegrew, J. W. (1998) *Life Sci.* **63,** 1807–1814.
23. Howlett, D. R., Perry, A. E., Godfrey, F., Swatton, J. E., Jennings, K. H., Spitzfaden, C., Wadsworth, H., Wood, S. J. & Markwell, R. E. (1999) *Biochem. J.* **340,** 283–289.
24. Tomiyama, T., Shoji, A., Kataoka, K., Suwa, Y., Asano, S., Kaneko, H. & Endo, N. (1996) *J. Biol. Chem.* **271,** 6839–6844.
25. Wanker, E. E., Scherzinger, E., Heiser, V., Sittler, A., Eickhoff, H. & Lehrach, H. (1999) *Methods Enzymol.* **309,** 375–386.
26. Sittler, A., Walter, S., Wedemeyer, N., Hasenbank, R., Scherzinger, E., Eickhoff, H., Bates, G. P., Lehrach, H. & Wanker, E. E. (1998) *Mol. Cell* **2,** 427–436.
27. Jimonet, P., Audiau, F., Barreau, M., Blanchard, J. C., Boireau, A., Bour, Y., Coleno, M. A., Doble, A., Doerflinger, G., Huu, C. D., *et al.* (1999) *J. Med. Chem.* **42,** 2828–2843.
28. Bensimon, G., Lacomblez, L. & Meininger, V. (1994) *N. Engl. J. Med.* **330,** 585–591.
29. Lacomblez, L., Bensimon, G., Leigh, P. N., Guillet, P. & Meininger, V. (1996) *Lancet* **347,** 1425–1431.
30. Rosas, H. D., Koroshetz, W. J., Jenkins, B. G., Chen, Y. I., Hayden, D. L., Beal, M. F. & Cudkowicz, M. E. (1999) *Movement Disorders* **14,** 326–330.
31. Schiefer, J., Landwehrmeyer, G. B., Lüesse, H.-G., Sprünken, A., Puls, C., Milkereit, A., Milkereit, E. & Kosinski, C. M. (2002) *Movement Disorders* **17,** 748–757.
32. Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehrach, H. & Wanker, E. E. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 4604–4609.
33. Gossen, M. & Bujard, H. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 5547–5551.
34. Lin, Y.-M., Raffen, R., Zhou, Y., Cassidy, C. S., Flavin, M. T. & Stevens, F. J. (2001) *Amyloid* **8,** 182–193.
35. Esler, W. P., Stimson, E. R., Ghilardi, J. R., Felix, A. M., Lu, Y.-A., Vinters, H. V., Mantyh, P. W. & Maggio, J. E. (1997) *Nat. Biotechnol.* **15,** 258–263.

Colloquium

Chaperoning brain degeneration

Nancy M. Bonini*

Department of Biology, University of Pennsylvania, Howard Hughes Medical Institute, Philadelphia, PA 19104-6018

Drosophila has emerged as a première model system for the study of human neurodegenerative disease. Genes associated with neurodegeneration can be expressed in flies, causing phenotypes remarkably similar to those of the counterpart human diseases. Because human neurodegenerative diseases, including Huntington's and Parkinson's diseases, are disorders for which few cures or treatments are available, Drosophila brings to bear powerful genetics to the problem of these diseases. The molecular chaperones were the first modifiers defined that interfere in the progression of such disease phenotypes in Drosophila. Hsp70 is a potent suppressor of both polyglutamine disease and Parkinson's disease in Drosophila. These studies provide the promise of treatments for human neurodegeneration through the up-regulation of stress and chaperone pathways.

Huntington's and Parkinson's diseases are late-onset, progressive human neurodegenerative diseases associated with selective neuronal loss and abnormal protein accumulations. Huntington's disease is one of a class of human diseases known as the polyglutamine repeat diseases (see ref. 1 for review). This class also includes dentatorubropallidoluysian atrophy (DRPLA), spinobulbar muscular atrophy (SBMA) and spinocerebellar ataxias type 1, 2, 3 (also known as Machado–Joseph disease, MJD), 6, 7, and 17. The polyglutamine diseases are characterized by the expansion of a run of the amino acid glutamine within the ORF of the respective proteins. The expanded polyglutamine domain confers dominant toxicity on the respective disease proteins, leading to neuronal dysfunction and degeneration. These diseases are also associated with abnormal protein accumulations containing the disease protein, typically in the form of nuclear inclusions. These inclusions immunostain for ubiquitin, suggesting that they contain misfolded or abnormally folded protein, potentially targeted for proteasomal degradation.

Dominant Parkinson's disease is characterized by selective loss of dopaminergic neurons in the substantia nigra pars compacta. Abnormal protein accumulations, known as Lewy bodies, typify the disease. Lewy bodies are cytoplasmic aggregates composed primarily of the protein α-synuclein (2); they contain ubiquitinated protein, suggesting that the accumulating protein has been targeted for degradation. Causal association of abnormal α-synuclein function with Parkinson's disease was found when two mutations in α-synuclein, A30P and A53T, were described in rare familial forms (3, 4).

Drosophila is a powerful genetic model system, which has been well studied as a developmental system. Many genes are conserved between humans and flies, including entire gene pathways (5). Drosophila has a complex nervous system and displays complex behaviors, including learning and memory. Many genes known to be involved in pathways of behavior, including learning and memory, circadian behavior, and phototaxis, were first described in Drosophila mutants (6–8). Given these striking homologies between Drosophila and humans, we reasoned that the power of Drosophila genetics could be brought to bear on the problem of human neurodegenerative disease.

Whereas mutations have been known for many years that lead to loss of integrity of the fly brain (8, 9), we reasoned that another way to generate such models of specific interest for their application to human neurodegeneration would be to express in the fly the pathogenic human disease gene. With the phenotype in the fly resembling that of the human disease in fundamental properties, this would indicate at least some aspects of the disease process are also conserved between flies and humans. This conservation therefore would allow fly genetics to be applied to define mechanisms of disease progression and modifiers that interfere with the disease process, thus opening up the realm of Drosophila neurogenetics toward the cure and treatment of these devastating human disorders. Here I present a review of previous findings on Drosophila models of neurodegeneration, with some additional new findings.

Materials and Methods

Details of the methods used in the studies summarized here are described in previously published research reports (see refs. 10–13). The Hsp70 dominant-negative transgene encoding Hsp70.K71E was generated by mutagenesis of the Hsp70 transgene described (12).

A *Drosophila* Model for Human Neurodegenerative Disease

To establish the fly as a model system for human neurodegeneration, we decided to express in the fly the normal form and a mutant disease form of the gene encoding spinocerebellar ataxia type 3, or MJD. We used a truncated form of the disease protein in these studies, as this protein had been shown to have effects when expressed in transgenic mice (14). To do this, we subcloned cDNAs encoding a protein with a polyglutamine repeat within the normal range, MJDtrQ27, and a protein with a polyglutamine repeat within the pathogenic range, MJDtrQ78, into fly transformation vectors. The two-component GAL4-UAS system was used for transgene expression (15). Transgenic flies were obtained, and expression was directed to neural tissues. Typically, expression is directed to the eye with gmr-GAL4 or to the entire nervous system with elav-GAL4.

Expression of the control protein MJDtrQ27 has no discernable phenotype—flies are born with eyes indistinguishable from normal. Expression of the disease form of the protein, MJDtrQ78, however, has profound effects. Flies are born with eyes mildly to strongly degenerate when the gmr-GAL4 eye driver is used, with loss of red pigmentation, loss of internal eye integrity, and severe degeneration of the photoreceptor neurons (Fig. 1 A–D) (13). The strength of the phenotype depends on expression level of the transgene encoding the pathogenic protein—weak expression induces mild degeneration, whereas strong expression is associated with severe degeneration.
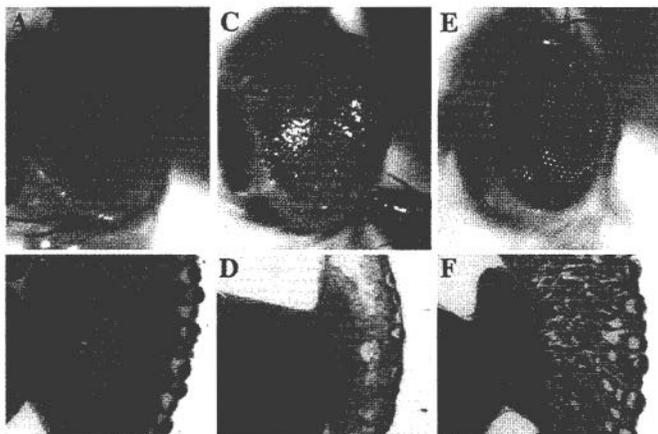
Fig. 1. Polyglutamine degeneration and suppression by the molecular chaperone Hsp70. (*Upper*) External eyes. (*Lower*) Horizontal sections through the eye to reveal the internal eye structure. (*A* and *B*) Normal fly with just the driver gmr-GAL4. (*C* and *D*) Fly expressing the pathogenic polyglutamine protein MJDtr-Q78 has severe eye degeneration, with loss of external and internal eye structure. (*E* and *F*) Flies co-expressing the toxic polyglutamine protein MJDtr-Q78 with Hsp70 have dramatically restored external and internal eye structure. Fly genotypes are *w; +/gmr-GAL4* (*A* and *B*), *w; gmr-GAL4 UAS-Hsp70/UAS-MJDtr-Q78(S)* (*C* and *D*), and *w; gmr-GAL4/UAS-MJDtr-Q78(S)* (*E* and *F*).

The phenotype was also progressive over time. Although the flies are born with various degrees of degeneration depending upon the specific transgenic insertion, degeneration becomes progressively more severe over the lifetime of the adult fly. This degeneration is seen as progressive loss of pigmentation, enhanced deterioration of internal eye integrity, and early death of the animal, associated with tremors and shaking movement.

Examination of the tissue for protein expression revealed that the pathogenic polyglutamine protein forms abnormal inclusions within the cell nuclei. Such abnormal inclusions are characteristic of the human polyglutamine diseases, where they are described as nuclear inclusions or cytoplasmic inclusions, depending upon their particular subcellular localization (16–18). In our MJD model, the inclusions are nuclear. Whereas these inclusions form early in the fly cells, degeneration of the cells does not occur for many days. The inclusions form in all cells in which the pathogenic protein is expressed, even in cells that do not degenerate and are insensitive to the pathogenic actions of the disease protein. This observation suggests that whereas the inclusions may be part of the disease process or indicative of the abnormal folding of the pathogenic protein, the mere presence of an inclusion is not sufficient for cellular degeneration. Nevertheless, such abnormal protein accumulations are characteristic of the human diseases, providing another point of similarity between the fly model and the human disorders.

These features of the fly model—late onset and progressive neurodegeneration accompanied by the formation of abnormal protein accumulations—are fundamental features of human polyglutamine disease. This fact indicates that *Drosophila* can display mechanisms of human polyglutamine degeneration. These findings indicate that *Drosophila* genetics can be applied toward defining mechanisms, cures, and treatments for such human neurodegeneration.

The Molecular Chaperone Hsp70 Is a Potent Suppressor of Polyglutamine Pathogenicity

Given that polyglutamine disease is associated with an abnormal protein conformation, the molecular chaperones may play a role in disease progression. In flies, the major stress-induced chaperone is Hsp70. Therefore, we asked whether Hsp70 might be involved in the disease process. We found that the nuclear inclusions immunostained, from initial stages of their formation, with antibodies that detect the Hsp70 proteins (12). This finding suggested the possibility that the cells were mounting a stress response against the pathogenic protein. Therefore, we asked whether it would make a difference to supply the cells with additional Hsp70 activity. To do this, we made transgenic flies that overexpress human Hsp70, which is highly conserved with fly Hsp70, but can be detected with species-specific antibodies.

Co-expression of Hsp70 dramatically suppresses the degeneration normally associated with the pathogenic polyglutamine protein MJDtr-Q78 (Fig. 1). The external eye structure is fully restored to normal, and internal eye structure is strongly restored. Moreover, not only is initial degeneration arrested but also progressive degeneration is prevented. We verified that there were no differences in transgene expression and that, rather, the added Hsp70 is protecting or compensating for the toxicity of the pathogenic disease protein.

To address whether the enzymatic ATPase activity of Hsp70 is important for the suppression, we examined transgenic flies that express a form of the constitutively expressed Hsp70, Hsc4, with a point mutation in the ATPase domain that acts *in vivo* in a dominant-negative manner (19). Co-expression of this protein with the disease protein not only fails to suppress, but actually enhances, degeneration. This finding suggests that toxicity to the polyglutamine protein is sensitive to the levels of the Hsp70 family of molecular chaperones, with added Hsp70 preventing degeneration, whereas interference with endogenous chaperone activity promotes degeneration.

We also investigated the role of the Hsp70 co-chaperone Hsp40 in protein pathogenicity, by creating transgenic flies that overexpress the fly counterpart of the human Hdj1 class of molecular chaperone, dHdj1. These flies, like those expressing Hsp70, also show strong suppression of polyglutamine toxicity (ref. 11, also ref. 20). The Hsp40 proteins show specificity in that dHdj1 is effective, whereas dHdj2 is poor at protecting against polyglutamine toxicity (11). This difference is consistent with idea that different Hsp40 class chaperones have distinct substrate specificities [see review by Hartl and colleagues in this issue (21) for more extensive discussion of Hsp70/Hsp40 functional interactions]. There appears to be selectivity for the polyglutamine protein, such that dHdj1 is more effective.

We also examined potential interactions between Hsp70 and dHdj-1. As dHdj-1 is presumably a co-chaperone for Hsp70, we anticipated that we might detect a synergy between the two proteins in suppression of pathogenicity. Indeed, although either Hsp70 or dHdj1 on its own is a strong suppressor, when they are co-expressed suppression of polyglutamine degeneration is even stronger (Fig. 2). These findings emphasize the importance of providing a sufficient complement of chaperones for suppression. In our disease model, it appears that there are sufficient levels of Hsp70 and Hsp40 alone to allow initial suppression. The late-onset nature of the degeneration may signify that the chaperone system eventually becomes overwhelmed. However, cells with a generally poor basal stress or chaperone system may require more than just Hsp70 or Hsp40 alone for significant suppression, instead requiring a complement of chaperones to effect protection.

It is of interest that the potent suppression of the adult eye degenerative phenotype by Hsp70 and dHdj1 occurs in the absence of an effect on the morphology of the aggregates formed by the pathogenic protein, as visualized by immunocytochemistry. Nuclear inclusions are formed in the fly upon chaperone suppression, and they are present in the same number and the same size as in the absence of additional chaperones. The nuclear inclusions appear the same, except that the exogenous Hsp70 and dHdj1 are now also found in the inclusions, suggesting an interaction with the pathogenic protein.
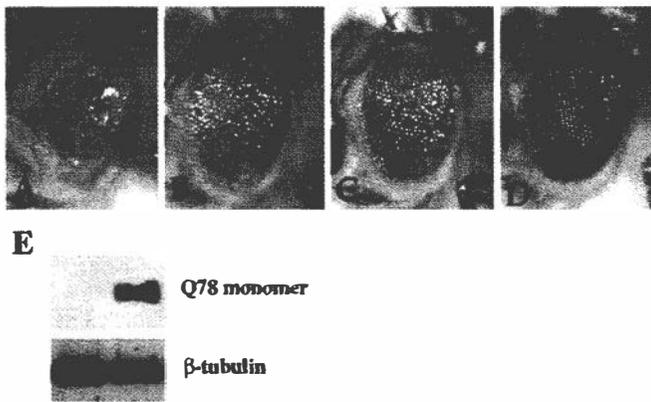
Fig. 2. Hsp70 and Hsp40 synergize in suppression of polyglutamine toxicity. Two copies of the polyglutamine protein are expressed in *A–D*, which makes the phenotype severe enough that synergy may readily be seen. (*A*) The polyglutamine protein MJDtr-Q78 causes severe degeneration. (*B* and *C*) Expression of either Hsp70 (*B*) or dHdj1 (*C*) alone has partial ability to rescue eye structure. (*D*) Expression of Hsp70 with dHdj1 results in full restoration of eye structure to normal (compare with Fig. 1 *A*). Fly genotypes are *w; gmr-GAL4 UAS-MJDtrQ78/UAS-MJDtr-Q78* (*A*), *w; gmr-GAL4 UAS-MJDtr-Q78/ UAS-MJDtrQ78 UAS-Hsp70* (*B*), *w; gmr-GAL4 UAS-MJDtrQ78/UAS-MJDtr-Q78 UAS-Hsp40* (*C*), and *w; gmr-GAL4 UAS-MJDtr-Q78 UAS-Hsp70/UAS-MJDtr-Q78 UAS-Hsp40*. (*E*) Chaperones increase the SDS-solubility of the pathogenic polyglutamine protein. Shown is a Western immunoblot of monomeric poly-glutamine protein extracted from the heads of flies expressing the disease protein alone (left lane), or with Hsp70 (right lane). As shown in the left lane, normally most of the polyglutamine protein is SDS-insoluble, remaining within the stacking gel and poorly transferring in a Western immunoblot (see ref. 11), such that little or no protein is present as a monomer. However, in the presence of chaperones (right lane), there is a significant amount of protein now SDS-soluble that runs as a monomer. Lower gel is β-tubulin control showing equal loading. Heads are from flies of genotype *w; gmr-GAL4 UAS-MJDtrQ78/+* (left lane) and *w; gmr-GAL4 UAS-MJDtr-Q78/UAS-Hsp70* (right lane).

Fig. 3. α-Synuclein Lewy-body-like aggregates in flies and in Parkinson's disease patient tissue. (*A* and *B*) Brain sections through a 20-day-old fly expressing wild-type α-synuclein show Lewy-body-like aggregates in the cortex (arrow) and neuropil (arrowheads) that immunolabel for α-synuclein (*A*) and the fly stress-induced Hsp70 (*B*). (*C* and *D*) Tissue from the substantia nigra of a patient with Parkinson's disease showing Lewy bodies and Lewy pathology that immunolabel for α-synuclein (*C*) and Hsp70 (*D*). The commonalities between fly and human suggest that chaperone activity may modulate human Parkinson's disease (see ref. 10). (Bar = 3 μm.)

To address this question in another manner, we performed Western immunoblot analysis on the flies and examined the solubility properties of the pathogenic protein. By this assay, the pathogenic protein remains largely insoluble in SDS-resistant complexes that fail to enter the protein gel, remaining within the stacking gel, and transferring poorly in immunoblot analysis. However, in flies that are co-expressing the chaperones, a large amount of the pathogenic protein is now SDS-soluble and detected as a monomeric protein by Western immunoblot (Fig. 2*E*) (11).

The degree of SDS-solubility strikingly correlates with pathogenicity of the protein—dHdj2, which suppresses poorly, shows little or no change in monomer, despite high levels of coexpressed chaperone. These data suggest that the properties of the pathogenic protein have changed in the presence of the chaperones. Potentially, the protein is being maintained in a more native or normal conformation, with toxic interactions being abated, seen as a change in SDS-solubility.

Chaperone Suppression of α-Synuclein Toxicity in a *Drosophila* Model for Parkinson's Disease

The demonstration that *Drosophila* can be used to model a human neurodegenerative disease by directed expression of the respective human disease protein opened the possibility of modeling human neurodegenerative diseases other than poly-glutamine diseases in the fly. Indeed, directed expression of α-synuclein, a component of Lewy bodies and mutated in familial forms of Parkinson's disease, causes adult-onset degeneration of dopaminergic neurons in *Drosophila*, thereby providing a model for Parkinson's disease (Fig. 3) (10, 22). We have
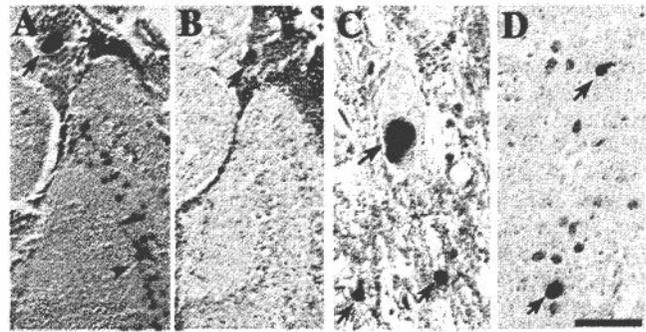
also determined whether we could apply the principles learned from chaperone suppression of polyglutamine pathogenicity to the problem of protein toxicity of α-synuclein in *Drosophila*.

When α-synuclein expression is directed to dopaminergic neurons in the *Drosophila* brain, select clusters of neurons show adult-onset progressive loss of cells. To define this cell loss, we prepared serial brain sections and immunostained for tyrosine hydroxylase expression, which selectively detects the dopaminergic neurons. We then counted the number of dopaminergic neurons present in the dorsomedial (DM) and dorsolateral (DL-1) clusters in the adult fly over time. We found a consistent 50% loss of dopaminergic neurons in the DM cluster, and a variable 0–50% loss of cells within the DL-1 cluster (10). Normal cell numbers were present at eclosion of the adult fly from the pupal case, with the cells degenerating over 20 days of adult life. We did not see further loss of cells, indicating that by 20 days all of the cells sensitive to α-synuclein toxicity had degenerated. Moreover, we did not detect a difference in the toxicity of normal α-synuclein, or the two mutant forms A30P and A53T.

In flies, normal α-synuclein and the mutant forms form abnormal Lewy-body-like and Lewy-neurite-like accumulations in the brain over time (Fig. 3). The Lewy-body-like aggregates appear as smaller, more loosely formed accumulations at 1 day, and become progressively larger by 20 days. As with human Lewy bodies, the fly Lewy-body-like aggregates immunolabel with antibodies to ubiquitin, indicating they may reflect accumulation of misfolded protein targeted for degradation by the proteasome. Most Parkinson's disease is sporadic and associated with accumulation of normal α-synuclein in Lewy bodies (2), consistent with the toxicity and Lewy-body-like formation by normal α-synuclein, in a manner similar to the mutant forms, in *Drosophila*.

We then asked whether co-expression of Hsp70 had an effect on α-synuclein toxicity. We co-expressed human Hsp70 with α-synuclein and counted the number of dopaminergic neurons in the DM clusters, where we detect a consistent loss of neurons upon α-synuclein expression alone.

Hsp70 had a dramatic effect to maintain dopaminergic neural numbers and prevent the degeneration of dopaminergic neurons (10). Whereas normally upon α-synuclein expression, 50% of neurons in the DM cluster were lost over 20 days in the adult, now all neurons were maintained over the 20-day period. This was the case upon expression of the wild-type α-synuclein, as well as the mutant forms A30P and A53T. We examined whether this

protection was accompanied by a change in formation of the Lewy-body-like aggregates; however, as with polyglutamine degeneration, we detected no change. This result indicates that Hsp70 fully protects against α-synuclein toxicity, despite the continued presence of aggregates. The aggregates immunolabel for the exogenous Hsp70, however, indicating a potential direct interaction of the chaperone with α-synuclein. We then asked whether there was a change in distribution of endogenous chaperones, and indeed we found that the aggregates immunolabel for the stress-induced form of fly Hsp70 (Fig. 3). This finding indicates that there might be an involvement of endogenous chaperones, and potentially a stress response, in α-synuclein toxicity.

Endogenous Chaperone Activity Plays a Role in α-Synuclein Toxicity in *Drosophila* and Potentially also in Parkinson's Disease

To ask whether endogenous chaperone levels may normally help protect against α-synuclein toxicity, the dominant-negative form of Hsc4 was co-expressed with the disease protein. This form of Hsc4 will interfere with endogenous activity of the Hsp70 family of chaperones, in effect lowering endogenous chaperone activity (19). In this situation, we noted an acceleration of α-synuclein toxicity. Whereas normally upon expression of α-synuclein, flies are born with the full complement of dopaminergic neurons in the DM clusters, in the presence of Hsc4.K71S, flies are born with a 50% loss of dopaminergic neurons. This cell loss did not progress further over time in the adult; rather, Hsc4.K71S accelerated the toxicity of α-synuclein to those cells sensitive to α-synuclein. However, we also noted that Hsc4.K71S has some toxicity to dopaminergic neurons when expressed in the absence of α-synuclein. The fact that α-synuclein and Hsc4.K71S are acting similarly with regard to dopaminergic neural loss indicates that the toxicity in both cases may share common mechanisms.

To extend our findings back to the human disease condition, we asked whether Parkinson's disease was associated with a change in chaperone function. To do this, we immunostained patient tissue with antibodies to Hsp70 and its co-chaperone Hsp40. Indeed, Lewy bodies and Lewy neurites in disease brain immunolabel for the chaperones (Fig. 3). The significance of this finding is best evaluated in the context of the fly study: in *Drosophila* the abnormal inclusions of α-synuclein immunolabel for Hsp70 and up-regulation of Hsp70 activity by directed transgenic expression mitigates the toxicity. This finding suggests that it may be of value in Parkinson's disease to up-regulate chaperone function.

Discussion

Our initial and other subsequent studies have established *Drosophila* as a model genetic system to bring to bear in the arsenal of approaches toward the combat of human neurodegenerative disease (10, 12, 13, 20, 22–26). The striking homology of a large number of fly genes with human genes—indeed of entire gene pathways—indicates that fundamental properties of degeneration modeled in flies may be conserved in humans. Whereas the demonstration of the fly as an outstanding *in vivo* model for human neurodegeneration is significant, of great importance is the use of those models to reveal disease mechanisms and pioneer ways to interfere in the disease process.

The implication that chaperones may be of interest in such neurodegenerative diseases finds its roots in an even simpler model system than *Drosophila*—the yeast *Saccharomyces cerevisiae*. Indeed, protein conformational changes relevant to human prion disease are found in yeast in the study of endogenous yeast prions such as Sup35 (27), where Hsp104 was described as a regulator of the prion state (28). The *Drosophila* studies establish that chaperone modulation can be applied to the nervous system *in vivo* in the context of neurodegeneration.

In *Drosophila*, the chaperones are found to be potent modulators both of polyglutamine toxicity and of α-synuclein toxicity in models for human polyglutamine disease and Parkinson's disease. Not only does added chaperone activity prevent disease, but interfering with endogenous chaperone activity accelerates pathogenesis. This finding indicates that chaperone activity is central to the disease process, being modulated upon both up-regulation and interference. Moreover, these studies in *Drosophila* for polyglutamine toxicity have been found to translate to mammalian models for polyglutamine toxicity. Up-regulation of Hsp70 in transgenic mice expressing the Ataxin-1 pathogenic polyglutamine disease protein leads to protection against behavioral and cellular pathology (29).

How are the chaperones modulating protein toxicity? *In vivo* in flies, chaperones modulate the solubility properties of the polyglutamine protein concomitant with a modulation in the toxicity. However, no morphological change in the aggregates is detected. This is the case for both polyglutamine and α-synuclein toxicity. Studies of polyglutamine aggregation and protein solubility in yeast remarkably parallel the fly findings, where chaperones modulate solubility but aggregates are still present (30). One possibility is that the chaperones are modulating the structure of the protein, and this is not visible in the large aggregates. Prefibrils or protofibrils may be the toxic entity (see refs. 31 and 32)—smaller clumps and misconformations of the disease protein that precede or are independent of large, visible aggregates. Chaperones may be modulating these abnormal and toxic conformations, thereby preventing neurodegeneration. This conceivably can happen in the absence of an effect on large, visible aggregates, which are a different form of the disease protein, perhaps even inert or protective.

Another possibility is that chaperones, by interacting with the disease protein, prevent abnormal interactions with other proteins in the cell that are causal in toxicity. However, clearly, just anything that physically interacts with the disease protein appears not to have an effect to suppress, as demonstrated by the dHdj2 studies for polyglutamine disease. Whereas dHdj-2 is in association with the polyglutamine protein in the aggregates, as it co-localizes by immunocytochemistry, it fails to suppress toxicity (11).

Another possibility is that of chaperone depletion. Because of an abnormal or misfolded conformation, pathogenic disease proteins may cause cellular depletion of chaperone activity. Chaperones are required for the proper folding and function of many cellular proteins with diverse roles. Therefore, slow depletion of chaperones from the cellular milieu could lead to the failure of many cellular processes.
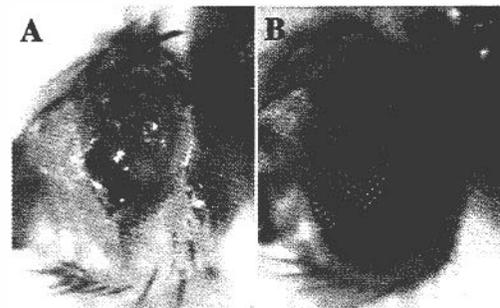

Fig. 4. Interfering with endogenous chaperone activity causes a severely degenerate eye phenotype. (*A*) Expression of a dominant-negative form of Hsp70 causes loss of pigmentation and deterioration of eye structure, similar to the expression of the polyglutamine protein in flies (see Fig. 1*C*). The genotype of the fly was *w; gmr-GAL4 UAS-Hsp70.K71E*. (*B*) Normal fly eye for comparison.

The role of chaperone depletion is intriguing because of the ability of chaperones to phenocopy aspects of disease protein expression. In the α-synuclein studies, expression of Hsc4.K71S on its own caused some loss of dopaminergic neurons, indicating that dopaminergic neurons are sensitive to compromised chaperone activity [also supported by other studies of α-synuclein and Parkinson's disease-associated proteins (33, 34)]. In the studies of the Hsc4.K71S transgene in flies, Elefant and Palter (19) noted that in some situations the protein was toxic, inducing a misfolded protein response accompanied by a degeneration reminiscent of neurodegenerative disease. We have also found that expressing Hsp70.K71E in the eye causes an external eye phenotype strikingly similar to the polyglutamine disease phenotype (Fig. 4). These studies suggest that compromising chaperone levels alone is phenotypically similar to the pathogenic actions of polyglutamine and α-synuclein proteins, indicating that chaperone interference is a major contributing factor to neurodegeneration. Cell specificity differs between polyglutamine and α-synuclein toxicity, with pathogenic polyglutamine protein appearing to be much more generally toxic in *Drosophila* than α-synuclein. However, this observation could be explained by cellular differences in the chaperone response to the specific protein in different tissues.

With the ever-accelerating development of fly models for various human neurodegenerative diseases, and tremendous interest in such models for both standard genetic and pharmacological approaches, *Drosophila* may reveal new cures and treatments of relevance to human neurodegeneration, including polyglutamine and Parkinson's diseases.

1. Zoghbi, H. Y. & Orr, H. T. (2000) *Annu. Rev. Neurosci.* **23,** 217–247.
2. Spillantini, M. G., Schmidt, M. L., Lee, V. M., Trojanowski, J. Q., Jakes, R. & Goedert, M. (1997) *Nature (London)* **388,** 839–840.
3. Kruger, R., Kuhn, W., Muller, T., Woitalla, D., Graeber, M., Kosel, S., Przuntek, H., Epplen, J. T., Schols, L. & Riess, O. (1998) *Nat. Genet.* **18,** 106–108.
4. Polymeropoulos, M. H., Lavedan, C., Leroy, E., Ide, S. E., Dehejia, A., Dutra, A., Pike, B., Root, H., Rubenstein, J., Boyer, R., *et al.* (1997) *Science* **276,** 2045–2047.
5. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287,** 2185–2195.
6. Benzer, S. (1967) *Proc. Natl. Acad. Sci. USA* **58,** 1112–1119.
7. Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G. & Benzer, S. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 1684–1688.
8. Benzer, S. (1971) *J. Am. Med. Assoc.* **218,** 1015–1022.
9. Heisenberg, M. & Bohl, K. (1979) *Z. Naturforsch.* **34,** 143–147.
10. Auluck, P. K., Chan, H. Y., Trojanowski, J. Q., Lee, V. M. & Bonini, N. M. (2002) *Science* **295,** 865–868.
11. Chan, H. Y., Warrick, J. M., Gray-Board, G. L., Paulson, H. L. & Bonini, N. M. (2000) *Hum. Mol. Genet.* **9,** 2811–2820.
12. Warrick, J. M., Chan, H. Y., Gray-Board, G. L., Chai, Y., Paulson, H. L. & Bonini, N. M. (1999) *Nat. Genet.* **23,** 425–428.
13. Warrick, J. M., Paulson, H. L., Gray-Board, G. L., Bui, Q. T., Fischbeck, K. H., Pittman, R. N. & Bonini, N. M. (1998) *Cell* **93,** 939–949.
14. Kawaguchi, Y., Okamoto, T., Taniwaki, M., Aizawa, M., Inoue, M., Katayama, H., Nakamura, S., Nishimura, M., Akiguchi, I., Kimura, J., *et al.* (1994) *Nat. Genet.* **8,** 221–228.
15. Brand, A. H. & Perrimon, N. (1993) *Development (Cambridge, U.K.)* **118,** 401–415.
16. Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L. & Bates, G. P. (1997) *Cell* **90,** 537–548.
17. DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P. & Aronin, N. (1997) *Science* **277,** 1990–1993.
18. Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J.-L., Fischbeck, K. H. & Pittman, R. N. (1997) *Neuron* **19,** 333–344.
19. Elefant, F. & Palter, K. (1999) *Mol. Biol. Cell* **10,** 2101–2117.
20. Kazemi-Esfarjani, P. & Benzer, S. (2000) *Science* **287,** 1837–1840.
21. Sakahira, H., Breuer, P., Hayer-Hartl, M. K. & Hartl, F. U. (2002) *Proc. Natl. Acad. Sci. USA* **99,** Suppl. 4, 16412–16418.
22. Feany, M. B. & Bender, W. W. (2000) *Nature (London)* **404,** 394–398.
23. Fernandez-Funez, P., Nino-Rosales, M. L., de Gouyon, B., She, W.-C., Luchak, J. M., Martinez, P., Turiegano, E., Benito, J., Capovilla, M., Skinner, P. J., *et al.* (2000) *Nature (London)* **408,** 101–106.
24. Jackson, G., Salecker, I., Dong, X., Yao, X., Arnheim, N., Faber, P., MacDonald, M. & Zipursky, S. (1998) *Neuron* **21,** 633–642.
25. Marsh, J. L., Walker, H., Theisen, H., Zhu, Y., Fielder, T., Purcell, J. & Thompson, L. M. (2000) *Hum. Mol. Genet.* **9,** 13–25.
26. Wittmann, C. W., Wszolek, M. F., Shulman, J. M., Salvaterra, P. M., Lewis, J., Hutton, M. & Feany, M. B. (2001) *Science* **293,** 711–714.
27. Serio, T. R. & Lindquist, S. L. (2000) *Trends Cell Biol.* **10,** 98–105.
28. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. (1995) *Science* **268,** 880–884.
29. Cummings, C. J., Sun, Y., Opal, P., Antalffy, B., Mestril, R., Orr, H. T., Dillmann, W. H. & Zoghbi, H. Y. (2001) *Hum. Mol. Genet.* **10,** 1511–1518.
30. Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K. & Hartl, F. U. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7841–7846.
31. Bucciantini, M., Glannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddel, N., Ramponi, G., Dobson, C. & Stefani, M. (2002) *Nature (London)* **416,** 507–511.
32. Walsh, D., Klyubin, I., Fadeeva, J., Cullen, W., Anwyl, R., Wolfe, M., Rowan, M. & Selkoe, D. (2002) *Nature (London)* **416,** 535–539.
33. Imai, Y., Soda, M., Inoue, H., Hattori, N., Mizuno, Y. & Takahashi, R. (2001) *Cell* **105,** 891–902.
34. Shimura, H., Schlossmacher, M. G., Hattori, N., Frosch, M. P., Trockenbacher, A., Schneider, R., Mizuno, Y., Kosik, K. S. & Selkoe, D. J. (2001) *Science* **293,** 263–269.

Colloquium

Molecular chaperones as modulators of polyglutamine protein aggregation and toxicity

Hideki Sakahira, Peter Breuer, Manajit K. Hayer-Hartl, and F. Ulrich Hartl*

Max-Planck-Institut für Biochemie, Department of Cellular Biochemistry, Am Klopferspitz 18a, D-82152 Martinsried, Germany

The formation of insoluble protein aggregates in neurons is a hallmark of neurodegenerative diseases caused by proteins with expanded polyglutamine (polyQ) repeats. However, the mechanistic relationship between polyQ aggregation and its toxic effects on neurons remains unclear. Two main hypotheses have been put forward for how polyQ expansions may cause cellular dysfunction. In one model neurotoxicity results from the ability of polyQ-expanded proteins to recruit other important cellular proteins with polyQ stretches into the aggregates. In the other model, aggregating polyQ proteins partially inhibit the ubiquitin–proteasome system for protein degradation. These two mechanisms are not exclusive but may act in combination. In general, protein misfolding and aggregation are prevented by the machinery of molecular chaperones. Some chaperones such as the members of the Hsp70 family also modulate polyQ aggregation and suppress its toxicity. These recent findings suggest that an imbalance between the neuronal chaperone capacity and the production of potentially dangerous polyQ proteins may trigger the onset of polyQ disease.

Expansions of CAG trinucleotide repeats encoding polyglutamine (polyQ) tracts in otherwise unrelated proteins are responsible for at least nine different neurodegenerative diseases (1–3). These diseases include Huntington's disease (HD), spinobulbar muscular atrophy, dentatorubral pallidoluysian atrophy, and spinocerebellar ataxia (SCA) types 1, 2, 3, 6, 7, and 17. With the exception of spinobulbar muscular atrophy, these neurodegenerative diseases are inherited in an autosomal dominant manner. All nine disorders show late onset of neurological symptoms with progressive neuronal dysfunction and eventual neuronal loss, although the susceptible regions in the nervous system differ among the various disorders. Generally, the pathologic length of the polyQ repeat is ≈40 or greater, whereas healthy individuals have polyQ repeats with fewer residues. Longer CAG repeats result in an earlier age of onset and a more severe pathology, consistent with a disease mechanism by gain of function. These fundamental observations point to a common molecular mechanism underlying the pathology of polyQ diseases. Exactly how polyQ expansions cause neuronal dysfunction is still obscure, however.

A characteristic feature of polyQ diseases is the formation of insoluble, granular, and fibrous deposits in affected neurons termed neuronal inclusions, which have been studied extensively in HD (4). HD is characterized by selective neuronal loss, primarily in the cortex and striatum, leading to motor impairment, personality changes, and dementia. The disease is caused by the expansion of a polyQ segment located within the first exon of the gene encoding huntingtin, an ≈350-kDa protein of unknown but essential function (5, 6). The neuronal inclusions in HD have fibrillar morphology and contain aggregated amino-terminal fragments of huntingtin (7). Similar inclusions containing aggregated polyQ proteins are detected also in other polyQ diseases (8, 9), suggesting a causal

relationship between these neurodegenerative disorders and amyloid fibrillogenesis (10). However, it remains unclear whether the aggregates themselves are pathogenic, epiphenomenal, or even beneficial. For example, large polyQ aggregates may provide an advantage over small oligomers by exposing less potentially dangerous protein surfaces.

It is thought that the aggregates result from the ability of long polyQ stretches to form self-associating β-sheets (11, 12). Interestingly, certain transcription factors containing polyQ segments in the nonpathological range, such as TATA-binding protein (TBP) and CREB-binding protein (CBP), are detected in neuronal inclusions (13–15). It has been proposed that sequestration of these essential proteins via polyQ–polyQ interactions may cause neuronal toxicity. Additional proteins detected in neuronal inclusions are ubiquitin, the 19S and 20S proteasome complexes, and several molecular chaperones (8, 16–19). Importantly, the cellular components involved in protein folding and degradation are associated also with intracellular inclusions in other neurodegenerative diseases not caused by polyQ expansion, including Alzheimer's disease, Parkinson's disease, and the prion diseases (20), which suggests that common pathomechanistic principles may underlie these misfolding diseases in general.

In this review, we focus mainly on the molecular mechanism of polyQ aggregation and its cellular toxicity. The functional relationship between molecular chaperones, the ubiquitin–proteasome system, and polyQ aggregation will be discussed.

Mechanism of polyQ Aggregation

In 1994, Perutz proposed that long sequences of polyQ might be able to form stable β-hairpins (11). These structures, also called "polar zippers," consist of polyQ-containing β-strands held together by hydrogen bonds between both main-chain and side-chain amides. PolyQ-containing hairpins may self-associate, forming stable β-sheet aggregates with fibrillar morphology. In a recent, refined structural model, expanded stretches of polyQ are proposed to form a cylindrical (helical), parallel β-sheet rather than an antiparallel β-sheet (12). In this model, the number of Q repeats per turn in polyQ fibrils is estimated at ≈20, but a single 20-residue helical turn would be unstable. In contrast, a polyQ segment with ≈40 residues could be stabilized by amide hydrogen bonds between successive turns and could act as a nucleus for further growth of a helical fibril. The external and internal diameters of a cylindrical fibril are supposed to be
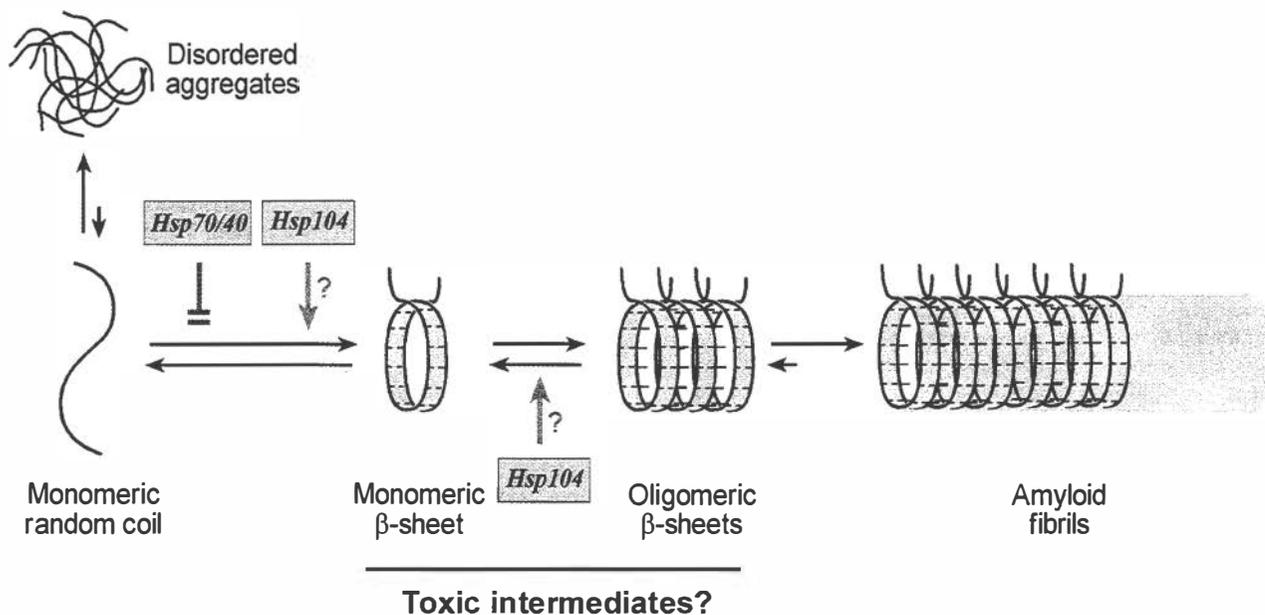
Fig. 1. Hypothetical model for the pathway of polyQ aggregation and its modulation by molecular chaperones. The first step in the aggregation process is thought to be the structural conversion of polyQ monomers from a random coil to β-sheet conformation, followed by oligomerization. Once oligomers have formed, polymerization into amyloid-like fibrils occurs rapidly. Monomers or soluble oligomers with β-sheet conformation may be toxic intermediates in the aggregation process, whereas amyloid fibrils as the end product of aggregation may be nontoxic. The Hsp70/Hsp40 chaperone system is proposed to prevent the initial conformational conversion and eventually chaperone-associated disordered aggregates form. In contrast, Hsp104 is proposed to promote conversion into β-sheet conformation at low relative concentration to polyQ protein but at high concentrations may dissociate soluble polyQ oligomers, thereby slowing the aggregation process.

≈30 and 10 Å, respectively. Thicker amyloid fibers may consist of two or more cylindrical β-sheet fibrils wound around one another (10). A parallel β-sheet structure has been suggested also for the amyloid fibrils formed by the yeast prion protein Sup35 and the Alzheimer's Aβ peptide (21–23).

The predicted length dependence of polyQ aggregation was confirmed by *in vitro* experiments. Wanker and coworkers showed that amino-terminal huntingtin fragments with polyQ tracts exceeding a critical length of ≈40 residues form SDS-resistant aggregates with a fibrillar morphology (24), similar to that of Aβ amyloid and yeast prion protein Sup35 fibrils. It has been proposed that fibril formation in general occurs by a mechanism of nucleated polymerization (10, 25). This mechanism is characterized by the rate-limiting formation of an oligomeric nucleus from monomers that have undergone a (transient) conformational change followed by rapid recruitment of further monomers or oligomers into highly ordered fibrils. A model of how such a mechanism may apply to polyQ fibrils is shown in Fig. 1. In this model, the first step in the aggregation process is proposed to be the repeat length-dependent conformational change of polyQ monomers from random coil to a parallel, helical β-sheet (12). This structural conversion then results in the formation of ordered polyQ oligomers that could function as nuclei for the rapid polymerization of amyloid-like fibrils. Either the initial conversion or the subsequent oligomerization may be the rate-limiting step in polyQ fibrillogenesis, dependent on protein concentration. The intramolecular formation of β-sheet structure is likely to be rate-limiting only at high concentrations of polyQ protein, as reported for the Sup35 protein (26). In the case of huntingtin exon 1, proteolytic cleavage of the polyQ-containing segment from a nonaggregating precursor construct *in vitro* initiates aggregation, presumably by relieving steric restrictions and facilitating both intramolecular conversions and formation of oligomeric nuclei (24, 27). Thus, after cleavage, the concentration-dependent formation of soluble polyQ oligomers is likely to

be rate-limiting in the aggregation of this protein. Indeed, both full-length huntingtin (350 kDa) and the SCA3 protein (ataxin-3, 42 kDa) are cleaved by as-yet-unidentified proteases *in vivo* (7, 28–30). The resulting production of polyQ-containing fragments dramatically enhances the formation of intracellular inclusions in both transgenic animals and cell-culture systems (28, 31, 32).

Relationship Between polyQ Aggregation and Toxicity

There are several lines of evidence that suggest a causal link between polyQ aggregation and the disease process. The numbers of neuronal inclusions in patient brains and the severity of neurological symptoms correlate with the polyQ repeat length of the expressed protein (33). Furthermore, the inclusions are present primarily in those neurons that are particularly vulnerable to the disease (7–9). These phenomena have been reproduced in both transgenic animals and cell-culture systems. In transgenic mice expressing polyQ-expanded HD exon 1, neuronal inclusions containing aggregated HD protein form before the onset of neurological symptoms (16). Regulated expression of HD exon 1 in a conditional transgenic mouse model resulted in a progressive neurological phenotype with polyQ inclusions in the striatum and cortex (34). Unexpectedly, when expression of HD exon 1 protein was switched off after polyQ aggregates had formed, neurological symptoms disappeared along with the intracellular inclusions. Although these observations do not establish a causal relationship between the inclusions and either disease initiation or progression, they link the process of polyQ protein aggregation to cellular dysfunction and disease.

Elucidating the mechanism(s) by which polyQ aggregation exerts cellular toxicity now represents one of the most challenging problems in the field. Interestingly, there are several reports that the neuronal inclusions themselves are not necessary for the initiation of symptoms and may even help to protect against cellular dysfunction (35, 36). Saudou *et al.* showed that expression of a dominant-negative ubiquitin-conjugating enzyme in cell culture reduces huntingtin aggregation but enhances its toxicity

(36). Zoghbi and coworkers reported similar results in SCA1 transgenic mice (37). These studies strongly indicate that microscopically detectable inclusions are not necessary for disease initiation. As mentioned above, the critical step in the aggregation process is thought to be an intramolecular conformational change in the polyQ protein that precedes the formation of an oligomeric nucleus (Fig. 1). It is possible that these nuclei or other early, i.e., prefibrillar, oligomeric intermediates in the aggregation process are the toxic agents. This idea parallels the recent demonstration by two groups that early stages in the process of protein misfolding and amyloid fibril formation are important for cellular toxicity, whereas the fibrils themselves are not toxic (38, 39). Specifically, soluble dimers and trimers of the Aβ peptide impair neuronal functions in rat brains but not Aβ monomers, protofibrils, and fibrils (39). The toxic aggregation intermediates of Aβ peptide accumulate extracellularly and may damage cell surface structures. Similarly, the unprotected β-sheets in intracellular polyQ oligomers or even in the monomeric polyQ fragments generated from full-length precursor proteins may interact unfavorably with the surfaces of other proteins, thereby impairing various cellular functions (Fig. 1). To critically test this "toxic-intermediate" hypothesis of polyQ disease, it will be important to define the biochemical and biophysical properties of the exposed β-sheets in the soluble states of polyQ proteins.

Transcriptional Dysregulation by polyQ Aggregation

Analyzing the cellular components contained in neuronal polyQ inclusions may offer valuable clues as to how polyQ expansions cause cellular dysfunction. To date, numerous proteins have been shown to be sequestered into the inclusions using immunohistochemical and biochemical approaches. One class of recruited proteins comprises essential transcription factors with nonpathological length polyQ tracts, such as TATA-binding protein and CBP (13–15, 40). For example, Ross and coworkers reported that CBP is depleted from its normal nuclear location and becomes sequestered into polyQ aggregates in HD cell-culture models, HD transgenic mice, and human HD postmortem brains (15). The CBP homolog p300, lacking a substantial polyQ stretch, is not recruited to neuronal inclusions, suggesting that sequestration occurs via polyQ–polyQ interactions. Importantly, the expression of expanded polyQ proteins specifically interfered with CBP-activated gene transcription, causing cellular toxicity. Gene-array studies also showed that expression of genes controlled by cAMP-response elements (CREs) is downregulated in HD transgenic mice or cell-culture models (41, 42). More recently, transgenic mice with disruptions in two CRE-binding proteins (CREB1 and CREM) in the adult forebrain were shown to develop progressive neurodegeneration in the dorsolateral striatum reminiscent of HD (43). Taken together, at least in HD, dysregulation of CREB-mediated transcription by sequestration of CBP into neuronal inclusions may cause cellular toxicity.

Considering that alterations in gene expression have been reported in several other polyQ disorders, including SCA1, SCA3, and dentatorubral pallidoluysian atrophy (44, 45), sequestration of transcription factors into neuronal inclusions provides an attractive explanation for the toxicity common to these diseases. Because transcription factors act in the nucleus, this model also would explain the greater toxicity of intranuclear versus cytosolic polyQ protein aggregation. Whether recruitment is mediated by (toxic) intermediates in the aggregation process or the final polyQ fibrils is still unclear. On the other hand, additional, alternative mechanisms also may be involved in polyQ-related transcriptional dysregulation. For example, TAF$_{II}$130, another important positive regulator of CREB-mediated transcription, was shown to bind directly to long polyQ stretches in SCA3 and dentatorubral pallidoluysian atrophy,

even though TAF$_{II}$130 lacks a polyQ tract (45). Long polyQ stretches also may cause a pronounced inhibition of the histone acetyltransferase (HAT) activities of CBP and other proteins such as P/CAF (p300/CBP-associated factor) by binding to their acetyltransferase domains, not their polyQ tracts (46). Interestingly, administration of histone deacetylase inhibitors had remarkable beneficial effects in a fly model of HD (46). Although it is unclear at the moment which mechanism(s) described above is primarily responsible for alterations in gene expression by polyQ expansion proteins, transcriptional dysregulation is clearly one important element of polyQ toxicity (Fig. 2A).

Role of Molecular Chaperones in polyQ Aggregation and Toxicity

Components involved in protein folding and degradation represent another group of proteins frequently recruited to polyQ inclusions. These factors include several molecular chaperones such as Hsp70 and Hsp40 as well as ubiquitin and the 20S and 19S proteasome complexes (8, 16–19, 40). Their presence in aggregates suggests that expanded polyQ tracts are recognized as misfolded conformers and that cellular quality-control mechanisms are activated in an attempt to prevent their accumulation (20).

Hsp70 chaperones promote protein folding by an ATP-dependent process of binding and release of extended polypeptide segments enriched in hydrophobic residues that are typically exposed by nonnative, i.e., fully or partially unfolded, proteins (47, 48). These structural features often give rise to intermolecular association (aggregation) mediated by hydrophobic interactions and β-sheet formation (49). Hsp70 binding may prevent protein aggregation directly by shielding the interactive surfaces of nonnative polypeptides and indirectly by inhibiting or reversing intramolecular misfolding. Hsp70 cooperates in this function with members of the Hsp40 family. These cochaperones are homologs of bacterial DnaJ and contain a so-called J domain. They have a critical role in mediating substrate binding to Hsp70 but alone are inefficient in preventing aggregation. Most Hsp40s recognize nonnative polypeptide segments and target them to Hsp70 by a direct interaction. In addition, the J domain of Hsp40 activates the Hsp70 ATPase, thereby catalyzing the formation of the ADP-state of Hsp70, which binds protein substrate tightly (48). This Hsp40 effect may be particularly important for the binding by Hsp70 of extended sequences not containing hydrophobic amino acid residues (50), such as polyQ. Expression of Hsp70 and Hsp40 chaperones is induced under various conditions of cell stress such as heat shock, which result in unfolding and aggregation of certain proteins.

There are several reports that increased expression of the Hsp70/Hsp40 chaperone system can suppress polyQ-induced neurotoxicity in fly models (51–54) and a mouse model of polyQ disease (ref. 55; Table 1) [reviewed by Bonini in this issue (ref. 56)]. These studies identified the Hsp40 protein dHdj1, the *Drosophila* homolog of human Hdj1, and the J-domain cochaperone dTPR2 as well as dHsp70, a *Drosophila* homolog of human Hsp70, as active components (51–54). Suppression of toxicity by expression of Hdj1 or dTPR2 alone is most likely caused by the ability of these cochaperones to activate the endogenous Hsp70 for polyQ protein binding, an effect that may be less pronounced with other Hsp40 homologs. In contrast, expression of a dominant negative mutant form of Hsp70 increased polyQ toxicity (52). The finding by Cummings *et al.* that overexpression of Hsp70 suppresses polyQ-induced toxicity in SCA1 transgenic mice (55) validates molecular chaperones and the components involved in regulating their expression as promising targets in developing a possible therapy for polyQ diseases (57).

Surprisingly, in all cases studied, overexpression of Hsp70/Hsp40 chaperones did not prevent the formation of polyQ aggregates, although polyQ toxicity was suppressed. Experiments by Muchowski *et al.* (58) helped to resolve this puzzle.
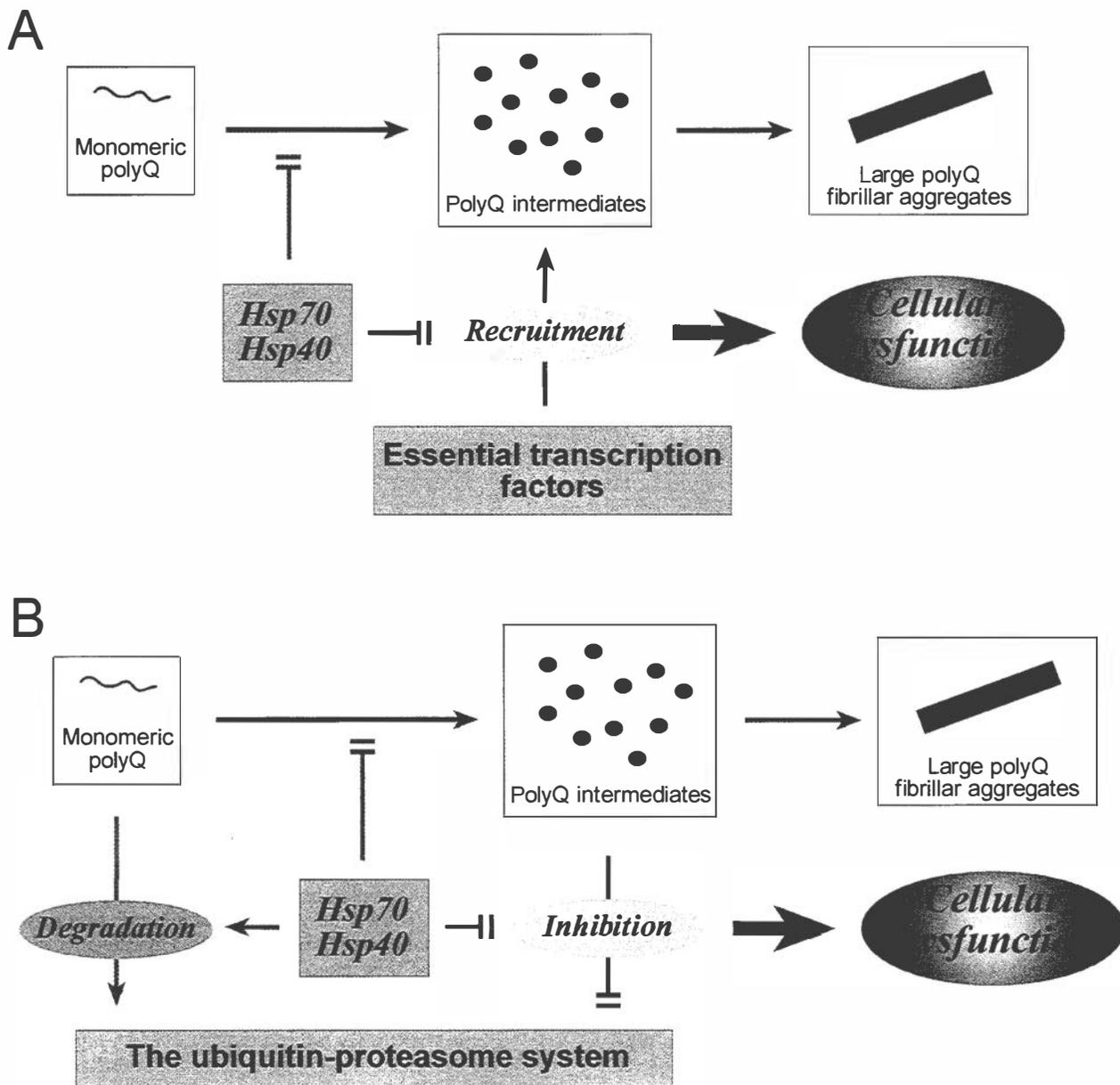
Fig. 2. Two models for polyQ-mediated toxicity and its suppression by the Hsp70/Hsp40 chaperone system. (*A*) Transcriptional dysregulation. Intermediates of polyQ aggregation recruit essential transcription factors, thereby inhibiting their transcriptional activities and causing cellular dysfunction. The Hsp70/Hsp40 chaperone system prevents this recruitment and thus mitigates polyQ toxicity. (*B*) Inhibition of the ubiquitin–proteasome system. PolyQ aggregation intermediates trap the 19S regulatory complex (and other components of the degradative machinery), resulting in a partial inhibition of proteasome-dependent proteolysis and eventually cellular dysfunction. The Hsp70/Hsp40 chaperone system may prevent or reduce this effect by stabilizing polyQ protein in a soluble, degradation-competent state and shielding aggregates.

They demonstrated that these molecular chaperones, when present at sufficient levels, profoundly modulate the aggregation process and the physical properties of the resulting polyQ inclusions without significantly changing their appearance in the fluorescence microscope. Purified mammalian Hsp70 and Hsp40 (Hdj1) suppressed, in an ATP-dependent manner, the *in vitro* assembly of polyQ-expanded HD exon 1 constructs into ordered, SDS-insoluble amyloid fibrils and instead allowed the formation of amorphous, SDS-soluble aggregates that were associated with Hsp70 (ref. 58; Fig. 1). This effect of the chaperones was reproduced in HD exon 1-expressing yeast (58) and mammalian COS-1 cells (57) by coexpressing Hsp70 and Hsp40 homologs. In addition, Bonini and coworkers demon-

strated in their *Drosophila* disease model that overexpression of Hsp70/Hsp40 strongly increases the SDS-solubility of polyQ aggregates (53). Considering that Hsp70 (in concert with Hsp40) binds to extended polypeptide segments, it may inhibit the formation of intramolecular β-sheet conformation and thus block ordered oligomerization and fibril growth (Fig. 1). On the other hand, binding of Hsp70 to the polyQ segments must be transient and of relatively low affinity, because Hsp70 cycles its substrates in an ATP-dependent manner, and glutamine is not a preferred residue in Hsp70-binding peptides (59). Consequently, the formation of irregular hydrogen bonds between polyQ sequences, resulting in amorphous aggregation, might not be suppressed efficiently except at a high molar excess of Hsp70.

Table 1. Effects of molecular chaperones on polyQ aggregation and toxicity *in vivo*

Chaperones	Organisms	PolyQ disease model	Aggregation	Toxicity	References
Hsp70 ↑	Mouse	SCA-1	→	↓	55
dHsp40 (dHdj1) ↑	*Drosophila*	HD	→	↓	51
dTPR2 ↑	*Drosophila*	HD	→	↓	51
dHsp70 ↑	*Drosophila*	SCA-3	→	↓	52
dHsp70 ↓	*Drosophila*	SCA-3	n.d.	↑	52
dHsp40 (dHdj1) ↑	*Drosophila*	SCA-3	n.d.	↓	53
dHsp40 (dHdj1) ↓	*Drosophila*	SCA-3	n.d.	↑	53
dHsp40 (dHdj2) ↑	*Drosophila*	SCA-3	n.d.	↓	53
dHsp70 ↑ + dHdj1 ↑	*Drosophila*	SCA-3	→	↓↓	53
dHsp40 (dHdj1) ↑	*Drosophila*	HD	n.d.	↓	53
dHsp40 (dHdj2) ↑	*Drosophila*	HD	n.d.	→	53
dHsp40 (dHdj1) ↑	*Drosophila*	SCA-1	→ (more compact)	↓	54
dVCP/Cdc48 ↓	*Drosophila*	SCA-3	→	↓	63
dVCP/Cdc48 ↑	*Drosophila*	SCA-3	→	↑	63
Hsp104 ↑	*C. elegans*	DRPLA	↓	↓	66
Hsp70 (SSA1) ↑	*S. cerevisiae*	HD	↓	(−)	62
Hsp104 ↑	*S. cerevisiae*	HD	↓	(−)	62
Hsp104 ↓	*S. cerevisiae*	HD	↓↓	(−)	62
Hsp104 ↓	*S. cerevisiae*	SCA-3	↓	(−)	67

n.d., not determined.

Assuming that this proposed mechanism is correct in outline, one would predict that Hsp70 and Hsp40 inhibit the recruitment of other polyQ-containing proteins such as CBP and TATA-binding protein into polyQ aggregates, explaining in part how the chaperones may mitigate polyQ-induced neurotoxicity (Fig. 2*A*). In addition, the observed association of Hsp70 with the aggregates may result in the coating of potentially dangerous surfaces. Indeed, the normal cellular levels of Hsp70/Hsp40 (and other chaperones) may be sufficient to control the damaging effects of polyQ-expanded proteins for decades, but eventually aging processes may result in a reduction of the available chaperone capacity (20, 60). Thus, a shift in the balance between cellular chaperone capacity and production of polyQ-expanded protein may be crucial in triggering the onset of disease ("chaperone hypothesis of polyQ disease"; ref. 58). Longer polyQ sequences would require more chaperone binding to avoid a toxic aggregation pathway, and therefore patients expressing such sequences would develop neuronal dysfunction earlier in life.

Interestingly, increased expression of Hsp70 also suppresses the toxicity induced by the non-polyQ-containing protein α-synuclein in a fly model for Parkinson's disease, again without altering the microscopic appearance of the neuronal inclusions formed (61). This observation points to the exciting possibility that the Hsp70/Hsp40 chaperone system has a general potential in mitigating the toxicity caused by misfolding proteins.

Thus far only two other chaperones structurally unrelated to Hsp70/Hsp40 have been identified as modulators of polyQ aggregation (Table 1). Both belong to the Hsp100/Clp family of AAA proteins (ATPases associated with various cellular activities; refs. 62 and 63). The yeast chaperone Hsp104 and its bacterial homolog, ClpB, can solubilize small protein aggregates in concert with Hsp70/Hsp40 (64, 65). Coexpression of Hsp104 with expanded polyQ protein reduced the formation of polyQ aggregates in yeast (62) and *Caenorhabditis elegans* (66). In *C. elegans* expression of polyQ protein in body wall muscle cells strongly impairs the ability of the animals to move, an effect that was shown to be alleviated by the expression of yeast Hsp104 (66). Interestingly, deletion of Hsp104 in yeast also prevented the aggregation of expanded polyQ proteins (62, 67), thus recapitulating the effects of Hsp104 on the aggregation of the yeast prion Sup35, which contains a Q/N-rich amino-terminal domain

that mediates aggregation (68). It is proposed that Hsp104 catalyzes a structural change of Sup35 into an aggregation-prone conformation (68). Similarly, Hsp104 may facilitate the conversion of a polyQ random coil into β-sheet conformation, perhaps by relieving a steric block exerted by sequences adjoining the polyQ repeat (Fig. 1). Considering that overexpression of Hsp104 also suppresses polyQ aggregation, it is possible that excess amounts of the chaperone, in cooperation with Hsp70/Hsp40, effectively dissociate polyQ oligomers that nucleate the aggregation process (Fig. 1). A mammalian counterpart of Hsp104 has not been identified yet, but VCP (valosin-containing protein, the mammalian homolog of yeast Cdc48) is a distantly related AAA protein involved in membrane fusion, protein disassembly, and degradation (69–72). Higashiyama *et al.* recently reported that loss-of-function mutants of VCP suppress polyQ-mediated toxicity in SCA3 transgenic flies without inhibiting the formation of visible polyQ aggregates (63). Further *in vitro* and *in vivo* studies will be required to investigate whether VCP/Cdc48 affects polyQ aggregation by a mechanism similar to that of Hsp104.

Role of the Ubiquitin–Proteasome System in polyQ Aggregation and Toxicity

The ability of molecular chaperones to prevent (or reverse) protein aggregation is also important in aiding the proteolytic degradation of proteins that cannot be refolded (48). Most cytosolic proteins destined for degradation are marked by covalent attachment of a polyubiquitin chain at lysine residues (73). In this process, ubiquitin is activated first by the ubiquitin-activating enzyme (E1) and then transferred to a ubiquitin-conjugating enzyme (E2). The latter then links the activated ubiquitin to the protein substrate in functional cooperation with an E3 ubiquitin ligase, which acts as a specificity factor. Polyubiquitinated protein substrate is recognized and degraded by a large molecular machine, the 26S proteasome, which consists of a barrel-shaped proteolytic core complex of 20S, capped at both ends by 19S regulatory complexes (74). The 19S cap can be divided further into two subcomplexes, the "lid" and the "base." The lid forms the distal part of the cap and functions in recognition and binding of polyubiquitinated substrate proteins. The base contacts the 20S core and contains a ring of six AAA

ATPases that mediate unfolding and translocation of the substrate into the proteolytic chamber of the 20S complex. Concomitantly with translocation, ubiquitin molecules are released from the substrate and recycled. Importantly, the ubiquitin–proteasome system does not only participate in normal protein turnover, but its activity is required also for essential regulatory functions in a variety of cellular processes (74). As shown recently, these functions also include a role of the 19S complex in transcriptional regulation (75), which may impact on the transcriptional dysregulation in polyQ disease described above.

The finding that polyQ inclusions stain positively for ubiquitin and the 20S and 19S complexes suggested that the ubiquitin–proteasome system may be involved in polyQ pathogenesis (17–19). There are several reports that formation of polyQ inclusions is accelerated when proteasome inhibitors are added to transfected cells (18, 19, 37). Indeed, soluble HD exon 1 proteins are degraded in a proteasome-dependent process, as demonstrated by pulse–chase experiments in a Chinese hamster ovary cell-culture model (P.B., unpublished results). Unexpectedly, in these experiments the half-lives of polyQ constructs were found to be similar, independent of polyQ repeat lengths. Comparable results were obtained after expression of polyQ proteins in spinobulbar muscular atrophy cells (76). On the other hand, a polyQ-expanded ataxin-1 protein translated *in vitro* was shown to be more resistant to proteasome-dependent degradation than versions with polyQ repeats in the normal range, although both short and long polyQ proteins seemed to be ubiquitinated with similar efficiencies (37). Further experiments will be necessary to resolve the differences between these studies. Given that soluble polyQ proteins are degraded by the proteasome, it is likely that this step is preceded by ubiquitination, although this has not been demonstrated yet *in vivo*. Interestingly, in the cellular system mentioned above (76), Hsp70 and Hsp40 were observed to enhance the degradation of an expanded polyQ protein. Transfer of polyQ-expanded proteins to the degradation machinery may be mediated by Hsp70 together with the newly discovered protein CHIP (carboxyterminus of Hsp70-interacting protein), which is thought to act as an E3 ligase in the ubiquitination of nonnative proteins in cooperation with Hsp70 and Hsp40 (77, 78). Because Hsp70 is likely to recognize an unstructured polyQ monomer (58), it will be interesting to investigate whether CHIP is involved in the ubiquitination of soluble polyQ proteins.

Intracellular aggregation of polyQ proteins has been proposed to impair the ubiquitin–proteasome system (reviewed in ref. 20). Once an expanded polyQ protein has escaped degradation and β-sheet oligomers are initiated, 19S regulatory complexes, unable to unfold these oligomers, may become trapped in the growing aggregates. This trapping of 19S particles may result in a partial inhibition of proteasomal activity and eventually cellular dysfunction (Fig. 2B). In contrast, large neuronal polyQ inclusions, the final product of the aggregation process, are compartmentalized in so-called aggresomes (ref. 19 and P.B., unpublished observations) and would no longer be able to trap 19S particles efficiently. Two recent reports generally support

this hypothesis (79, 80). Using a rapidly degraded version of green fluorescent protein as a reporter, Bence *et al.* showed that expression of an expanded polyQ protein partially inhibited the ubiquitin–proteasome system in cell culture (79). Navon and Goldberg provided proof of principle that a nondegradable model substrate, unrelated to polyQ proteins, can function as a dominant inhibitor of the unfolding and degradation of otherwise proteasome-degradable proteins *in vitro* (80). Further biochemical studies will have to reveal whether intermediates in the process of polyQ aggregation indeed inhibit the unfolding and degradation activities of the proteasome. Such a mechanism could explain how protein misfolding may cause cellular toxicity and how the Hsp70/Hsp40 chaperone system may mitigate this toxicity. There are several reports that the Hsp70/Hsp40 chaperone system is not only essential for proper folding but also for the rapid degradation of certain proteins (81–83). This effect may be attributed to the activity of these chaperones in preventing the formation of intermolecular β-sheets by nonnative proteins, maintaining them in a degradation-competent state (Fig. 2B).

Perspectives

Recent years have seen major strides toward understanding the molecular basis of polyQ diseases based on a combination of biochemical studies *in vitro* and the analysis of numerous animal and cell-culture models *in vivo*. However, the exact mechanisms by which expanded polyQ proteins exert cellular toxicity remain to be established. In testing current hypotheses, it will be important to dissect the process of polyQ aggregation into distinct steps. Significant progress would result from the development of techniques that allow the accumulation and isolation of aggregation intermediates at different stages of the process. These intermediates then could be tested for their ability to interact with molecular chaperones, to recruit other polyQ-containing proteins, or to inhibit the proteasome system.

It is striking that many other neurodegenerative diseases including Parkinson's disease, Alzheimer's disease, amyotrophic lateral sclerosis, and the prion diseases all involve the assembly of structurally unrelated proteins into intracellular or extracellular amyloid fibrils. Together with the recent finding that intermediates formed early in the aggregation pathway can be inherently highly cytotoxic (38, 39), even for proteins that are not disease-associated (38), it seems likely that common fundamental mechanisms underlie the toxicity of amyloid formation. This notion is strongly supported by the observation that molecular chaperones can suppress the toxicity of amyloidogenic proteins as different as α-synuclein in Parkinson's disease (61), Tau protein in Alzheimer's disease (H. Xu, personal communication), or the polyQ proteins. Searching for ways to pharmacologically induce the expression of molecular chaperones in neurons may open up a promising approach to the treatment of these diseases (57).

1. Orr, H. T. (2001) *Genes Dev.* **15**, 925–932.
2. Zoghbi, H. Y. & Orr, H. T. (2000) *Annu. Rev. Neurosci.* **23**, 217–247.
3. Perutz, M. F. (1999) *Trends Biochem. Sci.* **24**, 58–63.
4. Tobin, A. J. & Signer, E. R. (2000) *Trends Cell Biol.* **10**, 531–536.
5. The Huntington's Disease Collaborative Research Group (1993) *Cell* **72**, 971–983.
6. Nasir, J., Floresco, S. B., O'Kusky, J. R., Diewert, V. M., Richman, J. M., Zeisler, J., Borowski, A., Marth, J. D., Phillips, A. G. & Hayden, M. R. (1995) *Cell* **81**, 811–823.
7. DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P. & Aronin, N. (1997) *Science* **277**, 1990–1993.
8. Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. L., Fischbeck, K. H. & Pittman, R. N. (1997) *Neuron* **19**, 333–344.
9. Skinner, P. J., Koshy, B. T., Cummings, C. J., Klement, I. A., Helin, K., Servadio, A., Zoghbi, H. Y. & Orr, H. T. (1997) *Nature (London)* **389**, 971–974.
10. Rochet, J. C. & Lansbury, P. T., Jr. (2000) *Curr. Opin. Struct. Biol.* **10**, 60–68.
11. Perutz, M. F., Johnson, T., Suzuki, M. & Finch, J. T. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5355–5358.
12. Perutz, M. F., Finch, J. T., Berriman, J. & Lesk, A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5591–5595.
13. Perez, M. K., Paulson, H. L., Pendse, S. J., Saionz, S. J., Bonini, N. M. & Pittman, R. N. (1998) *J. Cell Biol.* **143**, 1457–1470.
14. Kazantsev, A., Preisinger, E., Dranovsky, A., Goldgaber, D. & Housman, D. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 11404–11409.

15. Nucifora, F. C., Jr., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V. L., Dawson, T. M. & Ross, C. A. (2001) *Science* **291**, 2423–2428.
16. Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L. & Bates, G. P. (1997) *Cell* **90**, 537–548.
17. Cummings, C. J., Mancini, M. A., Antalffy, B., DeFranco, D. B., Orr, H. T. & Zoghbi, H. Y. (1998) *Nat. Genet.* **19**, 148–154.
18. Chai, Y., Koppenhafer, S. L., Shoesmith, S. J., Perez, M. K. & Paulson, H. L. (1999) *Hum. Mol. Genet.* **8**, 673–682.
19. Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H. & Wanker, E. E. (2001) *Mol. Biol. Cell* **12**, 1393–1407.
20. Sherman, M. Y. & Goldberg, A. L. (2001) *Neuron* **29**, 15–32.
21. Balbirnie, M., Grothe, R. & Eisenberg, D. S. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 2375–2380.
22. Benzinger, T. L., Gregory, D. M., Burkoth, T. S., Miller-Auer, H., Lynn, D. G., Botto, R. E. & Meredith, S. C. (2000) *Biochemistry* **39**, 3491–3499.
23. Antzutkin, O. N., Balbach, J. J., Leapman, R. D., Rizzo, N. W., Reed, J. & Tycko, R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 13045–13050.
24. Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H. & Wanker, E. E. (1997) *Cell* **90**, 549–558.
25. Kelly, J. W. (2000) *Nat. Struct. Biol.* **7**, 824–826.
26. Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L., Arnsdorf, M. F. & Lindquist, S. L. (2000) *Science* **289**, 1317–1321.
27. Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G. P., Lehrach, H. & Wanker, E. E. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 4604–4609.
28. Lunkes, A. & Mandel, J. L. (1998) *Hum. Mol. Genet.* **7**, 1355–1361.
29. Wheeler, V. C., White, J. K., Gutekunst, C. A., Vrbanac, V., Weaver, M., Li, X. J., Li, S. H., Yi, H., Vonsattel, J. P., Gusella, J. F., *et al.* (2000) *Hum. Mol. Genet.* **9**, 503–513.
30. Yamamoto, Y., Hasegawa, H., Tanaka, K. & Kakizuka, A. (2001) *Cell Death Differ.* **8**, 871–873.
31. Ikeda, H., Yamaguchi, M., Sugai, S., Aze, Y., Narumiya, S. & Kakizuka, A. (1996) *Nat. Genet.* **13**, 196–202.
32. Cooper, J. K., Schilling, G., Peters, M. F., Herring, W. J., Sharp, A. H., Kaminsky, Z., Masone, J., Khan, F. A., Delanoy, M., Borchelt, D. R., *et al.* (1998) *Hum. Mol. Genet.* **7**, 783–790.
33. Becher, M. W., Kotzuk, J. A., Sharp, A. H., Davies, S. W., Bates, G. P., Price, D. L. & Ross, C. A. (1998) *Neurobiol. Dis.* **4**, 387–397.
34. Yamamoto, A., Lucas, J. J. & Hen, R. (2000) *Cell* **101**, 57–66.
35. Klement, I. A., Skinner, P. J., Kaytor, M. D., Yi, H., Hersch, S. M., Clark, H. B., Zoghbi, H. Y. & Orr, H. T. (1998) *Cell* **95**, 41–53.
36. Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. E. (1998) *Cell* **95**, 55–66.
37. Cummings, C. J., Reinstein, E., Sun, Y., Antalffy, B., Jiang, Y., Ciechanover, A., Orr, H. T., Beaudet, A. L. & Zoghbi, H. Y. (1999) *Neuron* **24**, 879–892.
38. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M. & Stefani, M. (2002) *Nature (London)* **416**, 507–511.
39. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J. & Selkoe, D. J. (2002) *Nature (London)* **416**, 535–539.
40. Suhr, S. T., Senut, M. C., Whitelegge, J. P., Faull, K. F., Cuizon, D. B. & Gage, F. H. (2001) *J. Cell Biol.* **153**, 283–294.
41. Luthi-Carter, R., Strand, A., Peters, N. L., Solano, S. M., Hollingsworth, Z. R., Menon, A. S., Frey, A. S., Spektor, B. S., Penney, E. B., Schilling, G., *et al.* (2000) *Hum. Mol. Genet.* **9**, 1259–1271.
42. Wyttenbach, A., Swartz, J., Kita, H., Thykjaer, T., Carmichael, J., Bradley, J., Brown, R., Maxwell, M., Schapira, A., Orntoft, T. F., Kato, K. & Rubinsztein, D. C. (2001) *Hum. Mol. Genet.* **10**, 1829–1845.
43. Mantamadiotis, T., Lemberger, T., Bleckmann, S. C., Kern, H., Kretz, O., Villalba, A. M., Tronche, F., Kellendonk, C., Gau, D., Kapfhammer, J., *et al.* (2002) *Nat. Genet.* **31**, 47–54.
44. Lin, X., Antalffy, B., Kang, D., Orr, H. T. & Zoghbi, H. Y. (2000) *Nat. Neurosci.* **3**, 157–163.
45. Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., Kimura, T., Koide, R., Nozaki, K., Sano, Y., *et al.* (2000) *Nat. Genet.* **26**, 29–36.
46. Steffan, J. S., Bodai, L., Pallos, J., Poelman, M., McCampbell, A., Apostol, B. L., Kazantsev, A., Schmidt, E., Zhu, Y. Z., Greenwald, M., *et al.* (2001) *Nature (London)* **413**, 739–743.
47. Bukau, B. & Horwich, A. L. (1998) *Cell* **92**, 351–366.
48. Hartl, F. U. & Hayer-Hartl, M. (2002) *Science* **295**, 1852–1858.
49. Dobson, C. M. & Karplus, M. (1999) *Curr. Opin. Struct. Biol.* **9**, 92–101.
50. Misselwitz, B., Staeck, O. & Rapoport, T. A. (1998) *Mol. Cell* **2**, 593–603.
51. Kazemi-Esfarjani, P. & Benzer, S. (2000) *Science* **287**, 1837–1840.
52. Warrick, J. M., Chan, H. Y., Gray-Board, G. L., Chai, Y., Paulson, H. L. & Bonini, N. M. (1999) *Nat. Genet.* **23**, 425–428.
53. Chan, H. Y., Warrick, J. M., Gray-Board, G. L., Paulson, H. L. & Bonini, N. M. (2000) *Hum. Mol. Genet.* **9**, 2811–2820.
54. Fernandez-Funez, P., Nino-Rosales, M. L., de Gouyon, B., She, W. C., Luchak, J. M., Martinez, P., Turiegano, E., Benito, J., Capovilla, M., Skinner, P. J., *et al.* (2000) *Nature (London)* **408**, 101–106.
55. Cummings, C. J., Sun, Y., Opal, P., Antalffy, B., Mestril, R., Orr, H. T., Dillmann, W. H. & Zoghbi, H. Y. (2001) *Hum. Mol. Genet.* **10**, 1511–1518.
56. Bonini, N. M. (2002) *Proc. Natl. Acad. Sci. USA* **99**, Suppl. 4, 16407–16411.
57. Sittler, A., Lurz, R., Lueder, G., Priller, J., Lehrach, H., Hayer-Hartl, M. K., Hartl, F. U. & Wanker, E. E. (2001) *Hum. Mol. Genet.* **10**, 1307–1315.
58. Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K. & Hartl, F. U. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 7841–7846.
59. Rudiger, S., Germeroth, L., Schneider-Mergener, J. & Bukau, B. (1997) *EMBO J.* **16**, 1501–1507.
60. Heydari, A. R., Takahashi, R., Gutsmann, A., You, S. & Richardson, A. (1994) *Experientia* **50**, 1092–1098.
61. Auluck, P. K., Chan, H. Y., Trojanowski, J. Q., Lee, V. M. & Bonini, N. M. (2002) *Science* **295**, 865–868.
62. Krobitsch, S. & Lindquist, S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1589–1594.
63. Higashiyama, H., Hirose, F., Yamaguchi, M., Inoue, Y. H., Fujikake, N., Matsukage, A. & Kakizuka, A. (2002) *Cell Death Differ.* **9**, 264–273.
64. Glover, J. R. & Lindquist, S. (1998) *Cell* **94**, 73–82.
65. Ben-Zvi, A. P. & Goloubinoff, P. (2001) *J. Struct. Biol.* **135**, 84–93.
66. Satyal, S. H., Schmidt, E., Kitagawa, K., Sondheimer, N., Lindquist, S., Kramer, J. M. & Morimoto, R. I. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5750–5755.
67. Kimura, Y., Koitabashi, S., Kakizuka, A. & Fujita, T. (2001) *Genes Cells* **6**, 887–897.
68. Serio, T. R. & Lindquist, S. L. (2000) *Trends Cell Biol.* **10**, 98–105.
69. Latterich, M., Frohlich, K. U. & Schekman, R. (1995) *Cell* **82**, 885–893.
70. Dai, R. M. & Li, C. C. (2001) *Nat. Cell Biol.* **3**, 740–744.
71. Ye, Y., Meyer, H. H. & Rapoport, T. A. (2001) *Nature (London)* **414**, 652–656.
72. Rape, M., Hoppe, T., Gorr, I., Kalocay, M., Richly, H. & Jentsch, S. (2001) *Cell* **107**, 667–677.
73. Pickart, C. M. (2001) *Annu. Rev. Biochem.* **70**, 503–533.
74. Voges, D., Zwickl, P. & Baumeister, W. (1999) *Annu. Rev. Biochem.* **68**, 1015–1068.
75. Gonzalez, F., Delahodde, A., Kodadek, T. & Johnston, S. A. (2002) *Science* **296**, 548–550.
76. Bailey, C. K., Andriola, I. F., Kampinga, H. H. & Merry, D. E. (2002) *Hum. Mol. Genet.* **11**, 515–523.
77. Demand, J., Alberti, S., Patterson, C. & Hohfeld, J. (2001) *Curr. Biol.* **11**, 1569–1577.
78. Murata, S., Minami, Y., Minami, M., Chiba, T. & Tanaka, K. (2001) *EMBO Rep.* **2**, 1133–1138.
79. Bence, N. F., Sampat, R. M. & Kopito, R. R. (2001) *Science* **292**, 1552–1555.
80. Navon, A. & Goldberg, A. L. (2001) *Mol. Cell* **8**, 1339–1349.
81. Lee, D. H., Sherman, M. Y. & Goldberg, A. L. (1996) *Mol. Cell. Biol.* **16**, 4773–4781.
82. Bercovich, B., Stancovski, I., Mayer, A., Blumenfeld, N., Laszlo, A., Schwartz, A. L. & Ciechanover, A. (1997) *J. Biol. Chem.* **272**, 9002–9010.
83. Ohba, M. (1997) *FEBS Lett.* **409**, 307–311.

Colloquium

Studies of the aggregation of mutant proteins *in vitro* provide insights into the genetics of amyloid diseases

Fabrizio Chiti*†, Martino Calamai*, Niccolò Taddei*, Massimo Stefani*, Giampietro Ramponi*, and Christopher M. Dobson†‡

*Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, Viale Morgagni 50, 50134 Florence, Italy; and †Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Protein aggregation and the formation of highly insoluble amyloid structures is associated with a range of debilitating human conditions, which include Alzheimer's disease, Parkinson's disease, and the Creutzfeldt–Jakob disease. Muscle acylphosphatase (AcP) has already provided significant insights into mutational changes that modulate amyloid formation. In the present paper, we have used this system to investigate the effects of mutations that modify the charge state of a protein without affecting significantly the hydrophobicity or secondary structural propensities of the polypeptide chain. A highly significant inverse correlation was found to exist between the rates of aggregation of the protein variants under denaturing conditions and their overall net charge. This result indicates that aggregation is generally favored by mutations that bring the net charge of the protein closer to neutrality. In light of this finding, we have analyzed natural mutations associated with familial forms of amyloid diseases that involve alteration of the net charge of the proteins or protein fragments associated with the diseases. Sixteen mutations have been identified for which the mechanism of action that causes the pathological condition is not yet known or fully understood. Remarkably, 14 of these 16 mutations cause the net charge of the corresponding peptide or protein that converts into amyloid deposits to be reduced. This result suggests that charge has been a key parameter in molecular evolution to ensure the avoidance of protein aggregation and identifies reduction of the net charge as an important determinant in at least some forms of protein deposition diseases.

A range of debilitating human diseases is known to be associated with the formation of stable highly organized protein aggregates known as amyloid fibrils. These diseases include cerebral conditions such as Alzheimer's disease, Parkinson's disease and Creutzfeldt–Jakob disease, and also a series of systemic amyloidoses in which amyloid deposition occurs in a wider variety of organs within the body (1, 2). In each of these pathological conditions, a specific peptide or protein that is normally soluble is deposited, either intact or in fragmented form, into insoluble fibrils which accumulate in one or more types of tissue. Some amyloid diseases are rare and are associated with specific mutations involving the peptide or protein associated with amyloid deposition. Examples are familial amyloidotic polyneuropathy (3), hereditary renal amyloidosis (4, 5), and the apoA1 amyloidosis (6, 7). Other diseases, such as Alzheimer's disease, frontotemporal dementia, and Parkinson's disease, are largely sporadic, with hereditary cases involving only a limited fraction of the patients suffering from the conditions in question (8–10). The identification of specific mutations associated with familial forms of a disease that is otherwise sporadic and the investigation of the mechanism by which the mutations result in pathological behavior have proved to be of

fundamental importance for identifying specific genes associated with the disease and for exploring the molecular basis of the underlying pathology (8, 10, 11).

The amino acid sequences and native structures of the proteins associated with amyloid diseases have been found to be highly variable, but structural studies have revealed that amyloid fibrils from different sources share a common ultrastructure (12). Electron microscopy has shown that amyloid fibrils are typically straight and unbranched and are formed from an assembly of protofilaments 2–5 nm wide (12). X-ray fiber diffraction studies indicate a characteristic structure in which the polypeptide chains form β strands oriented perpendicular to the long axis of the fibril, resulting in β-sheets propagating in the direction of the fibril (12).

It is increasingly recognized that the ability to form amyloid fibrils is not a property restricted to the relatively few amino acid sequences associated with specific diseases but is a generic phenomenon of polypeptide chains (13). A considerable number of proteins, including several that adopt α-helical structures under native conditions such as myoglobin and cytochrome c_{552}, have been shown to form amyloid fibrils *in vitro*, provided appropriate conditions are selected (14–17). Fibril formation involving globular proteins occurs when the native structure is at least partially unfolded under conditions in which the ability to form noncovalent interactions is retained (13, 15). Importantly, aggregates formed from such nondisease-related proteins have been found, at least in some cases, to be highly toxic to both neuronal and nonneuronal cells (18). Although the ability to form amyloid structures appears generic, the propensities of different protein sequences to aggregate under given conditions can differ very substantially (19–25).

The ability of a wide range of natural proteins to form amyloid fibrils *in vitro* provides a large variety of systems with which to study the underlying nature of the conversion from the soluble to the aggregated state of proteins. This opportunity allows the fundamental principles of a process central to human disease to be investigated using a set of proteins that can be chosen to have particular advantages for their study. Human muscle acylphosphatase (AcP) is a model system that has proved to be particularly suitable for studies of misfolding and aggregation (15, 25–28). It is a relatively simple protein with 98 residues consisting of a five-stranded antiparallel β-sheet packed against two par-
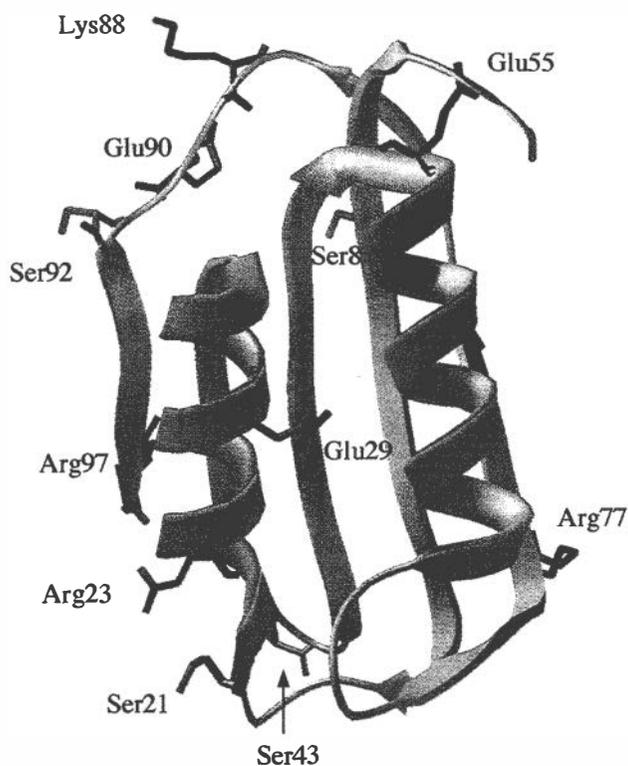

Fig. 1. Structure of AcP in its native state. Residues that have been mutated in the present study are labeled and their side chains shown. The various amino acid substitutions are listed in Table 1.

allel α-helices (Fig. 1). In addition to possessing a simple and well defined fold with no disulphide bridges or other complications such as bound cofactors, the normal folding behavior of AcP has been studied in a great detail at a residue-specific level (29, 30). This is an important issue, because folding and aggregation are potentially competitive events for a polypeptide chain in a biological environment. AcP has been shown to form readily, in the presence of moderate concentrations of denaturant such as 2,2,2, trifluoroethanol (TFE), amyloid fibrils structurally similar to those associated with disease (15). Importantly, a key step of the aggregation process, the initial formation of insoluble spherical and elongated protofibrils from soluble states, can be readily followed for AcP with a variety of biophysical techniques, and its rate is reproducible and easily measurable (25). Monitoring the first steps in the aggregation process leading to formation of prefibrillar aggregates is gaining in importance, because it is increasingly recognized that low-molecular weight oligomers that precede formation of mature amyloid fibrils, often referred to as protofibrils, represent the fundamental pathogenic species in at least many of the amyloid diseases (11, 18, 31–33).

In our initial studies of AcP, the propensity to form fibrils was investigated for a series of mutants under conditions in which the native states of the various protein variants were substantially populated, although significantly destabilized (26). The propensity to aggregate was found to correlate inversely with the conformational stability of the native state of the protein in the different mutants (26). Consistent with this finding, stabilization of the native state of AcP by ligand binding inhibits amyloid formation (27). These results show that the stability of the native state is a major factor preventing the conversion of a globular protein into amyloid fibrils under nonpathological conditions. This conclusion is in accord with studies that have shown that destabilization of the native state is a primary mechanism by which naturally occurring mutations promote their pathogenic

effect in at least some hereditary amyloid diseases (34–36). The overriding significance of the native state in preventing protein aggregation also suggests that strategies aimed at stabilizing the native states of amyloidogenic proteins could be of major value in the prevention of amyloid diseases (27, 37).

In a second study, the rates of protofibril formation of more than 30 mutants, with conservative amino acid substitutions spread throughout the sequence of the protein, were determined under conditions in which the native states of all mutants are fully destabilized (25). This approach has allowed different regions of the AcP sequence to be probed for their degree of involvement in promoting the aggregation process from an ensemble of partially denatured conformations. All mutations found to perturb significantly the rate of aggregation were found to be located in two specific regions of the protein sequence, residues 16–31 and 87–98. This result indicates that aggregation of AcP can be nucleated by specific regions of the protein sequence that are consequently directly involved in the rate-determining steps of this process (25). The measured rates of aggregation were found to correlate with changes in the hydrophobicity and in the propensity to convert from α-helical to β-sheet structure of the regions of the protein in which the mutations are located (25, 28). Interestingly, natural mutations of the prion protein that leave the conformational stability of the cellular form of the protein unaltered (38, 39) increase either the β-sheet propensity or hydrophobicity of the prion sequence.

In this paper, we have extended the protein engineering approach to investigate the role that charged residues play in the process of aggregation and amyloid formation. Electrostatic interactions have been suggested to be important in the modulation of the aggregation behavior of a number of specific disease-related proteins (14, 20, 40–42). Nevertheless, a systematic investigation of the importance in aggregation of the total charge state of a polypeptide chain or of particular electrostatic interactions involving specific residues within its sequence has not yet been reported. Furthermore, it is not yet clear whether electrostatic interactions play a key role in protein aggregation generally or whether they are particularly important for a limited number of protein systems. In addition to providing new information on the driving forces of protein aggregation, this study is also designed to give additional insights into the origin of heritable amyloid diseases, because a number of these conditions are associated with amino acid replacements that alter the charge state of the aggregating polypeptide chains.

Materials and Methods

Design, Production, and Purification of AcP Mutants. The 15 mutants were designed to perturb the α-helical and β-sheet propensities and the hydrophobicity of the protein sequence as little as possible. All replacements involving substitution of or by hydrophobic residues (Val, Ile, Leu, Ala, Gly, Tyr, Phe, Trp, Cys, Met, and Pro) were therefore discarded. To evaluate the β-sheet and α-helical propensities of the protein sequence before and after mutation, the scale of β-sheet propensities edited by Street and Mayo and the AGADIR algorithm were used, respectively (43, 44). Mutagenesis was carried out by using the QuikChange Site-Directed Mutagenesis kit (Stratagene). Protein expression and purification of wild-type and mutated AcP molecules were performed according to the procedures described previously (45). All proteins have the cysteine residue at position 21 replaced by a serine residue to avoid complexities arising from a free sulfhydryl group (29). DNA sequencing was used to ensure the presence of the desired mutation. Protein concentration was measured by UV absorption by using an ε_{280} value of 1.49 ml·mg^{-1}·cm^{-1}.

Equilibrium Unfolding Experiments. Equilibrium urea denaturation curves were obtained for each AcP variant by measuring the

intrinsic fluorescence of 25–30 equilibrated samples containing 0.02 mg·ml^{-1} protein and urea concentrations ranging from 0 to 8.1 M, in 50 mM acetate buffer, pH 5.5, 28°C. A Perkin–Elmer LS 55 with excitation and emission wavelengths of 280 and 335 nm, respectively, was used for the measurements. The data were analyzed according to the method of Santoro and Bolen (46) to yield the free energy of unfolding in the absence of denaturant ($\Delta G_{U\text{-}F}^{H2O}$), the dependence of $\Delta G_{U\text{-}F}$ on denaturant concentration (m value) and the urea concentration at which the protein is 50% denatured (C_m). To overcome the problems associated with accurate determination of m values in individual experiments, the conformational stabilities of all protein variants are expressed as $\Delta\Delta G_{U\text{-}F} = \langle m \rangle \cdot (C_m - C'_m)$, where $\langle m \rangle$ is the average m value of all mutants; C_m and C'_m are the midpoints of denaturation for the wild-type and mutated proteins, respectively (47).

Kinetics of Aggregation. The aggregation process was monitored as described previously (25). In brief, each AcP variant was incubated at a concentration of 0.4 mg·ml^{-1} in 25% (vol/vol) TFE/50 mM acetate buffer, pH 5.5, 25°C. At regular time intervals, aliquots of 60 μl of this solution were mixed with 440 μl of 25 mM phosphate buffer, pH 6.0, containing 25 μM thioflavine T (ThT). The resulting ThT fluorescence was measured by using excitation and emission wavelengths of 440 and 485 nm, respectively. Kinetic plots were fitted to single exponential functions to determine the aggregation rate constants (k_{AGG}) for the various proteins.

Results

Selection and Characteristics of the AcP Mutations. Fifteen variants of AcP, all having single replacements of charged or hydrophilic residues located on the surface of the protein, were purified for the present study (Table 1, Fig. 1). The mutations consist of substitutions of neutral residues with residues carrying a charge under the conditions of pH investigated here (S8H, S21R, S43E, and S92H), substitutions of charged residues with uncharged ones (R23Q, E29Q, E55Q, K88N, K88Q, and R97Q), and substitutions of charged residues with others of opposite sign (E29K, E29R, R77E, E90H, and R97E). All mutations involve an increase or a decrease in the charge state of the protein by 1 or 2 units (Table 1). The rate of aggregation of AcP from a denatured ensemble of conformations was found in a previous study to be sensitive to amino acid substitutions only when these occur within the two regions of the sequence comprising residues 16–31 and 87–98 (25). The majority of the mutations were therefore designed within such regions. Four mutations at positions outside these sequence regions were, however, also produced to act as controls (S8H, S43E, E55Q, and R77E).

The 15 mutations used here were chosen for their ability to leave the α-helical and β-sheet propensities of the protein sequence unchanged, because secondary structure-forming propensities have been shown to be major determinants of the aggregation behavior of AcP (see *Materials and Methods* for details) (25, 28). For the same reason, mutations in which hydrophilic residues are replaced by hydrophobic ones were not considered, because changes of hydrophobicity have been shown to modify considerably the aggregation rate of AcP (25). Through analysis of the 15 mutational variants selected here, we have therefore endeavored to relate effects on the aggregation process resulting from mutations simply to the changes in the charge state of the protein at the mutated position by minimizing mutational changes of hydrophobicity and secondary structural propensity of the protein.

The conformational stabilities of all of the 15 AcP variants were evaluated by means of equilibrium urea denaturation measurements (Fig. 2). The resulting parameters show that the mutations induce a destabilization of the native state of the

Table 1. Parameters describing the conformational stabilities and aggregation rates of AcP mutants

	Net charge of the mutant*	C_m, M†	$\Delta\Delta G_{U\text{-}F}$, kJ mol$^{-1\ddagger}$	ln k_{AGG}§
Wild-type	+5	4.0 ± 0.1	–	−6.98 ± 0.04
S8H	+6	1.8 ± 0.1	11.9 ± 0.6	−6.89 ± 0.12
S21R	+6	3.8 ± 0.1	1.4 ± 0.6	−9.04 ± 0.14
R23Q	+4	3.5 ± 0.1	3.0 ± 0.6	−6.31 ± 0.13
E29K	+7	3.3 ± 0.1	4.0 ± 0.6	−7.60 ± 0.13
E29Q	+6	4.1 ± 0.1	−0.6 ± 0.6	−7.39 ± 0.12
E29R	+7	3.3 ± 0.1	4.0 ± 0.6	−8.52 ± 0.14
S43E	+4	4.0 ± 0.1	−0.1 ± 0.6	−6.33 ± 0.14
E55Q	+6	3.3 ± 0.1	4.1 ± 0.6	−6.85 ± 0.12
R77E	+3	1.8 ± 0.1	11.7 ± 0.6	−6.16 ± 0.14
K88N	+4	3.7 ± 0.1	1.6 ± 0.6	−7.22 ± 0.12
K88Q	+4	3.9 ± 0.1	0.7 ± 0.6	−7.20 ± 0.12
E90H	+7	3.0 ± 0.1	5.5 ± 0.6	−9.50 ± 0.14
S92H	+6	3.8 ± 0.1	1.4 ± 0.6	−7.32 ± 0.12
R97E	+3	3.7 ± 0.1	1.8 ± 0.6	−6.57 ± 0.13
R97Q	+4	3.6 ± 0.1	2.2 ± 0.6	−6.98 ± 0.12

All experimental errors reported in the table are standard deviations unless stated otherwise.

*Calculated at pH 5.5 by using standard pKa values for protein side chains. Under the denaturing conditions used to study aggregation of AcP and its mutants, the pKa values of the various residues of AcP are assumed not to deviate significantly from normal ranges, because these residues are likely to be highly solvent exposed. At pH 5.5, residues of Asp, Glu, Arg, Lys, and His are therefore assumed to be charged.

†Concentration of urea at which the mutant is 50% denatured.

‡$\Delta\Delta G_{U\text{-}F}$ values were obtained by using $\Delta\Delta G_{U\text{-}F} = \langle m \rangle \cdot (C_m - C'_m)$, where $\langle m \rangle$ is the average m value of all mutants ($\langle m \rangle$ = 5.40 ± 0.15 kJ·mol^{-1}·M^{-1}); C_m and C'_m are the midpoint of denaturation for the wild-type and mutated protein, respectively. The $\Delta\Delta G_{U\text{-}F}$ values correspond to those obtained by subtracting the $\Delta G_{U\text{-}F}$ value of the mutant from that of the wild-type protein ($\Delta G_{U\text{-}F}$ = 21.7 ± 0.8 kJ·mol^{-1} for the wild type).

§The best estimate and experimental error reported for the wild-type protein are, respectively, the average value and standard error obtained from nine independent measurements.

protein ranging from 0 to 12 kJ mol^{-1} (Table 1). The two mutations S8H and R77E resulted in the largest values of the free energy of destabilization ($\Delta\Delta G_{U\text{-}F}$) of 11.9 and 11.7
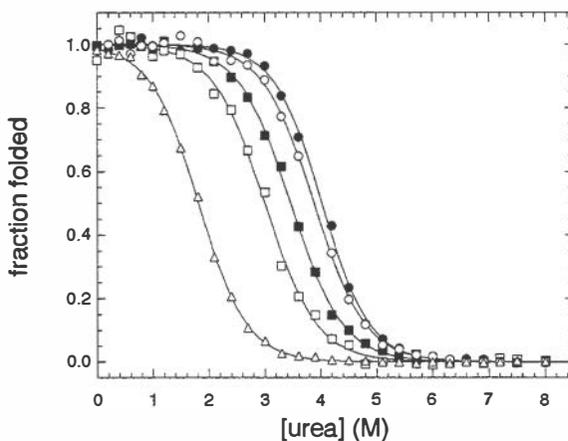

Fig. 2. Urea denaturation curves of representative AcP variants in 50 mM acetate buffer, pH 5.5, 28°C. Curves are normalized to the fraction of folded protein and correspond to those of wild-type AcP (filled circles), K88Q (open circles), R23Q (filled squares), E90H (open squares), and S8H (open triangles) mutants. The solid lines through the data represent the best fits of the data points to the equation given by Santoro and Bolen (46). The resulting thermodynamic parameters for all protein variants are listed in Table 1.

kJ·mol^{-1}, respectively. Apart from these mutations, however, none of the remaining 13 substitutions resulted in a loss of stability of >6 kJ mol^{-1}. A previous mutational study of AcP, in which residues in the hydrophobic core of the protein were replaced by other hydrophobic residues with a smaller size, indicated far larger $\Delta\Delta G_{\text{U-F}}$ values than those observed here (30). This difference is illustrated by the average values of $\Delta\Delta G_{\text{U-F}}$ determined for the substitutions at hydrophobic core positions and surface hydrophilic residues, which are 11.5 and 3.5 kJ·mol^{-1}, respectively. Charged or hydrophilic residues do not, therefore, play as significant a role in AcP conformational stability as residues in the hydrophobic core of the protein.

Effect of the Mutations on the Aggregation Process of AcP. Some of the mutations analyzed here cause a significant, albeit small, destabilization of the native state as described in the previous section. Amino acid replacements that destabilize the native state of a protein are known to favor the process of aggregation by populating unfolded or partially folded states that are more prone to aggregation than the fully native state; such effects have been observed for AcP as well as for other protein systems (26, 34–36, 48). To overcome this problem, we probed the rate of aggregation of the polypeptide chain under conditions where the native fold of the protein is substantially disrupted, i.e., in aqueous solutions containing 25% (vol/vol) TFE. These conditions were found to denature even the most stable AcP mutants within a few seconds but still to allow aggregation to occur (15, 25). This procedure permits any change in aggregation rate to be attributed entirely to the intrinsic effect of the amino acid substitutions on the aggregation process, without the complications of additional contributions arising from the destabilization of the native state.

Fig. 3a describes the increase of ThT fluorescence resulting from aggregation of wild-type AcP and some representative mutants under these conditions. Such increases in ThT fluorescence reflect the formation of small aggregates, revealed by electron microscopy to be spherical or elongated 4-nm-wide protofibrils (15, 25). The rate constants obtained by fitting the data points to single exponential functions are reported in Table 1 for all of the protein variants studied here. The majority of the mutations involving residues within the two regions of the sequence 16–31 and 87–98 change the rate of aggregation to a significant extent (Table 1). In addition to having a relatively high hydrophobicity and a considerable propensity to form β structure, these two regions are also characterized by a net charge of zero. The 16–31 region contains six charged residues, three with a positive charge and the other three with a negative charge. The 87–98 region is also neutral, because it contains two positively charged residues, one negatively charged residue, and the negatively charged C terminus of the protein. If the neutralities of these two regions were directly responsible for their critical role in the process of aggregation, all substitutions of charged residues located within these two regions would be expected to disfavor the process of aggregation. This does not appear to be the case, however, as two mutations (R23Q and R97E), both of which generate a local negative charge in these regions, accelerate the process of aggregation significantly (Table 1).

A trend of a different type, however, can be observed when the aggregation rates reported in Table 1 are examined from a different perspective. Mutations decreasing the aggregation rate invariably involve the addition to the protein of an extra positive charge and/or deletion of a negative one. Similarly, mutations that increase the aggregation rate involve deletion of positively charged residues and/or insertion of negatively charged groups. This behavior can be accounted for by considering the overall charge of AcP. Under these conditions of pH, the wild-type protein has a net charge of +5 (AcP contains 9 lysines, 6
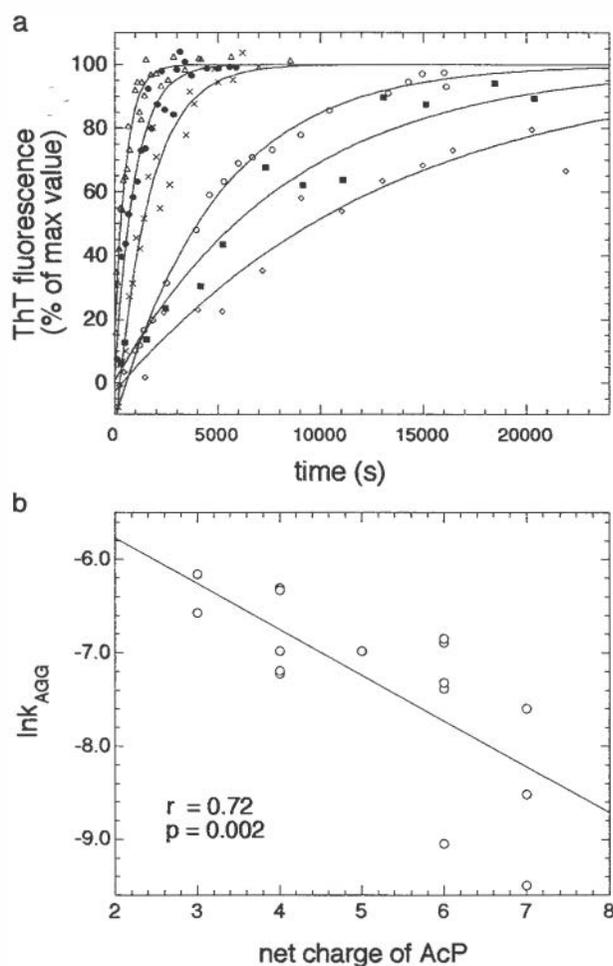

Fig. 3. (a) Aggregation of six representative AcP variants followed by ThT fluorescence. Aggregation was initiated in each case in 25% TFE/50 mM acetate buffer, pH 5.5, 25°C. Aliquots were withdrawn at regular time intervals for the ThT assay. The AcP variants shown are: wild-type (filled circles), R23Q (open triangles), E29Q (crosses), E29R (open circles), S21R (filled squares), and E90H (diamonds). The solid lines through the data points represent the best fits to single exponential functions. The resulting rate constant values are reported for all variants in Table 1. (b) Aggregation rate versus net charge constructed with the data points of the wild-type protein and the 15 mutants. Changes of net charge on mutation are calculated at pH 5.5 assuming standard pKa values for amino acid residues.

arginines, 6 glutamates, 4 aspartates, and no histidines). Mutations that disfavor the process of aggregation therefore increase the overall net charge, whereas those that favor aggregation reduce the overall net charge. This finding can be rationalized on a simple electrostatic argument, that self association will tend to be disfavored if the electrostatic repulsion between distinct molecules is increased, provided all other factors remain constant. For example, the mutation of the glutamate residue at position 29 to glutamine increases the net charge of the protein by 1 unit, from 5 to 6. This change results in an increase of the overall aggregation rate (Table 1). When Glu-29 is mutated to positively charged residues such as arginine or lysine, a more dramatic increase of the aggregation rate is observed (Table 1). The greater effect of the mutation to arginine relative to that to lysine can be attributed to the higher hydrophilicity of arginine as compared with lysine.

Unlike the hydrophobic interactions and secondary structural preferences, the effects of the net charge of the protein are not confined to local regions, again consistent with the simple

Chiti *et al.*

Table 2. Mutations associated with hereditary forms of amyloid disease involving changes in the net charge of the peptide or protein associated with each disease

Mutation*	Amyloid disease	Protein/peptide involved	Effect of mutation on net charge†	Reference
E693Q	Hereditary cerebral hemorrhage with amyloidosis	Aβ	From −3 to −2‡	57
E693K	Hereditary cerebral hemorrhage with amyloidosis	Aβ	From −3 to −1‡	58
E693G	Early-onset Alzheimer's disease	Aβ	From −3 to −2‡	24, 59
D694N	Dementia with cerebral amyloid angiopathy	Aβ	From −3 to −2‡	23, 60
G26R	apoA1 amyloidosis	apoA1	From −9 to −8§	61, 62
W50R	apoA1 amyloidosis	apoA1	From −9 to −8§	63
L60R	apoA1 amyloidosis	apoA1	From −9 to −8§	64
Δ70–72	apoA1 amyloidosis	apoA1	From −9 to −8§	40
Δ60–71¶	apoA1 amyloidosis	apoA1	From −9 to −8§	7
K257T	Frontotemporal dementia with Parkinsonism	τ	From +11 to +10‖	65
ΔK280	Frontotemporal dementia with Parkinsonism	τ	From +11 to +10‖	66
K369I	Frontotemporal dementia with Parkinsonism	τ	From +11 to +10§	67
G389R	Frontotemporal dementia with Parkinsonism	τ	From +11 to +12‖	68
R406W	Frontotemporal dementia with Parkinsonism	τ	From +11 to +10‖	69
E526V	Hereditary renal amyloidosis	Fibrinogen α chain	From −3 to −2**	4
R554L	Hereditary renal amyloidosis	Fibrinogen α chain	From −3 to −4**	5

This list does not include mutations for which a causative link with pathogenesis has been proposed (see text). The 17 mutations listed include only cases for which this link is not yet established or is still under debate. A recently compiled database (70) has been utilized to identify some of the mutations listed in this table.

*The numbering refers to the sequence of the intact proteins (τ) or of the precursors (Aβ, apolipoprotein A1, fibrinogen α chain). The numbering of Aβ and τ refers to the longest isoform in each case.

†Calculated at physiological pH 7.4, when only Lys, Arg, Glu, and Asp are assumed to be charged (71).

‡The net charge of −3 refers to the 40- or 42-residue form of the Aβ peptide.

§The net charge of −9 refers to the 93-residue form of apoA1, which is found in amyloid deposits. The charge is −6 if the alternative form of 82 residue is considered.

¶Consisting of the deletion of residues 60–71 and insertion of Val-Thr at the same position.

‖The net charge of +11 refers to the four-repeat domain that forms the core of the amyloid fibrils by τ (paired helical filaments) (72). The net charge becomes +2 if the whole sequence of the longest τ isoform is considered.

**The net charge of −3 refers to the fragment of the fibrinogen α chain (residues 500–580) that has been extracted from ex vivo fibrils (5).

electrostatic argument. Thus, two of the four substitutions that involve residues outside the two regions 16–31 and 87–98 result in aggregation rates significantly different from that of wild-type AcP (Table 1). The most marked rate change is for the R77E mutation, a substitution that results in a decrease of the net charge by 2 units. When all 15 variants are considered, a highly significant negative correlation exists between the net charge and the aggregation rate, with a linear correlation coefficient (r) and P values of 0.72 and 0.002, respectively (Fig. 3b).

Despite showing a highly significant correlation, the plot shown in Fig. 3b also indicates that mutations that cause the same change in net charge can exhibit considerably different aggregation rates. This variation may reflect a number of factors that influence this analysis such as the fact that the hydrophobicities and secondary structure preferences of the protein are not completely unchanged by any of the mutations. Such effects are expected to be particularly pronounced when the mutations are located in the two regions previously identified as key nucleation sites for aggregation. In accord with this expectation, the S21R and E90H mutations, both of which are located well inside the two key regions, decelerate the aggregation process more than do analogous replacements outside or at the edges of these regions. In addition, specific electrostatic effects may contribute to the aggregation rates, introducing further complications into the analysis. It is notable, however, that the presence of charged residues as such does not by itself act to inhibit aggregation. If this were the case, replacements of negatively charged residues by positive ones (E29K, E29R, E90H) and of positively charged residues by negative ones (R97E, R77E) would not be expected to produce significant effects on the aggregation kinetics. By contrast, these mutations result in substantial decelerations and accelerations of aggregation, respectively (Table 1).

Discussion

Net Charge, Hydrophobicity, and Secondary Structure Preferences in Protein Aggregation. The mutational study described here indicates that the total charge of the aggregation-prone state of a protein strongly influences its propensity to aggregate. The relevance of charged residues as "structural gatekeepers" against aggregation does not, however, appear to be based only on the ability of these residues to interrupt contiguous stretches of hydrophobic residues, as suggested (21), but also on their ability to generate electrostatic repulsions between protein molecules. This idea is also supported by separate observations that shielding of positively charged groups at low pH accelerates aggregation (14) and that aggregation of different proteins induced at neutral pH by preformed fibrils of a positively charged peptide correlates inversely with the isoionic point of the various proteins tested (41). Consistent with this view, proteins that are unfolded under physiological conditions ("natively unfolded proteins") generally have a total net charge that is significantly higher than proteins that fold into globular structures (49). This mechanism is likely to be a strategy through which proteins that do not fold into globular structures avoid aggregation and remain soluble in the crowded environment of the cell.

Unlike the amino acid replacements that alter the hydrophobicity, α-helical or β-sheet propensities of the AcP sequence, mutations that modify the charge of the protein are able to alter the aggregation rate even when these modifications occur outside the regions of sequence 16–31 and 87–98 that appear to be primarily responsible for regulating the aggregation process. Different regions may play a critical role at different stages of the aggregation process and therefore produce a change in the aggregation rate when mutated. However, the change of aggregation rate resulting from such mutations correlates well with the

overall change in the charge of the protein, rather than with that of specific regions of the sequence. The lack of sequence specificity of the charge effects suggests that the formation of specific electrostatic interactions is not an important determinant of the rate of aggregation of AcP, at least under the conditions examined in the present study. These data do not, of course, rule out the possibility that formation of salt bridges stabilizes the resulting aggregates; rather, they suggest that their formation plays a minor role in determining the kinetics of aggregation. The common ability of globular proteins to form amyloid fibrils at acidic pH values is not inconsistent with these general findings, because the primary effect of lowering the pH is the destabilization of the protein native fold, a process necessary for initiating aggregation (14, 50, 51). Indeed, aggregation of AcP is substantially slower at low pH than at neutral pH values if TFE is added to both solutions such that the protein remains partially denatured through this pH range (data not shown).

As mentioned above, the aggregation rate is highly sensitive to conservative mutations that alter the hydrophobicity or β-sheet propensity only when these occur within two relatively narrow regions of the AcP sequence, both of which have intrinsically high hydrophobicities and high propensities to form β-sheet structure (25). Because hydrogen bonding within β-structure and hydrophobic interactions between side-chains are likely to be the major stabilizing interactions within aggregates, increases in the propensities for such interactions are likely to enhance the rate at which aggregation occurs. As with the process of protein folding, however, formation of such interactions appear to be rather specific; indeed, both processes appear to involve well defined groups of residues that have the capability of nucleating the formation of either the intramolecular or intermolecular interactions (25). Interestingly, the residues that nucleate folding and aggregation are located in different regions of the polypeptide sequence, indicating that the two processes are highly distinct in natural proteins (25).

Origins of Familial Amyloid Diseases. Nearly 20 human diseases so far have been associated with amyloid deposition in either the brain or other organs in the body (1, 2). Hereditary forms have been described for many of these diseases, with single-point mutations often being the genetic changes that are responsible for their onset. The mechanisms through which the pathogenic effects of such naturally occurring mutations are mediated have been well established in a number of cases. Most of the mutations associated with early-onset Alzheimer's disease and related pathologies, for example, have been demonstrated to alter the efficiency or specificity of β- and γ-secretases, the two proteases that generate the amyloid β (Aβ) peptide from the precursor protein, APP (52–54). Such alterations are responsible for either an overproduction of the Aβ peptide or an increased release of the more amyloidogenic form of the peptide consisting of 42 rather than 40 residues. Alteration of proteolytic processing is also thought to be crucial in the pathogenesis of apolipoprotein A1 amyloidosis induced by the L174S mutation of this protein (6).

In other cases, destabilization of the native state of the globular protein responsible for amyloid deposition has been identified as the primary mechanism by which amino acid substitutions give rise to disease; this mechanism is particularly important in a number of systemic noncerebral amyloidoses such as familial amyloid polyneuropathy, lysozyme amyloidosis, and light chain amyloidosis (34–36). The two mutations of gelsolin associated with familial Finnish type amyloidosis (D187N and D187Y) have also been found to destabilize the native state of the protein, facilitating proteolytic attack that generates the highly amyloidogenic fragment (55, 56). An increased rate of oligomerization and an increased ability to permeate the cyto-

plasmic membrane have been suggested to be possible mechanisms by which the two natural mutations of α-synuclein, the A30P and A53T substitutions, cause familial Parkinson's disease (9, 11).

Despite the progress made recently in understanding the origins of a number of familial amyloid diseases, the mechanism of action through which the pathogenic effects of many amino acid replacements are mediated remains elusive. The conclusion from the present study raises the possibility that some disease-associated mutations act primarily through a reduction of the net charge of the corresponding peptides or proteins. To explore this possibility, we have examined pathogenic mutations involving substitution, or introduction, of charged residues within the proteins or peptides associated with amyloid diseases, and for which the causative link with the pathogenic effects is not yet established. The mutations of this type that we have been able to identify are listed in Table 2. Remarkably, 14 of 16 mutations listed in the table do indeed reduce the net charge of the polypeptides involved.

Unlike other mutations of APP, and despite the common assumption that all substitutions within the APP protein alter the processing of the protein, none of the four mutations of the Aβ peptide listed in the table have been found to act in this way (23, 24, 53). By contrast, all four mutations have been shown to increase the intrinsic propensity of Aβ to form fibrils or protofibrils *in vitro* (20, 22–24). That these mutations involve a reduction of the charge of Aβ suggests this perturbation could be the primary molecular basis for their pathogenic effect. In this regard, it is interesting to observe that the charge-preserving E693D substitution of APP is not pathogenic (20). Similarly, the six mutations of the apolipoprotein A1 listed in the table are not thought to stimulate the release of the N-terminal 82 residue fragment of the protein that is found in the fibrils, because they are too distant from the site of proteolysis (6). All six mutations, however, have the effect of reducing the negative charge within the amyloidogenic fragment (Table 2). It has, in fact, previously been suggested that the pathogenic action of these mutations could be related to alterations of the charge state (7, 40). This conclusion is well supported by the overall analysis of charge mutations described in the present work.

As far as the mutations associated with τ pathologies are concerned, a conflicting picture has emerged as to whether their pathogenicity results primarily from an impairment of the interactions of τ with microtubules or an accelerated rate of aggregation (19, 73, 74). An exception to this situation is the N279K variant of the protein, as this has been shown not to perturb the interaction with microtubules (75). The nucleotide substitution that leads to this particular mutation, however, has been shown to be pathogenic because of an alteration of the RNA splicing, which results in the production of a more aggregating variant (76, 77). Among the remaining five mutations that involve charge alterations within the τ protein, four of these reduce the net positive charge of the protein (K257T, ΔK280, K369I, and R406W). A mechanism unrelated to the reduction of charge must account for the pathogenicity of the mutation that increases the charge (G389R). Finally, the two amino acid substitutions of the fibrinogen α-chain associated with hereditary renal amyloidosis cause opposite effects on the net charge (Table 2); the E526V mutation reduces the charge of the protein, but the R554L increases it. In both cases, however, highly hydrophobic residues replace charged residues. The hydrophobic effect of the R554L substitution may outweigh its effect on the charge of the protein.

Overall, therefore, on the assumption that the pKa values are comparable to those of model compounds, 14 of the 16 mutations listed in Table 2 result in a reduction of the net charge of the corresponding peptide or protein under physiological conditions, whereas only two mutations increase it. This analysis

does not consider mutations for which an alternative causative link with pathogenesis, such as reduction of protein stability, alterations of splicing, or proteolytic processes, has been proposed; the 16 mutations discussed here include only cases for which this link is not established or is still under debate. For example, disease-related mutations involving globular proteins, such as the prion protein or transthyretin, are not considered, because in at least some cases they are likely to be pathogenic as a consequence of their destabilizing effect on the native fold of the precursor protein (34–36). A preliminary analysis, however, indicates that 12 of 13 and 19 of 25 mutations involving alterations of charge do, in fact, cause a reduction in the total charge for the prion protein and transthyretin, respectively (for the prion protein, the 125–231 domain is considered). The present analysis suggests that the mechanism of action through which the deposition of amyloid fibrils and the onset of the disease are mediated could be rather straightforward for the familial diseases associated with these particular mutations. Indeed, these mutations may stimulate pathological effects simply by increasing the intrinsic propensity of the aggregating polypeptide chains to self-assemble through a reduction in the electrostatic repulsion between the molecules.

Conclusion

Investigating the fundamentals of protein aggregation by using a model system not linked to amyloid disease and then using any principles that emerge to interpret the behavior of peptides and proteins that form amyloid deposits *in vivo* is a powerful approach to gaining an understanding of the general principles involved in the development of pathological conditions of this type. Such an approach may reveal common features in the process of protein aggregation in a particularly straightforward manner. The studies carried out on AcP show that mutations can favor aggregation either by destabilizing the native state, hence allowing unfolded or partially unfolded species prone to aggregation to be significantly populated, or by directly favoring the process of self-assembly through an increase in hydrophobicity,

β-sheet propensity, or a decrease of the net charge of the protein sequence. Examples of all these mechanisms can be found in the pathogenesis of the different familial amyloid diseases, indicating the value of studying such model systems.

Biological cells and extracellular spaces are highly crowded environments that enable the huge variety of interactions between molecules that are essential for life to take place efficiently and productively. Nevertheless, problems of self-association, or inappropriate interactions between different molecular species, will also be favored within such highly concentrated environments. The evolutionary selection of molecules able to interact specifically with their targets, but not with other species, is likely to have relied heavily on the values and distributions of charges on the surface of the molecules involved. For example, a very high net charge on a protein is likely to inhibit its interactions with other molecules of the same charge but might cause it to be unstable in a compact folded state as a result of large intramolecular repulsions (49, 78, 79). In addition, such high charges would presumably generate nonspecific association between proteins with opposite charges. It is therefore evident that protein sequences have evolved to optimize a balance between the attractions and repulsions between molecules within both the cellular and extracellular environments. The present findings on the critical role that charge plays in determining the onset of some forms of amyloid diseases, and presumably in modulating the aggregation process of proteins more generally, are a testament not only to the efficiency of the evolutionary selection of protein sequences but also to the fragility of biological organisms when such sequences are perturbed by mechanisms such as genetic mutations or misprocessing.

1. Kelly, J. W. (1996) *Curr. Opin. Struct. Biol.* **6,** 11–17.
2. Serpell, L. C., Sunde, M. & Blake, C. C. (1997) *Cell. Mol. Life Sci.* **53,** 871–887.
3. Plante-Bordeneuve, V. & Said, G. (2000) *Curr. Opin. Neurol.* **13,** 569–573.
4. Uemichi, T., Liepnieks, J. J., Alexander, F. & Benson, M. D. (1996) *Q. J. Med.* **89,** 745–750.
5. Benson, M. D., Liepnieks, J., Uemichi, T., Wheeler, G. & Correa, R. (1993) *Nat. Genet.* **3,** 252–255.
6. Obici, L., Bellotti, V., Mangione, P., Stoppini, M., Arbustini, E., Verga, L., Zorzoli, I., Anesi, E., Zanotti, G., Campana, C., *et al.* (1999) *Am. J. Pathol.* **155,** 695–702.
7. Booth, D. R., Tan, S. Y., Booth, S. E., Tennent, G. A., Hutchinson, W. L., Hsuan, J. J., Totty, N. F., Truong, O., Soutar, A. K., Hawkins, P. N., *et al.* (1996) *J. Clin. Invest.* **97,** 2714–2721.
8. Selkoe, D. J. (2001) *Physiol. Rev.* **81,** 741–766.
9. Conway, K. A., Lee, S. J., Rochet, J. C., Ding, T. T., Williamson, R. E. & Lansbury, P. T., Jr. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 571–576.
10. Goedert, M., Ghetti, B. & Spillantini, M. G. (2000) *Ann. N.Y. Acad. Sci.* **920,** 74–83.
11. Volles, M. J. & Lansbury, P. T., Jr. (2002) *Biochemistry* **41,** 4595–4602.
12. Sunde, M. & Blake, C. (1997) *Adv. Protein Chem.* **50,** 123–159.
13. Dobson, C. M. (1999) *Trends Biochem. Sci.* **24,** 329–332.
14. Zurdo, J., Guijarro, J. I., Jimenez, J. L., Saibil, H. R. & Dobson, C. M. (2001) *J. Mol. Biol.* **311,** 325–340.
15. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 3590–3594.
16. Fandrich, M., Fletcher, M. A. & Dobson, C. M. (2001) *Nature (London)* **410,** 165–166.
17. Pertinhez, T. A., Bouchard, M., Tomlinson, E. J., Wain, R., Ferguson, S. J., Dobson, C. M. & Smith, L. J. (2001) *FEBS Lett.* **495,** 184–186.
18. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M. & Stefani, M. (2002) *Nature (London)* **416,** 507–511.
19. Barghorn, S., Zheng-Fischhofer, Q., Ackmann, M., Biernat, J., von Bergen, M., Mandelkow, E. M. & Mandelkow, E. (2000) *Biochemistry* **39,** 11714–11721.
20. Melchor, J. P., McVoy, L. & Van Nostrand, W. E. (2000) *J. Neurochem.* **74,** 2209–2212.
21. Otzen, D. E., Kristensen, O. & Oliveberg, M. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9907–9912.
22. Sian, A. K., Frears, E. R., El-Agnaf, O. M., Patel, B. P., Manca, M. F., Siligardi, G., Hussain, R. & Austen, B. M. (2000) *Biochem. J.* **349,** 299–308.
23. Van Nostrand, W. E., Melchor, J. P., Cho, H. S., Greenberg, S. M. & Rebeck, G. W. (2001) *J. Biol. Chem.* **276,** 32860–32866.
24. Nilsberth, C., Westlind-Danielsson, A., Eckman, C. B., Condron, M. M., Axelman, K., Forsell, C., Stenh, C., Luthman, J., Teplow, D. B., Younkin, S. G., *et al.* (2001) *Nat. Neurosci.* **4,** 887–893.
25. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. (2002) *Nat. Struct. Biol.* **9,** 137–143.
26. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C. M. (2000) *EMBO J.* **19,** 1441–1449.
27. Chiti, F., Taddei, N., Stefani, M., Dobson, C. M. & Ramponi, G. (2001) *Protein Sci.* **10,** 879–886.
28. Taddei, N., Capanni, C., Chiti, F., Stefani, M., Dobson, C. M. & Ramponi, G. (2001) *J. Biol. Chem.* **276,** 37149–37154.
29. van Nuland, N. A., Chiti, F., Taddei, N., Raugei, G., Ramponi, G. & Dobson, C. M. (1998) *J. Mol. Biol.* **283,** 883–891.
30. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999) *Nat. Struct. Biol.* **6,** 1005–1010.
31. Lin, H., Bhatia, R. & Lal, R. (2001) *FASEB J.* **15,** 2433–2444.
32. Sousa, M. M., Cardoso, I., Fernandes, R., Guimaraes, A. & Saraiva, M. (2001) *Am. J. Pathol.* **159,** 1993–2000.
33. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J. & Selkoe, D. J. (2002) *Nature (London)* **416,** 535–539.
34. Hurle, M. R., Helms, L. R., Li, L., Chan, W. & Wetzel, R. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5446–5450.
35. McCutchen, S. L., Lai, Z. H., Miroy, G. J., Kelly, J. W. & Colon, W. (1995) *Biochemistry* **34,** 13527–13536.

36. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C., *et al.* (1997) *Nature (London)* **385,** 787–793.

37. Baures, P. W., Oza, V. B., Peterson, S. A. & Kelly, J. W. (1999) *Bioorg. Med. Chem.* **7,** 1339–1347.

38. Swietnicki, W., Petersen, R. B., Gambetti, P. & Surewicz, W. K. (1998) *J. Biol. Chem.* **273,** 31048–31052.

39. Liemann, S. & Glockshuber, R. (1999) *Biochemistry* **38,** 3258–3267.

40. Persey, M. R., Booth, D. R., Booth, S. E., van Zyl-Smit, R., Adams, B. K., Fattaar, A. B., Tennent, G. A., Hawkins, P. N. & Pepys, M. B. (1998) *Kidney Int.* **53,** 276–281.

41. Konno, T. (2001) *Biochemistry* **40,** 2148–2154.

42. Nielsen, L., Khurana, R., Coats, A., Frokjaer, S., Brange, J., Vyas, S., Uversky, V. N. & Fink, A. L. (2001) *Biochemistry* **40,** 6036–6046.

43. Street, A. G. & Mayo, S. L. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9074–9076.

44. Lacroix, E., Viguera, A. R. & Serrano, L. (1998) *J. Mol. Biol.* **284,** 173–191.

45. Taddei, N., Stefani, M., Magherini, F., Chiti, F., Modesti, A., Raugei, G. & Ramponi, G. (1996) *Biochemistry* **35,** 7077–7083.

46. Santoro, M. M. & Bolen, D. W. (1988) *Biochemistry* **27,** 8063–8068.

47. Matouschek, A. & Fersht, A. R. (1991) *Methods Enzymol.* **202,** 82–112.

48. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 8979–8984.

49. Uversky, V. N., Gillespie, J. R. & Fink, A. L. (2000) *Proteins* **41,** 415–427.

50. McParland, V. J., Kad, N. M., Kalverda, A. P., Brown, A., Kirwin-Jones, P., Hunter, M. G., Sunde, M. & Radford, S. E. (2000) *Biochemistry* **39,** 8735–8746.

51. Lai, Z., Colon, W. & Kelly, J. W. (1996) *Biochemistry* **35,** 6470–6482.

52. Scheuner, D., Eckman, C., Jensen, M., Song, X., Citron, M., Suzuki, N., Bird, T. D., Hardy, J., Hutton, M., Kukull, W., *et al.* (1996) *Nat. Med.* **2,** 864–870.

53. Farzan, M., Schnitzler, C. E., Vasilieva, N., Leung, D. & Choe, H. (1999) *Proc. Natl. Acad. Sci. USA* **97,** 9712–9717.

54. De Jonghe, C., Esselens, C., Kumar-Singh, S., Craessaerts, K., Serneels, S., Checler, F., Annaert, W., Van Broeckhoven, C. & De Strooper, B. (2001) *Hum. Mol. Genet.* **10,** 1665–1671.

55. Isaacson, R. L., Weeds, A. G. & Fersht, A. R. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 11247–11252.

56. Kazmirski, S. L., Isaacson, R. L., An, C., Buckle, A., Johnson, C. M., Daggett, V. & Fersht, A. R. (2002) *Nat. Struct. Biol.* **9,** 112–116.

57. Levy, E., Carman, M. D., Fernandez-Madrid, I. J., Power, M. D., Lieberburg, I., van Duinen, S. G., Bots, G. T., Luyendijk, W. & Frangione, B. (1990) *Science* **248,** 1124–1126.

58. Miravalle, L., Tokuda, T., Chiarle, R., Giaccone, G., Bugiani, O., Tagliavini, F., Frangione, B. & Ghiso, J. (2000) *J. Biol. Chem.* **275,** 27110–27116.

59. Kamino, K., Orr, H. T., Payami, H., Wijsman, E. M., Alonso, M. E., Pulst, S. M., Anderson, L., O'dahl, S., Nemens, E., White, J. A., *et al.* (1992) *Am. J. Hum. Genet.* **51,** 998–1014.

60. Grabowski, T. J., Cho, H. S., Vonsattel, J. P., Rebeck, G. W. & Greenberg, S. M. (2001) *Ann. Neurol.* **49,** 697–705.

61. Nichols, W. C., Gregg, R. E., Brewer, H. B., Jr., & Benson, M. D. (1990) *Genomics* **8,** 318–323.

62. Vigushin, D. M., Gough, J., Allan, D., Alguacil, A., Penner, B., Pettigrew, N. M., Quinonez, G., Bernstein, K., Booth, S. E., Booth, D. R., *et al.* (1994) *Q. J. Med.* **87,** 149–154.

63. Booth, D. R., Tan, S. Y., Booth, S. E., Hsuan, J. J., Totty, N. F., Nguyen, O., Hutton, T., Vigushin, D. M., Tennent, G. A., Hutchinson, W. L., *et al.* (1995) *Q. J. Med.* **88,** 695–702.

64. Soutar, A. K., Hawkins, P. N., Vigushin, D. M., Tennent, G. A., Booth, S. E., Hutton, T., Nguyen, O., Totty, N. F., Feest, T. G., Hsuan, J. J., *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7389–7393.

65. Rizzini, C., Goedert, M., Hodges, J. R., Smith, M. J., Jakes, R., Hills, R., Xuereb, J. H., Crowther, R. A. & Spillantini, M. G. (2000) *J. Neuropathol. Exp. Neurol.* **59,** 990–1001.

66. Rizzu, P., Van Swieten, J. C., Joosse, M., Hasegawa, M., Stevens, M., Tibben, A., Niermeijer, M. F., Hillebrand, M., Ravid, R., Oostra, B. A., *et al.* (1999) *Am. J. Hum. Genet.* **64,** 414–421.

67. Neumann, M., Schulz-Schaeffer, W., Crowther, R. A., Smith, M. J., Spillantini, M. G., Goedert, M. & Kretzschmar, H. A. (2001) *Ann. Neurol.* **50,** 503–513.

68. Murrell, J. R., Spillantini, M. G., Zolo, P., Guazzelli, M., Smith, M. J., Hasegawa, M., Redi, F., Crowther, R. A., Pietrini, P., Ghetti, B., *et al.* (1999) *J. Neuropathol. Exp. Neurol.* **58,** 1207–1226.

69. Hutton, M., Lendon, C. L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., *et al.* (1998) *Nature (London)* **393,** 702–705.

70. Siepen, J. A. & Westhead, D. R. (2002) *Protein Sci.* **11,** 1862–1866.

71. Creighton, T. E. (1993) in *Proteins* (Freeman, New York), 2nd Ed., pp. 6–7.

72. Wischik, C. M., Novak, M., Thogersen, H. C., Edwards, P. C., Runswick, M. J., Jakes, R., Walker, J. E., Milstein, C., Roth, M. & Klug, A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 4506–4510.

73. Goedert, M., Jakes, R. & Crowther, R. A. (1999) *FEBS Lett.* **450,** 306–311.

74. Friedhoff, P., von Bergen, M., Mandelkow, E. M. & Mandelkow, E. (2000) *Biochim. Biophys. Acta* **1502,** 122–132.

75. Hong, M., Zhukareva, V., Vogelsberg-Ragaglia, V., Wszolek, Z., Reed, L., Miller, B. I., Geschwind, D. H., Bird, T. D., McKeel, D., Goate, A., *et al.* (1998) *Science* **282,** 1914–1917.

76. Delisle, M. B., Murrell, J. R., Richardson, R., Trofatter, J. A., Rascol, O., Soulages, X., Mohr, M., Calvas, P. & Ghetti, B. (1999) *Acta Neuropathol.* **98,** 62–77.

77. Hasegawa, M., Smith, M. J., Iijima, M., Tabira, T. & Goedert, M. (1999) *FEBS Lett.* **443,** 93–96.

78. Goto, Y., Calciano, L. J. & Fink, A. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 573–577.

79. Luisi, D. L. & Raleigh, D. P. (2000) *J. Mol. Biol.* **299,** 1091–1100.

Chiti *et al.*

Colloquium

Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity

Per Hammarström*, Xin Jiang, Amy R. Hurshman, Evan T. Powers, and Jeffery W. Kelly†

Department of Chemistry and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road BCC265, La Jolla, CA 92037

Several misfolding diseases commence when a secreted folded protein encounters a partially denaturing microenvironment, enabling its self assembly into amyloid. Although amyloidosis is modulated by numerous environmental and genetic factors, single point mutations within the amyloidogenic protein can dramatically influence disease phenotype. Mutations that destabilize the native state predispose an individual to disease; however, thermodynamic stability alone does not reliably predict disease severity. Here we show that the rate of transthyretin (TTR) tetramer dissociation required for amyloid formation is strongly influenced by mutation (V30M, L55P, T119M, V122I), with rapid rates exacerbating and slow rates reducing amyloidogenicity. Although these rates are difficult to predict a *priori*, they notably influence disease penetrance and age of onset. L55P TTR exhibits severe pathology because the tetramer both dissociates quickly and is highly destabilized. Even though V30M and L55P TTR are similarly destabilized, the V30M disease phenotype is milder because V30M dissociates more slowly, even slower than wild type (WT). Although WT and V122I TTR have nearly equivalent tetramer stabilities, V122I cardiomyopathy, unlike WT cardiomyopathy, has nearly complete penetrance—presumably because of its 2-fold increase in dissociation rate. We show that the T119M homotetramer exhibits kinetic stabilization and therefore dissociates exceedingly slowly, likely explaining how it functions to protect V30M/T119M compound heterozygotes from disease. An understanding of how mutations influence both the kinetics and thermodynamics of misfolding allows us to rationalize the phenotypic diversity of amyloid diseases, especially when considered in concert with other genetic and environmental data.

Amyloid diseases are a large group of an even larger collection of misfolding disorders, the former including greater than 80 familial transthyretin (TTR)-based pathologies (1–11). The TTR missense mutations associated with familial amyloid disease display a wide range of diversity in age of disease onset, penetrance, etc. In diseases resembling and including the TTR amyloidoses, normally folded secreted proteins must first undergo partial denaturation to assemble into amyloid (7–10, 12). Although amyloidosis is modulated by numerous environmental and genetic factors (13), it is known that mutations that destabilize the native state predispose an individual to disease (7, 8, 14). However, thermodynamic stability alone does not reliably predict disease severity (15).

Tetramer dissociation is required for TTR amyloid fibril formation (12, 16, 17). However, the resulting normally folded monomer cannot form amyloid without undergoing partial denaturation (12, 16, 17), yielding the so-called monomeric amyloidogenic intermediate composed of a three-stranded antiparallel β-sheet structure (18). Herein we use chaotropic denaturation studies in an attempt to understand energetic differences between single site variants of TTR, all of which are tetrameric under physiological conditions. These studies demonstrate that kinetic and thermodynamic data, considered together, nicely rationalize why certain mutations lead to severe pathology, whereas others protect against disease or lead to mild pathology.

Methods

Variant TTR Production and Purification. Recombinant WT TTR and variants thereof were expressed in BL21/DE3 Epicurian gold *Escherichia coli* (Stratagene) transformed with pmmHα plasmid containing the TTR and ampicillin-resistance genes. Expression and purification were performed as described in detail previously (19). Recombinant expression of WT, L55P, V30M, V122I, and T119M TTR homotetramers in *E. coli* at 37°C each provide 30–50 mg/liter of purified native tetrameric protein.

TTR Fibril Formation Kinetics. TTR was buffer exchanged into 10 mM phosphate buffer (pH 7.2) containing 100 mM KCl, 1 mM EDTA, and 1 mM DTT. Solutions of TTR (0.40 mg/ml) were mixed with an equal volume of 200 mM acetate buffer (pH 4.3) containing 100 mM KCl, 1 mM EDTA to yield a final pH of 4.4. The samples were incubated at 37°C without stirring. The turbidity at 400 nm was measured as a function of time up to 360 h. The self-assembly reaction was also followed by thioflavin T (ThT) binding as described previously (20). In fibril formation facilitated by MeOH-mediated dielectric constant lowering, TTR was first buffer-exchanged into 50 mM Tris·HCl (pH 7.0) containing 100 mM KCl, 1 mM EDTA, and 1 mM DTT (Tris buffer). Two hundred microliters of this TTR solution (1.5 mg/ml) was then added to 2.8 ml of a methanol/Tris buffer solution at 25°C with constant stirring to yield TTR (0.10 mg/ml) solvated in 50% (vol/vol) MeOH in Tris buffer (pH 7.0). The turbidity at 330 and 400 nm was continuously monitored over the course of 3,600 s. The rate of fibril formation was evaluated by the initial slope of the turbidity at 400 nm (Fig. 4*b*).

Urea-Mediated TTR Dissociation Measured by Resveratrol Binding. Resveratrol displays a large increase in its fluorescence quantum yield and a blue shift on binding to tetrameric TTR but does not bind to the TTR monomer (X.J., P.H., and A. Sawkar, unpublished results). Resveratrol-binding curves to quantify the concentration of TTR tetramer were recorded for each TTR variant (0–0.12 mg/ml; 0–2.18 $\mu M_{tetramer}$) by using 18 μM resveratrol. The fluorescence intensity at 394 nm (I^{394}) was plotted versus the concentration of TTR, exhibiting a linear fit as expected. Standard curves are provided in Fig. 6, which is published as supporting information on the PNAS web site, www.pnas.org.

The resveratrol probe was added to TTR subjected to urea denaturation just before the fluorescence measurement in order not to shift the equilibrium toward tetramer significantly. Fortunately, this ligand does not noticeably perturb the reversible concentration-dependent tetramer/monomer equilibrium in urea based on quantification of the tetramer using crosslinking studies (concentration dependence and reversibility data are unpublished). Resveratrol fluorescence was recorded from 350 to 550 nm after excitation at 320 nm. Samples containing TTR (0.10 mg/ml/1.8 $\mu M_{tetramer}$) were incubated as a function of urea concentration, buffered with 50 mM phosphate (pH 7.0) containing 100 mM KCl, 1 mM EDTA, and 1 mM DTT. After incubation (96 h), 3.6 μl of resveratrol from a 2.5 mM stock solution in DMSO was added to a 500-μl protein sample just before the measurement yielding a final concentration of 18 μM resveratrol.

Urea Unfolding of TTR Measured by Tryptophan Fluorescence. Samples containing TTR (0.10 mg/ml) were incubated (25°C) in varying concentrations of urea buffered with 50 mM phosphate (pH 7.0) containing 100 mM KCl, 1 mM EDTA, and 1 mM DTT. Tryptophan fluorescence spectra were recorded between 310 and 410 nm with excitation at 295 nm. Equilibrium data were recorded after a 96-h incubation period sufficient to reach equilibrium in all cases except T119M. The fluorescence ratio at 355 and 335 nm was used as a structural probe as described previously (21).

Kinetics of Monomer Unfolding and Tetramer Dissociation as a Function of Urea. Unfolding of monomeric TTR (M-TTR; F87M/L110M) (17) was monitored by stopped-flow fluorescence (355 nm, 4°C) using a 1:10 dilution of M-TTR (50 μM) in 50 mM phosphate buffer containing 100 mM KCl, 1 mM EDTA, and 1 mM DTT into 5.55 M urea (Aviv Associates, Lakewood, NJ) ATF-105 spectrometer]. The unfolding rate of tetrameric WT TTR was also measured at 4°C [which is faster than at 25°C, due to lower stability (21)]. Because the monomer unfolds five to six orders of magnitude faster than tetrameric TTR dissociates, the rate-determining dissociation step was measured by purposefully linking the quaternary structural changes to tertiary structural changes (measured by tryptophan fluorescence) mediated by urea concentrations in the unfolding posttransition region. Time courses (25°C) were recorded for WT, V30M, L55P, and V122I TTR (27 s = dead time for manual mixing) up to 250–300 h. The final data points for the T119M variant were recorded after 480 h. No burst phase was observed for any TTR sequence. The kinetic data fit well to a single exponential function: $I^{355/335} = I_N^{355/335} + A(1 - e^{-k_{diss}t})$, where $I_N^{355/335}$ is the native protein fluorescence intensity ratio (355/335 nm), A is the amplitude difference, k_{diss} is the tetramer dissociation rate constant, and t is time in hours.

TTR Reassembly Kinetics. WT and T119M TTR were unfolded by incubation in 8 M urea at 4°C for 7 days [TTR is destabilized at low temperatures, facilitating denaturation (21)]. The proteins were confirmed to be unfolded by Trp fluorescence and the lack of resveratrol binding. The reassembly reaction was initiated by diluting unfolded TTR 10-fold (to 1.8 $\mu M_{tetramer}$) with phosphate buffer to achieve the desired final urea concentration in the presence of 18 μM resveratrol. Resveratrol fluorescence was monitored at 390 nm after excitation at 320 nm as a function of time. The TTR reassembly reaction was biphasic and the data best fit to a double exponential function. Under these conditions, the amplitudes of the two phases were equal for WT and depended on the urea concentration in the case of T119M. The final yield of reassembly for both variants is 90%, demonstrating that the process is reversible.
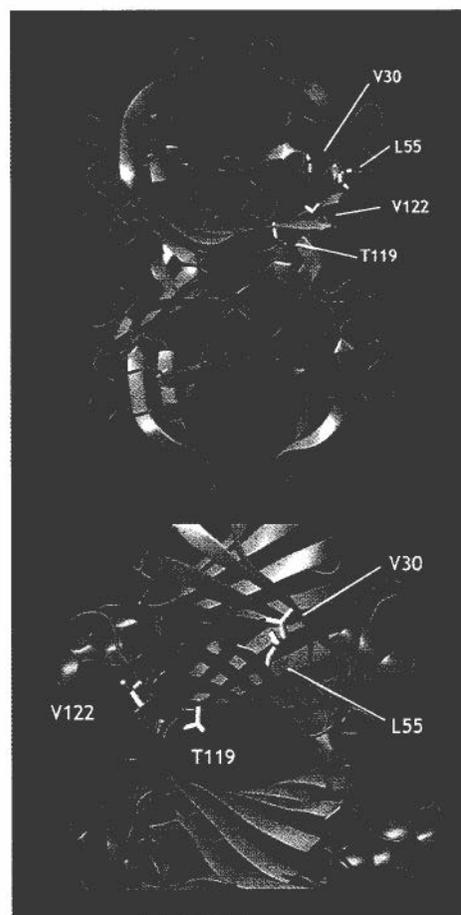

Fig. 1. The structure of tetrameric TTR, showing the location of the point mutations that change the kinetics and/or thermodynamics of partial denaturation and influence amyloidogenicity and disease phenotype (mutations identified in only one subunit). The V30M TTR mutation is located in the hydrophobic core, whereas the L55P substitution is located in β-strand D. The V122I TTR mutation is close to the C terminus and the dimer/dimer interface region of TTR. T119M contributes to the dimer/dimer interface and comprises the thyroxine-binding sites. The T119M TTR mutation appears to stabilize the quarternary structure and dramatically slows the rate of tetramer dissociation, enabling it to function as a transsuppressor of misfolding (20), apparently by increasing the hydrophobic surface area buried at the dimer/dimer interface (34).

Results and Discussion

Relative Thermodynamic Stability of the TTR Variants. In this study, we used TTR tetramers composed of identical monomer subunits to elucidate the effects of human mutations (Fig. 1) on thermodynamic stability, rate of tetramer dissociation, and rate of amyloid fibril formation. TTR tetramers do not denature in urea; hence dissociation to monomers is required for urea-induced tertiary structural changes detected by tryptophan fluorescence (Fig. 2b) (15, 21).

In principle, TTR quaternary and tertiary structural changes can be unlinked by conducting biophysical experiments at low TTR concentrations. Under these conditions, the tetramer-folded monomer equilibrium is concentration dependent, whereas the folded–unfolded monomer equilibrium should be concentration independent. Over the physiologically relevant range of concentrations studied thus far (0.1−0.5 mg/ml), the unfolding transitions detected by Trp fluorescence changes exhibit a TTR concentration dependence, strongly suggesting that the quaternary and tertiary structural transitions are linked for V30M, L55P, WT, V122I, and T119M TTR. Further evi-
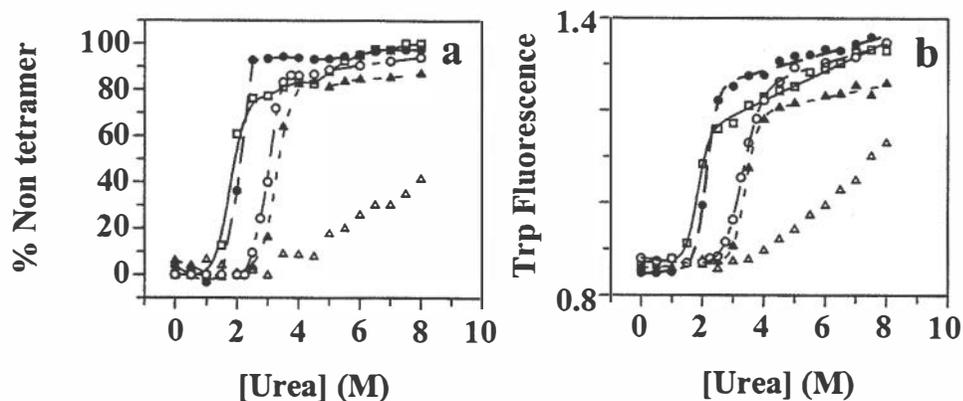
Fig. 2. Evaluation of the stability of TTR sequences as a function of urea concentration. WT, filled triangles; V30M, open squares; L55P, filled circles; V122I, open circles; T119M, open triangles. (a) Tetramer dissociation curve measured by resveratrol binding (96-h incubation). (b) Tertiary structure unfolding curve measured by intrinsic tryptophan fluorescence changes.

dence that these transitions are linked includes the nearly coincident quaternary and tertiary structural transitions mediated by increasing urea concentrations (Fig. 7, which is published as supporting information on the PNAS web site).

The fraction of TTR tetramer retained as a function of urea concentration (quaternary structural transition) was evaluated by using the small molecule resveratrol, which fluoresces when bound to at least one of the two thyroid-binding sites in the tetramer (Fig. 2a). The dissociation of TTR into monomeric subunits is not associated with intrinsic fluorescence or circular dichroism changes; therefore, ligand binding was used to probe the tetramer–monomer equilibrium (17). Although it is reasonable to hypothesize that this method may overestimate the fraction tetramer as a function of denaturant by shifting the equilibrium toward tetramer (Le Chatelier's principle), this was not noticeable when the resveratrol data were compared with analytical ultracentrifugation and glutaraldehyde crosslinking data (unpublished data). The dissociation curves depicted in Fig. 2a require a 96-h incubation period due to the high kinetic barrier associated with tetramer dissociation (see below). Trp fluorescence as a function of urea concentration is used to follow TTR tertiary structural changes (unfolding; Fig. 2b).

The tetramers composed exclusively of mutated subunits associated with disease are clearly less stable than WT based on their denaturation midpoints for dissociation or unfolding. Furthermore, all of the TTR variants have similar sensitivity to urea concentration, allowing direct comparisons. The stability of V30M and L55P is very similar and lower than WT, which is similar to V122I. The T119M suppressor tetramer appears to be the most stable (Fig. 2a); however, predictions about relative thermodynamic stability have to be made with caution because of its extremely slow dissociation rate (it does not reach equilibrium within 96 h). An identical rank ordering of stabilities is exhibited by comparing unfolding transitions evaluated by Trp fluorescence (Fig. 2b), consistent with linked tetramer dissociation and unfolding equilibria.

The Influence of Mutations on TTR Tetramer Dissociation Kinetics. It is perplexing that the disease-associated variants L55P and V30M, which are similarly destabilized relative to WT TTR, exhibit dramatically different clinical features, including disease penetrance and age of onset (22, 23). To understand this discrepancy, we explored the influence of mutations on the rate of tetramer dissociation and amyloid fibril formation. Given that the WT tetramer cannot undergo tertiary structural changes before it dissociates, one can detect the rate of quaternary structural changes by linking tetramer dissociation to TTR

tertiary structural changes, provided that the tertiary structural changes are much faster. The rate of monomeric TTR (17) denaturation ($t_{1/2}$ = 69 ms; in 5.0 M urea, 4°C, Fig. 3a *Inset*) is $\approx 5 \times 10^5$-fold faster than the dissociation rate of WT TTR ($t_{1/2}$ = 9.6 h; in 5.0 M urea, 4°C), demonstrating this to be the case. Using urea concentrations in the posttransition region for tertiary structural changes directly links the slow TTR quaternary structural changes to the rapid tertiary structural changes and renders unfolding irreversible. Dissociation time courses for the TTR variants described above fit well to a single exponential over a range of urea concentrations. Representative 6.0 M urea dissociation time courses are depicted in Fig. 3a.

Tetramer Dissociation Kinetics. It is clear that single amino acid changes in the TTR sequence can have a significant and not easily predicted influence on the rate of tetramer dissociation (Table 1). The logarithm of the rate constant ($\ln k_{diss}$) varies linearly with urea concentration (Fig. 3b), allowing extrapolations to more physiological conditions (0 M urea). The dissociation half life ($t_{1/2}$) of WT TTR is 42 h (0 M urea, 25°C). The most pathogenic familial amyloid polyneuropathy variant (L55P) dissociates 10-fold faster ($t_{1/2}$ = 4.4 h), whereas the V122I cardiac variant dissociates 2.2-fold faster than WT under identical conditions. Similar relative rates are exhibited under denaturing conditions (Fig. 3b). The L55P and V122I mutations lower the activation barrier for dissociation apparently by destabilizing the tetrameric ground state more than the transition state associated with dissociation. Amyloidogenic lysozyme mutations also destabilize the structure and increase the rate of denaturation relative to WT, suggesting that the rate of denaturation could be generally important in amyloidoses (24, 25). Furthermore, it has been shown that thermodynamic destabilization of the prion protein by familial point mutations is unlikely to be the determinant for disease phenotype (26).

In stark contrast to the disease-associated TTR sequences, the T119M TTR variant exhibits an exceedingly slow tetramer dissociation rate ($t_{1/2}$ of 1,534 h), demonstrating that this suppressor protects against amyloid disease by effectively precluding tetramer dissociation on a biologically relevant time scale. Hence, T119M misfolding is prevented by kinetic stabilization, because the activation barrier for dissociation is insurmountable (27, 28). The very slow dissociation rate exhibited by the T119M suppressor homotetramer is consistent with previous experiments that were unable to detect dissociation by subunit exchange (20, 29). The V30M mutant dissociates slightly slower than WT TTR (Fig. 3b), demonstrating that there are mutations that destabilize TTR significantly without increasing its disso-
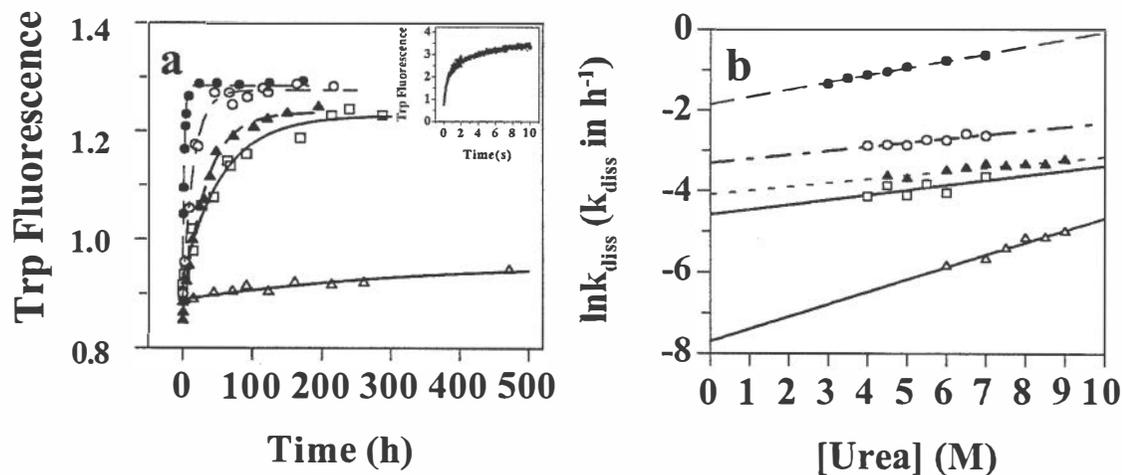
Fig. 3. TTR tetramer dissociation rates as a function of urea concentration (Fig. 2 symbols apply). (a) Unfolding time course measured by tryptophan fluorescence provides the rate of TTR tetramer dissociation in 6.0 M urea. *Inset* shows the rapid unfolding of monomeric TTR in 5.0 M urea measured by stopped-flow fluorescence. (b) The logarithm of the rate of tetramer dissociation, $\ln k_{diss}$ (k_{diss} in h^{-1}) plotted as a function of urea concentration. The $\ln k_{diss}$ vs. urea concentration plot is linear, allowing extrapolation to 0 M urea.

ciation rate (apparently V30M nearly equally destabilizes the ground and transition states of dissociation or changes the pathway for denaturation). That the V30M and WT tetramer dissociation rates are similar under physiological conditions is also supported by subunit exchange rates (20, 29). Even though the V30M tetramer is slightly more destabilized than L55P, the V30M disease phenotype is milder because V30M dissociates slowly, even more slowly than WT. For this and related arguments to be relevant to disease, tetramer dissociation rates have to correlate with amyloid formation rates, as demonstrated below.

Fibril Formation Rates Are Predicted by Tetramer Dissociation Rates. TTR will form amyloid fibrils *in vitro* under partially denaturing conditions imposed by lowering either the pH (simulating the endocytic pathway) or the dielectric constant of the aqueous medium (12, 20). The mechanism of TTR amyloid fibril formation is not yet fully understood; however, it is clear that it lacks a lag phase and is not seedable (J. White and J.W.K., unpublished data). The initial rates of TTR amyloid fibril formation for all of the variants shown in Fig. 4a (37°C, 0.20 mg/ml of TTR, pH 4.4) fit to single exponentials. This observation and others (15, 17) provide strong evidence that the rate-determining step for TTR amyloid formation is tetramer dissociation. In fact, the fibril formation rates of the TTR variants displayed in Fig. 4a are predictable from the tetramer dissociation rates, suggesting that the relative denaturation energy landscapes in two different denaturants (urea and acid) are similar. The most pathogenic variant, L55P, forms fibrils 9.2-fold faster that WT TTR, similar

to the V122I variant, which is 3.6-fold faster than WT (each characterized by 100% disease penetrance), whereas V30M is slightly slower than WT (V30M and WT TTR show incomplete disease penetrance). The T119M suppressor homotetramer forms fibrils 3,000-fold slower than WT, explaining how it likely protects against pathology in compound heterozygotes.

Lowering the dielectric constant of the aqueous medium solvating TTR by adding MeOH (50% vol/vol) also leads to partial denaturation and amyloid fibril formation (20). The fibril formation rate in aqueous MeOH is dramatically faster than that mediated by partial acid denaturation (compare Fig. 4 b to a). Nonetheless, the relative TTR tetramer dissociation rates still predict amyloid fibril formation velocity, implying that tetramer dissociation remains rate limiting. The L55P familial amyloid polyneuropathy (FAP) variant forms fibrils with a relative rate of 3.4, similar to the V122I familial amyloid cardiomyopathy variant (3.1), whereas the rate of the V30M FAP variant (0.92) is similar to WT (1), in stark contrast to T119M (0.048), which forms amyloid 20-fold slower than WT TTR.

The T119M Transsuppressor Exhibits High Kinetic Stability. The 40-fold slower dissociation rate of the T119M homotetramer relative to WT is consistent with a high kinetic barrier of dissociation (Table 1, Fig. 3). To provide further evidence for an increase in barrier height relative to WT, we monitored the reassembly kinetics of T119M and compared them to WT. The reassembly rate of T119M is 90- or 200-fold slower than WT, depending on which of the two phases are compared (Fig. 5). This demonstrates that the barriers for T119M dissociation and reassociation are both considerably higher than those characterizing WT. The increased T119M barriers lead to a tetramer exhibiting very high kinetic stability (27, 30) under amyloidogenic conditions. Preventing tetramer dissociation by increasing the kinetic barrier should be a very effective strategy to confer stability on a protein that can adopt a lower free energy amyloid state under denaturing conditions (28). The transsuppressor efficacy exhibited by inclusion of T119M subunits in V30M/T119M hybrid tetramers reported previously (20) is likely to be mediated by barrier height tuning, although this has not been directly demonstrated.

Sequence-Dependent Energetics Significantly Contribute to Amyloid Disease Diversity. The sequence-dependent variation in TTR tetramer dissociation rates (rate-determining step for amyloid
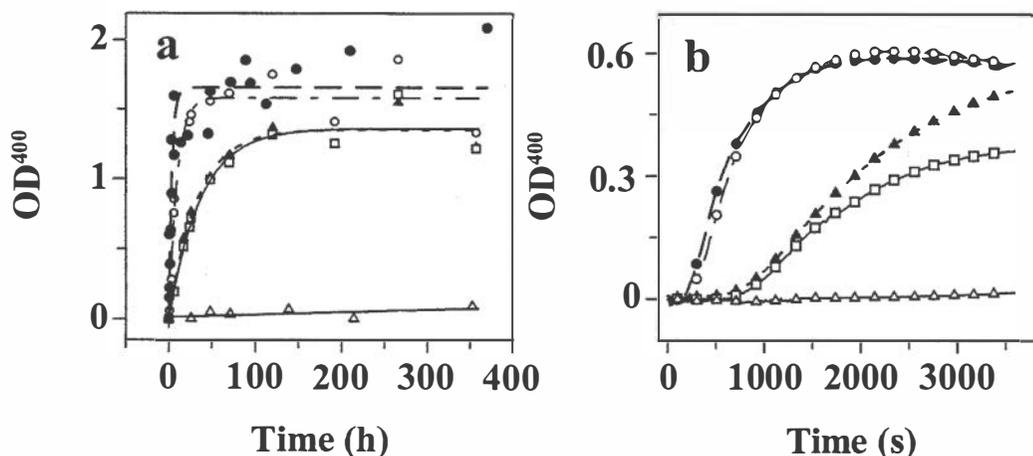
Table 1. TTR homotetramer dissociation rates derived from time courses as a function of urea concentration extrapolated to 0 M urea concentration

Sequence	$k_{diss}*$, h^{-1}	$m^{kin}*$, M^{-1}
WT	$1.68 \pm 0.15 \cdot 10^{-2}$	0.094 ± 0.013
V30M	$1.02 \pm 0.37 \cdot 10^{-2}$	0.121 ± 0.066
L55P	$15.7 \pm 0.58 \cdot 10^{-2}$	0.179 ± 0.008
V122I	$3.64 \pm 0.38 \cdot 10^{-2}$	0.100 ± 0.02
T119M	$4.52 \pm 1.18 \cdot 10^{-4}$	0.303 ± 0.03

*k_{diss} is the tetramer dissociation rate constant, and m^{kin} is the urea dependence of the tetramer dissociation rate constant.

Fig. 4. TTR fibril formation time courses detected by turbidity at 400 nm (OD^{400}). (a) Amyloid fibril formation mediated by partial acid denaturation of TTR (0.20 mg/ml) in acetate buffer (pH 4.4, 37°C, Fig. 2 symbols apply). (b) TTR (0.10 mg/ml) amyloid fibril formation enabled by MeOH-induced denaturation [50% (vol/vol) in Tris buffer (pH 7.0, 25°C)].

fibril formation), when considered in combination with the apparent thermodynamic stability of the tetramer (as judged by susceptibility to urea denaturation, a valid comparison due to the similarity in the m values presented by the denaturation curves in Fig. 2 a and b), nicely rationalizes the clinical data associated with the five TTR sequences studied herein. Although thermodynamic stability dictates whether amyloid formation is possible in a given denaturing environment, the rate at which the tetramer dissociates to the partially unfolded monomeric amyloidogenic intermediate governs the rate of fibril formation and therefore contributes significantly to disease severity. L55P TTR exhibits severe pathology, because the tetramer both dissociates rapidly and is highly destabilized, explaining why this mutation confers 100% disease penetrance with the earliest age of disease onset (15–25 years). Even though the V30M tetramer is slightly more destabilized than L55P, the disease phenotype is milder because V30M dissociates even more slowly than WT. Therefore, the

V30M monomeric amyloidogenic intermediate cannot form to the extent prescribed by thermodynamics because the slow tetramer dissociation rate limits its steady-state concentration— suggesting why penetrance of V30M disease can be as low as 2% (23, 31). Even though the V122I tetramer is similar in stability to WT TTR, the tetramer dissociates >2-fold faster than WT, causing cardiac amyloid disease with near 100% penetrance, unlike the <25% penetrance exhibited by WT cardiac disease (32). Compound heterozygotes expressing both T119M and V30M TTR develop a mild late-onset pathology—if at all (20). This may be explained by extending the slow dissociation rates exhibited by the T119M homotetramers studied within to the mixed V30M/T119M tetramers showing dramatically lower amyloidogenicity *in vitro* (20). The amyloidogenicity of TTR and disease phenotype is also likely influenced by the stoichiometry of TTR-binding partners and other factors whose concentrations are dictated by genetic background (13).
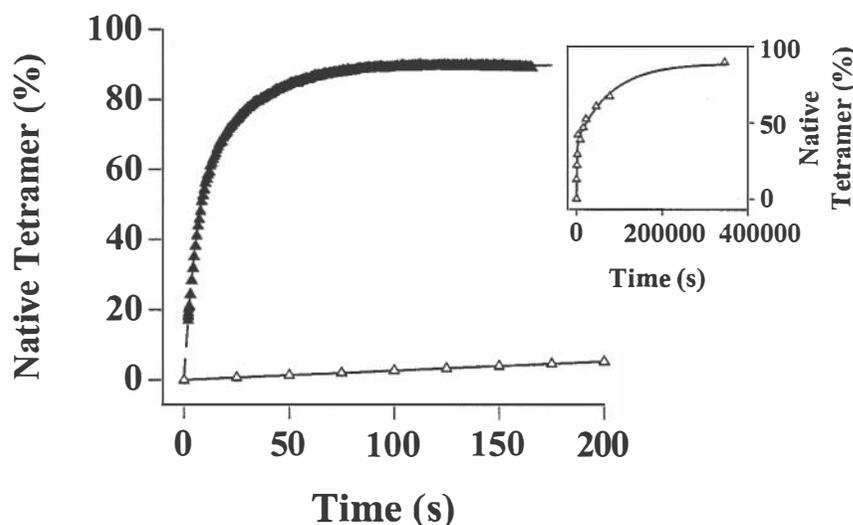

Fig. 5. The kinetics of WT (filled triangles) and T119M TTR (open triangles) reconstitution (folding and reassembly) monitored by resveratrol fluorescence (monitors tetramer formation). A 10-fold dilution of urea unfolded TTR (8 M) to a final TTR concentration of 1.8 $\mu M_{tetramer}$ initiates reconstitution (final urea concentration = 1.0 M). *Inset* shows the complete trace for the very slow T119M reassembly time course. The rate constants (extrapolated to 0 M urea) for the fast phase are 0.294 ± 0.014 s^{-1} (WT) and 3.16·10^{-3} ± 1.65·10^{-3}·s^{-1} (T119M). For the slow phase, the rate constants are 6.24·10^{-2} ± 0.62·10^{-2}·s^{-1} (WT) and 2.76·10^{-4} ± 2.3·10^{-4}·s^{-1} (T119M). Therefore, the reassembly rate of T119M (extrapolated to 0 M urea) is 93-fold slower (fast phase) and 226-fold slower (slow phase) relative to WT.

The 80 different TTR familial amyloid disease mutations coupled with the availability of clinical data will allow us to further scrutinize the hypothesis that it is necessary to consider both thermodynamics and the kinetics of partial denaturation to rationalize the spectrum of amyloid disease phenotypes. Interestingly, disease-associated mutations are also proposed to accelerate the conversion of the monomeric prion protein to a misfolded state possibly associated with prion pathology by analogous alterations in the free energy landscape (26, 33). From the data now available, it is clear that TTR mutations influencing amyloidogenicity can change thermodynamic stability without significant changes in the rate of partial denaturation or *vice versa*. Alternatively, mutations can change both the thermodynamics and the kinetics of partial denaturation. Mutations that destabilize the structure and increase the rate of monomer accumulation lead to severe amyloid diseases (e.g., L55P). Single amino acid sequence changes that increase the denaturation rate without significant structural destabilization (e.g., V122I) lead to highly penetrant diseases of intermediate severity. Mutations that destabilize without increasing the denaturation rate (e.g., V30M) lead to diseases with incomplete penetrance and intermediate severity, whereas mutations that apparently stabilize the homotetramer and slow the rate of partial denaturation (e.g., T119M) protect against amyloid disease onset in the context of compound heterozygotes.

How and where amyloid fibrils form in a human being are not yet established. That many amyloidogenic proteins and peptides form fibrils under acidic conditions suggests this may be an intracellular process, perhaps occurring within an organelle such as a lysosome. Despite significant effort, however, we have not been able to demonstrate intracellular TTR amyloidosis. The slow tetramer dissociation process required for amyloid fibril formation revealed within may provide some clues regarding how and where amyloidosis occurs in a human. The time scale of dissociation may be surprising if one assumes that the rate-limiting step of TTR amyloidosis is the formation of a high-energy multimeric intermediate referred to as a nucleus. Nucleated polymerizations are common for amyloidogenic peptides; however, TTR amyloidogenesis is different and very efficient in that it proceeds by a downhill polymerization mechanism not requiring high-energy multimeric nucleus formation (J. White and J.W.K., unpublished data). The monomeric misfolded form of TTR appears to be polymerization competent, consistent with the inability of TTR amyloidosis to be accelerated by seeding with preformed fibrils or protofilaments (fibril precursors). The slow dissociation process and the efficiency of TTR amyloidosis may provide useful constraints for discerning how and where amyloid fibrils form in mammals.

1. Kelly, J. W. (1996) *Curr. Opin. Struct. Biol.* **6**, 11–17.
2. Dobson, C. M. (1999) *Trends Biochem. Sci.* 329–332.
3. Goldberg, M. S. & Lansbury, P. T., Jr. (2000) *Nat. Cell Biol.* **2**, E115–E119.
4. Fink, A. L. (1998) *Folding Des.* **3**, R9–R23.
5. Uemichi, T. (1997) *Rinsho Kagaku (Nippon Rinsho Kagakkai)* **26**, 74–87.
6. Saraiva, M. J. M. (1995) *Hum. Mutat.* **5**, 191–196.
7. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. F. & Pepys, M. B. (1997) *Nature (London)* **385**, 787–793.
8. Hurle, M. R., Helms, L. R., Li, L., Chan, W. & Wetzel, R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 5446–5450.
9. McParland, V., Kad, N., Kalverda, A., Brown, A., Kirwin-Jones, P., Hunter, M., Sunde, M. & Radford, S. (2000) *Biochemistry* **39**, 8735–8746.
10. Chiti, F., Mangione, P., Andreola, A., Giorgetti, S., Stefani, M., Dobson, C., Bellotti, V. & Taddei, N. (2001) *J. Mol. Biol.* **307**, 379–391.
11. Jacobson, D. R., Pastore, R. D., Yaghoubian, R., Kane, I., Gallo, G., Buck, F. S. & Buxbaum, J. N. (1997) *New Engl. J. Med.* **336**, 466–473.
12. Colon, W. & Kelly, J. W. (1992) *Biochemistry* **31**, 8654–8660.
13. White, J. T. & Kelly, J. W. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 13019–13024.
14. McCutchen, S. L., Lai, Z., Miroy, G., Kelly, J. W. & Colon, W. (1995) *Biochemistry* **34**, 13527–13536.
15. Jiang, X., Buxbaum, J. N. & Kelly, J. W. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 14943–14948.
16. Lai, Z., Colon, W. & Kelly, J. W. (1996) *Biochemistry* **35**, 6470–6482.
17. Jiang, X., Smith, C. S., Petrassi, H. M., Hammarström, P., White, J. T., Sacchettini, J. C. & Kelly, J. W. (2001) *Biochemistry* **40**, 11442–11452.
18. Liu, K., Cho, H. S., Lashuel, H. A., Kelly, J. W. & Wemmer, D. E. (2000) *Nat. Struct. Biol.* **7**, 754–757.
19. Lashuel, H. A., Wurth, C., Woo, L. & Kelly, J. W. (1999) *Biochemistry* **38**, 13560–13573.
20. Hammarström, P., Schneider, F. & Kelly, J. W. (2001) *Science* **293**, 2459–2461.
21. Hammarström, P., Jiang, X., Deechongkit, S. & Kelly, J. W. (2001) *Biochemistry* **40**, 11453–11459.
22. Jacobson, D. R. & Buxbaum, J. N. (1991) *Adv. Hum. Genet.* **20**, 69–123.
23. Coelho, T. (1996) *Curr. Opin. Neurol.* **9**, 355–359.
24. Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V. & Dobson, C. M. (1999) *Biochemistry* **38**, 6419–6427.
25. Takano, K., Funahashi, J. & Yutani, K. (2001) *Eur. J. Biochem.* **268**, 155–159.
26. Liemann, S. & Glockshuber, R. (1999) *Biochemistry* **38**, 3258–3267.
27. Baker, D., Sohl, J. L. & Agard, D. A. (1992) *Nature (London)* **356**, 263–266.
28. Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. (2002) *Nature (London)* **415**, 343–346.
29. Schneider, F., Hammarström, P. & Kelly, J. W. (2001) *Protein Sci.* **10**, 1606–1613.
30. Sohl, J. L., Jaswal, S. S. & Agard, D. A. (1998) *Nature (London)* **395**, 817–819.
31. Holmgren, G., Costa, P. M., Andersson, C., Asplund, K., Steen, L., Bechman, L., Nylander, P. O., Teixeira, A., Saraiva, M. J. & Costa, P. P. (1994) *J. Med. Genet.* **31**, 351–354.
32. Westermark, P., Sletten, K., Johansson, B. & Cornwell, G. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2843–2845.
33. Baskakov, I. V., Legname, G., Prusiner, S. B. & Cohen, F. E. (2001) *J. Biol. Chem.* **276**, 19687–19690.
34. Sebastiao, M. P., Lamzin, V., Saraiva, M. J. & Damas, A. M. (2001) *J. Mol. Biol.* **306**, 733–744.

Colloquium

The insulation of genes from external enhancers and silencing chromatin

Bonnie Burgess-Beusse, Catherine Farrell, Miklos Gaszner, Michael Litt, Vesco Mutskov, Felix Recillas-Targa, Melanie Simpson, Adam West, and Gary Felsenfeld*

Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0540

Insulators are DNA sequence elements that can serve in some cases as barriers to protect a gene against the encroachment of adjacent inactive condensed chromatin. Some insulators also can act as blocking elements to protect against the activating influence of distal enhancers associated with other genes. Although most of the insulators identified so far derive from *Drosophila*, they also are found in vertebrates. An insulator at the 5′ end of the chicken β-globin locus marks a boundary between an open chromatin domain and a region of constitutively condensed chromatin. Detailed analysis of this element shows that it possesses both enhancer blocking activity and the ability to screen reporter genes against position effects. Enhancer blocking is associated with binding of the protein CTCF; sites that bind CTCF are found at other critical points in the genome. Protection against position effects involves other properties that appear to be associated with control of histone acetylation and methylation. Insulators thus are complex elements that can help to preserve the independent function of genes embedded in a genome in which they are surrounded by regulatory signals they must ignore.

Although DNA methylation has been considered the primary chromosomal modification involved in transmission of epigenetic information, it is becoming clear that the chromatin proteins, especially the histones, are also involved in this process. It is understood that the histones bound to transcriptionally active and inactive regions of the genome are chemically modified in different ways, and mechanisms have been proposed by which these modified states can be propagated along the chromosome and perhaps transmitted during replication.

We have for some time been interested in the boundary elements called insulators. Detailed study of these elements has led us to mechanisms that are coupled to allele-specific expression at an imprinted locus, as well as to mechanisms by which propagation of inactive chromatin states may be modulated.

An Insulator as an Enhancer-Blocking Element

Within the vertebrate genome, transcriptionally active genes are embedded in an environment containing in some cases extensive regions of condensed chromatin. In other cases active genes may be located near other, silent, genes that have a different program of expression. The possibilities thus arise that the active gene will be inappropriately silenced by the condensed chromatin or will inappropriately activate the adjacent silent gene. It is equally possible that in other tissues or at other developmental stages, where this gene is inactive, signals from adjacent extraneous enhancers could cause incorrect patterns of expression.

During the past several years studies begun initially in *Drosophila* and now extended to vertebrates have identified DNA sequence elements called insulators that appear to function as blocks against both kinds of signals from the outside. Two kinds of assay have been developed to measure these properties (Fig. 1). The first assay measures "enhancer blocking," the ability to shield a promoter from the action of a distal enhancer without preventing the enhancer from working on a proximal promoter. The second assay measures the "barrier" activity that prevents the advance of adjacent condensed chromatin. We have applied both of these assays in the study of vertebrate insulators.

Our attention was first drawn to this problem through our interest in the role of chromatin structure in the regulation of gene expression in the chicken β-globin locus (Fig. 2). The gene cluster contains four members of the β-globin family, expressed at different developmental stages. The regulatory elements of these genes are marked by a series of erythroid-specific DNase hypersensitive sites (HSs), but at the 5′ end of the locus there is a "constitutive" HS (5′HS4) present in all tissues that have been examined (1–3). We speculated that this HS might mark the 5′ boundary of the "open" globin chromatin domain, and indeed subsequent work (4) has shown (Fig. 2) that there is an abrupt change from a chromatin structure characterized by general heightened nuclease sensitivity and a high level of histone acetylation (signs of an active globin chromatin locus), to a region of condensed chromatin further upstream, that is nuclease resistant and underacetylated.

To test whether the DNA at 5′HS4 had properties of an insulator we devised a method to assay enhancer blocking activity (Fig. 3A), based on the ability of a 1.2-kb element that includes the HS to shield a reporter expressing a neomycin resistance gene from the action of a strong enhancer. As shown in Fig. 3B, the 1.2-kb element is quite effective in reducing the number of G418-resistant colonies, a measure of the strength of the blockade. Placing the 1.2-kb element outside the region between enhancer and promoter resulted in little blocking (2, 3), confirming that 5′HS4 has insulating properties. We dissected this region further and found that a 250-bp "core" sequence containing the HS was equally effective in enhancer blocking. Further subdivision, making use of the DNase I footprint patterns generated on the core by nuclear extracts, revealed that a single binding site (footprint II) was also active. This finding led to the identification of a known regulatory protein, CTCF, as the DNA binding factor responsible for enhancer blocking activity (5).

We asked where else CTCF sites with enhancer blocking activity could be found. It seemed plausible that if these were truly associated with boundaries there might also be one at the 3′ end of the β-globin locus. Indeed a constitutive HS with this activity, and binding CTCF, is found just upstream of the 3′ condensed chromatin region (ref. 6, Fig. 4). This region also harbors a gene for an odorant receptor (7). It seems possible that
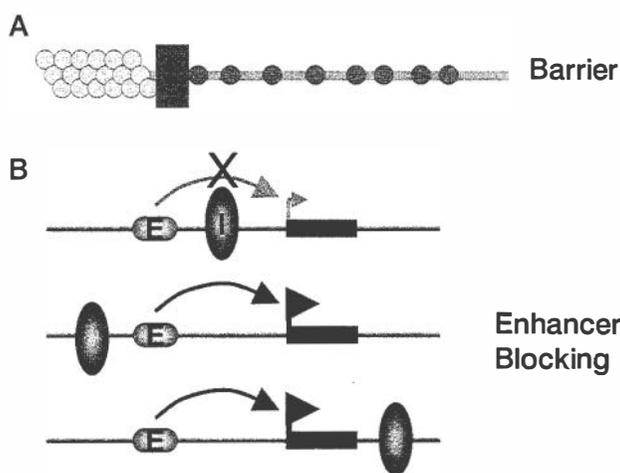
Fig. 1. Two kinds of insulator functions. (*A*) Some insulators may function as barriers against the encroachment of adjacent genomic condensed chromatin. (*B*) Some insulators may serve as positional enhancer-blocking elements that prevent enhancer action when placed between enhancer and promoter, but not otherwise.

the 3′ CTCF site could block cross-interaction between the regulatory elements of this receptor and those of the β-globin locus. CTCF sites are also present at conserved locations in the mouse and human loci that are embedded within clusters of odorant receptor genes (8). Further examination of the DNA upstream of the chicken 5′HS4 insulator reveals a somewhat different arrangement: Here, the upstream sequence is packaged as condensed chromatin containing CR1 repeat sequences. This sequence extends for about 16 kb and is followed by a gene (ref. 9, Fig. 4) for an erythroid-specific folate receptor (FR) that is expressed at a developmental stage preceding the point at which the globin genes are switched on. The FR gene and the globin genes are not expressed at the same stages, so that the presence of an enhancer blocking activity at 5′HS4 might once again be useful in avoiding cross-talk between the two gene systems.

A role for CTCF-mediated enhancer blocking activity has been demonstrated most clearly at the *Igf2/H19* locus in mouse

and human (10–12). In this imprinted locus the maternally transmitted allele expresses H19 but not Igf2, whereas the paternally transmitted allele expresses Igf2 but not H19. Furthermore the paternal allele is methylated at a site (the ICR or imprinted control region) located between the two genes (Fig. 5). Earlier work had suggested that the ICR might contain an enhancer blocking activity that would prevent downstream endodermal enhancers from activating *Igf2*. Direct examination reveals that the ICR contains four CTCF binding sites in the mouse ICR and seven in human. Methylation of these sites abolishes CTCF binding. These results indicate that CTCF plays an important role as an insulator protein in allele-specific regulation at this imprinted locus, and that the insulator function can be modulated by DNA methylation, thus making the CTCF sites susceptible to epigenetic regulation. Quite recently a cluster of differentially methylated CTCF sites has been identified at the *Xist* gene promoter, and it has been suggested that these are enhancer-blocking elements important for X chromosome inactivation (13).

Insulators as Barriers

A second property that some insulators possess is the ability to protect against silencing caused by formation of condensed chromatin. This is the barrier function, which can be detected by assays designed to measure protection of stably integrated transgenes against position effects (Fig. 1). We established an assay to test barrier function by constructing a reporter expressing a fragment of the IL-2 receptor (IL-2R) (Fig. 6) driven by an erythroid-specific promoter and enhancer, and integrating it into a chicken erythroid cell line, 6C2 (14). Typically after 80–100 days in culture expression was extinguished in most lines. This extinction is a manifestation of position effects, i.e., the dependence of expression on the site of integration. We repeated the experiment with the same reporter, but flanked on each side by two copies of the 1.2-kb 5′HS4 element; now expression was maintained in nearly all lines even after 80–100 days of incubation. The 5′HS4 element thus protects against position effects, a second property possessed by some insulators.

Silencing of gene expression can involve a number of chromatin modifications, and insulators might interfere with some of these. We turned again to the chicken β-globin locus and examined the state of modification of the histones over the entire
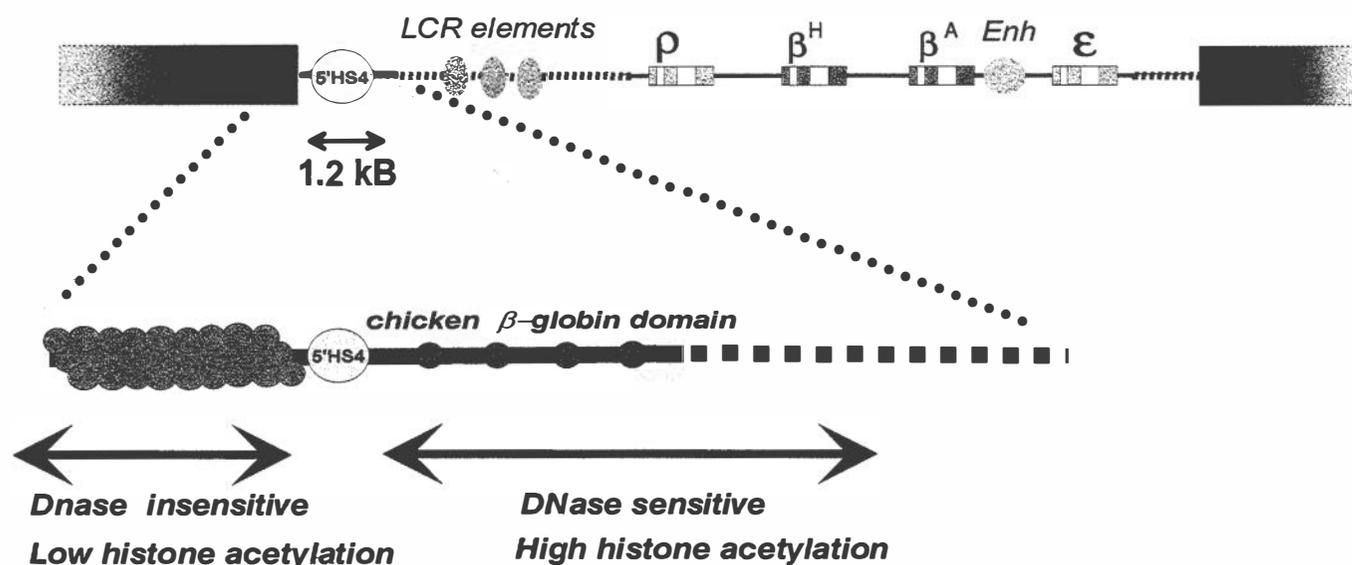

Fig. 2. The chicken β-globin locus (*Upper*) showing the four genes, the strong enhancer between the adult β gene (β^A) and the ε gene, and the constitutive HS, 5′HS4. The boundary between the open chromatin domain and the condensed chromatin domain further 5′, as determined by Hebbes et al. (4), is shown (*Lower*).
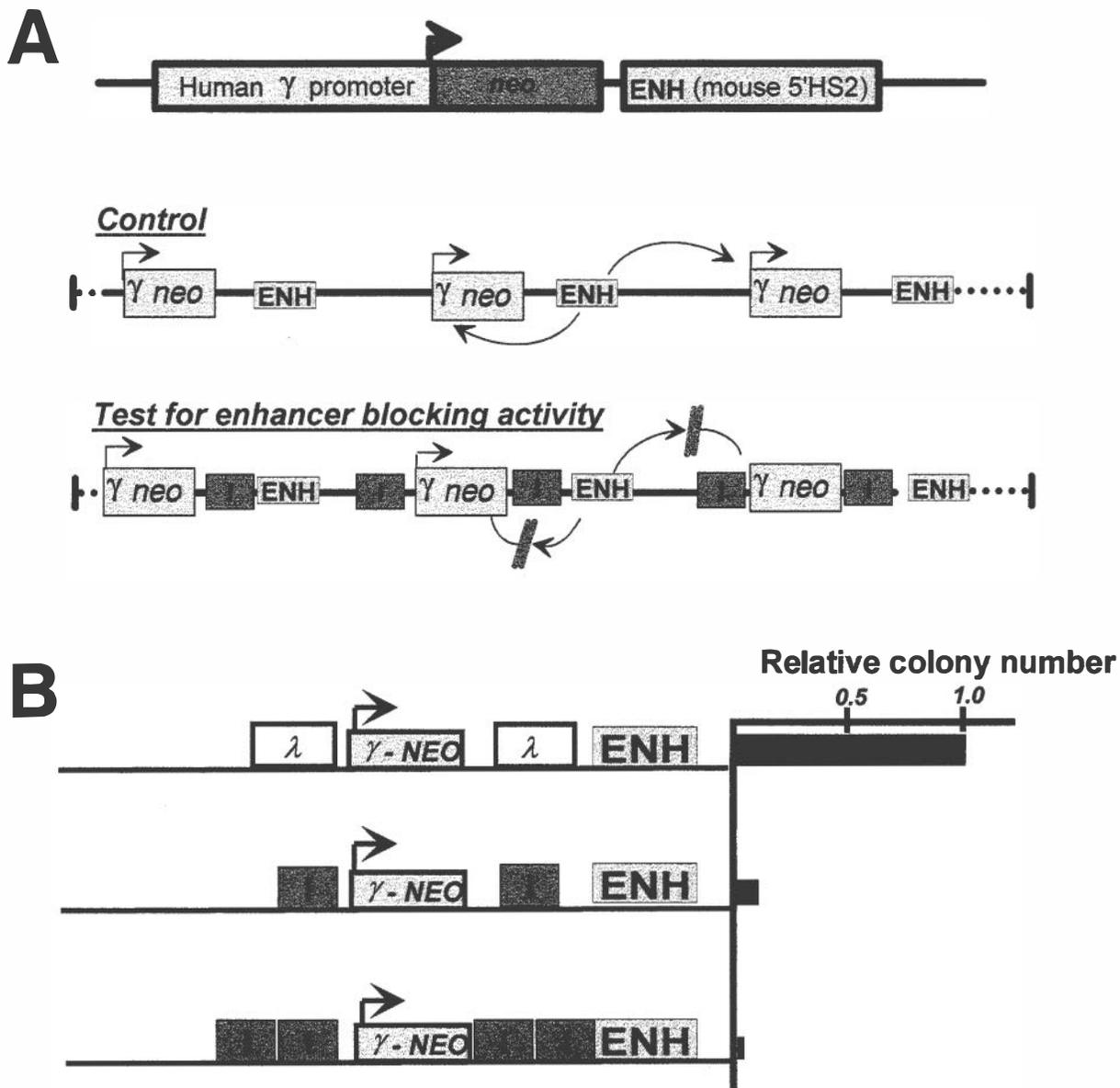
Burgess-Beusse *et al.*

Fig. 3. (A) Construction used in an assay for enhancer (ENH) blocking activity (2, 3). Expression of a gene conferring G418 resistance (*neo*) is driven by an erythroid-specific enhancer and promoter. This plasmid is stably transfected into K562 human erythroleukemia cells, and G418-resistant colonies are counted. Typically transfection produces tandem integrants (Control). The test for the putative insulator (I) is to insert it so that it can block enhancer action, and again to count colonies. (B) Results of inserting a 1.2-kb fragment (see Fig. 2) containing 5′HS4 on colony number in the above assay. The control has an approximately equal length of λ phage DNA on either side of the reporter to keep distances constant. The 5′HS4 element strongly reduces enhancer (ENH) activity. Other experiments show that it has a much smaller effect when placed on the other side of the enhancer, confirming the positional enhancer blocking activity (2, 3).

54-kb region containing the globin genes and the upstream FR (15, 16). Chromatin immunoprecipitation (ChIP) experiments with antibodies to acetylated histones H3 and H4 revealed elevated levels of acetylation associated with activation of the individual genes. Thus there was strong acetylation over the FR gene in 6C2 cells (Fig. 7), which express this gene, but in 10-day embryonic erythrocytes, corresponding to a later developmental stage, high levels of acetylation are shifted to the globin gene cluster (15). At all stages, the approximately 16-kb condensed chromatin region upstream of the globin genes remains unacetylated. Furthermore, there is a peak of acetylation over the 5′HS4 insulator element (as well as over the HSA enhancer of the FR gene) in all cells we examined, including a DT40 lymphocyte line in which neither globin nor FR genes are active

(15). Recent evidence (16–20) implicates histone methylation as well in the regulation of expression: Methyl group modifications at histone H3 lysines 4 or 9 are associated with active or inactive chromatin, respectively. We again used ChIP methods to measure patterns of histone methylation over the same region (Fig. 7). We found a striking correlation between previously observed patterns of histone acetylation and lysine 4 methylation and anticorrelation between acetylation and methylation of lysine 9. The clearest example of such distinct regions of modification is found in the mating type locus of *Schizosaccharomyces pombe*, where the heterochromatic and euchromatic domains of the mating type locus are distinguished by similar patterns of methylation and acetylation (18).

Work in a number of laboratories has led to models for propagation of the condensed chromatin state (15, 17, 18). These
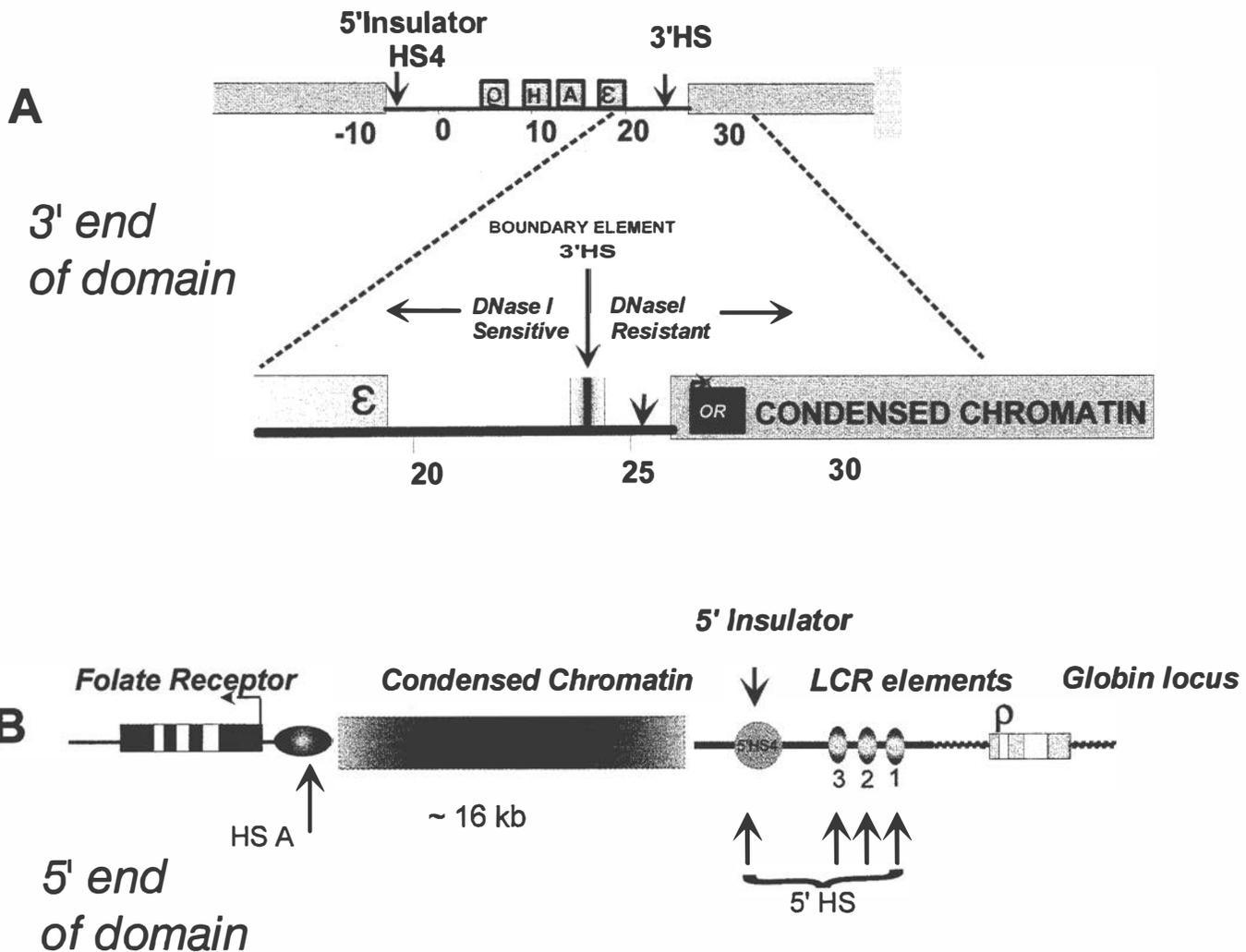
Fig. 4. The ends of the open β-globin domain. (A) 3'end. A second constitutive HS that binds CTCF and has enhancer blocking activity is found upstream of the beginning of a condensed chromatin region containing a gene for an odorant receptor (8). (B) 5' end. A condensed chromatin region extends for about 16 kb upstream of 5'HS4, and beyond that is an erythroid-specific FR gene (9).

are based on the following observations: (*i*) the heterochromatin protein HP1 [or in *S. pombe* its homolog Swi6 (18)] binds selectively to histone H3 that is methylated at lysine 9 in the amino terminal tail, and (*ii*) the enzyme responsible for methylating lysine 9, Suv39H1, interacts with HP1. This finding leads to the proposal (15, 17, 18) that a nucleosome methylated at that site will indirectly recruit Suv39H1 and promote methylation of sites on the adjacent nucleosome. Our observation that the 5'HS4 insulator is a major center of histone acetylation suggests that one of its roles at the 5' end of the β-globin locus is to continually acetylate the adjacent upstream nucleosome, in particular lysine 9 of H3. By doing so, it prevents methylation of that residue and thus terminates the propagation of the condensation signal. A related mechanism has been proposed for barrier activity in *Saccharomyces cerevisiae*, where a site that binds histone acetylases is sufficient to prevent extension of silencing from HMR-E (21).

These results suggest that *in vivo* the 5'HS4 insulator has two functions: it serves as an enhancer blocking element to screen out upstream signals, and it also acts as a barrier against the advance of the condensed chromatin structure immediately upstream. In fact, recent results indicate that these two functions are separable (22). The enhancer blocking function, as discussed
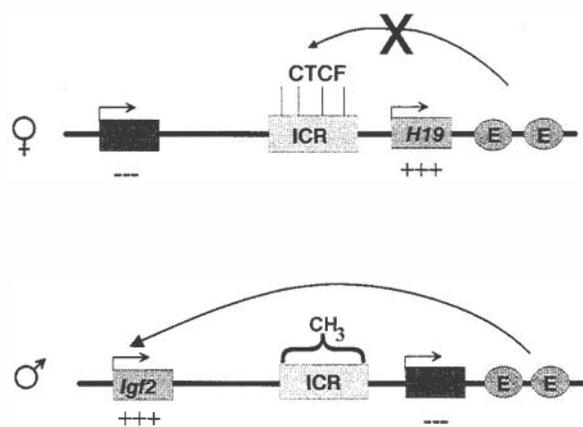

Fig. 5. The mouse *Igf2/H19* imprinted locus. In the maternally transmitted allele, *Igf2* is silent, but in the paternal allele it is expressed, and the ICR is methylated. The ICR has been shown to contain four CTCF binding sites, which have strong enhancer blocking properties; enhancer blocking is abolished by methylation of cytosines at CpG sites within the ICR. These and other results lead to a model in which the maternal ICR blocks the action of downstream endodermal enhancers (E) on the *Igf2* promoter. Methylation of the paternal ICR abolishes enhancer blocking and permits *Igf2* activation (10–12).
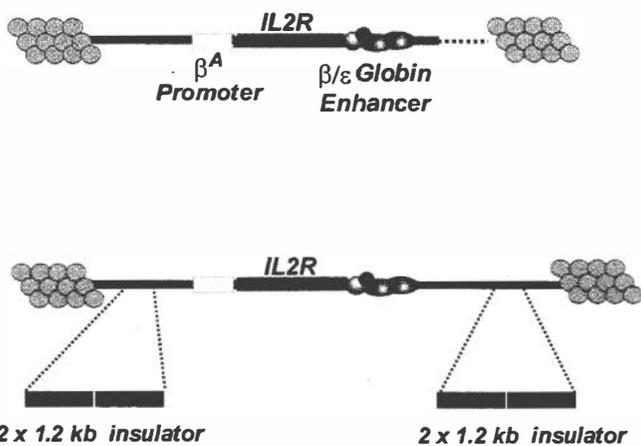
Fig. 6. Constructions to measure protection against position effects (barrier function). A reporter containing a fragment of the IL-2R driven by the chicken adult β-globin promoter and the downstream β/ε enhancer (see Fig. 2) is stably transformed into the avian erythroid line 6C2. Expression of IL-2R on the cell surface is monitored by FACS analysis. (*Upper*) The control construction. (*Lower*) The control reporter is surrounded by two copies of the 1.2-kb chicken 5′HS4 insulator on each side. In most lines transformed with the control construction, IL-2R expression was extinguished after 80–100 days in culture. Almost all lines carrying the insulated construction still expressed IL-2R after 80–100 days (14).

Fig. 7. Chromatin immunoprecipitation of modified histones across the β-globin locus. (*Top*) Diacetylated histone H3 and tetraacetylated histone H4 in 6C2 cells, arrested at the CFU-E stage of chicken erythroid development. (*Middle*) Map of the locus showing positions of HSs above the line and of probes used in PCR detection below the line. (*Bottom*) Patterns of histone H3 methylation at lysines 4 and 9 and diacetylated (lysines 9 and 14) histone H3 across the locus in 10-day embryonic chicken erythrocytes. (Adapted from refs. 15 and 16.)

above, depends on the CTCF site, whereas the barrier function depends on the other four subregions of the 250-bp insulator core and does not require CTCF. We note that the CTCF site at the 3′ end of the globin locus is not accompanied by the other subregions present in 5′HS4 and is also not a site of strong acetylation. The condensed chromatin at the 3′ end of the locus is facultative, i.e., it must open for expression of the odorant receptor, and it may therefore be distinct in properties from the constitutive condensed chromatin at the 5′ end and may not require an acetylated barrier.

How widely are insulators distributed in the genome? A considerable number of elements have been identified in *Drosophila*, each with its own characteristic site and associated proteins (for a recent general review of insulators see ref. 23). Barrier functions have been identified in both *S. pombe* and *S. cerevisiae* (20, 21). In vertebrates, the CTCF site appears to be widely distributed and to function in many cases as an enhancer blocking agent. However, no other barrier (position effect) elements have yet been identified similar to that found at the β-globin locus. Evidently many genes will not require insulators
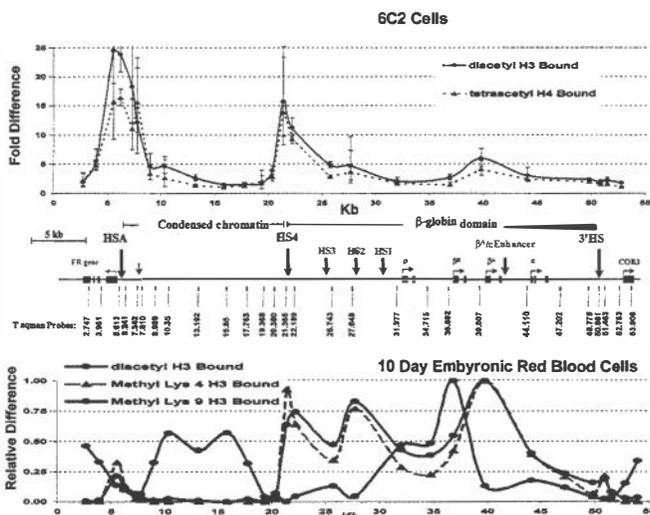
as a protection against inappropriate interaction of neighboring signals, but it seems reasonable to look for them in cases (such as the Ig and T cell receptor gene loci) where multiple regulatory elements are clustered. As the results in *Drosophila* and yeast have shown, diverse proteins and sites may be involved. There is no reason to think that vertebrates will be less complex, so it is likely that many novel insulators remain to be identified.

Conclusion

Insulator elements are clearly an important family of regulatory sequences, likely to be distributed widely in the genome. The connection of enhancer blocking activity with imprinted loci, and tentatively with the X-inactivation locus (13), suggests a role in the establishment of epigenetic imprinting marks. The barrier function is connected with the maintenance of boundaries that may also be established through epigenetic signals such as histone methylation. We are still not completely certain of detailed mechanisms of insulator action, but as we learn more, we will also understand more about how the cell controls and exploits epigenetic signals.

1. Reitman, M. & Felsenfeld, G. (1990) *Mol. Cell. Biol.* **10,** 2774–2786.
2. Chung, J. H., Whiteley, M. & Felsenfeld, G. (1993) *Cell* **74,** 505–514.
3. Chung, J. H., Bell, A. C. & Felsenfeld, G. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 575–580.
4. Hebbes, T. R., Clayton, A. L., Thorne, A. W. & Crane-Robinson, C. (1994) *EMBO J.* **13,** 1823–1830.
5. Bell, A. C., West, A. G. & Felsenfeld, G. (1999) *Cell* **98,** 387–396.
6. Saitoh, N., Bell, A. C., Recillas-Targa, F., West, A. G., Simpson, M., Pikaart, M. & Felsenfeld, G. (2000) *EMBO J.* **19,** 2315–2322.
7. Bulger, M., van Doorninck, J. H., Saitoh, N., Telling, A., Farrell, C., Bender, M. A., Felsenfeld, G., Axel, R., Groudine, M. & von Doorninck, J. H. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5129–5134.
8. Farrell, C. M., West, A. G. & Felsenfeld, G. (2002) *Mol. Cell. Biol.* **22,** 3820–3831.
9. Prioleau, M. N., Nony, P., Simpson, M. & Felsenfeld, G. (1999) *EMBO J.* **18,** 4035–4048.
10. Bell, A. C. & Felsenfeld, G. (2000) *Nature (London)* **405,** 482–485.
11. Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M. & Tilghman, S. M. (2000) *Nature (London)* **405,** 486–489.
12. Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C. F., Wolffe, A., Ohlsson, R. & Lobanenkov, V. V. (2000) *Curr. Biol.* **10,** 853–856.
13. Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S. & Lee, J. T. (2002) *Science* **295,** 345–347.
14. Pikaart, M. J., Recillas-Targa, F. & Felsenfeld, G. (1998) *Genes Dev.* **12,** 2852–2862.
15. Litt, M. D., Simpson, M., Recillas-Targa, F., Prioleau, M. N. & Felsenfeld, G. (2001) *EMBO J.* **20,** 2224–2235.
16. Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D. & Felsenfeld, G. (2001) *Science* **293,** 2453–2455.
17. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. (2001) *Nature (London)* **410,** 116–120.
18. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., *et al.* (2000) *Nature (London)* **406,** 593–599.
19. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. & Kouzarides, T. (2001) *Nature (London)* **410,** 120–124.
20. Noma, K., Allis, C. D. & Grewal, S. I. (2001) *Science* **293,** 1150–1155.
21. Donze, D. & Kamakaka, R. T. (2001) *EMBO J.* **20,** 520–531.
22. Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M., Gaszner, M. & Felsenfeld, G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6883–6888.
23. West, A. G., Gaszner, M. & Felsenfeld, G. (2002) *Genes Dev.* **16,** 271–288.

Colloquium

Histone H3 lysine 4 methylation is mediated by Set1 and promotes maintenance of active chromatin states in fission yeast

Ken-ichi Noma and Shiv I. S. Grewal*

Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724

Methylation of histone H3 at lysine 4 (H3 Lys-4) or lysine 9 (H3 Lys-9) is known to define active and silent chromosomal domains respectively from fission yeast to humans. However, in budding yeast, H3 Lys-4 methylation is also necessary for silent chromatin assembly at telomeres and ribosomal DNA. Here we demonstrate that deletion of *set1*, which encodes a protein containing an RNA recognition motif at its amino terminus and a SET domain at the carboxy terminus, abolishes H3 Lys-4 methylation in fission yeast. Unlike in budding yeast, Set1-mediated H3 Lys-4 methylation is not required for heterochromatin assembly at the silent mating-type region and centromeres in fission yeast. Our analysis suggests that H3 Lys-4 methylation is a stable histone modification present throughout the cell cycle, including mitosis. The loss of H3 Lys-4 methylation in *set1Δ* cells is correlated with a decrease in histone H3 acetylation levels, suggesting a mechanistic link between H3 Lys-4 methylation and acetylation of the H3 tail. We suggest that methylation of H3 Lys-4 primarily acts in the maintenance of transcriptionally poised euchromatic domains, and that this modification is dispensable for heterochromatin formation in fission yeast, which instead utilizes H3 Lys-9 methylation.

Dynamic changes in chromatin structure are directly influenced by the posttranslational modification of histones. Specific amino acids on histone amino-terminal tails that extend outward from nucleosome core particle are the targets of a number of modifications including acetylation, phosphorylation, ubiquitylation, ADP ribosylation, and methylation (1–5). The presence of a specific pattern of histone modifications has been linked to various chromosomal processes, such as the maintenance of gene expression patterns during development, recombination, chromosome condensation, and the proper segregation of chromosomes during mitosis. For example, hyperacetylated regions of chromatin contain active transcription units, whereas hypoacetylated chromatin is transcriptionally silent (6).

How histone modifications participate in the modulation of chromatin structure is not fully understood. It is believed that different combinations of histone modifications can determine the binding affinities of histone-interacting proteins whose chromosomal associations lead to discrete downstream events (2, 7). This is best illustrated by recent studies showing that modifications of the H3 tail by deacetylase and methyltransferase activities likely act in concert to establish the "histone code" essential for heterochromatin assembly (3, 4, 8). The chromodomains of heterochromatin proteins Swi6 and HP1 from fission yeast and *Drosophila*, respectively, are specific interaction motifs for the histone H3 amino-terminal tail modified by methylation on lysine 9 (9–11), and localization of these proteins to heterochromatic loci depends on H3 Lys-9 methylation (8, 12). Similarly, the bromodomain of many transcriptional coactivators binds specifically to the acetylated lysine residues on histone tails (13). Applying this general concept to other chromatin modulators, it is likely that unique combinations

of histone modifications serve as marks for the recruitment of different chromatin proteins or protein complexes to initiate the formation of defined chromosomal subdomains with distinct functions.

The methylation of H3 and H4 tails can occur on both arginine and lysine residues, and plays a critical role in transcriptional regulation (3, 5). A specific class of methyltransferases including CARM1 and PRMT1, which act as transcriptional coactivators, catalyze histone arginine methylation (14–16). However, enzymes that contain an evolutionary conserved SET domain, a 130-residue motif originally identified in Su(var)3–9, Enhancer-of-zeste and Trithorax proteins in *Drosophila* (17), are implicated in the lysine methylation of histones (5). The founding members of this class of histone methyltransferases are mammalian SUV39H1 and its fission yeast homolog Clr4, which specifically methylate H3 lysine 9 (H3 Lys-9) at heterochromatic loci (8, 18). Recently, several SET domain-containing proteins, such as G9a, SETDB1, and ESET, have been shown to methylate H3 Lys-9 in mammals (19–21).

In contrast to the H3 Lys-9 methylation that defines silent chromosomal domains, H3 Lys-4 methylation is specific to transcriptionally poised euchromatic regions in fission yeast, *Tetrahymena*, chicken, and mammals (22–26). Moreover, high-resolution mapping using chromatin immunoprecipitation (ChIP) has shown that distinct patterns of histone methylation marking heterochromatic and euchromatic domains are separated by boundary elements that protect against spreading of repressive chromatin into neighboring areas (22, 23). H3 Lys-4 is methylated by SET domain containing proteins—Set1 in *Saccharomyces cerevisiae* (27–29) and Set7/Set9 in mammals (30, 31). Although the precise function of H3 Lys-4 methylation is not known, it has been hypothesized to facilitate transcription by serving as a mark to recruit necessary transcription machinery components, or alternatively to protect euchromatic regions from the repressive effects of neighboring silent chromatin complexes (22, 23).

Interestingly, the deletion of *SET1* in *S. cerevisiae* leads to defects in silencing at telomeres, ribosomal DNA (rDNA), and the mating-type region (27, 32, 33). It has been suggested that H3 Lys-4 methylation in the context of other histone modifications might have a dual function both in transcriptional activation and silencing (27, 33). However, it remains to be investigated whether H3 Lys-4 methylation by Set1 homologs is also required for silent chromatin assembly in other species.
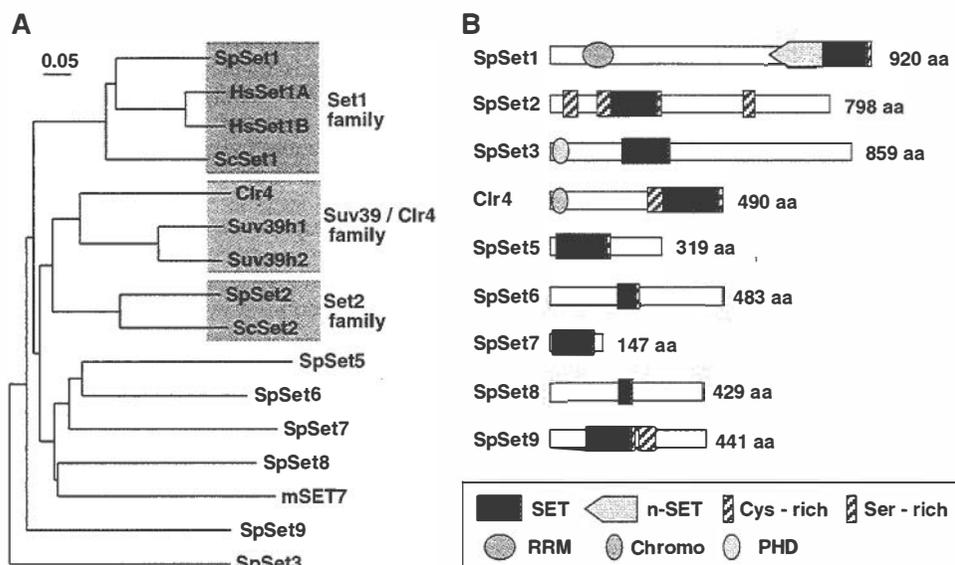
Fig. 1. SET domain proteins in *Sch. pombe* genome. (*A*) Phylogenetic tree of the SET domain proteins. The tree was constructed by the neighbor-joining method based on the amino acid sequences in the SET domains. The scale bar equals a distance of 0.05 aa. The SpSet (*Sch. pombe* SET) members (GenBank accession nos.: SpSet1, AL049728; SpSet2, Z99164; SpSet3, Z70043; SpSet5, AL031540; SpSet6, AL032684; SpSet7, AL049609; SpSet8, Z99568; SpSet9, AL132870) were identified by BLAST using previously identified SET domains as query. Abbreviations: Hs, *Homo sapiens*; m, *Mus musculus*. (*B*) Schematic representation of *Sch. pombe* SET domain proteins. The length of each protein (in aa) is noted on the right. Conserved domains are indicated as follows: RRM, RNA recognition motif; Chromo, Chromodomain; PHD, PHD finger; Cys-rich, Cysteine-rich domain; Ser-rich, Serine-rich domain.

In this study, we report identification and structural features of at least nine SET domain proteins present in the fission yeast genome. We show that one of these SET domain proteins that we named Set1 appears to be the exclusive H3 Lys-4-specific methyltransferase. Unlike in *S. cerevisiae*, Set1 is not required for heterochromatin assembly in fission yeast, consistent with our previous results showing that H3 Lys-4 methylation is specific to euchromatic regions. H3 Lys-4-methylaton is present throughout the cell cycle and important for the upkeep of transcriptionally poised domains in euchromatic regions, perhaps through the maintenance of histone acetylation levels.

Materials and Methods

Sequence Analysis. Database searches were performed with BLASTP. Multiple amino acid sequences were aligned by using CLUSTALW, Version 1.7. A phylogenetic tree was created by the neighbor-joining method based on the amino acid sequence alignment. The domains were characterized by using the MOTIFSCAN program against the PROSITE database. The resulting dendrogram showing the relationship between different SET domain proteins is shown in Fig. 1*A*, and structural features of different SET domain proteins in fission yeast are shown in Fig. 1*B*.

Strains. The genotypes of the *Schizosaccharomyces pombe* strains used in this study are listed in Table 1. The *set1Δ* strain was constructed by a PCR based method as described (34). Deletion was confirmed by PCR and Southern analysis. A strain containing deletion of a part of the *set1* ORF encoding RNA recognition motif (RRM) was constructed as follows: DNA fragments from upstream (0.9 kb) and downstream (2.6 kb) of RRM encoding region of *set1* were amplified by PCR with *Pfu*Turbo DNA polymerase (Strat-

agene) and cloned in frame into pCRII-TOPO (Invitrogen) to construct *set1* ORF minus RRM encoding region (*set1ΔRRM*). The resulting 3.5-kb fragment was gel-purified and used for transformation of SPK10 strain carrying *set1Δ::kanMX6* allele. Transformants were screened for G418 sensitivity, and colonies carrying *set1ΔRRM* were confirmed by using PCR analysis. Standard genetic crosses were used to construct all other strains.

Iodine Staining Assay. Efficiency of mating-type switching was analyzed by the iodine-staining assay. Individual colonies were replicated onto sporulation (PMA+) medium and then grown for 3 days at 26°C before being exposed to iodine vapors. The dark staining indicates efficient mating-type switching, which requires heterochromatin-mediated chromatin organization at the mating-type region (35). Defects in heterochromatin assembly at the mating-type region results in inefficient switching, causing a decrease in iodine staining.

Western Analysis of Histones. For isolation of bulk histones, fission yeast cells were grown to mid-log phase. Cells (5×10^8) were washed with 10 ml of NIB buffer (0.25 M sucrose/60 mM KCl/15 mM NaCl/5 mM MgCl$_2$/1 mM CaCl$_2$/15 mM Pipes, pH 6.8/0.8% Triton X-100) and resuspended in 500 μl of NIB buffer containing 10 ng/μl TSA, 2 mM ZnSO$_4$, Complete protease inhibitor mixture (Roche, 1 tablet per 10 ml), and 1 mM PMSF. Cells were disrupted by acid-washed glass beads (425–600 μM) using a minibeadsbeater (Biospec) for 4–5 min. Cell extracts were centrifuged at 11,000 × *g* for 10 min. The pellets were resuspended in 0.4 M H$_2$SO$_4$ and incubated on ice for 1 h with occasional mixing. The supernatant was collected by centrifugation at 8,000 × *g* for 5 min. The H$_2$SO$_4$ extraction was repeated. Pooled supernatants were trichloroacetic

Table 1. *Sch. pombe* strains used in this study

Strain	*ura4+* insertion	Genotype
FY498	*imr1R::ura4+*	*h+ leu1-32 ura4DS/E ade6-210*
FY648	*otr1R::ura4+*	*h+ leu1-32 ura4DS/E ade6-210*
PG925	*mat3M::ura4+*	*h90 leu1-32 ura4D18 ade6-210 clr3-735*
SP1464	*KΔ::ura4+*	*h90 leu1-32 ura4D18 ade6-210 clr6-1*
SPG1236	*Kint2::ura4+*	*h90 leu1-32 ura4DS/E ade6-216 his2*
SPK10	*Kint2::ura4+*	*h90 leu1-32 ura4DS/E ade6-216 his2 set1Δ::kanMX6+*
SPK108	*imr1R::ura4+*	*h+ leu1-32 ura4DS/E ade6-216 set1Δ::kanMX6+*
SPK111	*otr1R::ura4+*	*h+ leu1-32 ura4DS/E ade6-210 set1Δ::kanMX6+*
SPK121	*Kint2::ura4+*	*h90 leu1-32 ura4DS/E ade6-216 his2 set1ΔRRM*

acid precipitated and pellets were washed twice in 500 μl of cold acetone, air-dried and resuspended in 10 mM Tris·HCl, pH 8.0. Bulk histone samples were kept in $-70°C$ freezer until use. Ten micrograms of crude histone samples were resolved on an SDS/18% PAGE, transferred to a poly(vinylidene difluoride) membrane, and probed with site-specific acetyl- or methyl-histone antibodies. Antibodies to H3 Lys-4-methyl, H3 Lys-9-acetyl, and H3 Lys-14-acetyl were purchased from Upstate Biotechnology, whereas antibodies specific to acetylated Lys-5, Lys-8, Lys-12, or Lys-16 of histone H4 were purchased from Serotec. The band intensities were quantified with NIH IMAGE 1.62 software.

Immunofluorescence Analysis. Cells were grown to mid-log phase in yeast-extract adenine (YEA) medium. An equal volume of YEA + 2.4 M sorbitol was added and the culture was incubated further at 18°C for 5 min. For using α-H3 Lys-4-methyl, cells were fixed by adding paraformaldehyde to a final concentration of 1.7% and incubated at 18°C for 45 min. For staining by α-H3 Ser-10-phospho, cells were fixed in 3.0% paraformaldehyde for 30 min. Fixed cells were treated with Zymolyase to permeabilize the cell wall and then incubated overnight with primary antibodies, such as mouse α-tubulin TAT1 (1:150 dilution) or α-Nop1 (1:1,000 dilution) and rabbit α-H3 Lys-4-methyl (1:1,500) or α-H3 Ser-10-phospho (1:500). After extensive washing, cells were incubated for 6–8 h with Alexa Fluor 594 anti-rabbit IgG and Oregon Green 488 anti-mouse IgG (Molecular Probes) at a 1:2,000 dilution. After washing, cells were stained with 4′,6-diamidino-2-phenylindole (DAPI), mounted in Vectashield mounting medium (Vector Laboratories), and analyzed by a Zeiss Axioplan 2 fluorescence microscope.

ChIP. ChIP analysis was performed as described (36, 37). Fission yeast cells grown at 32°C in YEA (5×10^8 cells at 1×10^7 cells per ml for each reaction) were shifted to 18°C for 2 h before 30-min fixation in 3% paraformaldehyde. Soluble chromatin fractions prepared from fixed cells were sheared to ≈0.5- to 0.8-kb DNA fragments by sonication before immunoprecipitating by using antibodies to H3 Lys-4-methyl, H3 Lys-9-methyl, H3 Lys-14-acetyl, and Swi6. DNA fragments recovered from immunoprecipitated chromatin fractions or from whole cell crude extracts were subjected to PCR analyses (94°C for 30 s , 55°C for 30 s, 72°C for 1 min, 30 cycles). PCR products were labeled by including 0.25 μl of [α-^{32}P]deoxycytidine triphosphate (10 mCi/ml; 1 Ci = 37 GBq) in each reaction. PCR products were separated on a 4% polyacrylamide gel, and band intensities were quantified by using a Fuji PhosphoImager.

Results

SET Domain Proteins in *Sch. pombe* and Their Structural Features. A database search of the *Sch. pombe* genome (http://www.sanger.ac.uk/Projects/S_pombe/) was performed to identify the total number of SET domain proteins. The *Sch. pombe* genome has been sequenced, and contains 4,824 genes (38). A BLAST search with the SET domain of the Clr4 histone methyltransferase revealed that at least nine SET domain proteins, including Clr4, reside in the fission yeast genome (Fig. 1A). All of the *Sch. pombe* SET domain proteins contain a conserved NHSC motif, which, when mutated, has been shown to abolish histone methyltransferase activity (18). SpSet1, SpSet2, and SpSet3 were named after their *S. cerevisiae* counterparts based on homology. However, Clr4, SpSet5, SpSet6, SpSet7, SpSet8, and SpSet9 did not share significant similarities to *S. cerevisiae* proteins outside of their SET domain. The absence of a Clr4 homolog in *S. cerevisiae* is consistent with the fact that bulk histones isolated from budding yeast cells lack detectable levels of H3 Lys-9 methylation (24, 27).

We next carried out phylogenetic analysis to assess the relationships between different SET domain proteins (Fig. 1A). The SET domain sequences of *Sch. pombe* proteins and previously characterized proteins from other species were aligned together and a

phylogenetic tree was constructed. Based on the dendrogram (Fig. 1A), *Sch. pombe* SET domain proteins were not highly related to one another. As expected, SpSet1 and Clr4 proteins cluster with the previously described Set1 and SUV39 family of histone methyltransferases that have the capacity to methylate H3 Lys-4 and H3 Lys-9, respectively (5). The Set1 proteins from budding and fission yeast share 26% identity throughout their length and 63% in the SET domain region. Our analysis suggests that SpSet1 is more closely related to Set1 in *Homo sapiens* than it is to Set1 in *S. cerevisiae*. Another SET domain protein, SpSet2, is closely related to a H3 Lys-36-specific methyltransferase from *S. cerevisiae* (39).

Careful examination of the SET domains in *Sch. pombe* proteins revealed distinct structural features. As reported previously, the SET domain in Clr4 is surrounded by two cysteine-rich regions, referred to as preSET and postSET, that are essential for its catalytic activity (8). The preSET and postSET domains are also present in SpSet2, but only the postSET domain is present in SpSet1, SpSet5, SpSet6, and SpSet9 (Fig. 1B). SpSet1 contains a highly conserved 160-aa motif called n-SET at the preSET location. The SET domains of SpSet3, SpSet7, and SpSet8 do not have pre- or postSET motifs. It is possible that differences in structural features surrounding the SET domains might account for the altered substrate specificity of these proteins.

Further analysis using the MOTIFSCAN program revealed that in addition to the presence of a chromodomain in Clr4, as reported (40), SpSet1 and SpSet3 contain an RRM and PHD finger at their amino-terminal regions, respectively. Both the PHD finger that is involved in protein–protein interactions, and the RRM, which is known to interact with both RNA and proteins, are conserved motifs shown to be present in subunits of chromatin-modifying activities (41, 42).

***Sch. pombe* Set1 Is Required for H3 Lys-4 Methylation *in Vivo*.** As mentioned above, Set1 mediates H3 Lys-4 methylation in *S. cerevisiae* (27–29). Considering that Set1 proteins from budding and fission yeasts share considerable homology within and outside of their SET domains, it was possible that SpSet1 might also be involved in H3 Lys-4 methylation. To test this possibility, we constructed a strain containing a complete deletion of *set1* gene, replacing the entire ORF with the *KANMX6* gene. The resulting *set1Δ* mutant was viable, suggesting that SpSet1 is dispensable for cell growth.

To study the biological effects of *set1Δ* on H3 Lys-4 methylation, we performed immunofluorescence analysis using an antiserum specific to methyllysine 4 of histone H3 (22, 27). In wild-type cells, immunofluorescence signal corresponding to H3 Lys-4-methyl was preferentially enriched at the chromatin in the DAPI-stained areas but seemed to be excluded from the nucleolus containing rDNA repeats, as indicated by the staining of nucleolar marker protein Nop1 (Fig. 2A; ref. 43). In comparison to a high level of H3 Lys-4 methylation in the nuclei of wild-type cells, strikingly, methylation of histone H3 at lysine 4 was completely abolished in *set1Δ* cells (Fig. 2C), suggesting that SpSet1 is responsible for H3 Lys-4 methylation. This result was further confirmed by Western analysis of bulk histones prepared from wild-type and *set1Δ* cells by using H3 Lys-4-methyl-specific antibodies. Histones prepared from wild-type cells exhibit high levels of methylated H3 Lys-4, but we could not detect any signal in the *set1Δ* cells (Fig. 2B). These analyses suggest that Set1 mediates H3 Lys-4 methyaltion in *Sch. pombe*. Although it remains a possibility that low levels of H3 Lys-4 methylation exist in the nucleolus, histone H3 methylated at lysine 4 is clearly not enriched in the nucleolar compartment.

Previous studies in *S. cerevisiae* have suggested that H3 Lys-4 methylation is important for normal growth, and that mutations in *SET1* can result in a number of phenotypes including morphological abnormalities, perturbed DNA distribution, growth and sporulation defects (27, 32). We therefore investigated whether *set1Δ* cells display similar phenotypes in *Sch. pombe*. We noticed that colonies
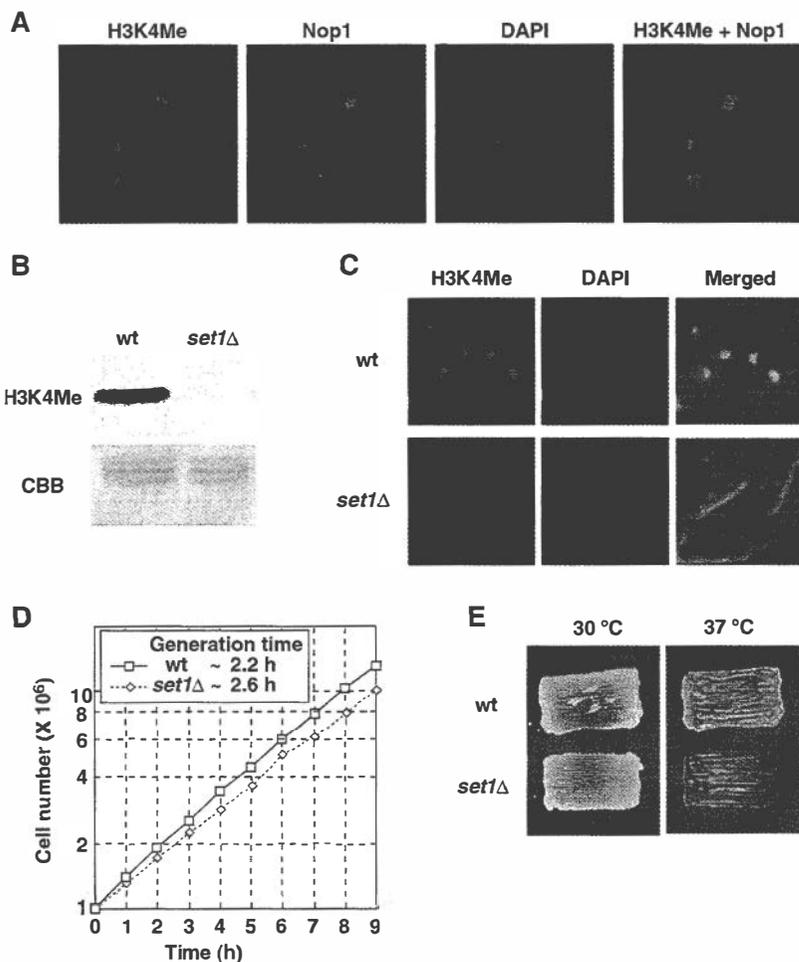
A
H3K4Me | Nop1 | DAPI | H3K4Me + Nop1

B
wt | set1Δ
H3K4Me
CBB

C
H3K4Me | DAPI | Merged
wt
set1Δ

D

Generation time
□ wt ~ 2.2 h
◇ set1Δ ~ 2.6 h

Cell number (X 10⁶)
Time (h)

E
30 °C | 37 °C
wt
set1Δ

Fig. 2. Deletion of *set1* abolishes H3 Lys-4 methylation. (*A*) Wild-type cells (SPG1236) were stained with anti-H3 Lys-4-methyl (H3K4Me) antibody (red), anti-Nop1 antibody (green), and DAPI (blue). (*B*) Western blot with H3K4Me antibody against crude bulk histones prepared from wild-type (SPG1236) and *set1Δ* (SPK10) strains. Identical samples were examined in parallel by Coomassie staining to show histone loading. (*C*) Wild-type and *set1Δ* (SPK10) cells were stained with anti-H3K4Me and DAPI. (*D*) Comparison in growth rate of wild-type (SPG1236) and *set1Δ* (SPK10) strain. Wild-type and mutant strains were grown in YEA-rich medium at 30°C. Cell numbers at fixed time points are plotted on a semilogarithmic graph. (*E*) Temperature sensitivity of *set1Δ* strain. The strains previously grown under permissive growth conditions (30°C) were replicated onto YEA plates and incubated overnight at 30°C or 37°C.

formed by *set1Δ* mutants are smaller in size when compared with their wild-type counterparts (data not shown). Furthermore, *set1Δ* cells have a slightly longer doubling time and exhibit temperature-sensitive growth defects (Fig. 2 *D* and *E*). However, deletion of *Sch. pombe set1* did not cause any obvious morphological abnormalities, sporulation defects, or abnormal DAPI staining patterns.

RRM of SpSet1 is Necessary for Its Role in H3 Lys-4 Methylation. As described above, the amino terminus of the SpSet1 contains a canonical RNA binding domain called RRM (Fig. 1*B*). Interestingly, this motif is also present in orthologs of SpSet1 in *S. cerevisiae*, *Caenorhabditis elegans*, *Drosophila*, and humans. We sought to investigate whether the RRM domain is essential for SpSet1-mediated H3 Lys-4 methylation. For this purpose, we constructed a strain in which a small part of the *set1* ORF, encoding the RRM, was deleted at its endogenous chromosomal location. The expression of *set1ΔRRM* is under the control of the native *set1* regulatory elements, so as to achieve wild-type levels of expression. Interestingly, Western analysis of bulk histones revealed that H3 Lys-4 methylation is severely defective in the *set1ΔRRM* strain as compared with wild-type (Fig. 3). Moreover, *set1ΔRRM* strain showed temperature-sensitive growth defects (data not shown). Although a possibility remains that deletion of RRM affects steady-state Set1 levels in the cells, we suggest that the RRM domain of SpSet1 might be required for its role in H3 Lys-4 methylation.

Deletion of *set1* Does Not Affect Silencing at the Mating-Type Region and Centromeres. Modifications of histone tails are known to play a critical role in heterochromatin assembly through their role in

dictating the interactions between nucleosome arrays and non-histone chromatin proteins (44). Set1-mediated H3 Lys-4 methylation has been shown to be required for transcriptional silencing of the silent mating-type loci, telomeres, and rDNA in *S. cerevisiae*. To investigate whether *set1* is also required for transcriptional silencing in *Sch. pombe*, we combined *set1Δ* with a *ura4+* marker gene
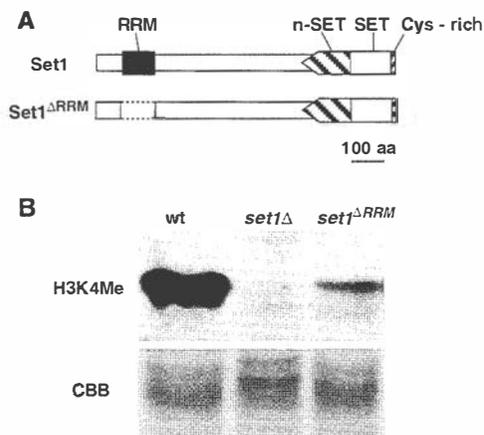

A
RRM | n-SET | SET | Cys - rich
Set1
Set1ΔRRM
100 aa

B
wt | set1Δ | set1ΔRRM
H3K4Me
CBB

Fig. 3. RRM of SpSet1 is required for its role in H3 Lys-4 methylation. (*A*) Schematic representation of the Set1 and Set1ΔRRM. (*B*) Western blot with H3K4Me antibody against crude bulk histones prepared from wild-type (SPG1236), *set1Δ* (SPK10), and *set1ΔRRM* (SPK121) strains. Coomassie brilliant blue (CBB) staining is shown as loading controls.
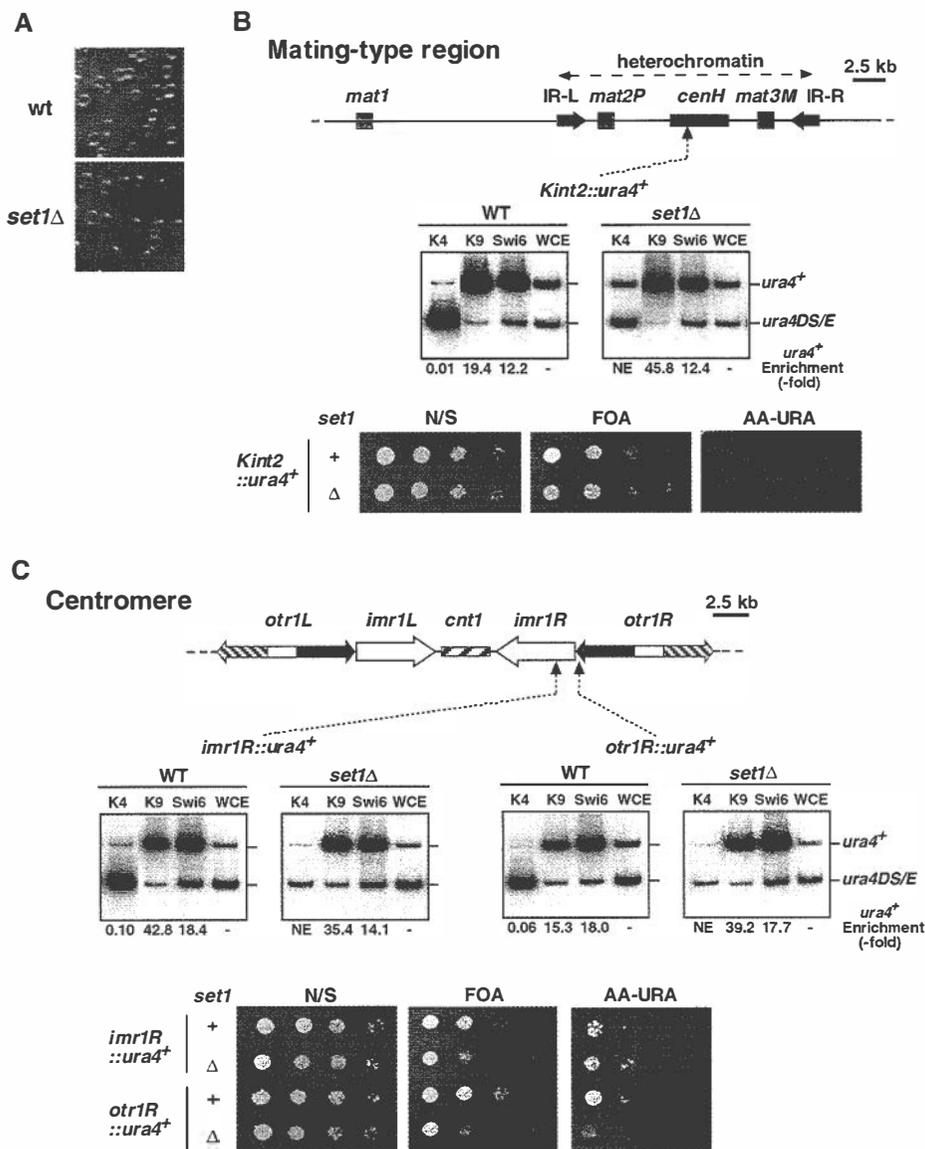
Fig. 4. Effects of *set1* deletion on silencing and heterochromatin assembly. (*A*) Iodine-staining assay. Wild-type and *set1*Δ colonies were sporulated on PMA⁺ medium at 25°C and exposed to iodine vapors before photography. (*B* and *C*) Deletion of *set1* does not affect silencing at the mating-type region and centromere. The strains that contain *ura4*⁺ genes inserted at the mating-type locus (*Kint2::ura4*⁺) or centromere region of chromosome 1 (*imr1R::ura4*⁺ and *otr1R::ura4*⁺) were used for ChIP assays with antibodies to H3 Lys-4-methyl (K4), H3 Lys-9-methyl (K9), or Swi6 protein. DNA from ChIP or WCE (whole cell crude extract) was analyzed by a competitive PCR strategy, whereby one set of primers amplifies different-sized products from the *ura4*⁺ marker gene and the control *ura4DS/E* minigene at the endogenous euchromatic location. The ratios of *ura4*⁺ and control *ura4DS/E* signals present in ChIP and WCE were used to calculate relative enrichment, shown beneath each lane. NE indicates no enrichment observed. The *ura4*⁺ expression levels at the mating-type and *cen1* region were evaluated by dilution analysis. Cells were suspended in water, and 10-fold serial dilutions were spotted onto nonselective (N/S), counterselective FOA, AA-URA medium, and grown 3 days before being photographed.

inserted at either the silent mating-type region (*Kint2::ura4*⁺) or at two different sites within *cen1*, one each at the outer (*otr1R::ura4*⁺) and inner (*imr1R::ura4*⁺) centromeric repeats that flank the central (*cnt*) domain (Fig. 4; ref. 45). Previous studies have shown that marker genes inserted within or adjacent to these heterochromatic locations are subject to transcriptional repression, as a consequence of repressive chromatin complexes spreading into the marker genes (46). Wild-type yeast cells carrying repressed *ura4*⁺ at a heterochromatic site cannot form colonies on medium lacking uracil (URA⁻) but grow efficiently on a counterselective medium containing 5′-fluoroorotic acid (FOA). However, cells defective in silencing grow on URA⁻ medium and are FOA sensitive. Dilution analysis of the wild-type and *set1*Δ strains carrying *ura4*⁺ marker gene at the silent mating-type region or at the centromeric locations revealed that SpSet1 is not required for silencing at these loci, as indicated by comparable growth of wild-type and *set1*Δ cells on URA⁻ and FOA media (Fig. 4 *B* and *C*).

We also tested whether *set1*Δ affects expression of the *KΔ::ura4*⁺ marker gene. The *KΔ::ura4*⁺ cells containing a substitution of part the interval between silent mating-type loci with the *ura4*⁺ marker gene exhibit variegated *ura4*⁺ expression (37). *ura4-off* and *ura4-on* epigenetic states are mitotically metastable. We used fluctuation

analysis to measure the effect of *set1*Δ on stability of the epigenetic states. The *set1*Δ had a subtle effect on the *ura4-off* to *ura4-on* transition (7.9 × 10⁻⁴ per cell division), as compared with wild-type cells (8.4 × 10⁻⁴ per cell division). Moreover, we observed that *set1*Δ caused a slight decrease (9.6 × 10⁻⁴ to 4.6 × 10⁻⁴ per cell division) in *ura4-on* to *ura4-off* conversion. This decrease in the "on" to "off" state conversion in *set1*Δ background could be caused by changes in the levels of *trans*-acting factors critical for establishment of the silenced state.

In addition to silencing, the efficiency of mating-type switching is also regulated by formation of a heterochromatic structure at the *mat* locus. Mutations that affect silencing and heterochromatin assembly at the mating-type region adversely affect mating-type interconversion (47, 48). We monitored the efficiency of mating-type switching in wild-type and *set1*Δ strains at the colony level by iodine-staining as described in the *Materials and Methods* section. We found that *set1*Δ has no effect on the efficiency of mating-type switching (Fig. 4*A*), consistent with our results that expression of the marker gene inserted at the silent mating-type region is not affected.

SpSet1 Is Dispensable for Heterochromatin Assembly. Silencing at the mating-type region and centromeres depends on Swi6 protein,

which is recruited to these loci through its binding to methylated H3 Lys-9 (46). We therefore explored whether $set1\Delta$ affects H3 Lys-9 methylation and/or Swi6 at heterochromatic loci. ChIP assays with antibodies to Swi6 or H3 Lys-9-methyl were used. DNA recovered from immunoprecipitated chromatin fractions was quantitated by using competitive PCR, whereby one primer pair amplifies 694-bp and 426-bp products from full-length $ura4^+$ inserted at the heterochromatic location and a mini-$ura4$ ($ura4DS/E$) at its endogenous euchromatic location. Our analysis revealed that the levels of H3Lys9 methylation and Swi6 at the silent mating-type region ($Kint2::ura4^+$) and centromeric repeats ($imr1R::ura4^+$ and $otr1R::ura4^+$) of $set1\Delta$ cells were comparable to their wild-type counterparts (Fig. 4 B and C). As expected, $set1\Delta$ abolished the preferential enrichment of H3 Lys-4 methylation at the euchromatic $ura4DS/E$ locus. Taken together, the results presented above suggest that Set1-mediated H3 Lys-4 methylation is not required for heterochromatin assembly in *Sch. pombe*.

H3 Lys-4 Methylation Is Present Throughout the Cell Cycle. In comparison to the acetylation of histones, lysine methylation is believed to be a relatively stable histone modification, which might serve as an epigenetic imprint for the long-term maintenance of chromatin states. If H3 Lys-4 methylation indeed serves as a molecular bookmark for inheritance of the active chromatin state, it is likely to be present throughout the cell cycle, even during mitosis. To address this issue, we performed immunofluorescence with antibodies to methylated H3 Lys-4 and tubulin used to visualize microtubules. As a control, we also studied H3 Ser-10 phosphorylation, which correlates with chromosome condensation during mitosis (49). At the G_2/M boundary, H3 Ser-10 phosphorylation was mainly localized to one or two discrete foci. However, Ser-10 phosphorylation spread throughout chromosomes by metaphase, the intensity of the signal diminished as cells enter anaphase, and almost all staining had disappeared in G_1/S cells (Fig. 5). In contrast to H3 Ser-10 phosphorylation, we found that H3 Lys-4 methylation levels remained unchanged throughout the cell cycle, including mitosis when chromosomes are highly condensed. Although changes in Lys-4 methylation at individual loci cannot be ruled out, H3 methylated at Lys-4 can be detected during different stages of the cell cycle. Because covalent modification by one enzyme can positively or negatively influence the efficiency of other enzymes responsible for modifying residues on the same histone tail, we also tested the effects of $set1\Delta$ on H3 Ser-10 phosphorylation. As shown in Fig. 5, the loss of H3 Lys-4 methylation in $set1\Delta$ strains did not affect H3 Ser-10 phosphorylation during mitosis.

Interplay Between H3 Lys-4 Methylation and Histone Acetylation *in Vivo*. It has been shown that Lys-4 methylation of H3 is preferentially associated with H3 acetylation in *S. cerevisiae*, chicken, and HeLa cells (23, 24). Here we examined whether SpSet1-mediated H3 Lys-4 methylation affects histone acetylation *in vivo*. Bulk histones prepared from wild-type and $set1\Delta$ were subjected to Western blot analysis with acetylation-site-specific H3 or H4 antibodies (Fig. 6). Interestingly, we found that the acetylation levels of histone tails, in particular of H3 Lys-9 and H3 Lys-14, were significantly decreased in $set1\Delta$ cells when compared with wild-type cells (Fig. 6). The $set1\Delta$ also results in a subtle but consistent decrease in H4 Lys-5 and H4 Lys-12 acetylation. We also investigated the possible effects of mutations in the histone deacetylases $clr3$ and $clr6$ on H3 Lys-4 methylation in bulk histones *in vivo*. As shown recently, $clr3$ specifically affects H3 Lys-14 acetylation, whereas mutation in $clr6$ results in elevated acetylation levels at all residues tested on the histone H3 and H4 tails (Fig. 6; ref. 50). Although H3 Lys-4 was slightly more methylated in $clr6$ mutant, mutation in $clr3$ had no effect on Lys-4 methylation. These data suggest that H3 Lys-4 methylation might help promote acetylation of histones in the transcriptionally poised regions of the chromosomes.
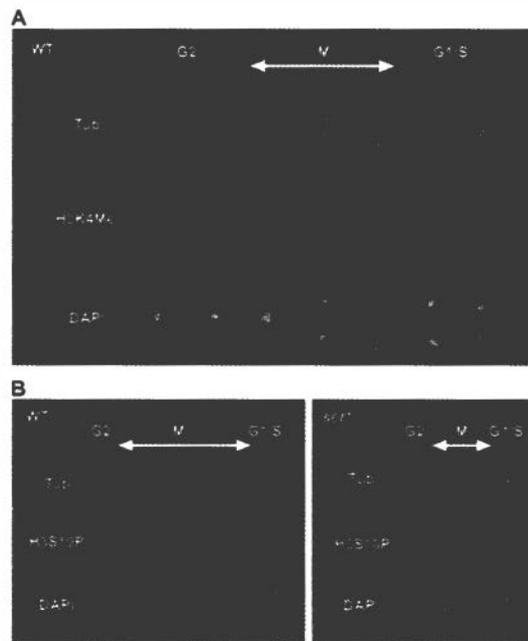

Fig. 5. H3 Lys-4 methylation is present throughout the cell cycle. (*A*) Wild-type cells were stained with anti-H3K4Me (red), anti-tubulin TAT-1 antibody (green), and DAPI (blue). For each cell, the corresponding panels are placed below each other. (*B*) Wild-type and $set1\Delta$ cells were stained with antibodies to phosphorylated H3 Ser-10 (H3S10P).

We also analyzed the effects of H3 Lys-4 methylation on H3 acetylation levels by ChIP assays. The loss of H3 Lys-4 methylation in $set1\Delta$ cells causes 50–60% reduction in H3 Lys-14 acetylation levels at the constitutively expressed loci $ura4$, $act1$, and $ade6$, as compared with wild-type background cells (Fig. 7). Consistent with decreased histone acetylation, we also observed that $set1\Delta$ causes subtle changes in $ade6$ expression. Because H3 Lys-4 methylation seems to globally affect active chromatin regions, changes in $ade6$ expression could not be quantified using Northern analysis because of the lack of appropriate controls. However, we observed that $set1\Delta$ cells carrying the $ade6-216$ allele at its endogenous chromosomal location formed deep red colonies on adenine-limiting medium, as compared with the pink colonies formed by their wild-type counterparts (Fig. 7). This phenotype indicating a decrease in $ade6$ expression consistently segregated with the $set1\Delta$ in more than 30 tetrads. Taken together, these analyses suggest that H3 Lys-4 methylation might be involved in the maintenance of active chromatin configurations, presumably by facilitating the acetylation of histones.

Discussion

Epigenetic control of higher-order chromatin assembly has been linked to the posttranslational covalent modifications of the histones tails. It has been formally suggested that distinct modifications on one or more of the histone tails act sequentially or in combination to form a histone code that is recognized by other chromatin-associated proteins (2, 7). Although histone-modifying enzymes such as acetyltransferases and deacetylases have been identified and characterized from a number of organisms, the factors that regulate methylation of histones are only now being discovered. Recent studies have identified a novel class of protein methyltransferases defined by an evolutionarily well-conserved structure, the SET domain (5, 18).

In this paper, we report the structural features of at least nine SET domain proteins present in fission yeast. The data presented demonstrate that a highly conserved SET domain protein named
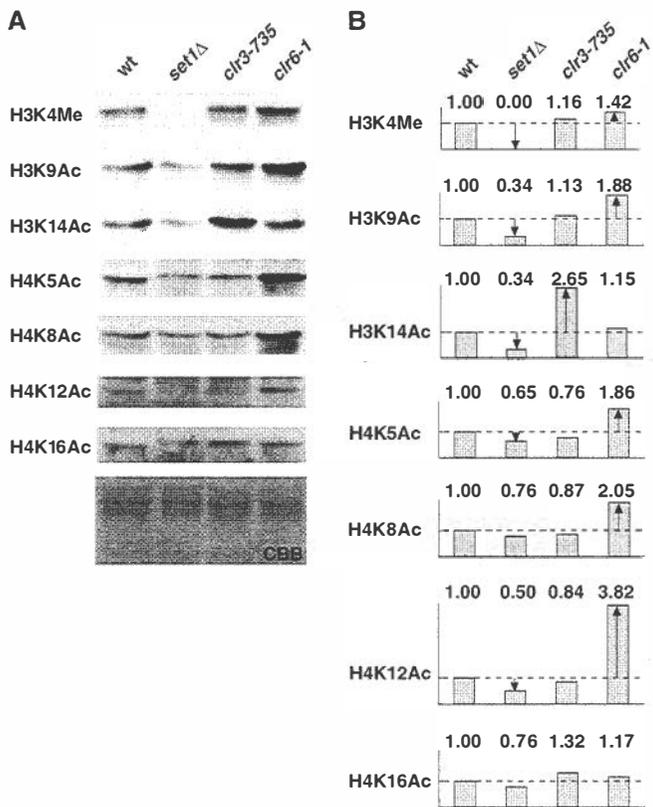
Fig. 6. *set1Δ* strain exhibits decreased levels of acetylation at lysine residues in H3 tail. (*A*) Bulk histones were prepared from wild-type, *set1Δ*, *clr3–735*, and *clr6-1* strains. Ten micrograms of crude histones were separated by SDS/PAGE and subjected to Western analysis with antibodies specific for H3K4me, or histone H3 or H4 acetylated at the indicated lysine residues. Coomassie staining (CBB) is shown as the loading control. Similar results were obtained in at least three independent experiments. (*B*) The intensity of the bands shown in *A*, quantified by using NIH IMAGE software, is summarized.

Fig. 7. *set1Δ* results in a decrease of H3 Lys-14 acetylation levels at genes. (*A*) ChIP analysis with antibodies to acetylated H3 Lys-14 (H3K14Ac). DNA fragments from the immunoprecipitated (IP) fraction and whole cell crude extract (W) were analyzed by using multiplex PCR. Relative enrichment of DNA fragments corresponding to the coding regions of *ura4*, *act1*, or *ade6* was examined. A DNA fragment from the silent mating-type region (*K*-region), which is known to lack H3 Lys-4-methyl and H3 Lys-14-acetyl modifications, was used as a control to normalize and calculate the relative enrichment of *ura4*, *act1*, and *ade6* sequences in IPed chromatin fractions. (*B*) Analysis of *ade6-216* phenotype. Wild-type and *set1Δ* cells were streaked onto adenine-limiting yeast extract medium and incubated at 30°C for 2–3 days. Deep red or pink color of colonies indicates *ade6-216* expression states. Colonies of *set1Δ* strains were deeper red than wild type, indicating decrease in *ade6* expression in mutant cells.

SpSet1 likely catalyzes the H3 Lys-4 methylation present at euchromatic regions. Based on genetic and biochemical studies, the SET domain and its flanking pre- and postSET domains are required for catalytic activity of SUV39H1 and Clr4 methyltransferases (8, 18). However, recent studies argue against the general requirement of the pre- and postSET domains for enzymatic activity. For example, human Set7/Set9 contains a SET domain but is devoid of the pre- and postSET motifs, and can efficiently methylate H3 Lys-4 *in vitro* and *in vivo* (30, 31), though it is possible that other sequences surrounding the SET domain might promote methyltransferase activity. The general consensus emerging is that the SET domain constitutes the catalytic motif, whereas flanking sequences might facilitate folding of the histone tails, providing specificity for a particular lysine residue. In this regard, the highly conserved n-SET motif and cysteine rich sequences flanking the SET domain in Set1 orthologs are likely to be critical for their specificity to H3 Lys-4. (Fig. 1; ref. 28). Based on conservation of the SET domain in yet uncharacterized SET domain proteins in fission yeast, it is probable that these proteins also act as methyltransferases.

In many cases, SET domain proteins also contain other conserved motifs such as the chromodomain, PHD finger, and RRM domain. For example, the SUV39 family of proteins including Clr4 is known to contain a chromodomain at their amino terminus (40). Although the SET domain of Clr4 and its surrounding sequences are sufficient for methyltransferase activity *in vitro*, both SET domains and chromodomains are required *in vivo* (8). We found that the RRM domain of *Sch. pombe* Set1 is required for its role in
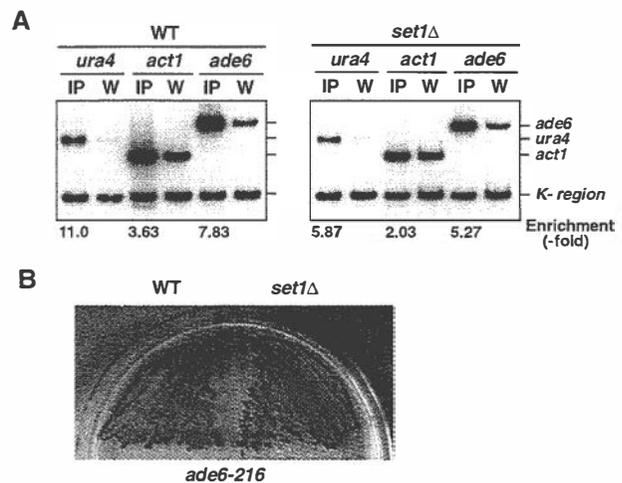
H3 Lys-4 methylation *in vivo*. The precise function of the RRM is not known, but it is possible that RRMs and chromodomains have related functions. These domains might help promote chromosomal targeting of their respective proteins, either through protein–protein interactions or through their binding to RNA. Supporting the possible role for RNAs in chromatin assembly, our recent work suggests that RNA interference (RNAi) mechanisms, through which small RNAs silence cognate genes, might be required for the targeting of histone-modifying activities to specific chromosomal domains in fission yeast. Specifically, RNAi machinery is essential for histone deacetylation and Clr4-mediated H3 Lys-9 methylation at centromeric repeats (I. Hall and S.I.S.G., unpublished data). Considering that certain chromodomains act as an RNA interaction module, it can be imagined that the binding of chromo- and/or RRM domains to RNA might guide the histone-modifying activities to homologous genomic sequences. In this scenario, RNA might provide specificity for the chromosomal targeting of these enzymes. Of course, it remains a possibility that chromodomains and RRM domains associated with these SET domain proteins are protein–protein interaction motifs. Future studies are necessary to address these possibilities.

Recent work suggests that H3 Lys-4 methylation can perform dual functions in *S. cerevisiae*. Interestingly, Set1-mediated H3 Lys-4 methylation at rDNA and telomeres in *S. cerevisiae* is required for transcriptional silencing (27, 32, 33). It has been suggested that H3 Lys-4 methylation in combination with other histone modifications could have a negative or positive effect on transcription at different chromosomal locations (27, 33). In this regard, our analysis suggests that *Sch. pombe* Set1 is not required for transcriptional silencing and heterochromatin assembly at centromeres or the silent mating-type interval. The differences in species of H3 Lys-4 that are mono-, di-, or tri-methylated might explain these seemingly contradictory results. A more likely explanation is that the histone code for silent chromatin assembly in *S. cerevisiae* and *Sch. pombe* is fundamentally different. H3 Lys-4 methylation is used in the process of silent

chromatin assembly in budding yeast, whereas H3 Lys-9 methylation is used for formation of repressive chromatin structures in fission yeast and higher eukaryotes. Consistent with this idea, H3 Lys-9 methylation has not been detected in *S. cerevisiae*, and homologs for neither the H3 Lys-9 methyltransferase nor the Swi6/HP1 protein that recognizes this modification are present (27).

The mechanism by which H3 Lys-4 methylation modulates chromatin structure is not clear. The budding yeast Set1 is found in a complex with homologs of the *Drosophila* protein Ash2 and Trithorax, which are essential for the stable maintenance of active gene expression states during development (28, 29). Furthermore, recent studies have shown that H3 Lys-4 methylation is preferentially associated with euchromatic regions and is correlated with H3 acetylation (22–25). Our analysis suggests that H3 Lys-4 methylation and acetylation at the H3 tail interact in *cis* at euchromatic regions *in vivo*. We found that *set1Δ* results in a decrease in H3 acetylation levels. Furthermore, we also observed that a mutation in the Clr6 histone deacetylase, which displays broad specificity to lysine residues at H3 and H4 tails (ref. 50 and this study), results in

a slight increase in H3 Lys-4 methylation. Based on these results, it is possible that H3 Lys-4 methylation either plays a facilitatory role in histone acetylation or it protects transcriptionally active regions from the effects of repressive chromatin remodelling activities such as histone deacetylases, or both. Supporting our *in vivo* analysis, it has been shown recently that H3 Lys-4 methylation facilitates subsequent acetylation of the histone tail by acetyltransferases *in vitro* (30). Furthermore, methylation of H3 Lys-4 interferes with interactions between the NuRD histone deacetylase and H3 tail, as well as precludes H3 Lys-9 methylation by SUV39H1 (31, 51). In conclusion, the results presented in this paper further extend the histone code hypothesis and suggest that Set1-mediated H3 Lys-4 methylation primarily acts in the maintenance of active chromatin configurations at euchromatic chromosomal domains in fission yeast.

1. van Holde, K. E. (1989) *Chromatin* (Springer, New York).
2. Strahl, B. D. & Allis, C. D. (2000) *Nature (London)* **403**, 41–45.
3. Zhang, Y. & Reinberg, D. (2001) *Genes Dev.* **15**, 2343–2360.
4. Berger, S. L. (2002) *Curr. Opin. Genet. Dev.* **12**, 142–148.
5. Kouzarides, T. (2002) *Curr. Opin. Genet. Dev.* **12**, 198–209.
6. Grunstein, M. (1997) *Nature (London)* **389**, 349–352.
7. Turner, B. M. (2000) *BioEssays* **22**, 836–845.
8. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. S. (2001) *Science* **292**, 110–113.
9. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. & Kouzarides, T. (2001) *Nature (London)* **410**, 120–124.
10. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. (2001) *Nature (London)* **410**, 116–120.
11. Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., Allis, C. D. & Khorasanizadeh, S. (2001) *EMBO J.* **20**, 5232–5241.
12. Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R. & Reuter, G. (2002) *EMBO J.* **21**, 1121–1131.
13. Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K. & Zhou, M. M. (1999) *Nature (London)* **399**, 491–496.
14. Ma, H., Baumann, C. T., Li, H., Strahl, B. D., Rice, R., Jelinek, M. A., Aswad, D. W., Allis, C. D., Hager, G. L. & Stallcup, M. R. (2001) *Curr. Biol.* **11**, 1981–1985.
15. Strahl, B. D., Briggs, S. D., Brame, C. J., Caldwell, J. A., Koh, S. S., Ma, H., Cook, R. G., Shabanowitz, J., Hunt, D. F., Stallcup, M. R. & Allis, C. D. (2001) *Curr. Biol.* **11**, 996–1000.
16. Wang, H., Huang, Z. Q., Xia, L., Feng, Q., Erdjument-Bromage, H., Strahl, B. D., Briggs, S. D., Allis, C. D., Wong, J., Tempst, P. & Zhang, Y. (2001) *Science* **293**, 853–857.
17. Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, G. & Reuter, G. (1994) *EMBO J.* **13**, 3822–3831.
18. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. & Jenuwein, T. (2000) *Nature (London)* **406**, 593–599.
19. Tachibana, M., Sugimoto, K., Fukushima, T. & Shinkai, Y. (2001) *J. Biol. Chem.* **276**, 25309–25317.
20. Yang, L., Xia, L., Wu, D. Y., Wang, H., Chansky, H. A., Schubach, W. H., Hickstein, D. D. & Zhang, Y. (2002) *Oncogene* **21**, 148–152.
21. Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G. & Rauscher, F. J., III. (2002) *Genes Dev.* **16**, 919–932.
22. Noma, K., Allis, C. D. & Grewal, S. I. (2001) *Science* **293**, 1150–1155.
23. Litt, M. D., Simpson, M., Gaszner, M., Allis, C. D. & Felsenfeld, G. (2001) *Science* **293**, 2453–2455.
24. Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 14967–14972.
25. Boggs, B. A., Cheung, P., Heard, E., Spector, D. L., Chinault, A. C. & Allis, C. D. (2002) *Nat. Genet.* **30**, 73–76.
26. Heard, E., Rougeulle, C., Arnaud, D., Avner, P., Allis, C. D. & Spector, D. L. (2001) *Cell* **107**, 727–738.
27. Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F. & Allis, C. D. (2001) *Genes Dev.* **15**, 3286–3295.
28. Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R. & Stewart, A. F. (2001) *EMBO J.* **20**, 7137–7148.
29. Nagy, P. L., Griesenbeck, J., Kornberg, R. D. & Cleary, M. L. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 90–94.
30. Wang, H., Cao, R., Xia, L., Erdjument-Bromage, H., Borchers, C., Tempst, P. & Zhang, Y. (2001) *Mol. Cell* **8**, 1207–1217.
31. Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P. & Reinberg, D. (2002) *Genes Dev.* **16**, 479–489.
32. Nislow, C., Ray, E. & Pillus, L. (1997) *Mol. Biol. Cell* **8**, 2421–2436.
33. Bryk, M., Briggs, S. D., Strahl, B. D., Curcio, M. J., Allis, C. D. & Winston, F. (2002) *Curr. Biol.* **12**, 165–170.
34. Bahler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., Wach, A., Philippsen, P. & Pringle, J. R. (1998) *Yeast* **14**, 943–951.
35. Grewal, S. I. & Klar, A. J. (1997) *Genetics* **146**, 1221–1238.
36. Ekwall, K. & Partridge, J. F. (1999) in *Chromosome Structural Analysis: A Practical Approach*, ed. Bickmore, W. (Oxford Univ. Press, Oxford), pp. 47–57.
37. Nakayama, J., Klar, A. J. & Grewal, S. I. (2000) *Cell* **101**, 307–317.
38. Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., *et al.* (2002) *Nature (London)* **415**, 871–880.
39. Strahl, B. D., Grant, P. A., Briggs, S. D., Sun, Z. W., Bone, J. R., Caldwell, J. A., Mollah, S., Cook, R. G., Shabanowitz, J., Hunt, D. F. & Allis, C. D. (2002) *Mol. Cell. Biol.* **22**, 1298–1306.
40. Ivanova, A. V., Bonaduce, M. J., Ivanov, S. V. & Klar, A. J. (1998) *Nat. Genet.* **19**, 192–195.
41. Burd, C. G. & Dreyfuss, G. (1994) *Science* **265**, 615–621.
42. Aasland, R., Gibson, T. J. & Stewart, A. F. (1995) *Trends Biochem. Sci.* **20**, 56–59.
43. Aris, J. P. & Blobel G. (1988) *J. Cell Biol.* **107**, 17–31.
44. Grewal, S. I. (2000) *J. Cell. Physiol.* **184**, 311–318.
45. Allshire, R. C., Nimmo, E. R., Ekwall, K., Javerzat, J. P. & Cranston, G. (1995) *Genes Dev.* **9**, 218–233.
46. Grewal, S. I. S. & Elgin, S. C. (2002) *Curr. Opin. Genet. Dev.* **12**, 178–187.
47. Ekwall, K. & Ruusala, T. (1994) *Genetics* **136**, 53–64.
48. Thon, G., Cohen, A. & Klar, A. J. (1994) *Genetics* **138**, 29–38.
49. Hendzel, M. J., Wei, Y., Mancini, M. A., Van Hooser, A., Ranalli, T., Brinkley, B. R., Bazett-Jones, D. P. & Allis, C. D. (1997) *Chromosoma* **106**, 348–360.
50. Bjerling, P., Silverstein, R. A., Thon, G., Caudy, A., Grewal, S. & Ekwall, K. (2002) *Mol. Cell. Biol.* **22**, 2170–2181.
51. Zegerman, P., Canas, B., Pappin, D. & Kouzarides, T. (2002) *J. Biol. Chem.* **277**, 11621–11624.

Colloquium

Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion [*PSI*+]

Jia-Jia Liu*†, Neal Sondheimer†‡, and Susan L. Lindquist§

Department of Molecular Genetics and Cell Biology, Howard Hughes Medical Institute, University of Chicago, Chicago, IL 60637

The yeast prion [*PSI*+] provides an epigenetic mechanism for the inheritance of new phenotypes through self-perpetuating changes in protein conformation. [*PSI*+] is a nonfunctional, ordered aggregate of the translation termination factor Sup35p that influences new Sup35 proteins to adopt the same state. The N-terminal region of Sup35p plays a central role in prion induction and propagation. The C-terminal region provides translation termination activity. The function of the highly charged, conformationally flexible middle region (M) is unknown. An M deletion mutant was capable of existing in either the prion or the nonprion state, but in either case it was mostly insoluble. Substituting a charged synthetic polypeptide for M restored solubility, but the prions formed by this variant were mitotically very unstable. Substituting charged flexible regions from two other proteins for M created variants that acquired prion states (defined as self-perpetuating changes in function transferred to them from wild-type [*PSI*+] elements), but had profoundly different properties. One was soluble in both the prion and the nonprion form, mitotically stable but meiotically unstable, and cured by guanidine HCl but not by alterations in heat shock protein 104 (Hsp104p). The other could only maintain the prion state in the presence of wild-type protein, producing Mendelian segregation patterns. The unique character of these M variants, all carrying the same N-terminal prion-determining region, demonstrate the importance of M for [*PSI*+] and suggest that a much wider range of epigenetic phenomena might be based on self-perpetuating, prion-like changes in protein conformation than suggested by our current methods for defining prion states.

Yeast prions represent a fundamentally different mechanism for the transmission of genetic information than DNA based inheritance. With prions, heritable changes in phenotype are produced by self-perpetuating changes in protein conformation rather than by any changes in nucleic acids (1–3). [*PSI*+] and other genetic elements of this type are called prions because of conceptual similarities between their modes of transmission and that postulated for the infectious agent of the mammalian prion diseases (1, 4). However, yeast prions play a different role in the biology of cells that harbor them. They are not generally pathogenic. Rather, they modify metabolism in an epigenetic manner that can be beneficial to the organism under certain circumstances (5, 6).

The protein determinant of [*PSI*+] is Sup35, a subunit of the translation termination factor (7). In [*psi*−] cells, which lack the prion, Sup35 protein (Sup35p) is soluble and functional. In [*PSI*+] cells, most Sup35p is found in self-perpetuating, ordered aggregates. In this state, the protein is nonfunctional. The reduced concentration of functional translation-terminator factor causes ribosomes to occasionally read-through stop codons (2, 3). Thus, the presence of [*PSI*+] is routinely monitored by suppression of nonsense-codon mutations in auxotrophic markers (8). The phenotype is heritable because Sup35p in the [*PSI*+]

state influences newly synthesized Sup35p to adopt the same state, and because the protein is passed from mother cell to daughter during mitosis. When the daughter cell starts to make her own Sup35 proteins, they are influenced by preexisting [*PSI*+] complexes (inherited from the mother's cytoplasm) to undergo conformational conversion. Thus, the change in Sup35p function is inherited cytoplasmically.

Sup35p can be divided into three regions based on sequence analysis and functional investigations. The C-terminal region (C, amino acids 254–685) is responsible for the translation termination activity and is essential for viability (9–12). The N-terminal region (N, amino acids 1–123) is required for the induction and maintenance of [*PSI*+] (11–13). Deletion of N eliminates [*PSI*+], whereas even transient over expression of N induces [*PSI*+] (12). N is also responsible for the species barrier: in chimeric Sup35 proteins created from different species, the prion state is efficiently transferred only between proteins that share the same critical region of N (14–16). The role of the region between N and C (M, amino acids 124–253) remains unclear.

In inter-specific comparisons of Sup35p amino acid sequences, N and M are less conserved than C (7, 17). However, general features of these regions have been retained for long periods of evolution (14–16, 18). N regions from even distantly related *Hemiascomycetes* are rich in glutamine and asparagine residues (16, 19). M regions are highly charged, and their sequences are heavily biased toward a subset of charged amino acids (9, 16, 18, 19). In *Saccharomyces cerevisiae*, 42% of the residues in M are charged. All positively charged residues are lysines, and these cluster at the N-terminal end of M. The negatively charged residues, mostly glutamates, are concentrated at the C-terminal end.

Although [*PSI*+] is inherited in an orderly way, both mitotically and meiotically, it is metastable. [*PSI*+] cells occasionally give rise to [*psi*−] cells and vice versa as the [*PSI*+] conformation is lost or gained (20). The rate at which [*PSI*+] elements are lost greatly increases during growth on media containing guanidine hydrochloride (Gdn·HCl) (21, 22). The inheritance of [*PSI*+] is

also modulated by several protein chaperones (23–26). This effect is most striking with heat shock protein 104 (Hsp104p), a protein remodeling factor. Both overexpression and deletion of the *HSP104* gene cure cells of [*PSI*+] (23). Hsp104p must be present at an intermediate concentration for [*PSI*+] propagation.

The ability of Sup35p to exist in two stable and heritable conformations is fundamental to the conversion of cells from [*psi*−] to [*PSI*+] and vice versa. It is known that the crucial features of Sup35p required for this transition are confined to N and M: transferring N and M to a heterologous protein is sufficient to confer all of the prion behaviors of [*PSI*+] to that hybrid protein (27). Moreover, the transition from the nonprion to the prion state *in vivo* has been modeled *in vitro* by using the purified NM fragment of Sup35p. This polypeptide can exist for extended periods in the soluble state with a high degree of random coil and converts to a β-sheet-rich amyloid by seeded polymerization (28–30). To date, all of the critical elements for prion induction and propagation have been mapped to N. However, the M region (amino acids 124–253) is critical for solubility of NM *in vitro*. M alone remains soluble for months and cannot be seeded by preformed amyloids. On the other hand, N alone is soluble only in denaturing buffers (29, 31). This observation would suggest that M might play an important role in prion biology but its function has not been investigated *in vivo*. We created several alterations in the M region. They produced very different effects, demonstrating that M plays crucial roles in [*PSI*+] biology. In fact, the self-perpetuating prion states of these altered Sup35 proteins are so strikingly different from those of wild-type (WT) protein, they suggest that a much broader range of behaviors might involve prion-like changes in proteins than has previously been suspected.

Materials and Methods

Strains, Cultivation Procedures, and Genetic Analysis. The [*PSI*+] and [*psi*−] isogenic strain pair used were 74-D694a [MAT**a**, *ade1-14*(UGA), *trp1-289*(UAG), *his3Δ-200*, *ura3-52*, *leu2-3,112*] and 74-D694α [MATα, *ade1-14*(UGA), *trp1-289*(UAG), *his3Δ-200*, *ura3-52*, *leu2-3,112*]. The *HSP104* disruption strain used was 74-D694:hspΔ-1L [*ade1-14*(UGA), *trp1-289*(UAG), *his3Δ-200*, *ura3-52*, *leu2-3,112*, *hsp104*::LEU2]. [*ETA*+] strain: SL1010-1A [*ETA*+] [*psi*−] [*ade1-14* (UGA), *his 5-2*, *met8-1*, *ura3-52*, *leu2-1*(UAA) *trp1* (1 or 289)] (The [*ETA*+] strain was a kind gift from S. W. Liebman, University of Illinois, Chicago).

Cells were grown at 30°C in rich media [yeast extract/peptone/dextrose (YPD)] until mid-logarithmic phase (32). [*PSI*+] was monitored by its ability to suppress the ade1-14 (UGA) mutation and allow cells to grow on synthetic medium (SC) lacking adenine (−ADE). The 74-D694 [*psi*−] cells are red and [*PSI*+] cells are white or pink on rich medium (13, 23). [*PSI*+] curing on medium containing Gdn·HCl, CuSO₄ induction of [*PSI*+], and GFP fluorescence analysis of Sup35p aggregates were performed as described (33). Standard procedures for crossing, sporulation, and tetrad analysis were used (34).

Plasmid Construction. The *SUP35* M deletion (amino acids 124–253) was generated by site-directed mutagenesis using oligonucleotide primers and a QuikChange Site-Directed Mutagenesis kit (Stratagene). The primers used were 5′-CAC AGT CTC AAG GTA TGT TTG GTG GTA AAG-3′ and 5′-CTT TAC CAC CAA ACA TAC CTT GAG ACT GTG-3′. The template plasmid for mutagenesis was p316CUPSup35GFP^SG. The middle region-deleted *SUP35* gene fragment was then PCR amplified and inserted at the *Eco*RI/*Bam*HI sites of the *SUP35* integrative construct pJLI-sup35. PCR primers were as follows: primer A 5′-CG GAA TTC ATG TCG GAT TCA AAC CAA GG-3′ and primer B 5′-CGC GGA TCC TTA CTC GGC AAT TTT AAC-3′. Thus, in the resultant construct pJLI-sup35ΔM, amino acid 123 was fused directly to amino acid 254.

The KDG tripeptide repeats were inserted between *SUP35* N and C by site-directed mutagenesis. Primers were as follows: 5′-CCA CAG TCT CAA GGT AAG GAT GGT AAG GAT GGTAAG GAT GGTAAG GAT GGTAAG GAT GGT AAG GAT GGT TTT GGT GGT AAA GAT-3′ and 5′-ATC TTT ACC ACC AAA ACC ATC CTT ACC ATC CTT ACC ATC CTT ACC ATC CTT ACC ATC CTT ACC ATC CTT ACC TTG AGA CTG TGG-3′. The *SUP35* N-KDG6-C fragment was then PCR amplified and inserted into the *SUP35* integrative construct pJLI-SUP35 at the *Eco*RI/*Bam*HI sites.

The N9C substitution was created by amplifying the region of Hsp90p coding for amino acids 210–262 from genomic 74D-694 DNA by using the primers 5′-CAT CTA GAA AGG AAG TTG AAA AGG AAG-3′ and 5′-CTC CGC GGC TTG TTT AGT TCT TCT ATC TC-3′. The resulting fragment was inserted into pN164 by using the *Sac*II and *Xba*I restriction sites. The NTC substitution was created by using the same strategy after amplification of amino acids 635–713 of human topoisomerase I from plasmid pKM10 (a kind gift from J. Champoux, University of Washington, Seattle). The primers used were 5′-GAT CTA GAG CAC CAC CAA AAA CTT TTG AG-3′ and 5′-CAC CGC GGC ACT GTG ATC CTA GGG TCC AG-3′. Fidelity of all constructs was confirmed by sequencing.

Yeast Transformation. One milliliter of overnight culture was pelleted and resuspended in 0.1 ml LiAc-PEG-TE solution (40% wt/vol PEG 4000/100 mM LiAc/10 mM Tris·HCl, pH 7.5/1 mM EDTA). Cells were mixed with 10 μg/μl carrier DNA (salmon testis DNA, boiled for 10 min and rapidly chilled on ice) and 0.5 μg/μl transforming DNA and incubated overnight at room temperature. Cells were heat shocked at 42°C for 15 min, rapidly chilled on ice, spread directly onto synthetic defined medium selective for transformation, and incubated at 30°C.

Gene Replacement. The two-step gene integration and replacement procedure was used to create M mutant strains as described (34, 35). Genotypes were confirmed by genomic PCR, and protein expression was confirmed by immunoblotting.

To disrupt *HSP104* in [*PSI*+]^ΔM and [*PSI*+]^NKC strains, an *hsp104*::URA3 disruption plasmid (pYSU2) was inserted in the genome. The plasmid was digested with *Eco*RI/*Hind*III to release the *hsp104*::URA3 fragment, and the restriction digest was used to transform yeast cells. Transformants were selected on SC-URA medium, and the *HSP104* disruption was confirmed by immunoblotting with anti-Hsp104p antibodies (mix of monoclonal 8-1 and 2-3; ref. 36). *HSP104* was disrupted in [*PSI*+]^NTC and *SUP35N9C* backgrounds by the short flanking homology method as described (37).

Sup35p Solubility Assay. Cells grown to mid-logarithmic phase in liquid YPD medium were washed with 0.1 M EDTA (pH 8.0), resuspended in spheroplasting buffer (1.0 M sorbitol/0.1 M EDTA, pH 7.5/50 mM DTT/0.166 mg/ml zymolyase 100 T) and incubated at 30°C for 1 h. Cells were collected, washed with the spheroplasting buffer without zymolyase, and lysed by resuspension in lysis buffer (50 mM Tris·HCl, pH 7.2/10 mM KCl/100 mM EDTA/1 mM DTT/0.2% wt/vol SDS/1% vol/vol Triton X-100/1 mM benzamidine/2 mM phenylmethylsulfonyl fluoride/10 μg/ml leupeptin/2 μg/ml pepstatin A) with incubation on ice for 5 min. Lysates were subjected to centrifugation at 100,000 × *g* (48,000 rpm in a TLA-100 rotor, Beckman Coulter) for 20 min at 4°C. Supernatant and pellet fractions were analyzed by SDS/10% PAGE (NOVEX, San Diego) and transferred to a poly(vinylidene difluoride) membranes (Millipore). Membranes were incubated with primary antibody (antiSup35p- rabbit polyclonal antipeptide against amino acids 55–68) 1:1,000 in PBS buffer with 0.1% Tween 20 for 1 h, washed with PBS with 0.1% Tween 20, and incubated with goat anti-rabbit antibody conju-
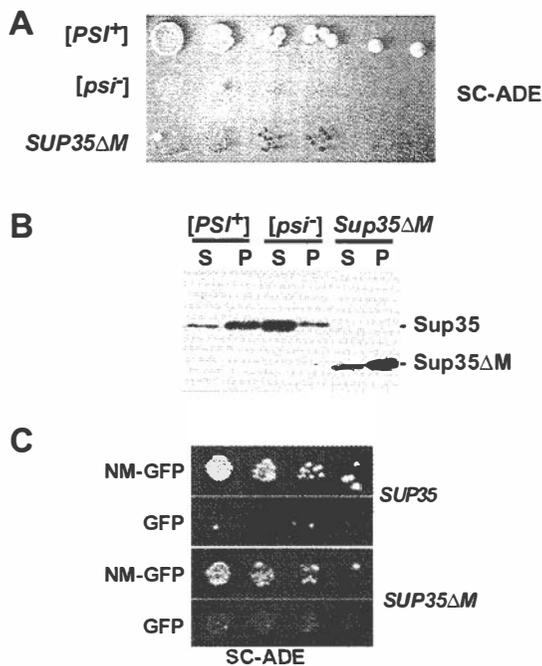
Fig. 1. Sup35ΔMp can convert to [*PSI*⁺]. (*A*) Five-fold serial dilution and growth of *SUP35ΔM* cells compared with WT [*PSI*⁺] and [*psi*⁻] cells on SC-ADE plates (30°C, 5 days). (*B*) Sup35p solubility assay of [*PSI*⁺], [*psi*⁻], and Sup35ΔMp. After cell lysis, high speed centrifugation and separation by SDS/10% PAGE Sup35ΔMp was detected by immunoblotting with anti-Sup35p antibody. S, supernatant fraction; P, pelleted fraction. (*C*) Induction of heritable [*PSI*⁺] factors in WT (*SUP35*) [*psi*⁻] and *SUP35ΔM* [*psi*⁻] strains. Five-fold serial dilutions of cells after 24-h induction of NM-GFP or GFP alone, plated on SC-ADE (30°C, 5 days).

gated to horseradish peroxidase (1:5,000), and immune complexes were visualized with enhanced chemiluminescence (ECL) reagent (Amersham Pharmacia).

Results

Sup35ΔMp Is Functional and Mostly Insoluble but Can Exist in Both Prion and Nonprion States. Starting with [*psi*⁻] cells, the WT copy of *SUP35* was replaced with a gene carrying a deletion of the middle region (*SUP35ΔM*). The strain contained a [*PSI*⁺]-suppressible nonsense mutation in the *ADE1* gene. In this background, [*PSI*⁺] cells grow on synthetic media deficient in adenine (SC-ADE) and are white on rich media. [*psi*⁻] cells do not grow on SC-ADE and produce red colonies on rich media because of the buildup of a colored byproduct of adenine biosynthesis.

Recombinant strains containing the *SUP35ΔM* gene ($n = 16$) at the *SUP35* locus were red on rich medium but showed a faint background of growth (dark red in color) on SC-ADE. Thus, at least some Sup35ΔM protein is functional in translation termination, but the protein is not as active as WT Sup35p (Fig. 1 *A* and *C*).

Differential centrifugation of cell lysates revealed that a much smaller fraction of Sup35ΔMp was present in the supernatant after a 100,000 × g spin, compared with WT Sup35p in [*psi*⁻] strains (Fig. 1*B*). Coomassie blue staining demonstrated equal loading of total proteins in these fractions and revealed no detectable changes in the solubility of other proteins (data not shown). The partial insolubility of Sup35ΔMp explains the weak suppressor phenotype of *SUP35ΔM* cells because less Sup35p is available for translational termination than in WT cells. Clearly, the highly charged M region provides a solubilizing function for

the Sup35 protein as a whole *in vivo*, as it does for the NM fragment *in vitro*.

Insolubility is a characteristic of Sup35p in the [*PSI*⁺] state (2) (Fig. 1*B*). The weak suppressor phenotype of *SUP35ΔM* could be due to a weak [*PSI*⁺] variant (see below). However the phenotype was unaffected by growth on Gdn·HCl, which efficiently cures [*PSI*⁺] (data not shown), suggesting that the insolubility of Sup35ΔMp might not be caused by prion formation. Also suggesting that the aggregates were not prion-like, the suppressor phenotype was recessive when *SUP35ΔM* cells were mated with WT [*psi*⁻] cells (data not shown). Moreover, differential centrifugation showed that aggregates of Sup35ΔMp in diploid cells did not cause WT Sup35p to fractionate into the pellet. That is, the Sup35ΔM protein in these cells could not recruit WT Sup35p to take on the [*PSI*⁺] state (data not shown). The aggregates had none of the characteristics of a prion and were more likely simply caused by loss of the solubilizing activity normally conferred by the highly charged M region.

Sup35ΔMp was, however, capable of acquiring a heritable [*PSI*⁺]-like state when *SUP35ΔM* cells were mated to WT [*PSI*⁺] cells. After mating, the strong suppressor phenotype of the [*PSI*⁺] parent was invariably dominant (data not shown), indicating that WT protein had converted Sup35ΔMp into a form that reduced its ability to function in translation. Moreover, the strong suppressor phenotype of the *SUP35ΔM* × [*PSI*⁺] diploid was cured by growth on medium containing Gdn·HCl (data not shown), indicating that Sup35ΔMp had acquired the [*PSI*⁺] state from WT protein, and could be cured of this state with Gdn·HCl, as is the WT protein.

Next, we asked whether Sup35ΔMp could acquire the prion state through another common mechanism, by transient over-expression of NM. In previous studies, we used NM fused to GFP to monitor the formation and propagation of [*PSI*⁺] in living cells (2). Sup35p in the [*PSI*⁺] state has the capacity to capture NM-GFP and induce it to adopt the same state, forming GFP aggregates visible by fluorescence microscopy. Furthermore, overexpression of NM-GFP induces new [*PSI*⁺] element formation in [*psi*⁻] cells (2). This state is retained even when the NM-GFP plasmid is lost.

SUP35ΔM cells were transformed with expression plasmids for GFP alone or NM-GFP. After 4 h of induction, intense coalescent foci were observed in many cells expressing NM-GFP, but never in cells expressing GFP alone (ref. 2 and data not shown). When plated to copper-free medium without selection for the plasmid, cells that had expressed NM-GFP produced colonies with a [*PSI*⁺] phenotype at a much higher frequency than those expressing GFP alone (Fig. 1*C*), suggesting that transient over-expression of NM had converted the Sup35ΔMp to the prion state. This was confirmed by 4:0 segregation of the suppressor phenotype in crosses to WT [*psi*⁻] cells and curing by growth on medium containing Gdn·HCl (data not shown). Therefore, the Sup35ΔMp can exist in two different states (we designate them [*PSI*⁺]ᐃM and [*psi*⁻]ᐃM) that are genetically analogous to the [*PSI*⁺] and [*psi*⁻] states of the WT Sup35 protein. Unlike WT Sup35p, however, the protein is largely insoluble in both cases (Figs. 1*B* and 2*B*).

Sup35ΔMp Can Maintain Different Prion Variants. Although haploid [*PSI*⁺]ᐃM strains were readily obtained by overexpressing NM-GFP, they were not readily obtained by other standard methods. For example, in one case *SUP35ΔM* was integrated at the site of WT *SUP35* (in tandem with it) in a [*PSI*⁺] strain (i.e., both *SUP35ΔM* and *SUP35* were present). These transformants retained the [*PSI*⁺] phenotype, as expected because WT Sup35p in the prion state converts Sup35ΔMp to that state. However, when one of the two genes was excised by selection against the *URA3* marker that had been used for transformation, clones carrying only *SUP35ΔM* did not survive. *SUP35ΔM* derivatives
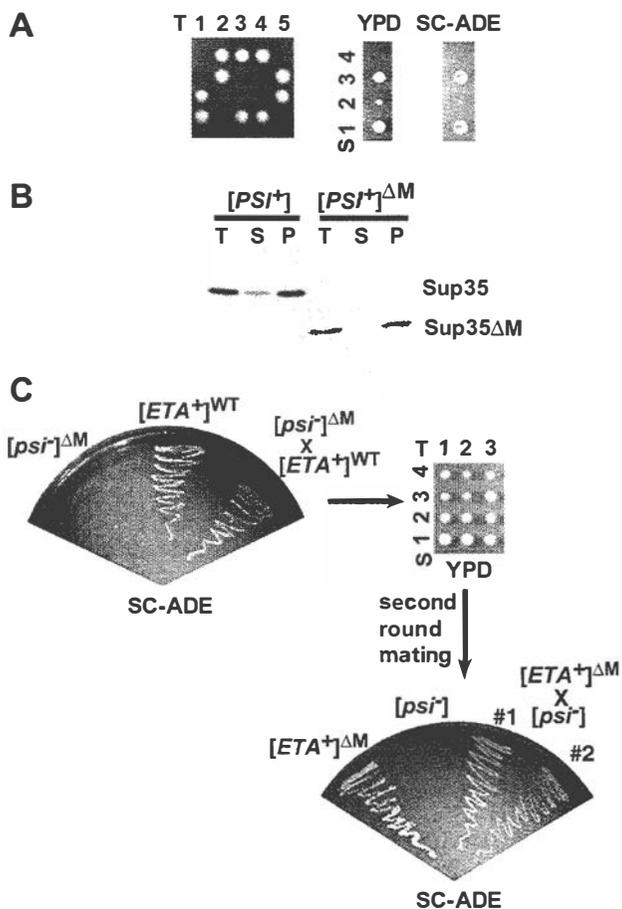
Fig. 2. Sup35ΔMp can maintain different [*PSI⁺*] variants. (*A*) Most *SUP35ΔM* spores from *SUP35ΔM* × WT [*PSI⁺*] crosses were not viable. (*Left*) Five representative tetrads from spores of a [*PSI⁺*] diploid from a *SUP35ΔM* [*psi⁻*] × *SUP35* [*PSI⁺*] cross, dissected on YPD medium. (*Right*) Growth of one set of tetrads with one surviving *SUP35ΔM* spore on YPD and SC-ADE. T, tetrad; S, spore. (*B*) Sup35p solubility properties of the WT spore and the *SUP35ΔM* spore (methods as described in Fig. 1). T, total protein; S, supernatant; P, pellet. (*C*) Sup35ΔMp can form and propagate [*ETA⁺*]. (*Upper Left*) Growth of *SUP35ΔM* [*psi⁻*] haploid strain ([*psi⁻*]ᐃᴹ), WT [*ETA⁺*] haploid strain ([*ETA⁺*]ᵂᵀ), and the diploid from the cross of these two strains on SC-ADE medium. (*Upper Right*) Three representative tetrads dissected from the diploid *SUP35ΔM[psi⁻]* × *SUP35[ETA⁺]* on YPD medium. (*Lower Right*) Growth on SC-ADE of haploid parental strains and two diploid strains from a cross of a *SUP35ΔM* spore from *SUP35ΔM* [*psi⁻*] ([*psi⁻*]ᐃᴹ)/*SUP35* [*ETA⁺*] ([*ETA⁺*]ᵂᵀ) diploid to WT [*psi⁻*].

were readily obtained when the initial insertion had been in a [*psi⁻*] background and other *SUP35* [*PSI⁺*] variants are readily obtained by this method (35). Similarly, sporulating *SUP35/SUP35ΔM* [*psi⁻*] diploids yielded the expected number of viable *SUP35ΔM* spores, but sporulating *SUP35/SUP35ΔM* [*PSI⁺*] diploids yielded very few (Fig. 2*A*). Those that were recovered grew very slowly, even on rich media (Fig. 2*A*), and by differential centrifugation, virtually all of their Sup35ΔM protein was found in the pellet (Fig. 2*B*). Genetic crosses between these slow-growing strains and the WT [*psi⁻*] strain generated [*PSI⁺*] diploids (data not shown). Thus, Sup35ΔMp in the prion state can transmit that state to WT protein. But [*PSI⁺*] cells in which the only copy of Sup35p is Sup35ΔMp can have unexpected problems with viability.

One explanation is that Sup35ΔMp, like WT Sup35p, can form prion variants with different "strengths." The phenomenon of different prion states (called prion "strains" or variants) that are strong, moderate, weak, and very weak is well characterized (13,

38–40). These variants are not caused by genetic differences, but are caused by epigenetic differences in the rates that prion variants capture and convert new Sup35p to the prion state. They have different levels of soluble Sup35p and different rates of translation termination (39–41). Because Sup35ΔMp is inherently less soluble than WT Sup35p, if it acquired a strong prion state there might be too little translation termination activity to keep cells viable. The haploid [*PSI⁺*]ᐃᴹ strains induced by transient overexpression of NM-GFP might represent weak variants, viable because a greater fraction of the Sup35ΔM protein remains soluble and active.

To test this possibility, we mated [*psi⁻*]ᐃᴹ cells to the weak [*PSI⁺*] variant [*ETA⁺*]. Conversion of Sup35ΔMp by this weak variant should leave a greater fraction of Sup35ΔMp in solution and produce more viable haploid *SUP35ΔM* [*PSI⁺*]ᐃᴹ spores. The diploid showed the same suppression of the *ade1–14* marker as the [*ETA⁺*] haploid parent (Fig. 2*C Left*), suggesting that the Sup35ΔM protein had converted to a weak prion state. In contrast to the poor viability of *SUP35ΔM* spores after mating to strong [*PSI⁺*] strains (Fig. 2*A*), nearly all *SUP35ΔM* progeny from the [*ETA⁺*] were viable (Fig. 2*C Upper Right*). When these progeny were mated to WT [*psi⁻*] tester strains, the diploids exhibited the suppressor phenotype characteristic of [*ETA⁺*] strains (Fig. 2*C Lower Right*). Thus, Sup35ΔMp could acquire, maintain, and transmit the [*ETA⁺*] state to WT protein. The N region is sufficient to form prion variants of different strengths.

The M Region Promotes Mitotic Stability of the [*PSI⁺*] State. On rich media, [*PSI⁺*]-mediated nonsense suppression is not required for growth, yet WT cells retain [*PSI⁺*] with high fidelity. In contrast [*PSI⁺*]ᐃᴹ strains lost the prion at a high rate (Fig. 3*A*). We asked whether we could restore stability simply by restoring solubility to the protein. To do this, a DNA fragment encoding a highly charged polypeptide rich in lysine and glutamic acid (6xKDG) was inserted in place of M creating the replacement *SUP35NKC*.

As expected, when the WT *SUP35* gene was replaced by *SUP35NKC* in [*psi⁻*] cells, they retained a [*psi⁻*] phenotype. The solubility of the Sup35NKC protein in this state was comparable to that of WT Sup35p (Fig. 3*B*). To determine whether this protein could acquire the prion state, *SUP135NKC* mutants were mated to a typical strong prion strain (Fig. 3*C*) and the diploid strain showed the suppressor phenotype. Sporulation of this diploid (data not shown) produced haploid *SUP35NKC* [*PSI⁺*] cells ([*PSI⁺*]ᴺᴷᶜ). Most Sup35NKCp was soluble in [*psi⁻*]ᴺᴷᶜ strains, and most of the protein became insoluble when it adopted the [*PSI⁺*]ᴺᴷᶜ state (Fig. 3*B*). Unlike [*PSI⁺*]ᐃᴹ cells, [*PSI⁺*]ᴺᴷᶜ cells exhibited no growth defect when streaked on rich medium (data not shown). However, even though Sup35NKCp appeared to be as soluble as WT protein and produced no general growth disadvantage, the [*PSI⁺*]ᴺᴷᶜ phenotype was highly unstable (Fig. 3*D*). Thus, replacement of the M region with a charged polypeptide that increases its inherent solubility *in vivo* is not sufficient to restore stability to the prion state. M provides more than a simple solubilizing function to Sup35p. It also promotes the mitotic stability of [*PSI⁺*].

The M region is highly charged and, in the soluble state, circular dichromism spectroscopy shows it to have a highly flexible structure (≈60% random coil; A. Cashikar and T. Scheibel, personal communication). Our next alterations were to replace the M region with two naturally occurring polypeptides that, like WT M, are highly charged and are known to have conformational flexibility.

The Human Topoisomerase Linker Restores Mitotic Stability but Causes Meiotic Instability. The human topoisomerase linker (T) has a percentage of charged residues similar to Sup35Mp. The linker has been characterized by x-ray crystallography and contains both structured and unstructured regions that link
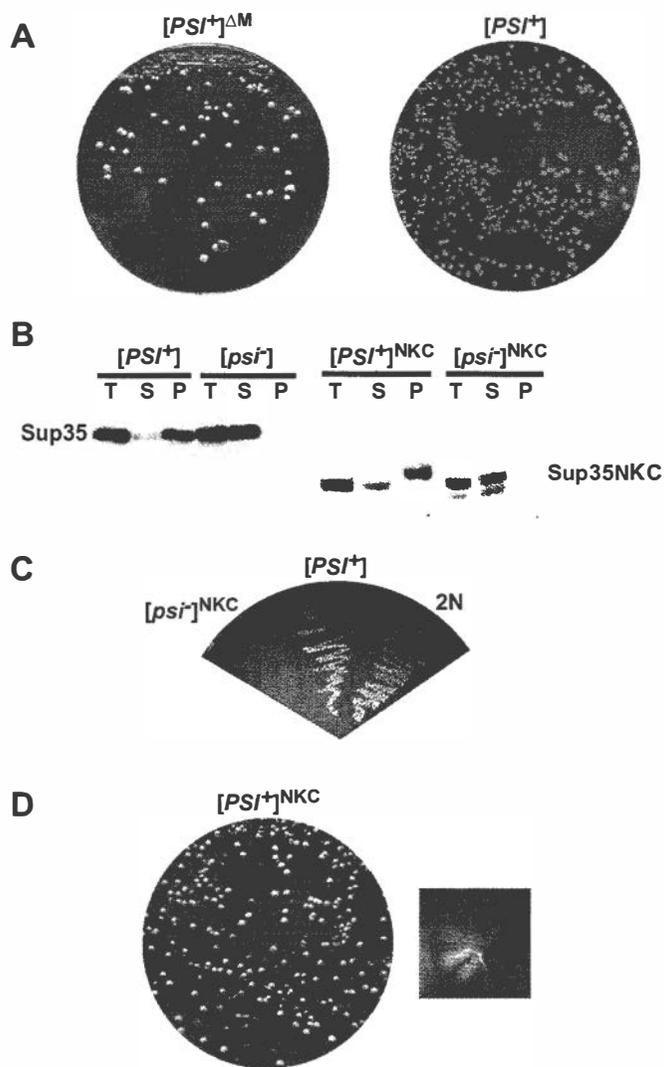
Fig. 3. The M region promotes mitotic stability of strong [*PSI*⁺].

Fig. 3. The M region promotes mitotic stability of strong [*PSI*⁺]. (*A*) Appearance of red sectors ([*psi*⁻]) out of white colonies when [*PSI*⁺]$^{\Delta M}$ cells were streaked onto YPD and incubated for several days (*Left*). Growth of [*PSI*⁺] cells on YPD shown for comparison (*Right*). (*B*) Sup35p solubility assay of the NKC mutant protein compared with that of WT Sup35p (methods as described in Fig. 1). (*Left*) WT [*PSI*⁺] and [*psi*⁻] strains. (*Right*) SUP35NKC [*PSI*⁺] and [*psi*⁻] ([*PSI*⁺]^{NKC} and [*psi*⁻]^{NKC}) haploid strains. (*C*) Mating SUP35NKC [*psi*⁻] ([*psi*⁻]^{NKC}) to [*PSI*⁺] generated [*PSI*⁺] diploid. Growth of the two parental strains and a diploid (2N) progeny on SC-ADE is shown. (*D*) [*PSI*⁺]^{NKC} cells were mitotically unstable. Cells were grown in liquid YPD for 16 h and plated onto YPD plates. Most colonies were red/white sectored (*Left*). A close-up image of one of the sectored colonies is shown (*Right*).

other functional domains of the protein (42). When an M to T replacement (*SUP35NTC*) was inserted in tandem with *SUP35* in [*psi*⁻] cells, strains retaining only *SUP35NTC* were obtained at an equal frequency to WT. *SUP35NTC* strains were phenotypically identical to WT [*psi*⁻] cells with respect to growth on rich media and SC-ADE. Sup35NTCp could readily be converted to the prion state, [*PSI*⁺]^{NTC}, by matings to WT [*PSI*⁺] strains. The haploid progeny of sporulation showed normal viability. [*PSI*⁺]^{NTC} strains were also obtained after transient overexpression of Sup35NTCp from an inducible plasmid in the *SUP35NTC* background.

[*PSI*⁺]^{NTC} was tested for other common prion properties including curability, mitotic stability, and non-Mendelian inheritance during meiosis (see Table 1). It was mitotically stable (Fig.

4*A*), capable of growth on SC-ADE (Fig. 4*B*) and was cured by growth on media containing Gdn·HCl, but was not cured by either the overexpression or the deletion of *HSP104* (Table 1, Fig. 4*C*). To further characterize [*PSI*⁺]^{NTC} and [*psi*⁻]^{NTC} states, we analyzed the solubility of the Sup35NTC protein. In both [*PSI*⁺]^{NTC} and [*psi*⁻]^{NTC} cells, most Sup35NTCp was soluble after a 100,000 × g spin (Fig. 4*D*). To test the aggregation of Sup35p by using GFP, we expressed a plasmid with a fusion of the N and T regions to the GFP marker (NT-GFP) in [*PSI*⁺]^{NTC} and [*psi*⁻]^{NTC} cells. NT-GFP showed a diffuse fluorescence pattern in both strain types confirming that the protein does not form large aggregates (Fig. 4*E*). Therefore this protein can exist in states that are genetically analogous to the prion states of Sup35p, but in both states most of the protein remains soluble after centrifugation at 100,000 × g for 20 min.

The [*PSI*⁺]^{NTC} state was dominant in crosses to [*psi*⁻]^{NTC} cells, indicating that it readily converted soluble Sup35NTCp to the prion state. When [*PSI*⁺]^{NTC} homozygous diploids were sporulated, the frequency of meiotic transmission of the suppressor phenotype to offspring was not always 4:0, the ratio typical for WT [*PSI*⁺] diploids, but it was clearly non-Mendelian (Fig. 4*F* and Table 1). This segregation pattern was similar to that of another yeast prion, [*URE3*] (43). These findings suggest that the highly charged M region also influences the accurate propagation of [*PSI*⁺] elements through meiosis. The different effects of M substitutions on mitotic and meiotic stability suggest that the mechanisms for maintaining meiotic and mitotic stability are, at least in part, distinct.

Substitution of the Hsp90p Linker for M Causes Another Distinct Genetic Behavior. The other M substitution we tested was derived from *S. cerevisiae* Hsp90 protein. This highly charged region (amino acids 210–262 of the polypeptide sequence) connects the two stably folded domains of Hsp90p and is degraded by even very short treatments with proteases, suggesting it is not inherently a tightly folded polypeptide (44). As with *SUP35NTC,* when M was replaced by this portion of the *HSP90* coding sequence (*SUP35N9C*), [*psi*⁻] cells retained a nonsuppressor state (data not shown).

In contrast to *SUP35NTC*, *SUP35ΔM*, and *SUP35NKC* cells, a suppressor state could not be induced in haploid *SUP35N9C* cells by overexpression of polypeptides containing the N region (data not shown). The protein could, however, acquire a [*PSI*⁺]-like state when *SUP35N9C* cells were mated to WT [*PSI*⁺] cells (Fig. 5*A*). The diploid strain had many other characteristics of [*PSI*⁺] strains, including a suppressor phenotype that was eliminated by plating to media containing Gdn·HCl (see Table 1). It also showed strong mitotic stability. But surprisingly, sporulation of this diploid always produced two [*PSI*⁺] colonies with a *SUP35* genotype and two *SUP35N9C* with the [*psi*⁻] phenotype (Fig. 5*A*).

These observations suggested that Sup35N9Cp could enter a [*PSI*⁺]-like state, but could only acquire that state from preformed [*PSI*⁺] elements and could not thereafter retain it on its own (when separated by sporulation from the WT protein). To more fully characterize these transitions we examined the solubility of the Sup35N9C protein in the haploid *SUP35N9C* strain, the diploid *SUP35/SUP35N9C* [*PSI*⁺] strain and the haploid progeny of sporulation. The Sup35N9C protein was almost entirely soluble in the *SUP35N9C* parent, but was insoluble (as was WT Sup35p) in the heterozygous [*PSI*⁺] diploid (Fig. 5*D*). After sporulation, Sup35N9Cp became soluble once again in the *SUP35N9C* haploid progeny, whereas the insoluble prion state was maintained in *SUP35* progeny (Fig. 5*D*, right lanes). This result was confirmed by the presence of small foci in the [*PSI*⁺] diploid after 2 h of expression of an N9-GFP fusion protein (Fig. 5*B*). In contrast, N9-GFP fluorescence in the nonsuppressed haploid *SUP35N9C* remained diffuse (Fig. 5*C*). Thus,

Table 1. M substitution mutants and their properties

Region length; % charge; no. of positive amino acid residues that are lysine	[PSI+] inducibility		Segregation pattern [PSI+]:[psi−]	Stable in mitosis	Gdn·HCl curable	Protein solubility		Hsp104 curability	
	Sup35 N terminus over express	Mate to [PSI+]				[PSI+]	[psi−]	ΔHSP104	Hsp104 over express
WT; 130 aa; 42%; 24 of 24	+	+	4:0	+	+	−	+	+	+
ΔM	+	+	4:0	−	+	−	−	+	+
KDG₆(NKC); 18 aa; 67%; 6 of 6	+	+	4:0	−	+	−	+	N.T.	N.T.
HuTop I (NTC); 79 aa; 44%; 17 of 21	+	+	4:0 17% 3:1 56% 2:2 22% 1:3 4%	+	+	+	+	−	−
Hsp90 (N9C); 53 aa; 77%; 17 of 17	−	+	2:2*	+*	+*	−*	+	+*	−

N.T., not tested.

*Tested in heterozygous [PSI+] diploid (WT/N9C).

Sup35N9Cp can readily enter a [PSI+]-like state under the influence of WT protein in that state, but it cannot maintain that state on its own.

Discussion

We have demonstrated that the M region of Sup35p makes important and diverse contributions to genetic and biochemical properties of [PSI+]. Sup35p mutants with a deletion of the M region or with substitutions in place of M can form prions, but these states are strikingly distinct from WT [PSI+] and from each other. A wide variety of prion states and behaviors can be conferred on the same C-terminal functional domain and N-terminal prion domain by intervening "auxiliary" sequences.

Prion proteins such as PrP and Sup35p aggregate when adopting the prion conformation (2, 37, 45, 46). However, large-scale aggregation is neither necessary nor sufficient for entry into the prion state. (The former has also been suggested by the analysis of certain URE3 prion variants, ref. 47.) We have shown that M helps maintain Sup35p in the soluble state and, as a result, Sup35ΔMp is found in the pellet after centrifugation of cell lysates, regardless of its prion state. This finding confirms the special nature of prion state protein.
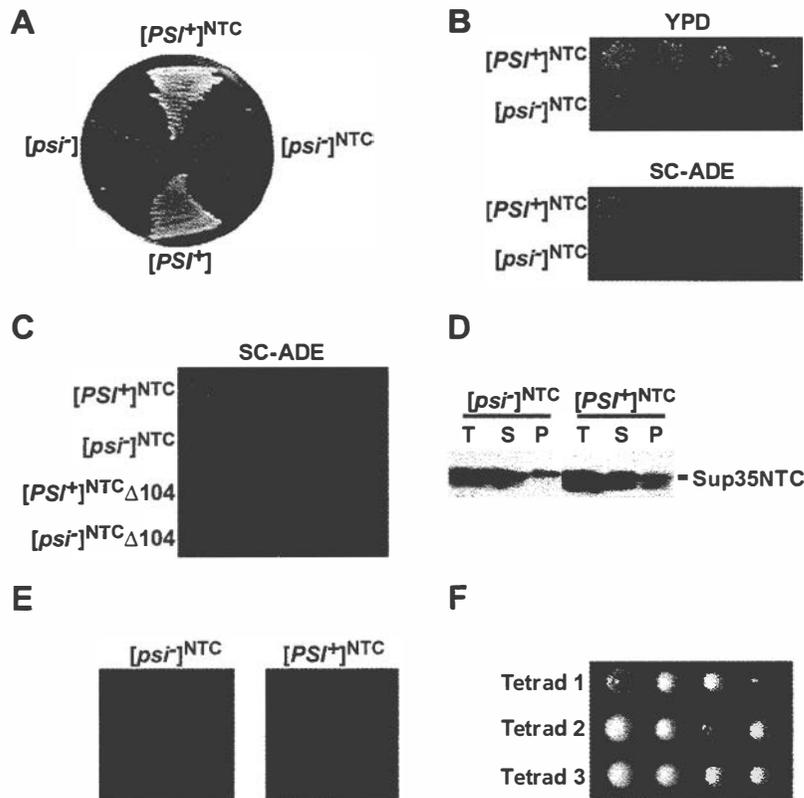

Fig. 4. Substitution of human topoisomerase I linker for the Sup35p M region causes meiotic instability. (A) [psi−]NTC and [PSI+]NTC strains containing the linker from human topoisomerase I in place of the M linker of Sup35p were plated onto YPD; [PSI+] and [psi−] were plated for comparison. (B) Growth phenotype of SUP35NTC cells on YPD and SC-ADE. (C) Deletion of HSP104 did not affect the suppressor state of [PSI+]NTC cells. (D) Sup35p solubility assay of [PSI+]NTC and [psi−]NTC (methods as described in Fig. 1). (E) Expression of NT-GFP in [psi−]NTC and [PSI+]NTC did not induce aggregation. (F) Tetrad dissection of a [PSI+]NTC diploid shows varying numbers of [PSI+]NTC and [psi−]NTC spores.
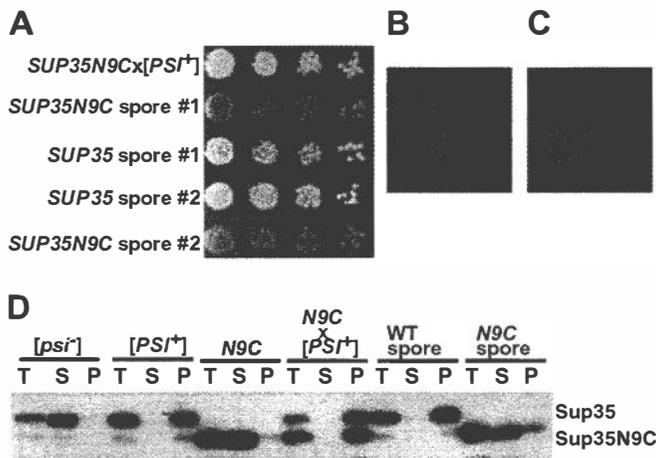
Fig. 5. Substitution of the Hsp90p linker for the Sup35p M region causes distinct genetic behavior. (A) Diploid [PSI+] cells with the *SUP35/N9C* genotype grew white on YPD (top row). On sporulation of this diploid, two red colonies and two white colonies were always obtained (following rows). (B) Expression of N9-GFP in the *SUP35/N9C* diploid [PSI+] strain causes aggregation. (C) Expression of N9-GFP in haploid cells expressing only *N9C* did not cause aggregation. (D) Sup35p solubility assay of N9C indicates that it pelleted only in the presence of Sup35p in the [PSI+] state. The protein returned to the soluble state after sporulation (methods as described in Fig. 1).

Differences between the Sup35ΔM protein in [PSI+]^ΔM cells and in [psi−]^ΔM cells are not simply a difference between aggregated and nonaggregated states. This point is also demonstrated by our experiments with Sup35NTCp. No aggregated state was detectable in [PSI+]^NTC cells. The prion state of Sup35NTCp may well involve higher-order complexes, but if so, they are clearly different in character from the large complexes of WT Sup35p in the [PSI+] state.

The M region is also important for the stabilization of [PSI+] during cell division. Cells with either the *SUP35ΔM* or the *SUP35NKC* replacement genotype could enter a prion state, but this state was not well maintained during mitotic division. Cells with the *SUP35NTC* genotype could also enter the prion state, and [PSI+]^NTC was mitotically stable. However, [PSI+]^NTC was not propagated after meiosis with the same fidelity as [PSI+]. Thus, the propagation of prion elements is quite sensitive to changes in the M region. Requirements for the maintenance of the prion during mitotic and meiotic cell division are distinct and M contributes to them both.

Altering the M region also had important consequences for prion curing. Because Hsp104p function is sensitive to Gdn·HCl, it has been suggested that Gdn·HCl treatment cures cells through the inactivation of Hsp104p (48–51), but this hypothesis is controversial (48, 50, 52). We have identified a prion state, [PSI+]^NTC, which can be cured by growth on Gdn·HCl but cannot be cured by *HSP104* deletion (Table 1). This finding indicates that curing by Gdn·HCl is not solely caused by Hsp104p inactivation. The results also suggest that some feature of M strongly

influences interactions with Hsp104p. In this respect, it is intriguing that 24 of 24 of the positively charged amino acids in M are lysines (Table 1). Polylysine binds Hsp104p with high affinity and triggers a series of changes in ATP hydrolysis and Hsp104p conformation (36). The deletion/insertion mutations we used also exhibit some conformational flexibility and lysine richness. However, the effect of Hsp104p on the prion-state conversions of these proteins differed. There must be something more than the mere presence of lysines and conformational flexibility that influences the interactions of the M region with Hsp104p. For example, length, specific sequence elements, residue spacings, and conformational predisposition could all influence these interactions.

The Sup35N9C protein showed an intriguing genetic property that was entirely unexpected. Unlike other M alterations, Sup35N9Cp was incapable of entering a [PSI+]-like state until cells expressing it were mated to cells already containing WT Sup35p in that state. Further, Sup35N9Cp could not support the prion state on its own. Once the two proteins were separated by sporulation, *SUP35N9C* cells reverted to a [psi−] phenotype and the Sup35N9C protein returned to the soluble state. Thus, Sup35N9Cp can participate in a heritable phenotypic change caused by a protein-only mechanism that exhibits Mendelian segregation, a striking departure from ordinary prion behavior.

The M region is not required for entry into the prion state. Yet in the context of proteins with the same N-terminal prion-determining region and C-terminal functional region, substitution of M with artificial linker regions confers a rich variety of genetic and biochemical characteristics to the prion state. Proteins with very different physical properties can undergo self-perpetuating conformational changes in state, but produce similar phenotypes (Sup35ΔMp, Sup35NTCp); proteins with similar physical characteristics can display very different genetic properties (WT Sup35p, Sup35N9Cp). Indeed, were it not for the fact that (*i*) the prion-determining nature of the N region of Sup35p has been extensively characterized by genetic, biochemical and cell biological methods, and (*ii*) the self-perpetuating changes in function obtained with each of our different mutants were acquired from WT Sup35p that was in its well-characterized prion state, it might be hard to argue that the changes in function we observed were caused by self-perpetuating prion conformations.

We have created prion variants with unusual properties by deliberate, artificial manipulations. We have no direct evidence that proteins with such properties exist in nature. However, given the divergence of prion sequences, which is particularly great in the M region, it seems reasonable to suppose that proteins that have such properties might well have appeared. Thus, we suggest that there may be many prion-like, self-perpetuating states not recognizable as such by the prion-defining criteria used to date. It may be that more epigenetic changes in biology are caused by prion-like processes than previously realized.

1. Wickner, R. B., Masison, D. C. & Edskes, H. K. (1995) *Yeast* **11,** 1671–1685.
2. Patino, M. M., Liu, J. J., Glover, J. R. & Lindquist, S. (1996) *Science* **273,** 622–626.
3. Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. (1996) *EMBO J.* **15,** 3127–3134.
4. Lindquist, S. (1997) *Cell* **89,** 495–498.
5. Eaglestone, S. S., Cox, B. S. & Tuite, M. F. (1999) *EMBO J.* **18,** 1974–1981.
6. True, H. L. & Lindquist, S. L. (2000) *Nature (London)* **407,** 477–483.
7. Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L. & Philippe, M. (1995) *EMBO J.* **14,** 4065–4072.
8. Cox, B. S., Tuite, M. F. & McLaughlin, C. S. (1988) *Yeast* **4,** 159–178.
9. Kushnirov, V. V., Ter-Avanesyan, M. D., Telckov, M. V., Surguchov, A. P., Smirnov, V. N. & Inge-Vechtomov, S. G. (1988) *Gene* **66,** 45–54.
10. Kushnirov, V. V., Ter-Avanesian, M. D., Dagkesamanskaia, A. R., Chernov Iu, O., Inge-Vechtomov, S. G. & Smirnov, V. N. (1990) *Mol. Biol.* **24,** 1037–1041.
11. Ter-Avanesyan, M. D., Kushnirov, V. V., Dagkesamanskaya, A. R., Didichenko, S. A., Chernoff, Y. O., Inge-Vechtomov, S. G. & Smirnov, V. N. (1993) *Mol. Microbiol.* **7,** 683–692.
12. Ter-Avanesyan, M. D., Dagkesamanskaya, A. R., Kushnirov, V. V. & Smirnov, V. N. (1994) *Genetics* **137,** 671–676.

13. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. (1996) *Genetics* **144**, 1375–1386.
14. Chernoff, Y. O., Galkin, A. P., Lewitin, E., Chernova, T. A., Newnam, G. P. & Belenkiy, S. M. (2000) *Mol. Microbiol.* **35**, 865–876.
15. Kushnirov, V. V., Kochneva-Pervukhova, N. V., Chechenova, M. B., Frolova, N. S. & Ter-Avanesyan, M. D. (2000) *EMBO J.* **19**, 324–331.
16. Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. (2000) *Cell* **100**, 277–288.
17. Stansfield, I., Jones, K. M., Kushnirov, V. V., Dagkesamanskaya, A. R., Poznyakovski, A. I., Paushkin, S. V., Nierras, C. R., Cox, B. S., Ter-Avanesyan, M. D. & Tuite, M. F. (1995) *EMBO J.* **14**, 4365–4373.
18. Kushnirov, V. V., Ter-Avanesyan, M. D., Didichenko, S. A., Smirnov, V. N., Chernoff, Y. O., Derkach, I. L., Novikova, O. N., Inge-Vechtomov, S. G., Neistat, M. A. & Tolstorukov, I. I. (1990) *Yeast* **6**, 461–472.
19. Nakayashiki, T., Ebihara, K., Bannai, H. & Nakamura, Y. (2001) *Mol. Cell* **7**, 1121–1130.
20. Cox, B. S. (1965) *Heredity* **20**, 505–521.
21. Singh, A., Helms, C. & Sherman, F. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 1952–1956.
22. Tuite, M. F., Mundy, C. R. & Cox, B. S. (1981) *Genetics* **98**, 691–711.
23. Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Liebman, S. W. (1995) *Science* **268**, 880–884.
24. Jung, G., Jones, G., Wegrzyn, R. D. & Masison, D. C. (2000) *Genetics* **156**, 559–570.
25. Kushnirov, V. V., Kryndushkin, D. S., Boguta, M., Smirnov, V. N. & Ter-Avanesyan, M. D. (2000) *Curr. Biol.* **10**, 1443–1446.
26. Newnam, G. P., Wegrzyn, R. D., Lindquist, S. L. & Chernoff, Y. O. (1999) *Mol. Cell. Biol.* **19**, 1325–1333.
27. Li, L. & Lindquist, S. (2000) *Science* **287**, 661–664.
28. DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. (1998) *Cell* **93**, 1241–1252.
29. Glover, J. R., Kowal, A. S., Schirmer, E. C., Patino, M. M., Liu, J. J. & Lindquist, S. (1997) *Cell* **89**, 811–819.
30. Serio, T. R., Cashikar, A. G., Kowal, A. S., Sawicki, G. J., Moslehi, J. J., Serpell, L., Arnsdorf, M. F. & Lindquist, S. L. (2000) *Science* **289**, 1317–1321.
31. King, C. Y., Tittmann, P., Gross, H., Gebert, R., Aebi, M. & Wuthrich, K. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6618–6622.
32. Kaiser, C., Michaelis, S. & Mitchell, A. (1994) *Methods in Yeast Genetics: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
33. Chernoff, Y. O., Uptain, S. M. & Lindquist, S. L. (2002) *Methods Enzymol.* **351**, 499–538.
34. Guthrie, C. & Fink, G. R., eds. (1991) *Guide to Yeast Genetics and Molecular Biology* (Academic, San Diego).
35. Liu, J. J. & Lindquist, S. (1999) *Nature (London)* **400**, 573–576.
36. Cashikar, A. G., Schirmer, E. C., Hattendorf, D. A., Glover, J. R., Ramakrishnan, M. S., Ware, D. M. & Lindquist, S. L. (2002) *Mol. Cell* **9**, 751–760.
37. Sondheimer, N. & Lindquist, S. (2000) *Mol. Cell* **5**, 163–172.
38. Liebman, S. W. & All-Robyn, J. A. (1984) *Curr. Genet.* **8**, 567–573.
39. Zhou, P., Derkatch, I. L., Uptain, S. M., Patino, M. M., Lindquist, S. & Liebman, S. W. (1999) *EMBO J.* **18**, 1182–1191.
40. Uptain, S. M., Sawicki, G. J., Caughey, B. & Lindquist, S. (2001) *EMBO J.* **20**, 6236–6245.
41. Kochneva-Pervukhova, N. V., Chechenova, M. B., Valouev, I. A., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. (2001) *Yeast* **18**, 489–497.
42. Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. & Champoux, J. J. (1998) *Science* **279**, 1534–1541.
43. Wickner, R. B. (1994) *Science* **264**, 566–569.
44. Scheibel, T., Siegmund, H. I., Jaenicke, R., Ganz, P., Lilie, H. & Buchner, J. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1297–1302.
45. Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. (1997) *Mol. Cell. Biol.* **17**, 2798–2805.
46. Prusiner, S. B. & Scott, M. R. (1997) *Annu. Rev. Genet.* **31**, 139–175.
47. Fernandez-Bellot, E., Guillemet, E., Ness, F., Baudin-Baillieu, A., Ripaud, L., Tuite, M. & Cullin, C. (2002) *EMBO Rep.* **3**, 76–81.
48. Ferreira, P. C., Ness, F., Edwards, S. R., Cox, B. S. & Tuite, M. F. (2001) *Mol. Microbiol.* **40**, 1357–1369.
49. Glover, J. R. & Lindquist, S. (1998) *Cell* **94**, 73–82.
50. Jung, G. & Masison, D. C. (2001) *Curr. Microbiol.* **43**, 7–10.
51. Jung, G., Jones, G. & Masison, D. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9936–9941.
52. Wegrzyn, R. D., Bapat, K., Newnam, G. P., Zink, A. D. & Chernoff, Y. O. (2001) *Mol. Cell. Biol.* **21**, 4656–4669.

Colloquium

Heritable chromatin structure: Mapping "memory" in histones H3 and H4

Christine M. Smith[†], Zara W. Haimberger[†], Catherine O. Johnson[†], Alex J. Wolf[†], Philip R. Gafken[‡], Zhongli Zhang[§], Mark R. Parthun[†¶], and Daniel E. Gottschling[†‖]

[†]Division of Basic Sciences and [‡]Proteomics Facility, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109; and [§]Department of Chemistry, University of Nebraska, Lincoln, NE 68588

Telomeric position effect in *Saccharomyces cerevisiae* is a chromatin-mediated phenomenon in which telomere proximal genes are repressed (silenced) in a heritable, but reversible, fashion. Once a transcriptional state (active or silenced) is established, however, there is a strong tendency for that state to be propagated. Twenty-five years ago, H. Weintraub and colleagues suggested that such heritability could be mediated by posttranslational modification of chromatin [Weintraub, H., Flint, S. J., Leffak, I. M., Groudine, M. & Grainger, R. M. (1977) *Cold Spring Harbor Symp. Quant. Biol.* 42, 401–407]. To identify potential sites within the chromatin that might act as sources of "memory" for the heritable transmission, we performed a genetic screen to isolate mutant alleles of the histones H3 and H4 genes that would "lock" telomeric marker genes into a silenced state. We identified mutations in the NH$_2$-terminal tail and core of both histones; most of the amino acid changes mapped adjacent to lysines that are known sites of acetylation or methylation. We developed a method using MS to quantify the level of acetylation at each lysine within the histone H4 NH$_2$-terminal tail in these mutants. We discovered that each of these mutants had a dramatic reduction in the level of acetylation at lysine 12 within the histone H4 tail. We propose that this lysine serves as a "memory mark" for propagating the expression state of a telomeric gene: when it is unacetylated, silent chromatin will be inherited; when it is acetylated an active state will be inherited.

Fig. 1. A genetic screen for histone H3 or H4 mutants that increase telomeric silencing. (*A*) Photograph of a colony of *S. cerevisiae* with a telomeric *ADE2* gene. Red (gray in photograph) sectors indicate cells in which the gene is silenced. (*B*) Outline of the principle of a screen for mutants in histone H3* or H4* that acted as bypass suppressors of the *cac1Δ* phenotype of telomeric silencing (see text for details). (*C*) Schematic representation of the "plasmid shuffle" method used to screen for histone H3 or H4 mutants that increased silencing. A mutagenized plasmid library of histone H3 and H4 genes (*HHT2*-*HHF2*-*TRP1*; pMP3) was transformed into strain UCC1371 (genotype is indicated). Those transformants that regained silencing (red and white sectored colonies that grew poorly, or not at all, in the absence of uracil) and lost the wild-type histone H3 and H4 genes plasmid (*HHT2-HHF2-LYS2*; pMP9) were chosen for further analysis.

In the beginning of the 20th century, the concept that genotype controlled inheritance of phenotype was defined and developed. However, in 1930, H. J. Muller described a mutant in *Drosophila* that was in apparent contradiction to the accepted dogma (1). The normal *Drosophila* compound eye is made of hundreds of red ommatidia, and typical eye-color mutants have a complete change in color of all of the ommatidia. However, Muller's "ever-sporting displacement" mutant resulted in the mosaic expression of red and white pigments in the *Drosophila* eye. This mosaic expression correlated with a specific chromosomal rearrangement that caused the *white* locus to be located near the centromere. In some eye cells of an individual, the *white* gene was expressed (red) whereas in others it was not (white). This phenomenon came to be known as position effect variegation and can now be heralded as the dawning realization that genotype did not necessarily predict phenotype (reviewed in refs. 2 and 3).

Telomeric position effect in the budding yeast *Saccharomyces cerevisiae* is very similar to position effect variegation in *Drosophila* (4). When a normal yeast gene is placed near a telomere, the gene undergoes transcriptional repression, or silencing, that is heritable over many cell generations. However, silencing is reversible, as the gene can become transcriptionally active. For example, when yeast cells with the *ADE2* gene placed near a telomere form a colony on solid medium, the colony is composed of subpopulations in which the *ADE2* gene is either expressed (white sectors) or repressed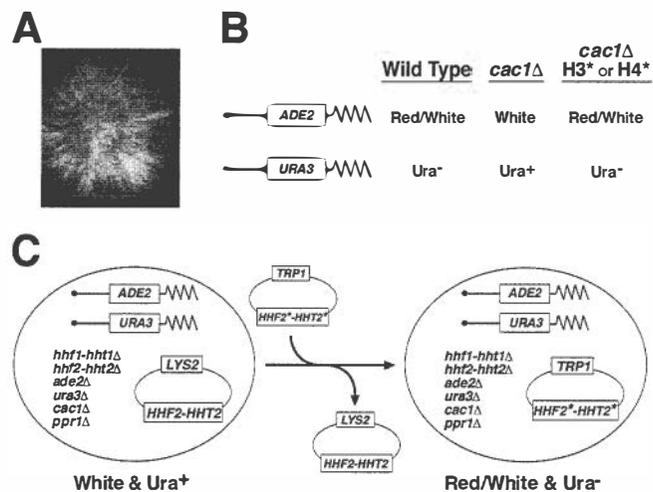 (red sectors, see Fig. 1*A*). The different phenotypes of the sectors in a colony reflect the ability of genetically identical cells to switch between phenotypic states. However, the fact that large sectors are phenotypically uniform reflects the ability of each state to be heritably propagated for multiple generations.

Telomeric silencing is the result of a repressive chromatin structure that initiates from the telomere and extends inward along the chromosome, rendering the enveloped DNA refractory to interaction with factors such as those of the transcriptional machinery (reviewed in ref. 5). The key structural components of telomeric silent chromatin are well defined. They

---

include telomeric DNA sequence, the telomere sequence DNA-binding proteins Rap1p and Ku, nucleosomal core histones H3 and H4, and nonhistone chromatin components Sir2p, Sir3p, and Sir4p (Silent Information Regulators). In a simple view of telomeric silencing, the Sir proteins are recruited to the telomeres through their interactions with Rap1p and Ku at the end of the chromosome, and with each other. They then "polymerize" along telomere-adjacent chromosome regions by binding the NH$_2$-terminal tails of histones H3 and H4 of the associated nucleosomes.

In addition to telomeres, the three Sir proteins act to repress the silent mating loci, *HML* and *HMR*. At these loci, additional factors are required that seem to be primarily involved in recruiting the Sir complex and helping to stabilize the silent state (6). In part because of this lack of redundancy, telomeric silencing is semistable and more sensitive to perturbation, particularly as a result of mutations in the NH$_2$ termini of histones H3 and H4. Mutations within residues 16–29 of histone H4 result in a defect in telomeric silencing (7). Deletion analysis of histone H3 implicates residues 4–20 as also being important for telomeric silencing (8).

Several lines of evidence indicate that posttranslational modifications (especially acetylation) of lysines in the NH$_2$-terminal tails of histone H3 and H4 are critical to telomeric silencing. For instance, the histone tails in silenced chromatin are hypoacetylated compared with the rest of the genome (9, 10). [However, there are conflicting reports with respect to the acetylated state of lysine 12 (K12) of histone H4.] Most likely this hypoacetylated state is achieved by the deacetylase activity of Sir2p, which can remove the acetyl groups from acetylated histone tail peptides *in vitro* (reviewed in ref. 11). The hypoacetylated state may be critical for Sir3p binding to chromatin. *In vitro*, a fragment of the Sir3 protein binds with greater affinity to histone H4 tail peptides that are completely unacetylated at K5, K8, K12, or K16, compared with those that are fully acetylated at these positions (12). Of these four lysines, *in vivo* analysis has implicated K16 as the "key" lysine in telomeric silencing (reviewed in ref. 13). From these studies, it has been interpreted that acetylation of K16 prevents silencing, i.e., precludes binding of Sir3p to chromatin. However, peptides with K16 as the sole site of acetylation still bind exceptionally well to Sir3p *in vitro* (12).

For histone H3 less is known and the situation is even more complex, in part because methylation and phosphorylation also occur within the tail (reviewed in ref. 13). The lysine residues in its NH$_2$-terminal tail are hypoacetylated when in silent chromatin as well, and when K9, K14, K18, and K23 all are mutated to arginine or glycine, telomeric silencing is reduced, although not to the extent that the same change at K16 on histone H4 reduces silencing (8). Thus, there is still ambiguity as to what posttranslational histone modification "code" is required for telomeric silencing.

Although a basic understanding of telomeric silent chromatin structure has been developed, its epigenetic character remains a mystery. In particular it is unclear how the heritable propagation of a state occurs within the context of a certain level of switching between expression states. The switching can be explained in part as the result of shifts in a competition between silencing components and transcriptional-activating factors for assembly onto telomere-proximal DNA (14). Despite this competition, examination of a colony of cells with *ADE2* at a telomere indicates that under normal circumstances the preexisting transcriptional state is most often inherited (see Fig. 1*A*). This finding indicates the existence of some mechanism to favor the status quo through successive cell cycles. In particular, assembly (or reassembly) of the silent chromatin must occur during, or shortly after, each round of DNA replication. Consistent with this idea, a number of unrelated mutations or drug treatments that lengthen S phase, and presumably affect the kinetics and

coordination of molecular events in S phase, are able to suppress defects in silencing (15, 16). Furthermore, telomeric silencing is sensitive to mutations in subunits of chromatin assembly factor 1 (CAF-1) or in ASF1; both facilitate assembly of newly replicated DNA into nucleosomes *in vitro* (17–20). Hence there appears to be an intimate coordination between silent chromatin assembly and DNA replication.

The role of histone modification as a means of "marking" chromatin to perpetuate the molecular memory of an expression state after DNA replication was suggested 25 years ago (21) and is likely to be germane to telomeric position effect in *S. cerevisiae*. Gene activation proceeds by a series of steps that includes recruitment of histone acetylases, which help produce the hyperacetylated state, and presumably prevents Sir3p binding (reviewed in ref. 22). Conversely, silent chromatin contains the Sir2p deacetylase and helps maintain Sir3p binding (5). Thus the acetylated state of histones could serve as the molecular mark that serves as "memory" for propagating the transcriptionally active state against a competing silencing complex, which once established, would maintain a deacetylated histone (reviewed in ref. 23). Other models for the heritable transmission of a chromatin state that include DNA methylation are likely irrelevant, as such modifications have not been detected in *S. cerevisiae* (24, 25). Similarly, those models that include methylation of specific histone residues such as K9 of histone H3 (a modification that is presumably irreversible), which then serve as binding sites for "silencing factors" such as HP1, are also unlikely; neither this histone modification, nor HP1 homologs, have been identified in *S. cerevisiae* telomeric chromatin (26).

Given the apparent complexity of interactions with histone H3 and H4 in the formation of telomeric chromatin, we sought to identify potential sites of reversible modification that might be particularly critical in imparting memory to a telomeric gene. That is, those residues and modifications that are responsible for the heritable bias of a telomeric gene's transcriptional state to be passed on from one cell to its progeny. Here we present our initial findings that combine a genetic screen and development of mass spectrometric methods to analyze these modified sites.

Materials and Methods

Plasmids, Oligos, and Strains. A 2.7-kb *Pst*I fragment of HHT2–HHF2 was inserted into *Pst*I sites of pRS317 and pRS314 to create pMP9 and pMP3, respectively (27). Histone H3 and H4 mutants were generated in three ways (see *Supporting Materials and Methods*, which are published as supporting information on the PNAS web site, www.pnas.org for details).

The yeast strain BY4705 was modified to create UCC1373 [*MAT**a** ade2Δ::hisG his3Δ200 leu2Δ0 lys2Δ0 met15Δ0 trp1Δ63 ura3Δ0 adh4::URA3-TEL(VII-L) ADE2-TEL(VR) hhf2-hht2Δ::MET15 hhf1-hht1Δ::LEU2 ppr1Δ::KanMX*]. UCC1371 is isogenic with UCC1373 but also contains *cac1Δ::HIS3*. Both strains use pMP9 (*LYS2* CEN ARS)-*HHF2-HHT2* as a covering plasmid to provide functional copies of histones H3 and H4.

Screening Procedure. Mutant histone H4 plasmids (pMP3) were transformed into UCC1371 as described (18). Pink/red colonies were picked and restreaked onto yeast complete media (YC)-Trp and replica-plated onto YC-Trp and YC-Lys to identify colonies that had lost pMP9. Plasmids were rescued from Lys$^-$/Trp$^+$ colonies and transformed into UCC1371 and UCC1373 to retest *ADE2* and *URA3* silencing. Plasmids were then sequenced to identify mutations in both *HHF2* and *HHT2*.

All silencing assays were carried out as described (28).

Transcriptional Microarrays. Isolation of mRNA, labeling, hybridization, and data analysis were carried out as described (29). GENESPRING software (Silicon Genetics, Redwood City, CA) was

used to analyze all transcript array data. A two-step process identified those loci indicated in Fig. 3. First, for each mutant allele, all genes that were calculated to be significantly different (>2 SD) than wild type were selected. Then the genes were ranked and those that were down-regulated the most were identified.

Histone Purification. Histones were isolated according to Waterborg (30). Histones were then separated on a Zorbax reverse-phase C-18 column with a Waters HPLC system using a 0.5%/min gradient from 40% to 60% acetonitrile (ACN)/0.1% trifluoroacetic acid (TFA). Histone H4 coeluted with H2A at ≈43% ACN/0.1%TFA and histone H3 eluted at ≈50% ACN/0.1%TFA. Purified histones were then dried with a speed-vac and stored at either 4° or −20°.

Chemical Acetylation/Trypsin Digestion. Purified and dried histone samples were resuspended in 50 μl deuterated acetic acid (Acros, Fair Lawn, NJ). Five microliters deuterated acetic anhydride was added, and the samples were left at room temperature for 6 h. Samples were then dried in a speed-vac, resuspended in 20 mM NH_4HCO_3 and 100 ng trypsin (Worthington), and incubated at 37° overnight. Digestion buffer was removed by speed-vac before mass spectrometric analysis.

MS and Analysis. MS and analysis are described in additional *Materials and Methods*, which are published as supporting information on the PNAS web site.

Results

Genetic Screen to Identify Histone H3 and H4 Mutants That Increase Telomeric Silencing.
As noted above, histones H3 and H4 are known to play a critical role in telomeric silencing, and their NH_2-terminal tails appear to be hypoacetylated in silenced chromatin (9). Conversely, the tails become hyperacetylated when a telomeric gene becomes transcriptionally active. We wanted to identify mutations in histones H3 and H4 that would increase the probability that a silenced state would occur, with the idea that these mutants might "lock" the histone into a posttranslational modification that was critical for silencing. The corresponding modified residue would be a candidate for a site on the histone in which the memory of a silenced or active transcriptional state might be stored.

We created and used *S. cerevisiae* strains in which the two chromosomal copies of the histone H3 and H4 genes were deleted and replaced with a wild-type copy of the histone genes (*HHT2-HHF2*) on a centromere plasmid containing a *LYS2* gene (pMP9; Fig. 1C) (27). In addition, the *ADE2* and *URA3* genes were integrated adjacent to telomeres on the right arm of chromosome V and left arm of chromosome VII, respectively (Fig. 1B) (18). In cells that were otherwise wild type, these markers served as independent reporters of telomeric position effect (28). The cells give rise to colonies that are red and white sectored because of variegated expression of *ADE2* and that grow poorly, or not at all, in the absence of uracil. [The transcriptional activator of the *URA3* gene, *PPR1*, is deleted, effectively crippling the *URA3* promoter and preventing it from being expressed when it is proximal to the telomere (14).]

To screen for histone mutants that would increase the frequency of telomeric silencing, the strain was "sensitized" by deletion of the *CAC1* gene (19, 20). This gene encodes a subunit of the chromatin assembly factor (CAF-I), which is a chaperone for histones H3 and H4 and likely helps in chromatin assembly. Loss of the gene product results in a loss of silencing. We hypothesize that in the absence of Cac1p, histones H3 and H4 are still deposited on the chromatin, but that they may be inappropriately modified on their way to deposition. Such modifications then prevent proper formation of silent chromatin. Thus we

Table 1. Mutations in the core region of histone H3 and H4 that increase telomeric silencing

Histone	Wild-type residue	Mutant residues isolated
H3	D77	A, G, N, V
	D81	G
H4		
	R39	K
	H75	Y

created a strain (UCC1371) that was defective in telomeric silencing (white colonies that could grow in the absence of uracil); this strain was used to screen for histone mutants that bypass the *cac1* defect and reinstate telomeric silencing (red and white sectored colonies and poor growth in the absence of uracil; Fig. 1 *B* and *C*).

Libraries of histone H3 and H4 mutants were transformed into the yeast strain described above (UCC1371), and the plated colonies were screened for increased redness (silencing of the telomeric *ADE2* gene) compared with colonies containing unmutagenized wild-type histones. Only colonies that had lost the wild-type histone/*LYS2* plasmid, which indicated that the mutant histones could replace the wild-type histones, were selected and examined further (see Fig. 1C). These strains were then tested for their ability to grow in media lacking uracil to address whether the telomeric *URA3* was also silenced. After verifying that the increase in telomeric silencing was plasmid-linked, the plasmids were sequenced to identify changes in histone H3 or H4. Those plasmids that had a single point mutation are presented (Table 1 and Fig. 2A).

The mutations were classified into four categories: those mapping to either the tails of histone H3 (Q5, A15, S22, and A24) or histone H4 (G7, L10, G11, G13, and A15) (see Fig. 2), or to the "core" of histone H3 (D77 and D81) or histone H4 (R39 and H75) (see Table 1). The mutant histones still required *SIR* gene function for silencing (data not shown), indicating that the mutant histone repression was still mediated by silent chromatin. We focused on the mutants within the histone H4 tail because of the documented role that acetylation of lysines within this domain plays in telomeric silencing, and because this tail was amenable to biochemical analysis of the modifications (described below).

Residues Adjacent to Histone H4 Tail Lysines Are Involved in Telomeric Silencing.
We chose two alleles at random from each of the mutated residues in the histone H4 tail (circled residues in Fig. 2A) for further characterization. The ability of these mutants to increase telomeric silencing at the *URA3* gene was quantified. Fig. 2B shows the fraction of cells that gave rise to a colony in the absence of uracil, which reflects the level of expression from *URA3*. As expected, the presence of *CAC1* in strains containing wild-type histones caused a significant reduction in the ability of cells to grow in the absence of uracil. However, each of the mutants increased silencing even further, regardless of whether or not *CAC1* was present. Mutants G7I, G7M, L10F, G13L, and G13A had the largest increase in silencing (≈50- to 500-fold reduction in uracil growth compared with wild type), whereas L10P, G11T and G11L, A15T, and A15V allowed a little more expression (≈10- to 40-fold reduction compared with wild type). In general, when *CAC1* was introduced into a histone mutant strain, silencing was increased further, with the possible exceptions of alleles L10F, L10P, and A15T, where no statistically significant difference was measured between *CAC1* and *cac1* strains. The two mutants that exhibited the most consistent *CAC1*-dependent increase in silencing were G11L and G11T (10- and 8-fold, respectively). Taken together these results indicated
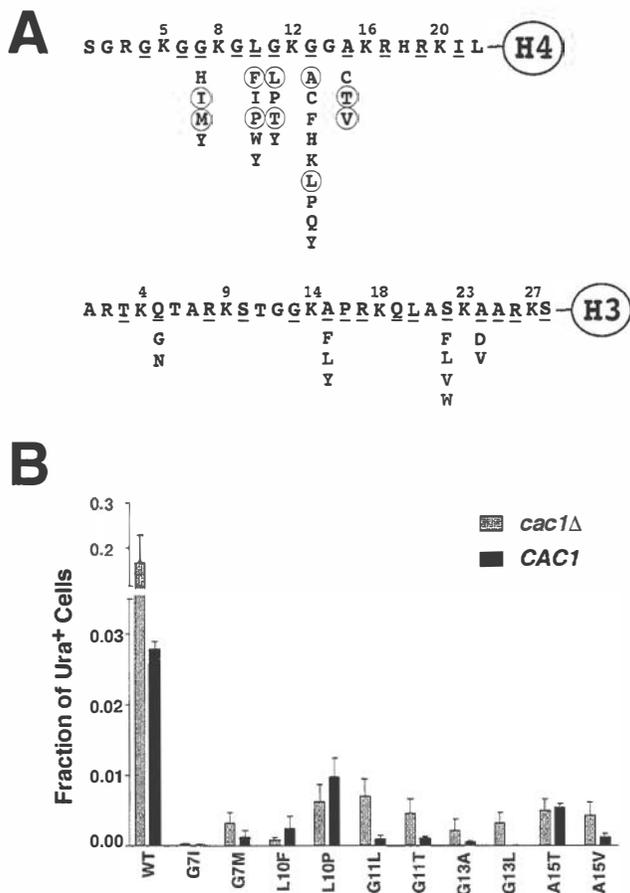
A

```
       5       8         12      16    20
  S G R G K G G K G L G K G G A K R H R K I L — (H4)
           H    (F)(L)    (A)        (C)
           (I)  (F)I  P   (C)        (T)
           (M)  (P)(T)    C          (T)
           Y    W   Y     F          (V)
                W         H
                          K
                          (L)
                          P
                          Q
                          Y

       4        9       14     18      23      27
  A R T K Q T A R K S T G G K A P R K Q L A S K A A R K S — (H3)
         G              F            F    D
         N              L            L    V
                        Y            V
                                     W
```

B

Fig. 2. Changes in amino acids adjacent to lysines in the tails of histone H3 and H4 increase telomeric silencing of marker genes. (A) Representation of the NH2-terminal tails of histones H3 and H4. Those residues that were specifically targeted for mutagenesis are underlined. Single amino acid changes that gave rise to increased silencing are indicated below the wild-type sequence. Those changes that were subjected to further analysis are circled. (B) Silencing of URA3 was measured by the fraction of cells in each strain that could form colonies on media lacking uracil for each of the histone H4 tail mutants circled in A. Gray and black bars represent data using cells from UCC1371 (cac1Δ) and UCC1373 (CAC1), respectively. Three independent transformants were used for each measurement.
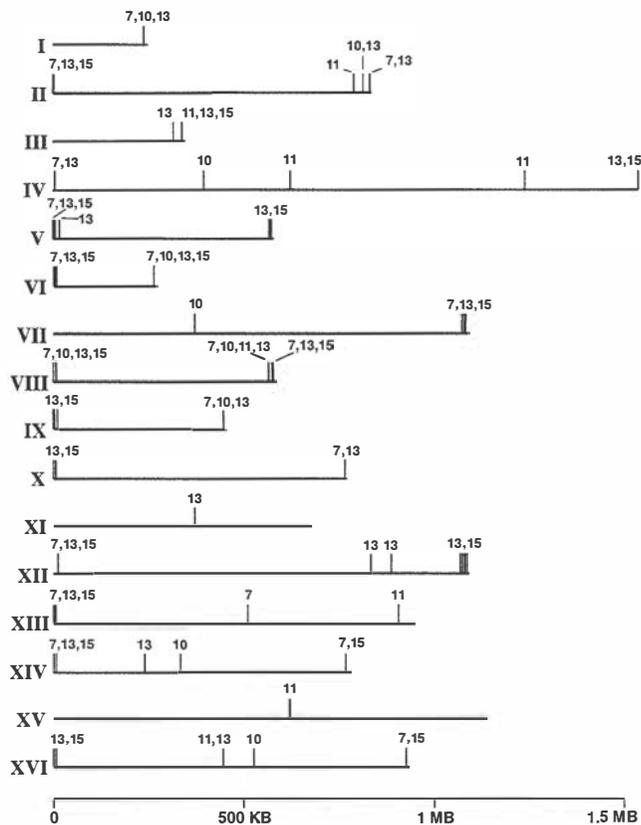

Fig. 3. Mutations in the histone H4 tail result in a general down-regulation of telomere proximal genes. Transcriptional changes between strains with mutant and wild-type histone H4 were determined by competitive hybridization to DNA microarrays. Each array contained >6,000 S. cerevisiae ORFs, and each hybridization was carried out a minimum of two times. A schematic of each S. cerevisiae chromosome is presented. Vertical lines above each chromosome mark loci that were transcriptionally down-regulated in the mutant strain. Numbers above these lines correspond with residue position number (e.g., G11T) in the mutants that showed an effect. For alleles L10F and G11T, the 10 loci that were down-regulated the most are indicated, for G7I and A15T ≈20 loci and for G13L ≈30 loci are shown.

that these 10 mutant alleles of histone H4 completely bypassed the need for CAC1 in telomeric silencing.

However, it was possible that the histone mutants increased silencing indirectly; for instance, they may have resulted in the misregulation of silencing proteins. Increased dosage of SIR3 causes increased silencing at telomeres and can compensate for some mutants that lose telomeric silencing (31). Currently there are fewer than 200 genes that have been implicated in telomeric silencing, according to the Yeast Proteome Database (www.incyte.com/proteome/databases.jsp). To examine whether these genes were differentially regulated by the mutant histones, a transcriptional microarray analysis was performed on one representative allele from each mutant site. For A15T, G13L, G11T, and L10F alleles, there was no significant (2-fold) increase or decrease in any of these genes (Table 2, which is published as supporting information on the PNAS web site). Only in the G7I allele was an effect observed: a 2-fold increase in expression of the EST2 and SAS4 genes. EST2 encodes the reverse transcriptase of telomerase (32), and SAS4 encodes a subunit of the SAS histone acetylase complex that is required for telomeric silencing (33). The overexpression of EST2 has been determined to have

no effect on telomeric silencing (34). The modest increase in SAS4 may impact telomeric silencing if this subunit is normally a limiting component, or if it is critical for targeting of the complex to telomeres.

The transcript microarray data also permitted us to map, on a genomewide scale, loci that were transcriptionally repressed (in addition to the telomeric ADE2 and URA3 genes) in the histone H4 mutants relative to wild type. Quite strikingly, when the genes were ranked by their expression level for each mutant compared with wild type, virtually all of the 20–40 genes that were down-regulated the most in the G7I, G13L, and A15T alleles were located in close proximity (within 20 kbp) to telomeres (Fig. 3 and Table 2). This bias was maintained in the L10F allele, although to a lesser extent; six of 10 genes were telomeric. Although telomeric loci were repressed in the G11T allele as well, they represented fewer of those most affected; only two of the 10 most repressed genes. Taken together, these data indicate that the increase in silencing on the two telomeric marker genes (ADE2 and URA3) caused by these mutant histone H4 alleles reflect an effect that occurred on telomeric chromatin in general.

Acetylation Analysis of Histone H4 Silencing Mutants by MS. Because the results of our screen pinpointed residues in proximity to the

histone H4 lysines at positions 5, 8, 12, and 16 as being important in controlling telomeric silencing, we suspected that acetylation of these residues in the mutants might be altered. Although antibodies to specific acetylated lysines have been used to determine levels of acetylation (9), we found these reagents to be unsuitable for our analysis. The mutations in histone H4 we created changed the recognition epitope for these antibodies. Furthermore, it is unclear how acetylation at a nearby lysine changes the affinity of the antibody for its epitope, thus making it difficult to truly quantify the level of acetylation at each lysine. Therefore we sought to eschew these issues by developing a mass spectrometric approach to measure the level of acetylation at each individual lysine residue in the tail of histone H4. Presented in Fig. 4A is an outline of this method. After initial isolation by standard procedures (30), purified histone H4 was prepared by reverse-phase HPLC (see *Materials and Methods*). The primary structure of the resulting population of histone H4 molecules was likely complex because any or all of the four tail lysine residues at positions 5, 8, 12, and 16 might be acetylated. To prepare a homogeneous, chemically identical population, purified histone H4 was treated with deuterated acetic anhydride, which labeled each unacetylated lysine with a deuterated acetyl moiety (35).

Trypsin efficiently cleaves after lysine or arginine in a peptide; however, acetylation of lysine protects these residues from cleavage (36). With each of the histone H4 lysines masked by either a protonated acetyl from endogenous acetylation or a deuterated acetyl from the chemical acetylation, digestion with trypsin produced peptides for which cleavage occurred only after arginine residues. For histone H4, the four acetylatable lysine residues within the NH_2 terminus, K5, K8, K12, and K16, remained together in a single peptide (residues 4–17) (data not shown).

Tandem MS was used to determine the endogenous level of site-specific acetylation at each lysine within the mutant histone H4 tail peptides (see Fig. 2A). This approach capitalized on the fact that the acetyl groups added *in vivo*, in the yeast cell, have a mass 3 Da less than the acetyl groups added *in vitro* by the reaction of the purified proteins with deuterated acetic anhydride. Tryptic products of histone H4 that corresponded to the fully acetylated peptide containing residues 4–17 were identified by HPLC electrospray ionization MS. Isotope patterns of this peptide indicated the extent of *in vivo* and *in vitro* acetylation in this segment (36, 37). Aided by "data-dependent" scanning, these molecular ions were fragmented by low-energy collision-induced dissociation (38), and the resulting b and y fragment ions were analyzed (39).

The level of *in vivo* acetylation at each lysine residue was determined from the relative intensities of isotope peaks caused by *in vivo* acetylation (Ac_H) and *in vitro* acetylation (Ac_D) for specific b and y ions. Because the peptide fragments used for these measurements were formed from chemically identical parent ions, fragmentation occurred independent of where *in vivo* acetylation occurred. The intensities of protonated and deuterated fragments ions were determined from the maximum ion currents of the appropriate isotope peaks of specific b and y ions (see *Materials and Methods*).

Measuring the level of endogenous acetylation at K16 was straightforward because it involved fragment ions (specifically, y3, y4, or y5 ions) that contained only this site of acetylation. Similarly, the level of acetylation at K5 was directly determined by using the fragment, b3. Quantification of acetylation at the internal lysine residues, however, was more complicated because b and y ions that include K8 and K12 also include K5 and K16, respectively (see Fig. 4B). Thus for fragments containing two sites of acetylation, there were three possible mass states: the ion can have two protonated acetylations (e.g., Ac_H at K12 and Ac_H at K16), one protonated acetylation and one deuterated acetylation (Ac_H at K12 and Ac_D at K16 or Ac_D at K12 and Ac_H at
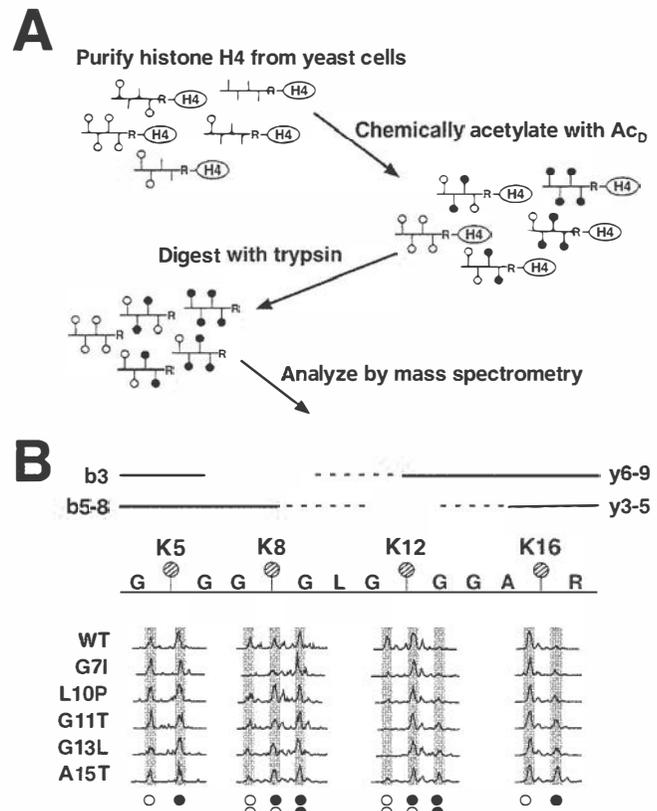

Fig. 4. Mass spectrometric analysis to measure acetylation at specific lysine residues in the histone H4 tail. (A) A schematic of the method used to determine site-specific acetylation of histone H4. Vertical lines represent lysines that can be acetylated, ○ represents those lysine residues acetylated *in vivo* (Ac_H), and ● represents lysine residues chemically acetylated by deuterated acetic anhydride *in vitro* (Ac_D); the mass difference between ○ and ● is 3 Da (42 versus 45, respectively). Tryptic digestion of purified and chemically modified histones produced a chemically identical population of tail peptides containing K5, K8, K12, and K16, beginning at G4 and terminating at R17. These peptides were then analyzed by MS. (B) Mass spectra showing altered levels of protonated versus deuterated acetylation in mutant and wild-type histone H4. A schematic of the histone H4 tryptic peptide is shown with each of the four lysine residues marked with a hatched circle. The lines (solid and dashed) above the peptide represent b and y ions. The b3 ion provided acetylation information for K5 and b ions 5-8 (b5–8) provided information for K8; similarly, y ions 3–5 (y3–5) were used to assess acetylation at K16 and y ions 6–9 (y6–9) were used for K12. (The b4 and y2 ions were indistinguishable in the spectra and were therefore omitted from analysis; the mass of the b2 ion was below the detection range.) Spectra of various b and y ions are shown for wild-type histone H4 along with mutants G7I, L10P, G11T, G13L, and A15T. The *x* axis and *y* axis of the mass spectra represent mass-to-charge ratio and relative abundance, respectively. Peaks are highlighted by shaded boxes and are aligned above symbols denoting whether the lysines are acetylated with a protonated (○) or deuterated (●) acetyl. ○○ represents fragment ions containing two lysine residues that are both modified with protonated acetyls; ○● represents fragments with one protonated and one deuterated acetyl; ●● represents fragments in which both acetyls are deuterated.

K16), or two deuterated acetylations (Ac_D at both K12 and K16). Thus, although we could determine the total relative number of protonated acetylations for a particular fragment, we could not tell directly how much acetylation is at either of the internal lysines. Given that we knew the acetylation level at one site, we could calculate the level at the other. By subtracting the contribution of the external site, information we obtained by analyzing fragments y3–5 or b3, the level of acetylation exclusively at the internal lysine was determined (see *Materials and Methods*).
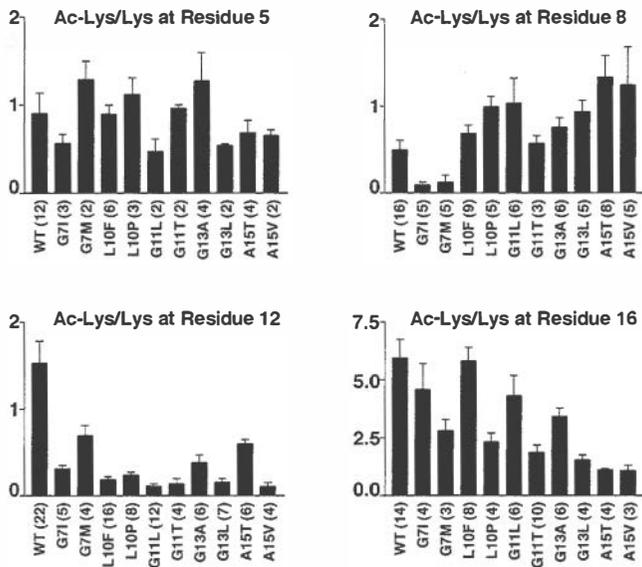
Fig. 5. Mutant histone H4 proteins exhibit decreased acetylation at specific lysine residues. The ratio of acetylated lysine to unacetylated lysine at positions 5, 8, 12, and 16 was determined for histone H4 mutants compared with wild-type histone H4. Levels of acetylation were calculated by comparing peak intensities of protonated vs. deuterated ions for specific y and b fragments (see *Materials and Methods*). Multiple b and y ions were used to calculate the fraction of acetylation at each site; the total number of intensity measurements, which is the sum of all ions measured from independent histone preparations, is shown in parentheses.

Presented in Fig. 4B are mass spectra of wild-type and selected mutant histone tryptic peptides; the spectra are placed beneath the lysine residue for which they provide acetylation information. There were two possible mass states for y3–5 ions: one at +42 (a single Ac_H represented by ○) and one at +45 (a single Ac_D represented by ●). In wild-type histone H4, the level of acetylation at K16 was very high, with only a small amount of deuterated acetylation. However, mutations close to K16 correlated with an increase in the amount of deuterated acetylation (note the change in peak areas in Fig. 4B). For the A15T mutant, there were equal levels of protonated and deuterated acetylation, indicating the level of endogenous acetylation dropped from ≈90% to ≈50%. Similarly, the amount of protonated acetylation dramatically decreased at K12 and K8, particularly for mutants containing amino acid changes near these positions. (See mutants G11T and G13L for K12 acetylation and mutant G7I for K8 acetylation.)

Summarized in Fig. 5 are the results of the quantitative mass spectrometric analysis of histone acetylation for each of the four lysine residues of histone H4. The data, presented as the ratio of acetylated lysine to unacetylated lysine, were based on the analysis of numerous b and y ions from multiple histone preparations; the number of data points that contributed to each set of data are marked in parentheses. The most dramatic observation was that acetylation of K12 was diminished in each of the mutants with the most drastic change occurring for mutants G11L, G11T, and G13L. However, in general, we found that mutation of a residue adjacent to a specific lysine decreased the ability of that lysine to be acetylated. For example, in addition to the decrease in acetylation at K12, G7I and G7M showed a striking decrease in the level of acetylation of K8, and A15T and A15V reduced acetylation at position 16. Acetylation of K5 appeared to be relatively unaffected by the histone tail mutations analyzed.

Discussion

We have identified 13 residues on histones H3 and H4 that are important in telomeric silencing by screening for mutants in these histone genes that bypass the loss of telomeric silencing phenotype of a *cac1Δ* strain. Single amino acid changes in the NH_2-terminal tails of histone H3 and H4, and within their core regions, reinstated telomeric silencing. Interestingly, mutations mapping to the core of these histones suggest another region for potential interaction between nucleosomes and the SIR complex.

In principle, there are several ways in which these mutants may be acting. They may increase the affinity of the SIR complex for the nucleosomes, thus stabilizing the silent chromatin state against competition by gene activation pathways. Such stabilization itself may proceed by several routes. Some may be very direct; a new ionic or hydrogen bond may be formed between the new residue and the SIR complex. This may be true for three of the changed residues, for which single amino acid substitutions were found. However, for the other 10 in which at least two substitutions were identified, each with very different side chains, this scenario is unlikely. Rather these residues may increase silencing by preventing the transcriptional activation pathway from competing effectively. In particular they may prevent the activation-associated, posttranslational modifications that diminish the ability of silent chromatin to form. Indeed, as discussed below, our data suggest that this is the case for the alleles that mapped to the tails of histone H4, where we were able to determine that the acetylation of lysines was reduced. Another possibility is that the amino acid substitutions may increase silencing by stabilizing the nucleosome, perhaps reducing the ability of the nucleosome to be "remodeled" for gene activation (22)—in essence these alleles would not increase telomeric silencing *per se*, but rather make activation more difficult. Also, it is unclear whether the histone H3 and H4 mutants are critical for increasing silencing by acting in steps of silent chromatin assembly and/or maintenance. But given that they all bypassed the need for *CAC1*, it is likely that they may influence a step in assembly, at least in part.

How then might these specific alleles be acting? Here we can gain potential insights from studying the structure of the yeast nucleosome (40). For example, R39 of histone H4 is located within the core of the nucleosome at an interface with the DNA. The R39K allele in histone H4 may cause the interaction between histone H4 and the DNA to change within the nucleosome. Thus this allele may increase silencing by enhancing the stability of the nucleosome. H75Y also maps to the core of histone H4; it, too, may be acting to stabilize the nucleosome. However, H75 is also located in proximity to the histone H3 loop that contains K79 and could be acting through this region of the nucleosome (see below).

The mutations in D77 and D81 of histone H3 are particularly intriguing, as they map on either side of K79, a residue we recently identified to be methylated by the methyltransferase Dot1p (41). This residue maps to the top and bottom surfaces of the nucleosome and has the potential to be another binding site for the Sir complex on the nucleosome. The methylation of K79 modulates the level of telomeric silencing; methylation seems to reduce binding of Sir proteins to the nucleosomes. Thus the mutant alleles at D77 and D81 may be acting either by reducing or eliminating Dot1p-mediated methylation of K79. Alternatively, they may change the surface of the nucleosome in a way that mitigates the effects of K79 methylation and instead facilitates Sir binding.

In the NH_2-terminal tail of histone H3, residues adjacent to all six lysines (K4, K9, K14, K18, K23, and K27) were heavily mutagenized. However, the only mutations that increased silencing were residues adjacent to K4, K14, or K23. Based on our

findings at the histone H4 tail, we suspect that posttranslational modification of these lysines was affected. At K14 and K23, a decrease or lack of acetylation is the probable effect, whereas at K4 it is likely to be a decrease in methylation (reviewed in ref. 13). The importance of K14 and K23 in telomeric silencing was also found when cells lacked the histone H4 acetylase Hat1p (27), thus supporting the significance of the results reported here. Given that mutations in *SET1* result in a decrease in telomeric silencing (42) and that Set1p mediates methylation of K4 (43–46), it is likely that the ability of Set1p to act on K4 is changed in the Q5G and Q5N alleles.

Directed mutagenesis of the residues adjacent to the five lysines (K5, K8, K12, K16, and K20) in the tail of histone H4 identified mutations only in those residues adjacent to K8, K12, and K16. K5 is reversibly acetylated, but apparently that has no consequence in telomeric silencing. No mutant alleles were identified at positions R17, R19, or I21. Given that virtually any change in the region encompassing residues 16–29 has been shown to result in a loss of telomeric silencing (8, 47, 48), it was not surprising that we did not identify mutant alleles in our screen for increased telomeric silencing.

The MS analysis after chemical acetylation by deuterated acetic anhydride allowed us to determine the relative level of acetylation at the four potentially acetylated lysines in the histone H4 tail. K16 was most dramatically underacetylated when A15 was changed, and K8 was similarly affected when G7 was mutated. This finding indicated that, on a global scale, these mutations prohibited or greatly reduced the efficacy of the major histone acetyltransferase(s) that act on these lysines. Although formally possible, it is doubtful that these mutations increased the efficacy of deacetylases that act on K8 or K16, because a variety of amino acid replacements at these residues (G7M, G7I, G7H, and G7Y; A15C, A15T, and A15V) all had the same phenotype.

However, the most striking result from the histone H4 tail mutants was the dramatic underacetylation of K12 in all of the mutants analyzed (Fig. 5). Returning to the idea that the acetylated status of a lysine might serve as a way for the chromatin state to be inherited, this result indicated that unacetylated K12 increased the probability that a silenced chromatin state would be inherited in a telomeric gene. We interpret this to mean that K12 is the site of a memory mark within the nucleosome for the heritable transmission of an active or silenced state in a telomeric gene.

If this idea is true, it has further implications with respect to histone H4 deposition in silent chromatin. It appears that shortly after synthesis of histone H4, it is acetylated at positions K5 and K12, most likely by the Hat1p histone acetylase (49–51). If an unacetylated K12 is needed for silent chromatin formation, then K12 must be rapidly deacetylated shortly after deposition onto the DNA in a silenced region of the genome. This may represent part of the silent chromatin maturation process that has been invoked to occur during, or shortly after, S phase (14, 52, 54).

Although the mass spectroscopic analysis of bulk histones has provided us with an intriguing insight about K12, the method is limited in that we have not determined the modified state of histones specifically within telomeric chromatin. We envision that the reduction in acetylation at specific lysines seen within total histone H4 mirrors the decreased likelihood that the lysine will be modified within telomeric chromatin as well. The application of this mass spectroscopic technology to isolated telomeric chromatin in the future will provide the greatest amount of information we seek. Nevertheless, by combining the analysis of bulk histones with the transcript microarray data for the five alleles of histone H4, in which we show there were no significant perturbations to expression of genes involved in telomeric silencing, we have alleviated the concern that the histone mutants increased telomeric silencing by indirect effects. Such indirect effects have plagued previous studies reporting the involvement of the *RPD3* and *HDA1* genes in telomeric silencing (55, 56). Mutants in these histone deacetylase genes changed the expression level of known silencing factors (57). Similarly, the earlier report of the requirement for *HAT1* in telomeric silencing (Hat1p acetylates K12 on histone H4) may also be indirect, particularly given that mutations in the histone H3 tail were required to uncover the *HAT1* dependency (27). Alternatively, the *HAT1* result may be caused by there being more than one pathway for depositing the histone H3/H4 tetramer into silenced chromatin (58) [both *ASF1*- and *CAC1*-dependent complexes are involved in telomeric silencing (17, 19, 20)]. Perhaps mutation of the histone H3 tail forces histone deposition to proceed via a pathway that relies on acetylation of K12.

The transcript array data showed a preponderance of telomeric loci in the list of genes down-regulated to the greatest extent in the histone H4 alleles, particularly the G7I, G13L, and A15T alleles (Fig. 3 and Table 2). Because none of the known silencing components were perturbed, this observation fortifies the hypothesis that these histone H4 mutants have a direct involvement in forming silent telomeric chromatin. It is also worth noting that the proportion of telomeric loci silenced was greatest in the A15T allele, which likely reflects that K16 also needs to be unacetylated for telomeric silencing. This finding suggests that both K16 and K12 may serve as key memory marks for silent chromatin. Because Sir-mediated silencing appears to require multiple weak interactions between the nucleosome and Sir complex, there are likely to be several such memory marks within the nucleosome. This finding also implies that there may not be a single place within the nucleosome where memory is stored, and we believe that our mutants will help to identify more of these marks.

One of the key strengths of this mutant hunt was that it allowed us to identify not only key residues that were modified, but also the particular modified state of the residue that appears to be relevant for telomeric silencing (i.e., acetylation states of lysines in histone H4). Previous attempts mutated the lysines and then inferred what the change meant (reviewed in ref. 13). The collection of mutants created in our study should also be useful in identifying those activities that are critical for switching a silenced gene to an active state and/or maintaining it as such. For instance, genetic screens designed to identify unlinked neomorphic alleles that could acetylate K12. Alternatively, the histone mutants could be used as a substrate in biochemical assays with histone acetylases and methylases that have been shown to be involved in telomeric silencing (44–46, 59). Ultimately it will be interesting to determine whether such factors are gene-specific or part of a general system of maintaining gene expression.

The analysis of these mutant histones may also define interdependency of modifications within the histones (reviewed in ref. 60). For instance, mutation of G7 in histone H4 confers reduced acetylation at K8 as well as at K12, which raises the possibility that unacetylated K8 may decrease the chance for K12 acetylation. That is, *in vivo*, there may be a hierarchy of acetylation along the histone H4 tail.

The application of mass spectrometric methods to quantitatively evaluate levels of histone modifications, as we have shown here, promises to enhance our understanding of chromatin structure regulation. We have found the combination of chemical isotopic acetylation of histones with trypsin cleavage and MS to be more robust and to require much less sample than other physical methods of examining acetylation at specific residues [i.e., Edman degradation (61)]. Furthermore, it provides greater flexibility in analysis than site-specific antibodies (53), which cannot bind to mutant histones. Future applications of this technology will lie in development of biochemical isolation procedures of locus-specific chromatin.

The combination of further analysis of the histone H3 and H4 mutants isolated in this study along with further development of mass spectroscopic methods will facilitate a greater understanding of how heritable states of telomeric genes in *S. cerevisiae* are propagated. These methods, and the paradigms that evolve from our studies, are likely to be applicable to a wide variety of systems in which chromatin states are responsible for propagating an epigenetic phenotype.

1. Muller, H. J. (1930) *J. Genet.* **22,** 299–324.
2. Henikoff, S. (1990) *Trends Genet.* **6,** 422–426.
3. Spofford, J. B. (1976) in *The Genetics and Biology of Drosophila*, eds. Ashburner, M. & Novitski, E. (Academic, New York), Vol. 1c, pp. 955–1018.
4. Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. (1990) *Cell* **63,** 751–762.
5. Moazed, D. (2001) *Mol. Cell* **8,** 489–498.
6. Loo, S. & Rine, J. (1995) *Annu. Rev. Cell Dev. Biol.* **11,** 519–548.
7. Grunstein, M. (1997) *Curr. Opin. Cell Biol.* **9,** 383–387.
8. Thompson, J. S., Ling, X. & Grunstein, M. (1994) *Nature (London)* **369,** 245–247.
9. Suka, N., Suka, Y., Carmen, A. A., Wu, J. S. & Grunstein, M. (2001) *Mol. Cell* **8,** 473–479.
10. Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M. & Broach, J. R. (1996) *Mol. Cell Biol.* **16,** 4349–4356.
11. Gottschling, D. E. (2000) *Curr. Biol.* **10,** R708–R711.
12. Carmen, A. A., Milne, L. & Grunstein, M. (2002) *J. Biol. Chem.* **277,** 4778–4781.
13. Richards, E. J. & Elgin, S. C. (2002) *Cell* **108,** 489–500.
14. Aparicio, O. M. & Gottschling, D. E. (1994) *Genes Dev.* **8,** 1133–1146.
15. Laman, H., Balderes, D. & Shore, D. (1995) *Mol. Cell. Biol.* **15,** 3608–3617.
16. Axelrod, A. & Rine, J. (1991) *Mol. Cell. Biol.* **11,** 1080–1091.
17. Tyler, J. K., Adams, C. R., Chen, S. R., Kobayashi, R., Kamakaka, R. T. & Kadonaga, J. T. (1999) *Nature (London)* **402,** 555–560.
18. Singer, M. S., Kahana, A., Wolf, A. J., Meisinger, L. L., Peterson, S. E., Goggin, C., Mahowald, M. & Gottschling, D. E. (1998) *Genetics* **150,** 613–632.
19. Kaufman, P. D., Kobayashi, R. & Stillman, B. (1997) *Genes Dev.* **11,** 345–357.
20. Enomoto, S., McCune-Zierath, P. D., Gerami-Nejad, M., Sanders, M. A. & Berman, J. (1997) *Genes Dev.* **11,** 358–370.
21. Weintraub, H., Flint, S. J., Leffak, I. M., Groudine, M. & Grainger, R. M. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42,** 401–407.
22. Narlikar, G. J., Fan, H. Y. & Kingston, R. E. (2002) *Cell* **108,** 475–487.
23. Turner, B. M. (2000) *BioEssays* **22,** 836–845.
24. Proffitt, J. H., Davie, J. R., Swinton, D. & Hattman, S. (1984) *Mol. Cell. Biol.* **4,** 985–988.
25. Bird, A. (2002) *Genes Dev.* **16,** 6–21.
26. Jenuwein, T. (2001) *Trends Cell Biol.* **11,** 266–273.
27. Kelly, T. J., Qin, S., Gottschling, D. E. & Parthun, M. R. (2000) *Mol. Cell. Biol.* **20,** 7051–7058.
28. van Leeuwen, F. & Gottschling, D. E. (2002) *Methods Enzymol.* **350,** 165–186.
29. Bedalov, A., Gatbonton, T., Irvine, W. P., Gottschling, D. E. & Simon, J. A. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 15113–15118.
30. Waterborg, J. H. (2000) *J. Biol. Chem.* **275,** 13007–13011.
31. Renauld, H., Aparicio, O. M., Zierath, P. D., Billington, B. L., Chhablani, S. K. & Gottschling, D. E. (1993) *Genes Dev.* **7,** 1133–1145.
32. Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V. & Cech, T. R. (1997) *Science* **276,** 561–567.
33. Xu, E. Y., Kim, S. & Rivier, D. H. (1999) *Genetics* **153,** 25–33.
34. Evans, S. K., Sistrunk, M. L., Nugent, C. I. & Lundblad, V. (1998) *Chromosoma* **107,** 352–358.
35. Riordan, J. F. & Vallee, B. L. (1967) *Methods Enzymol.* **11,** 565–570.
36. Steiner, R. F., Albaugh, S., Fenselau, C., Murphy, C. & Vestling, M. (1991) *Anal. Biochem.* **196,** 120–125.
37. Chen, J., Smith, D. L. & Griep, M. A. (1998) *Protein Sci.* **7,** 1781–1788.
38. Hayes, R. N. & Gross, M. L. (1990) *Methods Enzymol.* **193,** 237–263.
39. Biemann, K. (1990) *Methods Enzymol.* **193,** 886–887.
40. White, C. L., Suto, R. K. & Luger, K. (2001) *EMBO J.* **20,** 5201–5218.
41. van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. (2002) *Cell* **109,** 745–756.
42. Nislow, C., Ray, E. & Pillus, L. (1997) *Mol. Biol. Cell* **8,** 2421–2436.
43. Briggs, S. D., Bryk, M., Strahl, B. D., Cheung, W. L., Davie, J. K., Dent, S. Y., Winston, F. & Allis, C. D. (2001) *Genes Dev.* **15,** 3286–3295.
44. Krogan, N. J., Dover, J., Khorrami, S., Greenblatt, J. F., Schneider, J., Johnston, M. & Shilatifard, A. (2002) *J. Biol. Chem.* **277,** 10753–10755.
45. Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W., Wilm, M., Aasland, R. & Stewart, A. F. (2001) *EMBO J.* **20,** 7137–7148.
46. Nagy, P. L., Griesenbeck, J., Kornberg, R. D. & Cleary, M. L. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 90–94.
47. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. (1996) *Nature (London)* **383,** 92–96.
48. Johnson, L. M., Fisher-Adams, G. & Grunstein, M. (1992) *EMBO J.* **11,** 2201–2209.
49. Sobel, R. E., Cook, R. G., Perry, C. A., Annunziato, A. T. & Allis, C. D. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 1237–1241.
50. Kleff, S., Andrulis, E. D., Anderson, C. W. & Sternglanz, R. (1995) *J. Biol. Chem.* **270,** 24674–24677.
51. Parthun, M. R., Widom, J. & Gottschling, D. E. (1996) *Cell* **87,** 85–94.
52. Li, Y. C., Cheng, T. H. & Gartenberg, M. R. (2001) *Science* **291,** 650–653.
53. White, D. A., Belyaev, N. D. & Turner, B. M. (1999) *Methods* **19,** 417–424.
54. Kirchmaier, A. L. & Rine, J. (2001) *Science* **291,** 646–650.
55. Rundlett, S. E., Carmen, A. A., Kobayashi, R., Bavykin, S., Turner, B. M. & Grunstein, M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 14503–14508.
56. De Rubertis, F., Kadosh, D., Henchoz, S., Pauli, D., Reuter, G., Struhl, K. & Spierer, P. (1996) *Nature (London)* **384,** 589–591.
57. Bernstein, B. E., Tong, J. K. & Schreiber, S. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13708–13713.
58. Mello, J. A., Sillje, H. H., Roche, D. M., Kirschner, D. B., Nigg, E. A. & Almouzni, G. (2002) *EMBO Rep.* **3,** 329–334.
59. Marmorstein, R. & Roth, S. Y. (2001) *Curr. Opin. Genet. Dev.* **11,** 155–161.
60. Zhang, Y. & Reinberg, D. (2001) *Genes Dev.* **15,** 2343–2360.
61. Sobel, R. E., Cook, R. G. & Allis, C. D. (1994) *J. Biol. Chem.* **269,** 18576–18582.

Colloquium

Does heterochromatin protein 1 always follow code?

Yuhong Li*, Dawn A. Kirschmann†, and Lori L. Wallrath*‡

Departments of *Biochemistry and †Anatomy and Cell Biology, University of Iowa, Iowa City, IA 52242

Heterochromatin protein 1 (HP1) is a conserved chromosomal protein that participates in chromatin packaging and gene silencing. A loss of HP1 leads to lethality in *Drosophila* and correlates with metastasis in human breast cancer cells. On *Drosophila* polytene chromosomes HP1 is localized to centric regions, telomeric regions, in a banded pattern along the fourth chromosome, and at many sites scattered throughout the euchromatic arms. Recently, one mechanism of HP1 chromosome association was revealed; the amino-terminal chromo domain of HP1 interacts with methylated lysine nine of histone H3, consistent with the histone code hypothesis. Compelling data support this mechanism of HP1 association at centric regions. Is this the only mechanism by which HP1 associates with chromosomes? Interest is now shifting toward the role of HP1 within euchromatic domains. Accumulating evidence in *Drosophila* and mammals suggests that HP1 associates with chromosomes through interactions with nonhistone chromosomal proteins at locations other than centric heterochromatin. Does HP1 play a similar role in chromatin packaging and gene regulation at these sites as it does in centric heterochromatin? Does HP1 associate with the same proteins at these sites as it does in centric heterochromatin? A first step toward answering these questions is the identification of sequences associated with HP1 within euchromatic domains. Such sequences are likely to include HP1 "target genes" whose discovery will aid in our understanding of HP1 lethality in *Drosophila* and metastasis of breast cancer cells.

In eukaryotes, there are two major types of chromatin: heterochromatin and euchromatin (1). Heterochromatin corresponds to the relatively gene-poor, late-replicating, repetitious sequences found near centric and telomeric locations. In contrast, euchromatin replicates relatively early in the cell cycle and contains single copy sequences, including the majority of genes. Both euchromatin and heterochromatin are packaged into nucleosomes, the fundamental packaging unit consisting of a histone octamer. Euchromatin and heterochromatin can be distinguished by specific histone tail modifications. In general, the histone tails in heterochromatin are relatively hypoacetylated; however, acetylation of lysine twelve of histone H4 is a distinguishing mark for heterochromatin (2–4). In contrast, histone H3 and H4 tails found in euchromatin are generally acetylated (4). Histone H3 acetylation is often linked to H3 phosphorylation and is likely to represent a two-component code for high levels of gene expression (5, 6).

In addition to distinct differences in histone modification, euchromatin and heterochromatin show differences in nonhistone chromosomal protein constituents. One of the best-studied examples is heterochromatin protein 1 (HP1) first discovered in *Drosophila* and named for its predominant localization to centric heterochromatin (7) (Fig. 1*A*). Consistent with this localization, the gene encoding HP1, *Su(var)2–5*, was isolated as a dominant suppressor of position effect variegation (PEV) (8, 9). PEV is the mosaic pattern of expression exhibited by genes placed near centric heterochromatin by chromosomal rearrangements or transposition events (10). Overexpression of HP1 leads to enhanced silencing of variegating genes. Conversely, a decreased level of HP1 leads to reduced silencing of variegating genes. A

complete loss of HP1, as in homozygous *Su(var)2–5* null mutants, results in lethality. Larvae survive until the late third instar stage because of maternally supplied HP1 (11, 12). The cause of lethality is unknown. Given the centric localization of HP1, and the interaction between the *Schizosaccharomyces pombe* HP1-like protein Swi6 and a cohesion protein, chromosome segregation might be affected (13, 14). Thus, HP1 levels are critical for regulating the extent of heterochromatization within centric regions that is required for proper chromosome segregation.

In addition to centric regions, HP1 is observed at other regions of the genome known to be heterochromatic. The small fourth chromosome of *Drosophila melanogaster*, interspersed with heterochromatic domains, shows a banded pattern of HP1 localization (7, 15). Consistent with HP1 having a packaging function at these locations, transgenes inserted along the fourth chromosome exhibit PEV that is suppressed by *Su(var)2–5* mutations (15, 16). HP1 localization is also observed at *Drosophila* telomeres that terminate in repetitive arrays of retrotransposons (17). Telomeric association, however, appears to be independent of primary DNA sequence as broken chromosomes lacking terminal retrotransposons retain HP1 association (12). Telomere-telomere fusions occur in larval neuralblasts of *Su(var)2–5* mutants, suggesting HP1 plays a role in telomere capping (12).

In contrast to these chromosomal domains rich in repetitive DNA sequences, HP1 is present at approximately 200 sites within the euchromatic arms of polytene chromosomes that are relatively poor in repetitious DNA sequences. Do these sites represent small domains of repressive chromatin? Are there genes at these sites that are regulated by HP1? These questions are currently under investigation.

Here we describe current studies on the role of HP1 in gene regulation at both euchromatic and heterochromatic domains. We summarize the results from reports that have identified HP1 partner proteins and discuss implications for these findings. Last, we hypothesize about multiple mechanisms of HP1 chromosome association and their impact on gene expression.

HP1 Follows Code

HP1 is a highly conserved protein with family members found in a variety of eukaryotic organisms ranging from *S. pombe* to humans (18–21). In *Drosophila*, two additional HP1-like proteins, HP1b and HP1c, sharing amino acid sequence similarity and domain structure, have been identified (22) (Fig. 2). Whereas HP1b shows localization to both euchromatin and heterochromatin just as HP1, HP1c is found only in euchromatin (22). Mice and humans each have three HP1-like proteins that possess similarities in amino acid sequence, domain structure,
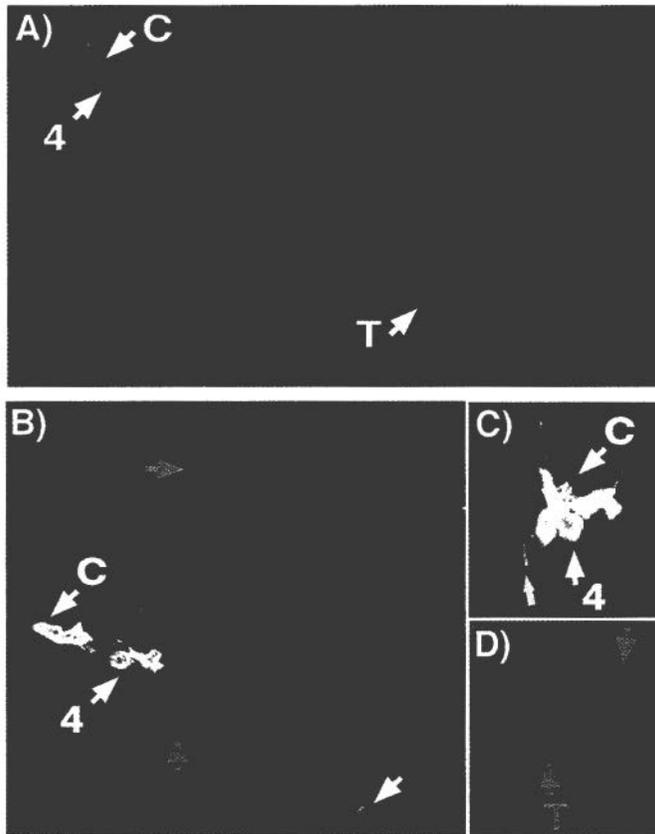
Fig. 1. (A) Pattern of HP1 distribution on *Drosophila* polytene chromosomes. *D. melanogaster* larval polytene chromosomes were stained with mouse monoclonal C1A9 antibodies against HP1 (gift of Sarah C. R. Elgin) and a secondary antibody conjugated with rhodamine. The chromocenter (C), the fourth chromosome (indicated by 4), telomeres (T), and euchromatic sites associated with HP1. (B) The pattern of HP1 and methylated lysine nine of histone H3 on *Drosophila* polytene chromosomes. *D. melanogaster* larval polytene chromosomes were stained with mouse monoclonal C1A9 antibody against HP1 and a rabbit polyclonal antibody that recognizes methylated lysine nine of histone H3 (gift of C. David Allis, University of Virginia, Charlottesville). A Cy5-conjugated rabbit secondary antibody and a FITC-conjugated mouse secondary antibody were used for detection. The chromocenter (C) and the fourth chromosome (indicated by 4) show strong colocalization (yellow). Example locations enriched in HP1 are denoted by green arrows; example locations enriched in methylated lysine nine of histone H3 are indicated by red arrows. (C) Same as in *B*, showing a closer view of the chromocenter region. (D) Same as in *B*, showing a closer view of a telomeric region.

and centric chromosomal localization properties as *Drosophila* HP1 (Fig. 2). Although there are minor differences in chromosomal localization and protein interaction partners for HP1-like proteins within a given species (23), it is not clear whether these proteins have specific or redundant functions. In flies, mice, and humans, the HP1-like proteins are small in size, ranging from 173 to 240 aa (Fig. 2). Overall the percent identity of HP1-like proteins to *Drosophila* HP1 is approximately 50% for mammalian HP1-like proteins. The majority of conserved amino acids are concentrated in two domains. The structure of HP1-like proteins can be simplified as two conserved domains separated by a less conserved hinge region (Fig. 2). The conserved amino-terminal region of HP1-like proteins is termed the chromo domain (CD) (24). This domain is present in 20 proteins in Flybase (www.ebi.ac.uk/proteome/DROME/interpro/stat.html), many of which play roles in gene regulation. The conserved carboxyl-terminal region, termed the chromo shadow

domain (CSD), is related to the CD in primary amino acid sequence (25). Both the CD and the CSD have been the subject of extensive structural analysis (26–30). Each domain forms a hydrophobic pocket. The CSD dimerizes (18, 27, 28, 31) as well as interacts with a wide variety of nuclear proteins (see below). Cross-species functional studies in which the mouse HP1-like protein M31 was expressed in *S. pombe* indicate that species-specific functions of HP1 reside within the CSD (32). The CD is required for chromosome association (33).

The mechanism(s) by which HP1 establishes the complex localization pattern on chromosomes remained a mystery for over a decade since its discovery. For many chromosomal proteins, localization is achieved through direct interaction with DNA sequences. Attempts to identify specific interactions between HP1 and DNA sequences, particularly repetitive DNA sequences found within heterochromatin, were not particularly revealing (34). For some chromosomal proteins, localization is achieved through interactions with DNA binding proteins. Therefore, a search for HP1 partner proteins might reveal the "missing link" between HP1 and the chromosome. A phage display assay was performed to identify peptides that interact with the CD and CSD (31). This assay revealed peptide sequences that showed a specific interaction with the CSD. Comparison of the peptide sequences allowed a consensus pentapeptide to be generated (31). Supporting these results, the consensus pentapeptide was found in several proteins shown to interact with HP1 by other types of assays (35–37). To date, the localization pattern of candidate interacting proteins cannot explain the entire localization pattern observed for HP1. In contrast to the results obtained for the CSD, no peptides were identified from the phage display assay that specifically interacted with the CD. These results were surprising because a point mutation within the CD of *Drosophila* HP1 eliminates the majority of chromosome association, suppresses PEV, and is homozygous lethal (9).

The mystery surrounding interactions of the CD was solved by studies of the murine Suv39h1 protein, a homologue of the *Drosophila* SU(VAR)3–9 protein (38, 39). A comparative genomic approach in combination with biochemical studies revealed that the SET [a conserved motif in *Drosophila* *Su(rar)3-9*, *Enhancer of Zeste*, and *trithorax*] domain of Suv39h1 contains methyltransferase activity specific for lysine nine of histone H3. This methylation mark on the histone H3 tail serves as a specific recognition code for the CD of HP1. This discovery supports the histone code hypothesis that proposes histone tail modifications serve as specific recognition motifs for chromatin proteins (40). The HP1 CD, but not the CD of several other proteins, binds methylated lysine nine of histone H3 (41). Therefore, the substrate specificity is likely caused by minor differences in the amino acid sequences of CDs from different proteins. The connection between Suv39h1 and HP1 is consistent with *Drosophila* research showing that the genes encoding HP1 and SU(VAR)3–9 genetically interact with the heterochromatin silencing system (8) and that the proteins physically interact (42). The relationship between HP1 and Suv39h1 has been maintained by the *S. pombe* homologues, Swi6 and Clr4, respectively (43, 44), suggesting evolutionary conservation in this mechanism of chromosome association and heterochromatic gene silencing.

In summary, HP1 serves as a bridging protein, connecting histones, through interactions with the CD, to nonhistone chromosomal proteins, through interactions with the CSD (Fig. 3*A*). In this case, Suv39h1 sets the histone code for HP1 association. Based on these findings, mechanisms for heterochromatin spreading have been proposed to involve recruitment of Suv39h1 by HP1 and propagation of the methylation mark along the chromosome (39). Details of such spreading mechanisms remain to be elucidated.
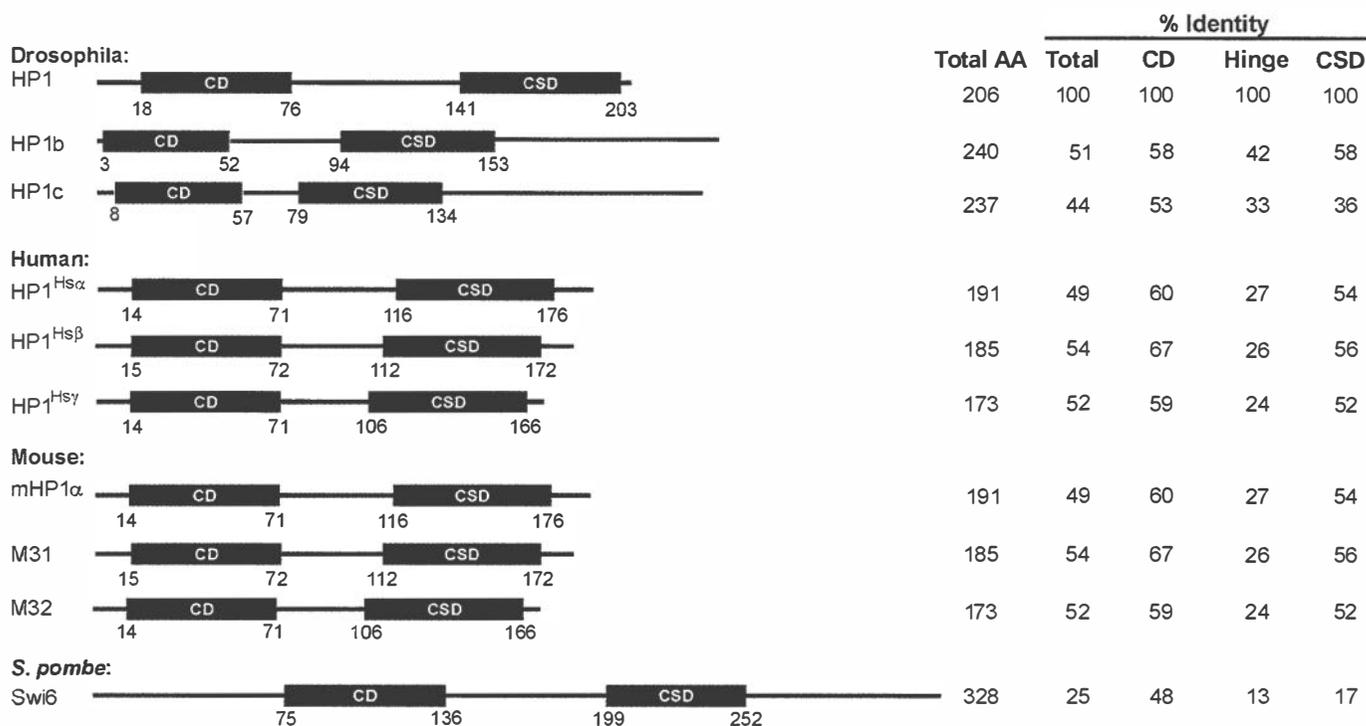
	Total AA	% Identity			
		Total	CD	Hinge	CSD
Drosophila:					
HP1	206	100	100	100	100
HP1b	240	51	58	42	58
HP1c	237	44	53	33	36
Human:					
HP1$^{Hs\alpha}$	191	49	60	27	54
HP1$^{Hs\beta}$	185	54	67	26	56
HP1$^{Hs\gamma}$	173	52	59	24	52
Mouse:					
mHP1α	191	49	60	27	54
M31	185	54	67	26	56
M32	173	52	59	24	52
S. pombe:					
Swi6	328	25	48	13	17

Fig. 2. Diagram of HP1 proteins in *Drosophila*, mouse, human, and *S. pombe*. Total length of each protein is indicated and drawn to relative scale. Percent identity when compared with *Drosophila* HP1 over the full length (total), or the CD, CSD, or the hinge region was calculated according to ref. 85.

Are There Multiple Mechanisms for HP1 Association?

HP1 localizes to distinctly different environments throughout the genome. Has the discovery of the interaction with the methylated lysine nine of histone H3 cracked the code, or are there alternative mechanisms of HP1 chromosomal association? The importance of this question is evident when reviewing data on the *Su(var)2-5^{02}* allele of the gene encoding HP1. This allele contains the amino acid substitution of a highly conserved valine to a methionine in the CD. Structural analysis indicates that this residue plays a critical role in the formation of the hydrophobic pocket of the CD (30). In most genetic silencing assays this allele behaves as an HP1 null; however, an important distinction between this allele and null alleles was revealed by a cytological analysis of HP1 staining on chromosomes from HP1 mutants. Whereas null alleles show no HP1 chromosome association, the *Su(var)2-5^{02}* allele shows diminished HP1 localization to centric regions, but retains association at euchromatic and telomeric sites (12). These data suggest an alternative mechanism of HP1 association might be operating at noncentric locations.

Further evidence for alternate mechanisms of association comes from cytological experiments on polytene chromosomes in wild-type flies. The pattern of staining observed by antibodies to HP1 and methylated lysine nine of histone H3 is not completely coincident. Both antibodies show colocalization to the chromocenter and along the fourth chromosome, but not throughout the euchromatic arms and at telomeric regions (Fig. 1 *B–D*) (45). One technical explanation for incomplete colocalization is that the epitopes recognized by either antibody are masked by fixation at specific genomic locations. However, if this is not the case, sites within the euchromatic arms that stain with only the HP1 antibody could be generated by HP1 associations through mechanisms independent of SU(VAR)3–9. Interactions of HP1 with unmodified histone tails, the histone-fold domain, and histone H1 might account for the staining pattern observed (34, 46). Such possibilities are diagrammed in Fig. 3*B*. Alternatively, in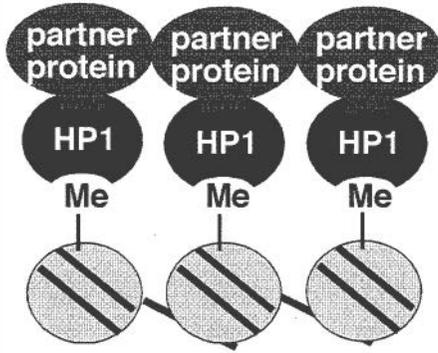teractions with nonhistone chromosomal proteins might serve as an additional mechanism of association (Fig. 3*C*). Interactions between HP1 and transcriptional corepressors that associate with DNA binding proteins (see below) (47, 48) support this hypothesis. The double staining also revealed sites within the euchromatic arms that are detected only by the methyl lysine nine histone H3 antibody. These sites could correspond to different degrees of methylation because the antibody recognizes only dimethylated lysine (Upstate Biotechnology, Lake Placid, NY); HP1 is thought to recognize both methylated states with relatively equal affinity (30). Another explanation for lack of complete colocalization of HP1 and the methyl lysine nine histone H3 antibody is that additional histone modifications might be present that do not permit HP1 association (Fig. 3*D*). Clearly the code for chromosomal protein association might have multiple components.

## HP1 Interacts with a Myriad of Proteins

Does the identification of proteins that associate with HP1 provide clues about the mechanisms of silencing? Genetic analysis of PEV in *Drosophila* provided a collection of mutations that encode candidate HP1 interaction partners: for example, SU(VAR)3–9, the histone methylase discussed above interacts with HP1 by two-hybrid analysis (42) (Table 1). A second example is SU(VAR)3–7, a zinc finger protein that associates with satellite DNA sequences (49). HP1 and SU(VAR)3–7 colocalize in the *Drosophila* embryo and on polytene chromosomes (50, 51) and interactions between the two proteins have been demonstrated by yeast two-hybrid analysis and coimmunoprecipitation from embryonic extracts (35, 50). More specifically, the CSD of HP1 interacts with multiple regions of SU(VAR)3–7, but it is not yet clear how these two proteins collaborate to form and/or spread heterochromatin.

In addition to a gene silencing function, HP1 is thought to play a role in nuclear organization. This hypothesis is based on the discovery that HP1 interacts with lamin B receptor, either directly (52, 53) or indirectly through interactions with histones
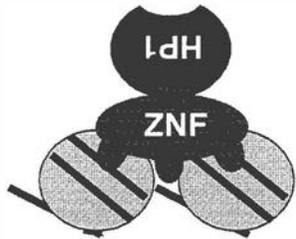
A) Interaction with methylated histones

B) Interactions with nucleosomes
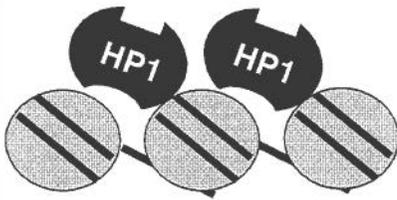
C) Interaction with DNA binding proteins

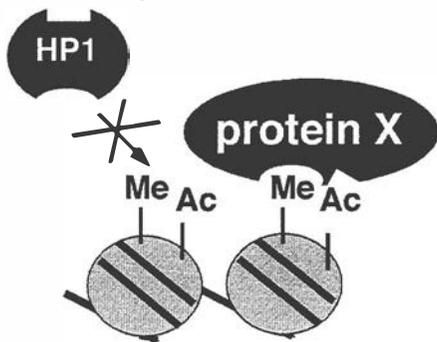# D) Exclusion by histone modifications

Fig. 3. Models for HP1 association and nonassociation with chromosomes. (*A*) Interaction between the HP1 CD with the methylated lysine nine of histone H3. HP1 serves as a bridge for partner proteins. (*B*) HP1 interacts with histones in a nonmethylated-dependent fashion. (*C*) HP1 associates with chromosomes through interactions of the CSD and DNA binding proteins, such as zinc-finger proteins (ZNF). (*D*) HP1 does not associate with methylated histones that have additional modifications such as acetylation or phosphorylation.

(54). In addition, experimental data support an interaction between HP1 and B-type lamin and Lap2β, lamin-associated protein, located within the nuclear envelope. *In vitro*, these interactions foster nuclear envelope assembly, suggesting HP1 plays a role in organizing nuclear architecture (55). Given that heterochromatin localizes to the nuclear periphery in many eukaryotic cell types, HP1 might tether heterochromatin to the nuclear envelope, leaving active regions of the genome free to coalesce into transcription factories within the interior of the nucleus (56).

The localization of HP1 to many sites throughout the *Drosophila* euchromatic arms brings to question the role of HP1 in the regulation of gene expression. Supporting a role for HP1 in transcriptional regulation, HP1 has been shown to interact with numerous proteins involved in modulating chromatin structure and gene expression (Table 1). In mammals, association of HP1, the retinoblastoma (Rb) protein and SUV39H1 with the *cyclin E* promoter correlates with gene silencing (57). Furthermore, HP1 has been implicated in gene repression mediated by Krüppel-associated box (KRAB) zinc finger proteins (47, 48). Taken together, these findings suggest that HP1 is recruited to specific genes by protein–protein interactions, resulting in gene silencing by an unknown mechanism.

In addition to transcriptional regulators, HP1 interacts with proteins involved in DNA replication and repair. Chromatin assembly factor 1 (CAF1) is a three-subunit complex that assembles histones H3 and H4 onto newly replicated DNA in both euchromatic and heterochromatic regions of the genome. In mammals, the large subunit, p150, contains a domain that interacts with the CSD of HP1 (37). Deletion of this domain does not alter CAF1-mediated chromatin assembly after replication *in vitro* or targeting of HP1 to heterochromatin *in vivo* during DNA replication. However, deletion of this domain reduces the amount of CAF1 present in heterochromatin outside of S phase. Although the significance of CAF1-HP1 interaction is not clear, the data suggest that CAF1 might stabilize heterochromatin structure during times of chromosome decondensation and transcription.

HP1 associates with origin recognition complex (ORC) proteins (58, 59). This presents an intriguing parallel to the situation in *Saccharomyces cerevisiae* where ORC proteins associate with silent information regulator (SIR) proteins to generate silent chromatin (60, 61). A high molecular weight complex isolated from *Drosophila* embryos was recently shown to contain HP1/ ORC-associated protein (HOAP) in addition to HP1 and ORCs. HOAP has sequence similarity to high mobility group proteins and binds to satellite sequences *in vitro* (62). Mutations in the genes encoding ORC proteins and HOAP are suppressors of PEV, suggesting a role in heterochromatin formation (62).

What regulates HP1 association with protein partners? Post-translational modifications are likely to govern interactions between HP1 and partner proteins and/or itself. In *Drosophila*, HP1 is multiply phosphorylated giving rise to at least eight differently charged isoforms (63). *Drosophila* embryonic extracts possess HP1-containing complexes that differ in HP1 phosphorylation status (58, 62). For example, hypophosphorylated HP1 is found in a complex containing ORC and HOAP (58, 62). In *Drosophila*, casein kinase II (CKII) is credited for the phosphorylation of serine residues at the amino and carboxyl termini of HP1. Mutation of these serine residues to alanine reduces the amount of HP1 localized to centric heterochromatin and reduces gene silencing, suggesting phosphorylation plays a role in chromosome association and/or complex stability (63, 64).

In mammals, HP1 phosphorylation changes through the cell cycle. HP1$^{Hs\alpha}$ and HP1$^{Hs\gamma}$ exhibit increased levels of phosphorylation during mitosis (23). HP1$^{Hs\gamma}$ is a substrate for Pim-1 kinase that phosphorylates a serine cluster in the center of the protein (65). Phosphorylation of HP1$^{Hs\alpha}$ is thought to disrupt protein–protein interactions that are necessary to maintain most of the centric localization. In G2, phosphorylated HP1$^{Hs\alpha}$ shifts from a centric location to being dispersed throughout the nucleus. Clearly, the role of phosphorylation needs further investigation to understand the biological significance of this dynamic process.

Table 1. HP1 interacting partners and candidate partners

Protein	Organism	HP1 variant	Methodology	HP1 domain	Ref(s).
Transcription regulation/ chromatin modifying proteins					
H1	*Drosophila*	HP1	rPD	nd	46
HP1-BP74 H1-like	Mouse	mHP1α	Y2H, FW, rPD	Hinge region	46, 86
H3	Mouse	mHP1α, M31, M32	FW, rPD, exIP	CD	46, 54
H3	*Drosophila*	HP1	rPD	nd	46
Methylated K9 of H3	*S. pombe*	Swi6	rPD, ChIP	CD	38, 43
Methylated K9 of H3	*Drosophila*	HP1	IF, FAITC, NMR	CD	26, 41
Methylated K9 of H3	Mouse	mHP1α, M31, M32	rPD	CD	39
Methylated K9 of H3	Human	HP1Hsα, HP1Hsβ, HP1Hsγ	rPD, SPRA	CD	38
H4	Mouse	M31	rPD	nd	54
H4	*Drosophila*	HP1	*In vitro* cross-linking	CSD	34
MacroH2A1.2*	Mouse	M31	IF	nd	87
SUVAR3–9	*Drosophila*	HP1	IF, Y2H, exIP	CSD	42
Suv39h1	Mouse	M31	IF, exIP, SED	nd	88, 89
SUV39H1	Human	HP1Hsβ	IF, exIP, SED	nd	88, 89
Suvar3–7	*Drosophila*	HP1	IF, Y2H, exIP	CSD	35, 50
KAP-1/TIF1β	Human	HP1Hsα, HP1Hsγ	IF, rPD, exIP, SPRA, GFC	CSD	28, 46--48
KAP-1/TIF1β	Mouse	mHP1α, M31, M32	IF, rPD, Y2H, exIP, GFC	CSD	28, 37, 47, 86, 90
TRF1/PIN2	Mouse	M31	IF	nd	91
TAFII130	Human	HP1Hsα, HP1Hsγ	Y2H, transPD, exPD	CSD	36
TIF1α	Mouse	mHP1α, M31, M32	Y2H, rPD	CSD	86, 90
mSNF2β	Mouse	mHP1α	Y2H	CSD	86
Rb	Human	HP1, HP1Hsγ	Y2H, exPD, exIP, ChIP	nd	57, 82
Rb	Maize	HP1γ	rPD, Y2H	nd	82
Dnmt3a	Mouse cells	mHP1α	IF	nd	92
Dnmt3b	Mouse cells	mHP1α	IF	nd	92
ATRX/HP1-BP38	Mouse	mHP1α, M31	Y2H, IF	CSD	86, 93
Pim-1	Human	HP1Hsγ	Y2H, exIP, rPD	CSD	65
CKII	*Drosophila*	HP1	*In vitro* phosphorylation	nd	63
dAF10	*Drosophila*	HP1	transPD	CSD	94
DNA replication and repair					
CAF-1 p150	Mouse	mHP1α, M31	IF, Y2H, rPD, GFC, NMR	CSD	28, 37
CAF-1 p150	Human	HP1Hsα	rPD	CSD	48
Ku70	Human	HP1Hsα	Y2H, rPD, exIP	CSD	95
BRCA-1*	Human	HP1Hsα	IF	nd	96
ORC1	*Drosophila*	HP1	transIP	CD, CSD	58
ORC2	*Drosophila*	HP1	IF, exPD, exIP, transIP	CD, CSD	58
ORC3	*Drosophila*	HP1	transIP	CD, CSD	58
ORC4	*Drosophila*	HP1	transIP	CD, CSD	58
ORC5	*Drosophila*	HP1	exIP, transIP	CD, CSD	58
ORC6	*Drosophila*	HP1	exIP, transIP	CD, CSD	58
Xorc1	*Xenopus*	XHP1α, XHP1γ	Y2H	nd	58
HOAP	*Drosophila*	HP1	IF, exIP	nd	62
Nuclear architecture					
Lamin B receptor	Human	HP1Hsα, HP1Hsβ, HP1Hsγ	Y2H, rPD, exPD, transPD, exIP	CSD	48, 52–54
HP1-BP84	Mouse	mHP1α, M31	Y2H	CSD	86
Lamin B	Mouse	M31	BA	CD	55
LAP2β	Mouse	M31	BA	CD	55
Nuclear envelope	Mouse	mHP1α, M31, M32	IF, BA	CD	55
Other chromosome-associated proteins					
Psc3	*S. pombe*	Swi6	IF, Y2H, exPD, ChIP	CD+Glu-rich	13
DDP1	*Drosophila*	HP1	IF	nd	97
Arp4/dArp6	*Drosophila*	HP1	IF	nd	98, 99
INCENP	Human	HP1Hsα, HP1Hsγ	Y2H, transPD	Hinge region	100
Ki-67	Human	mHP1α, M31, M32	Y2H, exPD, IF	CSD	101
SP100B	Human	HP1Hsα, HP1Hsβ, HP1Hsγ	IF, Y2H, rPD, transPD	CSD	48, 102, 103
EST AA153281	Mouse	mHP1α, M31	Y2H, rPD	CSD	37
EST AA003533	Mouse	mHP1α, M31	Y2H, rPD	CSD	37

BA, binding assay; ChIP, chromatin immunoprecipitation; exIP, co-immunoprecipitation using extract; exPD, pull-down assay using extracts; FAITC, fluorescence anisotropy, isothermal titration calorimetry; FW, far Western analysis; GFC, gel filtration chromatography; IF, immunofluorescence colocalization; nd, not determined; rIP, coprecipitation using recombinant proteins; rPD, pull-down assay using recombinant proteins; transIP, immunoprecipitation with *in vitro*-translated protein; transPD, pull-down assay using *in vitro* translated protein; SED, sedimentation assay; SPRA, surface plasmon resonance analysis; Y2H, yeast two-hybrid assay. *Denotes cell cycle-dependent association.

HP1 Regulates Gene Expression

Effects on Gene Expression Near Centric Heterochromatin. The gene encoding HP1, *Su(var)2–5*, was isolated in a screen for suppressors and enhancers of PEV of the *white*[+] gene brought into juxtaposition with heterochromatin through a chromosomal rearrangement (8). To determine the effects of HP1 on chromatin packaging, stocks containing the well-characterized *Drosophila hsp26* gene inserted within centric heterochromatin were used for chromatin structure analysis (16). These transgenes exhibit less accessibility to nucleases and are packaged into a more regular nucleosome array than euchromatic insertions. In an HP1 mutant background the transgenes become more accessible to restriction enzyme digestion, indicating a more "open" chromatin configuration (66). To determine the transcriptional mechanism impaired by packaging with HP1, high-resolution chromatin structure analysis was performed (67). The results indicated that general transcription factors such as TFIID and RNA polymerase II are not associated with heterochromatic transgenes exhibiting silencing. Thus, association of HP1 correlates with a "closed" chromatin configuration that limits the accessibility of regulatory sites to trans-acting factors. This is similar to the mechanism of X-chromosome inactivation in mammals where transcription factors are also absent from genes on the inactive X (68), but contrasts the mechanism of silencing at the mating type loci in *S. cerevisiae* and Polycomb-mediated silencing in *Drosophila* in which general transcription factors and RNA polymerase II are found in association with silenced promoters (69–71). Differences between these systems and HP1 silencing could reflect fundamentally different properties of the silencing systems or developmentally different stages of silent chromatin maturation.

In contrast to the silencing effects HP1 has on euchromatic genes, HP1 is required for the expression of genes that naturally reside within heterochromatin. Over the years genetic and molecular analyses have revealed genes that reside within centric heterochromatin (72, 73). Two well-characterized genes are *light*, an essential gene encoding a protein involved in the vesicle transport pathway (74, 75), and *rolled*, an essential mitogen-activated protein kinase (76). Heterochromatic genes are not specific for *Drosophila*; they also have been discovered in *Arabidopsis* (77) and are likely to be found in other organisms as genomic sequence analysis becomes more complete. In *Drosophila*, heterochromatic genes appear to be unrelated to each other in function, however, they do share some common properties. Structurally, many heterochromatic genes have introns containing middle repetitive DNA sequences (74) (D. E. Cryderman and L.L.W., unpublished data). Heterochromatic genes are inefficiently expressed and sometimes exhibit PEV when translocated to euchromatin (78, 79). In addition, heterochromatic genes require heterochromatin proteins such as HP1 for expression (11).

How does HP1 establish a chromatin configuration that hinders the expression of euchromatic genes while fostering the expression of heterochromatic genes? This question will be better addressed as the promoter regions of heterochromatic genes are analyzed. Assuming a role for HP1 in chromatin compaction, HP1 might bring distant regulatory elements in association with the promoter region of heterochromatic genes. Alternatively, HP1 may be required to set up a favorable chromatin configuration within the promoter proximal region and/or be involved in the recruitment of general transcription factors as suggested by a recent report showing an interaction between HP1 and the general transcription factor TF$_{II}$130 (36).

Effects of Tethering HP1. What are the effects of HP1 on gene expression at locations other than centric heterochromatin? One approach taken to address this question has been to generate HP1 fusion proteins containing heterologous DNA binding domains. In mammalian cell transient transfection assays, effects of HP1 fusion proteins on the expression of reporter genes possessing the appropriate DNA binding sequences are assayed. In these experiments both murine and human HP1 family members repress transcription when tethered to a small number of sites located in close proximity to the promoter (36). Transcriptional repression no longer occurs as the binding sites are moved to distances more than 2 kb from the promoter (80). Does this indicate that HP1-mediated repression has only short-range capabilities, perhaps operating on a gene-by-gene basis? If this is the case, HP1 located at euchromatic sites (Fig. 1) might play a role in the regulation of individual euchromatic genes, rather than entire domains.

The effects of tethering HP1 in a chromosomal context, as in transgenic *Drosophila*, demonstrate the complexities of gene silencing. A Gal4–HP1 fusion protein tethered upstream of a reporter gene caused silencing at only one of six genomic locations tested (51, 81). Interestingly, the site that supported silencing was flanked by middle repetitive DNA sequences, reminiscent of heterochromatin domains. In this case, silencing could spread in trans to a homologue lacking the tethering sites. These results suggest that not all chromosomal contexts will support the formation of silent chromatin by HP1. The inability to silence at certain locations might depend on the chromatin of the neighboring region, including the types of histone modifications, as well as gene density and transcriptional status of the region.

Identification of HP1 Target Genes. As a second approach to determine the effects of HP1 on gene expression at locations other than centric heterochromatin, investigators have identified potential target genes by their response to HP1 dosage. Representational difference analysis identified two genes that are up-regulated in homozygous HP1 mutant larvae (59). Interestingly, one of the genes misregulated in the HP1 homozygous mutant maps to cytological region 31, a chromosome division that stains intensely with antibodies to HP1 (7). Two additional randomly selected genes within region 31 show up-regulation in HP1 homozygous mutants (59). For all four candidate HP1 target genes, mutations in additional modifiers of PEV, including *Su(var)3–9*, cause increases in gene expression. These results suggest that HP1 might function to silence genes located within euchromatic domains in a mechanism similar to that operating in centric regions.

A microarray approach has also been used to identify candidate genes regulated by HP1. Several hundred genes mapping within the euchromatic arms are up-regulated in an HP1 mutant background (D. E. Cryderman and L.L.W., unpublished data). In addition, several hundred genes were down-regulated, a pattern similar to that of heterochromatic genes. For all of the candidate target genes identified in *Drosophila* to date, it remains to be determined whether misregulation is caused by a direct interaction with HP1. It will be of interest to determine whether interactions between HP1 and specific genes are conserved through evolution.

In mammals, it is also likely that HP1 plays a role in the regulation of genes at noncentric locations. HP1 has been identified as a partner protein for many promoter-associated factors involved in control of gene repression. These include the transcription intermediate factor TIF1β that interacts with zinc finger proteins containing Krüppel-associated box (KRAB) domains known to be involved in transcriptional repression. The role of HP1 in KRAB-mediated repression is not clear, but might involve recruitment of histone deacetylases (47, 48).

The first example of a direct association of HP1 with a promoter region came from studies on the *cyclin E* gene (57). Binding of the Rb protein upstream of the *cyclin E* promoter

causes gene silencing, partly through the recruitment of histone deacetylases. In addition, chromatin cross-linking and immuno-precipitation experiments place Rb at the promoter with HP1 and methylated lysine nine of histone H3 (57). Consistent with the finding, an "Rb binding motif" is present within the amino acid sequences of HP1 from a variety of species (82). It is unclear whether HP1 is recruited to the *cyclin E* promoter by Rb, methylated lysine nine of histone H3, SUV39H1, or any combination of these interacting molecules. In support of such interactions, HP1 possesses the ability to simultaneously interact with the methylated histone H3 tail and Rb (57). One hypothesis is that Rb recruits histone deacetylases first, because the histone H3 methyltansferase cannot use an acetylated lysine as a substrate for methylation (43), then SUV39H1 methylates the histone tail which serves as the substrate for HP1 binding. Interestingly, *Drosophila* SU(VAR)3–9 was recently purified in a complex with histone deacetylase HDAC1, suggesting that the two proteins might cooperate to methylate previously acetylated histone tails (83).

The identification of HP1 target genes has implications for understanding breast cancer metastasis in humans. HP1$^{Hs\alpha}$, but not HP1$^{Hs\beta}$ or HP1$^{Hs\gamma}$, is down-regulated in highly invasive/metastatic breast cancer cells compared with poorly invasive/nonmetastatic breast cancer cells (84). Introduction of a tagged HP1$^{Hs\alpha}$ into the highly invasive/metastatic cells, which normally have low levels of HP1$^{Hs\alpha}$, lead to a less *in vitro* invasive phenotype. These results imply that modulation of the levels of HP1$^{Hs\alpha}$ alters molecular properties of cells needed for invasion. Consistent with the cell culture studies, HP1$^{Hs\alpha}$ is down-regulated in tissues from distant metastatic sites in breast cancer patients (84). One hypothesis is that HP1$^{Hs\alpha}$ normally silences genes required for metastasis, making HP1$^{Hs\alpha}$ a candidate metastasis suppressor. Depending on when HP1$^{Hs\alpha}$ is down-regulated during tumor progression, HP1$^{Hs\alpha}$ could be used as a predictive/prognostic marker for metastatic breast cancer.

Since the discovery of HP1 over 12 years ago by the laboratory of Sarah C. R. Elgin (Washington University, St. Louis), HP1 has grown in popularity. In part, this has been caused by the fact that HP1 has unexpectedly been identified as an interacting partner for a wide variety of proteins with diverse nuclear functions. Dissecting the function of HP1 in association with its partner proteins lies ahead. These experiments will shed light on the connections between chromatin structure, gene expression, DNA replication and repair, and nuclear organization.

1. Richards, E. J. & Elgin, S. C. (2002) *Cell* **108**, 489–500.
2. Jeppesen, P., Mitchell, A., Turner, B. & Perry, P. (1992) *Chromosoma* **101**, 322–332.
3. Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M. & Broach, J. R. (1996) *Mol. Cell. Biol.* **16**, 4349–4356.
4. Turner, B. M., Birley, A. J. & Lavender, J. (1992) *Cell* **69**, 375–384.
5. Cheung, P., Tanner, K. G., Cheung, W. L., Sassone-Corsi, P., Denu, J. M. & Allis, C. D. (2000) *Mol. Cell* **5**, 905–915.
6. Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R. & Berger, S. L. (2000) *Mol. Cell* **5**, 917–926.
7. James, T. C., Eissenberg, J. C., Craig, C., Dietrich, V., Hobson, A. & Elgin, S. C. (1989) *Eur. J. Cell Biol.* **50**, 170–180.
8. Wustmann, G., Szidonya, J., Taubert, H. & Reuter, G. (1989) *Mol. Gen. Genet.* **217**, 520–527.
9. Eissenberg, J. C., James, T. C., Foster-Hartnett, D. M., Hartnett, T., Ngan, V. & Elgin, S. C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9923–9927.
10. Weiler, K. S. & Wakimoto, B. T. (1995) *Annu. Rev. Genet.* **29**, 577–605.
11. Lu, B. Y., Emtage, P. C., Duyf, B. J., Hilliker, A. J. & Eissenberg, J. C. (2000) *Genetics* **155**, 699–708.
12. Fanti, L., Giovinazzo, G., Berloco, M. & Pimpinelli, S. (1998) *Mol. Cell* **2**, 527–538.
13. Nonaka, N., Kitajima, T., Yokobayashi, S., Xiao, G., Yamamoto, M., Grewal, S. I. & Watanabe, Y. (2002) *Nat. Cell Biol.* **4**, 89–93.
14. Bernard, P., Maure, J. F., Partridge, J. F., Genier, S., Javerzat, J. P. & Allshire, R. C. (2001) *Science* **294**, 2539–2542.
15. Sun, F. L., Cuaycong, M. H., Craig, C. A., Wallrath, L. L., Locke, J. & Elgin, S. C. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5340–5345.
16. Wallrath, L. L. & Elgin, S. C. (1995) *Genes Dev.* **9**, 1263–1277.
17. Pardue, M. L. & DeBaryshe, P. G. (1999) *Genetica* **107**, 189–196.
18. Gaudin, V., Libault, M., Pouteau, S., Juul, T., Zhao, G., Lefebvre, D. & Grandjean, O. (2001) *Development (Cambridge, U.K.)* **128**, 4847–4858.
19. Couteau, F., Guerry, F., Muller, F. & Palladino, F. (2002) *EMBO Rep.* **3**, 235–241.
20. Eissenberg, J. C. & Elgin, S. C. (2000) *Curr. Opin. Genet. Dev.* **10**, 204–210.
21. Volpe, A. M., Horowitz, H., Grafer, C. M., Jackson, S. M. & Berg, C. A. (2001) *Genetics* **159**, 1117–1134.
22. Smothers, J. F. & Henikoff, S. (2001) *Mol. Cell. Biol.* **21**, 2555–2569.
23. Minc, E., Allory, Y., Worman, H. J., Courvalin, J. C. & Buendia, B. (1999) *Chromosoma* **108**, 220–234.
24. Eissenberg, J. C. (2001) *Gene* **275**, 19–29.
25. Aasland, R. & Stewart, A. F. (1995) *Nucleic Acids Res.* **23**, 3168–3174.
26. Jacobs, S. A. & Khorasanizadeh, S. (2002) *Science* **295**, 2080–2083.
27. Cowieson, N. P., Partridge, J. F., Allshire, R. C. & McLaughlin, P. J. (2000) *Curr. Biol.* **10**, 517–525.
28. Brasher, S. V., Smith, B. O., Fogh, R. H., Nietlispach, D., Thiru, A., Nielsen, P. R., Broadhurst, R. W., Ball, L. J., Murzina, N. V. & Laue, E. D. (2000) *EMBO J.* **19**, 1587–1597.
29. Ball, L. J., Murzina, N. V., Broadhurst, R. W., Raine, A. R., Archer, S. J., Stott, F. J., Murzin, A. G., Singh, P. B., Domaille, P. J. & Laue, E. D. (1997) *EMBO J.* **16**, 2473–2481.
30. Nielsen, P. R., Nietlispach, D., Mott, H. R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A. G., Murzina, N. V. & Laue, E. D. (2002) *Nature (London)* **416**, 103–107.
31. Smothers, J. F. & Henikoff, S. (2000) *Curr. Biol.* **10**, 27–30.
32. Wang, G., Ma, A., Chow, C. M., Horsley, D., Brown, N. R., Cowell, I. G. & Singh, P. B. (2000) *Mol. Cell. Biol.* **20**, 6970–6983.
33. Platero, J. S., Hartnett, T. & Eissenberg, J. C. (1995) *EMBO J.* **14**, 3977–3986.
34. Zhao, T., Heyduk, T., Allis, C. D. & Eissenberg, J. C. (2000) *J. Biol. Chem.* **275**, 28332–28338.
35. Delattre, M., Spierer, A., Tonka, C. H. & Spierer, P. (2000) *J. Cell. Sci.* **113**, 4253–4261.
36. Vassallo, M. F. & Tanese, N. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5919–5924.
37. Murzina, N., Verreault, A., Laue, E. & Stillman, B. (1999) *Mol. Cell* **4**, 529–540.
38. Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. & Kouzarides, T. (2001) *Nature (London)* **410**, 120–124.
39. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. (2001) *Nature (London)* **410**, 116–120.
40. Jenuwein, T. & Allis, C. D. (2001) *Science* **293**, 1074–1080.
41. Jacobs, S. A., Taverna, S. D., Zhang, Y., Briggs, S. D., Li, J., Eissenberg, J. C., Allis, C. D. & Khorasanizadeh, S. (2001) *EMBO J.* **20**, 5232–5241.
42. Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R. & Reuter, G. (2002) *EMBO J.* **21**, 1121–1131.
43. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. (2001) *Science* **292**, 110–113.
44. Grewal, S. I. & Elgin, S. C. (2002) *Curr. Opin. Genet. Dev.* **12**, 178–187.
45. Cowell, I. G., Aucott, R., Mahadevaiah, S. D., Borgoyne, P. S., Huskisson, N., Bongiorni, S., Prantera, G., Fanti, L., Pimpinelli, S., Wu, R., *et al.* (2002) *Chromosoma* **111**, 22–36.
46. Nielsen, A. L., Oulad-Abdelghani, M., Ortiz, J. A., Remboutsika, E., Chambon, P. & Losson, R. (2001) *Mol. Cell* **7**, 729–739.
47. Ryan, R. F., Schultz, D. C., Ayyanathan, K., Singh, P. B., Friedman, J. R., Fredericks, W. J. & Rauscher, F. J., 3rd (1999) *Mol. Cell. Biol.* **19**, 4366–4378.
48. Lechner, M. S., Begg, G. E., Speicher, D. W. & Rauscher, F. J., 3rd (2000) *Mol. Cell. Biol.* **20**, 6449–6465.
49. Cleard, F. & Spierer, P. (2001) *EMBO Rep.* **2**, 1095–1100.
50. Cleard, F., Delattre, M. & Spierer, P. (1997) *EMBO J.* **16**, 5280–5288.
51. Seum, C., Delattre, M., Spierer, A. & Spierer, P. (2001) *EMBO J.* **20**, 812–818.
52. Ye, Q. & Worman, H. J. (1996) *J. Biol. Chem.* **271**, 14653–14656.

53. Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J. C. & Worman, H. J. (1997) *J. Biol. Chem.* **272,** 14983–14989.
54. Polioudaki, H., Kourmouli, N., Drosou, V., Bakou, A., Theodoropoulos, P. A., Singh, P. B., Giannakouros, T. & Georgatos, S. D. (2001) *EMBO Rep.* **2,** 920–925.
55. Kourmouli, N., Theodoropoulos, P. A., Dialynas, G., Bakou, A., Politou, A. S., Cowell, I. G., Singh, P. B. & Georgatos, S. D. (2000) *EMBO J.* **19,** 6558–6568.
56. Pombo, A., Jones, E., Iborra, F. J., Kimura, H., Sugaya, K., Cook, P. R. & Jackson, D. A. (2000) *Crit. Rev. Eukaryotic Gene Expression* **10,** 21–29.
57. Nielsen, S. J., Schneider, R., Bauer, U. M., Bannister, A. J., Morrison, A., O'Carroll, D., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R. E. & Kouzarides, T. (2001) *Nature (London)* **412,** 561–565.
58. Pak, D. T., Pflumm, M., Chesnokov, I., Huang, D. W., Kellum, R., Marr, J., Romanowski, P. & Botchan, M. R. (1997) *Cell* **91,** 311–323.
59. Hwang, K. K., Eissenberg, J. C. & Worman, H. J. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 11423–11427.
60. Gross, D. S. (2001) *Trends Biochem. Sci.* **26,** 685–686.
61. Gasser, S. M. & Cockell, M. M. (2001) *Gene* **279,** 1–16.
62. Shareef, M. M., King, C., Damaj, M., Badagu, R., Huang, D. W. & Kellum, R. (2001) *Mol. Biol. Cell* **12,** 1671–1685.
63. Zhao, T. & Eissenberg, J. C. (1999) *J. Biol. Chem.* **274,** 15095–15100.
64. Zhao, T., Heyduk, T. & Eissenberg, J. C. (2001) *J. Biol. Chem.* **276,** 9512–9518.
65. Koike, N., Maita, H., Taira, T., Ariga, H. & Iguchi-Ariga, S. M. (2000) *FEBS Lett.* **467,** 17–21.
66. Cryderman, D. E., Cuaycong, M. H., Elgin, S. C. & Wallrath, L. L. (1998) *Chromosoma* **107,** 277–285.
67. Cryderman, D. E., Tang, H., Bell, C., Gilmour, D. S. & Wallrath, L. L. (1999) *Nucleic Acids Res.* **27,** 3364–3370.
68. Pfeifer, G. P. & Riggs, A. D. (1991) *Genes Dev.* **5,** 1102–1113.
69. Breiling, A., Turner, B. M., Bianchi, M. E. & Orlando, V. (2001) *Nature (London)* **412,** 651–655.
70. Sekinger, E. A. & Gross, D. S. (2001) *Cell* **105,** 403–414.
71. Sekinger, E. A. & Gross, D. S. (1999) *EMBO J.* **18,** 7041–7055.
72. Sinclair, D. A., Schulze, S., Silva, E., Fitzpatrick, K. A. & Honda, B. M. (2000) *Genetica* **109,** 9–18.
73. Carvalho, A. B., Dobo, B. A., Vibranovski, M. D. & Clark, A. G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13225–13230.
74. Devlin, R. H., Bingham, B. & Wakimoto, B. T. (1990) *Genetics* **125,** 129–140.
75. Warner, T. S., Sinclair, D. A., Fitzpatrick, K. A., Singh, M., Devlin, R. H. & Honda, B. M. (1998) *Genome* **41,** 236–243.
76. Eberl, D. F., Duyf, B. J. & Hilliker, A. J. (1993) *Genetics* **134,** 277–292.
77. Copenhaver, G. P., Nickel, K., Kuromori, T., Benito, M. I., Kaul, S., Lin, X., Bevan, M., Murphy, G., Harris, B., Parnell, L. D., *et al.* (1999) *Science* **286,** 2468–2474.
78. Weiler, K. S. & Wakimoto, B. T. (1998) *Genetics* **149,** 1451–1464.
79. Wakimoto, B. T. & Hearn, M. G. (1990) *Genetics* **125,** 141–154.
80. van der Vlag, J., den Blaauwen, J. L., Sewalt, R. G., van Driel, R. & Otte, A. P. (2000) *J. Biol. Chem.* **275,** 697–704.
81. Seum, C., Spierer, A., Delattre, M., Pauli, D. & Spierer, P. (2000) *Chromosoma* **109,** 453–459.
82. Williams, L. & Grafi, G. (2000) *Trends Plant Sci.* **5,** 239–240.
83. Czermin, B., Schotta, G., Hulsmann, B. B., Brehm, A., Becker, P. B., Reuter, G. & Imhof, A. (2001) *EMBO Rep.* **2,** 915–919.
84. Kirschmann, D. A., Lininger, R. A., Gardner, L. M., Seftor, E. A., Odero, V. A., Ainsztein, A. M., Earnshaw, W. C., Wallrath, L. L. & Hendrix, M. J. (2000) *Cancer Res.* **60,** 3359–3363.
85. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10915–10919.
86. Le Douarin, B., Nielsen, A. L., Garnier, J. M., Ichinose, H., Jeanmougin, F., Losson, R. & Chambon, P. (1996) *EMBO J.* **15,** 6701–6715.
87. Turner, J. M., Burgoyne, P. S. & Singh, P. B. (2001) *J. Cell. Sci.* **114,** 3367–3375.
88. Czvitkovich, S., Sauer, S., Peters, A. H., Deiner, E., Wolf, A., Laible, G., Opravil, S., Beug, H. & Jenuwein, T. (2001) *Mech. Dev.* **107,** 141–153.
89. Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P. B., *et al.* (1999) *EMBO J.* **18,** 1923–1938.
90. Nielsen, A. L., Ortiz, J. A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P. & Losson, R. (1999) *EMBO J.* **18,** 6385–6395.
91. Netzer, C., Rieger, L., Brero, A., Zhang, C. D., Hinzke, M., Kohlhase, J. & Bohlander, S. K. (2001) *Hum. Mol. Genet.* **10,** 3017–3024.
92. Bachman, K. E., Rountree, M. R. & Baylin, S. B. (2001) *J. Biol. Chem.* **276,** 32282–32287.
93. McDowell, T. L., Gibbons, R. J., Sutherland, H., O'Rourke, D. M., Bickmore, W. A., Pombo, A., Turley, H., Gatter, K., Picketts, D. J., Buckle, V. J., *et al.* (1999) *Proc. Natl. Acad. Sci. USA* **96,** 13983–13988.
94. Linder, B., Gerlach, N. & Jackle, H. (2001) *EMBO Rep.* **2,** 211–216.
95. Song, K., Jung, Y., Jung, D. & Lee, I. (2001) *J. Biol. Chem.* **276,** 8321–8327.
96. Maul, G. G., Jensen, D. E., Ishov, A. M., Herlyn, M. & Rauscher, F. J., 3rd (1998) *Cell Growth Differ.* **9,** 743–755.
97. Cortes, A., Huertas, D., Fanti, L., Pimpinelli, S., Marsellach, F. X., Pina, B. & Azorin, F. (1999) *EMBO J.* **18,** 3820–3833.
98. Frankel, S., Sigel, E. A., Craig, C., Elgin, S. C., Mooseker, M. S. & Artavanis-Tsakonas, S. (1997) *J. Cell. Sci.* **110,** 1999–2012.
99. Kato, M., Sasaki, M., Mizuno, S. & Harata, M. (2001) *Gene* **268,** 133–140.
100. Ainsztein, A. M., Kandels-Lewis, S. E., Mackay, A. M. & Earnshaw, W. C. (1998) *J. Cell Biol.* **143,** 1763–1774.
101. Scholzen, T., Endl, E., Wohlenberg, C., van der Sar, S., Cowell, I. G., Gerdes, J. & Singh, P. B. (2002) *J. Pathol.* **196,** 135–144.
102. Lehming, N., Le Saux, A., Schuller, J. & Ptashne, M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 7322–7326.
103. Seeler, J. S., Marchio, A., Sitterlin, D., Transy, C. & Dejean, A. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 7316–7321.

Self-perpetuating epigenetic pili switches in bacteria

Aaron Hernday*, Margareta Krabbe†, Bruce Braaten*, and David Low*‡

*Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, CA 93117; and †Karolinska Institutet, SE-171 77 Stockholm, Sweden

Bacteria have developed an epigenetic phase variation mechanism to control cell surface pili–adhesin complexes between heritable expression (phase ON) and nonexpression (phase OFF) states. In the pyelonephritis-associated pili (*pap*) system, global regulators [catabolite gene activator protein (CAP), leucine-responsive regulatory protein (Lrp), DNA adenine methylase (Dam)] and local regulators (PapI and PapB) control phase switching. Lrp binds cooperatively to three *pap* DNA binding sites, sites 1–3, proximal to the *papBA* pilin promoter in phase OFF cells, whereas Lrp is bound to sites 4–6 distal to *papBA* in phase ON cells. Two Dam methylation targets, GATC^{prox} and GATC^{dist}, are located in Lrp binding sites 2 and 5, respectively. In phase OFF cells, binding of Lrp at sites 1–3 inhibits methylation of GATC^{prox}, forming the phase OFF DNA methylation pattern (GATC^{dist} methylated, GATC^{prox} nonmethylated). Binding of Lrp at sites 1–3 blocks *pap* pili transcription and reduces the affinity of Lrp for sites 4–6. Together with methylation of GATC^{dist}, which inhibits Lrp binding at sites 4–6, the phase OFF state is maintained. We hypothesize that transition to the phase ON state requires DNA replication to dissociate Lrp and generate a hemimethylated GATC^{dist} site. PapI and methylation of GATC^{prox} act together to increase the affinity of Lrp for sites 4–6. Binding of Lrp at the distal sites protects GATC^{dist} from methylation, forming the phase ON methylation pattern (GATC^{dist} nonmethyated, GATC^{prox} methylated). Lrp binding at sites 4–6 together with cAMP-CAP binding 215.5 bp upstream of the papBA transcription start, is required for activation of pilin transcription. The first gene product of the *papBA* transcript, PapB, helps maintain the switch in the ON state by activating *papI* transcription, which in turn maintains Lrp binding at sites 4–6.

Cis- and Trans-Acting Pap Switch Components

Bacteria have developed phase variation mechanisms to control cell surface pili–adhesin complexes between expression (phase ON) and nonexpression (phase OFF) states. Pili phase variation can occur by site specific (1, 2) and homologous recombination (3) mechanisms. In addition, a large group of pili operons including pyelonephritis-associated pili (*pap*) are regulated by an epigenetic switch directly controlled by DNA methylation pattern formation (4, 5). The focus of this paper is on the mechanisms by which the phase OFF and phase ON states are perpetuated, the transition between phase OFF and phase ON states, and the external inputs that control phase switching.

The expression of *pap* is positively controlled by the local regulators PapI (8 kDa) and PapB (12 kDa) in concert with the global regulators leucine-responsive regulatory protein (Lrp) and catabolite gene activator protein (CAP). DNA adenine methylase (Dam) is also required for *pap* transcription (Table 1). Knockout mutations in each of the genes encoding these regulatory proteins inhibit the *pap* phase OFF to phase ON switch (Table 1). In addition, histone-like nucleoid structuring protein (H-NS) modulates *pap* phase switching because *hns* mutants show a decreased OFF to ON switch rate (6, 7).

The *pap* regulatory region encompasses the divergently transcribed *papI* and *papB* genes together with the 416-bp intergenic region (Fig. 1). There are six *pap* DNA Lrp binding sites, designated sites 1–6, within the *pap* regulatory region spaced three helical turns apart which control transcription at the *papBA* promoter. Two DNA GATC sites designated GATC^{prox} and GATC^{dist} (proximal and distal with respect to the papBA promoter) are located within Lrp binding sites 2 and 5, respectively. DNA GATC sites are target sites for Dam, which methylates the adenosine of the GATC sequence.

Mutational analyses showed that disruption of Lrp binding sites 2 or 3 resulted in a higher OFF to ON switch rate or a phase locked ON phenotype, respectively. In contrast, disruption of Lrp sites 4 or 5 resulted in a phase locked OFF phenotype (Fig. 1) (8). These results suggested that binding of Lrp proximal to the *papBAp* promoter inhibits transcription whereas binding of Lrp at the distal site activates transcription. *In vitro* DNA footprint analyses indicated that Lrp binds with highest affinity to Lrp sites 1–3, and with lower affinity to sites 4–6 (8–10). Insertion or deletion of a single base pair between Lrp binding sites 1 and 2 locks transcription in the ON phase, consistent with the observation that proper spacing between Lrp binding sites is necessary for the high cooperativity in binding (N. Kozak and D.L., unpublished results). Examination of the *pap* DNA methylation patterns showed that, in phase OFF cells, GATC^{prox} is nonmethylated and GATC^{dist} is methylated, whereas the converse pattern exists in phase ON cells (GATC^{dist} nonmethylated, GATC^{prox} methylated) (Fig. 2). These methylation patterns depend on Lrp (9, 11, 12). Moreover, addition of Lrp to *pap* DNA *in vitro* blocks methylation of the *pap* regulatory GATC sequences (12). Together, these data indicate that in phase OFF cells Lrp is bound at sites 1–3, blocking methylation of GATC^{prox} within site 2. In contrast, Lrp bound to sites 4–6 in phase ON cells blocks methylation of GATC^{dist} within site 5 (Fig. 2). Translocation of Lrp from sites 1–3 to sites 4–6 requires PapI. Lrp thus plays dual roles in regulating *pap* transcription. Lrp acts as a repressor when bound proximal to the pap BA pilin promoter. Lrp activates pilin transcription when, together with PapI, it is bound distal to the pilin promoter (13). Activation of *pap* transcription also requires binding of cAMP-CAP at a site 60 bp upstream of Lrp binding site 4 (14–16) and binding of the PapB regulatory protein at a site proximal to the *papI* promoter (17, 18) (Fig. 1).

Pap Phase Variation Model

A model for Pap phase variation is shown in Fig. 3. Specific details and supporting data for the model are discussed below.

The Self-Perpetuating Phase OFF State. Only the distal GATC site of the *pap* regulatory region is methylated in phase OFF cells

Table 1. Trans-acting factors that regulate Pap phase variation

Genotype	Description	Switch rates (OFF to ON)*
Wild type		7×10^{-4} per cell per generation
lrp⁻	Leucine-responsive regulatory protein	Locked OFF
crp⁻	Catabolite gene activator protein	Locked OFF
dam⁻	DNA adenine methylase	Locked OFF
papI⁻	Local regulatory protein	Locked OFF
papB⁻	Local regulatory protein	Locked OFF
hns⁻	Histone-like nucleoid structuring protein	2×10^{-4} per cell per generation

*Switch rates measured in M9 glycerol minimal medium. Data are from refs. 6, 7, 9–11, 13, and 22.

(GATC^dist methylated, GATC^prox nonmethylated). Initial analysis indicated that methylation of GATC^dist was required for maintenance of the OFF transcription state because introduction of an A to C transversion within GATC results in a phase-locked ON phenotype (Fig. 1) (10). Although the adenosine within GATC is obviously required for methylation of this site by Dam, the affinities of Lrp for wild-type *pap* DNA and DNA containing the GCTC^dist mutation appear similar (10), and dimethyl sulfate footprint analyses indicate that Lrp does not closely contact the adenosine of GATC (8). Further studies (10) showed that overproduction of Dam prevented the phase OFF to ON transition in cells containing a wild-type *pap* sequence but not in cells containing the GCTC^dist mutation (Fig. 4). Thus, overmethylation of GATC^dist prevents the phase OFF to ON transition, consistent with the hypothesis that methylation of this distal GATC site helps maintain cells in the OFF state.

Another factor that may contribute to maintenance of the phase OFF state is a mutual binding exclusion phenomenon. The affinity of Lrp is about 2 times higher for sites 1–3 compared with sites 4–6 when the sites are separated (Fig. 5A Lower). However, when the sites are linked (intact *pap* regulatory region) the affinity of Lrp for sites 4–6 is reduced 10-fold (Fig. 5A, compare *Upper* and *Lower*). These results indicate that binding of Lrp at sites 1–3 exerts a negative effect on Lrp binding at sites 4–6. This

mutual binding exclusion effect is reduced from 10-fold to only about 2-fold when unsupercoiled DNAs are used (unpublished data), showing a strong dependence on DNA topology. Because Lrp is known to form higher oligomers under certain conditions (19, 20) and bend DNA (21), one possible mechanism is that the conformation of *pap* sites 4–6 might be altered as a result of a binding of Lrp at sites 1–3 located 102 bp away (measured from site 2 to site 5) (Fig. 1). Because the binding of Lrp to *pap* DNA is highly cooperative (8), this could serve as a signal amplification mechanism. As discussed below, when the affinity of Lrp for sites 1–3 is reduced by mutation of Lrp binding site 3, mutual exclusion works in reverse, lowering the affinity of Lrp at sites 1–3 as a result of binding of Lrp at sites 4–6 (Fig. 5B).

Activation of *papBA* transcription requires binding of Lrp at sites 4–6 (8), and thus, binding of Lrp at sites 1–3 indirectly inhibits transcription because of mutual exclusion of Lrp binding at sites 4–6. In addition, Lrp binding at sites 1–3 appears to directly block *pap* transcription *in vivo* (13). Although *papBA* transcription is low in cells containing either *lrp*⁻ (3 Miller units, MU) or *hns*⁻ (59 MU) mutations, cells lacking both Lrp and H-NS display a basal transcription level (528 MU), which is about one-eighth that of phase ON cells (4,200 MU) (13). This transcription activity is similar to that observed for *hns*⁻ cells in which *pap* regulatory sequences upstream of the *papBA* pro-
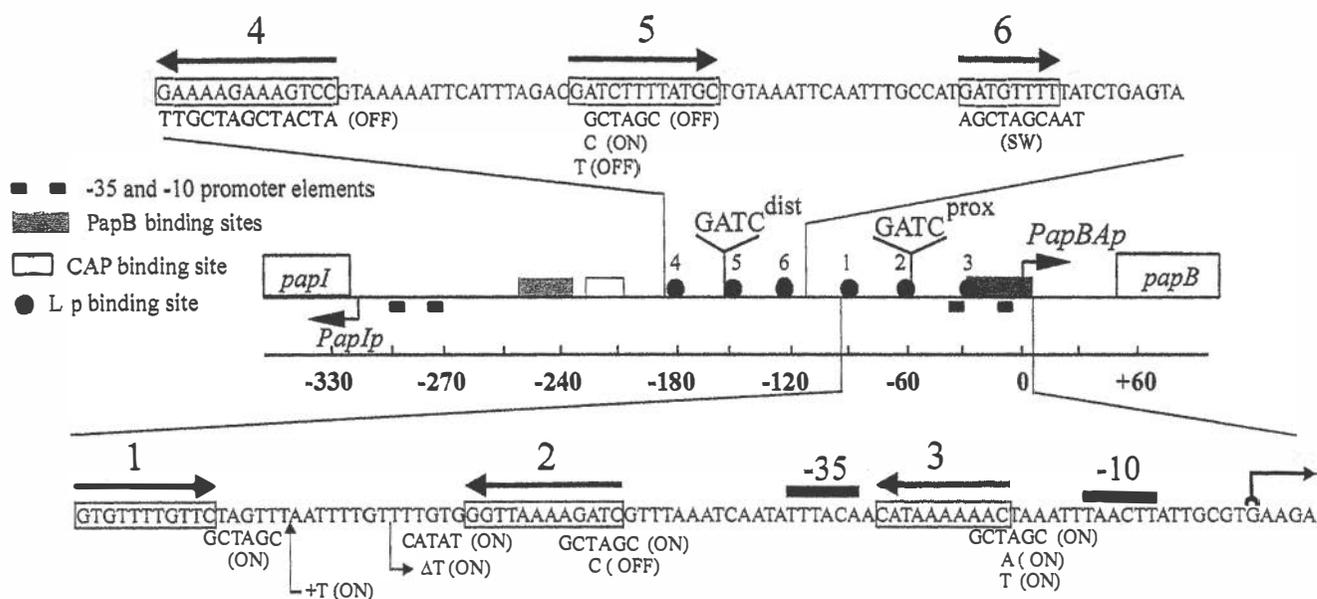

Fig. 1. Regulatory sequences of the *E. coli* pap operon. The *pap* regulatory between the divergently transcribed *papBA* and *papI* promoters is depicted. The two GATC sites subject to methylation by Dam are GATC^prox and GATC^dist, which are located within Lrp binding sites 2 and 5, respectively. The Lrp sites are shown as filled circles and the Lrp binding sites are shown as boxed regions on the expanded DNA sequence. The orientation of the Lrp sites [using a consensus sequence 5′-Gnn(n)TTTt-3′] is indicated with arrows above the sequence. The distance between sites 2 and 5 is 102 bp, and the distance between sites 1 and 6 is 32 bp, measured between conserved base-pairs within the Lrp binding sites (Fig. 6). The CAP and PapB binding sites are shown as open and hatched boxes, respectively. Substitution, deletion, and insertion mutations are shown below the wild-type sequence. Mutant switch phenotypes are indicated in parentheses.
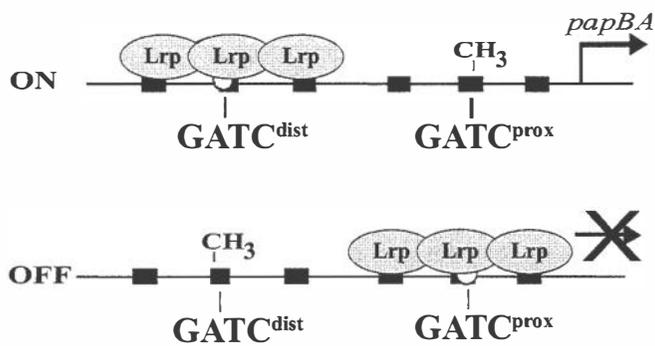
Fig. 2. DNA methylation states of phase ON and phase OFF cells. Binding of Lrp at sites 4–6 in phase ON cells and sites 1–3 in phase OFF cells controls the DNA methylation pattern by blocking methylation of the bound GATC site.

moter, including Lrp binding sites, have been deleted (486 MU), and thus likely represents basal *papBA* promoter activity (13). The mechanism by which H-NS represses *pap* basal transcription is not known, but likely involves specific binding to *pap* regulatory sequences (6).

Further evidence that Lrp directly represses *pap* transcription comes from *in vitro* analysis of *pap* transcription. Addition of

RNA polymerase-σ^{70} to a supercoiled *pap* DNA template containing the *papBA* promoter resulted in Lrp- and cAMP-CAP-independent transcription (intrinsic *papBA* promoter activity), with a transcription start site identical to that observed *in vivo* (22). Lrp was titrated while simultaneously monitoring transcription by primer extension and Lrp binding by *in vitro* methylation protection (IVMP). In IVMP, first applied by van der Woude *et al.* (12), binding of Lrp to the GATC^{dist} and GATC^{prox} sites is monitored by addition of Dam followed by restriction enzyme *Mbo*I, which cuts at fully nonmethylated GATC sites. It was observed that half-maximal inhibition of *papBA* transcription occurred at the same level of Lrp that gave half-occupancy of the GATC^{prox} site. Thus, binding of Lrp at the GATC^{prox} region (sites 1–3) correlated with repression of *pap* transcription. Moreover, mutational disruption of Lrp binding site 3, which reduced the affinity of Lrp for sites 1–3 by about 40-fold (from a K_d of 0.5 nM to 20 nM, measured using intact *pap* DNA, Fig. 5) also abrogated Lrp-dependent inhibition of *pap* transcription. Together, these results strongly indicate that binding of Lrp at sites 1–3 in phase OFF cells directly blocks the intrinsic *papBA* promoter activity.

The Phase OFF to Phase ON Transition. The Pap phase OFF to ON switch rate is about 100-fold lower than the phase ON to OFF rate (23), resulting in a bias in bacterial populations toward the
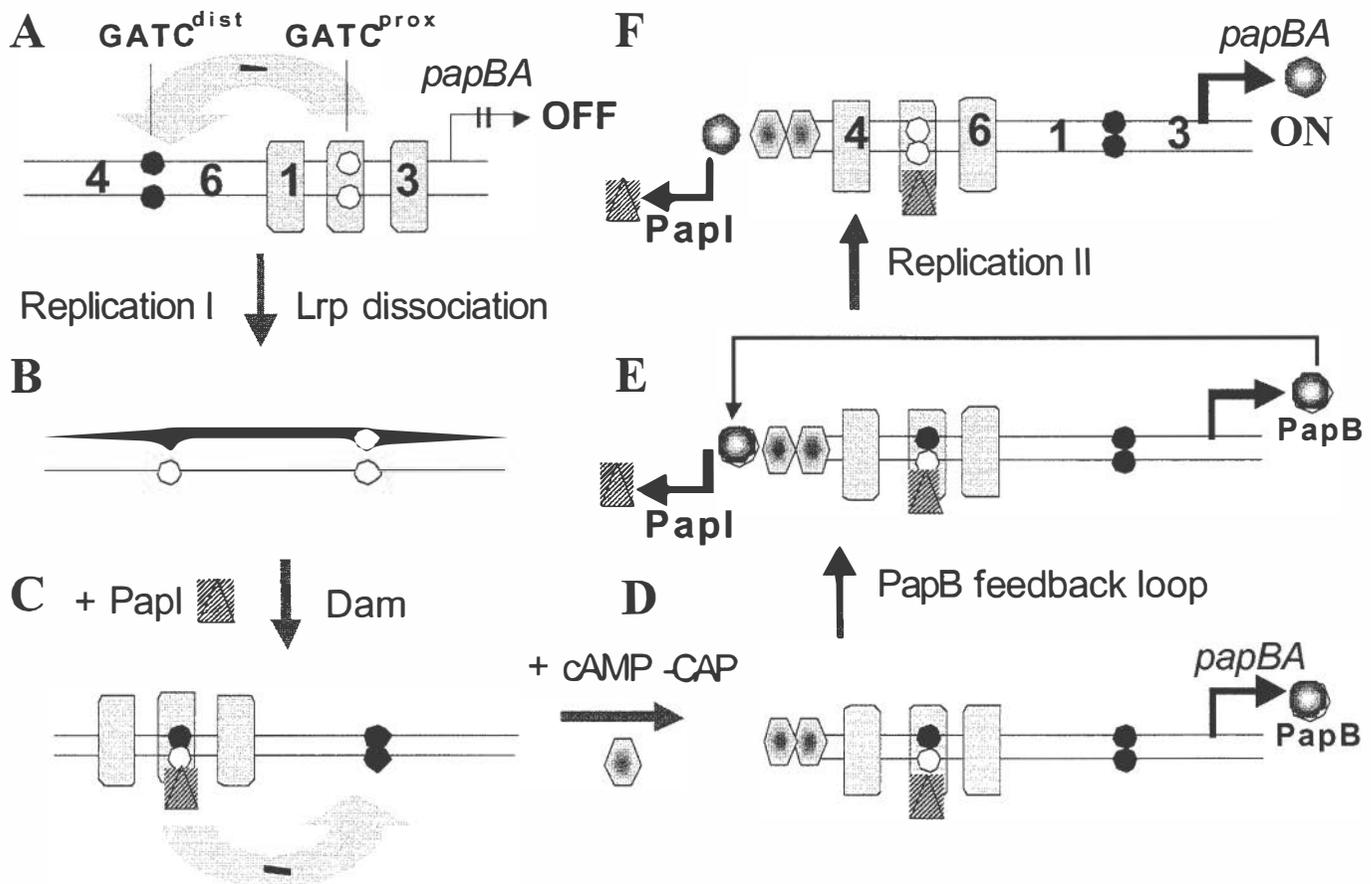

Fig. 3. Pap phase variation model. A model for the transition from phase OFF to phase ON is shown in panels *A–F*. In *A*, the OFF state is shown in which Lrp binds to sites 1–3, blocking methylation of GATC^{prox} and inhibiting *papBA* transcription. Nonmethylated *pap* GATC sites are depicted as open circles, and methylated sites are depicted as closed circles. The binding of Lrp at sites 1–3 reduces the affinity of Lrp for sites 4–6 ("mutual exclusion," see Fig. 5), and is depicted by an arrow with a negative sign. Transition to the phase ON state is hypothesized to require DNA replication, which dissociates Lrp from *pap* DNA generating two daughter duplexes of which one is shown (*B*). PapI and Dam work together to increase the affinity of Lrp for sites 4–6 as discussed in the paper (*C*). Activation of *pap* transcription requires cAMP-CAP (*D*), which stimulates PapB transcription (*E*). Because PapB activates *papI* transcription, the phase ON state is self-perpetuating (*F*).

		Dam Background	
	W.T.	Dam –	Dam+++
	Switching	OFF	OFF
	ON	ON	ON
GCTCprox	OFF	OFF	OFF
	OFF	OFF	OFF
GCTCdist	ON	OFF	ON ⇧

Fig. 4. The effects of DNA methylation on Pap phase variation. Wild-type and mutant *pap* DNA regulatory regions are shown at the left. Substitutions in Lrp binding sites 3 and 4 are depicted by an "X" and point mutations in GATCdist and GATCprox that disrupt DNA methylation are shown as "GCTC" sequences. The Pap switch phenotypes under variation levels of Dam are shown at the right. "Switching" indicates reversible phase variation, "OFF" indicates locked OFF, "ON" indicates locked ON, and "ON" with an arrow indicates up-regulation of *pap* pilin transcription. At the upper right, "W.T." indicates wild-type levels of Dam, "Dam–" indicates a deletion of the *dam* gene, and "Dam+++" indicates overproduction of Dam (>4-fold). Data are from refs. 8 and 10.

OFF state. This is generally true for most pili operons, including type 1 pili, regardless of their switch mechanisms (4).

We hypothesize that an opportunity for transition to the transcriptionally active ON phase only occurs after DNA replication (Fig. 3B). First, replication should dissociate Lrp from sites 1–3, removing the mutual exclusion effect on binding of Lrp at sites 4–6. Second, GATCdist will become transiently hemimethylated, providing an opportunity for binding of Lrp to sites 4–6 with the aid of PapI, which is required together with Lrp for methylation protection of GATCdist (9) and transition to the phase ON state (Fig. 3C). Third, dissociation of Lrp from sites 1–3 provides an opportunity for methylation of GATCprox by Dam, which is essential for *pap* transcription (10). These individual steps in the OFF to ON transition are discussed below.

Role of PapI. PapI is a small (8 kDa) coregulatory protein that shares homology only with other PapI-like genes present in many

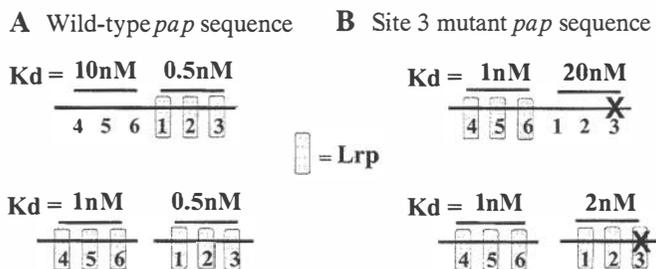**A** Wild-type *pap* sequence   **B** Site 3 mutant *pap* sequence

Fig. 5. Quantitation of the mutual exclusion effect on LRP binding by UV footprint analysis. DNA fragments containing the intact *pap* regulatory region (*Upper*) or unlinked regions containing only Lrp binding sites 1–3 or 4–6 (110 bp each) (*Lower*) were cloned into plasmid vector pTZ19U (41). Supercoiled plasmids were isolated from a Dam⁻ strain, and the affinity of Lrp for sites 1–3 and 4–6 was measured by UV footprinting as described (42). Briefly, samples were irradiated at 254 nm, and UV-induced pyrimidine dimers were detected by extension of ³²P-end-labeled primers with *Taq* DNA polymerase. The affinities of Lrp for *pap* DNA sites 1–3 and 4–6 were identical to affinities determined by using electrophoretic mobility shift analysis (unpublished data). The location of a 6-bp substitution mutation in site 3 (see Fig. 1) is depicted by an "X" in *B*.

Fig. 6. Sequence comparison of Lrp binding sites. The consensus Lrp binding sequence as determined by SELEX (43) is shown on the bottom line (Y = C or T, H = not G, W = A or T, D = not C, R = A or G). Lrp binding sites 1 through 6 of *pap* and *ilvIH* binding site 2 are shown above the SELEX sequence. The abilities of PapI to increase the affinity of Lrp for each site is shown at the far right, where "Yes" indicates an increase in Lrp affinity.

diverse pili operons in *Escherichia coli, Salmonella typhimurium*, and likely other enteric bacteria (5). Two functions of PapI have been identified: specific binding to Lrp (24) and specific binding to DNA sequences within *pap* sites 2 and 5 (A.H. and D.L., unpublished data). Notably, the affinity of PapI for *pap* DNA alone is very low, and cannot be detected by electrophoretic mobility shift analysis. Moreover, the affinity of PapI for free Lrp is also low based on protein crosslinking (24) and gel filtration analyses (24). PapI binds specifically with high affinity to Lrp in complex with *pap* sites 2 and 5, but not to Lrp bound to other *pap* sites or Lrp binding sites within the *ilvIH* regulatory sequence (ref. 24 and A.H. and D.L., unpublished data). This sequence selectivity appears to be caused by the presence of a conserved core sequence that includes "ACGATC" in *pap* sites 2 and 5 (Fig. 6) containing base pairs critical for PapI binding. Our data indicate that Lrp bound at adjacent and partially overlapping sites interacts with PapI, stabilizing PapI–DNA interactions. This is manifested by over 20-fold reduction in the dissociation rate of Lrp from *pap* DNA (M.K. and D.L., unpublished data) and a 10-fold increase in the affinity of Lrp for sites 2 or 5 (A.H. and D.L., unpublished data). Because binding of Lrp to *pap* DNA is highly cooperative, PapI increases the affinity of Lrp for these sites by forming PapI–Lrp–DNA complexes at sites 2 or 5.

Roles of Dam. The finding that PapI increases the affinity of Lrp for sites 4–6 based on *in vitro* binding data are consistent with the observation that PapI is required *in vivo* for methylation protection of GATCdist present in site 5. However, PapI also increases Lrp's affinity for sites 1–3, which raises the problem of how Lrp moves from promoter proximal to promoter distal sites in the phase OFF to ON transition. A possible answer involves DNA methylation. As shown in Fig. 4, wild-type *pap* is transcriptionally inactive in the absence of Dam. In addition, a normally locked ON GCTCdist mutant also requires Dam for transcription (10) and a GCTCdist GCTCprox double mutant is locked OFF, indicating that methylation of GATCprox is essential for transition to the phase ON state. Apparently, methylation of GATCprox in phase ON cells is not saturating because overproduction of Dam significantly increases transcription in the phase-locked ON GCTCdist mutant (10). Dam, however, is not required for *pap* transcription when Lrp binding site 3 is disrupted by substitution (Fig. 4), suggesting that methylation at GATCprox might inhibit binding of Lrp and/or PapI-Lrp to sites 1–3. Recent data indicate that PapI-dependent binding of Lrp at GATCprox is blocked by methylation at this site, although Lrp binding is unaffected (A.H. and D.L., in preparation). In contrast, methylation at GATCdist blocks binding of Lrp, but has much less affect on PapI dependent Lrp binding. Based on these data, we conclude that methylation of GATCprox may be required

to provide directionality to the switch by reducing the affinity of PapI–Lrp for sites 1–3.

Roles of CAP. CAP plays important roles in activation at the *papBA* and *papI* promoters (14–16, 25). Binding of CAP to a single site located at −215.5 bp from the *papBA* start and −115.5 bp from the *papI* start is essential for activation of both *papBA* and *papI* transcription. CAP directly activates transcription at the *papBA* promoter (16). However, recent data suggest that CAP activates *papI* transcription indirectly by means of expression of PapB regulatory protein because activation of *papI* transcription occurs in the absence of CAP when *papB* is expressed from an independent promoter (manuscript in preparation). CAP appears to affect Pap phase switching indirectly by means of its control of PapI transcription because movement of Lrp from sites 1–3 to 4–6 occurs *in vivo* in the absence of CAP under conditions in which transcription is PapI-independent (16). In addition, although the CAP binding site is centered 36 bp from Lrp binding site 4, Lrp and CAP bind independently to their respective binding sites (16). The mechanism by which PapB activates *papI* transcription appears to involve binding of PapB at a high affinity binding site adjacent to the CAP binding site (17, 18) (Fig. 1).

Recently, the mechanism by which CAP stimulates transcription at the *papBA* promoter has been explored (16). In many respects, activation of *papBA* transcription by CAP is similar to activation of class I promoters by CAP (*lac* for example), even though the distance between CAP and the *papBA* transcription start site (215.5 bp) is much larger than *lac* (61.5 bp) and other class I operons (see Fig. 1). CAP-dependent activation of transcription of the *papBA* promoter and class I operons (26) share a requirement for the following: (*i*) activating region 1 (AR1) of CAP, (*ii*) α C-terminal domain of RNA polymerase, (*iii*) helical phase dependence between CAP and RNA polymerase, and (*iv*) only the promoter–proximal subunit of the CAP homodimer is required. These results clearly show that CAP plays a direct role in activation of *papBA* transcription by contact with the transcription apparatus. Although a previous study suggested that CAP may activate *papBA* transcription indirectly by means of antagonism with the histone-like protein H-NS (25), this does not appear to be the case because CAP AR1 is required for transcription even in the absence of H-NS (16).

Role of PapB. Early studies by Uhlin's laboratory established that PapB plays an essential role in activation of *pap* transcription (14). There is a least one high affinity PapB binding site, located between the CAP binding site and the *papI* promoter in each *pap* operon that has been studied (Fig. 1) (18). Binding of PapB to this high-affinity site is essential for *papI* transcription and subsequent Pap pili expression. PapB is a 12-kDa regulatory protein that binds to the minor grooves of DNA targets containing the nonamer GACACAAAC (18). PapB appears to bind as a multimer of 8–10 subunits protecting a DNA region of 50–70 nucleotides (18). The role of PapB in activation of Pap transcription is solely due to activation at the *papI* promoter because PapB is not essential for Pap phase variation under conditions in which PapI is constitutively expressed (5). PapB thus acts as a feedback link between the *papBA* and *papI* promoters, a link that is involved in self-perpetuation of the phase ON state (see below).

Role of H-NS. The H-NS, unlike Lrp, CAP, PapI, Dam, and PapB, is not essential for Pap phase variation (13, 16). However, this 15.5-kDa global regulatory protein does modulate Pap phase switching, and appears to play an essential role in the repression of Pap expression by environmental signals including low temperature (27). At temperatures below 26°C, H-NS appeared to bind to the *pap* regulatory region as evidenced by methylation

protection of the two *pap* GATC sites (6). This conclusion was supported by the finding that H-NS binds specifically to *pap* DNA and blocks methylation of GATC^dist and GATC^prox *in vitro* (6). Simultaneous analysis of *pap* transcription and the *pap* DNA methylation pattern after a shift to low temperature showed that transcription shut off before switching to the OFF state, indicating that H-NS blocks transcription from phase ON cells at low temperature, possibly as a result of the formation of a nucleo-protein repression complex.

At 37°C, H-NS appears to play both positive and negative roles in modulating Pap expression. In the absence of *hns*, the *pap* phase OFF to ON switch rate is reduced 2- to 3-fold (7, 13), indicating that H-NS helps promote transition to the phase ON state. In contrast, H-NS was shown to inhibit the basal activity of the *papBA* promoter in the absence of Lrp (13). The mechanism by which H-NS exerts these affects is not known, but likely involves specific binding within the *pap* regulatory region. H-NS could positively regulate OFF to ON switching by reducing the affinity of Lrp for sites 1–3, and negatively regulate *papBA* transcription by altering the interactions of RNA polymerase with the promoter.

The Self-Perpetuating Phase ON State. Addition of Lrp, cAMP-CAP, and RNA polymerase-σ^70 with *pap-13* template DNA (*pap-13* contains a Lrp binding site 3 mutation making transcription PapI-independent) activates *papBA* transcription about 12-fold compared with basal transcription measured with RNA polymerase alone (16). This relative activation level is similar to that observed *in vivo* (13), indicating that Lrp (bound at sites 4–6) and cAMP-CAP are sufficient for the activation of *papBA* transcription observed in phase ON cells. As described above, a PapB feedback loop ensues because of transcription and expression of PapB, which activates *papI* transcription. PapI, in turn, will facilitate movement of Lrp to sites 4–6, maintaining the switch in the ON state (Fig. 3E). If this model is correct, then *papI* and *papBA* transcription should be coordinated ON or OFF, depending on the phase state. Although this has not yet been tested, analysis of a *papI–lacZ* fusion shows that the *papI* promoter is subject to phase variation control with similar OFF and ON rates to the *papBA* promoter. Moreover, mutations in Lrp binding sites 4 and 5 block both *papI* and *papBA* transcription (A. Brinkman, N. Weyand, and D.L., manuscript in preparation). These data support the hypothesis that *papI* and *papBA* transcription is coordinated by a PapB feedback loop.

The PapB feedback loop is subject to autoregulation, which may prevent spiraling up of *papI* and *papB* transcription (17). Titration of PapB showed that, at low levels, *papBA* transcription was enhanced, whereas at higher levels transcription was inhibited. Autoregulation appears to be caused by the presence of a low-affinity PapB binding site located overlapping the −10 hexamer RNA polymerase binding site in the *papBA* promoter (Fig. 1), though this has not been directly shown. Under certain conditions, PapI might also act as an autoregulator. Overexpression of PapI in normally phase-locked ON cells containing a *pap* GCTC^dist mutation results in a switch phenotype in which both phase ON and OFF colonies are present after plating on solid media, indicating that PapI induces some cells to turn off (unpublished data). This could occur by means of PapI-assisted binding of Lrp to sites 1–3, which normally is inhibited by methylation of GATC^prox. After DNA replication, the fully methylated GATC^prox in phase ON cells becomes hemimethylated for a short period before remethylation by Dam. It is possible that a single methyl group on the top or bottom *pap* GATC^prox site may not inhibit PapI-dependent binding of Lrp at sites 1–3 when high PapI levels are present. This hypothesis could be tested by titration of PapI with analysis of phase switch rates by fluorescent activated cell sorting using a fluorescent reporter

for *papBA* expression. High PapI levels should increase the switch rate to the OFF phase.

Switch Inputs

Pap pili-phase variation is controlled by a variety of environmental stimuli (7, 14). Growth of *E. coli* in glucose reduces the OFF to ON switch rate by 35-fold as a result of lowered cAMP level, which prevents CAP-dependent activation of *papBA* and subsequent *papI* transcription (23). Pap pilin transcription is also significantly repressed by growth at low temperature (<26°C) (27, 28) and rich medium (LB broth) (7). It is not clear whether these growth conditions directly alter the phase switch itself or alternatively inhibit *pap* transcription in phase ON cells. The mechanism by which these conditions repress *pap* transcription is not known, but appears to involve H-NS because introduction of an *hns* mutation partially relieves repression (7, 27).

In addition to negative regulatory inputs, recently a positive input to *pap* transcription was described by the Silhavy and Hultgren laboratories (29, 30). The CpxAR two-component system consists of the CpxA membrane sensor and the CpxR response regulator. This sensor system appears to be activated by misfolded proteins in the periplasm such as unchaperoned Pap pilin subunits as well as binding of Pap pili to solid surfaces (29), which initiate a phosphotransfer relay from CpxA to CpxR. Once phosphorylated, CpxR-P controls transcription of a number of genes, including *degP* and the *pap* operon. It was shown that when the Cpx pathway is activated, *papI* and *papBA* transcription was enhanced 2-fold. Moreover, Pap pili were expressed under conditions of catabolite repression (+ glucose) when the Cpx pathway was activated (30). Recent data from our laboratory have indicated that under conditions of constitutive phosphorylation of CpxR, the phase OFF to ON rate increases 3-fold. Our data show that CpxR-P binds specifically to the *pap* site 1–3 region, which could increase phase ON switching by reduction of the affinity of Lrp for binding sites 1–3, similar to the phenotypes of mutations in *pap* sites 2 and 3 (see Fig. 1) (P. Engelbert and D.L., in preparation).

A number of non-*pap* pili operons including *sfa* (S pili), *daa* (F1845 pili), and *fae* (K88 pili) in *E. coli* and *pef* (Pef pili) in *S. typhimurium* share common regulatory features with *pap* (4). These include conserved GATC^dist and GATC^prox sites as well as *papI* and *papB* homologues. Although some of these regulatory sequences (*sfa* for example) contain a conserved CAP binding site at the same position as that in *pap*, others such as *fae* and *pef* do not (4). This finding raises the possibility that additional regulatory inputs could control these pili operons. This is the case with *pef* expression, which, though normally heavily biased to the OFF expression state, is induced by growth under acidic conditions to switch ON (pH 4.5) (4, 31). The mechanism by which Pef pili are induced by low pH is unknown, but induction cannot be caused by activation of PefI expression (a PapI homologue) because PefI acts as a negative regulator of *pef* transcription (31). Compared with *pap*, the organization of *pef* may be reversed so that the binding of Lrp to *pef* sites 4–6 occurs with the highest affinity, with PefI facilitating transition to the OFF state by movement of Lrp to sites 1–3 (31).

Outputs

As a result of the PapB feedback loop, phase ON cells expressing Pap pili also express relatively high levels of PapB and PapI, which control other genes in *E. coli* besides *pap*. Many uropathogenic *E. coli* contain multiple pili operons, many of which share the core control mechanisms of the *pap* operon including cross-complementing PapI and PapB homologues. Analysis of the *pap*-related fimbiae (*prf*) operon of the uropathogenic *E. coli* strain 536 showed that deletion of the *prfI* and *prfB* genes,

homologues of *papI* and *papB*, reduced the expression of S pili encoded by the unlinked *sfa* operon (32). Introduction of constitutively expressed PrfB or PrfI restored S pili expression, indicating that cross-activation of *sfa* was occurring. In contrast, little if any cross-talk between the *pap-17* and *pap-21* pili operons in *E. coli* C1212 appears to occur based on comparison of Pap-17 and Pap-21 pili expression on individual cells when only one versus both operons are present (33). Further work needs to be done to understand the extent of cross-talk between *pap*-like operons.

More recently, it has been shown that PapB greatly reduces the expression of type 1 pili by a form of regulatory cross-talk (34, 35). Type 1 pili are regulated by a DNA inversion-mediated phase variation switch catalyzed by two DNA recominbinases, FimB and FimE. The former enzyme enhances both OFF to ON and ON to OFF switching, whereas the latter enzyme mediates ON to OFF switching only (36, 37). PapB was found to increase FimE-dependent ON to OFF switching by 2-fold, and inhibited FimB-dependent switching by over 50-fold, thus blocking switching to the ON phase while speeding up switching to the OFF phase (34). Although it has not been directly shown, these data suggest that expression of Pap pili and type 1 pili is mutually exclusive.

Biological Relevance. As described above, the Pap pili phase variation mechanism is highly complex and tightly regulated by many components that contribute to reversible OFF–ON switching of Pap pili. One reason for this complexity may be that it allows for global (Lrp, CAP, Dam, H-NS) as well as local (PapI, PapB) regulatory inputs that provide a means for environmental factors to regulate Pap phase switch rates (see above). If only a small fraction of *E. coli* express Pap pili in a specific environment and if that confers a selective advantage on *E. coli*, then the cell population will rapidly convert to phase ON cells because of the heritable nature of Pap pili expression states. Thus, environmental stimulation of Pap phase OFF to ON switching in the appropriate milieu should accelerate *E. coli*'s colonization of that environmental niche. In addition, environmental control would help to conserve cell resources when pili are not needed because pili expression requires a significant fraction of the cells energy.

Pap pili phase variation can be thought of as a "simple" developmental switch mechanism that controls cell differentiation. Not only do *pap* phase ON cells express Pap pili at their surface, but type 1 pili expression is turned off as a result (34). Although the physiological relevance of this effect is not known, Pap and type 1 pili bind to different receptors and have distinct roles in pathogenesis: type 1 pili are required for colonization of the lower urinary tract by uropathogenic *E. coli* (38), whereas Pap pili play an essential role in colonization of the upper urinary tract (39). Because inappropriate expression of pili may be deleterious to *E. coli* by enhancing detection by the immune system, PapB-mediated shut-off of type 1 pili may be important in the biology of *E. coli*. Recent data suggest that *pap* gene expression may also negatively control cell motility by means of inhibition of flagellar expression (40). This regulatory mechanism might facilitate colonization of mucosal surfaces of the intestines or urinary tract by cycling between motile and sessile states that could counteract bacterial removal by peristalsis and urine output, respectively. Comparative microarray analysis of gene expression in Pap phase OFF and phase ON cells should reveal whether additional regulatory outputs are present.

1. Fulks, K. A., Marrs, C. F., Stevens, S. P. & Green, M. R. (1990) *J. Bacteriol.* **172,** 310–316.
2. Blomfield, I. C., Kulasekara, D. H. & Eisenstein, B. I. (1997) *Mol. Microbiol.* **23,** 705–717.
3. Mehr, I. J. & Seifert, H. S. (1998) *Mol. Microbiol.* **30,** 697–710.
4. Krabbe, M., Weyand, N. & Low, D. (2000) in *Bacterial Stress Responses*, eds. Storz, G. & Hengge-Aronis, R. (Am. Soc. Microbiol., Washington, DC), pp. 305–321.
5. van der Woude, M., Braaten, B. & Low, D. (1996) *Trends Microbiol.* **4,** 5–9.
6. White-Ziegler, C. A., Angus Hill, M. L., Braaten, B. A., van der Woude, M. W. & Low, D. A. (1998) *Mol. Microbiol.* **28,** 1121–1137.
7. White-Ziegler, C. A., Villapakkam, A., Ronaszeki, K. & Young, S. (2000) *J. Bacteriol.* **182,** 6391–6400.
8. Nou, X., Braaten, B., Kaltenbach, L. & Low, D. A. (1995) *EMBO J.* **14,** 5785–5797.
9. Nou, X., Skinner, B., Braaten, B., Blyn, L., Hirsch, D. & Low, D. (1993) *Mol. Microbiol.* **7,** 545–553.
10. Braaten, B. A., Nou, X., Kaltenbach, L. S. & Low, D. A. (1994) *Cell* **76,** 577–588.
11. Braaten, B. A., Blyn, L. B., Skinner, B. S. & Low, D. A. (1991) *J. Bacteriol.* **173,** 1789–1800.
12. van der Woude, M., Hale, W. B. & Low, D. A. (1998) *J. Bacteriol.* **180,** 5913–5920.
13. van der Woude, M. W., Kaltenbach, L. S. & Low, D. A. (1995) *Mol. Microbiol.* **17,** 303–312.
14. Baga, M., Goransson, M., Normark, S. & Uhlin, B. E. (1985) *EMBO J.* **4,** 3887–3893.
15. Goransson, M., Forsman, P., Nilsson, P. & Uhlin, B. E. (1989) *Mol. Microbiol.* **3,** 1557–1565.
16. Weyand, N. J., Braaten, B. A., van der Woude, M., Tucker, J. & Low, D. A. (2001) *Mol. Microbiol.* **39,** 1504–1522.
17. Forsman, K., Goransson, M. & Uhlin, B. E. (1989) *EMBO J.* **8,** 1271–1277.
18. Xia, Y., Forsman, K., Jass, J. & Uhlin, B. E. (1998) *Mol. Microbiol.* **30,** 513–523.
19. Chen, S., Hao, Z., Bieniek, E. & Calvo, J. M. (2001) *J. Mol. Biol.* **314,** 1067–1075.
20. Chen, S., Rosner, M. H. & Calvo, J. M. (2001) *J. Mol. Biol.* **312,** 625–635.
21. Wang, Q. & Calvo, J. M. (1993) *EMBO J.* **12,** 2495–2501.
22. Weyand, N. J. & Low, D. A. (2000) *J. Biol. Chem.* **275,** 3192–3200.
23. Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H. & Low, D. A. (1989) *EMBO J.* **8,** 613–620.
24. Kaltenbach, L. S., Braaten, B. A. & Low, D. A. (1995) *J. Bacteriol.* **177,** 6449–6455.
25. Forsman, K., Sonden, B., Goransson, M. & Uhlin, B. E. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9880–9884.
26. Busby, S. & Ebright, R. H. (1999) *J. Mol. Biol.* **293,** 199–213.
27. Goransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K. & Uhlin, B. E. (1990) *Nature (London)* **344,** 682–685.
28. Goransson, M. & Uhlin, B. E. (1984) *EMBO J.* **3,** 2885–2898.
29. Otto, K. & Silhavy, T. J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 2287–2292.
30. Hung, D. L., Raivio, T. L., Jones, C. H., Silhavy, T. J. & Hultgren, S. J. (2001) *EMBO J.* **20,** 1508–1518.
31. Nicholson, B. & Low, D. (2000) *Mol. Microbiol.* **35,** 728–742.
32. Morschhauser, J., Vetter, V., Emody, L. & Hacker, J. (1994) *Mol. Microbiol.* **11,** 555–566.
33. Low, D., Robinson, E. N., Jr., McGee, Z. A. & Falkow, S. (1987) *Mol. Microbiol.* **1,** 335–346.
34. Xia, Y., Gally, D., Forsman-Semb, K. & Uhlin, B. E. (2000) *EMBO J.* **19,** 1450–1457.
35. Holden, N. J., Uhlin, B. E. & Gally, D. L. (2001) *Mol. Microbiol.* **42,** 319–330.
36. Gally, D. L., Leathart, J. & Blomfield, I. C. (1996) *Mol. Microbiol.* **21,** 725–738.
37. McClain, M. S., Blomfield, I. C. & Eisenstein, B. I. (1991) *J. Bacteriol.* **173,** 5308–5314.
38. Martinez, J. J., Mulvey, M. A., Schilling, J. D., Pinkner, J. S. & Hultgren, S. J. (2000) *EMBO J.* **19,** 2803–2812.
39. Roberts, J. A., Marklund, B. I., Ilver, D., Haslam, D., Kaack, M. B., Baskin, G., Louis, M., Mollby, R., Winberg, J. & Normark, S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11889–11893.
40. Li, X., Rasko, D. A., Lockatell, C. V., Johnson, D. E. & Mobley, H. L. (2001) *EMBO J.* **20,** 4854–4862.
41. Mead, D. A., Szczesna-Skorupa, E. & Kemper, B. (1986) *Protein Eng.* **1,** 67–74.
42. Becker, M. M. & Grossmann, G. (1993) in *Footprinting of Nucleic Acid–Protein Complexes*, ed. Revzin, A. (Academic, San Diego), pp. 129–157.
43. Cui, Y., Wang, Q., Stormo, G. D. & Calvo, J. M. (1995) *J. Bacteriol.* **177,** 4872–4880.

Colloquium

Histone H3 variants specify modes of chromatin assembly

Kami Ahmad and Steven Henikoff*

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, A1-162, Seattle, WA 98109

Histone variants have been known for 30 years, but their functions and the mechanism of their deposition are still largely unknown. *Drosophila* has three versions of histone H3. H3 packages the bulk genome, H3.3 marks active chromatin and may be essential for gene regulation, and Cid is the characteristic structural component of centromeric chromatin. We have characterized the properties of these histones by using a *Drosophila* cell-line system that allows precise analysis of both DNA replication and histone deposition. The deposition of H3 is restricted to replicating DNA. In striking contrast, H3.3 and Cid deposit throughout the cell cycle. Deposition of H3.3 occurs without any corresponding DNA replication. To confirm that the deposition of Cid is also replication-independent (RI), we examined centromere replication in cultured cells and neuroblasts. We found that centromeres replicate out of phase with heterochromatin and display replication patterns that may limit H3 deposition. This confirms that both variants undergo RI deposition, but at different locations in the nucleus. How variant histones accomplish RI deposition is unknown, and raises basic questions about the stability of nucleosomes, the machinery that accomplishes nucleosome assembly, and the functional organization of the nucleus. The different *in vivo* properties of H3, H3.3, and Cid set the stage for identifying the mechanisms by which they are differentially targeted. Here we suggest that local effects of "open" chromatin and broader effects of nuclear organization help to guide the two different H3 variants to their target sites.

Nucleosomes are the fundamental units of chromatin, consisting of 146 bp of DNA wrapped around an octamer of four core histones. Histone deposition occurs primarily as DNA replicates to complete chromatin doubling (1). During S phase of the cell cycle, new histones are produced in abundance for immediate replication-coupled deposition. In most metazoans, this abundant S-phase synthesis results from the tight regulation of tens to hundreds of intronless histone genes that have special 3′ untranscribed regions instead of poly(A) tails (2). However, some histones are produced from orphan genes outside of S phase. In *Drosophila*, orphan genes encode two H3 variants: one encodes Cid, the centromeric histone (3), and two encode H3.3, the replacement variant (Fig. 1; refs. 4 and 5). These variants have equivalents in many other eukaryotes (6, 7). The H3.3 histone is nearly identical to H3, differing at only four amino acid positions. Cid differs profoundly from H3 in sequence, showing some significant identity only within the histone fold domain. Surprisingly, these three histones have different deposition properties. H3 and H3.3 are deposited as DNA replicates, but both H3.3 and Cid can be deposited at sites that are not undergoing DNA replication (Fig. 2; refs. 8 and 9). Whereas only a minor fraction of the bulk genome is packaged into Cid- and H3.3-containing nucleosomes, each variant is targeted to different specialized sites, with Cid localizing to centromeres and H3.3 to transcriptionally active genes. Specific localization of centromeric H3-like histones (CenH3s) has been observed in various animals (10), fungi (11, 12), and plants (13). Also, an H3.3-like histone targets the transcriptionally active macronucleus in

Fig. 1. *Drosophila* produces three versions of H3 histones. H3 and H3.3 are nearly identical throughout their N-terminal tail and histone fold domains (HFD), with only four amino acid residue differences. Cid is much more diverged, and can only be aligned to H3 in the HFD (49). Black boxes indicate identities to H3.

ciliates (14). Thus, the targeting of H3 variants is likely a feature of every eukaryotic cell, where centromeres and transcribed regions are the major loci of activity in metaphase and interphase, respectively. Both kinds of loci use a distinct pathway for nucleosome assembly (8, 9), and here we explore the properties of this process.

H3 Variants Determine the Nucleosome Assembly Pathway

Studies of histone deposition have generally been done using crude extracts, purified components or pools of cells from which bulk chromatin is extracted (15). These methods reveal the average properties of chromatin, and have shown that the bulk of chromatin doubles as DNA replicates. Extensive *in vitro* work has demonstrated that the assembly of nucleosomes is a stepwise process in which deposition of an $(H3 \cdot H4)_2$ tetramer is followed by addition of two H2A·H2B dimers (1). The new histones are brought to the replication fork in a complex with chromatin assembly factor 1 (CAF1). CAF1 appears to be recruited to the replication fork by binding to the ring-shaped proliferating cell nuclear antigen (PCNA) that encircles the DNA template at each replication fork (16). Histones from the parent DNA are distributively segregated to the two sister chromatids behind the replication fork, and the gaps in their nucleosomal arrays are rapidly filled by step-wise assembly of new nucleosomes. These nucleosomes are then matured by addition of linker histones and covalent modification of histone tails to complete chromatin.

Nucleosomes containing H3 variants comprise only a small proportion of bulk chromatin, and thus their properties have been generally undetectable. However, replacement H3 variants can become enriched in the chromatin of nonreplicating cells (17–19). This means that other ways of depositing histones must exist; but because such variant enrichment was only detectable in unusual cell types (such as long-lived neurons or spermato-
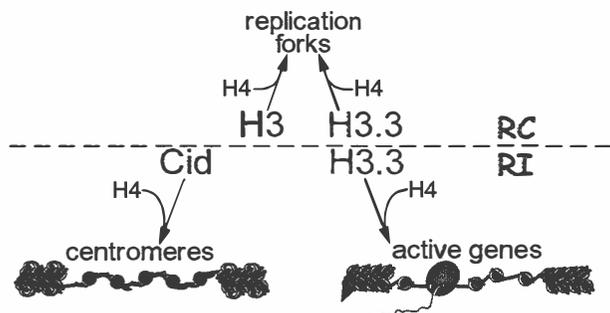
Fig. 2. The three versions of histone H3 determine the mode of nucleosome assembly. The deposition of H3 is strictly replication-coupled (RC), and H3 is recruited to replication forks for chromatin doubling. Deposition of Cid (blue) is exclusively replication-independent (RI), and normally occurs only at centromeres. H3.3 (green) undergoes RC deposition, and RI deposition at active loci. Open chromatin at centromeres and at active genes may promote histone replacement. Transcriptional activity, chromatin remodeling factors, and RNA polymerases (orange) will unfold the chromatin fiber and disrupt nucleosome (gray) structure (open chromatin). Transcriptionally inactive regions are not subjected to these forces and remain in a closed configuration. Flanking heterochromatin and H3-containing blocks within the centromeric domain presents the H3K9me epitope, thereby binding HP1 (red) and resulting in a compacted, closed chromatin structure. Cid-containing nucleosomes cannot be methylated in this way, and thus remain comparatively open. The specialized N-terminal tail of Cid may alter the linker DNA between nucleosomes, also contributing to the open chromatin configuration. RI deposition of H3.3 is limited to the open chromatin at active genes and RI deposition of Cid is limited to the open chromatin in centromeric domains.

Fig. 3. *Drosophila Kc* cells. (*A*) The cell cycle is ≈20 h long, and S phase has two distinct periods: early S phase, when all euchromatin (gray) replicates, and late S phase, when all heterochromatin (black) replicates. Eighty percent of cells show ≈15 chromosomes, and this karyotype has been stable for >2 years. (*B*) The morphology of a *Drosophila* interphase nucleus. All heterochromatin typically associates into a chromocenter (black). Centromeres (red) are enclosed within the chromocenter, with the nucleolus (light gray) next to it. The active rDNA genes (green) are located within the nucleolus.

cytes), studies of the phenomenon have been limited. The ability to tag histones and examine their deposition properties in single cells has allowed us to gain insight into chromatin assembly processes.

We developed a cytological assay system for studying replication and chromatin assembly by using *Drosophila Kc* cells, a cell line that displays a regular cell division schedule (Fig. 3*A*) and a consistent tetraploid karyotype. Organization of the *Drosophila* nucleus is visually simple, because the late-replicating heterochromatin typically coalesces into a compartment in the nucleus, termed the chromocenter (Fig. 3*B*). This provides both a temporal and spatial distinction between the early replicating, gene-rich euchromatin, and the late-replicating heterochromatin. DNA replication can be tracked either by pulse-labeling with nucleotide analogs or by using anti-PCNA antibody. Furthermore, by introducing histone-GFP fusion constructs and producing a pulse of the tagged protein, we can track histone deposition during the cell cycle. Using this system, we have been able to quantitatively examine DNA replication and histone deposition in unsynchronized populations of cells (3, 8, 9).

GFP-tagged H3 shows exclusively replication-coupled deposition, displaying co-localization with replication markers and showing no detectable deposition in cells in which replication has been blocked (9). The N-terminal tail of H3 is required, suggesting that the H3 tails of tetramer particles interact with accessory factors at some early step in nucleosome assembly *in vivo*.

In contrast to the properties of GFP-tagged H3 in cells, tagged H3.3 deposits in a replication-independent manner at actively transcribing loci (9). Deposition can occur in any stage of the cell cycle, and we demonstrated that it is not accompanied by unscheduled DNA synthesis. Incorporation of H4 also occurs at these target sites, as expected for deposition of (H3.3·H4)$_2$ tetramers; but how replication-independent (RI) histone deposition occurs is virtually unknown.

Tagged Cid can also deposit throughout the cell cycle (8), suggesting that its deposition is also replication-independent.
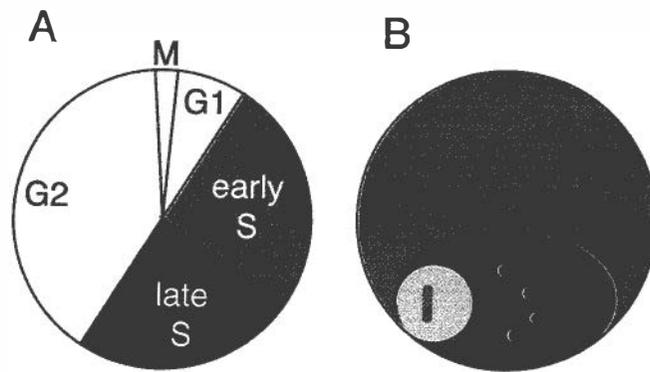
However, this conclusion depends on knowing the timing of centromere replication. We have shown that centromeres replicate within a defined portion of S phase (8). The evidence for this conclusion has been challenged (20), and so here we examine the available data on centromere replication timing. We confirm that *Drosophila* centromeres replicate as isolated domains within later-replicating heterochromatin.

Centromeres Replicate Before Their Surrounding Heterochromatin

Historically, centromeres have been thought to replicate very late in the cell cycle. This is because they are embedded within pericentric heterochromatin, which replicates late. Analysis has usually relied on visualization at mitosis; but mitotic chromosomes have inherently low resolution because they are highly condensed. Indeed, a recent study showed that *Drosophila* centromeres cannot be resolved from heterochromatin in 44% of spread mitotic chromosomes (21). Despite this limitation, Sullivan and Karpen (20) concluded from the analysis of normal mitotic chromosomes that Cid-containing chromatin replicates on the same late schedule as pericentric heterochromatin. However, this could be late replication in pericentric heterochromatin that was mis-scored as replication of centromeres.

We have addressed this uncertainty by analyzing mitotic chromosome replication patterns, providing brief 15-min pulses to *Kc* cells and examining mitotic figures after a chase. This provides a "snapshot" of replication at single points in the cell cycle. We observed examples of heterochromatin replication patterns similar to those previously reported (20), where labeling overlaps Cid spots (Fig. 4*A*). However, we also observed unambiguous examples of chromosomes that were intensely labeled throughout the euchromatic arms, with foci directly coinciding with centromeres (Fig. 4*B*). These centromeric foci are surrounded by heterochromatin that did not replicate during the labeling pulse. We attribute the inability of Sullivan and Karpen to observe early centromeric replication foci to the continuous labeling protocol they used, where all surrounding heterochromatin will always be labeled when earlier-replicating sites acquire label. When late-replicating heterochromatin is labeled, this will often overlap Cid-containing chromatin. If overlapping heterochromatin accounts for the apparent late replication of centromeres when analyzed on mitotic chromosomes, then its removal should improve the visualization of earlier-replicating
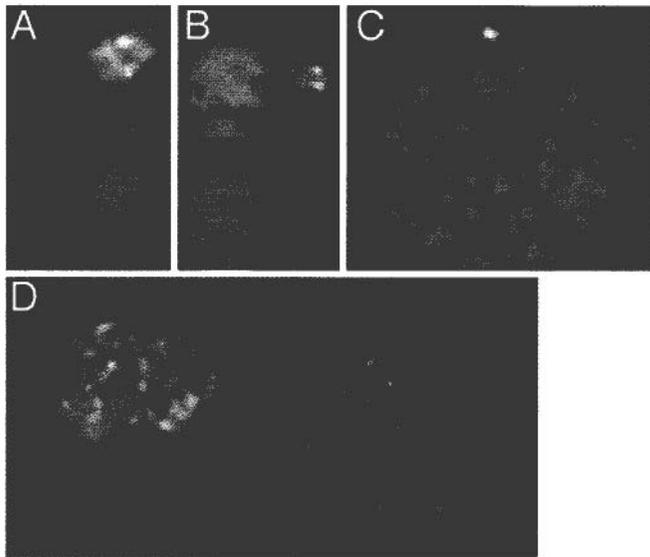
Fig. 4. Centromeres replicate with euchromatin in tetraploid *Kc* cells and in larval diploid neuroblasts. (*A*) Mitotic X chromosomes from cells pulsed with dig-dUTP nucleotide analog (green) and then chased for 4 h show heavy labeling in the heterochromatin surrounding centromeres (Cid, red), as expected for incorporation during late S phase. (*B*) Mitotic X chromosomes from cells pulsed with dig-dUTP and then chased for 10 h were in early S phase at the time of the pulse, because they show heavy labeling in the euchromatic arms. There are also foci of incorporation corresponding to both sister centromeres. (*C*) Pulse-labeling and imaging of interphase *Kc* cells shows that centromeres replicate in the early S-phase period when euchromatin is also replicating. We tracked cell survival and S-phase progression over a 5-h period, and mitotic index over a 25-h period in all labeling experiments. These parameters were indistinguishable from control, untreated cultures. In labeled cultures after 7 h we observed ≈98% labeling of mitotic figures, indicating that virtually all cells in S phase at the time of the pulse received the nucleotide analog. (*D*) Neuroblast centromeres are contained within one to three heterochromatic chromocenters (H3K9me, blue). Pulse-labeling with dig-dUTP reveals foci of DNA replication in two centromeric spots and in euchromatin. Cultured cells and dissected larval brains were labeled and prepared as described (8).

Fig. 5. Replication within centromeric domains. (*A*) Cid and H3 appear to be interspersed as alternating blocks in the centromeric domain. Replication within the domain occurs before replication of the flanking heterochromatin. We consider three possible arrangements of replication origins within this domain. 1: Multiple origins coincide precisely with each block of Cid-containing chromatin. Firing of these origins would replicate every Cid block first, after which replication forks proceed into H3-containing chromatin. 2: Multiple origins are distributed throughout the domain, without regard to the kind of chromatin. At any one time, some Cid-containing and some H3-containing chromatin would be replicating. 3: A single origin lies in the centromeric domain, and bidirectional replication duplicates the entire region. Labeling from short nucleotide analog pulses can distinguish these arrangements. (*B*) A pulse-labeled centromeric domain fiber. *Kc* cells were labeled with dig-dUTP (8) and fibers prepared according to (30), except that a high salt buffer described in (60) was used. The stretched fiber shows an array of Cid spots (red). In this case, the nucleotide analog (green) has incorporated in the intervening gaps between Cid chromatin. These replication tracts are scattered throughout the centromeric domain, and this pattern demonstrates that multiple origins are scattered throughout the domain. The statistically significant association of replication tracts with non-Cid chromatin across the domain suggests that origins have a fixed relationship to the chromatin blocks, and that H3-containing blocks replicate out of phase from Cid-containing blocks.

centromeres. Indeed, this has been observed for minichromosomes deficient in flanking heterochromatin (20).

Our previous experiments using interphase *Kc* cells revealed that ≈90% of centromere replication occurs when euchromatin is replicating. The remaining 10% may have been late replication in centromeric regions, but is more likely the result of nearby heterochromatic replication foci that could not be resolved from sites with Cid. Such early replication of centromeres is not limited to tetraploid *Kc* cells (Fig. 4*C*)—we have observed similar replication patterns in diploid larval neuroblasts (Fig. 4*D*)—although the much shorter cell cycle time and the more irregular chromocenter limits quantitative analysis. Therefore, this early timing of centromere replication appears to be general for *Drosophila* cells.

This feature of centromeres extends to other eukaryotes. It has long been known that budding yeast centromeres replicate early in the cell cycle (22). In mammals and plants, centromeres appear to replicate conspicuously earlier than similarly repetitive heterochromatic DNA (23–25). Thus, whereas the absolute timing of centromere replication in the cell cycle appears variable (26), the relative timing of replication in euchromatin, centromeres, and heterochromatin is consistent.

High-Resolution Mapping of Centromere Replicons

A series of progressively more direct experiments have provided insight into the fine structure in the centromere region. A model
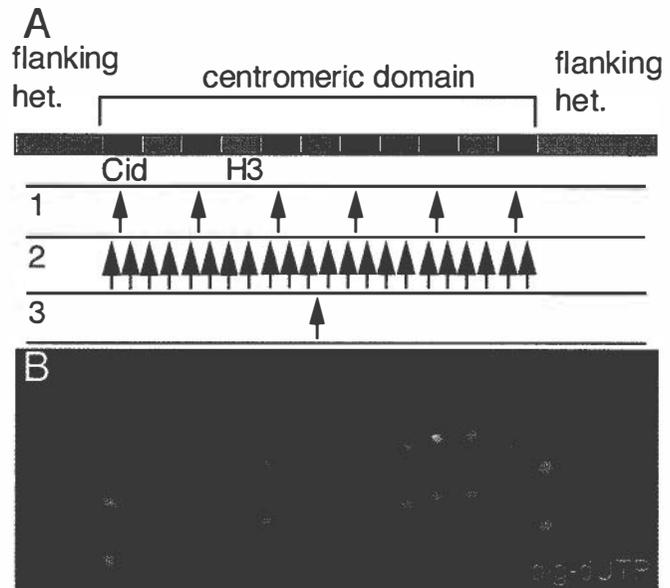
for the centromeric constriction has suggested that loops of DNA coil through the constriction, with centromeric nucleosomes lying in the outward parts of these coils, and conventional nucleosomes in the interior portions (21, 27, 28). This would account for the polar structure of the entire centromere if centromeric nucleosomes nucleate kinetochore formation (and thus microtubule capture) and conventional nucleosomes recruit cohesins (and thus centromeric cohesion). The linear arrangement of nucleosomes along centromeric DNA would then be alternating blocks of centromeric and conventional nucleosomes within the centromeric domain (Fig. 5*A*). A recent study using stretched chromatin fibers has demonstrated that Cid and H3 are interspersed in *Drosophila*, although these are not included in the same nucleosome (21). Apparently, blocks packaged in one kind of nucleosome alternate with blocks packaged in the other. How could the duplication of such regular but discontinuous arrays of nucleosomes occur?

The alternating pattern of nucleosomes on stretched chromatin fibers is reminiscent of replication patterns on fibers from normal chromatin (29). Replication origins within a chromatin domain often appear to be regularly spaced with an interval of 50–100 kb, and these origins fire synchronously. Perhaps the

nucleosome blocks in the centromeric regions correspond to an underlying regular arrangement of replication origins throughout the entire centromeric domain. If Cid-containing blocks include the origins for these domains, and if replication initiates at a time when H3 is not available, ultimately only the RI deposition of Cid will package these blocks. The later replicating stretches would incorporate H3 as it becomes available. In this way, the fine pattern of replication would maintain the discontinuous Cid arrays over an extended region.

Our model for maintaining the higher-order chromatin structure of the entire centromere has precise requirements for replication patterns in this region: a discontinuously spaced arrangement of origins must correspond to the blocks of Cid-containing chromatin (Fig. 5A, pattern 1). At least two other patterns of replication in this region can be imagined. Firstly, all Cid- and H3-containing blocks might replicate simultaneously (pattern 2). Secondly, a single origin might replicate the entire domain (pattern 3).

In a study of stretched centromeric fibers (21), Blower *et al.* showed that centromeric domains replicate out of phase with surrounding heterochromatin. This result is consistent with our previous demonstration that centromere replication precedes that of surrounding heterochromatin (8). Within centromeric domains, it was reported that H3- and CenH3-containing chromatin replicate concurrently (21). However, any fine structure to the replication patterns in these domains might have been obscured by the 2–2.5-h labeling period this study used, because this is enough time for even just two bidirectional replication forks from a single replicon to transit the entire domain. Despite this qualification, a data trace in this work (figure 5C in ref. 21) was intriguing. This trace appears to show that edges of CenpA signal significantly coincided with edges of nucleotide incorporation, with gaps between the CenpA blocks. If confirmed, this might be an example of discontinuous replication tracks corresponding to blocks of centromeric chromatin.

We investigated this possibility by pulse-labeling cells for only 15 min. To prepare stretched chromatin fibers, we disrupted nuclei spread on a glass slide in a high-salt buffer. As the buffer runs off the slide, it pulls chromatin fibers behind it. We identified stretched centromeres and examined those fibers in which nucleotide incorporation was unambiguous. In each of these cases it was clear that replication was occurring in discrete patches scattered throughout the centromeric domain (Fig. 5B). These replication tracks must arise from multiple origins, and thus we can rule out the two possibilities that the entire domain replicates from a single origin, or that the whole domain replicates simultaneously.

These patches corresponded significantly with the segments *between* Cid-containing chromatin. Thus, from published experiments (21) and the experiments described here it appears that replication occurs in two discrete phases: all CenH3-containing chromatin within a domain replicates, and at a different time all H3-containing chromatin replicates. Therefore, replication within this domain is discontinuous and initiates from multiple origins.

Identifying the precise location of origins within the centromeric domain is problematic, because of three difficulties inherent to interpreting stretched chromatin fibers. Firstly, although these fibers are stretched to about 50–100 times their interphase size (21), the radius of H3 or CenH3 spots is substantially less than the resolution of light microscopes. Secondly, the intensity of spots along a fiber is variable, implying that the fiber is unevenly stretched. Thirdly, fibers are inherently spotted—even DNA *in situ* hybridization with probes that should uniformly label an extended region always appear spotty (30). These effects mean that a lack of overlap between different signals could result from artifacts in fiber preparation, and conversely that overlap can result from unstretched segments. It

is apparent from fiber preparations that artifacts are occurring, because some sites on fibers do not appear to be packaged with any histones (21). These concerns can only be addressed with improvements in fiber technology.

Identifying the position of origins within centromeric domains is a critical issue to address, because the deposition of H3 must occur as its DNA substrate replicates. The maintenance of this interspersed arrangement of chromatin must involve the differential regulation of the replication-coupled deposition of H3 and the RI deposition of Cid.

Inferring the RI Machinery

Given that deposition of any H3 must occur in the form of $(H3\cdot H4)_2$ tetramers, there must be discrimination of H3-containing tetramers from tetramers containing variants. Our analysis of RI assembly initiated the mapping of discriminating sites within the histone variants (9). We found that one type of discrimination is a cluster of three residues within the histone fold domain (HFD) of H3 that limits it to replication-coupled deposition. Furthermore, because both Cid and H3.3 undergo RI deposition but have mutually exclusive targets, there must be additional discrimination between these variants.

Replication-coupled nucleosome assembly is aided by accessory factors that are recruited to the replication fork by binding to PCNA. However, the process of RI deposition must be different, because RI deposition of H3.3 does not require portions of the histone that are required for replication-coupled deposition (9). Furthermore, the lack of PCNA during gap phase deposition raises the question of what is recruiting histones to the sites. The phenomenon of CenH3 targeting has raised expectations that a specific, localized chromatin assembly factor or histone modification will be involved in the targeting of CenH3s (26, 31, 32). Indeed, a chromatin remodeler of the RSC family, P/BAF, localizes to kinetochores during mitosis of mammalian cells (33). Furthermore, RSC mutations in budding yeast have been previously described to have altered chromatin structure specifically around centromeres (34), and perhaps RSC activity is involved in assembly of centromeric nucleosomes. Mutations in CAF and Hir genes also give centromere defects, and it has been suggested that these factors are involved in loading the yeast CenH3 Cse4p (35). However, a role for any of these factors does little to explain the specific targeting of CenH3s, because these factors are all widely distributed in the nucleus (33, 36). The best candidate for a uniquely centromere-localized chromatin assembly factor is the Mis6 protein in fission yeast (12). This protein is required for centromeric localization of the CenH3 SpCENP-A, but Mis6 homologs in budding yeast (Ctf3; ref. 37) and in mammals (CENP-I; ref. 38) localize to centromeres but are not required for targeting CenH3s. Thus, Mis6 proteins appear to be structural components of centromeres, not histone assembly factors.

An alternative model is that some feature of centromeric chromatin facilitates the targeting of its specialized histones. An obvious candidate for this feature is that centromeric nucleosomes themselves bind to and thereby recruit new CenH3 tetramers for future deposition. Such an interaction is a possible molecular mechanism for direct templating of centromere duplication (39). Regardless of whether CenH3 targeting involves specialized co-factors, templating, or both, the question remains as to why it should use an RI pathway.

The targeted deposition of H3.3 to active genes is likewise replication-independent, although transcription-coupled assembly may facilitate $(H3.3\cdot H4)_2$ deposition. Perhaps H3.3 targeting is mediated by a component of RNA polymerase complexes. Because RNA polymerases move processively along the DNA during transcription, a contiguous transcribed segment of DNA might incorporate the H3.3 variant. Alternatively, RI deposition of H3.3 may be facilitated by any of a number of ATP-dependent

chromatin remodeling complexes to target specific sites near transcription units. Any candidate factor might be expected to preferentially use H3.3 instead of H3, but whether there is any such discriminating factor is unknown, because all *in vitro* studies of higher eukaryotic chromatin assembly have been performed with H3. We anticipate that this will soon be addressed. However, the prospects for identifying a unique remodeler that is required for RI deposition are uncertain, because budding yeast mutants that eliminate any known chromatin assembly factors do not eliminate chromatin assembly (1). Thus, we need to consider the possibility that RI deposition at active genes and at centromeres uses generic remodeling activities, and that components or structural aspects common to *both* centromeres and actively transcribed genes may result in RI histone deposition at both kinds of sites.

Opening a Space for Histone Replacement

The deposition of histones throughout the cell cycle by a replication-independent process implies that previously existing nucleosomes are unraveled, and their histones released. It is known that the process of transcription results in a local unfolding of the chromatin fiber and an "open" chromatin configuration (Fig. 2; ref. 15). Although transcription of nucleosomal templates with bacterial polymerases can occur *in vitro* without displacing histone octamers from DNA (40), *in vivo* assays demonstrated that a measurable amount of transcription-dependent histone displacement does occur in eukaryotic nuclei (41). In fact, recent experiments revealed that, even *in vitro*, RNA polymerase II is virtually unable to transcribe nucleosomal DNA under physiological conditions (42). Transcription requires that histone–DNA contacts be broken for polymerase to transit the nucleosomal DNA. Although transcription can occur without histone displacement if the histone octamer releases some contacts with DNA and maintains others (40), at some frequency all contacts might be released. The histone octamer would then simply fall off. Additionally, localized remodeling factors will disrupt nucleosome structure as they act. The *in vitro* and *in vivo* observations can be reconciled if histone displacement occurs occasionally as nucleosomes are disrupted.

Constraints on nucleosomes in a compacted chromatin fiber (i.e., "closed" chromatin) would limit histone displacement. Although internucleosome forces within inactive chromatin are uncharacterized, they have been inferred from numerous experiments, including the tendency of nucleosomes within heterochromatin to form extremely regular and fixed arrays (43). A likely constraint in heterochromatin arises from the multimeric associations that occur between heterochromatin-specific non-histone chromatin proteins. Attention has focused on the heterochromatin protein-1 (HP1). HP1 is recruited to heterochromatic DNA by binding, through its chromodomain, to the H3 tail when it is methylated at lysine-9 (H3-K9me; ref. 44). The chromo shadow domain of HP1 mediates associations between HP1 molecules, and multimers of HP1 bound to methylated histone tails provides one basis for constraining arrays of nucleosomes.

Although the state of chromatin in heterochromatin and in actively transcribed regions is well known, less is known about the chromatin fiber packaged by centromeric nucleosomes. However, these regions appear to be open. Centromeric DNA is sensitive to micrococcal nuclease digestion both in budding yeast (45) and in the central core region of fission yeast centromeres where SpCCENP-A-containing nucleosomes reside (12), and plant meiotic centromeres appear decondensed (46). In addition, early replication is a feature of open chromatin, and centromeric chromatin replicates before surrounding heterochromatin. An open configuration may arise from at least three sources. First, all CenH3s lack a canonical H3 tail (3). Because methyl-modification of lysine-9 appears to be the key epitope to maintain heterochromatin, the lack of this site in centromeric

nucleosomes means that such regions cannot become heterochromatic. Indeed, the heterochromatin protein HP1 is not associated with chromatin packaged by CenH3s (47, 48). Second, our recent study of Cid homologs in drosophilids has uncovered DNA minor-groove binding motifs in the Cid tail outside of the nucleosome core (49). Extension of the Cid tail along linker DNA between nucleosomes may inhibit compaction of the nucleosome strand, thus maintaining these regions in an open configuration (Fig. 2). Third, chromatin remodeling factors that destabilize nucleosomes are found both at active genes and centromeres, and their activity will promote histone replacement. We suggest that an open chromatin configuration is the common basis for RI deposition at centromeres and at actively transcribed genes.

The RI Target Sites for H3 Variants Are in Distinct Nuclear Compartments

If open chromatin were the sole basis for RI deposition, then we would expect that active genes and centromeres would incorporate both H3.3 and CenH3s. However, their deposition is mutually exclusive. This exclusivity is likely to rely on multiple mechanisms that act on all steps in nucleosome assembly. Factors that discriminate between H3.3 and Cid would be the best candidates for directing these variants to their targets. However, the organization of the nucleus provides a clue as to another way in which exclusive targeting may be accomplished. Centromeric DNA in *Drosophila* is flanked by repeated sequences that are packaged into heterochromatin, and this forms a compartment at interphase in which centromeres are embedded in heterochromatin (Fig. 3B). The active rDNA genes are the primary sites of H3.3 deposition and they are also found in a distinct nuclear compartment, the nucleolus, next to the chromocenter. This functional nuclear organization is very simple to see in *Drosophila*, where all heterochromatin typically associates into one large chromocenter, and the active rDNA arrays also often associate to present one large nucleolus. In fact, this general compartmentalization is almost invariant in eukaryotes, and has led to the idea that heterochromatin somehow protects centromeres and NORs (50). Although both Cid and H3.3 undergo RI deposition, their exclusive targeting could in part be accomplished by restricting one or both variants within the nucleus. For example, unincorporated $(Cid \cdot H4)_2$ tetramers might be sequestered within the heterochromatic chromocenter. Cid deposition would then appear targeted to the centromere, because this is the only site within the chromocenter with open chromatin.

Whether $(Cid \cdot H4)_2$ tetramers are actually sequestered in this way is unknown. Indeed, whether sequestering substrates can have *any* effect on reactions within the nucleus has become a pressing issue (51). Many nuclear components remain mobile, but functional experiments argue that certain effects in the nucleus actually only occur when components are sequestered (52). It is likely that some reactions in the nucleus are relatively independent of localization because they associate efficiently with their partners and their reactions proceed quickly. Conversely, reactions that involve weak interactions or multiple steps may require raising the effective concentration of their substrates by nuclear sequestration.

We have previously suggested that the heterochromatic compartment is involved in histone traffic within the nucleus (8). The basis of this hypothesis was our realization that Cid-containing chromatin behaves unusually during S phase. Generally, the deposition of H3 quickly follows DNA replication. However, the replication of Cid-containing centromeric DNA occurs without H3 deposition (8), implying that the normal coupling between replication components and nucleosome assembly components must be broken. Because this coupling is thought to result from an interaction between chromatin assembly factor 1 (CAF1)-
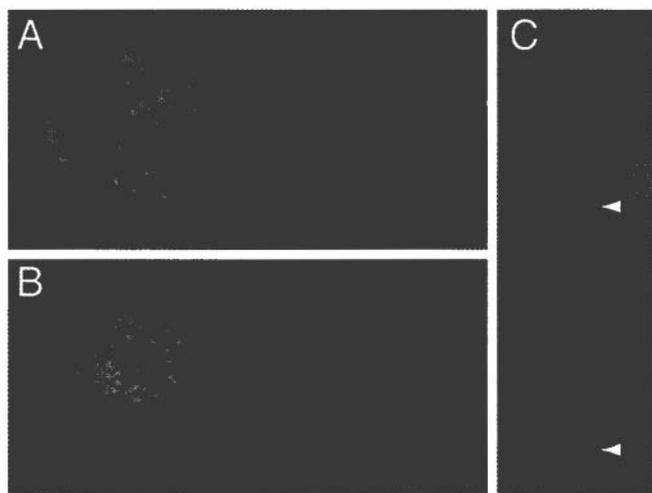
Fig. 6. Overexpression of Cid mis-localizes to euchromatin by RI deposition. *Kc* cells were transfected with a *HS-CidGFP* construct (3) with a modified translational start sequence (61). Cells were induced to produce high levels of Cid-GFP (red), and then immediately pulse-labeled with nucleotide analog (green) to identify cell cycle stages. The heterochromatic compartment is labeled with an anti-HP1 antibody (blue; ref. 61). (*A* and *B*) Overexpressed Cid incorporates at centromeres but also mis-incorporates throughout euchromatin. Mis-incorporation in euchromatin occurs by a replication-independent process, because it occurs both in early S phase (*A*) and in late S phase (*B*) cells. (*C*) After a chase of 6 h, mitotic chromosomes that were induced during S phase show labeling at both sister centromeric foci (arrowhead), and throughout the euchromatic arms.

histone complexes and PCNA, the simplest explanation for uncoupling the two processes would be to sequester replicative nucleosome assembly factors away from centromeres. We imagined that unincorporated H3-containing tetramers might be sequestered in euchromatin in the first half of S phase, and would thus never (productively) see the replication forks at centromeres within the heterochromatic compartment. This uncoupling might be necessary to prevent dilution of centromeric nucleosomes by conventional nucleosomes that would assemble after replication-coupled deposition. Genetic experiments in budding yeast and *Drosophila* suggest that CenH3s and H3 do compete for assembly (21, 53).

One way that a competition between CenH3 and H3 histones can be probed is to change their relative concentration. We have previously reported that a tagged Cid protein exclusively deposits at centromeres when it was ectopically expressed at low levels from a heat-shock-inducible promoter (3). However, it was apparent that expression from this construct remained low. Re-engineering the transcriptional start region of the construct to include a translational initiation consensus site now allows overproduction of Cid in cells.

To analyze the behavior of excess quantities of Cid protein, we introduced an overexpression construct into *Drosophila Kc* cells (Fig. 6). Cells receive varying amounts of transfected DNA, and thus express Cid over a wide range of levels. In cells that express low amounts of the ectopic protein, Cid localizes to centromeres, as expected. However, a new localization pattern for Cid is seen at high expression levels: the tagged protein localizes to centromeres and throughout euchromatin. The incorporation pattern of ectopic Cid is especially clear on mitotic chromosomes from these transfections, where the tagged protein is incorporated throughout the euchromatic arms as well as at centromeres (Fig. 6*C*). We conclude from this result that excess Cid can be deposited at sites other than centromeres. Normal cells must have mechanisms to prevent euchromatic deposition, but overexpression is sufficient, by itself, to overcome this restriction.

The mis-incorporation pattern of Cid shows an interesting specificity: Cid can deposit at centromeres and euchromatin but not in heterochromatin (Fig. 6). Therefore, heterochromatin must lack the feature that tolerates mis-incorporation, or must actively exclude Cid. As we have argued above, centromeres and euchromatin share the feature of open chromatin, which we have proposed is the first prerequisite for RI deposition of histone variants. Indeed, the mis-incorporation of Cid into euchromatin is replication-independent, because it occurs both when euchromatin is replicating in early S phase (Fig. 6*A*), and in late S phase when euchromatic replication is complete (Fig. 6*B*). We suggest that Cid is contaminating open chromatin in the euchromatic compartment when it is overexpressed.

What normally prevents the deposition of Cid into euchromatin? Endogenous Cid is present only at low levels, and mis-incorporation could be avoided if Cid were sequestered away from euchromatin in the nucleus. If unincorporated Cid were sequestered in the heterochromatic chromocenter, it would be unable to deposit in the closed chromatin of this compartment. Thus, sequestration might serve two purposes: deposition in euchromatin would be prevented and deposition at centromeres would be promoted. Overexpression of CenpA in mammalian cells also mis-incorporates into euchromatin (54). Although it has not been examined whether CenpA mis-incorporation is replication-independent, we expect this to be the case, because this is how CenpA deposits at centromeres (24).

The idea that histone variants may respect nuclear compartments was first raised by our experiments expressing heterologous CenH3s in *Drosophila Kc* and human HeLa cells (3). These extremely diverged heterologous histones did not localize to centromeres in these cells, implying that there is some kind of specificity for depositing the correct CenH3 at centromeres. Surprisingly, heterologous histones were preferentially enriched in the heterochromatic blocks. We suggested that it is a default ability of cells to enrich diverged H3 variants in the heterochromatic compartment. Perhaps heterochromatic enrichment is a normal first step in the deposition of the endogenous CenH3s. Those experiments and our overexpression results encourage the view that nuclear compartments may guide histone variants to the correct subset of their potential deposition sites. Compartment effects may also affect the RI deposition of H3.3 in an inverse way to Cid: i.e., sequestering to promote H3.3 deposition at active genes, and preventing its deposition at centromeres. Because H3.3 is largely identical to H3, the hypothetical element that is recognized in H3 and results in its exclusion from chromocenters during centromere replication may also be present in H3.3. Perhaps this discrimination against canonical H3 histones also serves to prevent the RI deposition of H3.3 at centromeres.

Why RI Assembly?

RI assembly permits immediate chromatin repair. The unfolding of chromatin during transcription may be damaging, in that the forces RNA polymerases apply to their template DNA should at least occasionally displace histone octamers from DNA (55). Additionally, histone octamers may sometimes be displaced by chromatin remodeling factors associated with transcriptional activity. In either case, these regions must be repackaged into nucleosomes. Similarly, replacement of CenH3s may be required to maintain the nucleosomal configuration of centromeres after mitosis. Bundles of microtubules drag a chromosome to the pole during anaphase, and the forces they apply (56) may be sufficient to occasionally pull off histone octamers. Chromatin would then be stripped of some CenH3 histone octamers. RI deposition allows repair of this damage. In fact, the RI deposition of CenpA in mammalian cells seems to occur around the time of mitosis (24). The deposition of Cid in *Drosophila* cells occurs throughout the cell cycle, but may only be required at two points: as
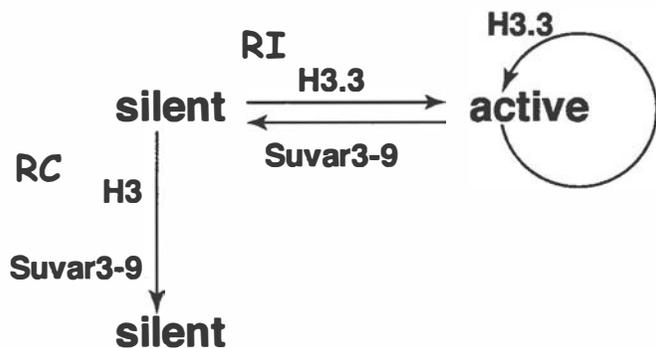
Fig. 7. RI deposition allows switching of heritable chromatin states. Nucleosomes in silent heterochromatin are distinctively modified by methylation, and thereby recruit the HMT Suvar3-9. The silencing epitope can be perpetuated by Suvar3-9 through the cell cycle by the methylation of H3 after replication-coupled deposition (vertical arrow). A gene can be activated (rightward arrow) at any time in the cell cycle, and the unraveling of methylated nucleosomes and RI deposition of H3.3 will remove the silencing epitope. This abolishes Suvar3-9 recruitment and allows stable activation. RI deposition of H3.3 will continue as long as the gene is transcribed. Switching from an active to a silent state (leftward arrow) can occur by repressing transcription and methylating the N-terminal tail of H3.3 at Lys-9, once again recruiting the Suvar3-9 complex.

Fig. 8. The relationship of the canonical H3 histones in animals and fungi. Three residues in the HFD differ between H3 and the replacement H3.3 histone in metazoans. Basidiomycete fungi (*Cryptococcus*) also have two canonical H3 histones that can be classed as an H3 and a replacement histone. Ascomycetes (*Saccharomyces*, *Neurospora*, and *Schizosaccharomyces*) have only one type of canonical H3, resembling the replacement histone in basidiomycetes. This phylogeny identifies that ascomycetes have lost their H3 histone, and retain only a replacement variant.

centromeric DNA replicates to double its chromatin, and after mitosis to repair stripped chromatin.

The process of RI assembly at active genes provides a novel level of control over histone modifications. Replacement of nucleosomes in one modification state by new histones could switch chromatin to an active state. Initiation of transcription would start this process, and successive transits of RNA polymerases would promote RI assembly. The replacement H3 histone in alfalfa is hyperacetylated (57), and RI assembly with acetylated histones could enrich such modifications in active chromatin. However, histone modification by methylation has appeared more problematic (43). A number of histone methyltransferases (HMTs) have been characterized (44), but no histone demethylase is known. Methylated lysine-9 in the H3 tail (H3K9me) is a critical epitope for recruiting heterochromatic chromatin proteins, because this is the binding site for HP1. HP1 recruits additional heterochromatic proteins including the Suvar3-9 HMT. Therefore, it is straightforward to imagine how these recruited proteins could perpetuate a heterochromatic state through replication-coupled nucleosome assembly and cell division (Fig. 7).

Because an irreversible methyl modification appears to specify the heterochromatic state, it has been unknown how a heterochromatic site could switch to an active state. One route for switching might be to prevent the methylation of nucleosomes assembled during replication. Successive cell cycles could then dilute methylated nucleosomes, allowing eventual activation. However, more rapid mechanisms for activating silenced chromatin must exist. Induction of silenced genes can occur within a single cell cycle; for example, X chromosomes become reactivated and lose H3K9me during diplotene in the *Caenorhabditis* ovary (58). In addition, our work using a reporter for heterochromatic gene silencing suggests that switching to an active state can occur in somatic cells without cell division (59). Thus, H3K9me can be removed without replication-coupled nucleosome assembly.

RI deposition implies that the entire heterochromatic nucleosome may be unraveled and replaced (9). The process of transcriptional activation may force the disassembly of H3K9me-containing nucleosomes, followed by RI assembly of an unmarked nu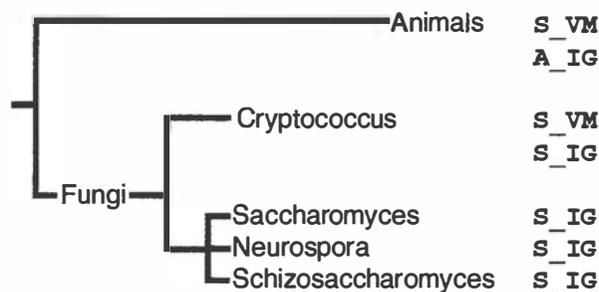cleosome. Although we do not know the fate of the displaced methylated H3, we do know that RI deposition can occur at any time in the cell cycle, and thus should be able to rapidly derepress silencing (Fig. 7). Conversely, an active gene could be silenced by methylating the tail of H3.3, which presents the same lysine-9 epitope. The stability of histone methylation gives it a distinct advantage over other histone modifications for heritable effects on chromatin. The possibility of RI deposition circumvents the irreversible nature of methylation, thus retaining the potential to switch the heritable chromatin state at a later time.

Conclusions

H3 variants are used to package functionally specialized chromatin, where they play vital functional roles. Localizing these variants to centromeres and to transcriptionally active regions utilizes an RI process that is distinct from the nonspecific, replication-coupled method of packaging the bulk genome. We have argued that RI deposition is the consequence of the activities that impinge on these sites in the genome and create an open chromatin structure. This flexibility in histone deposition may be necessary to maintain the nucleosomal structure of these regions. In higher eukaryotes, the RI deposition process allows specialized chromatin to be distinguished at the most basic level, where histone variants are incorporated into chromatin. The differences between the generic H3—which packages the bulk of the genome—and the H3 variants may contribute to the physical properties of specialized regions and recruit particular non-histone chromatin proteins. Because histones remain associated with DNA through mitosis, these variants establish heritable distinctions in chromatin.

Centromeres are a defining feature of eukaryotes, and all are likely to have a CenH3. However, the utilization of two conserved versions like H3 and H3.3 is not universal. For example, budding yeast has only one canonical H3 histone, which undergoes both replication-coupled and RI deposition (1). Surprisingly, this is H3.3: phylogenetic analysis reveals that ascomycetes have lost H3, whereas their sister clade basidiomycetes have both H3 and H3.3, as do animals (Fig. 8; ref. 9). Therefore, an H3.3 gene performs all general functions in some organisms. The extraordinary conservation of H3.3, which is identical from mollusks to mammals, speaks to its fundamental role in the eukaryotic nucleus.

1. Verreault, A. (2000) *Genes Dev.* **14,** 1430–1438.
2. Osley, M. A. (1991) *Annu. Rev. Biochem.* **60,** 827–861.
3. Henikoff, S., Ahmad, K., Platero, J. S. & van Steensel, B. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 716–721.
4. Akhmanova, A. S., Bindels, P. C. T., Xu, J., Miedema, K., Kremer, H. & Hennig, W. (1995) *Genome* **38,** 586–600.
5. Fretzin, S., Allan, B. D., van Daal, A. & Elgin, S. C. (1991) *Gene* **15,** 341–342.
6. Henikoff, S., Ahmad, K. & Malik, H. S. (2001) *Science* **293,** 1098–1102.
7. Waterborg, J. H. & Robertson, A. J. (1996) *J. Mol. Evol.* **43,** 194–206.
8. Ahmad, K. & Henikoff, S. (2001) *J. Cell Biol.* **153,** 101–110.
9. Ahmad, K. & Henikoff, S. (2002) *Mol. Cell.* **9,** 1191–1200.
10. Warburton, P. E., Cooke, C. A., Bourassa, S., Vafa, O., Sullivan, B. A., Stetten, G., Gimelli, G., Warburton, D., Tyler-Smith, C., Sullivan, K. F., *et al.* (1997) *Curr. Biol.* **7,** 901–904.
11. Stoler, S., Keith, K. C., Curnick, K. E. & Fitzgerald-Hayes, M. (1995) *Genes Dev.* **9,** 573–586.
12. Takahashi, K., Chen, E. S. & Yanagida, M. (2000) *Science* **288,** 2215–2219.
13. Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L. & Henikoff, S. (2002) *Plant Cell.* **14,** 1053–1066.
14. Allis, C. D. & Wiggins, J. C. (1984) *Dev. Biol.* **101,** 282–294.
15. Wolffe, A. P. (1998) *Chromatin: Structure and Function* (Academic, San Diego), 3rd Ed.
16. Shibahara, K. & Stillman, B. (1999) *Cell* **96,** 575–585.
17. Akhmanova, A., Verkerk, T., Langeveld, A., Grosveld, F. & Galjart, N. (2000) *J. Cell Sci.* **113,** 4463–4474.
18. Akhmanova, A. S., Miedema, K., Wang, Y., van Bruggen, M., Berden, J. H. M., Moudrianakis, E. N. & Hennig, W. (1997) *Chromosoma* **106,** 335–347.
19. Pina, B. & Suau, P. (1987) *Dev. Biol.* **123,** 51–58.
20. Sullivan, B. A. & Karpen, G. H. (2001) *J. Cell Biol.* **154,** 683–690.
21. Blower, M. D., Sullivan, B. A. & Karpen, G. H. (2002) *Dev. Cell* **2,** 319–330.
22. McCarroll, R. M. & Fangman, W. L. (1988) *Cell* **54,** 505–513.
23. O'Keefe, R. T., Henderson, S. C. & Spector, D. L. (1992) *J. Cell Biol.* **116,** 1095–1110.
24. Shelby, R. D., Monier, K. & Sullivan, K. F. (2000) *J. Cell Biol.* **151,** 1113–1118.
25. Jasencakova, Z., Meister, A. & Schubert, I. (2001) *Chromosoma* **110,** 83–92.
26. Choo, K. H. A. (2001) *Dev. Cell* **1,** 165–177.
27. Zinkowski, R. P., Meyne, J. & Brinkley, B. R. (1991) *J. Cell Biol.* **113,** 1091–1110.
28. Buchwitz, B. J., Ahmad, K., Moore, L. L., Roth, M. B. & Henikoff, S. (1999) *Nature (London)* **401,** 547–548.
29. Berezney, R., Dubey, D. D. & Huberman, J. A. (2000) *Chromosoma* **108,** 471–484.
30. Haaf, T. & Ward, D. C. (1994) *Hum. Mol. Genet.* **3,** 697–709.
31. Sullivan, K. F. (2001) *Curr. Opin. Genet. Dev.* **11,** 182–188.
32. Sullivan, B. A., Blower, M. D. & Karpen, G. H. (2001) *Nat. Rev. Genet.* **2,** 584–596.
33. Xue, Y., Canman, J. C., Lee, C. S., Nie, Z., Yang, D., Moreno, G. T., Young, M. K., Salmon, E. D. & Wang, W. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13015–13020.
34. Tsuchiya, E., Hosotani, T. & Miyakawa, T. (1998) *Nucleic Acids Res.* **26,** 3286–3292.
35. Sharp, J. A., Franco, A. A., Osley, M. A. & Kaufman, P. D. (2002) *Genes Dev.* **16,** 85–100.
36. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. (2002) *Genes Dev.* **16,** 806–819.
37. Measday, V., Hailey, D. W., Pot, I., Givan, S. A., Hyland, K. M., Cagney, G., Fields, S., Davis, T. N. & Hieter, P. (2002) *Genes Dev.* **16,** 101–113.
38. Nishihashi, A., Haraguchi, T., Hiraoka, Y., Ikemura, T., Regnier, V., Dodson, H., Earnshaw, W. C. & Fukagawa, T. (2002) *Dev. Cell* **2,** 463–476.
39. Willard, H. F. (1998) *Curr. Opin. Genet. Dev.* **8,** 219–225.
40. Studitsky, V. M., Clark, D. J. & Felsenfeld, G. (1994) *Cell* **76,** 371–382.
41. Jackson, V. (1990) *Biochemistry* **29,** 719–731.
42. Kireeva, M. L., Walter, W., Tchernajenko, V., Bondarenko, V., Kashlev, M. & Studitsky, V. M. (2002) *Mol. Cell* **9,** 541–552.
43. Grewal, S. I. S. & Elgin, S. C. R. (2002) *Curr. Opin. Genet. Dev.* **12,** 178–187.
44. Jenuwein, T. & Allis, C. D. (2001) *Science* **293,** 1074–1080.
45. Bloom, K. S. & Carbon, J. (1982) *Cell* **29,** 305–307.
46. Fransz, P. F., Armstrong, S., de Jong, J. H., Parnell, L. D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T. & Jones, G. H. (2000) *Cell* **100,** 367–376.
47. Partridge, J. F., Borgstrom, B. & Allshire, R. C. (2000) *Genes Dev.* **14,** 783–791.
48. Blower, B. D. & Karpen, G. H. (2001) *Nat. Cell Biol.* **3,** 730–739.
49. Malik, H. S., Vermaak, D. & Henikoff, S. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 1449–1454.
50. Manuelidis, L. (1990) *Science* **250,** 1533–1540.
51. Carmo-Fonseca, M. (2002) *Cell* **108,** 513–521.
52. Feuerbach, F., Galy, V., Trelles-Sticken, E., Fromont-Racine, M., Jacquier, A., Gilson, E., Olivo-Marin, J.-C., Scherthan, H. & Nehrbass, U. (2002) *Nat. Cell Biol.* **4,** 214–221.
53. Glowczewski, L., Yang, P., Kalashnikova, T., Santisteban, M. S. & Smith, M. M. (2000) *Mol. Cell. Biol.* **20,** 5700–5711.
54. Van Hooser, A. A., Ouspenski, I. I., Gregson, H. C., Starr, D. A., Yen, T. J., Goldberg, M. L., Yokomori, K., Earnshaw, W. C., Sullivan, K. F. & Brinkley, B. R. (2001) *J. Cell Sci.* **114,** 3529–3542.
55. Bennink, M. L., Leuba, S. H., Leno, G. H., Zlatanova, J., de Grooth, B. G. & Greve, J. (2001) *Nat. Struct. Biol.* **8,** 606–610.
56. King, J. M. & Nicklas, R. B. (2000) *J. Cell Sci.* **113,** 3815–3823.
57. Waterborg, J. H. (1993) *J. Biol. Chem.* **268,** 4912–4917.
58. Kelly, W. G., Schaner, C. E., Dernberg, A. F., Lee, M.-H., Kim, S. K., Villeneuve, A. M. & Reinke, V. (2002) *Development (Cambridge, U.K.)* **129,** 479–492.
59. Ahmad, K. & Henikoff, S. (2001) *Cell* **104,** 839–847.
60. Balajee, A. S. & Geard, C. R. (2001) *Nucleic Acids Res.* **29,** 1341–1351.
61. Smothers, J. F. & Henikoff, S. (2001) *Mol. Cell. Biol.* **21,** 2555–2569.

Colloquium

Induction and maintenance of nonsymmetrical DNA methylation in *Neurospora*

Eric U. Selker*[†], Michael Freitag*, Gregory O. Kothe*, Brian S. Margolin*[‡], Michael R. Rountree*[§], C. David Allis[¶], and Hisashi Tamaru*

*Institute of Molecular Biology, University of Oregon, Eugene, OR 97403-1229; and [¶]Department of Biochemistry and Molecular Genetics, University of Virginia Health Sciences Center, Charlottesville, VA 22908-0733

One can imagine a variety of mechanisms that should result in self-perpetuating biological states. It is generally assumed that cytosine methylation is propagated in eukaryotes by enzymes that specifically methylate hemimethylated symmetrical sites (e.g., 5'CpG/GpC5' or 5'CpNpG/GpNpC5'). Although there is wide support for this model, we and others have found examples of methylation that must be propagated by a different mechanism. Most methylated regions of the *Neurospora* genome that have been examined are products of repeat-induced point mutation, a premeiotic genome defense system that litters duplicated sequences with C·G to T·A mutations and typically leaves them methylated at remaining cytosines. In general, such relics of repeat-induced point mutation are capable of triggering methylation *de novo*. Nevertheless, some reflect a mechanism that can propagate heterogeneous methylation at nonsymmetrical sites. We propose that *de novo* and maintenance methylation are manifestations of a single mechanism in *Neurospora*, catalyzed by the DIM-2 DNA methyltransferase. The action of DIM-2 is controlled by the DIM-5 histone H3 Lys-9 methyltransferase, which in turn is influenced by other modifications of histone H3. DNA methylation indirectly recruits histone deacetylases, providing the framework of a self-reinforcing system that could result in propagation of DNA methylation and the associated silenced chromatin state.

histone methylation | chromatin | RIP | epigenetics | DNA methyltransferase

More than 25 years ago, Holliday and Pugh (1) and Riggs (2) pointed out that the symmetrical nature of methylated sites (5'CpG/GpC5') observed in animals could allow for propagation of methylation patterns by a "maintenance methylase" specific for cytosines symmetrically opposed to 5-methylcytosines. In principle, a pattern of methylation, perhaps established early in development, could be faithfully propagated indefinitely. Results from transfection experiments in several animal systems are generally consistent with this model. Methylated transforming sequences tend to remain methylated through repeated cycles of DNA replication, whereas unmethylated transforming sequences typically remain unmethylated (3–6). Strong evidence for maintenance methylation also comes from observations in *Arabidopsis* (7). Findings that eukaryotic DNA methyltransferases (DMTs) such as Dnmt1 preferentially methylate hemimethylated CpG/GpC sites provided additional support for the model (8, 9). Nevertheless, some observations seem to conflict with the maintenance methylase model. The model predicts that in a clonal population of cells engaged only in maintenance methylation, (*i*) any given site should be either methylated or unmethylated in all cells, and (*ii*) methylation should be confined to symmetrical sequences. Violations of these predictions have been demonstrated in fungi, plants, and animals (10–23).

The fungus *Ascobolus immersus* offers a particularly clear example of propagation of methylation at nonsymmetrical sites and maintenance of heterogeneous methylation (24, 25). Sequences exposed to methylation induced premeiotically (MIP) in the sexual stage of the life cycle of *Ascobolus* retain their methylation through numerous vegetative cell divisions. As discussed below, DNA methylation that depends on preexisting methylation also has been found in the more thoroughly studied filamentous fungus, *Neurospora crassa*. Although mechanisms for maintenance methylation that do not rely on faithful copying at symmetrical sites can be imagined (e.g., see refs. 14, 26, 27), none have yet been demonstrated. Nevertheless, recent studies that have identified components of the methylation machine provide clues. In the picture that is emerging, DNA methylation patterns reflect a combination of superimposed mechanisms, at least some of which involve modifications of histones.

Maintenance Methylation in *Neurospora*

Ninety-eight percent of cytosines in the *Neurospora* genome are unmethylated, and most DNA methylation is found in relics of the genome defense system called repeat-induced point mutation (RIP; E.U.S., N. Tountas, S. Cross, B.S.M., J. Murphy, A. P. Bird, and M.F., unpublished work; refs. 25, 28, and 29). RIP detects duplicated sequences in the haploid genomes of specialized dikaryotic cells resulting from fertilization (30), and then riddles both copies of the duplicated sequence with C·G to T·A mutations (31). Frequently, but not invariably, sequences altered by RIP are left methylated, perhaps by the action of a putative DMT that is essential for RIP, RIP defective (RID) (32). This scenario is consistent with the possibility that the mutations occur by deamination of cytosines methylated in the sexual phase of the life cycle (29, 33). In principle, the methylation of relics of RIP that is observed in vegetative cells could either reflect creation of a signal for *de novo* methylation or could reflect propagation of methylation established earlier. To distinguish between these possibilities, eight alleles of the *am* gene that were generated by RIP were tested with two assays for their capacity to induce methylation *de novo* (34). First, alleles were demethylated by using the methylation inhibitor 5-azacytidine (5AC) and then scored for their ability to reestablish methylation. DNA
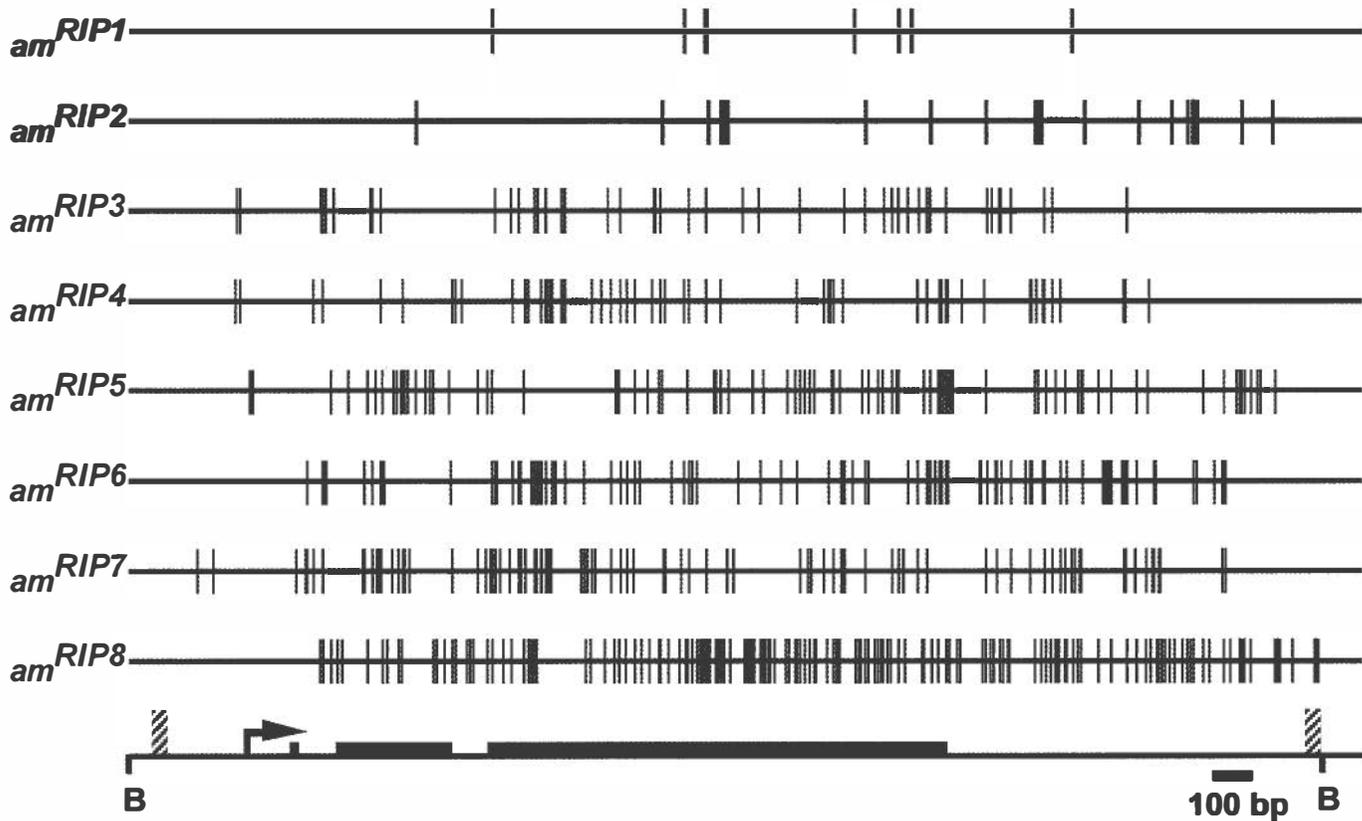
Fig. 1. Mutations from RIP and methylation status of eight different *am* alleles. This figure was adapted from ref. 34 and readers should refer to that article for details of the experiments. Vertical lines indicate mutations. Alleles shown in black were not methylated. Alleles in blue were initially methylated but, after induced loss of methylation (with 5AC or by cloning and gene replacement), did not serve as *de novo* methylation signals. Alleles shown in red were methylated after the cross and served as *de novo* methylation signals. The horizontal line at the bottom represents the 2.7-kb region that was sequenced, including the 2.6-kb *Bam*HI (B) fragment. Exons (thick bars) of the *am* gene, the extent of the duplication (striped bars), and the transcription start (arrow) are indicated.

from each allele also was cloned in bacteria, transformed into *Neurospora* to replace the *am*⁺ allele, and scored for *de novo* methylation. Equivalent results were obtained with both tests and are summarized in Fig. 1. The four alleles with the greatest number of mutations (58 to 158 in the 2.6-kb duplicated region) all triggered *de novo* methylation in vegetative cells, which is consistent with prior evidence that products of RIP can signal DNA methylation (35, 36). We were surprised to find, however, that two alleles (*am*^RIP3^ and *am*^RIP4^) that were initially methylated were incapable of inducing methylation *de novo* at their native chromosomal location (34). Apparently this methylation, which was not limited to symmetrical sites and was rather heterogeneous, depended on preexisting methylation established in the sexual phase by RIP.

As a first step in exploring the mechanism of this maintenance methylation, we tested whether *Neurospora* would propagate methylation established *in vitro*. A 5.2-kb DNA fragment including the *am*^RIP4^ sequence was prepared by PCR using 2′-deoxyribosyl-5-methylcytosine triphosphate (dmCTP) in place of dCTP to ensure that every cytosine was methylated. Then, we introduced this, or an unmethylated control DNA, into an *am*⁺ strain by cotransformation with a fragment including the hygromycin resistance gene, *hph,* and enriched for transformants in which the native *am* gene had been replaced (37). Transformants bearing a single copy of the *am*^RIP4^ sequence in place of *am*⁺ were identified and analyzed by Southern hybridization. In contrast to our experience with a variety of other DNA fragments introduced at *am,* the methylation status of the *in vitro*-methylated sequences varied from transformant to transformant. Two of four transformants generated with methylated

am^RIP4^ sequences showed stable incomplete methylation of these sequences (Fig. 2). Although heterogeneous, the observed methylation persisted through numerous cell cycles; only slight changes in the extent of methylation were observed after three serial passages of both strains (e.g., see Fig. 2, lanes 3 and 8). To verify that the methylation depended on preexisting methylation, i.e., that it was self-propagating, we treated each strain with 5AC for 48 h, verified that no detectable methylation remained (Fig. 2, lanes 4 and 9), and then grew the strains in the absence of the drug to ascertain whether the sequences would trigger methylation *de novo*. Methylation was not reestablished (Fig. 2, lanes 5 and 10), confirming the occurrence of a previously uncharacterized form of maintenance methylation that can propagate heterogeneous regional methylation. The loss of methylation that occurred in some of the transformants obtained with methylated DNA (Fig. 2, strain 3) suggests that establishment of stable methylation in our transformation system involves a stochastic step, perhaps associated with chromatin formation.

The observed fractional methylation at all restriction sites examined limits possible models to account for this propagation of methylation. Most surprisingly, the methylation did not noticeably spread; in every case examined, the methylation seemed to cover an ≈5-kb segment of DNA. We considered the possibility that the mutations from RIP, although insufficient to trigger methylation *de novo*, were nevertheless essential for propagation of methylation and, thereby, controlled the extent of methylation. To test this possibility, we performed a set of transformations with fully methylated *am* DNA lacking mutations by RIP. Because our gene-targeting scheme required selection for loss of *am* function, we used a nonfunctional *am*
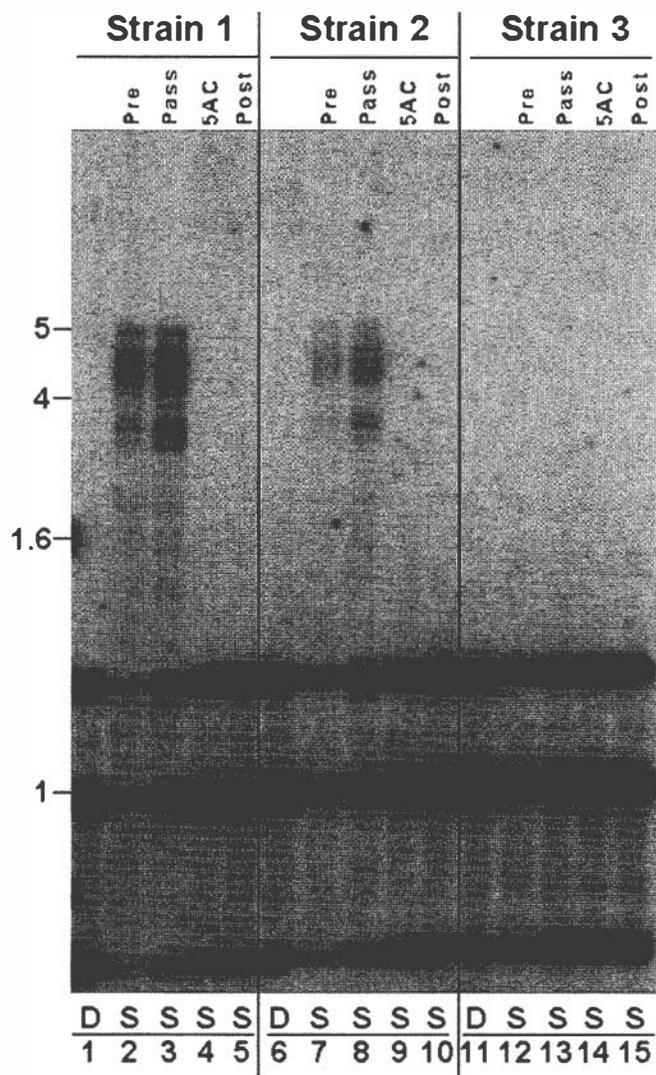
Fig. 2. Maintenance of DNA methylation in *Neurospora*. A fully methylated 5.2-kb DNA fragment containing the *am^RIP4* allele was generated by PCR with dmCTP and targeted to the *am* locus, as described (37). Each transformant was analyzed by Southern hybridization to verify that transforming DNA had integrated properly and without extraneous copies (data not shown) and to assess methylation, as described (40). DNA from the original transformant (Pre), the transformant after three passages (Pass), the transformant after a 48-h treatment with 5-azacytidine (5AC; 24 μM added at 0 and 24 hours), and the transformant grown for 2 days (≈12–15 cell doublings) in the absence of 5AC after treatment with this drug (Post) was digested with the 5-mC-insensitive restriction endonuclease *Dpn*II (D; shown for the original transformant only) or its 5-mC-sensitive isoschizomer Sau3AI (S), fractionated and probed for *am*. Results for three transformants are shown. Size markers (kb) are shown on the left.

Fig. 3. Methylation of *in vitro*-methylated wild-type *am* alleles is not maintained. Transformants with ectopic copies of *am* were generated with fully methylated DNA by cotransformation with *hph* (data not shown). Ten arbitrary transformants bearing one or more copies of *am* are illustrated. Digestions with *Dpn*II (D) and *Sau*IIIAI (S) revealed no differences indicative of methylation, indicating that wild-type *am* sequences are unable to maintain methylation previously established *in vitro* and that a high copy number of gene fragments *per se* does not signal DNA methylation. Size markers (kb) are shown on the left.

allele with an engineered *Cla*I site in place of the *Bgl*II site (plasmid pMS4). Transformants that integrated methylated DNA of this allele at the native *am* locus showed no propagation of the methylation (data not shown). Similarly, all cotransformants that integrated nonselected, fully methylated wild-type *am* DNA failed to retain noticeable methylation (Fig. 3). These findings support the hypothesis that the mutations by RIP play an essential role in the propagation of methylation on *am^RIP4* sequences. We considered the possibility that those alleles that could propagate methylation were those that could not be transcribed but found that even the heavily mutated alleles could support robust transcription when stripped of their methylation,
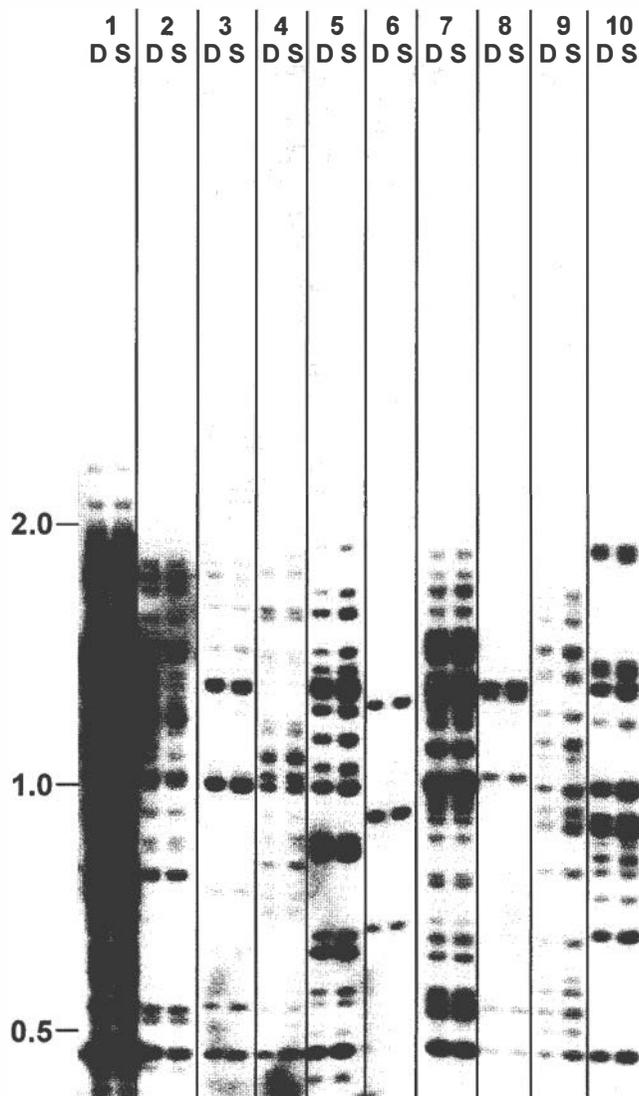
which is consistent with the fact that their promoters were not mutated (38). We conclude that sequences differ in their ability to serve as substrates for maintenance methylation in *Neurospora*.

The findings described above raised the possibility that maintenance and *de novo* methylation are different manifestations of a common mechanism. Several observations support this idea. First, we found that *am^RIP4* DNA is frequently subjected to *de novo* methylation when it becomes integrated in multicopy arrays at ectopic sites (data not shown). This finding is in stark contrast to the case with *am*^+ DNA, which remains free of methylation even when many copies were integrated into the genome (e.g., Fig. 3, transformant 1). Second, we found that *am^RIP4* DNA, but not *am*^+ DNA, can induce methylation of a 200-bp segment of the ζ−η region that includes numerous mutations by RIP but is
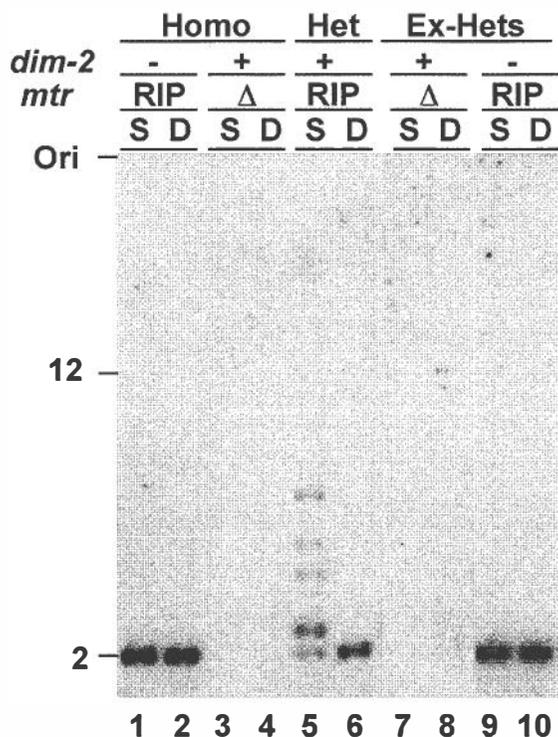
Fig. 4. DIM-2 is the sole DMT required for DNA methylation in vegetative tissues. Heterokaryons (Het) were generated from two homokaryons (Homo): strain N1931, which has a deletion allele of *dim-2* (*dim-2⁻*) and an RIP-mutated allele of the *mtr* gene (*mtr^RIP*) and strain N534, which is *dim-2⁺* and has *mtr* deleted (*mtr^Δ*). DNA methylation was assayed by Southern blotting of genomic DNA digested with *Sau*IIIAI (S) and *Dpn*II (D). The RIP-mutated *mtr* allele is unmethylated in the *dim-2* strain (lanes 1 and 2), but complementation allows for normal DNA methylation in the heterokaryon (lanes 5 and 6). The heterokaryon was broken, and resulting homokaryotic strains (Ex-Hets; lanes 7 to 10) were analyzed for DNA methylation. DNA methylation was not maintained in the absence of DIM-2 (lanes 9 and 10).

unable to trigger *de novo* methylation alone (H.T. and E.U.S., unpublished work). Third, we found that one DMT, DIM-2, is responsible for both *de novo* and maintenance methylation in vegetative cells of *Neurospora* (39). Mutants of *dim-2* are devoid of DNA methylation in all tissues that can be assayed, including methylation in cells that are normally capable of rapid *de novo* methylation (40). To verify that DIM-2 is responsible for maintenance methylation, we first built a heterokaryon between a *dim-2* deletion strain carrying an RIP-mutated allele of the *mtr* gene, which works as a *de novo* methylation signal, and a strain with a wild-type *dim-2* allele carrying a deletion of the *mtr* gene, and confirmed *de novo* methylation (Fig. 4). Then, we isolated homokaryotic derivatives representing the two original constituents and showed that methylation was not maintained in the *dim-2* isolate (Ex-Hets; Fig. 4, lanes 9 and 10).

Clues to Involvement of Methyl/RIP-DNA Binding Proteins and Histone Deacetylation in DNA Methylation

Two important, potentially related questions remain. How do mutations by RIP trigger DNA methylation? How does DNA methylation promote its own propagation? One possibility is that one or more proteins recognizes some feature(s) of sequences modified by RIP and induces DNA methylation. Clues to the existence of such proteins came from band-shift experiments with methylated DNA with mutations from RIP and cellular extracts of *Neurospora*. For example, an activity that we call methyl/RIP binding protein 1 (MRBP-1) binds strongly to
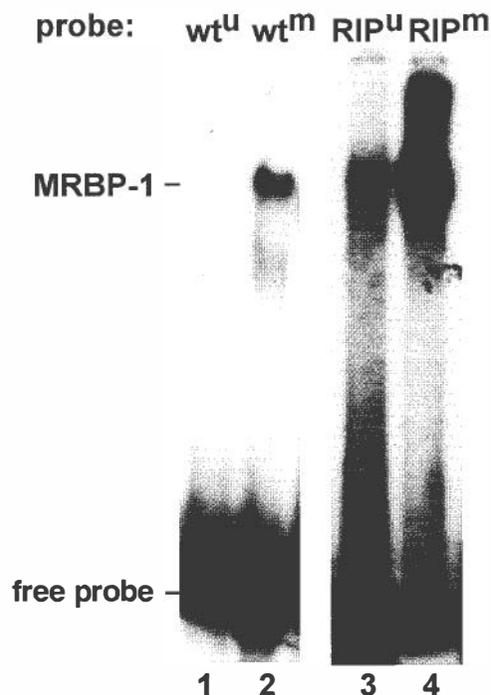

Fig. 5. Gel mobility shift assay using partially purified methyl-RIP binding protein-1 (MRBP-1). MRBP-1 was purified by chromatography on CM Macroprep, SP Fast Flow Sepharose, and heparin Sepharose (data not shown) and incubated 10 min on ice [in 10 mM Hepes, pH 7.9/50 mM KCl/1 mM EDTA/5 mM DTT/10% (vol/vol) glycerol] with 1–2 ng of a 250-bp probe (third exon of *am*) prepared in methylated (m) or unmethylated (u) form by PCR from template with either a wild-type (wt) or RIP-mutated (RIP) allele. Each reaction contained 1 μg of poly dI·dC as competitor and was analyzed by electrophoresis in 4% polyacrylamide gels containing 0.5× TBE buffer (52).

methylated DNA with mutations by RIP (Fig. 5, lane 4) and less strongly to methylated native DNA (lane 2) or unmethylated mutated DNA (Fig. 5, lane 3; G.O.K., M.R.R., A. McCormack, L. David, and E.U.S., unpublished work). In principle, a factor such as MRBP-1 could directly or indirectly trigger DNA methylation. Recent observations favor the latter possibility. First, we found that treatment of *Neurospora* with the histone deacetylase (HDAC) inhibitor trichostatin A (TSA) can reduce methylation in some chromosomal regions (26). One model is that HDACs are recruited to regions with numerous mutations by RIP, perhaps by MRBP-1 or a similar factor, and that the resulting hypoacetylated state leads directly or indirectly to methylation of DNA. Fig. 6 illustrates several possible relationships among DNA methylation, histone acetylation, and transcription. Model 1, which suggests that hypoacetylation is a simple consequence of DNA methylation accounts for indications from animal systems that methyl-DNA binding proteins and DMTs recruit HDACs (see ref. 41) and a preliminary observation that mutation of *dim-2* leads to hyperacetylation of histone H4 (data not shown) but does not account for the observed hypomethylation induced by TSA. Conversely, model 2 does not account for the effects of DNA methylation on histone acetylation. The selectivity of demethylation by TSA could reflect an absence of histone acetyltransferases in regions that do not lose methylation in response to TSA (e.g., in model 3) or may reflect an indirect effect, e.g., through transcription (model 4).

Preliminary information on the relationship between histone acetylation and DNA methylation came from chromatin immunoprecipitation with antibodies prepared against tetra-acetylated histone H4 to assess acetylation of histone H4 molecules associated with particular genomic regions. A region
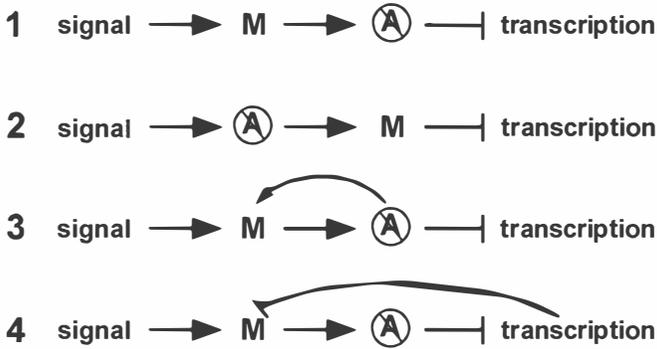
1 signal ⟶ M ⟶ (A) ⊣ transcription

2 signal ⟶ (A) ⟶ M ⊣ transcription

3 signal ⟶ M ⟶ (A) ⊣ transcription

4 signal ⟶ M ⟶ (A) ⊣ transcription

Fig. 6. Four potential relationships of methylation signal, DNA methylation (M), histone hypoacetylation (slashed A), and transcription. Arrows and blocked lines indicate stimulation and inhibition, respectively. See text for details.

whose DNA methylation is reduced by TSA (am^{RIP}) showed corresponding hyperacetylation, as expected, whereas regions whose DNA methylation was unaffected by TSA (e.g., Ψ_{63}) showed no change in histone acetylation (data not shown). This finding is consistent with the idea that not all DNA regions have the capability to become hyperacetylated, possibly because they lack targets for histone acetyltransferases. Interestingly, we also identified an unmethylated DNA region that is hypoacetylated, suggesting that hypoacetylation of histone H4 *per se* does not directly lead to DNA methylation, which conflicts with models 2 and 3 (Fig. 6).

The DIM-5 Histone Methyltransferase Integrates Histone Code Information and Controls DNA Methylation

Results from investigating a *Neurospora* mutant, *dim-5*, which is completely defective in DNA methylation, provided the strongest evidence that histones are intimately involved in DNA methylation (42). Isolation and characterization of the *dim-5* gene showed that it encodes a histone H3 methyltransferase specific for Lys-9 (ref. 42 and data not shown). To confirm that histone H3 is the biologically significant target of DIM-5, we constructed mutant histone H3 genes that replaced Lys-9 with Leu or Arg and introduced the mutant genes into a strain bearing an allele of the bacterial *hph* gene that was silenced by DNA methylation (43). The transformants showed reactivation of *hph* and dramatic loss of methylation of all chromosomal regions examined (42). Thus methylation of histone H3 controls DNA methylation in *Neurospora*.

It seems likely that DIM-5 serves to integrate information relevant to whether DNA in a particular chromosome region should be methylated. *In vitro* assays of DIM-5 activity on various substrates showed that the enzyme is strongly inhibited by methylation of Lys-4, phosphorylation of Ser-10 or acetylation of Lys-14 of histone H3 (E. Schmidt, H.T., and E.U.S., unpublished work), similar to observations with other K9 histone H3 methyltransferases (44–46). This provides a possible mechanistic explanation for our observation that an inhibitor of an HDAC can inhibit DNA methylation (26). Interestingly, and quite unexpectedly, DIM-5 showed strong activity on a peptide dimethylated at Lys-9 (data not shown). This contrasts with available information on previously described H3 histone methyltransferase (44, 45) and raises the possibility that trimethyl-K9 is the specific mark for cytosine methylation in *Neurospora*.

Our working model (Fig. 7) summarizes the likely elements of the DNA methylation machinery of *Neurospora*. The aberrant base composition resulting from RIP may be recognized by MRBP-1 or a similar factor that recruits an HDAC complex. Based on findings in other eukaryotes that have revealed con-
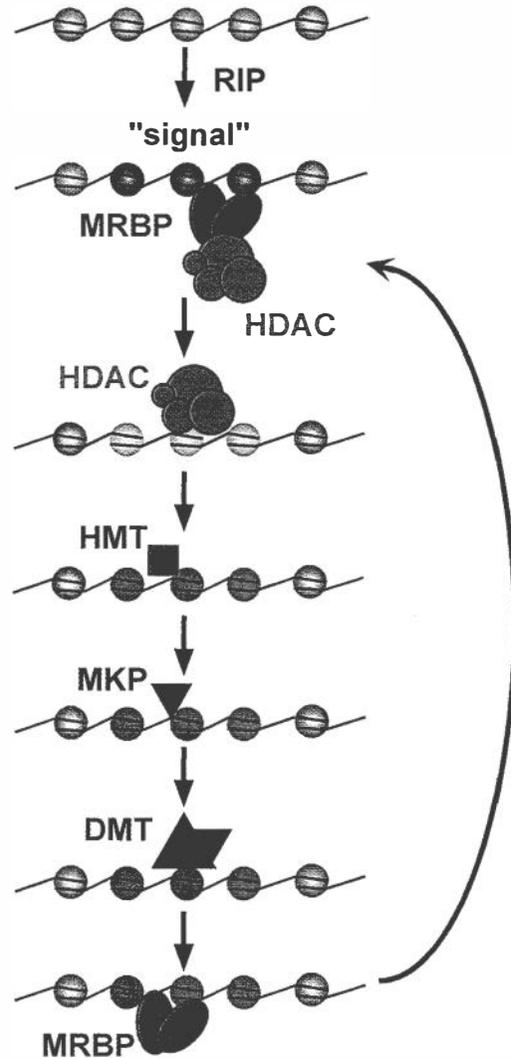

Fig. 7. Stylized working model for the induction and maintenance of DNA methylation in *Neurospora crassa*. The precise nature of the signal for DNA methylation resulting from heavy mutagenesis by RIP (indicated by black DNA on gray nucleosomes) is unknown but seems to consist, at least in part, of structural features of A·T-rich DNA (H.T. and E.U.S., unpublished work; ref. 53). The signal may be recognized by a factor such as MRBP-1, which recognizes RIP-mutated and/or methylated DNA (Fig. 5). Such a factor may directly or indirectly (e.g., through chromatin remodeling factors) recruit one or more HDACs. Histone hypoacetylation (green nucleosomes) of particular residues is postulated to be one prerequisite for DIM-5 histone methyltransferase (HMT) activity. Methylated K9 (orange nucleosomes), which results from the action of DIM-5 on histone H3 (Fig. 7), is presumably recognized by a chromo-domain protein (47) or some other methyl-lysine recognition protein (MKP). The MKP then would directly or indirectly cause the DMT, DIM-2 (39), to methylate (red) the associated DNA. Once methylated, the DNA may bind more strongly to MBRP and/or other methyl-DNA binding proteins, and these, in turn, could again recruit HDAC complexes, resulting in a maintenance loop (red arrow). Whether the various factors that are (for simplicity) shown transiently associated with chromatin actually remain associated with the chromatin is not yet established.

nections between such complexes and chromatin remodeling factors (CRFs; ref. 47) and on indications from plants (48) and animals (49, 50) that putative CRFs are involved in DNA methylation, it seems likely that one or more such factors are involved in this step. At present, we do not know which amino acid residues of histones must be hypoacetylated to allow for the action of DIM-5, but they probably include K9 and K14 of

histone H3 (44) and perhaps one or more of K5, K8, K12, and K16 of histone H4. In addition, we have preliminary evidence that DIM-5, or a DIM-5 complex, is sensitive to other modifications, including the methylation, phosphorylation, and acetylation state of K4, S10, and K14 of histone H3, respectively (data not shown), consistent with observations on histone methyltransferases from other organisms (44, 45, 46). The product of DIM-5 may be recognized by a chromo-domain protein, such as a homologue of HP1 (47). There is at least one candidate for such a protein in the *Neurospora* proteome (www-genome.wi.mit.edu/annotation/fungi/*Neurospora*/). The hypothetical *Neurospora* methyl-lysine recognition protein (MKP) may directly or indirectly recruit the DMT, DIM-2. Conceivably, DIM-2 could directly recruit chromatin modification factors, as in animals (51). In any case, once methylated, the DNA should attract methyl-binding proteins, such as MRBP-1, which would again recruit chromatin modification enzymes, strengthening the epigenetic loop and resulting in the observed propagation of DNA methylation.

1. Holliday, R. & Pugh, J. E. (1975) *Science* **187**, 226–232.
2. Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14**, 9–25.
3. Harland, R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2323–2327.
4. Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A. & Cedar, H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 61–65.
5. Wigler, M., Levy, D. & Perucho, M. (1981) *Cell* **24**, 33–40.
6. Lorincz, M. C., Schubeler, D. & Groudine, M. (2001) *Mol. Cell. Biol.* **21**, 7913–7922.
7. Jeddeloh, J. A., Bender, J. & Richards, E. J. (1998) *Genes Dev.* **12**, 1714–1725.
8. Bestor, T. H. & Ingram, V. M. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2674–2678.
9. Bolden, A. H., Nalin, C. M., Ward, C. A., Poonian, M. S. & Weissbach, A. (1986) *Mol. Cell. Biol.* **6**, 1135–1140.
10. Selker, E. U. & Stevens, J. N. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8114–8118.
11. Selker, E. U., Fritz, D. Y. & Singer, M. J. (1993) *Science* **262**, 1724–1728.
12. Toth, M., Müller, U. & Doerfler, W. (1990) *J. Mol. Biol.* **214**, 673–683.
13. Woodcock, D. M., Crowther, P. J. & Diver, W. P. (1987) *Biochem. Biophys. Res. Commun.* **145**, 888–894.
14. Selker, E. U. (1990) *Trends Biochem. Sci.* **15**, 103–107.
15. Yisraeli, J. & Szyf, M. (1984) in *DNA Methylation: Biochemistry and Biological Significance*, eds. Razin, A., Cedar, H. & Riggs, A. D. (Springer, New York), pp. 353–378.
16. Turker, M. S., Swisshelm, K., Smith, A. C. & Martin, G. M. (1989) *J. Biol. Chem.* **264**, 11632–11636.
17. Finnegan, E. J., Brettell, R. I. S. & Dennis, E. S. (1993) in *DNA Methylation: Molecular Biology and Biological Significance*, eds. Jost, J. P. & Saluz, H. P. (Birkäuser, Basel), pp. 218–261.
18. Kishimoto, N., Sakai, H., Jackson, J., Jacobsen, S. E., Meyerowitz, E. M., Dennis, E. S. & Finnegan, E. J. (2001) *Plant Mol. Biol.* **46**, 171–183.
19. Pelissier, T., Thalmeir, S., Kempe, D., Sanger, H. L. & Wassenegger, M. (1999) *Nucleic Acids Res.* **27**, 1625–1634.
20. Goubely, C., Arnaud, P., Tatout, C., Heslop-Harrison, J. S. & Deragon, J. M. (1999) *Plant Mol. Biol.* **39**, 243–255.
21. Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S. & Jacobsen, S. E. (2001) *Science* **292**, 2077–2080.
22. Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. & Jaenisch, R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5237–5242.
23. Vu, T. H., Li, T., Nguyen, D., Nguyen, B. T., Yao, X. M., Hu, J. F. & Hoffman, A. R. (2000) *Genomics* **64**, 132–143.
24. Rossignol, J.-L. & Faugeron, G. (1994) in *Gene Silencing in Higher Plants and Related Phenomena in Other Eukaryotes*, ed. Meyer, P. (Springer, Heidelberg), Vol. 197, p. 26.
25. Selker, E. U. (2002) *Adv. Genet.* **46**, 439–450.
26. Selker, E. U. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 9430–9435.
27. Rountree, M. R., Bachman, K. E., Herman, J. G. & Baylin, S. B. (2001) *Oncogene* **20**, 3156–3165.
28. Selker, E. U. & Garrett, P. W. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6870–6874.
29. Selker, E. U. (1990) *Annu. Rev. Genet.* **24**, 579–613.
30. Selker, E. U., Cambareri, E. B., Jensen, B. C. & Haack, K. R. (1987) *Cell* **51**, 741–752.
31. Cambareri, E. B., Jensen, B. C., Schabtach, E. & Selker, E. U. (1989) *Science* **244**, 1571–1575.
32. Freitag, M., Williams, R., Kothe, G. O. & Selker, E. U. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 8802–8807.
33. Watters, M. K., Randall, T. A., Margolin, B. S., Selker, E. U. & Stadler, D. R. (1999) *Genetics* **153**, 705–714.
34. Singer, M. J., Marcotte, B. A. & Selker, E. U. (1995) *Mol. Cell. Biol.* **15**, 5586–5597.
35. Selker, E. U., Richardson, G. A., Garrett-Engele, P. W., Singer, M. J. & Miao, V. (1993) *Cold Spring Harbor Symp. Quant. Biol.* **58**, 323–329.
36. Cambareri, E. B., Singer, M. J. & Selker, E. U. (1991) *Genetics* **127**, 699–710.
37. Miao, V. P. W., Singer, M. J., Rountree, M. R. & Selker, E. U. (1994) *Mol. Cell. Biol.* **14**, 7059–7067.
38. Rountree, M. R. & Selker, E. U. (1997) *Genes Dev.* **11**, 2383–2395.
39. Kouzminova, E. A. & Selker, E. U. (2001) *EMBO J.* **20**, 4309–4323.
40. Foss, H. M., Roberts, C. J., Claeys, K. M. & Selker, E. U. (1993) *Science* **262**, 1737–1741.
41. Dobosy, J. R. & Selker, E. U. (2001) *Cell Mol. Life Sci.* **58**, 721–727.
42. Tamaru, H. & Selker, E. U. (2001) *Nature (London)* **414**, 277–283.
43. Irelan, J. T. & Selker, E. U. (1997) *Genetics* **146**, 509–523.
44. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D. & Grewal, S. I. (2001) *Science* **292**, 110–113.
45. Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. & Jenuwein, T. (2000) *Nature (London)* **406**, 593–599.
46. Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P. & Reinberg, D. (2002) *Genes Dev.* **16**, 479–489.
47. Richards, E. J. & Elgin, S. C. (2002) *Cell* **108**, 489–500.
48. Jeddeloh, J. A., Stokes, T. L. & Richards, E. J. (1999) *Nat. Genet.* **22**, 94–97.
49. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. (2001) *Genes Dev.* **15**, 2940–2944.
50. Gibbons, R. J., McDowell, T. L., Raman, S., O'Rourke, D. M., Garrick, D., Ayyub, H. & Higgs, D. R. (2000) *Nat. Genet.* **24**, 368–371.
51. Rountree, M. R., Bachman, K. E. & Baylin, S. B. (2000) *Nat. Genet.* **25**, 269–277.
52. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
53. Miao, V. P., Freitag, M. & Selker, E. U. (2000) *J. Mol. Biol.* **300**, 249–273.

Locus-specific control of asymmetric and CpNpG methylation by the *DRM* and *CMT3* methyltransferase genes

Xiaofeng Cao* and Steven E. Jacobsen*[†][‡]

*Department of Molecular, Cell, and Developmental Biology, and [†]Molecular Biology Institute, University of California, Los Angeles, CA 90095-1606

Many plant, animal, and fungal genomes contain cytosine DNA methylation in asymmetric sequence contexts (CpHpH, H = A, T, C). Although the enzymes responsible for this methylation are unknown, it has been assumed that asymmetric methylation is maintained by the persistent activity of *de novo* methyltransferases (enzymes capable of methylating previously unmodified DNA). We recently reported that the *DOMAINS REARRANGED METHYLASE* (*DRM*) genes are required for *de novo* DNA methylation in *Arabidopsis thaliana* because *drm1 drm2* double mutants lack the *de novo* methylation normally associated with transgene silencing. In this study, we have used bisulfite sequencing and Southern blot analysis to examine the role of the *DRM* loci in the maintenance of asymmetric methylation. At some loci, *drm1 drm2* double mutants eliminated all asymmetric methylation. However, at the *SUPERMAN* locus, asymmetric methylation was only completely abolished in *drm1 drm2 chromomethylase 3* (*cmt3*) triple mutant plants. *drm1 drm2* double mutants also showed a strong reduction of CpNpG (*n* = A, T, C, or G) methylation at some loci, but not at others. The *drm1 drm2 cmt3* triple mutant plants did not affect CpG methylation at any locus tested, suggesting that the primary CpG methylases are encoded by the *MET1* class of genes. Although neither the *drm1 drm2* double mutants nor the *cmt3* single mutants show morphological defects, *drm1 drm2 cmt3* triple mutant plants show pleiotropic effects on plant development. Our results suggest that the *DRM* and *CMT3* genes act in a partially redundant and locus-specific manner to control asymmetric and CpNpG methylation.

Cytosine DNA methylation plays a major role in gene silencing and heterochromatin formation (1). In mammals, methylation is largely restricted to CpG dinucleotides, but low levels of methylation at asymmetric sites are found in some cell types (2). Asymmetric DNA methylation is also found in some fungal genomes. For instance, *Neurospora crassa* shows dense asymmetric methylation associated with a phenomenon called Repeat-Induced Point mutation (RIP) (3), and *Ascobolus immersus* shows asymmetric methylation associated with the Methylation Induced Premeiotically (MIP) phenomenon (4). Plant genomes contain high levels of asymmetric methylation, and also contain abundant CpNpG methylation (5–8).

The symmetry of the CpG site was proposed to be important for stable maintenance of methylation patterns after DNA replication (9, 10). Replication of symmetrical sites would produce hemimethylated sequences, which were proposed to be preferred targets for a maintenance methyltransferase that would methylate cytosines in the newly synthesized DNA strand (11). Consistent with these early ideas, Dnmt1, the major mammalian CpG methyltransferase (12), is known to prefer DNA substrates containing hemimethylated CpG dinucleotides (13). Dnmt1 also localizes to DNA replication foci (14) consistent with the notion that maintenance methylation and DNA

replication are tightly coupled. The CpNpG site methylated in plants is also symmetric. Although it is attractive to imagine that maintenance of CpNpG methylation is similar to that of CpG methylation, essentially nothing of the mechanism is known. Methylation of asymmetric or nonpalindromic sequences, on the other hand, could well be maintained by a mechanism different from that of symmetric sites. Asymmetric methylation must be reestablished after each DNA replication cycle, because there is no complementary sequence to serve as a guide for remethylation of particular cytosines (2, 15, 16).

The function of asymmetric methylation remains unclear. In one study, all symmetric CpG and CpNpG sites were removed from a 35S transgene that is normally subject to transsilencing by a second transgene which contains the 35S promotor present in a repetitive array (17). The symmetric site-free transgene became heavily methylated at the asymmetric sites. Furthermore, the removal of symmetric sites did not prevent transcriptional silencing of this transgene, but did prevent the maintenance of silencing in the absence of the repetitive array. *In vitro* methylation of the asymmetric cytosines also reduced the transcriptional activity of this 35S promotor in protoplasts (18). These studies show that asymmetric methylation can exist in the absence of symmetric methylation and may contribute to gene silencing.

Arabidopsis thaliana has at least three classes of DNA methyltransferase genes that are possible candidates for controlling asymmetric methylation, the *MET1* class, the *CMT3* (*CHROMOMETHYLASE 3*) class, and the *DOMAINS REARRANGED METHYLASE* (*DRM*) class (19). Although it seems likely that these genes encode enzymes that are active methyltransferases, this has not yet been directly demonstrated by studying the *in vitro* enzymatic properties of the proteins.

The *MET1* class of genes (20) is most similar to *Dnmt1* in both sequence and function. Loss-of-function *met1* mutants, (also called *ddm2* mutants) and antisense-*MET1* transgenic plants lack the majority of CpG methylation (20–25). In contrast to mutant *Dnmt1* mice, which die after 9 days of development, antisense-*MET1* or *met1* mutant plants are viable, but display a number of developmental abnormalities that become progressively more extreme as the mutants are inbred (22, 23). Surprisingly, some of these abnormalities are caused by ectopic hypermethylation of particular genes, such as *SUPERMAN* (*SUP*) and *AGAMOUS* (8, 26), a phenomenon that superficially resembles

---

the hypermethylation of human tumor suppressor genes in otherwise hypomethylated genomic backgrounds (27).

The *CMT3* class of genes are specific to the plant kingdom and encode methyltransferase proteins containing a chromo domain (28). *CMT3* loss-of-function mutants were isolated in three independent studies, and show a genome-wide loss of CpNpG methylation and a reduction of asymmetric methylation at some loci (24, 29, 30). CpNpG methylation is also affected by histone methylation, because loss-of-function mutations in the *KRYP-TONITE* (*KYP*) histone H3 lysine 9 methyltransferase gene show a loss of DNA methylation in CpNpG contexts (31).

The third class of genes, composed of *DRM1* and *DRM2*, contain catalytic domains showing sequence similarity to those of the mammalian Dnmt3 methyltransferases (32). *Dnmt3a* and *Dnmt3b* encode important *de novo* methyltransferases (33–36). For instance, mice harboring mutations in both enzymes lack the ability to *de novo* methylate retroviruses when introduced into embryonic stem cells (35). Furthermore, in humans, *Dnmt3b* mutations reduce CpG methylation at some pericentromeric sequences and are the cause of a rare recessive disorder called ICF syndrome, for immunodeficiency, centromeric instability, and facial anomalies (35, 37, 38). We recently reported that the *DRM* genes are required for the initial establishment of methylation of cytosines in all known sequence contexts: CpG, CpNpG, and asymmetric (39). We found that *drm1 drm2* double mutants lacked *de novo* methylation of the direct repeats of the *FWA* locus, which normally occurs when *FWA* is transformed into wild-type plants (39). *drm1 drm2* double mutants also blocked *de novo* methylation of the *SUP* locus, which occurs in the presence of a *SUP* inverted repeat. However, the *drm* mutants did not show reactivation of previously methylated and silenced *FWA* or *SUP* epigenetic alleles, suggesting that the *DRM* genes are required for the establishment but not the maintenance of gene silencing (39).

In this report, we show that the *DRM* loci are required for the maintenance of asymmetric DNA methylation. However, at some loci such as *SUP*, we find that *DRMs* act redundantly with *CMT3*, so that only in *drm1 drm2 cmt3* triple mutants is all asymmetric methylation lost. Furthermore, we find that at some loci the *DRMs* are more important for the maintenance of CpNpG methylation than *CMT3*. Our data suggest that the *DRM* and *CMT3* genes encode partially redundant methyltransferases, and that different sequence or chromatin contexts can modulate their function.

Materials and Methods

Genetic Analysis. The *drm1 drm2, clk-st, drm1 drm2 clk-st,* and *cmt3-7 clk-st* strains used in these studies, and the PCR-based molecular markers used to follow each gene, were recently described (24, 39). We constructed *drm1 drm2 cmt3-7* triple mutants in a homozygous *clk-st* background by crossing a *cmt3-7 clk-st* plant to a *drm1 drm2 clk-st* plant and PCR genotyping the F2 progeny. We identified five independent *drm1 drm2 cmt3-7 clk-st* plants. Seeds of the *met1* (*ddm2*) mutant were a kind gift of Eric Richards (Department of Biology, Washington University, St. Louis).

Genomic Bisulfite Sequencing. Two micrograms of genomic DNA from the appropriate genotype was digested with restriction enzymes that cut just outside of the region of interest. Bisulfite treatment was performed as described (26). Because our previous studies found similar patterns of methylation on the two strands of DNA (8, 40), only one strand was analyzed for each locus. The bottom strand of *MEA-ISR* analyzed corresponds to positions 68067–68320 of bacterial artificial chromosome (BAC) clone T14P4 (GenBank accession no. AC022521). Primers for *MEA-ISR* are: JP1026 5'-AAA GTG GTT GTA GTT TAT GAA AGG TTT TAT and JP1027 5'-CTT AAA AAA TTT

TCA ACT CAT TTT TTT TAA AAA A. PCR products were cloned by using the TOPO TA cloning kit (Invitrogen), and 15 individual clones were sequenced. Bisulfite sequencing on the top strand of *SUP* was performed as described (24). Bisulfite sequencing of the top strand of *FWA* was as described (40).

Southern Blot Analysis. *MEA-ISR* probe was amplified from genomic DNA with following primers: JP980 5'-AAA CCT TTC GTA AGC TAC AGC CAC TTT GTT-3'; JP981 5'-TCG GAT TGG TTC TTC CTA CCT CTT TAC CTT-3', corresponding to positions 68348–69438 on BAC clone T14P4 (GenBank accession no. AC022521). *FWA* probe was a 1.7-kb PCR-generated DNA fragment corresponding to positions 498-2281 in GenBank accession no. AF178688. PCR primers were JP 1084 5'-CTT CGC CTT TCT CTT CCT CAT CTG CGC TT-3' and JP1085 5'-GAG TTT GAT AAG CAG TGC GCC TAT GGG TT-3'. *Ta3* probe was as described (41) and amplified by JP1154 5'-GAT CTA TCT GGC CCC AGA CGT AGA TCT AA-3' and JP1155 5'-CCG GCA ATC TAC TAT ATG AGA TCT TTA CAA-3'; 180-bp centromere repeat probes were described in ref. 21.

Results

DRMs Control Asymmetric and CpNpG Methylation at *FWA* and *MEA-ISR*. To determine the effect of the *DRM* loci in the maintenance of asymmetric methylation, we used bisulfite genomic sequencing to study the levels of methylation at two endogenous sequences that are methylated at asymmetric sites, *FWA* and *MEA-ISR*. The *FWA* locus encodes a homeodomain-containing protein whose expression is silenced in the vegetative tissues of wild-type plants. This silencing is associated with methylation of two direct repeats in the 5' region of the gene (40). When this methylation is lost, either in spontaneous hypomethylated epigenetic mutants such as *fwa-1*, or in the methylation mutants *ddm1* and *met1*, the *FWA* gene is overexpressed causing a dominant late flowering phenotype (24, 40, 42). In wild type, the *FWA* direct repeats contain 89% CpG, 18% CpNpG, and 4% asymmetric methylation (Fig. 1 and Table 2, which is published as supporting information on the PNAS web site, www.pnas.org) (40). *MEA-ISR* is an approximately 183-bp sequence present in seven direct repeats lying in an intergenic region between the imprinted medea (*MEA*) gene (43, 44) and the aldehyde oxidase gene on BAC clone T14P4 near the upper end of chromosome 1, approximately 500,000 bp from the end. These repeats are also found in 12 other genomic locations, all of which are also subtelomeric. For this reason, we named this sequence *MEA-ISR* for Intergenic Subtelomeric Repeat. We found that this repeat shows high levels of DNA methylation in wild-type strains, namely 87% CpG, 47% CpNpG, and 18% asymmetric (Fig. 1).

We compared the parental Wassilewskija (WS) strain with a *drm1 drm2* double mutant strain by using bisulfite genomic sequencing and found that the *drm1 drm2* mutations eliminated all asymmetric methylation of both the *FWA* and *MEA-ISR* sequences. Therefore, the *DRM* loci are important for the methylation of asymmetric cytosines.

The *drm1 drm2* mutations also eliminated the CpNpG methylation of both *FWA* and *MEA-ISR* (Fig. 1). Because *cmt3* mutants were previously found to strongly reduce methylation of CpNpG sites of all loci tested (24, 30), we next analyzed *FWA* and *MEA-ISR* in the *cmt3-7* mutant, a null *cmt3* allele (24). We found that *cmt3-7* reduced, but did not eliminate, the CpNpG methylation (Fig. 1). Therefore, both *DRM* and *CMT3* are required for proper maintenance of CpNpG methylation patterns, and at some loci, such as *FWA* and *MEA-ISR*, the *drm1 drm2* double mutant is more effective than *cmt3-7* at reducing CpNpG methylation.

We also analyzed *drm1 drm2 cmt3-7* triple mutants by bisulfite sequencing (Fig. 1) and found that, like *drm1 drm2* double
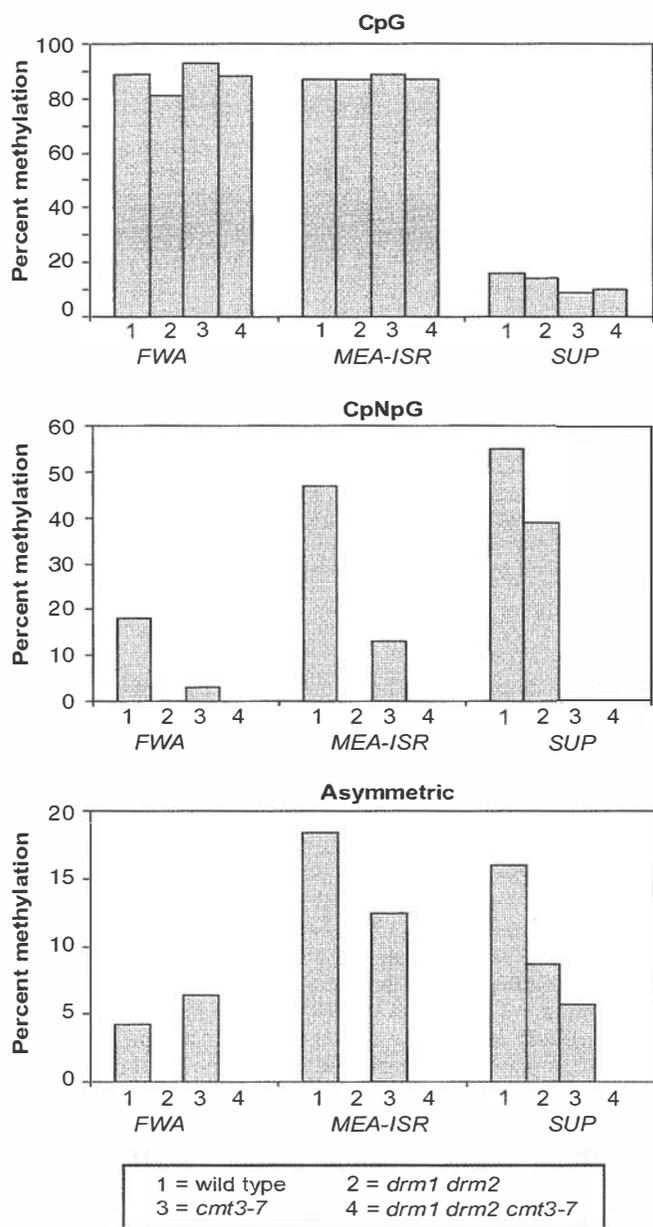
Cao and Jacobsen

Fig. 1. Diagram represents bisulfite sequencing of a 500-nt region of the top strand of the *FWA* gene, 219-nt region of the bottom strand of the *MEA-ISR* region, and 1,028-nt region of the top strand of the *SUP* gene in different mutant backgrounds. Detailed data supporting the graphical presentation can be found in Table 2 and Fig. 6. For the analysis of *FWA* and *MEA-ISR*, the wild-type strain is WS. For the *SUP* gene, the wild-type strain is *clk-st* in the Landsberg *erecta* background. *drm1 drm2*, *cmt3-7*, and *drm1 drm2 cmt3* mutants are homozygous for the *clk-st* inverted repeat *SUP* transgene. Height of the bars represents the percentage of methylation at each site of 15 clones analyzed by bisulfite sequencing.

mutants, the triple mutants lacked all traces of asymmetric and CpNpG methylation at *FWA* and *MEA-ISR*. However, CpG methylation levels were similar to the wild type.

We used Southern blot analysis with methylation-sensitive restriction enzymes to confirm some of the bisulfite sequencing results at *FWA* and *MEA-ISR*. In these analyses, we also included the *drm1* and *drm2* single mutants as well as the *met1* mutant. For the *FWA* blots, we included the *fwa-1* epigenetic mutant (40) as a completely unmethylated control. Methylation at the *FWA* gene was assayed with two methylation-sensitive restriction
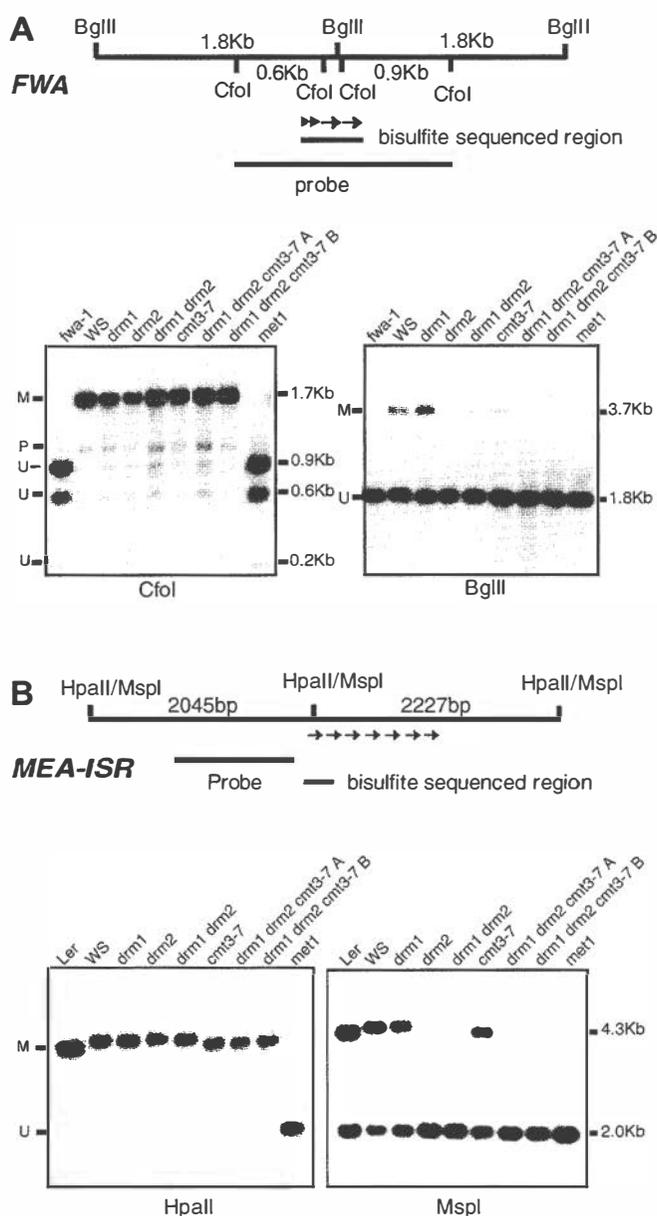
Fig. 2. Southern blot analysis of *FWA* and *MEA-ISR*. (*A Upper*) A diagram of the *Cfo*I and *Bgl*II restriction fragments present in the *FWA* gene. The inner two *Cfo*I sites and the *Bgl*II site are within the methylated direct repeats of the *FWA* promoter, shown as arrows above. The region analyzed by bisulfite sequencing is also noted. (*Lower*) DNA blots of equivalent amounts of *Cfo*I (*Left*) and *Bgl*II (*Right*) digested genomic DNA probed with a 1.74-kb fragment corresponding to the diagram on top. The positions of the methylated (M), unmethylated (U), and partially methylated (P) bands are shown, as well as the size of the bands in kilobases. *drm1 drm2 cmt3-7* A and B are two independently isolated triple mutant strains. (*B Upper*) A diagram of the *Hpa*II/*Msp*I restriction fragments present in the *MEA-ISR* region. The inner *Hpa*II/*Msp*I site lies in the first of seven direct repeats (shown as arrows) and the probe lies in the unique sequence between these repeats and the *MEA* locus. (*Lower*) DNA blots of equivalent amounts of *Hpa*II (*Left*) and *Msp*I (*Right*) digested genomic DNA probed with a 1.1-kb fragment corresponding to the diagram on top.

enzymes, *Cfo*I and *Bgl*II, which are found within the methylated direct repeats (Fig. 2*A*). *Cfo*I recognizes the sequence GCGC and is inhibited by CpG methylation of the site. Previous Southern blot results showed that *drm1 drm2* double mutants (39) and the *cmt3-7* single mutant (31) did not affect CpG methylation at *FWA*. Here we compared two independently

isolated *drm1 drm2 cmt3* triple mutants (named A and B) with all of the single mutants, and found that all mutants or combinations of mutants showed the same level of CpG methylation as the wild-type control WS (Fig. 2*A*). This finding is consistent with the bisulfite sequence data indicating that the *drm1 drm2 cmt3* triple mutants do not affect preexisting CpG methylation at *FWA*. CpG methylation was completely lost in a *met1* mutant line that had developed an *fwa* late flowering phenotype (Fig. 2*A*). This finding further confirms that *MET1* maintains CpG methylation (20, 22–25).

The restriction enzyme *Bgl*II is inhibited by cytosine methylation within its recognition site AGATCT (45), and we have used this enzyme to detect CpNpG methylation of *FWA*. Because of the sequence context at *FWA* (A*AG ATCT*G), *Bgl*II will detect CpTpG methylation at a site that we previously found to be methylated by bisulfite sequencing (40). *Bgl*II would also be inhibited by methylation of the asymmetric CpTpT site on the other strand, but because we have not found methylation at this site by bisulfite sequencing, this is unlikely to complicate the analysis. A 3.7-kb *Bgl*II fragment caused by cytosine methylation at the *Bgl*II site appears in wild-type WS but not in *fwa-1* (Fig. 2*A*). A faint 3.7-kb methylated band is detected in *cmt3-7*, consistent with bisulfite sequence data showing that CpNpG methylation was reduced but not completely lost in *cmt3-7* (Fig. 1). *drm1 drm2* double mutants and *drm1 drm2 cmt3-7* triple mutants eliminated methylation of the *Bgl*II site, consistent with the bisulfite data showing a lack of CpNpG methylation in these strains (Fig. 1).

Methylation of the *Bgl*II site was also lost in the *drm2* single mutant, but not in the *drm1* single mutant (Fig. 2*A*). This finding fits with our previous observations that *drm2* but not *drm1* blocked *FWA* transgene *de novo* methylation (39), and with the fact that *DRM2* RNA is expressed at a much higher level than *DRM1* RNA (32). Thus *DRM2* is likely to encode the predominant enzyme in the DRM family.

For Southern blot analysis of the *MEA-ISR* region, we used a single *Hpa*II/*Msp*I site (CCGG) present within the first direct repeat. We used a probe within unique sequence adjacent to these repeats to assay methylation of this site (Fig. 2*B*). *Hpa*II is inhibited by methylation of either cytosine within its recognition site allowing detection of CpG and CpNpG methylation, whereas *Msp*I is only inhibited by methylation of the outer cytosine allowing detection of CpNpG methylation. Similar to the wild-type strains WS and Ler, *drm1 drm2*, *cmt3-7* and *drm1 drm2 cmt3-7* triple mutant plants were not cut by *Hpa*II, confirming the bisulfite data showing that these mutants do not affect CpG methylation (Fig. 1). However, the *met1* mutant showed complete digestion with *Hpa*II suggesting a loss of methylation of the site (Fig. 2*B*).

The *Msp*I Southern blot of the *MEA-ISR* sequence revealed about 50% cleavage of wild-type Ler and WS DNAs (Fig. 2*B*), consistent with the 47% CpNpG methylation determined by bisulfite sequence (Fig. 1). Similar to results with the *FWA* gene, *drm2* but not *drm1* caused a loss of all CpNpG methylation at *MEA-ISR*. *cmt3-7* retained a fair amount of CpNpG methylation, again consistent with the bisulfite data (Fig. 1). These results confirm that, at the *MEA-ISR* locus, *cmt3-7* single mutants and *drm1 drm2* double mutants both reduce methylation at CpNpG sites, but *drm1 drm2* does so much more efficiently.

In conclusion, *drm1 drm2* double mutants showed a complete loss of non-CpG methylation at both *FWA* and *MEA-ISR*. Interestingly, *drm1 drm2* plants do not show a late flowering phenotype typical of plants in which expression of the *FWA* gene has been reactivated (39). This suggests that non-CpG methylation does not play major role in the silencing of *FWA*.

Both DRM and CMT3 Control Asymmetric Methylation at SUP. Epigenetic silenced alleles of the *SUP* locus (the *clark kent* or *clk*

Table 1. 5equence context of the asymmetric methylation found at the *SUP* locus within 15 cloned PCR products of bisulfite-treated DNA

| Site | % methylation of site | | | | No. of sites |
	Wild type	cmt3-7	drm1 drm2	drm1 drm2 cmt3-7	
All sites*	16.2	5.7	8.7	0	204
CpA	17.1	4.6	8.7	0	70
CpT	22	9.5	13.0	0	86
CpC	4.7	0.7	1.0	0	48
Ap**C**	8.6	0.6	4.5	0	46
Tp**C**	18.3	7.1	8.9	0	83
Gp**C**	13.3	4.7	5.8	0	31
Cp**C**	22.4	9.1	14.7	0	44
CpNp**C**†	22.5	8.4	12.7	0	61
CpCp**C**	40.0	15.6	30.4	0	9

The genotypes analyzed were either the unmutagenized *clk-st* line (wild-type), or *clk-st* containing the *cmt3-7* and/or *drm1 drm2* mutations.
*Asymmetric methylation (All sites) are cytosines within the context CpHpH, where H = A, T, or C. In each context shown, the methylated cytosine is in bold.
†N = A, T, C, or G.

alleles) are densely methylated in all sequence contexts (8, 24, 25). Whereas the originally isolated *clk* alleles spontaneously revert to an unmethylated state ≈3% of the time (8), a transgenic allele called *clk-st* shows a stable nonreverting phenotype, making it more suitable for genetic studies (24). *clk-st* contains a single inverted repeat of the *SUP* locus, which can induce *de novo* methylation of itself as well as of the endogenous *SUP* locus (described in detail in reference 39). The methylation consists of 16% CpG, 55% CpNpG and 16% asymmetric (Fig. 1) (24). We previously found that *cmt3-7* eliminated the CpNpG methylation, and reduced the asymmetric methylation by approximately 60% (Fig. 1) (24). To test the effect of the *drm* mutations on *SUP* methylation, we compared *clk-st* with a *drm1 drm2 clk-st* strain (39). Contrary to what was found with *FWA* and *MEA-ISR*, we observed that *drm1 drm2* double mutants reduced but did not eliminate asymmetric methylation (Fig. 1 and Fig. 6 and Table 2, which are published as supporting information on the PNAS web site) (24, 39). However, this residual asymmetric methylation was not detected in *drm1 drm2 cmt3-7 clk-st* strains (Fig. 1). Therefore, *DRM* and *CMT3* act redundantly to control *SUP* asymmetric methylation.

Because *drm1 drm2 cmt3-7* triple mutants eliminate all asymmetric methylation at *SUP*, an analysis of the methylation remaining in the *cmt3-7* mutants should reflect the residual activity of *DRM*, and that remaining in *drm1 drm2* should reflect the residual activity of *CMT3*. Fig. 4*A* shows that the positions of the asymmetric methylation of *SUP* remaining in either *cmt3-7* plants or *drm1 drm2* plants were largely overlapping (also see Fig. 6). An analysis of the sequence context of the asymmetric methylation in these mutants suggests that both CMT3 and DRM prefer to methylate sites that follow cytosine residues (Table 1). However, both CMT3 and DRM showed a bias against sites that immediately precede cytosines; CpA and CpT methylation were much more frequent than CpC methylation (Table 1). These results suggest that CMT3 and DRM can methylate the same asymmetric sites, and both show roughly the same preference for particular DNA sequence contexts.

The effects of the *drm* mutations on CpNpG methylation at *SUP* were also different from that seen at *FWA* and *MEA-ISR*. *drm1 drm2* only reduced CpNpG methylation at the *SUP* locus by about 30%, whereas *cmt3-7* completely abolished this methylation. Thus, the effect of *drm1 drm2* on CpNpG methylation is
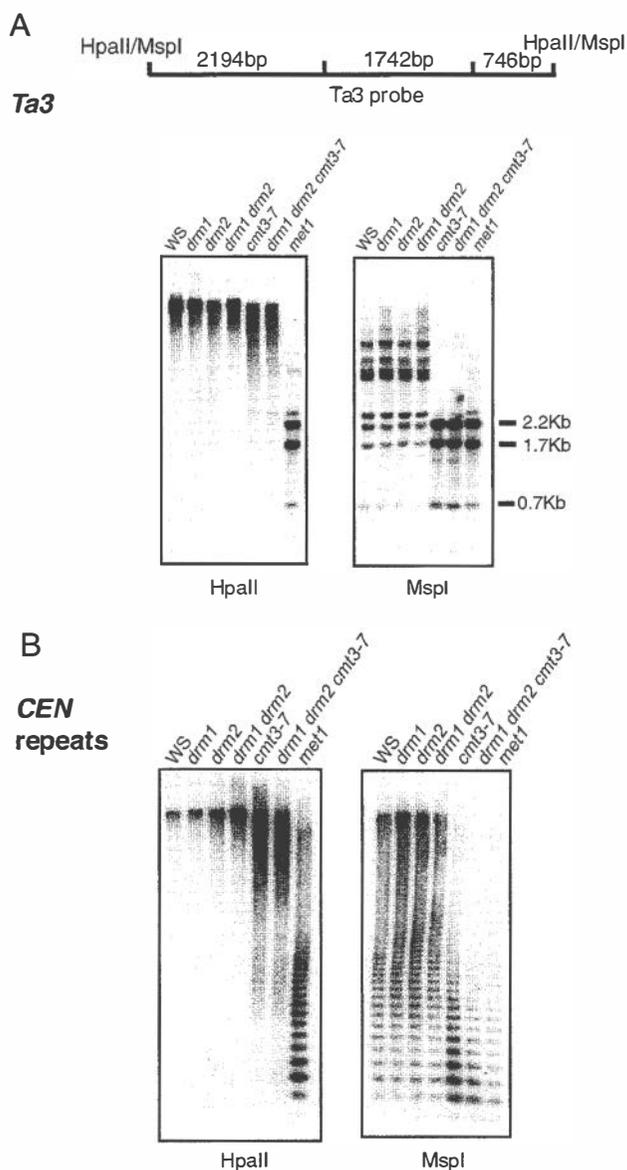
Fig. 3. Southern blot analysis of the *Ta3* retrotransposon and 180-bp centromeric repeat sequences. Each panel shows lanes that contain equivalent amounts of genomic DNA of the indicated genotype digested with *Hpa*II (*Left*) and *Msp*I (*Right*). (*A Upper*) A diagram of the *Hpa*II/*Msp*I restriction fragments present in the *Ta3* probe. (*Lower*) A DNA blot probed with *Ta3* probe. (*B*) Blot probed with a 180-bp centromeric repeat probe.

locus-specific, completely eliminating the CpNpG methylation of *FWA* and *MEA-ISR*, but only reducing it moderately at *SUP*.

CMT3 Controls CpNpG Methylation at the Ta3 Retrotransposon and at Centromeric Repeats. We tested the effect of the various single, double, and triple methyltransferase mutants on two pericentromeric sequences, the *Ta3* retrotransposon, and the 180-bp centromeric repeat sequence, using Southern blot analysis with *Hpa*II and *Msp*I (Fig. 3). Neither the *drm* single mutants nor the *drm1 drm2* double mutants affected the pattern of enzyme digestion, suggesting that the *DRM* genes do not play a role in maintaining CpG or CpNpG methylation at these sequences. In contrast, *cmt3-7* and *drm1 drm2 cmt3-7* triple mutants showed nearly complete digestion by *Msp*I, but not *Hpa*II, showing that that *CMT3* is responsible for the maintenance of CpNpG methylation at both *Ta3* and the centromeric repeats (24, 31).

We also found that, in contrast to *cmt3* mutants (24) and *met1* mutants (X.C, L. Johnson, and S.E.J. unpublished observation), *drm1 drm2* double mutants did not show reactivation of RNA expression of the *Ta3* sequence (data not shown). Therefore, the *DRM* genes do not appear to play a role the maintenance of gene silencing of *Ta3*.

MET1 Likely Affects CpNpG and Asymmetric Methylation Indirectly. Bisulfite sequencing and Southern blot data show that methylation of CpG sites are unaffected in *drm1 drm2 cmt3-7* triple mutant strains at all sequences tested (Figs. 1–3). In contrast, at every sequence tested, the *met1* mutant showed a strong reduction in CpG methylation (Figs. 2 and 3). Therefore, like its mammalian counterpart Dnmt1, MET1 appears to be the primary methyltransferase for the maintenance of CpG methylation. However, consistent with earlier studies (22, 23, 46), we observed that *met1* mutants greatly reduced CpNpG methylation at *FWA*, *MEA-ISR*, *Ta3*, and the centromeric repeat sequences (Figs. 2 and 3), and eliminated the asymmetric methylation of *FWA* and *MEA-ISR* as detected by bisulfite sequencing (data not shown). Because *MET1* cannot substitute for the maintenance of CpNpG and asymmetric methylation in *drm1 drm2 cmt3-7* strains, it seems most likely that the losses of CpNpG and asymmetric methylation in *met1* are not directly caused by promiscuous enzymatic activity of the MET1 enzyme, but are instead a secondary effect caused by the primary loss of CpG methylation.

A Pleiotropic Morphological Phenotype Marks DRM and CMT3 Redundancy. Although neither the *cmt3* mutants (24) nor the *drm1 drm2* double mutants (39) show morphological differences from wild type, *drm1 drm2 cmt3-7* plants showed a pleiotropic phenotype including developmental retardation, reduced plant size, and partial sterility (Fig. 4*B*). We did not observe, however, several defects commonly seen in the *ddm1* and *met1* methylation mutants, such as *clavata*-like flowers, *apetala2*-like flowers, *agamous*-like flowers, *sup*-like flowers, or extreme late flowering (22, 23, 26, 42). Thus, *DRM* and *CMT3* act in a redundant fashion to control some aspects of plant growth and development, which may be different from those affected in *ddm1* and *met1* mutants.

Discussion

Locus-Specific Action of the DRM and CMT3 Methyltransferase Genes. Our results suggest that the *DRM* and *CMT3* genes encode methyltransferase enzymes that show overlapping roles in the control of asymmetric and CpNpG methylation (Fig. 5*A*). However, the activities of these methyltransferases are highly dependent on the locus under study, giving a surprising number of different patterns of dependence on either *DRM*, *CMT3*, or both. Fig. 5*B* summarizes the results of this study, showing the dependence of each methylated gene on *DRM* and/or *CMT3* for both asymmetric and CpNpG methylation. What are the signals that target these methylases to vary their function? It seems likely that several factors could be involved, including chromatin modifications present at specific loci, the DNA sequences involved, and cross talk between CpG methylation and non-CpG methylation.

Chromatin modifications can be considered at two levels: overall chromatin structure, and specific histone modifications. Examples of overall chromatin structure are provided by the *Ta3* retrotransposon and centromeric repeats, which are nested within the highly condensed, pericentromeric, constitutive heterochromatin. The higher order chromatin structures present near centromeres may modulate methyltransferase function in such a way that CMT3 is solely responsible for the methylation of CpNpG sites (Fig. 5*B*). In contrast, *MEA-ISR* is subtelomeric, and *SUP* and *FWA* are in the midst of single copy sequence, which is presumably euchromatic in nature. An example of a
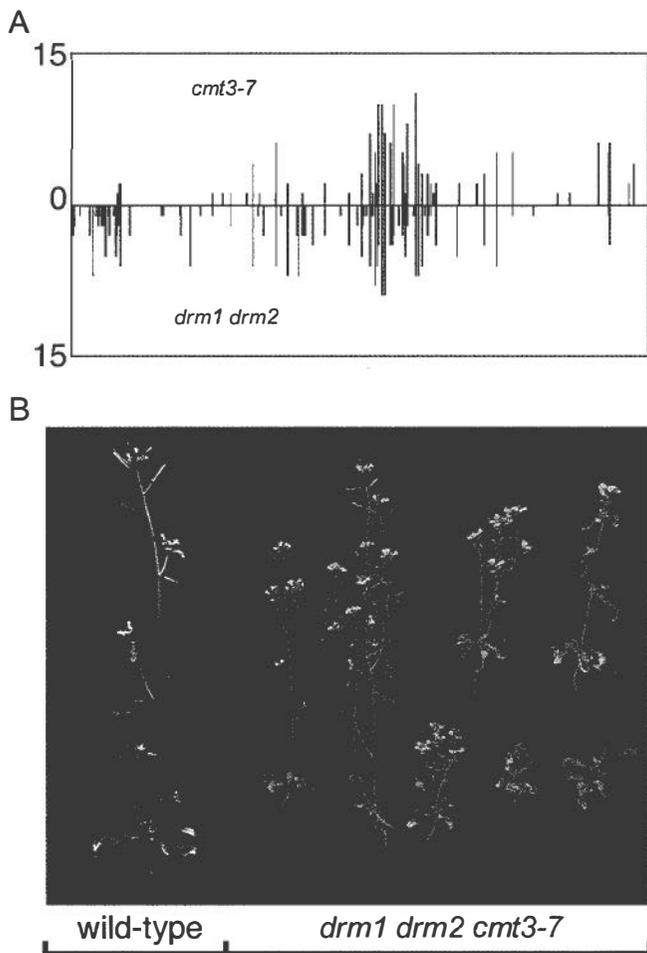
Fig. 4. Genetic redundancy of the *DRM* and *CMT3* genes. (*A*) Graph showing the level and positions of asymmetric methylation of *SUP* in the *cmt3-7* mutant or the *drm1 drm2* double mutant, both in the *clk-st* background. Height of the bars represents the frequency of methylation at each site of 15 clones analyzed by bisulfite sequencing. The *x* axis represents the 1,028-nt methylated region on the top strand of *SUP*. (*B*) Photographs of a four-week-old wild-type Ler plant (*Left*) and several *drm1 drm2 cmt3-7* triple mutant plants (*Right*).

Fig. 5. Model illustrating the complex relationship between DRM and CMT3 activities. (*A*) DRM and CMT3 both have the capacity to methylate asymmetric and CpNpG sites, but these activities are locus-specific, and therefore may be regulated by various factors as discussed in the text. (*B*) Summary of the activities of DRM and CMT3 inferred from the analysis of the types of methylation lost at each locus in the *drm1 drm2*, *cmt3-7*, or *drm1 drm2 cmt3-7* mutants. Thicker arrows denote more activity, which is inferred from the greater loss of methylation in a particular mutant.

specific chromatin modification is histone H3 Lys-9 methylation. Mutations in the *KYP* gene, which encodes a histone H3 Lys-9-specific methyltransferase, reduce CpNpG methylation (31). This observation suggests that CpNpG DNA methylation is at least partially controlled by histone H3 Lys-9 methylation, through an interaction of CMT3 with methylated chromatin. One possibility is that H3 Lys-9 methylation simply acts to more efficiently recruit CMT3 to particular loci, possibly through an interaction with a plant homolog of the H3 Lys-9 binding protein HP1 (31). A second possibility is that H3 Lys-9 methylation and/or HP1 binding acts to modulate CMT3's biochemical properties, stimulating its CpNpG methylating activity or modulating its preference for unmethylated or hemimethylated CpNpG sites.

At least two DNA sequence parameters can be imagined that could modulate methyltransferase function; overall sequence architecture and sequence composition. An example of the overall sequence architecture could be the presence of direct or inverted repeats, which are in fact found in all of the genes in this study. *FWA*, *MEA-ISR,* and the 180-bp centromeric repeat sequences all contain direct repeats that are immediately adjacent to one another. The *Ta3* retrotransposon contains long
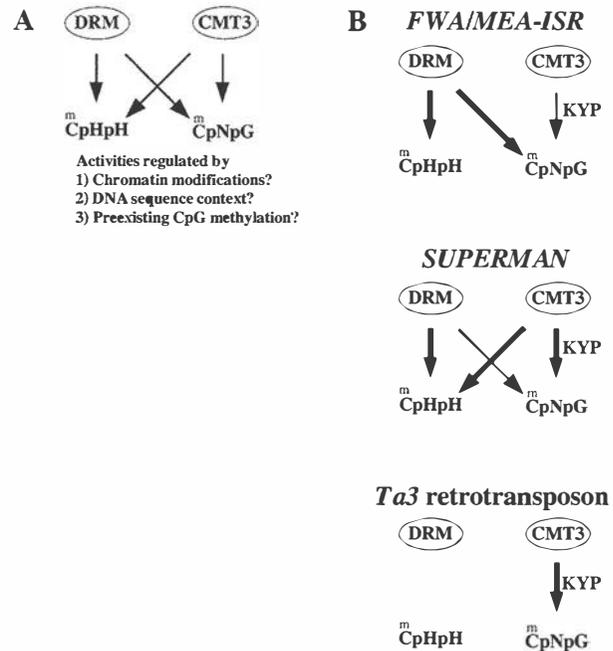
terminal repeat sequences at each end of the element, which are direct repeats separated by several kilobases. Finally, the *clk* allele used in this study, *clk-st*, contains an inverted repeat of the *SUP* gene in addition to the endogenous *SUP* locus (39). With respect to sequence composition, it is possible that the density of particular target sites is important. In particular, we note that *SUP*, a locus whose methylation and silencing is highly dependent on *CMT3* function, is rich in CpNpG sites but has few CpG sites. Within the methylated 1,028-bp region there are 9 CpGs and 27 CpNpG sites. Conversely, the *FWA* locus, whose methylation and silencing is mostly dependent on *MET1* function, is rich in CpG. The 500-bp methylated region of *FWA* contains 20 CpG sites (more than four times the density at *SUP*), and 14 CpNpG sites.

Another possibility to be considered is that CpG methylation could target non-CpG methylation. The results of this and other studies suggest that MET1 is specific for CpG methylation, and yet *met1* mutants show a strong reduction in non-CpG methylation at most loci examined (20, 22–25). The strongest evidence that MET1 does not directly methylate non-CpG sites, is that all traces of non-CpG methylation are lost in *drm1 drm2 cmt3-7* triple mutants, which contain a wild-type *MET1* gene. In addition, at the *SUP* locus *met1* mutants cause a large reduction of CpG methylation without major losses of CpNpG and asymmetric methylation (24, 25). These results suggest that MET1 is specific for CpG methylation, and that the losses of CpNpG and asymmetric methylation in *met1* mutants are indirect. One interpretation of these findings is that preexisting CpG methylation generates signals for the further methylation of cytosines in other contexts. An example is *FWA*, which is methylated primarily at CpG sites (89%), and less so at CpNpG sites (18%) and asymmetric sites (4%). The *DRM* loci are mainly responsible

Cao and Jacobsen

for maintaining this non-CpG methylation, because it is lost in the *drm1 drm2* double mutants. However, in *fwa* mutants induced in a *met1* background, all traces of both CpG and non-CpG methylation are lost (24, 40). Thus, DRM may methylate CpNpG and asymmetric sites at *FWA*, either because of the preexisting CpG methylation itself, or because of other chromatin modifications that result from the presence of CpG methylation. A second possibility is that the loss of CpG methylation of *FWA* is sufficient to reactivate gene expression, and that transcription produces a chromatin structure that is repressive for non-CpG methylation.

An interesting aspect of this study is that CMT3 appears to methylate asymmetric cytosines at the *SUP* locus, but not at the other tested loci. One possible explanation is that the inverted repeat structure of the *SUP* locus present in the transgenic *clk-st* line activates CMT3's ability to methylate asymmetric sites. However, we found residual asymmetric methylation in *drm1 drm2 clk-3* triple mutant plants (ref. 39 and data not shown), suggesting that CMT3 can methylate asymmetric sites on the single *SUP* endogene present in this strain. Therefore, it seems that some aspect of the chromatin or DNA structure of *SUP* activates CMT3's asymmetric methylation activity. We previously found that the most densely methylated region of *SUP* contains a pyrimidine-rich sequence, which is predicted to form a small hairpin (26). Similar sequences were also found in both hypermethylated regions of the agamous locus. It seems possible that the unusual DNA structure of the *SUP* hairpin, or the potential of this hairpin to be transcribed into a small RNA, activates dense non-CpG methylation of *SUP*. One might predict that the multiple copy *PAI* loci (47) would show a relationship with DNA methyltransferases similar to *SUP*, with asymmetric methylation depending on both DRM and CMT3 activities. This is because previous observations of the methylation patterns of the *PAI* genes showed high levels of asymmetric methylation that were significantly but not completely reduced in *cmt3* mutants (30).

In summary, several chromatin or DNA sequence parameters may regulate non-CpG methylation, either alone in combination, and further insight into this problem may only come by studying additional genetic modifiers of non-CpG methylation.

Asymmetric Methylation Sequence Specificity. An analysis of the sequence context of the asymmetric sites methylated by DRM and CMT3 suggests that both enzymes prefer to methylate CpA and CpT sites relative to CpC sites (Table 1). These results are consistent with those of other studies of asymmetric methylation in plants, where CpA and CpT have been found in more abundance than CpC methylation (6, 17, 48). These results are also consistent with nearest neighbor analysis of plant DNA (5), and with studies of the *in vitro* substrate preference of purified plant methyltransferases (49, 50). Interestingly, mouse embryonic stem cells also contain a significant amount CpA and CpT methylation and lower amounts of CpC methylation, and indirect evidence suggests an important role for *Dnmt3* genes (the closest mammalian *DRM* homologs) in maintaining this methylation (2). For example, mouse embryonic stem cells mutant for *Dnmt1* still contain a significant level of CpA and CpT methylation, and the high level of expression of *Dnmt3a* implicates it as the methyltransferase responsible (2). Furthermore, Dnmt3a can methylate non-CpG sites *in vitro*, or when ectopically

expressed in *Drosophila* cells (2, 16) and in both cases CpA and CpT sites are methylated more frequently than CpC sites. Finally, the site $C^{me}C(A/T)GG$ is also found to be methylated in mammals (51–55). Thus a preference for the methylation of CpA and CpT sites may be a conserved property of asymmetric methyltransferases.

De Novo Methylation vs. the Maintenance of Non-CpG Methylation. Previous characterization of the *drm* mutants suggested that the DRMs encode the major *de novo* methyltransferases in *Arabidopsis* (39). For instance, *drm1 drm2* double mutants completely blocked *de novo* CpG methylation and transgene silencing of *FWA*, which normally occurs when *FWA* is transformed into plants. Furthermore, *drm1 drm2* double mutants blocked *de novo* CpNpG and asymmetric methylation and gene silencing of the endogenous *SUP* locus, which normally occurs in the presence of a *SUP* inverted repeat transgene locus. These experiments suggest that the *DRM* genes encode enzymes capable of methylating previously unmodified DNA, and that *CMT3* cannot substitute for this function. However, in this study we show that, with respect to genes whose silencing is already established, the *DRM* and *CMT3* loci act redundantly to maintain the proper patterns of asymmetric and CpNpG methylation. Thus, a distinction must be made between the initial methylation and silencing of a gene (normally termed *de novo* methylation) and the maintenance of overall patterns of non-CpG methylation once a gene has been initially methylated. DRM appears to participate in both of these processes, and CMT3 only in the latter.

Further, although the *DRM* and *CMT3* loci are functionally redundant in terms of maintenance of overall patterns of non-CpG methylation, it is possible, particularly in the case of CpNpG methylation that the two act by different mechanisms. For instance, it is possible that DRM2 is more efficient at methylating completely unmodified CpNpG sites, whereas CMT3 may show a preference for hemimethylated CpNpG sites.

The Biological Function of Non-CpG Methylation. We found that although neither the *drm1 drm2* double mutants nor the *cmt3-7* single mutant showed morphological differences from the wild type, *drm1 drm2 cmt3-7* plants showed pleiotropic developmental defects. These *drm1 drm2 cmt3-7* plants retained CpG methylation at all of the sequences tested (Figs. 1–3), suggesting that the pleiotropic phenotype is caused by a reduction of non-CpG methylation, which is important for some endogenous plant process(es). In plants, high levels of non-CpG methylation are associated with the RNA-directed DNA methylation that is observed during transcriptional and posttranscriptional gene silencing, and this methylation may contribute to the silencing of viral and transposon sequences (15, 56–62). Thus, one possibility is that the pleiotropic phenotype of *drm1 drm2 cmt3-7* plants is caused by defects in genome defense/gene silencing processes. Further study of these triple mutant plants may help to reveal possible biological functions of non-CpG methylation.

1. Richards, E. J. & Elgin, S. C. (2002) *Cell* **108,** 489–500.
2. Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. & Jaenisch, R. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 5237–5242.
3. Selker, E. U., Fritz, D. Y. & Singer, M. J. (1993) *Science* **262,** 1724–1728.
4. Goyon, C., Nogueira, T. I. & Faugeron, G. (1994) *J. Mol. Biol.* **240,** 42–51.
5. Gruenbaum, Y., Naveh-Many, T., Cedar, H. & Razin, A. (1981) *Nature (London)* **292,** 860–862.
6. Meyer, P., Niedenhof, I. & ten Lohuis, M. (1994) *EMBO J.* **13,** 2084–2088.
7. Ingelbrecht, I., Van Houdt, H., Van Montagu, M. & Depicker, A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10502–10526.
8. Jacobsen, S. E. & Meyerowitz, E. M. (1997) *Science* **277,** 1100–1103.
9. Riggs, A. D. (1975) *Cytogenet. Cell Genet.* **14,** 9–25.
10. Holliday, R. & Pugh, J. E. (1975) *Science* **187,** 226–232.
11. Bird, A. P. (1978) *J. Mol. Biol.* **118,** 49–60.
12. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. (1988) *J. Mol. Biol.* **203,** 971–983.

13. Bestor, T. H. (1992) *EMBO J.* **11,** 2611–2617.
14. Leonhardt, H., Page, A. W., Weier, H. U. & Bestor, T. H. (1992) *Cell* **71,** 865–873.
15. Pelissier, T., Thalmeir, S., Kempe, D., Sanger, H. L. & Wassenegger, M. (1999) *Nucleic Acids Res.* **27,** 1625–1634.
16. Gowher, H. & Jeltsch, A. (2001) *J. Mol. Biol.* **309,** 1201–1208.
17. Dieguez, M. J., Vaucheret, H., Paszkowski, J. & Mittelsten Scheid, O. (1998) *Mol. Gen. Genet.* **259,** 207–215.
18. Dieguez, M. J., Bellotto, M., Afsar, K., Mittelsten Scheid, O. & Paszkowski, J. (1997) *Mol. Gen. Genet.* **253,** 581–588.
19. Finnegan, E. J. & Kovac, K. A. (2000) *Plant Mol. Biol.* **43,** 189–201.
20. Finnegan, E. J. & Dennis, E. S. (1993) *Nucleic Acids Res.* **21,** 2383–2388.
21. Vongs, A., Kakutani, T., Martienssen, R. A. & Richards, E. J. (1993) *Science* **260,** 1926–1928.
22. Finnegan, E. J., Peacock, W. J. & Dennis, E. S. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 8449–8454.
23. Ronemus, M. J., Galbiati, M., Ticknor, C., Chen, J. & Dellaporta, S. L. (1996) *Science* **273,** 654–657.
24. Lindroth, A. M., Cao, X., Jackson, J. P., Zilberman, D., McCallum, C. M., Henikoff, S. & Jacobsen, S. E. (2001) *Science* **292,** 2077–2080.
25. Kishimoto, N., Sakai, H., Jackson, J., Jacobsen, S. E., Meyerowitz, E. M., Dennis, E. S. & Finnegan, E. J. (2001) *Plant Mol. Biol.* **46,** 171–183.
26. Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X. & Meyerowitz, E. M. (2000) *Curr. Biol.* **10,** 179–186.
27. Baylin, S. B. & Herman, J. G. (2000) *Trends Genet.* **16,** 168–174.
28. Henikoff, S. & Comai, L. (1998) *Genetics* **149,** 307–318.
29. McCallum, C. M., Comai, L., Greene, E. A. & Henikoff, S. (2000) *Nat. Biotechnol.* **18,** 455–457.
30. Bartee, L., Malagnac, F. & Bender, J. (2001) *Genes Dev.* **15,** 1753–1758.
31. Jackson, J. P., Lindroth, A. M., Cao, X. & Jacobsen, S. E. (2002) *Nature (London)* **416,** 556–560.
32. Cao, X., Springer, N. M., Muszynski, M. G., Phillips, R. L., Kaeppler, S. & Jacobsen, S. E. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 4979–4984.
33. Okano, M., Xie, S. & Li, E. (1998) *Nat. Genet.* **19,** 219–220.
34. Hsieh, C. L. (1999) *Mol. Cell. Biol.* **19,** 8211–8218.
35. Okano, M., Bell, D. W., Haber, D. A. & Li, E. (1999) *Cell* **99,** 247–257.
36. Lyko, F., Ramsahoye, B. H., Kashevsky, H., Tudor, M., Mastrangelo, M. A., Orr-Weaver, T. L. & Jaenisch, R. (1999) *Nat. Genet.* **23,** 363–366.
37. Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X. Y., Russo, J. J. & Viegas-Pequignot, E. (1999) *Nature (London)* **402,** 187–191.
38. Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. & Gartler, S. M. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 14412–14417.
39. Cao, X. & Jacobsen, S. E. (2002) *Curr. Biol.* **12,** 1138–1144.
40. Soppe, W. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M. & Peeters, A. J. (2000) *Mol. Cell* **6,** 791–802.
41. Kakutani, T., Munakata, K., Richards, E. J. & Hirochika, H. (1999) *Genetics* **151,** 831–838.
42. Kakutani, T., Jeddeloh, J. A., Flowers, S. K., Munakata, K. & Richards, E. J. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12406–12411.
43. Grossniklaus, U., Vielle-Calzada, J. P., Hoeppner, M. A. & Gagliano, W. B. (1998) *Science* **280,** 446–450.
44. Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J. J., Goldberg, R. B. & Fischer, R. L. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 4186–4191.
45. McClelland, M., Nelson, M. & Raschke, E. (1994) *Nucleic Acids Res.* **22,** 3640–3659.
46. Bartee, L. & Bender, J. (2001) *Nucleic Acids Res.* **29,** 2127–2134.
47. Bender, J. & Fink, G. R. (1995) *Cell* **83,** 725–734.
48. Luff, B., Pawlowski, L. & Bender, J. (1999) *Mol. Cell* **3,** 505–511.
49. Theiss, G., Schleicher, R., Schimpff-Weiland, G. & Follmann, H. (1987) *Eur. J. Biochem.* **167,** 89–96.
50. Adams, R. L., Lindsay, H., Reale, A., Seivwright, C., Kass, S., Cummings, M. & Houlston, C. (1993) *EXS* **64,** 120–144.
51. Clark, S. J., Harrison, J. & Frommer, M. (1995) *Nat. Genet.* **10,** 20–27.
52. Franchina, M. & Kay, P. H. (2000) *DNA Cell Biol.* **19,** 521–526.
53. Lorincz, M. C., Schubeler, D., Goeke, S. C., Walters, M., Groudine, M. & Martin, D. I. (2000) *Mol. Cell. Biol.* **20,** 842–850.
54. Malone, C. S., Miner, M. D., Doerr, J. R., Jackson, J. P., Jacobsen, S. E., Wall, R. & Teitell, M. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10404–10409.
55. Lorincz, M. C. & Groudine, M. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10034–10036.
56. Wassenegger, M., Heimes, S., Riedel, L. & Sanger, H. L. (1994) *Cell* **76,** 567–576.
57. Jones, A. L., Thomas, C. L. & Maule, A. J. (1998) *EMBO J.* **17,** 6385–6393.
58. Mette, M. F., van der Winden, J., Matzke, M. A. & Matzke, A. J. (1999) *EMBO J.* **18,** 241–248.
59. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. (2000) *EMBO J.* **19,** 5194–5201.
60. Wassenegger, M. (2000) *Plant Mol. Biol.* **43,** 203–220.
61. Matzke, M., Matzke, A. J. M. & Kooter, J. M. (2001) *Science* **293,** 1080–1083.
62. Jones, L., Ratcliff, F. & Baulcombe, D. C. (2001) *Curr. Biol.* **11,** 747–757.

Colloquium

RNA-directed DNA methylation in *Arabidopsis*

Werner Aufsatz*, M. Florian Mette*, Johannes van der Winden, Antonius J. M. Matzke, and Marjori Matzke†

Institute of Molecular Biology, Austrian Academy of Sciences, Billrothstrasse 11, A-5020 Salzburg, Austria

In plants, double-stranded RNA that is processed to short RNAs ≈21–24 nt in length can trigger two types of epigenetic gene silencing. Posttranscriptional gene silencing, which is related to RNA interference in animals and quelling in fungi, involves targeted elimination of homologous mRNA in the cytoplasm. RNA-directed DNA methylation involves *de novo* methylation of almost all cytosine residues within a region of RNA–DNA sequence identity. RNA-directed DNA methylation is presumed to be responsible for the methylation observed in protein coding regions of posttranscriptionally silenced genes. Moreover, a type of transcriptional gene silencing and *de novo* methylation of homologous promoters in *trans* can occur if a double-stranded RNA contains promoter sequences. Although RNA-directed DNA methylation has been described so far only in plants, there is increasing evidence that RNA can also target genome modifications in other organisms. To understand how RNA directs methylation to identical DNA sequences and how changes in chromatin configuration contribute to initiating or maintaining DNA methylation induced by RNA, a promoter double-stranded RNA-mediated transcriptional gene silencing system has been established in *Arabidopsis*. A genetic analysis of this system is helping to unravel the relationships among RNA signals, DNA methylation, and chromatin structure.

The term "RNA silencing" refers to epigenetic gene silencing effects that are initiated by double-stranded RNA (dsRNA) (1). Discovered independently in plants, fungi, and animals, RNA silencing phenomena are revealing new ways to repress gene expression and to subdue transposable elements and viruses that produce dsRNA during their replication cycle (2–8). A fundamental step in RNA silencing pathways is cleavage of dsRNA into short RNAs (9), which are believed to act as guides for enzyme complexes that either degrade or modify homologous nucleic acids.

The most familiar type of RNA silencing occurs primarily in the cytoplasm and is termed posttranscriptional gene silencing (PTGS) in plants, quelling in *Neurospora*, and RNA interference (RNAi) in animals. PTGS/RNAi involves a dsRNA that is processed by an RNase III-like enzyme called Dicer into short interfering (si) RNAs 21–22 nt in length. The antisense siRNAs associate with a ribonuclease complex and guide sequence-specific degradation of complementary mRNAs (5–8).

A second form of RNA silencing involves sequence-specific changes at the genome level. RNA-directed DNA methylation (RdDM) (10), which has been described so far only in plants, leads to *de novo* methylation of almost all cytosine residues within the region of sequence identity between the triggering RNA and the target DNA. Similarly to PTGS/RNAi, RdDM requires a dsRNA that is cleaved to short RNAs ≈21–24 nt in length (11). It is not yet certain whether the short RNAs or dsRNA guide methylation of homologous DNA sequences, although the length of short RNAs is consistent with the minimum DNA target size of RdDM (≈30 bp) (12).

RdDM is assumed to be the source of methylation observed in protein coding regions in many cases of PTGS, where it can contribute in an unknown way to the maintenance of silencing (13, 14). In addition, RdDM has been implicated in a type of transcriptional gene silencing (TGS) that is initiated by dsRNAs containing promoter sequences. Promoter dsRNAs that trigger TGS and RdDM of homologous promoters in *trans* can be produced in the nucleus by transcription through inverted repeats (IRs) of promoter sequences (11, 15) or in the cytoplasm by a replicating RNA virus that is engineered to contain sequences identical to the promoter of a nuclear gene (16, 17).

Although the phenomenon of RdDM is well established in plants, a number of questions remain. One concerns the identity of the DNA methyltransferases (MTases) that are required for establishing and maintaining the unusual pattern of methylation characteristic of RdDM. A second issue concerns the relationship between DNA methylation and changes in chromatin structure. Given the close links between DNA methylation, chromatin remodeling (18–20), and histone modifications, such as acetylation (21) and methylation (22, 23), it might be anticipated that alterations in chromatin structure would be required to initiate and/or retain methylation induced by RdDM.

To address these questions, we carried out a genetic analysis of a promoter dsRNA-mediated TGS system that we have established in *Arabidopsis*. In this paper, we describe this system and the impact of several mutations that impair DNA methylation and/or possible chromatin remodeling processes. We discuss whether RdDM might occur in animals and whether RNA might direct chromatin modifications in organisms that do not methylate their DNA.

Experimental Procedures

T-DNA Constructs and Plant Transformation. *Arabidopsis thaliana* plants (ecotype Columbia) were grown at 22°C in a 16 h/8 h day/night cycle. Transformation was done by the floral dip method (24). The nopaline synthase promoter (NOSpro) target construct [NOSpro–*NPTII* (neomycin phosphotransferase) NOSter–NOSpro–*NOS* (nopaline synthase) NOSter] was introduced into *Arabidopsis* by using *Agrobacterium* strain A208 harboring a disarmed Ti-plasmid (25). A line homozygous for the target locus was retransformed by using *Agrobacterium* harboring a binary vector with a NOSpro IR (in which the halves were separated by ≈250 bp of the α' promoter of soybean β-conglycinin; ref. 26) driven by the 35S promoter of cauliflower mosaic virus (35Spro), a pUC18 plasmid vector, and a 19Spro-*HPT* (hygromycin phosphotransferase) gene as selectable marker (11). The 35Spro was flanked by *lox* sites in direct orientation to allow removal by Cre recombinase. Selection of transgenic target and silencer plants was done as described (11).
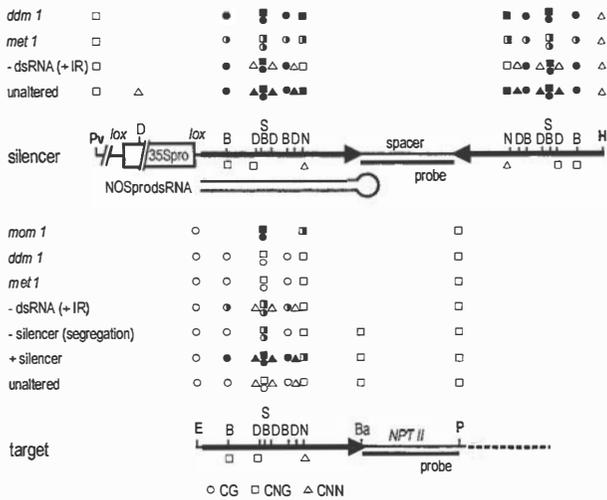
Fig. 1. Structures and methylation analysis of the silencer NOSpro IR (*Upper*) and target NOSpro–*NPTII* gene (*Lower*). NOSpro sequences are depicted as heavy black arrows. Enzymes and probes used for DNA blot analyses are indicated. Abbreviations: Pv, *Pvu*II; D, *Dde*I; S, *Sac*II; N, *Nhe*I; Hi, *Hind*III; E, *Eco*RI; B, *Bst*UI; Ba, *Bam*HI; P, *Pst*I. To assess methylation in the target NOSpro, an E and P double digest was performed (the minus lanes in Figs. 4 *A*, *C*, *E*, and *G*) and one of several methylation-sensitive enzymes (B, D, S, N, Ba) was added. Methylation in the NOSpro IR was tested by digesting with Pv and Hi (the minus lanes in Figs. 4 *B* and 6 *B*, *D*, *F*, and *H*), together with either B, D, S, or N. Filled, half-filled, and open circles, squares, and triangles (CG, CNG, and CNN, respectively) indicate >90%, ≈50%, and <10% cytosine (C) methylation, respectively. Open squares or triangles below each map for the enzymes D, N, and B indicate that the top and bottom DNA strands contain C residues in different sequence contexts (e.g., CG and CNG, or CNG and CNN). The *Nhe*I site (underlined) is in the sequence context: 5′-CAGCTACGᵐCAA-3′ (top); and 3′-GTᵐCGATCGTT-5′ (bottom).

Cre/*lox*-Mediated Deletion of the 35Spro.

To delete the 35Spro by Cre recombinase, plants doubly homozygous for target and silencer were supertransformed with a third T-DNA construct expressing Cre recombinase from the 35Spro (27). The Cre construct encodes glufosinate resistance (BAR), which allows the herbicide BASTA to be used for selection directly on soil. Soil-grown triple transformants were selected by spraying with BASTA (Celaflor, Hoechst, Vienna, Austria; 400 mg/liter ammonium glufosinate) twice a week for 2 weeks. T2 seeds from BASTA-resistant T1 plants were plated on MS agar containing 40 mg/liter kanamycin (Kan) (Sigma), and/or 40 mg/liter Kan plus 20 mg/liter hygromycin B (Hyg) (Calbiochem). Resistant T2 seedlings were genotyped by PCR using BAR primers to confirm the absence of the Cre construct because the presence of Cre-recombinase interferes with Southern analysis of DNA-fragments containing *lox* sites (M.F.M., unpublished observations). Genotype-PCR-grade DNA was isolated from *Arabidopsis* leaves as described (28). BAR primers were 5′-CGAGACAAGCACGGTCAACTTC-3′ and 5′-ACCCACGTCATGCCAGTTCC-3′. BAR-negative plants were allowed to set seeds, and DNA was extracted from T3 progeny plants by using a DNAeasy plant maxi kit (Qiagen, Hilden, Germany). The DNA was subjected to restriction digests and Southern hybridization as described in the Fig. 1 and Fig. 6 legends. Data in Fig. 3*C* and Fig. 6 were obtained for plants 2 generations after Cre-mediated removal of the 35Spro.

Mutant Crosses.

The following *Arabidopsis* mutants were used in this study: *ddm1-5/som8* (decrease in DNA methylation/somniferous) (29) in ecotype Zürich; *mom1* (Morpheus' molecule) (30) in ecotype Zürich; and *met1/ddm2–1* (DNA methyltransferase 1) (E. Richards, personal communication) in ecotype La-er. Because the strength of NOSpro silencing varied some-

what in different ecotypes (La-er > Col-0 > Zürich), control crosses of double homozygous target/silencer plants with the respective wild-type backgrounds were always performed. Reactivation of the NOSpro–*NPTII* target gene in a given mutant was assessed by survival of seedlings in the F₂ generation and/or advanced generations on Murashige and Skoog (MS) medium containing 40 mg/liter Kan alone or 40 mg/liter Kan plus 20 mg/liter Hyg. The *mom1* and *met1* mutations are not genetically linked to either the target locus or the silencing locus, and could be tested for their effects on NOSpro–*NPTII* silencing immediately in the F2 generation; *ddm1/som8* is linked to the target locus and had to be introgressed into the NOSpro target/silencer line as described below.

ddm1-5/som8 is a fast neutron-generated allele that is distinguishable from the wild-type gene by an 82 bp insertion into the second exon (18). The *ddm1-5/som8* mutation was separated from transgene locus A (which harbors 35S-*HPT* genes) (29) by two outcrosses to ecotype Col-0. Plants homozygous for *ddm1-5/som8* and lacking locus A were screened out from selfed progeny of the second outcross. Genotyping reactions were done by using *som* primers: 5′-AAGCGACGGAGACGACT-GTTTG-3′ and 5′-TTTCACAAAGCAACCACACTACG-3′. 35S-HPT primers were 5′-CCCACTATCCTTCGCAAGA-3′and 5′-CGTCTGCTGCTCCATACAAGC-3′. Because the *DDM1* gene is linked to the target transgene locus (the physical distance is ≈270 kb), the *ddm1-5* mutation was introgressed into the genetic background of the NOSpro target/silencer transgenic plant. A *ddm1-5/ddm1-5* plant lacking transgene locus A was crossed with a plant doubly homozygous for the NOSpro target/silencer. An F₂ plant homozygous for the NOSpro silencing locus, hemizygous for the NOSpro target locus and heterozygous for the *ddm1-5* allele was selected. This plant was allowed to self-pollinate, and 193 F₃ seedlings were analyzed for their *DDM1* genotype with the *som* primers described above. Forty-eight F₃ seedlings were homozygous for the *ddm1-5* allele and were further subjected to PCR-genotyping for the presence of target NOSpro–*NPTII* sequences by using NOSpro–*NPTII* primers: 5′-GAGAATTAAGGGAGTCACG-3′and 5′-TCGTCCT-GCAGTTCATTC-3′. Two of the 48 *ddm1-5/ddm1-5* plants were hemizygous for the target locus indicating a recombination event between the *ddm1-5* allele and the NOSpro target locus during meiosis of the parental F₂ plant. Progeny seeds from this genotype (i.e., the second generation of homozygosity for *ddm1-5*; the target is still segregating) were analyzed for reactivation of the NOSpro–*NPTII* gene on medium containing Kan (*n* = 308 seeds) or Kan and Hyg (*n* = 317). Only weak reactivation was observed in a few progeny. Eventually, in progeny that were third generation homozygous for the *ddm1-5* allele (and homozygous for the target and silencing loci), resistance generally improved and stronger Kan resistant (Kanᴿ) seedlings appeared stochastically in populations of seedlings that also contained moderately and weakly Kanᴿ members (Fig. 3*D*).

Plants homozygous for the *mom1* allele without transgene locus A were crossed with a doubly homozygous NOSpro target/silencer plant. F₂ seeds were plated on medium containing Kan, Hyg, or both to test for an immediate effect of the *mom1* mutant allele. No immediate effect was observed. Among the F₃ progeny of this cross, a triple homozygous line for the NOSpro target, NOSpro silencer, and the *mom1* allele was selected. The genotype for the NOSpro silencer was determined by selecting seeds on Hyg-containing medium. The genotype with respect to the *mom1* allele, which is tagged with a T-DNA conferring BASTA resistance, was determined by spraying seedlings with the herbicide as described above. To assess the genotype for the NOSpro target, PCR genotyping with NOSpro–*NPTII* primers (see above) was performed. The progeny F₄ seeds of the triple homozygous plant (i.e., the third generation of homozygosity for

mom1) were plated on medium containing Kan ($n = 440$) or Kan and Hyg ($n = 441$). No resistant seedlings were observed.

met1/ddm2–1 was obtained as progeny from a heterozygous plant and homozygous plants were screened out by using the demethylation assay of centromeric and ribosomal DNA (rDNA) repeats (ref. 31; E. Richards, personal communication). A homozygous *met1* plant was crossed to plants homozygous for both the NOSpro target locus and the NOSpro silencing locus. Resulting F_1 plants were allowed to self-pollinate, and the F_2 seeds were plated on medium containing Kan, Hyg, or both antibiotics. The F_2 seedlings showed weak to moderate resistance on Kan and Kan–Hyg, consistent with a partial release of silencing. Double resistant seedlings were recovered on soil, and *met1* homozygosity was confirmed by using cleaved amplified polymorphic sequence markers (CAPS) markers as described (32). F_4 progeny lines that are triple homozygous for *met1* (i.e., third generation of *met1* homozygosity), the target locus and the silencing locus were selected for further analysis. These seedlings display weak to strong Kan^R (Fig. 3D).

Methylation Analysis. Plant genomic DNA was extracted as described (11). For the DNA methylation analyses in *met1*, *ddm1*, and *mom1* mutants, DNA was isolated from plants that had been homozygous for a given mutation for two generations (i.e., F_3 generation for *mom1* and *met1*; F_4 generation for *ddm1*). Methylation in plants containing the Cre-altered, nontranscribed silencing locus were performed after the Cre-expressing locus had been segregated out, eliminating possible background transcription of the Cre locus. Restriction digests were done according to the instructions of the manufacturers. DNA blot analysis using ^{32}P-labeled RNA probes has been described (33). As probes, subcloned 0.19 kb *NPTII* and 0.25 kb α' pro fragments were transcribed *in vitro*. Methylation analysis of centromeric repeats and rDNA repeats was conducted as described by others (32). Bisulfite sequencing was performed as described (12, 34). The following degenerate primers, which allowed for unconverted or converted cytosines, were used to amplify the top strand of the NOS-NPTII target: 5′ primer 5′-YATGAGYG-GAGAATTAAGGGAGT-3′ (Y = C or T); 3′ primer 5′-CCRAATARCCTCTCCACCCAA-3′ (R = G or A).

Cloning of Transgene Inserts. Genomic λ clones from target and silencer transgenes were obtained and sequenced as described (35). The silencer transgene complex comprises a single copy of the T-DNA construct with a complete 35Spro–NOSpro IR that was integrated in chromosome 4, BAC clone F10M10 (GenBank accession no. AL035521) with the right border downstream of nucleotide 21681 and the left border upstream of nucleotide 21693. Between plant sequence and the right border, the filler sequence "TTTTT" was inserted. The target locus was originally screened out genetically as a single transgene locus that was readily inactivated when the silencing locus was introduced, and largely reactivated the first generation after segregating away from the silencing locus. More detailed structural analysis by λ cloning revealed that the target locus contains several complete and incomplete copies of the T-DNA construct flanked by *Arabidopsis* DNA from chromosome 5 on the left and chromosome 3 on the right. This finding suggests that a rearrangement of plant DNA, which occasionally occurs during T-DNA integration (36, 37), had occurred. All of the bands visualized in Southern blot analyses using an *NPTII* probe always cosegregated in multiple, independent crosses, which is consistent with a single, multicopy transgene locus. In the absence of the silencing locus, the target locus was stably expressed over multiple homozygous generations. The moderate structural complexity probably enhanced its susceptibility to silencing, as has been observed in a NOSpro silencing system in tobacco (38). Genetic analysis revealing linkage to *ddm1* indicated that the

actual T-DNA insertion site was on chromosome 5 (data not shown). Nucleotide 143 of the T-DNA right border region (GenBank accession no. J01826) was fused upstream of nucleotide 66153 of BAC clone F2103 (GenBank accession no. AC009853) on chromosome 3. Nucleotide 98 of the left T-DNA border region (GenBank accession no. J01825) was fused to sequences at the distal end of chromosome 5, BAC clone K9I9 (GenBank accession no. AB013390) upstream of nucleotide 4316. The rearrangement had no visible phenotype effects or impact on target NOSpro expression. Mitotic chromosome counts (39) revealed a normal diploid number of $2n = 10$.

RNA Analyses. Total RNA was extracted, electrophoresed, and blotted as described (40). Control hybridization of tobacco with an actin probe was carried out following published procedures (40). *Arabidopsis* RNA was hybridized with a eukaryotic protein synthesis initiation factor 4A (*eIF-4A*) fragment from *Arabidopsis* (41). Analysis of NOSpro dsRNA and small RNA was carried out as described (11). Transcriptional run-on analysis was performed as described in a former report (42).

Results

A promoter dsRNA-mediated *trans*-silencing system based on the NOSpro was originally established in tobacco (11) and has served as a model for setting up a similar system in *Arabidopsis*. A homozygous line that stably expresses a NOSpro–*NPTII* target gene encoding resistance to Kan (Fig. 1 *Lower*) was produced. The target line was then retransformed with a silencing construct, which contains a NOSpro IR under the control of the 35Spro (Fig. 1 *Upper*) together with a gene encoding resistance to Hyg driven by the 19S promoter of cauliflower mosaic virus (11). The silencing locus produces NOSpro dsRNA (Fig. 2A, *Arabidopsis* target + silencer) that is processed into short RNAs ≈21–24 nucleotides in length (Fig. 2C, *Arabidopsis* target + silencer), similar to those observed in tobacco transformed with the same construct (Fig. 2 A and C, tobacco target + silencer) (11).

In the presence of the silencing locus, the target NOSpro–*NPTII* gene is inactivated, as revealed by cultivation of seedlings on media containing different antibiotics. When self fertilized, a plant that is homozygous for the active NOSpro–*NPTII* target gene produces, as expected, 100% Kan^R progeny (Fig. 3A, target–Kan). When the silencing locus, which encodes NOSpro dsRNA and resistance to Hyg, is introduced into the target line and is present in the hemizygous condition, selfing produces 75% Hyg-resistant seedlings (Fig. 3A, target + silencer-Hyg). Even though the parent is homozygous for the target locus, however, only 25% of the seedlings are Kan^R (Fig. 3A, target + silencer-Kan). Any seedling that is Kan^R has not inherited the silencing locus, as indicated by the lack of double resistance (Fig. 3A, target + silencer, Kan/Hyg). Conversely, a seedling that has inherited the silencing locus is Kan-sensitive because of silencing of the NOSpro–*NPTII* target gene. Silencing of the target gene occurs at the transcriptional level as demonstrated by a nuclear run-on analysis (Fig. 2D).

Transcriptional silencing of the NOSpro–*NPTII* target gene is accompanied by *de novo* methylation of the target NOSpro. When active, the target gene is normally unmethylated in the NOSpro region, as indicated by nearly complete digestion with the methylation-sensitive restriction enzymes *Sac*II ($^mC^mCG^m$-CGG), *Bst*UI ($^mCG^mCG$), and *Nhe*I(GCTAG^mC) (Fig. 4A, unmeth. control; a superscript "m" indicates a methylated cytosine that can inhibit cleavage). In the presence of the silencing locus, the NOSpro region specifically becomes methylated in both symmetrical (CG and CNG) and nonsymmetrical (CNN) cytosines as demonstrated, respectively, by negligible digestion with *Sac*II and *Bst*UI, and approximately 50% digestion with *Nhe*I (Fig. 4A, target + silencer). This pattern of
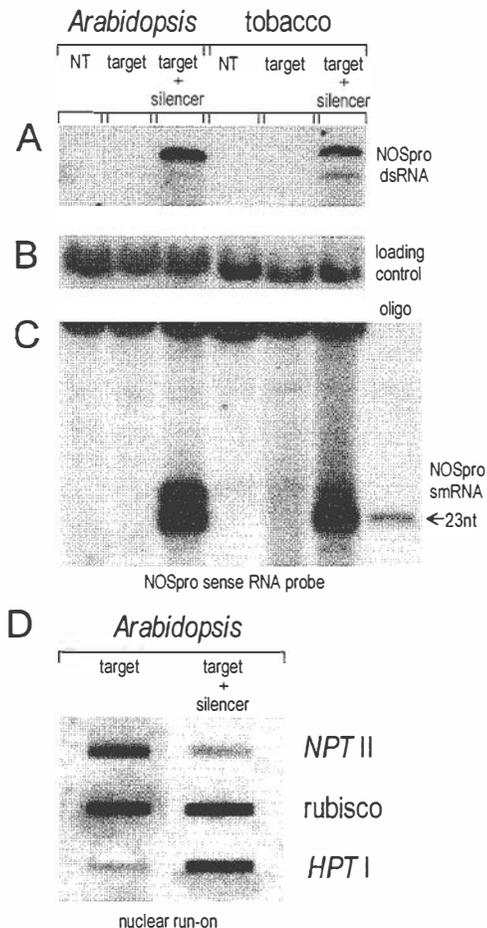
Fig. 2. RNA analysis. (A) RNase protection reveals the ≈0.3-kb NOSpro dsRNA transcribed from the silencer NOSpro IR. NT, normal untransformed plants. (B) Total RNA used in A probed with an actin probe from tobacco and an eIF-4A probe from *Arabidopsis* as loading controls. (C) Detection of NOSpro short RNAs (sense probe) produced by means of dsRNA cleavage. Identical results were obtained with an antisense probe. (D) Nuclear run-on analysis demonstrating transcriptional down-regulation of the NOSpro–*NPTII* target gene in the presence of the silencing locus, which encodes *HPT* and NOSpro dsRNA. A constitutively expressed ribulose 1,5-bisphosphate carboxylase (rubisco) gene was used as a control. Positive controls in A and C were prepared from tobacco plants transformed with the 35Spro–NOSproIR construct.

methylation, in which C residues in any sequence context are modified specifically in the region of RNA-DNA sequence identity, is characteristic of RdDM and was confirmed when bisulfite sequencing was used to examine methylation in more detail (Fig. 5).

Methylation of the target NOSpro–*NPTII* gene is largely eliminated when the target locus and silencing locus segregate in progeny, as indicated by nearly complete digestion with *Bst*UI and *Nhe*I, and about 50% cleavage with *Sac*II (Fig. 4A, target minus silencer). The remaining methylation at the *Sac*II site, which is correlated with the mottled phenotype of many Kan[R] seedlings (Fig. 3B), is presumably caused by maintenance of some CG and/or CNG methylation through meiosis (42).

A requirement for NOSpro dsRNA in silencing and methylation of the target NOSpro–*NPTII* gene was demonstrated by removing the transcribing 35Spro, which is flanked by *lox* sites (Fig. 1 *Upper*), with Cre recombinase. The removal of the 35Spro and retention of the NOSpro IR at the Cre-modified "silencing" locus was monitored by a shift to a smaller band of the expected
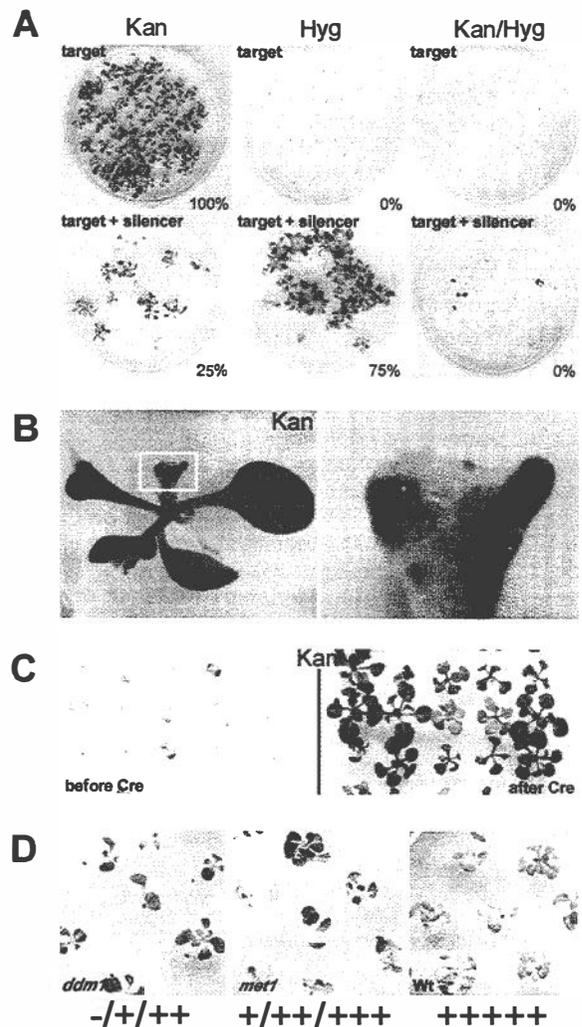

Fig. 3. Phenotypic analysis of silencing. NOSpro–*NPTII* target gene expression is assayed by Kan[R]; the silencing locus by Hyg[R]. (A) Selfing a plant homozygous for an active target gene produces 100% Kan[R] progeny. Selfing a plant homozygous for the target locus and hemizygous for the silencing locus, revealed by 75% Hyg[R], produces only 25% Kan[R] progeny. Kan[R] seedlings lack the silencer, indicated by 0% (Kan/Hyg)[R]. (B) Mottled Kan[R] seedling in the first generation after crossing out the silencing locus (*Right* is an enlargement of the boxed region in *Left*; white and green patches represent Kan[S] and Kan[R] regions, respectively). (C Left) Kan[S] seedlings before removing the 35Spro with Cre recombinase. (*Right*) Kan[R] seedlings two generations after removing the 35Spro. (D) Ranges of phenotypes on Kan-containing medium (plus signs, different degrees of Kan[R]; minus sign, Kan[S]) in seedlings after three generations of homozygosity for the *ddm1* and *met1* mutations, based on 5 plus signs for wild-type levels of Kan[R] in seedlings containing the target locus in the unsilenced state.

size with all enzymes tested in a DNA blot analysis (Fig. 6 *B*, *D*, *F*, and *H*; compare minus lanes in unaltered silencer panels with minus lanes in Cre-altered silencer panels) and confirmed by cloning and sequencing the Cre-modified silencing locus (data not shown).

After deletion of the 35Spro from the silencing locus, NOSpro short RNAs are no longer detectable, even after long exposures of the respective Northern blots (data not shown). Consequently, the target NOSpro–*NPTII* gene is active in the presence of the nontranscribed NOSpro IR, as indicated by the Kan[R] phenotype of seedlings that are doubly homozygous for the target locus and Cre-modified silencing locus (Fig. 3C, after Cre). Furthermore, methylation of the NOSpro–*NPTII* target gene is reduced ap-
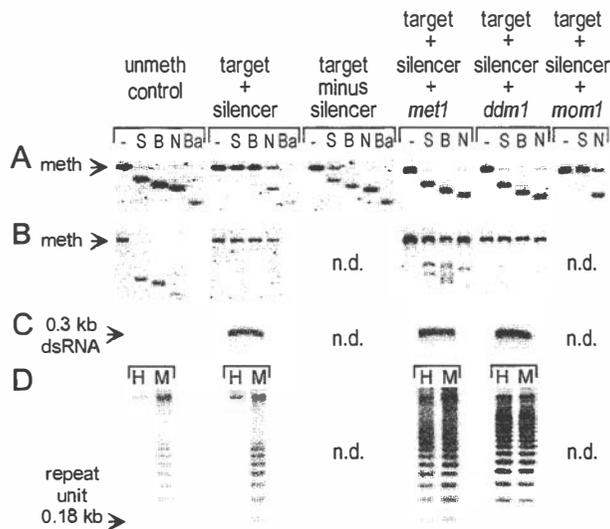
Fig. 4. Methylation analysis. (*A*) Target NOSpro. (*B*) Silencer NOSpro IR. (*C*) NOSpro dsRNA. (*D*) Centromeric repeats. Methylation of the target and silencing loci were analyzed by using the enzymes and probes described in Fig. 1. For *met1*, *ddm1*, and *mom1* mutants, methylation was analyzed by using DNA isolated from plants that had been homozygous for the respective mutation for two generations. Methylation of centromeric repeats was analyzed by using *Hpa*II (H: $^mC^mCGG$) and *Msp*I (M: ^mCCGG). The unmethylated control for the silencer NOSpro IR consisted of a λ genomic clone containing the silencing locus. Shifts to the smaller fragments indicate no methylation at a particular site. Arrows in *A* and *B* indicate position of methylated fragment; in *C* and *D*, arrows represent the sizes of the indicated features. n.d., not determined.

Fig. 5. Bisulfite sequencing. The ≈300-bp NOSpro sequence is shown with the region of identity to NOSpro dsRNA underlined. Methylation (filled symbols) in 10 cloned PCR fragments from the upper DNA strand is indicated. Symbols are described in the Fig. 1 legend. The positions of restriction enzyme sites used in the DNA blot analyses are indicated (abbreviations are given in the Fig. 1 legend). The four boxed regions represent transcriptional regulatory elements (61), which contain short IRs (arrows). The transcription start site is indicated by the bent arrow at −1. The sequence of the primers used is indicated. Methylation does not infiltrate significantly into *NPTII* coding sequences.

proximately 30% at symmetrical Cs in the *Sac*II (Fig. 6*A*, target + Cre-altered silencer) and *Bst*UI (Fig. 6*C*, target + Cre-altered silencer) sites and almost completely at nonsymmetrical C residues in the *Nhe*I site (Fig. 6*E*, target + Cre-altered silencer) and *Dde*I sites (^mCTNAG) (Fig. 6*G*, target + Cre-altered silencer).

The NOSpro dsRNA not only triggers methylation and silencing of the target NOSpro in *trans*, it also contributes to methylation in *cis* of the NOSpro copies in the IR at the silencing locus. This was demonstrated by examining methylation of the NOSpro IR before and after removing the transcribing 35Spro with Cre recombinase. The transcribed NOSpro IR at the unaltered silencing locus is heavily methylated at both symmetrical and nonsymmetrical Cs within the repeated region as indicated, respectively, by lack of digestion with *Sac*II and *Bst*UI (Fig. 6 *B* and *D*, unaltered silencer panels), and *Nhe*I and *Dde*I (Fig. 6 *F* and *H*, unaltered silencer panels). In contrast, the nontranscribed NOSpro IR at the Cre-altered silencing locus loses methylation at nonsymmetrical C residues, which was revealed by substantial digestion with *Nhe*I and *Dde*I (Fig. 6 *F* and *H*; compare unaltered silencer panel with Cre-altered silencer panels). At the same time, methylation at symmetrical C residues is almost completely retained, as indicated by poor digestion with *Sac*II and *Bst*UI (Fig. 6 *B* and *D*; compare unaltered silencer panel with Cre-altered silencer panels).

The effects of several mutations that release TGS in other systems (14, 17, 29, 30, 32, 43, 44) were tested on the NOSpro dsRNA-mediated TGS system. For these experiments, the double homozygous target/silencer line was crossed with lines homozygous for the following recessive mutations: the *som8* allele (29) of *ddm1*, which encodes a putative component of a SWI/SNF2 chromatin remodelling complex (18); *met1* (*ddm2*), which encodes a DNA MTase (E. Richards, personal communication) that maintains methylation in CG dinucleotides (45);
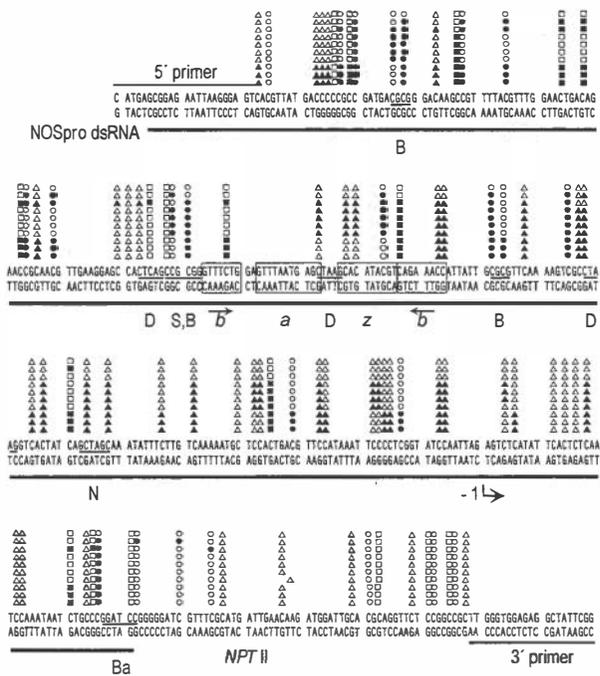
and *mom1*, which encodes a possible chromatin remodeling protein (30). F_1 progeny obtained from these crosses were selfed and the extent of silencing evaluated in the F_2, F_3, and F_4 generations. If a given mutation has no effect (and assuming no linkage between a mutation and the target locus or silencing locus), the percentages of antibiotic resistance in F_2 progeny should be 19% Kan^R, 75% Hyg^R, and 0% (Kan–Hyg)^R. If a mutation releases silencing and is fully penetrant, these percentages would change to 33% Kan^R, 75% Hyg^R, and 14% (Kan–Hyg)^R. In other words, impaired silencing would be indicated by an increase in the percentage of Kan^R F_2 progeny and by the appearance of some F_2 progeny that display double resistance.

The *mom1* mutation is the only one of the three tested that did not visibly affect NOSpro silencing, as indicated by no recovery of (Kan–Hyg)^R progeny, even in F_3 and F_4 generations (data not shown). Methylation of the target NOSpro is also not reduced in *mom1* mutants, as demonstrated by levels of methylation at the *Sac*II and *Nhe*I sites that approximate those in the silenced state (Fig. 4*A*, compare target + silencer + *mom1* with target + silencer).

The *met1* mutation partially released silencing of the NOSpro–*NPTII* gene in F_2 progeny, as indicated by an increase in the percentage of Kan^R seedlings (29%, $n = 428$) and weak resistance of some of these seedlings on medium containing both Kan and Hyg (11% Kan-Hyg^R, $n = 912$). Because the *ddm1* mutation is linked to the target locus on chromosome 5, it had to be introgressed into the double homozygous target/silencer line. In the first generation, when the strength of antibiotic resistance could be tested in *ddm1* mutants, sporadic weak reactivation of NOSpro–*NPTII* gene expression was observed (data not shown). In both *met1* and *ddm1* mutants, Kan^R resistance could improve
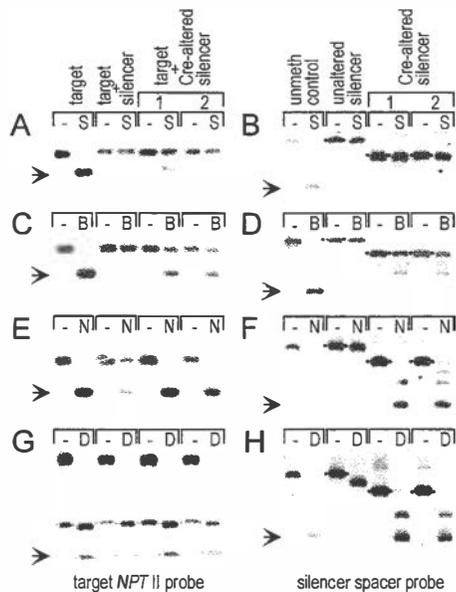
Fig. 6. Methylation analysis in the presence and absence of NOSpro dsRNA. (*A*, *C*, *E*, and *G*) Target NOSpro. (*B*, *D*, *F*, and *H*) Silencer NOSpro IR. The enzymes and probes used are described in Fig. 1. The arrows to the left of each blot show the position of the expected unmethylated fragment. Results from two lines containing the Cre-altered silencer are shown. (*G*) The NOSpro–*NPTII* bands of interest are flanked by blue dots. The large hybridizing fragment in the minus lanes and the fragment in the *D* lanes running just below the NOSpro–*NPTII* band should be disregarded as they are caused by a second *NPTII* gene (not visible in *A*, *C*, and *E*) used for bacterial selection during cloning (25). The size of the fragment in the minus lanes in *B*, *D*, *F*, and *H* is shifted after the Cre cross (indicated by red dots) because of removal of the 35Spro and is independent of methylation. Because of an unmethylated *Dde*I site in the 35Spro (Fig. 1 *Upper*), the size of the fragment of the unaltered silencer shifts after addition of *Dde*I (*H*).

in advanced generations, although the strength of NOSpro–*NPTII* gene expression continued to be nonuniform in genotypically identical seedlings (Fig. 3*D*). The strongest Kan^R plants sustained significant losses of methylation from the target locus, as indicated by substantial digestion with *Sac*II, *Bst*UI, and *Nhe*I (Fig. 4*A*, target + silencer + *met1* and + *ddm1*). These plants are indeed homozygous for the respective mutations, which cause global demethylation, as revealed by the loss of CG methylation at centromeric repeats (Fig. 4*D*, target + silencer + *met1* and + *ddm1* panels) and rDNA repeats (data not shown).

In contrast to the substantial reduction of methylation of the NOSpro target locus in plants showing relatively strong Kan^R in the *met1* and *ddm1* backgrounds, the NOSpro IR at the silencing locus retains considerable methylation in these mutants. This was particularly evident in *ddm1* mutant plants, where—similarly to wild type plants—virtually no digestion by *Sac*II, *Bst*UI, and *Nhe*I was observed (Fig. 4*B*, compare target + silencer + *ddm1* with target + silencer). In *met1* plants, methylation is reduced ≈20–30% at both symmetrical (*Sac*II, *Bst*UI) and nonsymmetrical (*Nhe*I) sites (Fig. 4*B*, target + silencer + *met1* panels). NOSpro dsRNA continues to be synthesized at wild-type levels in the *met1* and *ddm1* mutant plants (Fig. 4*C*).

Discussion

To dissect the mechanism of RdDM and dsRNA-mediated TGS, we have established a two component silencing system based on the NOSpro in *Arabidopsis*. A NOSpro dsRNA transcribed from a NOSpro IR at the silencing locus is processed to short RNAs 21–24 nt in length. Either the dsRNA or the short RNAs can locate and interact with the homologous NOSpro at the unlinked

target locus and trigger TGS and promoter methylation. We report here an analysis of methylation of the target NOSpro and the NOSpro IR at the silencing locus in the presence and absence of NOSpro dsRNA, and in several mutants deficient in DNA methylation and/or putative chromatin remodeling proteins. The results are summarized in Fig. 1.

NOSpro dsRNA induces *de novo* methylation of the target NOSpro at Cs in any sequence context within the region of RNA-DNA sequence identity (Fig. 1 target + silencer). Removing the source of the dsRNA, either by segregating away the silencing locus [Fig. 1 target, − silencer (segregation)] or by removing the transcribing 35Spro via Cre/*lox*-mediated recombination [Fig. 1 target, − dsRNA (+IR)], results readily in loss of methylation in nonsymmetrical C residues, indicating that they require continuous *de novo* methylation induced by RdDM. In contrast, methylation in symmetrical CG and CNG nucleotide groups, which can be maintained, respectively, by the DNA MTases MET1 (45) and chromomethylase3 (CMT3) (46, 47), is retained to varying degrees in the target NOSpro after it segregates away from the silencing locus or in the presence of the Cre-altered silencing locus. The persistent target promoter methylation resembles paramutation, which involves meiotically heritable changes in gene expression induced by allelic interactions (48). Indeed, some paramutation-like phenomena are probably caused by maintenance of RNA-induced CG and CNG methylation through meiosis (17), suggesting that MET1 and CMT3 can function during this period. Interestingly, methylation in symmetrical C residues is not lost as readily in the presence of the nontranscribed Cre-altered silencing locus as it is following segregation of the unaltered silencing locus. This difference cannot be ascribed to NOSpro dsRNA, which is not synthesized to detectable levels following Cre-mediated excision of the 35Spro, nor to the somewhat repetitive nature of the target locus, which remains unchanged regardless of the presence of the silencing locus or the Cre-altered derivative. One possibility is that the nontranscribed IR enhances maintenance methylation of the target NOSpro, perhaps through DNA–DNA pairing interactions (49).

The *met1* and *ddm1* mutations, which reduce global methylation and release silencing to varying extents in other TGS systems (14, 17, 29, 32, 43, 44), partially alleviate silencing and reduce methylation of the NOSpro–*NPTII* target gene but these effects can only be considered indirect. With *met1*, partial, nonuniform recovery of Kan resistance was observed in the F$_2$ generation, followed by progressive improvement in subsequent generations. The *ddm1* mutation appeared somewhat less efficient at releasing silencing than *met1*, though it is difficult to make a strict comparison because the *ddm1* mutation had to be introgressed into our NOSpro target/silencer line. In both mutants, losses of methylation in the target NOSpro can be substantial in F$_3$ and F$_4$ progeny that show the strongest Kan^R (Fig. 1 target, *met1*, *ddm1*). Any slight methylation that persists is presumably caused by continued *de novo* methylation induced by NOSpro dsRNA. The lack of effect of *mom1* on NOSpro silencing and methylation (Fig. 1 target, *mom1*) is not unexpected, because this mutation affects a subset of transcriptionally inactivated genes that are silenced by a methylation-independent pathway and, unlike *met1* and *ddm1*, does not cause global demethylation (30, 50).

The two copies of the NOSpro in the IR at the silencing locus are methylated substantially at symmetrical and nonsymmetrical C residues (Fig. 1, unaltered silencer). When dsRNA synthesis terminates following Cre-mediated removal of the transcribing 35Spro, methylation in CG and CNG nucleotide groups is maintained. In contrast, nonsymmetrical CNN methylation is substantially reduced, demonstrating its dependence on transcription of NOSpro dsRNA [Fig. 1 silencer − dsRNA (+IR)]. After withdrawal of the triggering dsRNA, the nontranscribed

Aufsatz *et al.*

IR maintains CG and CNG methylation better than singlet copies of NOSpro at target locus, suggesting that some intrinsic feature of the IR helps to maintain methylation independently of dsRNA. One possibility is that pairing of the IR in *cis* generates an unusual structure (51) that is recognized by the maintenance MTase activities. Consistent with a critical role for maintenance MTases in retaining methylation in the NOSpro IR, reductions in CG and CNG methylation in this region were greater in *met1* than in *ddm1* mutant plants (Fig. 1 silencer, *met1* and *ddm1*). A similar stronger effect of *met1* compared with *ddm1* on methylation of an IR has been noted previously on studies with the *PAI* gene family in *Arabidopsis* (32).

Overall, the results from the mutant analysis indicate that efficient maintenance of methylation triggered by RdDM requires MET1 and the activity of DDM1, perhaps as part of a chromatin remodeling complex. Despite the continued presence of NOSpro dsRNA, significant losses of target NOSpro methylation were observed after several generations in *met1* and *ddm1* mutants. This suggests that in the absence of a maintenance MTase and chromatin restructuring activities, which can help to reinforce silencing, methylation induced by RdDM is lost more rapidly than it can be regenerated *de novo*.

It is not yet known which DNA MTase catalyzes the *de novo* methylation step of RdDM, though MET1 is considered unlikely because of the somewhat delayed influence of the *met1* mutation on our NOSpro silencing system. A CMT was initially a promising candidate for RdDM (1, 52) because of the presence in these enzymes of a chromodomain, which can serve as an RNA-protein interaction module (53). Initial results with the *cmt3* mutation, however, suggest negligible effects on NOSpro silencing in F$_2$ generation (W.A., X. Cao, S. and Jacobsen, M.M., unpublished results). NOSpro methylation must still be examined in the *cmt3* mutants. Another candidate for RdDM is a member of the domain rearranged (DRM) class, which is the major *de novo* DNA MTase family in plants (54). Mutants defective in DRM2 (X. Cao and S. Jacobsen, personal communication) are currently being tested with the NOSpro system. A final possibility is a member of the Dnmt2 family, which is also present in vertebrates, *Drosophila* and—in a mutated form—in *Schizosaccharomyces pombe* (55). Mutations in this class of putative DNA MTases remain to be assessed in our NOSpro system.

There are so far no reports that RNA directs DNA methylation in animals. This apparent deficiency may reflect differences between plants and animals with respect to specific requirements for RdDM. Factors to consider include whether the unique pattern of methylation triggered by RdDM can be detected at a particular developmental stage in animals, and whether the required DNA MTase is available at that time. In both mammals and *Drosophila*, non-CG methylation, which is conceivably directed by RNA (56), is present in early embryos (57, 58). This methylation might be catalyzed by Dnmt3a, which is the major *de novo* DNA MTase active early in mammalian development (57), or by Dnmt2, which is also primarily active during the initial stages of development in mammals and in *Drosophila* (58). Both of these enzymes have been implicated in the catalysis of non-CG methylation (57, 58), which would be consistent with RdDM. Thus, if RdDM occurs in animals, it might be limited to early stages of development when the appropriate DNA MTase(s) is active. In contrast, the occurrence of RdDM throughout plant development (56) suggests the continuous activity of the necessary DNA MTase, a feature that probably facilitated the detection of RdDM in adult plants.

Even for organisms that do not methylate their DNA, there is growing evidence that chromatin modifications are targeted by components of the RNAi machinery. In *Drosophila*, transgene TGS and PTGS are both dependent on the piwi protein, which is a member of the Argonaute family required for RNAi (59). TGS is associated with complexes of polycomb-group proteins, which are perhaps directed to the transgene promoter by short RNAs containing transcriptional regulatory sequences. In *S. pombe*, homologs of three proteins required for RNAi—dicer, a putative RNA-dependent RNA polymerase, and argonaute—are needed for histone methylation and localization of the heterochromatin protein Swi6 at centromeric repeats (S. Grewal, personal communication). RNAi-based genetic screens to find genes required for RNAi in *Caenorhabditis elegans* identified several ORFs that are predicted to encode chromatin-associated proteins (60).

Genetic screens are required to recover novel mutations affecting NOSpro dsRNA-mediated TGS. We have recently identified one prospective mutant, *rts-1* (RNA-mediated transcriptional silencing), in which silencing is substantially alleviated, whereas target NOSpro methylation is only reduced about 50% (W.A., M.F.M., and A.J.M.M., unpublished results). The *rts-1* mutation does not map to a region of the *Arabidopsis* genome known to encode a DNA MTase, suggesting that it might encode a chromatin factor. Identification of the *RTS-1* gene and further genetic analyses using the NOSpro dsRNA-mediated TGS system should continue to provide insights into the relationship between RdDM and chromatin modifications.

1. Matzke, M. A., Matzke, A. J. M. & Kooter, J. (2001) *Science* **293**, 1080–1083.
2. Vance, V. B. & Vaucheret, H. (2001) *Science* **292**, 2277–2280.
3. Waterhouse, P., Wang, M. B. & Lough, T. (2001) *Nature (London)* **411**, 834–842
4. Voinnet, O. (2001) *Trends Genet.* **17**, 449–459.
5. Bernstein, E., Denli, A. M. & Hannon, G. J. (2001) *RNA* **7**, 1509–1521.
6. Cogoni, C. (2001) *Annu. Rev. Microbiol.* **55**, 381–406.
7. Chicas, A. & Macino, G. (2001) *EMBO Rep.* **2**, 992–996.
8. Hutvágner, G. & Zamore, P. (2002) *Curr. Opin. Genet. Dev.* **12**, 225–232.
9. Hamilton, A. & Baulcombe, D. C. (1999) *Science* **286**, 950–952.
10. Wassenegger, M. (2000) *Plant Mol. Biol.* **43**, 203–220.
11. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. M. (2000) *EMBO J.* **19**, 5194–5201.
12. Pélissier, T. & Wassenegger, M. (2000) *RNA* **6**, 55–65.
13. Kovarik, A., Van Houdt, H., Holy, A. & Depicker, A. (2000) *FEBS Lett.* **467**, 47–51.
14. Morel, J. B., Mourrain, P., Béclin, C. & Vaucheret, H. (2000) *Curr. Biol.* **10**, 1591–1594.
15. Sijen, T., Vijn, I., Rebocho, A., van Blokland, R., Roelofs, D., Mol, J. N. M. & Kooter, J. (2001) *Curr. Biol.* **11**, 436–440.
16. Jones, L., Hamilton, A. J., Voinnet, O., Thomas, C. L., Maule, A. J. & Baulcombe, D. C. (1999) *Plant Cell* **11**, 2291–2301.
17. Jones, L., Ratcliff, F. & Baulcombe, D. C. (2001) *Curr. Biol.* **11**, 747–757.
18. Jedelloh, J., Stokes, T. & Richards, E. (1999) *Nat. Genet.* **22**, 94–97.
19. Gibbons, R., McDowell, T., Raman, S., O'Rourke, D., Garrick, D., Ayyub, H. & Higgs, D. (2000) *Nat. Genet.* **24**, 368–371.
20. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. (2001) *Genes Dev.* **15**, 2940–2944.
21. Dobosy, J. & Selker, E. (2001) *Cell Mol. Life Sci.* **58**, 721–727.
22. Tamaru, H. & Selker, E. (2001) *Nature (London)* **414**, 277–283.
23. Jackson, J., Lindroth, A., Cao, X. & Jacobsen, S. (2002) *Nature (London)* **416**, 556–560.
24. Clough, S. J. & Bent, A. F. (1998) *Plant J.* **16**, 735–743.
25. Matzke, A. J. M. & Matzke, M. (1986) *Plant Mol. Biol.* **7**, 357–365.
26. Chen, Z. L., Schuler, M. A. & Beachy, R. N. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8560–8564.

27. De Buck, S., De Wilde, C., Van Montagu, M. & Depicker, A. (2000) *Mol. Plant–Microbe Interact.* **13**, 658–665.
28. Edwards, K., Johnstone, C. & Thompson, C. (1991) *Nucleic Acids Res.* **19**, 1349.
29. Mittelsten Scheid, O., Afsar, K. & Paszkowski, J. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 632–637.
30. Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O. & Paszkowski, J. (2000) *Nature (London)* **405**, 203–206.
31. Vongs, A., Kakutani, T., Martienssen, R. & Richards, E. (1993) *Science* **260**, 1926–1928.
32. Bartee, L. & Bender, J. (2001) *Nucleic Acids Res.* **29**, 2127–2134.
33. Matzke, M., Primig, M., Trnovsky, J. & Matzke, A. J. M. (1989) *EMBO J.* **8**, 643–649.
34. Pélissier, T., Thalmeir, S., Kempe, D., Sänger, H. L. & Wassenegger, M. (1999) *Nucleic Acids Res.* **27**, 1625–1634.
35. Jakowitsch, J., Mette, M. F., van der Winden, J., Matzke, M. & Matzke, A. J. M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 13241–13246.
36. Nacry, P., Camilleri, C., Courtial, B., Caboche, M. & Bouchez, D. (1998) *Genetics* **149**, 641–650.
37. Tax, F. E. & Vernon, D. M. (2001) *Plant Physiol.* **126**, 1527–1538.
38. Jakowitsch, J., Papp, I., Moscone, E., van der Winden, J., Matzke, M. & Matzke, A. J. M. (1999) *Plant J.* **17**, 131–140.
39. Jones, G. H. & Heslop Harrison, J. S. (2000) in *Arabidopsis: A Practical Approach*, ed. Wilson, Z. A. (Oxford Univ. Press, Oxford), pp. 105–124.
40. Mette, M. F., van der Winden, J., Matzke, M. & Matzke, A. J. M. (1999) *EMBO J.* **18**, 241–248.
41. Metz, A. M., Timmer, R. T. & Browning, K. S. (1992) *Gene* **120**, 313–314.
42. Park, Y. D., Papp, I., Moscone, E., Iglesias, V., Vaucheret, H., Matzke, A. J. M. & Matzke, M. (1996) *Plant J.* **9**, 183–194.
43. Jedelloh, J., Bender, J. & Richards, E. (1998) *Genes Dev.* **12**, 1714–1725.
44. Vielle-Calzada, J., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M. & Grossniklaus, U. (1999) *Genes Dev.* **13**, 2971–2982.
45. Kishimoto, N., Sakai, H., Jackson, J., Jacobsen, S., Meyerowitz, E., Dennis, E. & Finnegan, E. J. (2001) *Plant Mol. Biol.* **46**, 171–183.
46. Bartee, L., Malagnac, F. & Bender, J. (2001) *Genes Dev.* **15**, 1753–1758.
47. Lindroth, A., Cao, X., Jackson, J., Zilberman, D., McCallum, C., Henikoff, S. & Jacobsen, S. (2001) *Science* **292**, 2077–2080.
48. Chandler, V. L., Eggleston, W. & Dorweiler, J. (2000) *Plant Mol. Biol.* **43**, 121–145.
49. Matzke, M., Mette, M. F., Jakowitsch, J., Kanno, T., Moscone, E., van der Winden, J. & Matzke, A. J. M. (2001) *Genetics* **158**, 451–461.
50. Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O. & Paszkowski, J. (2000) *Plant Cell* **12**, 1165–1178.
51. Smith, S. S., Kann, J. L., Baker, D. J., Kaplan, B. E. & Dembek, P. (1991) *J. Mol. Biol.* **217**, 39–54.
52. Habu, Y., Kakutani, T. & Paszkowski, J. (2001) *Curr. Opin. Genet. Dev.* **11**, 215–220.
53. Akhtar, A., Zink, D. & Becker, P. B. (2000) *Nature (London)* **407**, 405–409.
54. Cao, X., Springer, N., Muszynski, M., Phillips, R., Kaeppler, S. & Jacobsen, S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 4979–4984.
55. Bestor, T. H. (2000) *Hum. Mol. Genet.* **9**, 2395–2402.
56. Matzke, M., Mette, M. F., Kanno, T., Aufsatz, W. & Matzke, A. J. M. (2002) in *Gene Silencing*, ed. Hannon, G. (Cold Spring Harbor Lab. Press, Plainview, NY), in press.
57. Ramsahoye, B., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. & Jaenisch, R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5237–5242.
58. Lyko, F. (2001) *Trends Genet.* **17**, 169–172.
59. Pal-Bhadra, M., Bhadra, U. & Birchler, J. (2002) *Mol. Cell* **9**, 315–327.
60. Dudley, N., Labbé, J. & Goldstein, B. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 4191–4196.
61. Mitra, A. & An, G. (1989) *Mol. Gen. Genet.* **215**, 294–299.

Arthur M. Sackler
C O L L O Q U I A
OF THE NATIONAL ACADEMY OF SCIENCES

Self-Perpetuating Structural States in Biology, Disease, and Genetics

March 22–24, 2002

National Academy of Sciences, Washington, DC
Organized by Susan Lindquist and Steve Henikoff

Program

Friday, March 22

Sackler Lecture

Susan Lindquist, Whitehead Institute
Mad Cows Meet Psi-Chotic Yeast: New Paradigms in Genetics and Disease

Saturday, March 23

Session I. Self-Perpetuating Protein Conformations

Keynote Speaker
Charles Weissmann, Imperial College School of Medicine at St. Mary's
Mammalian Prion Diseases

Reed Wickner, National Institutes of Health
The Yeast Prion [URE3] Is a Heritable Amyloid of Ure2p: Genetics and Mechanisms

Susan Liebman, University of Illinois
Prion–Prion Interactions, Predisposing Factors

Eric Kandel and Kausik Si, Columbia University
Self-Perpetuating Protein Conformations and Memory

Session II. Protein Conformation Changes, Amyloid and Disease

Erich Wanker, Max Delbruck Center for Molecular Medicine
Huntington Protein Assembly, Trafficking and Drug Interactions

Nancy Bonini, University of Pennsylvania
*Drosophil*a **Model for Neurodegenerative Disease, Protein–Protein Interactions**

Ulrich Hartl, Max Planck Institute for Biochemistry
Amyloid Disease Proteins and Chaperone Interactions

Chris Dobson, University of Cambridge
Links Between Protein Folding and Human Disease

Jeffrey W. Kelly, The Scripps Research Institute
Understanding the Energy Landscape Associated with Transthyretin Amyloid Diseases and Manipulating It to Prevent Amyloidosis

Sunday, March 24

Session III. Chromatin Structural States

Gary Felsenfeld, National Institutes of Health
Insulation of Genes from Silencing Chromatin

Shiv Grewal, Cold Spring Harbor Laboratory
Epigenetic Control of Higher-Order Chromatin Assembly

Susan Lindquist, Whitehead Institute
Prion Diversity: Multiple Functions Including Transcription States

Dan Gottschling, Fred Hutchinson Cancer Research Center
The Challenge of Creating a Silent Chromatin State at Yeast Telomeres: Heritability and Promiscuity

Lori Wallrath, University of Iowa
Connections Between Breast Cancer Metastasis and Gene Silencing from Studies on Heterochromatin Protein 1 in Humans and *Drosophila*

Session IV. Self-Perpetuating Genetic Interactions

David Low, University of California, Santa Barbara
Epigenetic Pili Switches in Bacteria

Steve Henikoff, Fred Hutchinson Cancer Research Center
Replication-Independent Assembly of Chromatin Utilizes Variant H3 Histones

Eric Selker, University of Oregon
**Control of DNA Methylation: Lessons from Fungi**

Steve Jacobsen, University of California, Los Angeles
Genetics of DNA Methylation in *Arabidopsis*

Marjori Matzke, Austrian Academy of Sciences
RNA-Based Epigenetic Silencing in Plants

Arthur M. Sackler
COLLOQUIA
OF THE NATIONAL ACADEMY OF SCIENCES

National Academy of Sciences
Sackler Colloquia Series

The Arthur M. Sackler Colloquia of the National Academy of Sciences address scientific topics of broad and current interest, cutting across the boundaries of traditional disciplines. Each year, five or six such colloquia are scheduled, typically two days in length and international in scope. Colloquia are organized by a member of the Academy, often with the assistance of an organizing committee, and feature presentations by leading scientists in the field and discussions with a hundred or more researchers with an interest in the topic. Colloquium proceedings are published in the *Proceedings of the National Academy of Sciences* (PNAS) and are available for purchase from the National Academy Press. These colloquia are made possible by a generous gift from Mrs. Arthur M. Sackler, in memory of her husband.
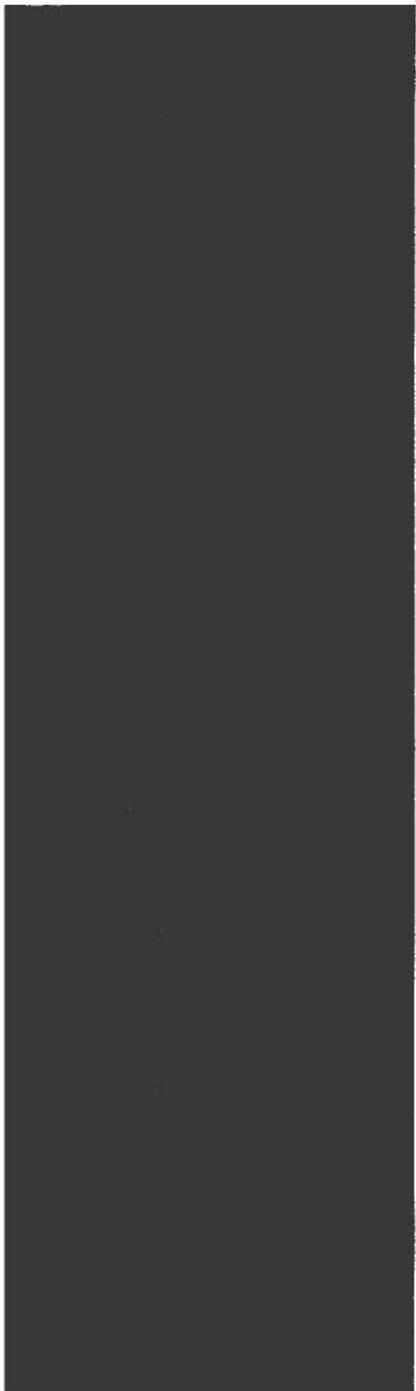
Arthur M. Sackler
C O L L O Q U I A
OF THE NATIONAL ACADEMY OF SCIENCES

National Academy of Sciences
500 Fifth Street, NW
Washington, DC 20001