


## (Sackler NAS Colloquium) Chemical Communication in a Post-Genomic World

ISBN  
978-0-309-09089-6

108 pages  
8 1/2 x 11  
2003

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

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences. Request reprint permission for this book

# Chemical Communication in a Post-Genomic World

National Academy of Sciences  
Washington, D.C.

ch. 12 part  
#389.1

## Arthur M. Sackler, M.D.

1913–1987

**B**orn 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.



## Contents

### Papers from the Arthur M. Sackler Colloquium of the National Academy of Sciences

#### INTRODUCTIONS

- 14513 **Chemical communication in a post-genomic world**  
May R. Berenbaum and Gene E. Robinson
- 14514 **Understanding the chemistry of chemical communication: Are we there yet?**  
Jerrold Meinwald
- 14517 **Chemical ecology: Can it survive without natural products chemistry?**  
Thomas Eisner

#### COLLOQUIUM PAPERS

- 14519 **Pheromone-mediated gene expression in the honey bee brain**  
Christina M. Grozinger, Noura M. Sharabash, Charles W. Whitfield, and Gene E. Robinson
- 14526 ***Drosophila Gr5a* encodes a taste receptor tuned to trehalose**  
Sylwester Chyb, Anupama Dahanukar, Andrew Wickens, and John R. Carlson
- 14531 **Mammalian TRPV4 (VR-OAC) directs behavioral responses to osmotic and mechanical stimuli in *Caenorhabditis elegans***  
Wolfgang Liedtke, David M. Tobin, Cornelia I. Bargmann, and Jeffrey M. Friedman
- 14537 **Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster***  
Hugh M. Robertson, Coral G. Warr, and John R. Carlson
- 14543 **A genomic perspective on nutrient provisioning by bacterial symbionts of insects**  
Nancy A. Moran, Gordon R. Plague, Jonas P. Sandström, and Jennifer L. Wilcox

- 14549 **Chemical communication among bacteria**  
Michiko E. Taga and Bonnie L. Bassler
- 14555 **Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species**  
Gregory L. Challis and David A. Hopwood
- 14562 **Efficient oxidative folding of conotoxins and the radiation of venomous cone snails**  
Grzegorz Bulaj, Olga Buczek, Ian Goodsell, Elsie C. Jimenez, Jessica Kranski, Jacob S. Nielsen, James E. Garrett, and Baldomero M. Olivera
- 14569 **Non-self recognition, transcriptional reprogramming, and secondary metabolite accumulation during plant/pathogen interactions**  
Klaus Hahlbrock, Pawel Bednarek, Ingo Ciolkowski, Björn Hamberger, Andreas Heise, Hiltrud Liedgens, Elke Logemann, Thorsten Nürnberger, Elmon Schmelzer, Imre E. Somssich, and Jianwen Tan
- 14577 **Systemins: A functionally defined family of peptide signals that regulate defensive genes in Solanaceae species**  
Clarence A. Ryan and Gregory Pearce
- 14581 ***Manduca sexta* recognition and resistance among allopolyploid *Nicotiana* host plants**  
Yonggen Lou and Ian T. Baldwin
- 14587 **Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus**  
Juergen Kroymann, Susanne Donnerhacke, Domenica Schnabelrauch, and Thomas Mitchell-Olds
- 14593 **Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate**  
Weimin Li, Mary A. Schuler, and May R. Berenbaum
- 9179 **Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera**  
Wendell L. Roelofs and Alejandro P. Rooney



# Chemical communication in a post-genomic world

May R. Berenbaum\* and Gene E. Robinson

Department of Entomology, University of Illinois, 320 Morrill Hall, 505 South Goodwin, Urbana, IL 61801

With genome sequences accumulating at a rapid pace, one major goal of biology is to understand the function of genes. Many gene functions are comprehensible only within the context of chemical communication, and emerging research on genomics and chemical communication has catalyzed development of this highly productive interface. Many of the most abundantly represented genes in the genomes characterized to date encode proteins mediating interactions among organisms, including odorant receptors and binding proteins, enzymes involved in biosynthesis of pheromones and toxins, and enzymes catalyzing the detoxification of defense compounds. Chemosensory signaling shares several features irrespective of taxon; components of the vast majority of chemosensory signaling systems include a receptor that interacts with a signal molecule, a signal transducer, an amplifier, and a receiver. Across a wide range of organisms, many of the same classes of molecules perform these functions, even if the precise identity of the molecule in particular systems differs. Genome sequencing projects are bringing these similarities into sharp focus for the first time, particularly in the area of chemical communication.

Determining the molecular underpinnings of the component elements of chemical communication systems in all of their forms has the potential to explain a vast array of ecological and evolutionary phenomena. By the same token, ecologists who elucidate the environmental challenges faced by the organisms are uniquely well equipped to characterize natural ligands for receptors and substrates for enzymes. Thus, partnerships between genome biologists and chemical ecologists can be extremely synergistic. To date, these groups have rarely had

opportunities to interact within a single forum. Such interactions are vital given the considerable practical benefits potentially stemming from these studies, including the development of biorational products for agricultural and forest pest management, for disease treatment, and for improving the quality of ecosystem health (1).

In 1994, the National Academy of Sciences sponsored a colloquium on chemical ecology; that colloquium, the proceedings of which were published in PNAS, resulted in the publication of a book edited by Thomas Eisner and Jerrold Meinwald, and generated considerable interest in and excitement about the field (2). Since that time, with the availability of genomic tools, the field has metamorphosed and progressed in unprecedented quantum leaps. We organized this Arthur M. Sackler colloquium, held almost a decade later, to aid in the effort to integrate chemical ecology into the broader context of modern molecular biology. The articles in this volume, representing the proceedings of the colloquium held January 17–19 at the Arnold and Mabel Beckman Center in Irvine, California, provide not only examples of cutting-edge work at the interface between molecular and organismal biology but also useful perspectives and guidelines for future work in these and other systems. We thank Dr. James Langer, vice president of the National Academy of Sciences, and the Sackler selection committee for providing the resources to support the colloquium; Miriam Glaser Heston, program officer for the colloquium series, for superb logistical support and advice; and Andrew Pillifant and Christina Colosimo of the editorial staff of PNAS for invaluable assistance in assembling this special issue. This volume is dedicated to Drs. Eisner and Meinwald, trailblazers and intellectual leaders in this field.

1. Eisner, T. & Berenbaum, M. (2002) *Science* **295**, 1973.

2. Eisner, T. & Meinwald, J., eds. (1995) *Chemical Ecology: The Chemistry of Biotic Interaction* (Natl. Acad. Press, Washington, DC).

This paper serves as an introduction to the following papers, which result from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

\*E-mail: maybe@uiuc.edu.

© 2003 by The National Academy of Sciences of the USA

## Introduction

# Understanding the chemistry of chemical communication: Are we there yet?

Jerrold Meinwald\*

Department of Chemistry and Chemical Biology, Baker Laboratory, Cornell University, Ithaca, NY 14853-1301

Small molecules often carry vital information. Compounds as simple as nitric oxide, carbon dioxide, and ethylene play essential roles as hormones or pheromones in plants and animals. Extensive research during the past century has revealed that chemical signaling is not only the primary means whereby organisms coordinate and regulate their internal activities, but also the chief channel through which organisms interact with one another. In this Introduction, I will examine briefly the evolution of the science of organic chemistry, which lies at the heart of chemical communication, and consider what some future goals for organic chemistry in a postgenomic world may be.

By the start of the 19th century, chemists had begun to make sense of the inorganic world. Elements were being discovered at a rapid rate, and the importance of quantitative research had become apparent. Chemists' curiosity naturally extended to the study of compounds making up living organisms as well, but "organic" chemistry, the chemical study of the biotic world, presented essentially insurmountable difficulties. In the preface to his pioneering treatise, *Lectures on Animal Chemistry* (1806), J. J. Berzelius wrote "... Of all of the sciences contributing to medicine, chemistry is the primary one, and apart from the general light it throws on the entire art of healing, it will soon bestow on some of its branches a perfection such as one never could have anticipated." Despite this exuberance, worthy of the introductory section of any contemporary chemist's National Institutes of Health research grant application, it was only 4 years later that Berzelius himself became deeply discouraged with the unreliability of his own analytical work on animal products, declaring in 1810 that he had written his last paper on nasal mucus, bile, urine, and related subjects (1). Although by 1831 Justus Liebig was able to report elemental analyses of such fascinating and important plant-derived natural products as morphine, quinine, and strychnine, the results he obtained were disorienting, because there was no understanding of what the complex ratios of carbon to hydrogen, oxygen, and nitrogen in these alkaloids implied about their chemical nature. The prevailing confusion was dramatically summarized in an 1835 letter to Berzelius by his distinguished student, colleague, and friend, Friedrich Wöhler, who wrote "... Organic chemistry just now is enough to drive one mad. It gives me an impression of a primeval tropical forest, full of the most remarkable things, a monstrous and boundless thicket, with no way to escape, into which one may well dread to enter..." (2). The problem with organic chemistry was 2-fold. On the experimental front, only the crudest techniques of distillation and crystallization were available for the isolation and purification of organic compounds, and combustion analyses required grams of material simply for the determination of carbon, hydrogen, and nitrogen percentages in an unknown sample. On the theoretical side, matters were even worse. It was not until 1859 that August Kekulé provided a clear insight into molecular structure theory, thereby explaining the nature of structural isomerism and providing a rationale for structure determination. And it

was only in 1874 that J. A. LeBel and J. H. van't Hoff discerned the fundamentals of stereochemistry, which proved indispensable for the modeling of essentially all biologically relevant chemistry.

With the realization that organic compounds are almost universally compounds of carbon, and that carbon-based chemistry offers a practically limitless domain of study entirely independent of biology, the subject of organic chemistry redefined itself much more broadly as simply the chemistry of carbon compounds. Methods for making and breaking carbon-carbon bonds, assembling natural and non-natural structures by rational synthesis, determining both molecular structures and stereochemistry, relating molecular structures to chemical and physical properties, and understanding the detailed mechanisms by which organic reactions occur all took center stage. Organic chemistry blossomed into a self-contained, inward-looking science, bringing order and understanding to the synthesis and reactions of millions of compounds. Pursuit of this young science also led to the discovery of synthetic dyes, plastics, fibers, pesticides, explosives, and pharmaceuticals that have repeatedly changed the quality of human life on Earth. It may not be going too far to suggest that the appeal of the abstract logic of organic chemistry, closely akin to that of Euclidean geometry, combined with the intriguing aromas of volatile compounds and the visual beauty of various crystals, inspires a sense of chemiphilia in many chemists, analogous in some ways to E. O. Wilson's biophilia (3).

It is impossible to overestimate the extent to which the spectacular development of organic chemistry has depended on a relatively small number of experimental techniques that were discovered or significantly improved during the 20th century. Chromatography in its various forms permits undreamed-of separations of complex mixtures from large-scale preparative experiments down to the nanogram scale. UV, IR, and NMR spectroscopy, MS, and single crystal x-ray crystallography have revolutionized the art of structure determination. The amount of compound required for a structure determination in Kekulé's time compared with ours has decreased at least a million-fold. Furthermore, a structure determination that might have required 20 or more years, starting in 1900, might be accomplished in <1 hour today, using readily available NMR spectroscopic techniques.

The advances in separation and structure determination developed chiefly in the second half of the 20th century had a special impact on the subdiscipline of "natural products" chemistry. They made possible the isolation and characterization of tens of thousands of the most important "secondary metabolites" found in nature. Even more significant from the biologist's

---

This paper serves as an introduction to the following papers, which result from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17-19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

\*E-mail: circe@cornell.edu.

© 2003 by The National Academy of Sciences of the USA

**Table 1. The relationship of sample mass of a hypothetical compound (molecular weight of  $\approx 300$ ) to the number of molecules in the sample, and to the applicability of the three most powerful experimental techniques for structure determination**

Sample size, g	Sample size, no. of molecules	X-ray crystallography	NMR spectroscopy	MS
$\sim 300 \times 10^0$	$6.23 \times 10^{23}$	+	+	+
$50 \times 10^0$	$10^{23}$	+	+	+
$50 \times 10^{-3}$	$10^{20}$	+	+	+
$50 \times 10^{-6}$	$10^{17}$	+	+	+
$50 \times 10^{-9}$	$10^{14}$	-	-	+
$50 \times 10^{-12}$	$10^{11}$	-	-	+
$50 \times 10^{-15}$	$10^8$	-	-	-
$50 \times 10^{-18}$	$10^5$	-	-	-
$50 \times 10^{-21}$	$10^2$	-	-	-

+, applicable; -, inapplicable.

viewpoint, these novel techniques could be coupled not only with one another but also with newly developed biological experimental methods, such as the "electroantennogram" technique, to provide insights into interactions at the molecular level, thereby supporting the emergent discipline of chemical ecology.

It is instructive to contrast present-day research techniques with Butenandt's pioneering research on chemical communication, carried out over a roughly two-decade period in the mid-20th century and leading to the characterization of bombykol, the first pheromone to be chemically characterized. His painstaking preparation of  $\approx 10$  mg of the pure 4'-nitroazobenzene-4-carboxylic ester of bombykol from extracts of the terminal segments of a half million virgin female silkworm moths (*Bombyx mori*) is a classic of natural products chemistry (4). With these beautiful red-orange crystals, he was able to carry out a microscale oxidative degradation that allowed him to deduce the structure of bombykol itself. (Significantly, the stereochemistry of natural bombykol could only be established on the basis of stereospecific syntheses of all four of the possible E/Z isomers.)

A contemporary researcher would find a problem of this sort vastly easier to solve. Using the combined techniques of gas chromatography, electroantennogram detection, and MS, a typical moth sex pheromone might be chemically characterized in a day's work by using only a microgram or less of material obtainable from a single female. (Again, synthetic work might be needed to ascertain stereochemistry and provide quantities useful for further studies, such as field trapping experiments.) Clearly, we have come a long way in recent decades.

Does this mean that chemists have completed their part of the job, and that the chemical aspects of chemical communication research are essentially routine? If we examine chemical signaling from the viewpoint of numbers of signal molecules required for state-of-the-art structure determination, we can see that we have the potential to do very much better. Consider a signal compound (pheromone, allomone, kairomone, hormone, or neurotransmitter) with a modest molecular mass of  $\approx 300$ . In 1950, structure proof for such a compound might have been carried out by using a sample size of  $\approx 50$  mg. Today, this task might be accomplished by using only  $\approx 50$   $\mu$ g or perhaps even 50 ng for a moderately complex unknown, which would represent a reduction of between  $10^3$  and  $10^6$  in sample size. However, even as little as 50 ng of our hypothetical signal compound corresponds to a hundred trillion ( $10^{14}$ ) molecules (see Table 1). If we wanted to study chemical communication in mites, or pursue the chemistry of still smaller organisms that

are not readily cultured, these quantities would be difficult to obtain. Because  $10^2$  or fewer signal molecules are certainly sufficient in many situations to trigger behavioral responses, there is still a quantitative gap of at least a factor of  $10^{12}$  between biologically significant quantities of a molecular messenger and the quantities that a chemist skilled in the art can characterize today.

Of course, any chemist working with natural products will recognize that up to this point we have completely ignored several extremely important factors in our discussion, thereby greatly oversimplifying the task at hand. Most natural compounds of interest are found in a matrix of other compounds, from which they need to be separated or else analyzed *in situ* (for example by UV, IR, or NMR spectroscopy) (5, 6). *In situ* analysis has significant advantages over any procedure involving a preliminary separation (chromatographic or otherwise) during which unstable, volatile, or irreversibly absorbed components might be lost. Unfortunately, such a direct approach is not universally applicable. In any case, it is important to bear in mind that the handling of ultra-small quantities becomes increasingly difficult as sample size decreases. Finally, the fact that biological information transfer frequently depends on mixtures of two or more components, sometimes in well-defined ratios, adds still another level of complexity to the analysis of chemical communication systems.

Overall, recent technical advances in analytical techniques enable the characterization of chemical signals with a facility that could not have been anticipated even a few decades ago. There is every reason to assume that future advances will continue to make the chemist's job easier, ultimately permitting both separations and chemical characterization to be achieved by using a small number of molecules. Although it is not within the scope of this discussion, it is also likely that chemical biologists will soon be able to model the interactions of many signal molecules with their receptor proteins, which should provide invaluable insights into the workings of any chemical communication system. We can further expect that genomic and proteomic research will go a long way toward elucidating the evolution and regulation of biosynthetic pathways and toward defining the mechanisms of reception and transduction.

What kinds of problems are accessible with the techniques already in hand? If a quantitative bioassay can be designed, it should be possible to characterize any biologically active natural product that can be obtained in microgram quantities. One example taken from the current literature illustrates the kind of opportunity that is open to imaginative researchers. In an exciting reversal of the traditional approach of natural products chemists, Spehr *et al.* (7) have found and characterized an olfactory receptor protein in the tails of human spermatozoa. Those researchers could demonstrate that a simple aromatic aldehyde, bourgeonal, at a concentration as low as  $10^{-7}$  M, can guide sperm swimming. They have also shown that this response is blocked by a low concentration of *n*-undecanal. However, it still remains to find and characterize the natural ligand for this olfactory receptor, presumably a low molecular weight pheromone secreted by the human egg. This intriguing task will very likely test the limits of small-scale chemical characterization. And it will certainly not be without wide-ranging impact!

There can be no doubt that Berzelius, Liebig, Wöhler, *et al.*, were they with us today, would be absolutely delighted to take up once again the original mission of organic chemistry, *i.e.*, to provide the chemical basis for the understanding of life. They would certainly be amazed by the incredibly powerful experimental and theoretical tools now at hand and available for this purpose. Those of us who have had the privilege of pursuing problems involving natural products chemistry and chemical communication have enjoyed the endeavor immensely. It is hard



to imagine that future generations of chemiphiles will not be at least equally intrigued by the opportunities that the explorations of future biophiles are sure to provide. If one were to ask whether chemistry right now is anywhere near reaching its ultimate goal with respect to providing a full molecular understanding of chemical communication, the answer must be a resounding NO! Future opportunities far outweigh present accomplishments,

1. Jorpes, J. E. (1966) *Jac. Berzelius, His Life and Work* (Almqvist and Wiksell, Stockholm).
2. Fieser, L. F. & Fieser, M. (1961) *Advanced Organic Chemistry* (Reinhold, New York) p. 11.
3. Wilson, E. O. (1984) *Biophilia* (Harvard Univ. Press, Cambridge, MA).
4. Hecker, E. & Butenandt, A. (1984) in *Techniques in Pheromone Research*, eds.

which are best viewed as a promising start. So, we are not there yet, but we are certainly on our way!

My research in chemical communication, pursued for more than four decades in close collaboration with Dr. Thomas Eisner, along with many dozens of dedicated predoctoral and postdoctoral students, has been generously supported by National Institutes of Health Grant GM53830 and the Schering Plough Research Institute.

- Hummel, H. E. & Miller, T. A. (Springer, New York), pp. 1–44.
5. Schröder, F. C., Sinnwell, V., Bauman, H., Kaib, M. & Francke, W. (1997) *Angew. Chem. Int. Ed.* **36**, 77–80.
6. Schröder, F. C., Farmer, J. J., Attygalle, A. B., Smedley, S. R., Eisner, T. & Meinwald, J. (1998) *Science* **281**, 428–431.
7. Spehr, M., Gisselman, G., Poplawski, A., Riffell, J. A., Wetzel, C. H., Zimmer, R. K. & Hatt, H. (2003) *Science* **299**, 2054–2058.

## Introduction

# Chemical ecology: Can it survive without natural products chemistry?

Thomas Eisner\*

Department of Neurobiology and Behavior, Cornell University, Ithaca, NY 14853

New disciplines arise by convergence of interests. Chemical ecology is the product of a partnership between biologists and natural products chemists united by a shared vision and empowered by complementary skills. The vision stems from the realization that all organisms emit chemical signals and that all, in their respective ways, respond to the chemical emissions of others. Nature, in accord with this construct, is a dynamic assemblage of vast complexity, driven by interactions that are, for the most part, mediated by molecules. The natural products chemist brings to the partnership the ability to decipher these chemical signals. Not surprisingly, the collaboration between the chemical ecologist and the natural products chemist is a close one (see Fig. 1).

Chemical ecology came into its own in the midst of the molecular biological revolution, in the 1950s, at the same time that vastly improved techniques and instruments came on line by which chemicals could be isolated and characterized. Natural products chemistry traditionally had been applied in its orientation. Its primary goal was the discovery of molecules of use, substances of medical, industrial, and agricultural interest, and it was highly successful in such endeavors. Its expansion into the domain of chemical ecology represented a shift in goals, to problems of fundamental rather than applied significance, but the challenge was immense and the potential significance of the findings enormous. At stake was the understanding of the chemical basis of biotic interaction.

Think of how profoundly chemical ecology has been affected by what relatively little we have learned so far about the signaling agents of nature. Think of how this knowledge has affected the way we view the biotic world. And think also of how profoundly this knowledge has affected our view of the physical world. The air, the oceans, and the inland waters can no longer be viewed as simple matrices. All, in accord with the grander view, are seen as carriers, as the vehicles by which the communicative messages of life are conveyed.

Chemical ecology was to derive further advantage by what can be termed its "molecularization," that is, its ever-widening linkage with various molecular subdisciplines of the biological sciences. The signal molecules that convey information from one organism to another are biosynthesized under genetic control, deciphered at specific receptor sites, transduced into neuronal, neuroendocrine, or phytoendocrine signals, and, eventually, after triggering intermediate cascading effects, translated into behavioral or morphogenetic responses. Each step in this sequence of events lends itself to interpretation in molecular terms. Not surprisingly, chemical ecology is poised to broaden its inquiries in accord with opportunities offered by advances in molecular biology.

It is clear from all these developments that chemical ecology has major exploratory potential and that the path ahead for the discipline is rich in opportunity. Yet there is a disturbing development that needs attention, lest it prevent that path from being taken. Ironic as it may seem, natural products chemistry is currently slated for deemphasis. This turn of events is partic-

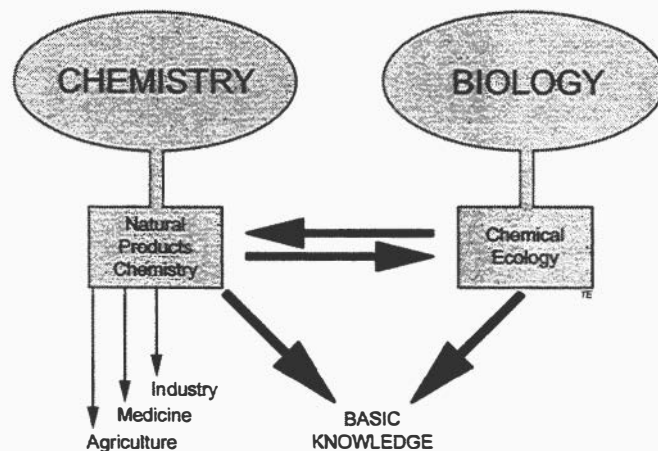


Fig. 1. Chemical ecology and natural products chemistry are linked in a productive partnership aimed at clarifying the chemical basis of ecological and behavioral interactions in nature. To curtail natural products chemistry now is to put on hold the acquisition of this fundamental knowledge.

ularly disconcerting because it comes at a time when, technically, the discipline is ideally positioned to meet its goals. Thanks to vast improvements in techniques and instrumentation, compounds can now be isolated and characterized on the basis of minute amounts of material. One can now literally reach into the air or the waters to capture, selectively, and in the desired amounts, the messenger molecules of one's choice. Where gram or kilogram quantities of starting material were once needed, microgram or nanogram quantities now suffice. Given the promise and the significance of the discoveries at stake, does it make sense that one should now put constraints on natural products chemistry?

Some recent events are worth pondering. At the very time when one is finally in possession of the proper analytical tools, chemistry departments and funding agencies, not to mention industrial concerns, are backing away from support of natural products chemistry. The very universities that in years past provided a home for the birth of chemical ecology are now relinquishing leadership in the area. Chemistry departments, caught up in the campuswide struggle for control of the genomic, postgenomic, and other molecular programs, are seeking glory in name change. "Department of Chemistry and Chemical Biology" is what more than one university chemistry department now calls itself, even after taking the paradoxical step of

This paper serves as an introduction to the following papers, which result from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

\*E-mail: te14@cornell.edu.

© 2003 by The National Academy of Sciences of the USA

eliminating natural products chemistry from the curriculum. And industries themselves are increasingly inclined to curtail their analytical and synthetic natural products programs.

Chemical ecologists, and for that matter biologists generally, will do well to stand in opposition to this trend. The biological research effort will wither on many fronts unless the capability to characterize natural products is maintained. If chemistry departments are unwilling to house the necessary instrumentation, or recruit faculty in the area, then these requirements will need to be met by the biological establishment itself. There is precedent for the incorporation into biology of a discipline that could just as well formally have been linked to chemistry. Biochemistry, at the time of its emergence, did not gain accep-

tance in chemistry but in biology, and the affiliation, quite obviously, has been a successful one. I would suggest that the same could be done for chemical ecology. If this discipline is to live up to its promise, it will need to find a home that is sympathetic both to its biological and to its chemical needs. Chemical ecology is now embarking on the most ambitious and inventive phase of its existence. To stand by and allow natural products chemistry to vanish, or even to be weakened, is to deny chemical ecology its future.

My research in chemical ecology is supported by National Institutes of Health Grant AI-02908.

# Pheromone-mediated gene expression in the honey bee brain

Christina M. Grozinger<sup>\*†‡§</sup>, Noura M. Sharabash<sup>†</sup>, Charles W. Whitfield<sup>†‡</sup>, and Gene E. Robinson<sup>†‡</sup>

<sup>\*</sup>Beckman Institute for Advanced Science and Technology, <sup>†</sup>Department of Entomology, and <sup>‡</sup>Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, IL 61801

We tested the hypothesis that queen mandibular pheromone (QMP) causes changes in gene expression in the brain of the adult worker honey bee, and that these changes can be correlated to the downstream behavioral responses induced by QMP. In support of the first hypothesis, cage experiments revealed that QMP transiently regulated expression of several hundred genes and chronically regulated the expression of 19 genes. Several of these genes were also affected by QMP in experiments with bee colonies in the field, demonstrating robust gene regulation by pheromone. To evaluate the second hypothesis, we focused on one function of QMP: delaying the transition from working in the hive (e.g., brood care, or "nursing") to foraging. We compared the list of QMP-regulated genes with the lists of genes differentially regulated in nurse and forager brains generated in a separate study. QMP consistently activated "nursing genes" and repressed "foraging genes," suggesting that QMP may delay behavioral maturation by regulating genes in the brain that produce these behavioral states. We also report here on an ortholog of the *Drosophila* transcription factor *kruppel homolog 1* that was strongly regulated by QMP, especially in the mushroom bodies of the bee brain. These results demonstrate chronic gene regulation by a primer pheromone and illustrate the potential of genomics to trace the actions of a pheromone from perception to action, and thereby provide insights into how pheromones regulate social life.

Many animal species, from insects to mammals, communicate via pheromones, chemicals that cause dramatic alterations in physiology and behavior. The recent identification of olfactory and pheromone receptors in *Drosophila melanogaster* and the mouse have provided new insights into how the olfactory system senses and encodes odorants (reviewed in refs. 1 and 2). It also has been demonstrated that pheromones affect gene expression in brain neurons, particularly with respect to immediate-early genes (3, 4). However, the molecular mechanisms by which pheromones are further transduced in the brain to influence behavior are only beginning to be understood (1). Particularly interesting are long-term changes in brain gene expression that might result from exposure to primer pheromones. These long-term changes in gene expression may be responsible for inducing long-term changes in physiology and behavior, a hallmark of primer pheromone action. Here we report on our efforts to use the honey bee (*Apis mellifera*) to study the effects of a primer pheromone on brain gene expression. We also have begun to correlate these gene expression changes with pheromone-mediated behavioral changes.

Honey bees show complex social organization that is controlled to a large extent by pheromones, many of which have been well characterized, both chemically and with respect to their specific behavioral effects (5–8). We studied the best understood bee pheromone, queen mandibular pheromone (QMP), a well-characterized blend that is part of a recently identified nine-component pheromone that attracts workers to attend the queen (9). QMP consists of five chemicals: (E)-9-keto-2-decenoic acid

(9-ODA), (R,E)-(-)- and (S,E)-(+)-9-hydroxy-2-decenoic acid (9HDA), methyl *p*-hydroxybenzoate (HOB), and 4-hydroxy-3-methoxyphenylethanol (HVA) (7). QMP plays many roles in social regulation (7). It prevents reproduction both by inhibiting worker ovary development (10) and the rearing of new queens (11). It may control brain development in the olfactory system of young bees (12). QMP also influences age-related division of labor among worker bees. Bees perform in-hive tasks such as brood care (nursing) for the first 2–3 weeks of their adult life and switch to foraging for food outside the hive in the last weeks of their life. Honey bees show many changes in physiology in association with this behavioral maturation, including structural changes in the mushroom bodies of the brain (which are sites of multimodal sensory integration and higher-order functions such as learning and memory), and differences in brain gene expression in nurses compared with foragers (13, 14). QMP delays honey bee behavioral maturation (15).

In this article we report on experiments that test two hypotheses: exposure to QMP causes changes in gene expression in the brain, and these changes correlate with the downstream behavioral effects of the pheromone. We used a recently developed honey bee brain microarray (16) to profile pheromone-induced changes in brain gene expression. Experiments were performed both with cage studies under controlled conditions in the laboratory and natural colony conditions in the field, to determine the robustness of the QMP effects. Because differences in gene expression associated with nursing and foraging behaviors have been analyzed extensively with microarrays (14), we used this information to explore our second hypothesis, that pheromone-regulated changes in gene expression are correlated with pheromone-induced changes in behavior. Finally, we began to characterize the effects of QMP on an ortholog of the *Drosophila* transcription factor *kruppel homolog 1* (*Kr-h1*) (17), a particularly promising candidate gene that emerged from our microarray analyses.

## Materials and Methods

**Genetic Manipulation.** Worker bees were derived from queens that were instrumentally inseminated with semen from a single, different drone, according to established procedures (18). Because male bees are haploid, the coefficient of relatedness

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: QMP, queen mandibular pheromone; *Kr-h1*, *kruppel homolog 1*; qRT-PCR, quantitative RT-PCR.

Data deposition: Gene expression data meet Minimum Information About a Microarray Experiment (MIAME) standards and have been deposited at ArrayExpress ([www.ebi.ac.uk/arrayexpress](http://www.ebi.ac.uk/arrayexpress)).

<sup>§</sup>To whom correspondence should be addressed. E-mail: [grozinge@staff.uiuc.edu](mailto:grozinge@staff.uiuc.edu).

© 2003 by The National Academy of Sciences of the USA

among offspring of such an instrumentally inseminated queen is 0.75. Bees were obtained from three “source colonies,” each headed by one of these queens, R8, R11, or R16. Bees from “genotype” R8 were used for the microarray experiments, and bees from R11 and R16 were used to confirm some of the microarray results with real-time quantitative RT-PCR (qRT-PCR), as described below.

**Rearing.** Colonies were maintained at the University of Illinois Bee Research Facility according to standard commercial procedures. To provide bees of known age, honeycombs containing late-stage pupae were removed from source colonies and placed in an incubator to emerge (33°C, 95% relative humidity).

For cage studies, bees were collected 16 h after eclosion and placed in small (10 × 10 × 7 cm) Plexiglas cages (35 bees per cage) for 8 h before pheromone exposure began. Bees were provided with water and food (45% honey, 45% pollen, 10% water). Cages were kept in a humidity-controlled (50% relative humidity), dark incubator at 33°C. Pheromone exposure consisted of 0.1 queen equivalent (the typical amount found in one queen) of QMP (QMP<sup>+</sup>) or a solvent control (QMP<sup>-</sup>) introduced on a glass slide. This dose is sufficient for inhibiting ovary development in caged worker bees (10). The duration of pheromone exposure was 1, 2, 3, or 4 days. For each replicate, one entire QMP<sup>+</sup> and one QMP<sup>-</sup> cage were flash-frozen in liquid nitrogen to prevent any additional changes in gene expression (19). Heads were removed and stored at -80°C until dissection. Bees were sampled at the same time of day to minimize any variation caused by circadian rhythms.

For field studies, the source colony was split into three colonies, allocating roughly equal quantities of adult bees, brood, and pollen and nectar stores. One colony retained the original queen (QR), one was left queenless (QL), and the third was given a strip that contained 10 queen equivalents (QMP<sup>+</sup>), a dose shown to mimic a live queen (20). The three colonies were transferred to a different apiary >2 miles away so they would not return to the site of their natal hive. Before the colony split, bees ( $n \approx 1,500$ ) were collected 0–36 h after eclosion, marked on the dorsal thorax with a paint dot (Testor’s Paint, Rockford, IL), and  $\approx 500$  were placed in each of the three colonies. Two days later, the marked bees were collected ( $n = 100$ ). Bees were collected into liquid nitrogen, and heads were stored as above.

For gene expression analysis in the mushroom bodies, foragers were collected from two colonies (R5, R11) upon return from foraging flights, and 1-day-old bees were collected <24 h after eclosion.

**Pheromones.** QMP was obtained from PheroTech (Delta, Canada). For cage experiments, 0.1 QMP was applied to glass slides in 10  $\mu$ l of 1% H<sub>2</sub>O/isopropyl alcohol and allowed to dry (9). QMP was applied to a fresh slide at the same time every day. For field experiments, commercially formulated QMP strips (Phero-Tech) were used and replaced daily (20).

**Confirmation of Pheromonal Activity.** To confirm QMP activity in cages, the effect of QMP on ovary development was assessed (10). Cages of bees were set up as described above and maintained for 10 days, then collected and stored at -20°C until dissection. The degree of ovary development was assessed according to Veltuis (21). Two cages from each of the three genotypes were analyzed. A total of 12 of 30 QMP<sup>-</sup> bees developed ovaries (with developed eggs and follicle cells), whereas 0 of the 30 QMP<sup>+</sup> bees developed ovaries ( $P < 0.01$ ,  $\chi^2$  test), demonstrating that the QMP was active and the bees were sensitive to it.

To confirm QMP activity in the field, the effect of QMP on the number of queen cells built was determined. Queen cells are distinctive large cells on the honeycomb built by worker bees to

rear new queens (8); QMP partially inhibits this behavior (11). The number of queen cells built in QL colonies was 31, 29, and 12 (R8, R16, and R11, respectively) compared with 1, 5, and 5 queen cells built in the QMP<sup>+</sup> colonies. In QR colonies, no queen cells were built in R8 and R16, whereas three were built in R11. QMP clearly suppressed queen cell-building behavior in the field.

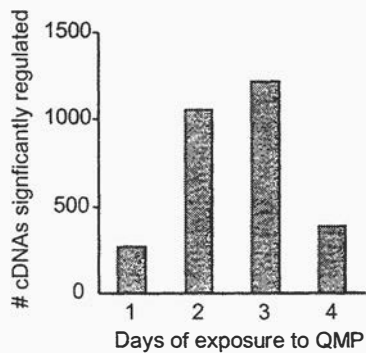
**Brain Dissection.** Whole bee heads were partially lyophilized to facilitate brain dissection (22). Dissections were performed over dry ice so tissue never thawed. Because ocelli and the subesophageal ganglion frequently fractured during dissection, these were removed during all dissections, while the remainder of the brain was included. Mushroom bodies were dissected from freeze-dried brains as in ref. 22.

**Microarrays and Data Analysis.** The microarray contained  $\approx 9,000$  cDNAs, representing  $\approx 7,600$  different genes, 40% of which have been annotated primarily by using comparisons to *Drosophila* genes and the molecular function classification scheme of the Gene Ontology Consortium (16). This number is estimated to account for 50% of the genes in the honey bee genome, based on comparisons with the *Drosophila* genome.

For cage studies, direct competitive hybridization comparisons were made for matched samples of QMP<sup>+</sup> and QMP<sup>-</sup> bees. Each sample consisted of 10 brains pooled (10 bees taken randomly from a cage of 35 bees after the entire cage was killed). Eight arrays were analyzed for each time point (1, 2, 3, or 4 days of pheromone exposure), a total of 32 microarrays. The eight replicates were comprised of four biological replicates (different cages) and four technical replicates. For field studies, each sample consisted of 10 brains pooled. Samples from the QR, QMP<sup>+</sup>, and QL colonies split from the R8 source colony were compared with each other in a loop design that used three microarrays, with four technical replicates, for a total of 12 arrays (23).

Dissected bee brains were homogenized in Trizol (Invitrogen Life Technologies), and RNA was extracted according to the manufacturer’s protocols. mRNA was then purified by using an Oligotex mRNA kit (Qiagen, Valencia, CA). cDNA synthesis and hybridization to microarrays followed protocols modified from ref. 24 and described in ref. 14. Arrays were scanned with GENEPIX software and normalized to the median intensity with regional and intensity-dependent Lowess normalization. Dyes used to label the QMP<sup>+</sup> and QMP<sup>-</sup> samples were reversed in half of the replicates to control for dye-by-gene interactions. The following filtering protocols were followed, as in Whitfield *et al.* (14). cDNAs with average expression intensities across each set of eight arrays <350 or absent from  $\geq 1$  array were removed from the analysis. cDNAs expressed at equal levels in hypopharyngeal glands in the honey bee head (14) were removed because material from these glands can sometimes contaminate dissected brains. Only cDNAs expressed at all time points and also in the Whitfield *et al.* (14) nurse vs. forager analysis were analyzed, to facilitate comparisons across the two studies. This left 6,360 cDNAs, representing  $\approx 5,000$  unique genes.

Bayesian analysis (25) was used to determine which cDNAs were significantly regulated by QMP. This method generates an estimated mean and 95% confidence interval (CI) for the relative level of expression of each cDNA for each experimental group of bees. cDNAs were classified as differentially expressed if their 95% CIs did not overlap between groups (one-tailed  $t$  test). This approach enables identification of small, but reproducible, differences in gene expression (26, 27). Notably, the expression of most significantly regulated genes changed by  $\approx 10\%$ . For example, of 2,607 cDNAs found to be significantly regulated during the time course, only 1,158 showed differences of >10%, and only 111 showed differences of >20%.



**Fig. 1.** Dynamic regulation of gene expression in the honey bee brain by QMP. Bees were maintained in cages in the laboratory either with or without QMP. The number of cDNAs showing significant differences in expression on each day is shown.

**mRNA Quantification by Real-Time qRT-PCR.** Confirmation of some of the results obtained from microarray analysis was performed with real-time qRT-PCR in individual brains (19) with an ABI Prism 7900 sequence detector and the SYBR green detection method (Applied Biosystems). *Rp49* or *eIF3-S8*, two housekeeping genes that did not vary in expression levels on these microarrays, were used as loading controls. Quantification was as described (19). The sequences for the primers used are given in Table 4, which is published as supporting information on the PNAS web site.

## Results

**QMP Influences Brain Gene Expression in the Laboratory.** There were 1,210 cDNAs with significantly higher expression and 1,397 cDNAs with significantly lower expression in the brains of QMP<sup>+</sup> bees relative to QMP<sup>-</sup> bees. This is about eight times more than the number of false positives expected (25). Several hundred cDNAs were significantly regulated on each day of the 4-day time course, with maximum numbers on days 2 and 3 and the fewest on day 1 (Fig. 1). The magnitude differences were smallest on day 1 as well: expression of only 3 ESTs changed by 40% on day 1, whereas 21, 20, and 21 ESTs, respectively, were regulated by 40% on the other three days. Expression levels for most cDNAs changed during the time course, but a small subset was chronically regulated across days 2, 3, and 4 (13 up-regulated

and 6 down-regulated). Eight of these chronically regulated genes are annotated, with high sequence similarity to known *Drosophila* or human genes (Table 1). They include the transcription factor *Kr-h1* (28), the G protein-coupled receptor *frizzled 2* (29), and a transmembrane protein involved in regulating glial cell thickness (*push/poe*) (30). See Table 5, which is published as supporting information on the PNAS web site, for a complete list of regulated genes.

**QMP-Mediated Gene Expression Correlates with Downstream Behavioral Effects.** If QMP regulation of gene expression in the brain is related to the QMP-mediated delay in the transition from hive work to foraging (15), then the following pattern is predicted: exposure to QMP activates genes in the brain associated with nursing and represses genes associated with foraging. The prediction was tested by comparing our data with results from a previous study (14). That study showed widespread differences in gene expression between nurses and foragers, with 1,360 showing significantly more mRNA abundance in nurse brains (“nurse list”) and 1,310 significantly more mRNA abundance in forager brains (“forager list”). Results of this comparison agree with the prediction. A significantly larger proportion of QMP-up-regulated cDNAs was on the nurse list than were QMP-down-regulated cDNAs (Table 2). Likewise, a significantly larger proportion of QMP-down-regulated cDNAs were on the forager list than were QMP-up-regulated cDNAs. (Table 2). These trends were even more striking if only the chronically regulated cDNAs are considered (Table 2). See Table 5 for a complete listing of all QMP-regulated genes that are also on the nurse or foragers lists.

**QMP Influences Brain Gene Expression in Bee Colonies.** There were 697 cDNAs that showed significant differences in brain expression in bees sampled from QMP<sup>+</sup> and QL colonies (Fig. 2). Overall, these expression differences were most similar to the results from cages sampled on day 3. For example, of the cDNAs significantly up-regulated in the field, 59 were also significantly up-regulated in the cages (vs. 20 that were down-regulated in the cages), and of the cDNAs down-regulated in the field, 74 were down-regulated in cages (vs. 7 that were up-regulated in the cages). Relative to cage bees, there were fewer cDNAs with significant differences in colony bees, and they showed expression differences of smaller magnitude. Among the genes significantly regulated in the brain by QMP in both cage and colony

**Table 1. Annotated genes chronically regulated by QMP in the honey bee brain**

Closest <i>Drosophila</i> (or human) match	Similarity score	Fold difference QMP <sup>+</sup> /QMP <sup>-</sup>	Possible function	Nurse/forager
<i>Q9H2Y7</i> (human)	2E-08	1.11	WD-repeat protein	Nurse
<i>CG14168</i>	1E-13	1.12	PDZ domain	Nurse
<i>clt</i>	5E-23	1.16	Carboxylesterase	
<i>Traf1</i>	1E-24	1.16	Tumor necrosis factor receptor-associated factor	Forager
<i>CG7474</i>	4E-85	1.17	Tubuliny-tyrosine ligase	
<i>poe</i>	9E-34	0.86	Calmodulin binding, synpatogenesis	
<i>frizzled 2</i>	5E-96	0.70	G protein-coupled receptor	Forager
<i>kr-h1</i>	1E-100	0.59	Zn-finger transcription factor	Forager

Genes found to be chronically regulated by QMP with high sequence similarity to known *Drosophila* or human genes are listed below. Similarity score: BLAST E value (this corresponds to the EST sequence only, except for *Kr-h1*, for which full sequence was obtained, and *frizzled 2*, for which additional sequence was identified from the genome project, <http://hgsc.bcm.tmc.edu/blast/?organism=Ameliffiera>). Fold difference column: the maximum difference between values obtained for bees maintained in cages with or without QMP. Nurse/forager column indicates in which behavioral group the gene was found to be significantly upregulated in an independent microarray-based study (14).

**Table 2. QMP activates genes in the bee brain associated with nursing and represses genes associated with foraging**

Location	QMP	No. of cDNAs	% on nurse list	% on forager list
Cage	Total QMP ↓	1,397	14	27
	Total QMP ↑	1,210	33	12
	Chronically ↓	6	0	67
	Chronically ↑	13	62	8
Colony	QMP ↓	374	13	26
	QMP ↑	323	42	16
Cage and colony	QMP ↓	59	15	42
	QMP ↑	74	53	3

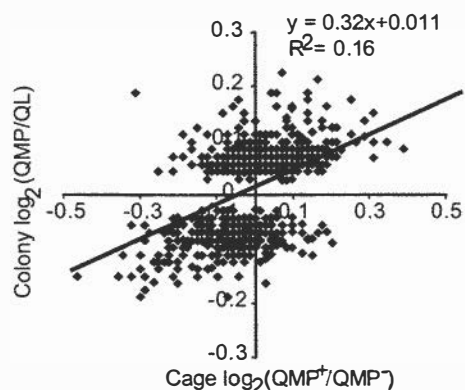
QMP-regulated cDNAs were compared to the lists of cDNAs significantly up-regulated in nurse bees or up-regulated in foragers. A significantly larger proportion of QMP ↑ (up-regulated) cDNAs were on the nurse list than were QMP ↓ cDNAs. Likewise, a significantly larger proportion of QMP ↓ cDNAs were on the forager list than were QMP ↑ cDNAs ( $P < 0.001$ ,  $\chi^2$  tests). Genes chronically regulated by QMP showed the same trends.

were: transcription factors *HLH3B* and *klumpfuss*, gap junction proteins *inx2* and *inx3*, *Tor*, a phosphatidylinositol 3-kinase, *Nrv2*, an  $\text{Na}^+/\text{K}^+$  exchanging ATPase complex, *RacGAP*, a GTPase activator, and *Rab6*, a GTPase.

As was the case in the laboratory experiments, a significantly larger proportion of QMP-up-regulated cDNAs were on the nurse list than were QMP-down-regulated cDNAs, and likewise for QMP-down-regulated cDNAs on the forager list (Table 2). These trends were even more striking if only the cDNAs regulated in both the cages and colonies are considered (Table 2).

The field experiments also allowed us to compare the effects of a live queen to the effects of QMP on gene expression in the brain. A total of 1,047 cDNAs were differentially expressed between the QR and QMP colonies; thus, QMP did not completely mimic the live queen. A total of 335 cDNAs were coregulated in the QR and QMP colonies relative to the QL colony; these genes may be specifically regulated by QMP released by the queen.

**QMP Effects on Transcription Factor.** The gene with the biggest difference in mean expression levels in the cage experiments was



**Fig. 2.** Comparison of QMP-regulated brain gene expression in bees from colonies in the field vs. cages in the laboratory. The cDNAs showing significant differences in expression in bees from QMP vs. QL (queenless) colonies are shown (minus three outliers,  $n = 694$ ). The expression levels (log-transformed) of these cDNAs under colony and cage (day 3) conditions are plotted, and the regression line is shown. In general, cDNAs down-regulated in colony bees were down-regulated in cage bees, and likewise for up-regulated cDNAs. Expression differences were generally lower in colony bees.

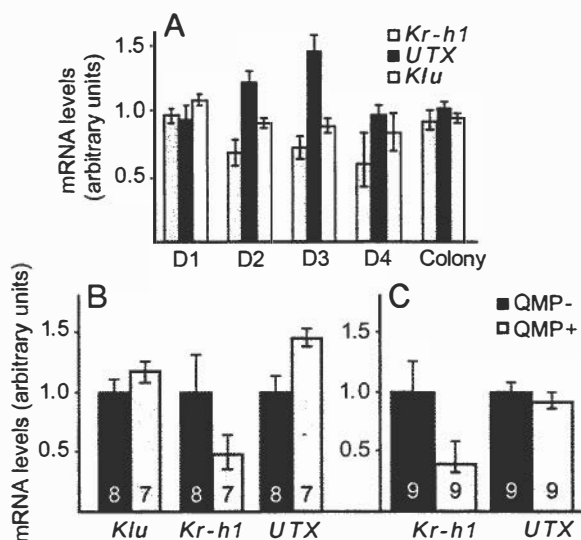
**Table 3. Effects of QMP on different functional categories of genes**

Functional category	No. annotated	% Regulated
Protein phosphatases	35	2.9
Synaptic vesicle transport	66	1.5
Ion channels	73	6.8
Protein kinases	87	8.0
Cytoskeletal organization and biogenesis	89	4.5
Receptors	106	5.7
Peptidases	108	5.6
RNA-binding proteins	127	3.9
Oxidoreductases	129	8.5
Transcription factors	129	13.2

For each functional category, the number of genes significantly regulated by QMP (>10%) for at least 1 day in cage experiments are shown. Categories of molecular function are from the Gene Ontology Consortium; annotation is from ref. 16.

a transcription factor, *Kr-h1*. Because transcription factors regulate the expression of other genes, they may serve as markers for regulation of pheromone-mediated “transcriptional programs” that control downstream behavioral effects. A total of 39 of the 129 cDNAs annotated as transcription factors (16) were significantly regulated by exposure to QMP (Table 6, which is published as supporting information on the PNAS web site), whereas 17 were significantly regulated by >10%. This proportion was relatively high compared with other functional categories of genes (Table 3).

Microarray analysis showed that brain expression of three transcription factors (*Kr-h1*, *Utx*, and *klumpfuss*) was significantly affected by QMP over more than 1 day (Fig. 3A). In the colony experiments, only *Klumpfuss* was significantly regulated. The cage results were verified with qRT-PCR, using individual bee



**Fig. 3.** Confirmation of microarray results. mRNA levels for three transcription factors (*kr-h1*, *utx*, and *klu*) shown by microarray analysis to be significantly regulated by QMP were quantified by real-time qRT-PCR in individual brains. (A) Expression levels (QMP<sup>+</sup>/QMP<sup>-</sup> ratio) determined by microarray analysis on days (D) 1–4 for cage bees and colony bees. (B) Expression levels determined by qRT-PCR for bees exposed to QMP for 3 days from the same genotype as in A. Number of individually analyzed brains is indicated on bars. Data are means  $\pm$  SE (converted to the same arbitrary scale as the mean), normalized to the QMP<sup>-</sup> sample. Statistical analysis was done by one-tailed t test. (C) Same as B, except bees were from a second genotype.

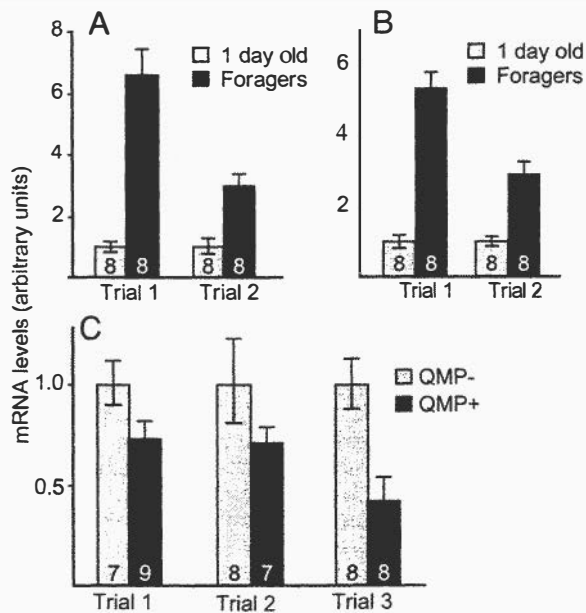


Fig. 4. Behavior- and QMP-related *Kr-h1* expression in whole brains and mushroom bodies in the honey bee. mRNA levels of *Kr-h1* in 1-day-old bees and foragers were quantified with qRT-PCR in whole brains (A) and mushroom bodies (B) in two genotypes (trials 1 and 2). Differences were significant ( $P < 0.001$ , one-tailed *t* test). (C) Effect of QMP on *Kr-h1* mRNA levels in mushroom bodies. Differences were significant in trial 1 ( $P < 0.05$ ) and trial 3 ( $P < 0.01$ ) but not in trial 2 ( $P = 0.09$ ). Other notations are as in Fig. 3.

brains rather than samples of pooled brains (Fig. 3B and C). In genotype R8 (Fig. 3B), which was the same as that used for the microarray studies, brain expression of *Kr-h1* and *Utx* was significantly decreased and increased, respectively, in QMP<sup>+</sup> bees, as in the microarray analyses, whereas *klumpfuss* levels were not significantly different. In a second genotype (Fig. 3C), *Kr-h1* brain expression again was significantly lower in QMP<sup>+</sup> bees, but there was no effect of QMP on *Utx* expression levels. Because *Utx* was up-regulated for only 2 days in the microarray experiments, it is possible that its temporal pattern of QMP-mediated expression was different in this genotype. Regardless, *Kr-h1* clearly showed the most robust and consistent pattern of QMP regulation.

Two additional experiments were conducted to further establish *Kr-h1* as a candidate gene to study long-term, behaviorally relevant changes in brain gene expression caused by pheromone exposure. First, we determined whether QMP effects on *Kr-h1* expression in the brain were associated with behaviorally related differences in expression. Previous microarray studies indicated that *Kr-h1* expression is higher in foragers relative to nurses (14). qRT-PCR analysis agreed, showing that brain *Kr-h1* expression was increased in foragers relative to 1-day-old bees (Fig. 4A). Second, we determined whether *Kr-h1* is expressed in the mushroom bodies, as a first step toward exploring the effects of pheromone on gene expression in brain regions involved in higher-order integration and processing. *Kr-h1* was expressed in the mushroom bodies, with expression significantly higher in foragers relative to 1-day-old bees (Fig. 4B) and also significantly higher in QMP<sup>-</sup> bees than QMP<sup>+</sup> bees (Fig. 4C). *Kr-h1* expression levels were not significantly different in optic lobes of QMP<sup>-</sup> vs. QMP<sup>+</sup> bees (data not shown), suggesting that expression of this gene is specifically regulated in the mushroom bodies rather than the whole brain. Cloning *A. mellifera kr-h1* (GenBank accession no AY338499, see Fig. 5, which is published as supporting information on the PNAS web site) revealed that it contains eight zinc finger domains from amino acids 62–289 and

has significant sequence similarity to the *Drosophila* *Kr-h1* protein and the partial *Anopheles gambiae* protein. There are two transcripts for *Kr-h1* present in *Drosophila*, *Kr-h1 RA* and *RB* (28). The *RA* transcript is present in larvae and adults, whereas the *RB* transcript is present only in embryos and has an additional 55 aa on the N terminal. The cloned *A. mellifera kr-h1* does not contain this N-terminal extension, and thus matches the *RA* transcript most closely. However, it is  $\approx 200$  aa shorter than the *RA*, because of deletions throughout the gene.

## Discussion

This study has demonstrated that QMP causes changes in expression levels of many genes in the brain of adult honey bees, and that these changes correlate with some of the downstream behavioral effects of the pheromone. QMP effects on gene expression were detected both in a controlled laboratory environment and in bee colonies in the field, which represent a more natural environment. QMP causes transient changes in expression of several hundred genes, but causes chronic changes in only a small subset of genes in the brain.

The effects of QMP on brain gene expression changed over time. One possible explanation for this is that the transcriptional response to QMP is dynamic, with one set of genes activating/repressing expression of another downstream set of genes in “transcriptional cascades” reminiscent of what occurs during *Drosophila* neural development (31). These genes may only need to act transiently to produce the necessary response to queen pheromone, such as neuronal remodeling of pheromone-responsive networks. For example, bees raised in queenless conditions for 4 days have smaller synaptic structures in the antennal lobe, suggesting that QMP may influence synaptic growth and remodeling during this time frame (12). Thus, the dynamic nature of the response to QMP might be related to the fact that the bee’s olfactory system continues developing for the first few days of adult life (32).

Alternatively, it simply might be difficult to detect small, but statistically significant, differences in gene expression across consecutive days, even with the relatively large number of replicates used in this study. In most cases, the observed significant changes in gene expression were relatively subtle (<20%). Perhaps the relatively small QMP-induced expression differences are caused by effects of QMP on a small set of neurons. Small, but statistically significant, differences in gene expression are detectable in microarrays studies such as ours with numerous replicates (14, 27) and can have biological significance (26, 27).

Some genes chronically regulated by QMP may be involved in stably altering neuronal activity and responsiveness. For example, one of the chronically down-regulated genes encodes a homolog of the *Drosophila* protein Pushover (also named Poe), which has been shown to play a role in regulating glial cell thickness (30) and neuronal excitability (33). Such alterations in central brain processing regions could lead to changes in responsiveness to various stimuli, thereby altering behavior (34).

There was considerable overlap in the genes regulated by QMP in the cages and in colonies, but gene expression changes in the colonies were smaller and involved fewer genes. This might be because the “QMP strip” used in the colony releases a lower amount of pheromone per bee than the glass slide in the cages; thus, the colony bees may have been exposed to an almost 2-fold lower dose of QMP (K. N. Slessor, personal communication). Another factor is that the colony environment is more varied with respect to pheromonal and other stimuli. In particular, the brood in the colonies release a pheromone that produces many of the same effects as QMP (35–38), and this may have attenuated the response to QMP in the colony experiment. If so, then the set of genes that were found to be regulated in both the colony and the cage experiments may be robustly regulated by or



only sensitive to QMP. Further cage experiments using brood pheromone can help address these questions.

QMP and a live queen had similar effects on some, but not all, genes. This finding is consistent with observations showing that QMP is not as effective as a full queen extract in producing behavioral and physiological responses. For example, an additional four components were recently identified that improve the performance of synthetic queen pheromone in the attraction of workers to the queen, and even this nine-component blend is not quite as effective as the full extract (9). The genes that were specifically regulated only under queenright conditions may be responding to other unidentified components of queen pheromone and may prove to be reliable markers for future elucidation of a more complete pheromone blend.

QMP consistently activated genes correlated with nursing behavior and repressed genes correlated with foraging behavior, in both cage and colony experiments. This finding suggests that the effects of QMP on the timing of the transition from working in the hive to foraging (15) may be caused by QMP regulation of genes in the brain that produce these behavioral states. At least some QMP-regulated genes may be involved in regulating the transitions to nursing and foraging, rather than the maintenance of these states, because the bees we used were too young or simply unable to perform these behaviors in laboratory cages. The fact that many QMP-regulated genes were not found to be associated with nursing or foraging behavior may reflect the fact that QMP also is involved in the regulation of several other behavioral and physiological processes besides age-related division of labor, such as inhibiting ovary development and the rearing of new queens. Further microarray experiments may identify distinct transcriptional programs that relate to other pheromone-regulated processes.

The proportion of transcription factors regulated by QMP was relatively high compared with other functional groups of genes. This result suggests transcription factors may be important targets of pheromone activation. To aid in the future identification of specific transcriptional programs, we catalogued the transcription factors whose expression levels were significantly regulated by QMP. Only one transcription factor was found to be chronically regulated; most showed significant changes in expression on only 1 day. Transiently regulated transcription factors may simply initiate a downstream program and be subsequently turned off in a matter of hours as is the case for CREB (cyclic AMP-response element binding protein), which initiates changes in neuronal plasticity (reviewed in refs. 39 and 40).

*Kr-h1* was the most highly and robustly regulated gene identified in this study. Furthermore, *Kr-h1* has the unique trait of being chronically regulated by a pheromone. Kruppel homologs are zinc finger transcription factors that play important roles in orchestrating development and cell differentiation (reviewed in (17), including neural development (31). *Kr-h1* in particular was

identified as an ecdysone-sensitive transcript during *Drosophila* morphogenesis (28), and in a gain-of-function screen for genes involved in motor axon guidance and synaptogenesis in *Drosophila* larvae (41). Differences in expression patterns and sequence divergences within the DNA-binding regions of Kruppel-like proteins across mammalian lineages suggest that members of this family have undergone duplication and acquired novel functions in different species during evolution (42).

The *A. mellifera* *Kr-h1* protein is highly similar to *Drosophila* *Kr-h1* and the *Anopheles gambiae* ortholog, suggesting that it functions similarly in all three insects. However, in the case of the honey bee, a species with a highly derived form of social organization, it has evolved to be regulated by a pheromone, perhaps in addition to intrinsic factors common to the dipteran species with solitary lifestyles. QMP regulation of *Kr-h1* occurs in the mushroom bodies, the sites of integration of sensory information, so *Kr-h1* may be involved in organizing stable changes in gene expression and neuron structure that are necessary to transduce the chemosensory QMP stimulus to downstream changes in behavior and physiology. Further experiments will be necessary to map *Kr-h1* expression throughout the brain to further elucidate the function of *Kr-h1* in response to pheromone. Given that *Kr-h1* is strongly down-regulated by QMP, it is puzzling that *Kr-h1* is repressed by QMP in young honey bees but up-regulated in older bees, because both groups were taken from colonies with a queen. Perhaps this is because foragers typically contact the queen less than do younger bees, or foragers are simply less responsive to QMP. Further experiments will be needed to explore this issue.

QMP affects the expression of many genes in the bee brain, and in particular changes expression of genes associated with behaviors that also are regulated by this pheromone. Some effects on gene expression are relatively transient, which may reflect short-term modifications in synaptic plasticity, and some are more chronic, which may reflect long-term changes in neuronal responsiveness and behavioral state. Some of the genes identified here may represent transcriptional programs regulating particular pheromone-mediated physiological and behavioral processes. Further genomewide analysis of the transcriptional effects of QMP and other pheromones should lead to new insights into how pheromones regulate social life.

We thank K. Pruiett for expert assistance with the bees; A. Barron, C. Schook, and L. Wraight for technical assistance; R. Velarde for samples of 1-day-old and forager bees; S. Rodriguez-Zas for help with statistical analysis; M. Band for fabricating microarrays of outstanding quality; C. Keeling, S. Hoover, K. N. Slessor, and M. L. Winston for helpful advice; and R. Maleszka and members of the Robinson laboratory for comments that improved the manuscript. This work was supported by a Beckman Institute postdoctoral fellowship (to C.M.G.), a National Science Foundation Bioinformatics postdoctoral fellowship (to C.W.W.), and grants from the Burroughs Wellcome Trust, the National Institutes of Health, and the U.S. Department of Agriculture (to G.E.R.).

1. Dulac, C. & Torello, A. T. (2003) *Nat. Rev. Neurosci.* **4**, 551–562.
2. Vosshall, L. B. (2001) *Chem. Senses* **26**, 207–213.
3. Halem, H. A., Baum, M. J. & Cherry, J. A. (2001) *J. Neurosci.* **21**, 2474–2480.
4. Brennan, P. A., Schellinck, H. M. & Keverne, E. B. (1999) *Neuroscience* **90**, 1463–1470.
5. Free, J. B. (1987) *Pheromones of Social Bees* (Cornell Univ. Press, Ithaca, NY).
6. Slessor, K. N., Kaminski, L.-A., King, G. G. S., Borden, J. H. & Winston, M. L. (1988) *Nature* **332**, 354–356.
7. Winston, M. L. & Slessor, K. N. (1998) *Apidologie* **29**, 81–95.
8. Winston, M. L. (1987) *The Biology of the Honey Bee* (Harvard Univ. Press, Cambridge, MA).
9. Keeling, C. I., Slessor, K. N., Higo, H. A. & Winston, M. L. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 4486–4491.
10. Hoover, S. E. R., Keeling, C. I., Winston, M. L. & Slessor, K. N. (2003) *Naturwissenschaften* **90**, 477–480.
11. Winston, M. L., Higo, H. A., Colley, S. J., Pankiw, T. & Slessor, K. N. (1991) *J. Insect Behav.* **4**, 649–660.
12. Morgan, S. M., Butz Huryn, V. M., Downes, S. R. & Mercer, A. R. (1998) *Behav. Brain Res.* **91**, 115–126.
13. Robinson, G. (2002) *Am. Nat.* **160**, S160–S172.
14. Whitfield, C. W., Cziko, A.-M. & Robinson, G. (2003) *Science* **302**, 296–299.
15. Pankiw, T., Huang, Z.-Y., Winston, M. L. & Robinson, G. E. (1998) *J. Insect Physiol.* **44**, 685–692.
16. Whitfield, C. W., Band, M. R., Bonaldo, M. F., Kumar, C. G., Liu, L., Pardinas, J. R., Robertson, H. M., Soares, M. B. & Robinson, G. E. (2002) *Genomè Res.* **12**, 555–566.
17. Kaczynski, J., Cook, T. & Urrutia, R. (2003) *Genome Biol.* **4**, 206.1–206.8.
18. Laidlaw, J., H. H. (1987) *Bee World*, 17–38.
19. Ben-Shahar, Y., Robichon, A., Sokolowski, M. B. & Robinson, G. E. (2002) *Science* **296**, 741–744.
20. Ledoux, M. N., Winston, M. L., Higo, H. A., Keeling, C. I., Slessor, K. N. & Le Conte, Y. (2001) *Insectes Soc.* **48**, 14–20.
21. Veltuis, H. H. W. (1970) *Entomol. Exp. Appl.* **13**, 377–394.
22. Schulz, D. J. & Robinson, G. E. (1999) *J. Comp. Physiol. A* **184**, 481–488.

23. Kerr, M. K. & Churchill, G. A. (2001) *Genet. Res.* **77**, 123–128.
24. Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J. E., Snesrud, E., Lee, N. & Quackenbush, J. (2000) *BioTechniques* **29**, 548–550, 552–554, 556.
25. Townsend, J. P. & Hartl, D. L. (2002) *Genome Biol.* **3**, 0071.1–0071.6.
26. Yan, X. D., Hanson, A. J., Nahreini, P., Koustas, W. T., Andreatta, C. & Prasad, K. N. (2002) *In Vitro Cell Dev. Biol. Anim.* **38**, 529–537.
27. Gibson, G. (2002) *Mol. Ecol.* **11**, 17–24.
28. Pecasse, F., Beck, Y., Ruiz, C. & Richards, G. (2000) *Dev. Biol.* **221**, 53–67.
29. Bhanot, P., Brink, M., Samos, C. H., Hsieh, J. C., Wang, Y., Macke, J. P., Andrew, D., Nathans, J. & Nusse, R. (1996) *Nature* **382**, 225–230.
30. Yager, J., Richards, S., Hekmat-Safe, D. S., Hurd, D. D., Sundaresan, V., Caprette, D. R., Saxton, W. M., Carlson, J. R. & Stern, M. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 10445–10450.
31. Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. (2001) *Cell* **106**, 511–521.
32. Pham-Delengue, M. H., Trouiller, J., Caillaud, C. M., Roger, B. & Masson, C. (1993) *Apidologie* **24**, 267–281.
33. Richards, S., Hillman, T. & Stern, M. (1996) *Genetics* **142**, 1215–1223.
34. Beshers, S. N., Robinson, G. E. & Mittenthal, J. E. (1999) in *Information Processing in Social Insects*, ed. Detrain, C. (Birkhäuser, Boston), pp. 115–139.
35. Arnold, G., Le Conte, Y., Trouiller, J., Herve, H., Chappe, B. & Masson, C. (1994) *C. R. Acad. Sci. Ser. III* **317**, 511–515.
36. Le Conte, Y., Areski, M. & Robinson, G. E. (2001) *Proc. R. Soc. London Ser. B* **268**, 1–6.
37. Mohammedi, A., Crauser, D., Paris, A. & Le Conte, Y. (1996) *C. R. Acad. Sci. Ser. III* **319**, 769–772.
38. Mohammedi, A., Paris, A., Crauser, D. & Le Conte, Y. (1998) *Naturwissenschaften* **85**, 455–458.
39. Clayton, D. F. (2000) *Neurobiol. Learn. Mem.* **74**, 185–216.
40. West, A. E., Griffith, E. C. & Greenberg, M. E. (2002) *Nat. Rev. Neurosci.* **3**, 921–931.
41. Kraut, R., Menon, K. & Zinn, K. (2001) *Curr. Biol.* **11**, 417–430.
42. Shannon, M., Hamilton, A. T., Gordon, L., Branscomb, E. & Stubbs, L. (2003) *Genome Res.* **13**, 1097–1110.

# *Drosophila Gr5a* encodes a taste receptor tuned to trehalose

Sylwester Chyb\*, Anupama Dahanukar†, Andrew Wickens\*, and John R. Carlson†\*

\*Imperial College London, Wye Campus, Kent TN25 SAH, United Kingdom; and †Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520

Recent studies have suggested that *Drosophila* taste receptors are encoded by a family of G protein-coupled receptor genes comprising at least 56 members. One of these genes, *Gr5a*, has been shown by genetic analysis to be required by the fly for behavioral and sensory responses to a sugar, trehalose. Here, we show that *Gr5a* is expressed in neurons of taste sensilla located on the labellum and legs. Expression is observed in most if not all labellar sensilla and suggests that many taste neurons express more than one receptor. We demonstrate by heterologous expression in a *Drosophila* S2 cell line that *Gr5a* encodes a receptor tuned to trehalose. This is the first functional expression of an invertebrate taste receptor.

The majority of taste sensilla in *Drosophila* are located on the labellum, a gustatory organ of the proboscis, and the tarsal segments of the legs. Most sensilla house four gustatory neurons, one sensitive to sugars, one strongly sensitive to salt, one weakly sensitive to salt, and one responsive to water, as well as one mechanosensory neuron (1, 2). Previous work has identified a large family of G protein-coupled receptor genes, the *Gr* genes, which have been proposed to encode gustatory receptors in *Drosophila* (3). The *Gr* family comprises at least 56 members, many of which are expressed in subsets of taste neurons in the different taste organs (3–5).

Perception of sugars is a critical taste modality in animals such as fruit flies that ingest sweet substances for nutrition. Among the sugars, one that plays a particularly important role in insects is the disaccharide trehalose [1-*O*-( $\alpha$ -D-glucopyranosyl- $\alpha$ -D-glucopyranose)], which is composed of two glucose molecules connected by an unusual 1,1' linkage. Also called mycose, this disaccharide is abundant in yeasts and fungi, which are present in fermented fruit, an important food source of *Drosophila*. Trehalose is especially critical to insects in that it is the principal sugar found in the hemolymph, where it is involved in regulation of osmolarity, and in many winged insects it is an easily transported and accessible energy source metabolized during flight (6).

Previous genetic studies have identified a locus on the X chromosome called *Tre*, whose alleles confer differing levels of sensitivity to trehalose (7, 8). Recently we have shown that a member of the *Gr* family, *Gr5a*, maps to this locus and is necessary for trehalose response *in vivo* (9). Deletion mutants of *Gr5a* have a greatly diminished response to trehalose when assayed by electrophysiological recordings from single taste sensilla, or by a behavioral test. This defect was rescued by reintroducing a functional copy of *Gr5a* on a transgene, but not by introducing a mutant copy of *Gr5a*. Consistent with our results, Ueno *et al.* (10) showed that polymorphisms in *Gr5a* correlate with trehalose responses. However, these results do not exclude the possibility that *Gr5a* plays an indirect role in trehalose response.

Here we show that *Gr5a* is expressed in taste neurons of the labellum and the tarsi, supporting its identity as a taste receptor

gene. We then provide direct evidence that *Gr5a* encodes a trehalose receptor by expressing it in *Drosophila* S2 cells and showing that stimulation with trehalose evokes changes in intracellular calcium ( $Ca^{2+}$ ) levels in these cells. We further show that *Gr5a* is narrowly tuned to trehalose, showing little if any response to other disaccharides.

## Methods

**Expression Analysis.** An 8.5-kb fragment upstream of *Gr5a* was amplified from genomic DNA of Canton-S flies and inserted upstream of the GAL4-coding sequence in the pG4PN vector (C. Warr and J.R.C., unpublished data). Transgenic flies were generated by using standard procedures. Heterozygous flies carrying one copy each of *Gr5a-GAL4* and *UAS-lacZ* were stained for LacZ activity. For visualization of GFP, flies carrying two copies of *Gr5a-GAL4*, as well as two copies of *UAS-mCD8:GFP*, were examined by using confocal microscopy.

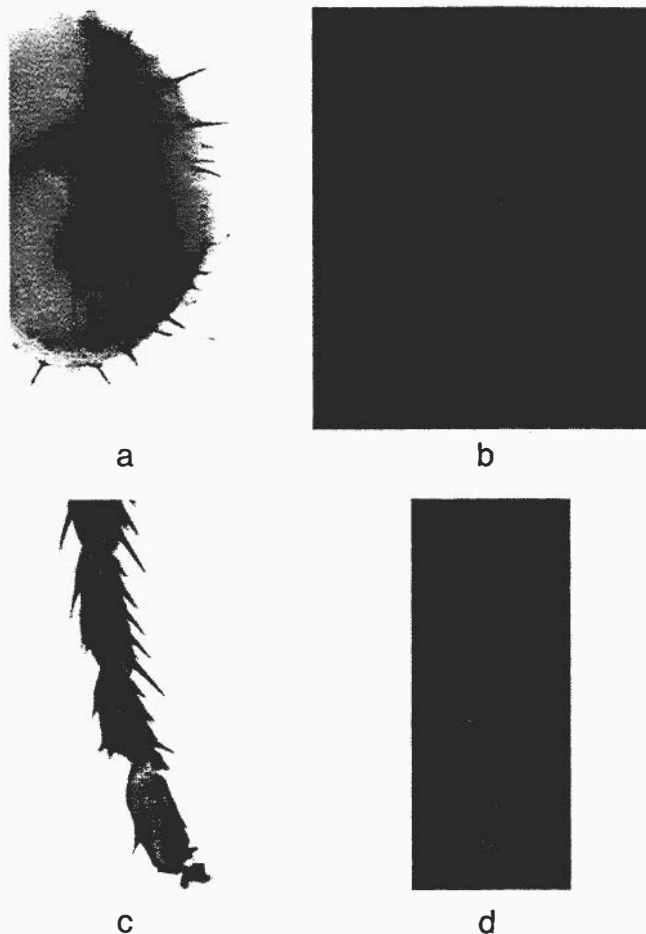
**Heterologous Expression.** A 1.2-kb fragment of full-length *Gr5a* cDNA was amplified from head mRNA by using the Smart RACE cDNA Amplification kit (Clontech). This fragment was inserted into a unique *EcoRI* site in the pRmHa3 vector (11). *Drosophila* S2 cells were grown at 25°C in Shields and Sang M3 insect medium supplemented with 10% FCS and antibiotic/antimycotic solution. S2 cells were cotransfected with pRmHa3-*Gr5a* and pIZT-V5/His vector (Invitrogen) encoding GFP and zeocin-resistance to facilitate recognition of transfectants and antibiotic selection, at a 3:1 ratio by using liposomal formulation (CellFectin, Invitrogen) according to the manufacturer's instructions. Expression of *Gr5a* was induced by adding 0.6 mM  $Cu^{2+}$  to the cell culture media 48 hrs before an experiment. Levels of *Gr5a* expression in uninduced and induced cells were examined by RT-PCR to confirm the induction of *Gr5a*.

**Calcium Imaging.** Transfected cells were grown in 96-well plates and loaded with membrane-permeable fura 2-acetoxymethyl ester (fura 2-AM) as described elsewhere (12). Briefly, after being washed in Hanks' balanced salt solution (HBSS), cells were incubated for 1 h with 1–2  $\mu$ M fura 2-AM (in 10% Pluronic F-127) at room temperature and under low light conditions. Subsequently, fura 2-AM solution was removed and cells were incubated for 1 h in HBSS (50  $\mu$ l) to allow endogenous esterases to cleave AM ester. Cells were stimulated by addition of 50  $\mu$ l of 2 $\times$  tastant solution. Response kinetics were measured from cells loaded with 100  $\mu$ M fura 2 via patch pipette and stimulated

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

\*To whom correspondence should be addressed. E-mail: john.carlson@yale.edu.

© 2003 by The National Academy of Sciences of the USA

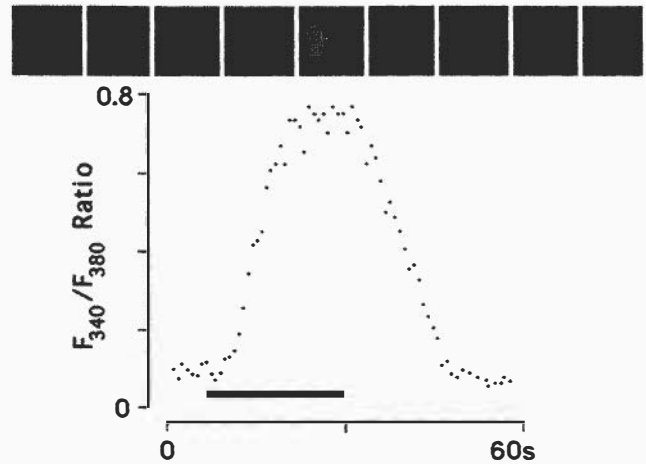


**Fig. 1.** Expression of *Gr5a* in taste neurons in the labellum and distal segments in the leg. Shown here are whole mount samples of labella or tarsi. Genotypes examined were as follows: *Gr5a-GAL4/+; UAS-lacZ/+* (a and c) and *Gr5a-GAL4/UAS-mCD8:GFP; Gr5a-GAL4/UAS-mCD8:GFP* (b and d). Shown in b is a composite of a series of confocal images. Reporter gene expression was observed in  $\approx 30$  cells in each half of the labellum and 2 cells in each of the two to three distal-most segments of the leg.

with  $1\times$  tastant solution via a puffer pipette under control of PicoPump 830 (World Precision Instruments, Sarasota, FL). All chemicals were purchased from Sigma. Imaging experiments were conducted on a Nikon TE300 inverted microscope equipped with a Nikon superfluor lens ( $\times 10/0.5$  and  $\times 20/0.75$ ). Cells were stimulated with 340 and 380 nm UV light from a DeltaRAM monochromator, and resulting images were collected by using an IC-200 intensified charge-coupled device camera; data acquisition and analysis were performed by using an ImageMaster suite (PTI, South Brunswick, NJ). Individual responses were recorded for 60 s.  $F_{340}/F_{380}$  ratios were analyzed to measure  $Ca^{2+}$  release. A responder cell was defined as one showing a ratio of 0.34 or above. In general,  $>90\%$  of responses showed a ratio of at least 0.51.

## Results

To test the hypothesis that *Gr5a* plays a direct role in trehalose reception, we first asked whether it is expressed in taste neurons. Because previous attempts at *in situ* hybridization have proved unsuccessful with the great majority of *Gr* genes (3, 4) we generated *Gr5a* promoter-GAL4 lines. An 8.5-kb genomic region upstream of *Gr5a* was used to supply a promoter, and GAL4 was used to drive expression of both *UAS-lacZ* and *UAS-GFP*



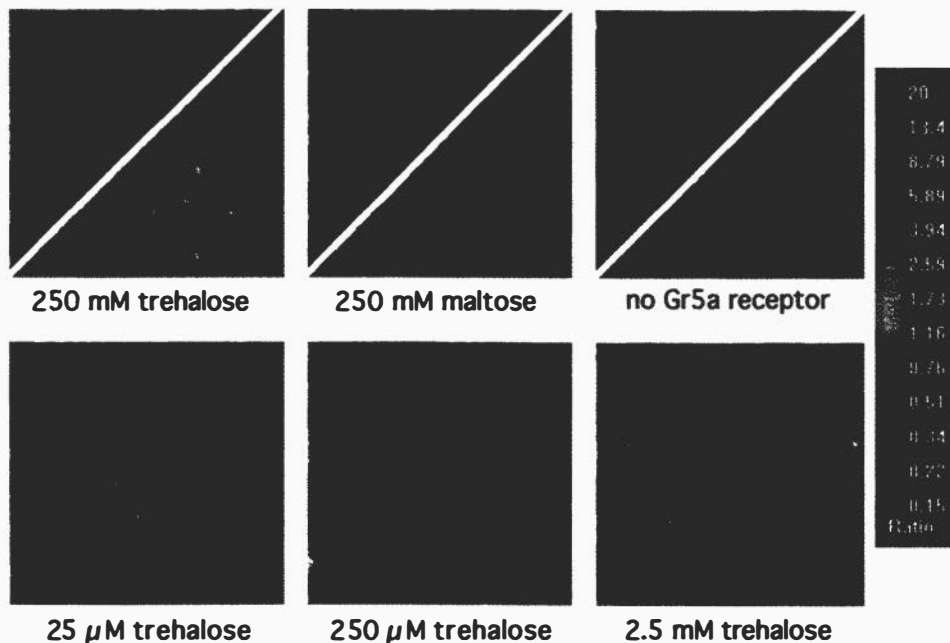
**Fig. 2.** Time course of trehalose response in S2-*Gr5a* cells. (Upper) A series of images of a single fura 2-loaded S2-*Gr5a* cell, taken at 5-s intervals. The first image is taken 5 s before the application of 100 mM trehalose. (Lower) A quantitative representation of the response of the same cell. Bar indicates stimulus period.

reporters. We observed wide expression in taste neurons of the labellum as well as in four to six neurons in the tarsi of adult flies (Fig. 1). No sexual dimorphism was observed in the expression pattern. Six independently derived lines were examined and all gave equivalent results.

To examine *Gr5a* function at the cellular level, we expressed *Gr5a* cDNA in *Drosophila* S2 cells. This cell line was chosen for two reasons. First, chemosensory receptors have been notoriously difficult to express in heterologous systems and we predicted that use of a *Drosophila* cell line might improve the possibility of functional expression of a *Drosophila* receptor. Second, previous studies have documented  $Ca^{2+}$  release after activation of G protein-coupled receptors that couple to the endogenous  $G_q$  protein of S2 cells (13–15). In this system, ligand binding to the receptor results in the activation of the phosphoinositide (PI) pathway: hydrolysis of  $PIP_2$  by phospholipase C into  $InsP_3$  and 4,5-diacylglycerol, and release of  $Ca^{2+}$  from intracellular stores. The stimulus-activated change in  $[Ca^{2+}]_i$  can be monitored with  $Ca^{2+}$ -sensitive fluorescent ratiometric indicators, such as fura 2 (16).

We transiently expressed *Gr5a* in S2 cells, loaded them with 100  $\mu M$  fura 2, and applied 100 mM trehalose via puffer pipette (Fig. 2). Stimulation evoked  $Ca^{2+}$  release: cell response developed within  $\approx 5$  s of ligand application and reached a peak intensity within  $\approx 15$  s of application. Upon removal of the ligand, the level of intracellular calcium gradually returned to the baseline. These data provided initial evidence that *Gr5a* encodes a functional trehalose receptor when expressed in S2 cells. The results also suggested the possibility that *Gr5a*-encoded receptor protein couples to the endogenous phosphoinositide pathway. We then cotransfected S2 cells with *Gr5a* and promiscuous G proteins:  $G_{\alpha 15}$ ,  $G_{\alpha 16}$ , or  $G_{\alpha 15}$  and  $G_{\alpha 16}$  together (17, 18). These G proteins are known to couple a wide variety of G protein-coupled receptors to intracellular  $Ca^{2+}$  release. There was no significant increase in the response intensity to trehalose stimulation compared with the S2-*Gr5a* cells, consistent with the possibility that *Gr5a* couples efficiently to  $G_q$ .

The response of S2-*Gr5a* cells depended on the presence of both an appropriate ligand and *Gr5a* (Fig. 3). We tested a wide range of trehalose concentrations, varying from 0.025  $\mu M$  to 250 mM (Figs. 3 and 4a). Responses to trehalose show a steep dose-dependence over this range. Responses were weak at 25



**Fig. 3.** Dose-dependence of trehalose response in S2-Gr5a cells. (Upper) Divided panels of S2-Gr5a cells (Left and Center) or negative controls, transfected with GFP vector alone (Right), before and after application of either 250 mM trehalose (Left and Right) or 250 mM maltose (Center). (Lower) Images of fields of S2-Gr5a cells taken on application of different concentrations of trehalose (indicated below).

$\mu\text{M}$ , and saturation was observed in the low mM range (Fig. 4a). The Hill coefficient was  $n_H = 1.92$ , suggesting the possibility that Gr5a functions as a homodimer.

We have investigated the ligand specificity of Gr5a by challenging S2-Gr5a cells with other sugars. We first tested two other disaccharides, maltose (composed of two glucose units) and sucrose (one glucose unit linked to fructose), and found that even when tested at high concentrations they elicited little if any response, as measured in terms of the percentage of cells that yielded a  $F340/380 \geq 0.34$  (Figs. 3 and 4). We then systematically tested a number of common disaccharides structurally related to trehalose, each composed of two glucose units (Fig. 4b). These isomers varied only in the positions of their glycosidic bond (e.g., 1,1', 1,4', or 1,6') and/or the configuration ( $\alpha$  or  $\beta$ ) of the D-glucose subunits. None of these other disaccharides evoked significant responses (Fig. 4c). The isomers tested included isotrehalose and neotrehalose, which, like trehalose, contain 1,1' linkages; however, isotrehalose contains a  $\beta,\beta$  linkage, and neotrehalose contains an  $\alpha,\beta$  linkage, whereas trehalose contains an  $\alpha,\alpha$  linkage. D-glucose, a monosaccharide component of all of the disaccharides tested, evoked no detectable  $\text{Ca}^{2+}$  release even at the highest dose tested (250 mM). The simplest interpretation of these results is that Gr5a recognizes moieties close to the  $1\alpha,1'\alpha$  glycosidic bond.

Trehalose is found in certain drought-adapted organisms such as yeasts and is thought to play a role in protecting the membrane during dehydration (19, 20). Although trehalose has no effect on  $\text{Ca}^{2+}$  levels in control cells that do not express Gr5a, we carried out two further experiments to control for the formal possibility that trehalose interacts with the plasma membrane and somehow activates the receptor nonspecifically. We found first that trehalose had no effect on cells transfected with another Gr gene, Gr64f, and, second, that other molecules believed to have similar effects on membranes, glycerol and 1,2 propanediol, have no effect on Gr5a-expressing cells (tested at 100 mM each, data not shown). These results are consistent with the conclusion that trehalose is a ligand for Gr5a.

## Discussion

This study provides direct evidence that Gr5a, a member of a large family of G protein-coupled receptors, functions as a trehalose taste receptor. Gr5a is expressed in neurons in taste sensilla of both the labellum and tarsi, consistent with a role as a taste receptor. When expressed in *Drosophila* S2 cells, it confers a response to trehalose in a dose-dependent manner. The response depends both on the expression of Gr5a and on a specific stimulus, trehalose. Other disaccharides tested evoke little if any response in this system.

We have provided evidence that Gr5a is expressed in all, or almost all, of the  $\approx 33$  sensilla present on the labellum. Because the labellum responds to a variety of sugars, and because the sensilla each contain a single sugar-sensitive neuron, the broad expression we have observed is consistent with a model in which many, if not all, of the sugar-sensitive taste neurons express more than one receptor. This model is supported by our earlier finding that mutation of Gr5a affected the physiological response of the sugar cell to trehalose, but not to sucrose, as if many of the sugar-sensitive cells contain both a trehalose receptor, Gr5a, and a sucrose receptor (9).

The expression pattern of Gr5a is broader than that observed for previously described GAL4 lines established by using promoters of other Gr genes (4, 5). The broad pattern is consistent with our earlier physiological data (9), which indicated that Gr5a is required for trehalose response in all L- and M-type sensilla (21). Hiroi *et al.* (22) also found that most sensilla on the labellum respond to trehalose.

We note that the response threshold of S2-Gr5a cells to trehalose is lower than in taste neurons *in vivo*, as determined in single-unit electrophysiological recordings (9). There are several possible explanations for this difference. One is that the two experiments measure different parameters, i.e.,  $\text{Ca}^{2+}$  levels vs. action potential frequency, and the  $\text{Ca}^{2+}$  level we have established as a criterion for scoring a response may be less than that required to initiate action potentials. Another possibility concerns access to tastant: in the expression system,

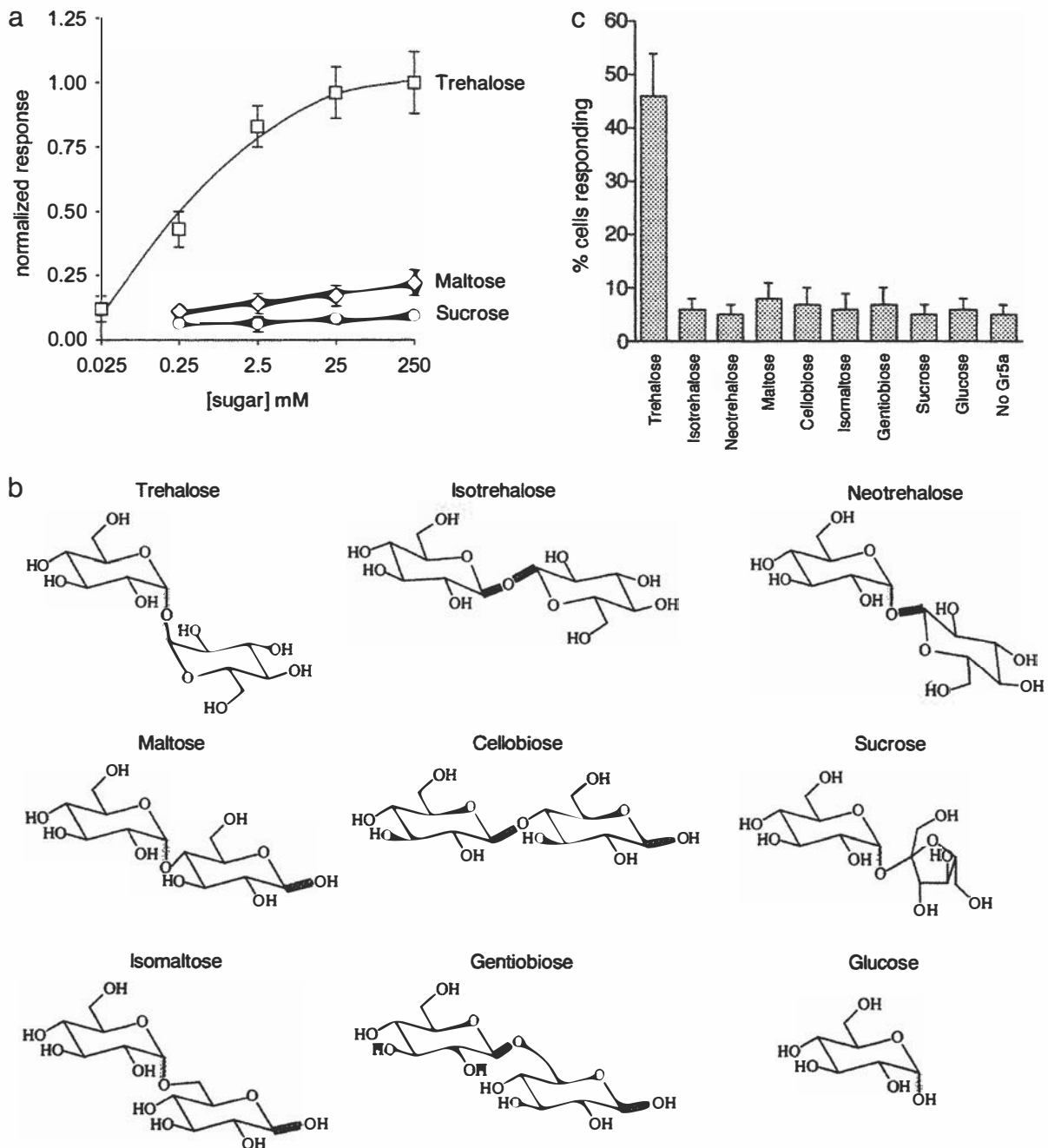


Fig. 4. (a) Dose-response curves for selected disaccharides. Responses are normalized to response at 250 mM trehalose.  $4 \leq n \leq 6$ ; error bars = SEM. (b) Structures of selected disaccharides related to trehalose. With the exception of sucrose, all are composed of 2 units of glucose. The  $\alpha$ -configuration is highlighted in yellow, and the  $\beta$ -configuration of the C1 carbon is highlighted in red. Isotrehalose and neotrehalose do not occur naturally. (c) Specificity of trehalose response in S2-Gr5a cells. Shown here are the responses of Gr5a-S2 cells to the various disaccharides illustrated in b. Compounds were tested at a concentration of 250 mM.  $5 \leq n \leq 9$ . "No Gr5a" cells were stimulated with 250 mM trehalose.

the cells are bathed in tastant, whereas *in vivo* the tastant must enter a pore in a sensillum and diffuse into the lymph surrounding a dendrite, where its final concentration may be lower than that of the test solution. A third possible explanation is that the density of receptor protein, or of another signaling component, may be greater in the heterologous expression system than *in vivo*.

The simplest interpretation of our results is that Gr5a functions as a homodimer, unlike the mammalian sweet receptors, which function as heterodimers (12). Furthermore, in contrast to the T1R2/T1R3 mammalian receptor that is rather broadly

tuned to diverse sweet-tasting molecules such as sucrose, saccharin, dulcin, and acesulfame-K, the Gr5a receptor is tuned to trehalose and shows much less, if any, response to other sugars, such as sucrose, fructose, and glucose, which the fly encounters in its natural habitat.

The relatively narrow tuning of Gr5a has implications for the mechanism of taste coding. If other *Drosophila* taste receptors are as specific as Gr5a, then an individual tastant is likely to be encoded largely by the activity of one or a small number of receptors, as opposed to the integrated activity of many receptors, each exhibiting a varying degree of response to a ligand. In

the olfactory system of *Drosophila*, many individual odorants activate several distinct classes of receptor neurons (23, 24), each expressing distinct odor receptors (ref. 25 and J.R.C., unpublished results). This model of taste coding is also supported by the severe loss of trehalose response after mutation of a single receptor gene, *Gr5a*. Analysis of further *Gr* proteins will be required to determine whether the narrow tuning of *Gr5a* is representative of *Gr* receptors at large or of those that recognize

tastants of particular metabolic significance to the fly, such as trehalose.

We thank Dr. K. Morris and Mrs. M. Chyb for technical assistance and Dr. R. C. Hardie, Dr. S. Simon, Prof. J. L. Frazier, Dr. W. M. van der Goes van Naters, and A. Ray for discussion or comments on the manuscript. This work was supported by Biotechnology and Biological Sciences Research Council (S.C.), a National Research Service Award (to A.D.), and National Institutes of Health grants and a McKnight Investigator Award (to J.R.C.).

1. Dethier, V. (1976) *The Hungry Fly* (Harvard Univ. Press, Cambridge, MA).
2. Rodrigues, V. & Siddiqi, O. (1978) *Proc. Indian Acad. Sci. Sect. B* **87**, 147–160.
3. Clyne, P. J., Warr, C. G. & Carlson, J. R. (2000) *Science* **287**, 1830–1834.
4. Scott, K., Brady, R., Jr., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C. & Axel, R. (2001) *Cell* **104**, 661–673.
5. Dunipace, L., Meister, S., McNealy, C. & Amrein, H. (2001) *Curr. Biol.* **11**, 822–835.
6. Chapman, R. F. (1982) in *The Insects: Structure and Function* (Harvard Univ. Press, Cambridge, MA), 3rd Ed., p. 919.
7. Tanimura, T., Isono, K., Takamura, T. & Shimada, I. (1982) *J. Comp. Physiol.* **147**, 433–437.
8. Tanimura, T., Isono, K. & Yamamoto, M. (1988) *Genetics* **119**, 399–406.
9. Dahanukar, A., Foster, K., Van der Goes van Naters, W. M. & Carlson, J. R. (2001) *Nat. Neurosci.* **4**, 1182–1186.
10. Ueno, K., Ohta, M., Morita, H., Mikuni, Y., Nakajima, S., Yamamoto, K. & Isono, K. (2001) *Curr. Biol.* **11**, 1451–1455.
11. Bunch, T. A., Grinblat, Y. & Goldstein, L. S. B. (1988) *Nucleic Acids Res.* **16**, 1043–1061.
12. Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y., Ryba, N. J. P. & Zuker, C. S. (2001) *Cell* **106**, 381–390.
13. Hardie, R. C., Reuss, H., Lansdell, S. J. & Millar, N. S. (1997) *Cell Calcium* **21**, 431–440.
14. Chyb, S., Raghu, P. & Hardie, R. C. (1999) *Nature* **397**, 255–259.
15. Graziano, M. P., Broderick, D. J. & Tota, M. R. (1999) in *Identification and Expression of G Protein-Coupled Receptors*, ed. Lynch, K. L. (Wiley-Liss, New York), pp. 181–195.
16. Gryniewicz, G., Poenie, M. & Tsien, R. Y. (1985) *J. Biol. Chem.* **260**, 3440–3450.
17. Offermanns, S. & Simon, M. I. (1995) *J. Biol. Chem.* **270**, 15175–15180.
18. Milligan, G., Marshall, F. & Rees, S. (1996) *Trends Pharmacol. Sci.* **17**, 235–237.
19. Crowe, J. H., Crowe, L. M. & Chapman, D. (1984) *Science* **223**, 701–703.
20. Crowe, L. M. (2002) *Comp. Biochem. Physiol.* **131**, 505–513.
21. Ray, K., Hartenstein, V. & Rodrigues, V. (1993) *Dev. Biol.* **155**, 26–37.
22. Hiroi, M., Marion-Poll, F. & Tanimura, T. (2002) *Zool. Sci.* **19**, 1009–1018.
23. De Bruyne, M., Clyne, P. J. & Carlson, J. R. (1999) *J. Neurosci.* **19**, 4520–4532.
24. De Bruyne, M., Foster, K. & Carlson, J. R. (2001) *Neuron* **30**, 537–552.
25. Dobritsa, A. A., Van der Goes van Naters, W., Warr, C. G., Steinbrecht, R. A. & Carlson, J. R. (2003) *Neuron* **37**, 1–20.

# Mammalian TRPV4 (VR-OAC) directs behavioral responses to osmotic and mechanical stimuli in *Caenorhabditis elegans*

Wolfgang Liedtke<sup>\*†</sup>, David M. Tobin<sup>†‡</sup>, Cornelia I. Bargmann<sup>\*§¶</sup>, and Jeffrey M. Friedman<sup>\*||</sup>

<sup>\*</sup>Laboratory of Molecular Genetics and <sup>||</sup>Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10021; and <sup>‡</sup>Departments of Anatomy and of Biochemistry and Biophysics, Programs in Developmental Biology, Genetics, and Neuroscience, and <sup>§</sup>Howard Hughes Medical Institute, University of California, San Francisco, CA 94143

All animals detect osmotic and mechanical stimuli, but the molecular basis for these responses is incompletely understood. The vertebrate transient receptor potential channel vanilloid subfamily 4 (TRPV4) (VR-OAC) cation channel has been suggested to be an osmo/mechanosensory channel. To assess its function *in vivo*, we expressed TRPV4 in *Caenorhabditis elegans* sensory neurons and examined its ability to generate behavioral responses to sensory stimuli. *C. elegans* ASH neurons function as polymodal sensory neurons that generate a characteristic escape behavior in response to mechanical, osmotic, or olfactory stimuli. These behaviors require the TRPV channel OSM-9 because *osm-9* mutants do not avoid nose touch, high osmolarity, or noxious odors. Expression of mammalian TRPV4 in ASH neurons of *osm-9* worms restored avoidance responses to hypertonicity and nose touch, but not the response to odorant repellents. Mutations known to reduce TRPV4 channel activity also reduced its ability to direct nematode avoidance behavior. TRPV4 function in ASH required the endogenous *C. elegans* osmotic and nose touch avoidance genes *ocr-2*, *odr-3*, *osm-10*, and *glr-1*, indicating that TRPV4 is integrated into the normal ASH sensory apparatus. The osmotic and mechanical avoidance responses of TRPV4-expressing animals were different in their sensitivity and temperature dependence from the responses of wild-type animals, suggesting that the TRPV4 channel confers its characteristic properties on the transgenic animals' behavior. These results provide evidence that TRPV4 can function as a component of an osmotic/mechanical sensor *in vivo*.

Animals perceive and avoid danger through sensory nociception, the detection of noxious stimuli in the external environment or noxious conditions in internal tissues. Noxious stimuli are polymodal: a stimulus can be recognized as noxious based on its specific chemical composition, as in inflammatory cytokines, based on its physicochemical properties such as pH and osmolarity, based on purely physical properties like pressure or heat, or based on a combination of properties. Some types of nociceptive neurons are intrinsically polymodal, detecting many noxious stimuli, whereas others are specialized to detect particular physical or chemical stimuli (1, 2). In mammals, dorsal root ganglion neurons sense noxious chemical, thermal, osmotic, acidic, or physical cues, or a combination of such cues (3). The more specific inner ear hair cells sense sound and acceleration (4), and neurons in circumventricular organs of the brain sense systemic osmotic pressure (5, 6). Although sensory transduction mechanisms for chemosensation and thermosensation have been extensively characterized, the molecular mechanisms for sensing mechanical or physicochemical stimuli are less well understood.

Nociception has been genetically characterized in the invertebrates *Drosophila melanogaster* and *Caenorhabditis elegans*. In these organisms, mutants with deficits in the response to noxious mechanical, osmotic, and chemical stimuli have been character-

ized (1, 7). In *C. elegans*, genetic studies have identified two distinct mechanosensory nociceptive pathways (8, 9). In one mechanosensory pathway, a mechanical stimulus applied to the animal's body elicits a withdrawal response. Members of the mechanically activated channel, degenerin channel, epithelial sodium channel (MEC/DEG/EnaC) family are required for body touch mechanosensation, and are likely to form the mechanosensory channels in these neurons (8, 10, 11). In another mechanosensory pathway, an avoidance response to nose touch is mediated by the ciliated ASH sensory neurons in the head (9, 12). ASH is a polymodal nociceptor: stimulation of the ASH cilia by either light touch, a hyperosmolar solution, or repulsive odorants leads the animal to reverse its direction of movement (13). Animals with mutations in *osm-9* or *ocr-2* do not avoid any of these stimuli. The *osm-9* and *ocr-2* genes encode putative ion channels belonging to the transient receptor potential (TRP) channel superfamily, vanilloid subfamily (TRPV) (14–16) and are proposed to encode the sensory transduction channel in ASH (17). The extent to which OSM-9 responds directly to sensory stimuli or indirectly to signal transduction pathways is unknown. Interestingly, the *Drosophila* ortholog of the *ocr* genes is required for fly hearing, a mechanosensory process (18). Another TRP family channel, the *Drosophila* *nompC* channel, is also thought to encode a mechanosensory channel, and a zebrafish *nompC* channel is implicated in mechanosensory hair cell function (19, 20).

TRPV4, also known as vanilloid receptor related osmotically activated channel (VR-OAC), OTRPC4, TRP12, or VRL-2, is a vertebrate nonspecific cation channel that is gated by osmotic stimuli in heterologous expression systems (15, 21–25). TRPV4 belongs to the TRPV subfamily of the TRP ion channel superfamily (15, 16, 25, 26) and shows similarity to the *C. elegans* channels OSM-9 (26% amino acid identity, 44% identity or conservative change) and OCR-2 (24% identity, 38% identity or conservative change). TRPV4 is also related to the vertebrate vanilloid receptor 1 (VR1; TRPV1), the vanilloid receptor-like channel (VRL-1; TRPV2), and the recently reported TRPV3 (27–32). TRPV1–3 are cation channels that are gated by warm and hot thermal stimuli.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: TRPV, transient receptor potential channel, vanilloid subfamily; VRL-1, vanilloid receptor-like receptor 1; VR1, vanilloid receptor 1.

<sup>†</sup>W.L. and D.M.T. contributed equally to this work.

<sup>‡</sup>To whom correspondence should be addressed. E-mail: cori@itsa.ucsf.edu.

© 2003 by The National Academy of Sciences of the USA



The expression of TRPV4 in the circumventricular organs of the vertebrate CNS suggests that this channel could sense systemic osmotic pressure. In transfected cells, TRPV4 is gated by hypotonicity within the physiological range (22, 23), and hypotonicity-gated TRPV4 function is suggested as one mechanism of pain sensation (33). However, in the circumventricular organ, hypertonicity is thought to be the physiological stimulus. This difference could result from accessory molecules in osmosensitive cells that modify TRPV4 function. TRPV4 is also expressed in vertebrate mechanosensory cells (1). Most models of mechanosensation propose the existence of multiple accessory proteins to mechanosensory channels (6). A full understanding of TRPV4 function might best be accomplished in cells with intrinsic mechanosensory or osmosensory function, rather than entirely heterologous cells.

To better understand the function of TRPV4 *in vivo*, we have expressed rat TRPV4 in the ASH neurons of *C. elegans*. Our results indicate that the rat channel can direct osmotic and mechanosensory behaviors when integrated into the normal ASH sensory transduction apparatus.

## Methods

**C. elegans Strains and Transgenic Animals.** *C. elegans* was grown at 20°C and maintained by using standard methods (34, 35). Wild-type animals were *C. elegans* variety Bristol, strain N2. Mutations used in this work were *osm-10(n1602) III, glr-1(ky176) III, osm-9(ky10) IV, ocr-2(ak47) IV, and odr-3(n1605) V*. All are strong loss-of-function mutations that are believed to represent null alleles. Transgenic arrays used were *kyEx575(sra-6::trpv4 elt-2::GFP)*, *kyEx594(sra-6::trpv4::GFP odr-1::dsRED)*, *kyEx609(sra-6::trpv4ΔN elt-2::GFP)*, *kyEx596(sra6::trpv4ΔC elt-2::GFP)*, *kyEx608(sra6::trpv4ΔNΔC elt-2::GFP)*, *kyEx612(sra-6::trpv4 D672K elt-2::GFP)*, *kyEx597* and *kyEx605(sra-6::trpv4 M680K elt-2::GFP)*, *kyEx598* and *kyEx611(sra-6::trpv4 D682K elt-2::GFP)*.

The coding region of the rat TRPV4 cDNA was expressed under the *sra-6* promoter in the *C. elegans* expression vector pPD49.26 (17). This promoter directs strong expression in the ASH and PVQ neurons, and weak expression in ASI neurons (35). Germ-line transformation was carried out by injecting DNA at 50 ng/μl (35) together with the intestinal *elt-2::GFP* marker at 10 ng/μl (17). Transgenic lines were maintained in an *osm-9(ky10)* genetic background; *ky10* is a null allele of *osm-9* that results from an early stop codon.

Details of molecular biology are available upon request. OSM-9::GFP5 lines were as described (34).

Neurons were identified for laser ablation by using differential interference contrast optics and a combination of positional and morphological cues as described (36). Cell kills of ASH were confirmed by bilateral absence of dye-filling with 1,1'-dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine perchlorate (DiD).

**Imaging** Animals expressing *sra-6::trpv4::gfp*, a C-terminally tagged protein, were analyzed with the Deltavision imaging system, which deconvolutes multiple sections of fluorescent micrographs to generate a projection and a 3D reconstruction (37).

Immunohistochemistry of fixed animals was performed by using an antibody raised against the first 445 aa of TRPV4 (a generous gift from Stefan Heller, Harvard University, Boston) and secondary goat anti-rabbit FITC antisera by using described methods (38).

**Behavioral Assays.** All behavioral assays were performed by investigators blinded to the genotype of the animals. Osmotic avoidance assays were performed by exposing individual animals to a drop of 1 M fructose or glycerol (39). A graded series of osmotic stimuli was obtained through serial dilution of 1 M glycerol. A reversal of more than half a body length before the

animal had left the drop area was scored as a response (17). For nose touch response, the tip of the nose of a forward-moving animal was touched with a fine hair, and reversal was scored as a response (9, 13). For chemical avoidance, 2-octanone and 1-octanol were used. A microcapillary containing 5 μl of 2-octanone or 1-octanol was placed immediately in front of a freely moving adult animal (17), and reversal within 3 s was scored as a response. For all avoidance assays, 10 trials per animal were recorded and the percentage of positive responses for each animal was compiled per group.

Temperature was modulated by using an over-the-counter heat lamp (75-W bulb) at a distance of 10–12 cm from the center of the assay plate under the dissecting microscope. Plate temperature was recorded and calibrated with a precision thermometer (YSI Temperature, Dayton, OH).

For further details, see *Supporting Methods*, which is published as supporting information on the PNAS web site, [www.pnas.org](http://www.pnas.org).

**Tissue Culture.** Cultivation of Chinese hamster ovary (CHO) cells, transfection, and stimulation were carried out as described (22). For further details, see *Supporting Methods*.

**Statistical Analysis.** Pairwise comparison was performed by *t* test, multigroup comparison by ANOVA analysis in combination with Dunnett's posttest analysis. The statistical program PRISM4 for Macintosh was used (GraphPad, San Diego).

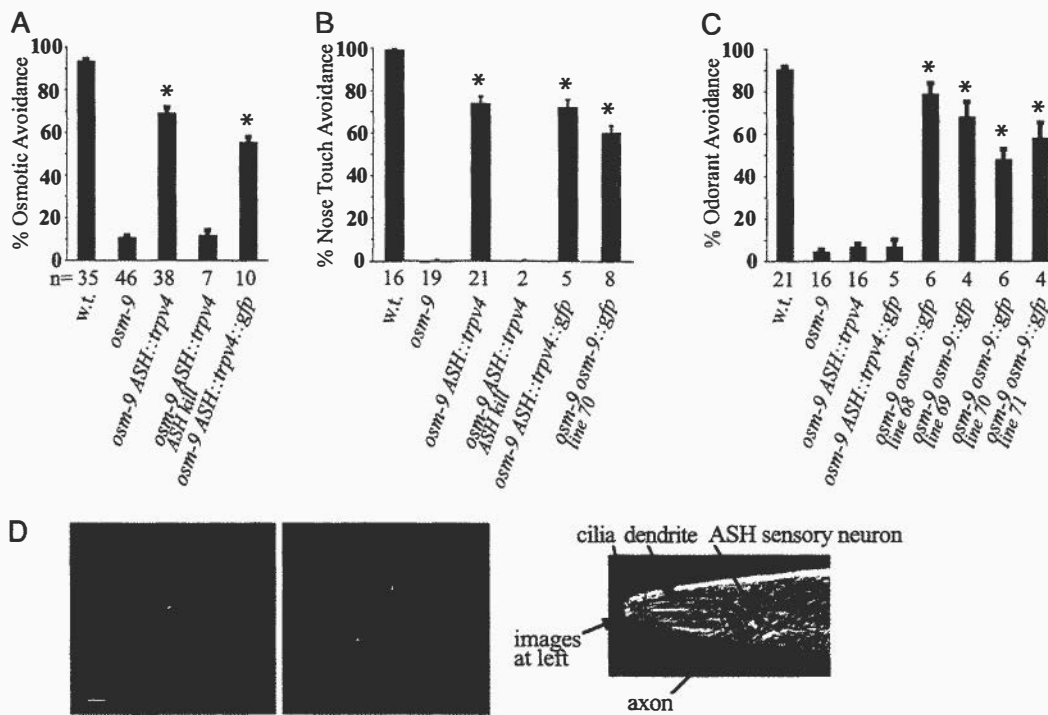
## Results and Discussion

To establish the properties of TRPV4 in a sensory system *in vivo*, we tested its ability to function within the ciliated ASH sensory neurons of *C. elegans* in the presence of an *osm-9* mutation that eliminates all endogenous ASH functions. *osm-9* animals are almost completely defective in their response to noxious hyperosmotic stimuli, nose touch and several aversive olfactory stimuli. By contrast, transgenic animals expressing TRPV4 or TRPV4::GFP in ASH neurons avoided both osmotic stimuli ( $P < 0.01$ ) and nose touch ( $P < 0.01$ ) (Fig. 1*A* and *B* and Movies 1–3, which are published as supporting information on the PNAS web site). The osmotic avoidance response of *osm-9 ASH::trpv4* animals was slightly delayed relative to the endogenous *C. elegans* behavior; this result is consistent with the relatively slow activation of vertebrate osmosensory channels.

To confirm that osmotic avoidance and nose touch avoidance were generated by TRPV4 expression in ASH, the ASH neurons were ablated in *osm-9 ASH::trpv4* animals by using a laser microbeam. Ablation of the ASH neurons eliminated osmotic avoidance and nose touch avoidance in the transgenic strain (Fig. 1*A* and *B*).

The biologically active TRPV4::GFP protein was localized to the cilia of ASH, where mechanical and osmotic stimuli are sensed (Fig. 1*D*). It was also present in ASH cell bodies, but not in axons or dendrites. Both OSM-9 and OCR-2 are similarly localized to sensory cilia (17).

TRPV4 failed to restore ASH-mediated odorant avoidance of the volatile repellents 2-octanone and 1-octanol to *osm-9* mutants (Fig. 1*C* and data not shown). Thus, the TRPV4 channel appears to confer osmosensitivity and mechanosensitivity, but not chemosensitivity, to the ASH neuron in *osm-9* mutants. This property is consistent with the expected functions of TRPV4, and different from the normal functions of OSM-9. As a control for this experiment, we tested an *osm-9* strain rescued with *osm-9::gfp* in parallel to the *osm-9 ASH::trpv4* and *osm-9 ASH::trpv4::gfp* strains. *osm-9 [osm-9::gfp]* animals were rescued for osmosensation, mechanosensation, and odor sensation, unlike *osm-9 ASH::trpv4* animals. The difference between OSM-9 and TRPV4 in odorant response appears to be qualitative rather than quantitative. The *osm-9 [osm-9::gfp]* strain exhibited relatively weak rescue of nose touch compared with the



**Fig. 1.** TRPV4 expression directs osmotic and nose touch avoidance in *osm-9* mutants. (A) Osmotic avoidance of 1 M fructose or glycerol. (B) Nose touch avoidance. (C) Avoidance of the odorant 2-octanone. Wild-type (w.t.) and *osm-9(ky10)* animals with or without an *ASH::trpv4* transgene, an *ASH::trpv4::gfp* transgene, or an *osm-9::gfp5* transgene were tested. "ASH kill" denotes bilateral laser ablation of the ASH neuron. In all panels, asterisks denote significant differences between the indicated group and the *osm-9* group ( $P < 0.01$ , one-way ANOVA with Dunnett's posttest analysis). Error bars denote SEM.  $n$  = number of animals tested, 10 trials each. (D) TRPV4::GFP expression in the nociceptive ASH neurons. (Left) Lateral view. (Center) Dorsal view. TRPV4::GFP in ASH appears green, and an ODR-1::dsRED fusion protein expressed in the dendrite and weakly in the cilium of the adjacent AWC sensory neuron appears red. Yellow arrow, ASH cilia; blue arrow, base of dendrite. (Scale bar = 5  $\mu$ m.) (Right) Schematic diagram of *C. elegans* ASH sensory neuron (red), 1 of 12 amphid sensory neurons (blue) that extend dendrites to the nose, where they terminate in sensory cilia. Two amphids each contain an ASH polymodal nociceptive neuron; only the left amphid is shown. The area depicted in the fluorescent micrographs is highlighted.

*ASH::trpv4* line (Fig. 1B), presumably because of lower expression of the transgene, yet the line still rescued odorant avoidance (Fig. 1C). Thus, the specific rescue of the osmotic and mechanical sensing deficits of *osm-9* by TRPV4 represents an intrinsic difference between TRPV4 and OSM-9.

TRPV4-expressing animals avoided hyperosmotic stimuli, but in heterologous expression systems, TRPV4-expressing cells are activated only in response to hypoosmolar stimuli (22, 23). This difference suggests that the ASH sensory neuron might contain accessory molecules that allow TRPV4 to sense hyperosmotic stimuli. *ocr-2*, *odr-3*, and *osm-10* genes are required for endogenous ASH osmosensory behaviors. *ocr-2* encodes another TRPV ion channel subunit, *odr-3* encodes a  $G_{\alpha}$ -protein (40), and *osm-10* encodes a cytoplasmic protein in the ASH neuron (41). To assess whether these genes were also required for TRPV4 function in ASH neurons, *osm-9 ASH::trpv4* transgenic animals were crossed to animals carrying *ocr-2*, *odr-3*, and *osm-10* mutations. TRPV4 was unable to generate osmosensory behaviors in *ocr-2*, *odr-3*, and *osm-10* mutants (Fig. 2A). Thus, endogenous osmosensory molecules in ASH collaborate with TRPV4 to generate an osmotic response. OCR-2 is required for OSM-9 localization to ASH sensory cilia, but is not required for cilia localization of TRPV4 (Fig. 5, which is published as supporting information on the PNAS web site). This result suggests that the requirement for both OCR-2 and OSM-9 or TRPV4 reflects activity of both channel subunits, and not only their mutual localization to cilia.

Endogenous ASH mechanosensation requires *ocr-2* and *odr-3*, but not *osm-10*. ASH mechanosensation also requires the *glr-1*

AMPA-type glutamate receptor, which is not required for osmosensation (42, 43). Mechanosensation mediated by *ASH::trpv4* required the function of *ocr-2*, *odr-3*, and *glr-1* genes (Fig. 2B). As in the endogenous response, *glr-1* was not required for *ASH::trpv4*-mediated osmosensation, and *osm-10* was not required for *ASH::trpv4*-mediated mechanosensation.

Expression of TRPV1 (VR1) does not rescue osmosensation or mechanosensation in *osm-9* mutants, but does confer a strong avoidance behavior to capsaicin, a TRPV1 ligand (17). The capsaicin response of *osm-9 ASH::trpv1* transgenic animals was retained in *ocr-2*, *odr-3*, *osm-10*, and *glr-1* mutant backgrounds. Thus, TRPV4, but not TRPV1, exploits all of the characterized endogenous osmosensory and mechanosensory signaling proteins in ASH to generate behavior.

If TRPV4 acts as a mechanosensory or osmosensory channel, its behavioral functions should correlate with its ability to function as an ion channel. Three point mutations in the predicted pore-loop domain of TRPV4, D672K, M680K, and D682K were made based on published studies of other family members, TRPV1 (VR1) and TRPV5 (EcaC) (Fig. 3A and Fig. 6, which is published as supporting information on the PNAS web site) (44, 45). The replacement of a methionine with lysine at position 680 (M680K) eliminated the ability of TRPV4 to confer nose touch avoidance and osmotic avoidance to the *osm-9* mutant (Fig. 3B and C). TRPV4<sub>M680K</sub> lacked channel activity when expressed in mammalian tissue culture cells stimulated with a TRPV4-activating phorbol ester or hypotonic solution (Fig. 7, which is published as supporting information on the PNAS web site). This observation is consistent with previous

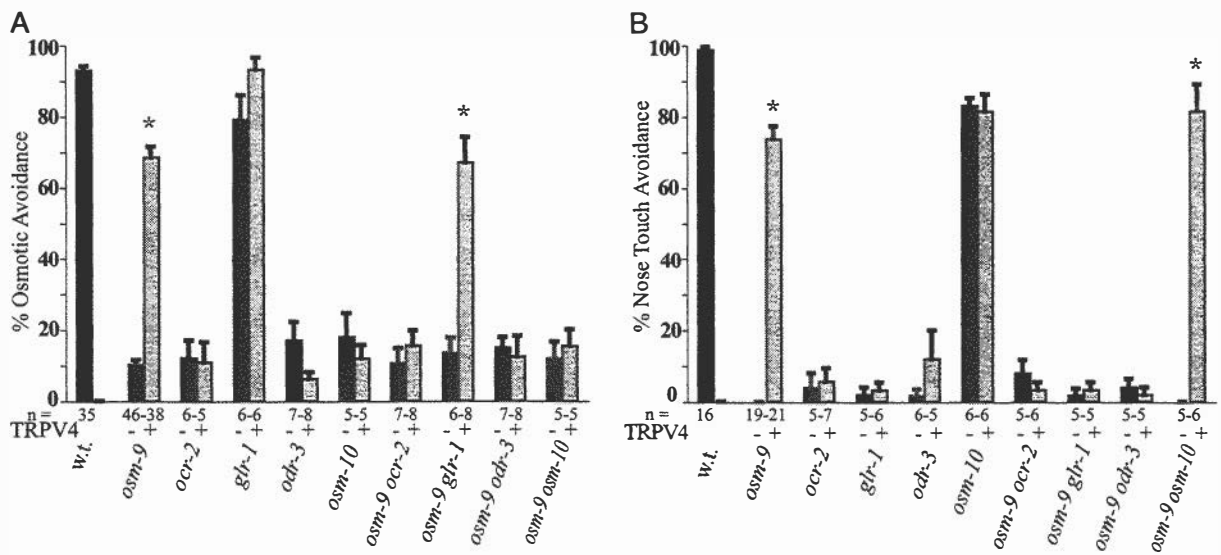


Fig. 2. TRPV4 functions with endogenous *C. elegans* osmo- and mechanosensory genes. (A) Osmotic avoidance of 1 M fructose or glycerol by single mutants (Left) or double mutants with *osm-9* (Right), with (gray) or without (black) the *ASH::trpv4* transgene. The first three bars represent the same results depicted in Fig. 1A. (B) Nose touch avoidance. Strains as are in A. The first three bars represent the same results depicted in Fig. 1B. In both panels, asterisks denote statistically significant differences between the *trpv4* transgenic group and the parallel nontransgenic group ( $P < 0.01$ , one-way ANOVA with Dunnett's posttest analysis).  $n$  = number of animals tested, 10 trials each; w.t., wild type.

studies showing that residue M680 is required for TRPV ion channel function (44–46). TRPV4-mediated behaviors were greatly diminished in animals expressing TRPV4 mutations at position 682 (D682K) and at position 672 (D672K) (Fig. 3B and C). Deletion of the TRPV4 N terminus (residues 1–410), C terminus (residues 741–871), or both N- and C-termini reduced but did not eliminate its behavioral function (Fig. 3B and C). The  $\Delta N\Delta C$  mutant lacking both N and C termini localized to the ASH cilia, as did the M680K mutant channel (Fig. 8, which is published as supporting information on the PNAS web site).

TRPV4 osmotic responses are enhanced at physiological temperatures in transfected mammalian cells (22, 47). We assessed the temperature-dependence of avoidance behaviors of wild-type and *osm-9 ASH::trpv4* animals at room temperature

(23°C) and 34–35°C, the upper limit of *C. elegans* viability. Wild-type animals responded significantly more frequently to nose touch at room temperature, whereas *osm-9 ASH::trpv4* responded more strongly at 35°C (Fig. 4A and Fig. 9, which is published as supporting information on the PNAS web site); a smaller effect was observed for osmotic avoidance (Fig. 4B). Although TRPV4-induced behaviors appeared to reflect the temperature modulation of TRPV4, TRPV4 did not generate avoidance behavior to thermal stimuli alone (48, 49), whereas TRPV1, a channel implicated in noxious heat responses, conferred a significant thermal avoidance response to isotonic M13 buffer at 37°C (Fig. 10, which is published as supporting information on the PNAS web site). Thus, the transduction of osmotic and mechanical stimuli in *osm-9 ASH::trpv4* animals bears the molecular signature of TRPV4. Somatosensory

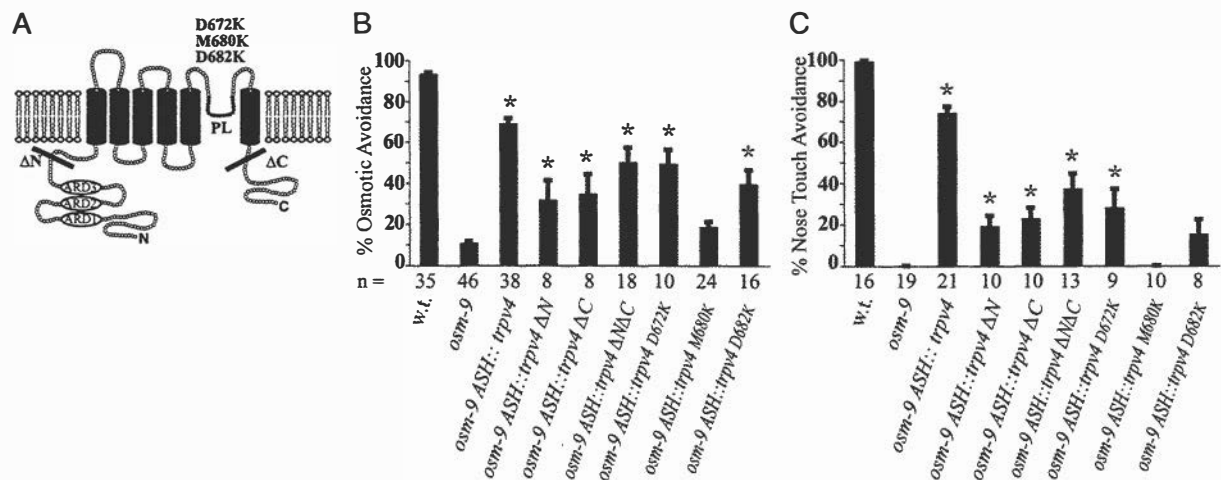


Fig. 3. TRPV4 mutations affect *C. elegans* behavior and channel function. (A) Schematic drawing of TRPV4 including sites of mutations. ARD, ankyrin-repeat domain; PL, pore-loop domain;  $\Delta N$ , extent of deleted N-terminal domain;  $\Delta C$ , extent of deleted C-terminal domain. N and C termini are intracellular. (B) Osmotic avoidance of 1 M fructose or glycerol. Bar graphs depict pooled data of three independent transgenic lines for each mutant. The first three bars represent the same results depicted in Fig. 1A. (C) Nose touch avoidance. Bar graphs depict the combined results for three independent transgenic lines of each mutant. The first three bars represent the same results depicted in Fig. 1B. For B and C, asterisk denotes significant differences between the indicated group and the *osm-9* group (one-way ANOVA with Dunnett's posttest analysis; for B:  $P < 0.01$  except  $\Delta N$  group and  $\Delta C$  group, here  $P < 0.05$ ; for C:  $P < 0.01$  except  $\Delta N$  group).  $n$  = number of animals tested, 10 trials each; w.t., wild type.

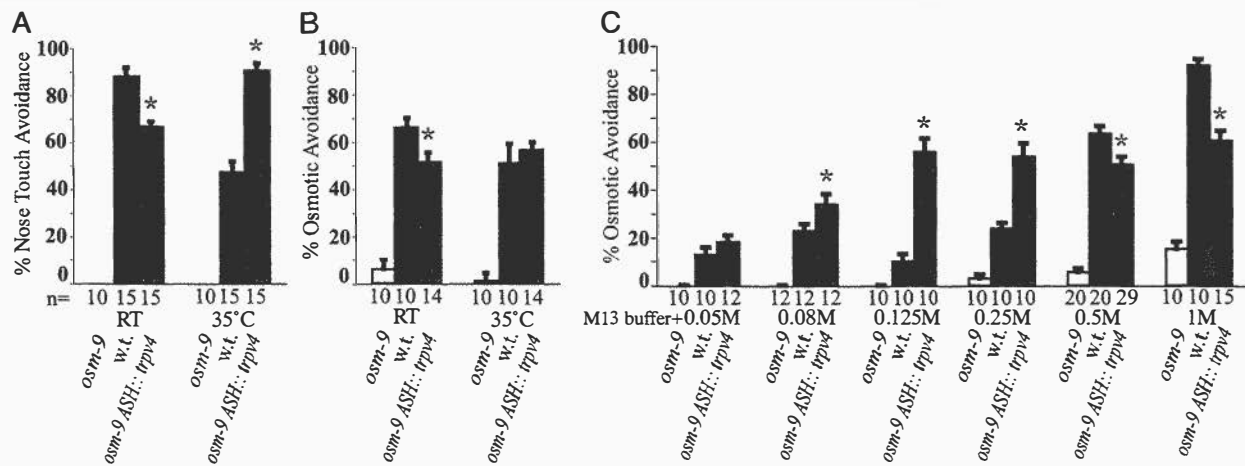


Fig. 4. TRPV4 avoidance behaviors are modulated by stimulus strength and temperature. (A) Nose touch avoidance at different temperatures. (B) Osmotic avoidance of 0.5 M glycerol at different temperatures. (C) Osmotic avoidance of different concentrations of glycerol. The osmotic strength of M13 buffer is 295 mOsmol/liter; i.e., a 0.05 M osmotic stimulus yields a final concentration of 345 mOsmol/liter. Asterisks denote significant differences between *osm-9 ASH::trpv4* and wild-type (w.t.) groups ( $P < 0.05$ ,  $t$  test for the 0.08 M group in C,  $P < 0.01$  for all other groups in A–C).  $n$  = number of animals tested, 10 trials each.

perception of mechanical stimuli in humans and seals is modulated by temperature (50–52), and TRPV4 may recapitulate this effect at the molecular level.

Vertebrate osmosensation in the central nervous system is exquisitely sensitive to small increases in systemic osmolarity. By contrast, *C. elegans* osmosensation detects relatively large increases in external osmolarity that may be associated with salty or brackish water. This difference in physiological function might predict a difference in the threshold of osmosensory channels in *C. elegans* and vertebrates. The osmotic avoidance behaviors of wild-type, *osm-9*, and *osm-9 ASH::trpv4* animals were assessed by using solutions with different osmotic strengths (Fig. 4C). Wild-type *C. elegans* responded weakly to increased osmolalities of 0.25 M and below, with a maximal behavioral response at a 1 M osmotic stimulus. By contrast, *osm-9 ASH::trpv4* animals responded equally well to 0.125 M, 0.25 M, 0.5 M, and 1 M osmotic stimuli. Thus, the *osm-9 ASH::trpv4* animals exhibited a significantly lower threshold to hyperosmotic stimuli than wild-type animals, as would be consistent with the high sensitivity of vertebrate osmosensation.

TRPV4 allows ASH neurons to respond to hypertonicity, but it responds to hypotonicity when expressed in transfected cells. One possible explanation for this difference is suggested by the properties of the mechanosensitive ion channel gramicidin A, which behaves either as a stretch-inactivated or as a stretch-activated channel depending on the lipid composition of the surrounding lipid bilayer (53). An alternate possibility is that TRPV4 forms heteromultimeric complexes with *C. elegans* proteins. TRP ion channels can form heteromeric complexes with related family members (17, 25, 54, 55); notably, the two TRPV channels OCR-2 and OSM-9 may associate in ASH (17).

Expression of rat TRPV4 in ASH rescued the transduction of osmotic and mechanical stimuli in *osm-9* but not *ocr-2* mutants, whereas rat TRPV1 did not rescue either mutant. These results suggest that TRPV4 and OSM-9 have orthologous functions,

indicating a phylogenetic conservation of function. It is interesting that TRPV4 could not rescue the odorant avoidance deficit of *osm-9* mutants for the two odors tested. *C. elegans* senses odorants by using G protein-coupled receptors; it is possible that TRPV4 does not interact with an essential component of the G protein-coupled receptor signaling pathway, although it does interact with the components required for sensing physical stimuli. Among several models consistent with these results (Fig. 11, which is published as supporting information on the PNAS web site), the conserved role of TRPV ion channels in osmosensory pathways suggests a central, perhaps direct, role for these channels in osmosensation and mechanosensation. However, the essential role of the G protein  $\alpha$  subunit ODR-3 in mechanosensation and chemosensation leaves open the possible involvement of G protein-coupled receptors as mediators or regulators of these sensory signals.

Our results provide evidence that mammalian TRPV4 can function as a central component of the sensor for osmotic and mechanical stimuli in ASH sensory neurons in *C. elegans*. Expression of TRPV4 directs a behavioral response to osmotic and mechanical stimuli *in vivo*, and specific properties of the behavior are conferred by the mammalian channel. This conclusion is consistent with findings in *trpv4* null mice (56–58), and supports the hypothesis that TRPV4 functions as part of a mammalian osmotic and mechanical sensor.

The Rockefeller University Bio-Imaging Resource Center (Alison North, director), provided assistance with imaging, and Jim Hudspeth, Shai Shaham (both from The Rockefeller University), Justin Blau (New York University, New York), Stefan Heller (Harvard University, Boston) and Sebastian Martinek (Pennie Partners, New York) provided valuable suggestions and support. W.L., C.I.B., and J.M.F. were supported by the National Institutes of Health. C.I.B. and J.M.F. are Investigators with the Howard Hughes Medical Institute. D.M.T. received a National Science Foundation predoctoral fellowship.

- Goodman, M. B. & Schwarz, E. M. (2003) *Annu. Rev. Physiol.* **65**, 429–452.
- Gillespie, P. G. & Walker, R. G. (2001) *Nature* **413**, 194–202.
- Gardner, E. P., Martin, J. H. & Jessell, T. M. (2000) in *Principles of Neural Science*, ed. Kandel E. R., Schwartz, J. H. & Jessell, T. M. (McGraw-Hill, New York), pp. 430–450.
- Hudspeth, A. J. (1989) *Nature* **341**, 397–404.
- Bourque, C. W. & Oliet, S. H. (1997) *Annu. Rev. Physiol.* **59**, 601–619.
- Denton, D. A., McKinley, M. J. & Weisinger, R. S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 7397–7404.

- Duggan, A., Garcia-Anoveros, J. & Corey, D. P. (2000) *Curr. Biol.* **10**, R384–R387.
- Goodman, M. B., Ernstrom, G. G., Chelur, D. S., O'Hagan, R., Yao, C. A. & Chalfie, M. (2002) *Nature* **415**, 1039–1042.
- Kaplan, J. M. & Horvitz, H. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 2227–2231.
- Garcia-Anoveros, J. & Corey, D. P. (1997) *Annu. Rev. Neurosci.* **20**, 567–594.
- Price, M. P., Lewin, G. R., McIlwrath, S. L., Cheng, C., Xie, J., Heppenstall, P. A., Stucky, C. L., Mannsfeldt, A. G., Brennan, T. J., Drummond, H. A., et al. (2000) *Nature* **407**, 1007–1011.

12. Bargmann, C. I. & Kaplan, J. M. (1998) *Annu. Rev. Neurosci.* **21**, 279–308.
13. Bargmann, C. I., Thomas, J. H. & Horvitz, H. R. (1990) *Cold Spring Harbor Symp. Quant. Biol.* **55**, 529–538.
14. Montell, C., Birnbaumer, L., Flockerzi, V., Bindels, R. J., Bruford, E. A., Caterina, M. J., Clapham, D. E., Harteneck, C., Heller, S., Julius, D., et al. (2002) *Mol. Cell* **9**, 229–231.
15. Clapham, D. E., Runnels, L. W. & Strubing, C. (2001) *Nat. Rev. Neurosci.* **2**, 387–396.
16. Gunthorpe, M. J., Benham, C. D., Randall, A. & Davis, J. B. (2002) *Trends Pharmacol. Sci.* **23**, 183–191.
17. Tobin, D., Madsen, D. M., Kahn-Kirby, A., Peckol, E., Moulder, G., Barstead, R., Maricq, A. V. & Bargmann, C. I. (2002) *Neuron* **35**, 307–318.
18. Kim, J., Chung, Y. D., Park, D. Y., Choi, S., Shin, D. W., Soh, H., Lee, H. W., Son, W., Yim, J., Park, C. S., et al. (2003) *Nature* **424**, 81–84.
19. Walker, R. G., Willingham, A. T. & Zuker, C. S. (2000) *Science* **287**, 2229–2234.
20. Sidi, S., Friedrich, R. W. & Nicolson, T. (2003) *Science* **301**, 96–99.
21. Delany, N. S., Hurler, M., Facer, P., Alnadaf, T., Plumpton, C., Kinghorn, I., See, C. G., Costigan, M., Anand, P., Woolf, C. J., et al. (2001) *Physiol. Genomics* **4**, 165–174.
22. Liedtke, W., Choe, Y., Marti-Renom, M. A., Bell, A. M., Denis, C. S., Sali, A., Hudspeth, A. J., Friedman, J. M. & Heller, S. (2000) *Cell* **103**, 525–535.
23. Strotmann, R., Harteneck, C., Nunnenmacher, K., Schultz, G. & Plant, T. D. (2000) *Nat. Cell Biol.* **2**, 695–702.
24. Wissenbach, U., Boding, M., Freichel, M. & Flockerzi, V. (2000) *FEBS Lett.* **485**, 127–134.
25. Montell, C., Birnbaumer, L. & Flockerzi, V. (2002) *Cell* **108**, 595–598.
26. Harteneck, C., Plant, T. D. & Schultz, G. (2000) *Trends Neurosci.* **23**, 159–166.
27. Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D. & Julius, D. (1997) *Nature* **389**, 816–824.
28. Caterina, M. J., Rosen, T. A., Tominaga, M., Brake, A. J. & Julius, D. (1999) *Nature* **398**, 436–441.
29. Caterina, M. J., Leffler, A., Malmberg, A. B., Martin, W. J., Trafton, J., Petersen-Zeitz, K. R., Koltzenburg, M., Basbaum, A. I. & Julius, D. (2000) *Science* **288**, 306–313.
30. Peier, A. M., Reeve, A. J., Andersson, D. A., Moqrich, A., Earley, T. J., Hergarden, A. C., Story, G. M., Colley, S., Hogenesch, J. B., McIntyre, P., et al. (2002) *Science* **296**, 2046–2049.
31. Smith, G. D., Gunthorpe, M. J., Kelsell, R. E., Hayes, P. D., Reilly, P., Facer, P., Wright, J. E., Jerman, J. C., Walhin, J. P., Ooi, L., et al. (2002) *Nature* **418**, 186–190.
32. Xu, H., Ramsey, I. S., Kotecha, S. A., Moran, M. M., Chong, J. A., Lawson, D., Ge, P., Lilly, J., Silos-Santiago, I., Xie, Y., et al. (2002) *Nature* **418**, 181–186.
33. Alessandri-Haber, N., Yeh, J., Boyd, A. E., Parada, C. A., Chen, X., Reichling, D. B. & Levine, J. D. (2003) *Neuron* **39**, 497–511.
34. Colbert, H. A., Smith, T. L. & Bargmann, C. I. (1997) *J. Neurosci.* **17**, 8259–8269.
35. Troemel, E. R., Chou, J. H., Dwyer, N. D., Colbert, H. A. & Bargmann, C. I. (1995) *Cell* **83**, 207–218.
36. Bargmann, C. I. & Avery, L. (1995) *Methods Cell Biol.* **48**, 225–250.
37. Chikashige, Y., Ding, D. Q., Funabiki, H., Haraguchi, T., Mashiko, S., Yanagida, M. & Hiraoka, Y. (1994) *Science* **264**, 270–273.
38. L'Etoile, N. D. & Bargmann, C. I. (2000) *Neuron* **25**, 575–586.
39. Hilliard, M. A., Bargmann, C. I. & Bazzicalupo, P. (2002) *Curr. Biol.* **12**, 730–734.
40. Roayaie, K., Crump, J. G., Sagasti, A. & Bargmann, C. I. (1998) *Neuron* **20**, 55–67.
41. Hart, A. C., Kass, J., Shapiro, J. E. & Kaplan, J. M. (1999) *J. Neurosci.* **19**, 1952–1958.
42. Hart, A. C., Sims, S. & Kaplan, J. M. (1995) *Nature* **378**, 82–85.
43. Maricq, A. V., Peckol, E., Driscoll, M. & Bargmann, C. I. (1995) *Nature* **378**, 78–81.
44. Nilius, B., Vennekens, R., Prenen, J., Hoenderop, J. G., Droogmans, G. & Bindels, R. J. (2001) *J. Biol. Chem.* **276**, 1020–1025.
45. Garcia-Martinez, C., Morenilla-Palao, C., Planells-Cases, R., Merino, J. M. & Ferrer-Montiel, A. (2000) *J. Biol. Chem.* **275**, 32552–32558.
46. Voets, T., Prenen, J., Vriens, J., Watanabe, H., Janssens, A., Wissenbach, U., Boedding, M., Droogmans, G. & Nilius, B. (2002) *J. Biol. Chem.* **277**, 33704–33710.
47. Gao, X., Wu, L. & O'Neil, R. G. (2003) *J. Biol. Chem.* **278**, 27129–27137.
48. Guler, A. D., Lee, H., Iida, T., Shimizu, I., Tominaga, M. & Caterina, M. (2002) *J. Neurosci.* **22**, 6408–6414.
49. Watanabe, H., Vriens, J., Suh, S. H., Benham, C. D., Droogmans, G. & Nilius, B. (2002) *J. Biol. Chem.* **277**, 47044–47051.
50. Dehnhardt, G., Mauck, B. & Hyvarinen, H. (1998) *J. Exp. Biol.* **201**, 3023–3029.
51. Fucci, D., Crary, M. & Wilson, H. (1976) *Percept. Mot. Skills* **43**, 263–266.
52. Weitz, J. (1941) *J. Exp. Psychol.* **28**, 21–36.
53. Martinac, B. & Hamill, O. P. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 4308–4312.
54. Xu, X. Z., Chien, F., Butler, A., Salkoff, L. & Montell, C. (2000) *Neuron* **26**, 647–657.
55. Xu, X. Z., Li, H. S., Guggino, W. B. & Montell, C. (1997) *Cell* **89**, 1155–1164.
56. Suzuki, M., Mizuno, A., Kodaira, K. & Imai, M. (2003) *J. Biol. Chem.* **278**, 22664–22668.
57. Mizuno, A., Matsumoto, N., Imai, M. & Suzuki, M. (2003) *Am. J. Physiol.* **285**, C96–C101.
58. Liedtke, W. & Friedman, J. M. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 13698–13703.

# Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*

Hugh M. Robertson\*<sup>†</sup>, Coral G. Warr\*<sup>‡</sup>, and John R. Carlson<sup>§</sup>

\*Department of Entomology, University of Illinois, 505 South Goodwin Avenue, Urbana, IL 61801; <sup>‡</sup>School of Biological Sciences, Monash University, Clayton VIC 3800, Australia; and <sup>§</sup>Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520

The insect chemoreceptor superfamily in *Drosophila melanogaster* is predicted to consist of 62 odorant receptor (Or) and 68 gustatory receptor (Gr) proteins, encoded by families of 60 Or and 60 Gr genes through alternative splicing. We include two previously undescribed Or genes and two previously undescribed Gr genes; two previously predicted Or genes are shown to be alternative splice forms. Three polymorphic pseudogenes and one highly defective pseudogene are recognized. Phylogenetic analysis reveals deep branches connecting multiple highly divergent clades within the Gr family, and the Or family appears to be a single highly expanded lineage within the superfamily. The genes are spread throughout the *Drosophila* genome, with some relatively recently diverged genes still clustered in the genome. The *Gr5a* gene on the X chromosome, which encodes a receptor for the sugar trehalose, has transposed from one such tandem cluster of six genes at cytological location 64, as has *Gr61a*, and all eight of these receptors might bind sugars. Analysis of intron evolution suggests that the common ancestor consisted of a long N-terminal exon encoding transmembrane domains 1–5 followed by three exons encoding transmembrane domains 6–7. As many as 57 additional introns have been acquired idiosyncratically during the evolution of the superfamily, whereas the ancestral introns and some of the older idiosyncratic introns have been lost at least 48 times independently. Altogether, these patterns of molecular evolution suggest that this is an ancient superfamily of chemoreceptors, probably dating back at least to the origin of the arthropods.

odorant receptor | gustatory receptor | olfaction | taste | gustation

Chemoreception in insects has long been a major focus of insect chemical ecology; however, despite many efforts, it was only with the sequencing of the genome of *Drosophila melanogaster* that candidate receptor proteins mediating olfaction and gustation were identified. These discoveries depended on the use of bioinformatic methods to identify genes encoding novel candidate G protein-coupled seven-transmembrane receptor proteins (1–5).

Members of the first family of these genes, the odorant receptor (Or) genes, were found to be expressed in subsets of olfactory neurons in the antenna and maxillary palp, the olfactory organs of this fly (1, 3, 6, 7). Completion of the genome sequence allowed extension of the Or family to 60 receptors with a unified naming system based on their chromosomal location (8). Immunolocalization showed expression in dendrites, as expected of odorant receptors (9, 10). Functional evidence for a role in odor reception was provided by Wetzel *et al.* (11) and Störtkuhl and Kettler (12), who used heterologous expression in *Xenopus* oocytes and overexpression in the *Drosophila* antenna, respectively, to show that Or43a mediates responses to a subset of odorants. Recently Dobritsa *et al.* (10) have shown through mutant and transgenic rescue analysis that *Or22a* is required *in vivo* for response to ethyl butyrate and certain other odorants. Moreover, several other Or genes were shown to confer response

to particular odorants or were mapped to particular functional classes of neurons, either by receptor substitution experiments in a mutant neuron or by analysis of strains in which Or promoters were used to drive reporter genes (10). Finally, we note that this family is conserved in other insects. *Anopheles gambiae* contains as many as 79 Or genes (13, 14), with few simple orthologs of *Drosophila* Or genes and largely species-specific expansion of gene subfamily lineages. In a study of four of these genes, all were detected exclusively in the antenna, and one is female-specific and down-regulated after a bloodmeal, as expected of a receptor for host odors (13).

Clyne *et al.* (2) subsequently used bioinformatics to identify another set of 42 seven-transmembrane genes, the gustatory receptor (Gr) genes. A role in gustatory reception was suggested by their expression profile. RT-PCR analysis revealed expression primarily in the proboscis but also in other organs containing gustatory neurons; moreover, expression was absent in mutants lacking gustatory neurons. After completion of the genome sequence, Scott *et al.* (15) and Dunipace *et al.* (16) extended the Gr family to at least 54 and 56 members, respectively, and used *in situ* hybridization and reporter gene constructs to reveal the detailed expression patterns of a subset of the Gr genes. Members of this family are expressed in subsets of neurons in proboscis, pharynx, and leg as well as in larval chemosensory organs. Some members are expressed in the antenna, suggesting a role for some members of the Gr family in olfaction. Evidence that a Gr gene functions in taste perception was provided by genetic analysis of *Gr5a*, which showed that it is required for response to the sugar trehalose (17, 18), and heterologous expression experiments show that *Gr5a* is a taste receptor tuned to trehalose (19).

Dunipace *et al.* (16) noted that the Gr proteins are distantly related to Or83b, whereas Scott *et al.* (15) suggested that the Or and Gr families belong together in a superfamily of insect chemoreceptors based on conservation of a few amino acid residues in transmembrane domain 7 (TM7). Here we extend the Or family to 62 receptors and the Gr family to 68 receptors, and we analyze their molecular evolution in an insect chemoreceptor superfamily.

## Materials and Methods

The public DNA database of the *Drosophila* genome sequences at National Center for Biotechnology Information (20) was searched with all available Or and Gr proteins by using TBLASTN (21) to find additional genes encoding proteins in these families,

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviation: TM, transmembrane domain.

<sup>†</sup>To whom correspondence should be addressed. E-mail: hughrobe@uiuc.edu.

© 2003 by The National Academy of Sciences of the USA

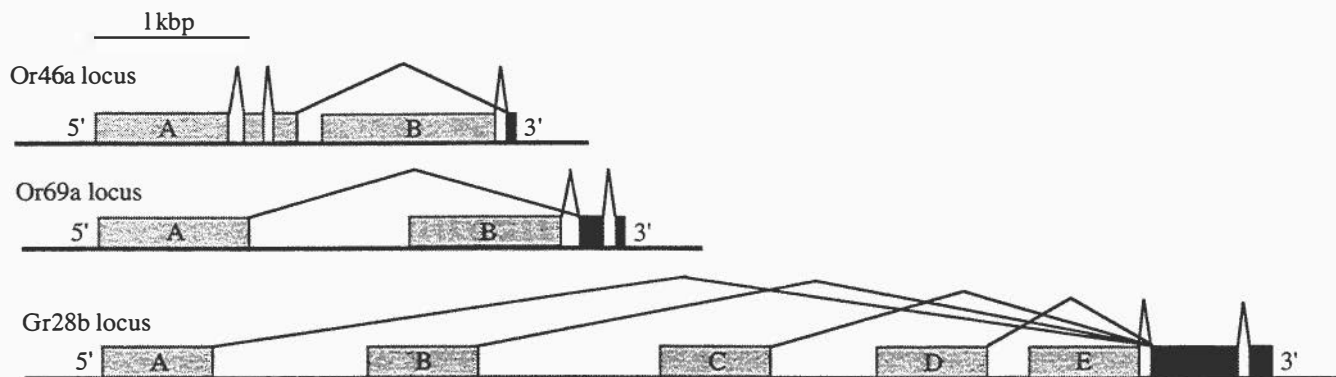


Fig. 1. Alternative splicing of *Or46a*, *Or69a*, and *Gr28b*. The gray boxes indicate the N-terminal exons that are unique to the differently spliced products, labeled with a letter designating the splice product, whereas the black boxes indicate the shared exons. In the case of *Or46a*, the encoded proteins share TM7; for *Or69a*, they share TM6 and -7; and for *Gr28b*, they share TM5–7. See Clyne *et al.* (2) for alternative splicing of *Gr23a* and -39a.

which were in turn used in searches to find more genes in an iterative process. These searches included the updated genomic sequences available as Release 3.1 of the genome (22). Multiple PSI-BLAST searches were initiated with divergent Ors and Grs to find any additional already annotated proteins that might belong to these families, and up to 10 iterations were used. The genes were reconstructed manually in the PAUP editor (23) by using the expected exon/intron structures as guides and the SPL program (Softberry, [www.softberry.com/berry.phtml](http://www.softberry.com/berry.phtml)) to locate predicted introns. In addition, protein alignments were used to indicate instances of unusual gene structure, as were comparisons with orthologs in the draft *Drosophila pseudoobscura* genome sequence. Proteins were aligned by using CLUSTALX (24), with considerable testing of alternative settings (see ref. 25). To facilitate alignment, the unusually long extracellular loop 2 between TM4 and -5 was removed from *Or83a*, *83b*, and *85e*, and *Gr33a*, -43a, and -66a, as were the long N and C termini of *Gr5a*, -21a, -32a, -61a, -63a, and *Gr64a*, -e, and -f. Relaxing the pairwise and multiple alignment gap and extension penalties by 10% to 9 and 0.09, respectively, yielded the best alignment of the seven TMs. Amino acid distances calculated between each pair of proteins were corrected for multiple amino acid changes in the past by using the maximum likelihood model in TREE-PUZZLE Ver. 5 (26), with the BLOSUM62 amino acid exchange matrix and uniform rates based on the actual sequences. A phylogenetic tree was constructed by using neighbor joining followed by a heuristic search for better trees by using tree-bisection-reconnection branch-swapping in PAUP\* Ver. 4.0b10 (23). Bootstrap analysis was performed by using 1,000 neighbor-joining replications with uncorrected distances. RT-PCR was performed as in Clyne *et al.* (1).

## Results

**The *Or* Family.** Fifty-four members of the *Or* family had been annotated by Celera Genomics and the *Drosophila* Annotation Jamboree (27) when this work was begun in mid-2000, whereas six more were recognized subsequently (6, 8). In the original Celera Genomics scaffold sequences, there were two identical copies of *Or19a* in inverted orientation  $\approx 50$  kb apart; however, resequencing of the genome by the Berkeley *Drosophila* Genome Project showed that these copies differ at seven nucleotide positions, yielding three changes in amino acid sequence; the proximal gene retains the *Or19a* name, and the distal gene we have named *Or19b*. The duplication extends  $\approx 850$  bp beyond the predicted N termini and  $\approx 700$  bp beyond the predicted C termini. This gene pair represents an unusually recent segmental duplication of the kind that might have been responsible for some of the expansion of the family; however, the separated and

inverted nature of the duplication is not typical of the tandem pairs and triplets of related genes seen for the rest of the family. One additional divergent *Or* gene, *Or67d*, was discovered in TBLASTN searches, located near the three known *Or* genes in chromosomal division 67. Twenty-eight of the annotations provided by Celera Genomics for the 54 annotated proteins seemed unlikely to be entirely correct, as judged from alignment of the protein family and common features of their structures, particularly a short final exon that encodes part of TM7 and that follows a final intron at a conserved location. These revisions were communicated to Swiss-Prot and FLYBASE, and most are incorporated in Release 3.1 of the genome annotations (22).

Despite intensive searching, satisfactory C-terminal exons could not be identified for two of the genes, *Or46a* and *Or69b*, leading us to hypothesize that they undergo alternative splicing similar to that noted by Clyne *et al.* (2) for two *Drosophila* *Gr* genes and Hill *et al.* (14) for several *Anopheles* *Gr* genes. An appropriate final exon encoding the end of TM7 was not found for *Or46a*, suggesting that the exons encoding most of *Or46a* are spliced to the final exon of *Or46b*. *Or69a* might similarly be spliced to the final two exons of *Or69b* (Fig. 1). We have confirmed both of these models by RT-PCR analysis, by using maxillary palp cDNA for *Or46a* and antennal cDNA for *Or69a*. These annotation changes require name changes from *Or46a* and *Or46b* to *Or46aA* and *Or46aB*, and from *Or69b* and *Or69a* to *Or69aA* and *Or69aB*.

Pseudogenes are rare in the *Drosophila* genome; however, several are found in the superfamily. In the sequenced Canton-S-derived *y; cn bw sp* strain, *Or85e* has suffered a deletion of the 3' end of the gene relative to an intact cDNA obtained by Vosshall *et al.* (3) from the Oregon-R strain. In addition, a fragmentary pseudogene was found just upstream of, and in tandem with, *Or98a* and was named *Or98P*. Its sequence is comparable to that of *Or98a*, but it has suffered a  $\approx 1$ -kbp internal deletion that leaves 138 bp encoding the 46 N-terminal amino acids, the final intron, and the final exon of 69 bp that encodes the 23 C-terminal amino acids. Large deletions like this are thought to be responsible for the paucity of pseudogenes in the *Drosophila* genome and its small size (e.g., ref. 28).

The locations and orientations of these *Or* genes are shown in Fig. 2. In addition to the recent *Or19a/b* duplication, a number of them are in short tandem arrays of two or three genes, indicating relatively recent gene duplication, and indeed in some cases the encoded proteins are closely related (Fig. 3; *Or22a/b*, *Or33a-c*, *Or59b/c*, *Or65a-c*, *Or85b-d*, and *Or94a/b*). However, the majority of *Or* genes are widely spread through the genome, indicating that, in agreement with their high sequence divergence, they are old members of this gene family that have been

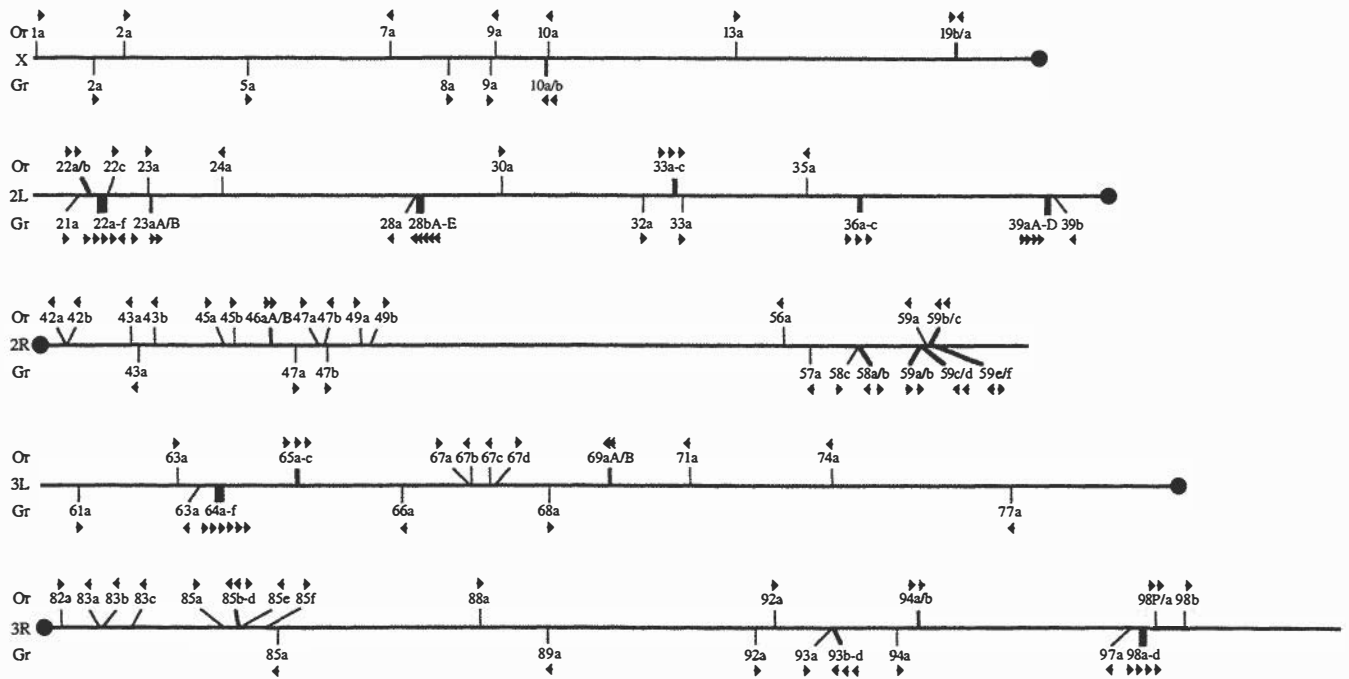


Fig. 2. Genomic locations of the *Or* and *Gr* genes. The *Or* genes are shown above, and the *Gr* genes below, central lines representing each of the five major chromosome arms drawn to scale following Adams *et al.* (27). Genes with inferred independent origins are designated by thin lines, whereas clusters of related adjacent genes, or alternatively spliced genes, are shown by thick lines. Orientation of transcription is shown with an arrow; the arrows for the alternatively spliced products are contiguous. The fragmentary *Or* pseudogene is indicated as 98P. All gene locations and orientations are based on data from Release 3.1 of the genome.

distributed around the genome by the processes of genome flux (e.g., ref. 29). This genomic distribution of the family is in contrast with the patterns observed with the mammalian olfactory receptors (e.g., refs. 30, 31) and the nematode chemoreceptors in the *str*, *srh*, and *srj* families (32, 33), which commonly are highly clustered on particular chromosomes, in part reflecting the relatively recent expansions of these chemoreceptor families.

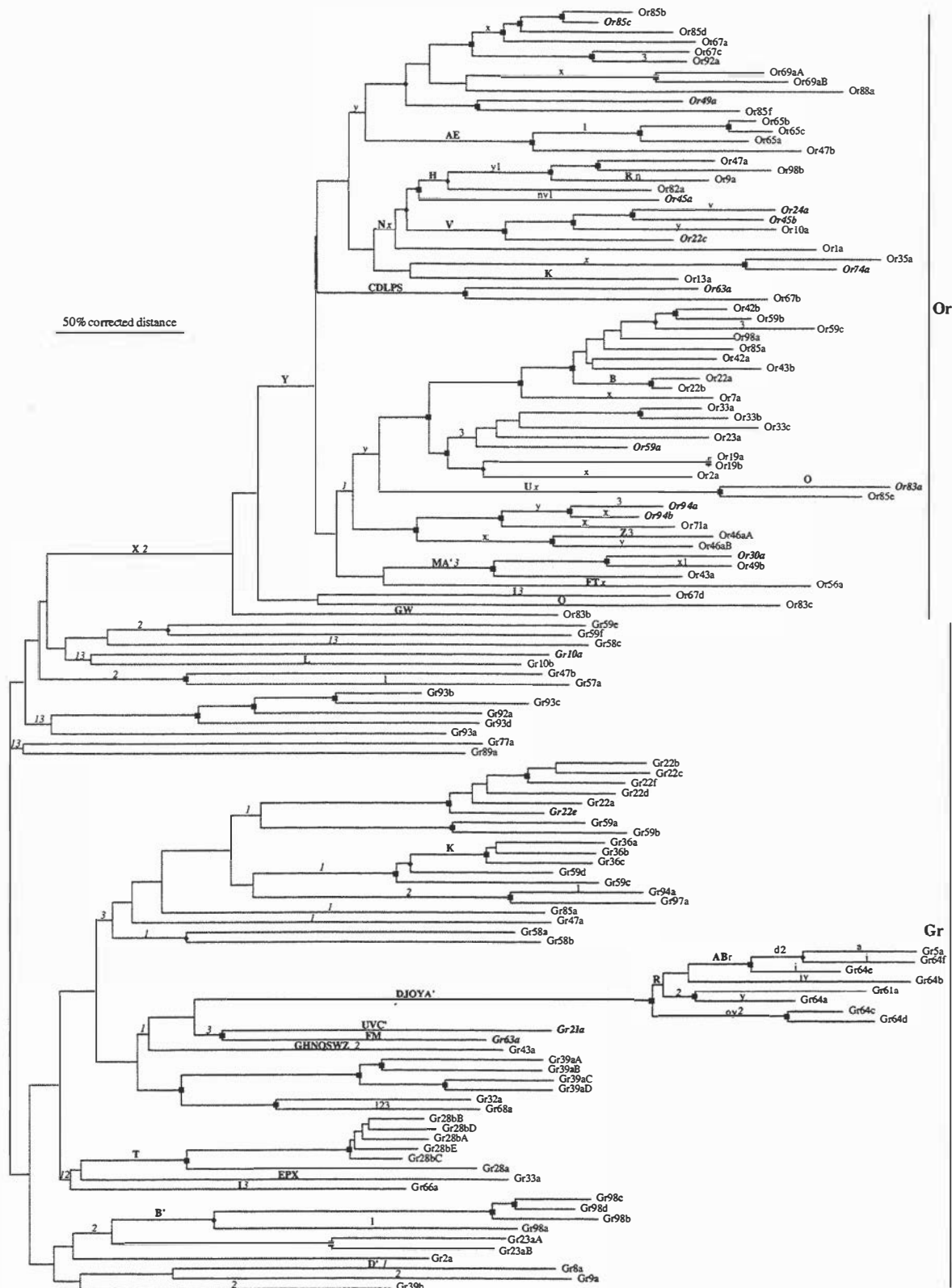
The ancestral and idiosyncratic intron locations within the coding regions of the *Or* genes are shown schematically in Fig. 4, along with those of the *Gr* genes. Three intron locations appear to be ancestral within the superfamily, as determined by their common location and insertion phase among multiple highly divergent *Or* and/or *Gr* lineages. We have named these intron positions 1, 2, and 3 (intron 2 is not present in the *Or* family, as if it were lost at the base of the *Or* lineage). There are 27 idiosyncratic *Or* introns that are not shared in the same location and phase with any *Gr* lineage but are generally present in one or only a few closely related *Or* genes. The exceptions are introns x and y, in phases 2 and 0, respectively, which are present in divergent *Or* lineages and were likely acquired near the base of the *Or* family tree. It seems likely that at least 25 introns were independently acquired within these single genes or small lineages relatively recently. They are unlikely to be ancient, because then the original *Or* gene must have been extraordinarily fragmented by introns, and multiple independent losses must have occurred in multiple different *Or* lineages. Twenty-seven independent losses of introns are inferred on the tree in Fig. 3. Twelve *Or* genes have lost all but one of their older introns without acquiring any new ones and hence have only one intron within their coding regions; at the other extreme, *Or63a* and *Or67b* acquired five new introns in addition to four older ones for a total of nine introns each.

Vosshall *et al.* (6) detected expression of 40 of the 57 *Or* genes they examined in sensory neurons of the antenna and maxillary

palp by *in situ* hybridization, and we have been able to detect mRNAs representing 48 of 61 *Or* transcripts (the analysis did not distinguish between *Or19a* and *-b*) in olfactory organs by a combination of *in situ* hybridization and RT-PCR (ref. 1; C.G.W., unpublished data), leaving 13 transcripts for which there is no evidence of expression in adult olfactory organs. There is no phylogenetic pattern to these 13 *Or* transcripts in the tree in Fig. 3 (highlighted in bold italics), suggesting that they do not represent a single lineage of genes. It is possible that they have been recruited to expression in other cell types or life stages.

**The *Gr* Family.** The combined efforts of Clyne *et al.* (2), Scott *et al.* (15), and Dunipace *et al.* (16) led to the recognition of  $\approx 64$  proteins in this family. In addition to the examples described by Clyne *et al.* (2) of alternatively spliced transcripts from two genes (*Gr23a* and *-39a*), together encoding six substantially different proteins, *Gr28b* encodes five predicted proteins (Fig. 1), which were only partially recognized by Scott *et al.* (15) and Dunipace *et al.* (16). We also add two previously undescribed members to the family, *Gr9a* and *-89a*, bringing the total to 68 proteins encoded by 60 genes. Most of these were poorly annotated or missed by the automated Celera Genomics annotation. The principal authors of these three papers (2, 15, 16) have coordinated a naming convention for these proteins analogous to that agreed to for the *Or* proteins; improved annotations for them have been agreed to by all groups and submitted to Swiss-Prot, and most are available in Release 3.1 of the *Drosophila* annotation (three genes originally designated as members of this family, *Gr36d*, *-43b*, and *-65a*, are no longer considered to be members of the superfamily). Comparison with *Gr* orthologs in the draft *D. pseudoobscura* genome indicated that eight of these annotations require further revision; the updated versions have been communicated to FLYBASE and are utilized here (see *Or/Gr Proteins*, which is published as supporting information on the





**Fig. 3.** Tree of the insect chemoreceptor superfamily. The tree is rooted at the midpoint. The Or and Gr families are indicated on the right, and the scale bar indicates 50% divergence in corrected distances (far larger than the uncorrected distances when comparing distantly related proteins). Branches with 75–100% bootstrap support are indicated with a square and can be considered to be confident, whereas branches with 60–75% bootstrap support are indicated with a diamond and can be considered somewhat confident. Inferred intron gains within the superfamily are indicated above branches in bold uppercase letters, whereas inferred intron losses are shown in lowercase letters (intron losses that are not confidently independent according to the bootstrap support for branches are shown in italics). The Or and Gr families have separate sets of intron letter designations (see Fig. 4), and the putatively ancient ancestral phase-0 C-terminal introns 1–3 are shown as numbers. The Or genes for which no evidence of expression in antenna or maxillary palp has been detected by Vosshall *et al.* (6) or by us are highlighted in bold italics, as are the four Gr genes that Scott *et al.* (15) and Dunipace *et al.* (16) showed to be expressed in the antenna and/or maxillary palp.

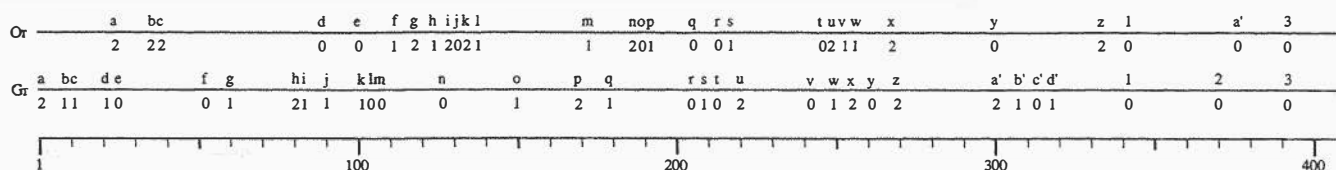


Fig. 4. Locations and phases of introns in the *Or* and *Gr* genes. The intron locations (above the lines) and phases (below the lines) are shown separately for the two families, relative to a scale of the average receptor size in amino acids (determined by excluding the large insertions in some of the receptors). The ancestral phase-0 C-terminal introns 1 and 3 are shared between the two families.

PNAS web site; see also *ClustalX Multiple Sequence Alignment* and Table 1 (CG numbers, locations, lengths, and intron numbers), which are published as supporting information on the PNAS web site.

Dunipace *et al.* (16) noted that *Gr22b* and *-d* are pseudogenes because of an in-frame stop codon and a single base-pair deletion in their first exons, respectively. Scott *et al.* (15) avoided these mutations by starting their annotations downstream, but the resulting proteins do not have TMI. We amplified and sequenced the relevant regions of these two genes from the Oregon-R strain, and both are intact genes in this strain. The apparent TGA stop codon in *Gr22b* corresponds to a CGA encoding arginine, whereas the single base deletion in *Gr22d*, located in a string of four Ts, corresponds to a string of five Ts in Oregon-R. Like *Or85e*, these differences therefore reflect strain polymorphisms also seen in some *Anopheles* receptors (14); the intact versions are used herein.

The genomic locations of the *Gr* genes are shown in Fig. 2. Like the *Or* genes, they are distributed widely around the genome; however, there are some larger clusters (*Gr22a-f* and *Gr64a-f* contain six genes each). Some *Gr* genes appear to have transposed relatively recently: within the *Gr22a-f* cluster, *Gr22e* is in the opposite orientation, and *Gr22f* is  $\approx 28$  kb downstream (see also ref. 16). Perhaps the most interesting cases are *Gr61a* and *-5a*, which phylogenetically cluster confidently with the *Gr64a-f* set but are now located elsewhere. *Gr5a* is the only *Gr* gene for which a function has been shown; it is required for response to trehalose (17–19), and it is possible that *Gr61a* and *Gr64a-f* encode additional sugar receptors. These genes share five idiosyncratic introns and are highly divergent from other receptors within the family. In contrast, the *Gr58a-c* genes are adjacent to each other, yet only *Gr58a/b* weakly cluster in the tree, and all three share as little amino acid similarity with each other as many other pairs of *Gr* proteins (11–16%). We note that *Or10a* is immediately upstream of and in the same orientation as *Gr10a/b*, with only  $\approx 350$  bp separating the stop codon of *Or10a* and the start codon of *Gr10a*. This proximity might be related to the expression of *Gr10a* in the antenna (15).

Most *Gr* proteins are extraordinarily divergent, sharing as little as 8% amino acid identity. Multiple alignment and phylogenetic analysis of such highly divergent proteins is difficult. We have used the entire lengths of the proteins, as aligned by CLUSTALX by using slightly relaxed gap penalties, which are particularly capable of achieving alignment of the hydrophobic TMs. In this manner, relationships of the more closely related proteins are well resolved, possibly at the expense of the more distant relationships of the backbone of the tree. As with the *Or* family, more distant relationships have no bootstrap support (Fig. 3).

The *Gr* genes contain even more idiosyncratic introns than the *Or* family. The most recently shared ancestral *Gr* gene is inferred to have had a long exon encoding TM1–5 followed by three phase 0 introns separating three exons encoding the C-terminal region beyond TM6 (Figs. 3 and 4; see also figure 1 in ref. 2). Mapping of intron losses on the tree is complicated by uncertainty in the backbone of their relationships. However, it seems most likely

that there have been 30 idiosyncratic intron gains and at least 21 intron losses, bringing the superfamily totals to 57 gains beyond the ancestral three phase 0 introns and at least 48 losses. Within the *Gr* family, two genes, *Gr68a* and *-94a*, have lost all introns within the coding region. Both of these genes are within the introns of other genes, and *Gr68a* may have lost its three introns simultaneously through reverse transcription and insertion of a retrocopy. It is possible that these two *Gr* genes and others have introns in their 5' UTRs that cannot easily be recognized bioinformatically. In contrast, *Gr64e* has eight introns.

### Discussion

We describe here what may be the full complement of 60 *Or* and 60 *Gr* family genes and 130 predicted chemoreceptor proteins that they encode in *D. melanogaster*. We cannot exclude the possibility of additional highly divergent, or evolutionarily independent, insect chemoreceptors. Scott *et al.* (15) introduced the notion that these two families of odorant and gustatory receptors are evolutionarily related in an insect chemoreceptor superfamily, and we endorse this view. This superfamily provides a remarkable diversity of receptors that could underlie the entire range of chemoreceptive capabilities of this fly. The *Or* genes appear to be one lineage, albeit highly expanded in gene number, within the larger *Gr* family (Fig. 3). To preserve nomenclatural clarity, however, we prefer to retain the *Or* and *Gr* designations. In addition to this phylogenetic interrelationship, Scott *et al.* (15) and Dunipace *et al.* (16) found that four *Gr* genes are expressed in subsets of neurons in the antenna and/or maxillary palp (indicated in bold italics in Fig. 3). Neurons expressing *Gr21a* project axons to glomeruli in the antennal lobe of the brain (15). If these antennal *Gr*s in fact function as odorant receptors, then it would appear that olfactory receptor function has evolved separately several times within the superfamily, perhaps in conjunction with the evolution of terrestrial insects from aquatic arthropod ancestors  $\approx 400$  million years ago.

*Or83b* is extremely divergent from the other *Or* proteins and is expressed in most olfactory receptor neurons (3). *Or83b* is also unusual in having an ortholog in *A. gambiae* that shares 78% amino acid identity, much higher than any other orthologous pair in these two distantly related dipterans (14), as well as 68% identity to CR2 from the moth *Heliothis virescens* (34). This conservation is sustained throughout the Endopterygota (metamorphosing insects) (ref. 35; G. New, H. Patch, K. Walden, and H.M.R., unpublished results). Dunipace *et al.* (16) noted that among the *Or*s, *Or83b* appears most similar to the *Gr* family proteins. Our phylogenetic analysis supports this observation (Fig. 3), and although this placement is not supported by bootstrapping, it is obtained regularly when the alignment and tree methodologies are modified considerably. This extraordinary conservation suggests that *Or83b* serves a function unlike that of other chemoreceptors (14, 35).

The antiquity of this insect chemoreceptor superfamily is supported by several lines of evidence. First, the genes encoding these proteins are roughly evenly spread throughout the genome (Fig. 2). Although there are a few clusters of related proteins that represent recent *in situ* expansions of gene lineages, the pro-

cesses of genome flux that led to the current distribution of the genes are clearly evident (for example, the translocation of *Gr5a* to the X chromosome from the *Gr64* cluster on chromosome 3L). This flux is reminiscent of several other ancient gene families in the *Drosophila* genome, e.g., the tetraspanin superfamily (36).

Second, the amino acid divergences between the Gr and Or proteins, and particularly among the Gr proteins, are extremely high; indeed, Gr proteins commonly share only 8–12% amino acid identity. Some of this divergence could be attributed to an evolving need to adapt to new ecological niches. Nevertheless, the extreme divergence within the family is consistent with an ancient origin. Identification of Or and Gr family members in the moths *H. virescens* (34) and *Manduca sexta* (H. Patch, K. Walden, and H.M.R., unpublished results) confirms the antiquity of the families.

Third, the vast majority of introns appear to have been idiosyncratically acquired by limited lineages of genes and commonly single genes (Fig. 3). This pattern of intron evolution is found in other old gene superfamilies (e.g., ref. 36). The ancestral insect chemoreceptor genes appear to have had only three phase 0 introns near their C termini.

Fourth, extended PSI-BLASTP searches initiated with various Grs detected similarities with proteins encoded by five *gustatory related (gur)* genes, putative seven-TM receptors in the nematode *Caenorhabditis elegans* (C. Stoezner and H.M.R., unpublished results). The *gur* genes are quite distinct from the ≈1,000 candidate chemoreceptors already identified in this nematode (32, 33, 37, 38). The similarity between Gr and GUR proteins suggests that the superfamily predates the arthropod/nematode split.

**Note Added in Proof.** Bray and Amrein (39) demonstrate that Gr68a is expressed in neurons of male-specific contact-chemosensory sensilla on male forelegs and implicate Gr68a in recognition of females in an early step of courtship, when males tap the abdomen of a female with their forelegs, presumably sampling their sex- and species-specific cuticular hydrocarbons. This study provides further support for a gustatory role for most Gr proteins.

We thank P. Clyne, J. Daniels, J. Kim, E. Moriyama, and A. Ray for assistance with annotation of these genes; and H. Amrein and K. Scott for cooperation in annotation/naming of the Grs. This work was funded by National Science Foundation Grant 9604095 (to H.M.R.) and National Institutes of Health Grants DC04729 and GM63364 (to J.R.C.).

- Clyne, P. J., Warr, C. G., Freeman, M. R., Lessing, D., Kim, J. & Carlson, J. (1999) *Neuron* **22**, 327–338.
- Clyne, P. J., Warr, C. G. & Carlson, J. R. (2000) *Science* **287**, 1830–1834.
- Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A. & Axel, R. (1999) *Cell* **96**, 725–736.
- Kim, J., Moriyama, E. N., Warr, C. G., Clyne, P. J. & Carlson, J. R. (2000) *Bioinformatics* **16**, 767–775.
- Kim, J. & Carlson, J. R. (2002) *J. Cell. Sci.* **115**, 1107–1112.
- Vosshall, L. B., Wong, A. M. & Axel, R. (2000) *Cell* **102**, 147–159.
- Gao, Q. & Chess, A. (1999) *Genomics* **60**, 31–39.
- Drosophila* Odorant Receptor Nomenclature Committee (2000) *Cell* **102**, 145–146.
- Elmore, T. & Smith, D. P. (2001) *Insect Biochem. Mol. Biol.* **31**, 791–798.
- Dobritsa, A., van der Goes van Naters, W., Warr, C., Steinbrecht, A. & Carlson, J. R. (2003) *Neuron* **37**, 827–841.
- Wetzel, C. H., Behrendt, H. J., Gisselmann, G., Stortkuhl, K. F., Hovemann, B. & Hatt, H. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 9377–9380.
- Stortkuhl, K. F. & Kettler, R. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 9381–9385.
- Fox, A., Pitts, R., Robertson, H., Carlson, J. & Zwiebel, L. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 14693–14697.
- Hill, C. A., Fox, A. N., Pitts, R. J., Kent, L. B., Tan, P. L., Chrystal, M. A., Cravchik, A., Collins, F. H., Robertson, H. M. & Zwiebel, L. J. (2002) *Science* **298**, 176–178.
- Scott, K., Brady, R., Jr., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C. & Axel, R. (2001) *Cell* **104**, 661–673.
- Dunipace, L., Meister, S., McNealy, C. & Amrein, H. (2001) *Curr. Biol.* **11**, 821–835.
- Dahanukar, A., Foster, K., van der Goes van Naters, W. M. & Carlson, J. R. (2001) *Nat. Neurosci.* **4**, 1182–1186.
- Ueno, K., Ohta, M., Morita, H., Mikuni, Y., Nakajima, S., Yamamoto, K. & Isono, K. (2001) *Curr. Biol.* **11**, 1451–1455.
- Chyb, S., Dahanukar, A., Wickens, A. & Carlson, J. R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 14526–14530.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2002) *Nucleic Acids Res.* **30**, 17–20.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, Z., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E., et al. (2002) *Genome Biol.* **3**, RESEARCH0083.1–0083.22; Epub 2002 Dec 31.
- Swofford, D. L. (2001) PAUP\*: *Phylogenetic Analysis Using Parsimony and Other Methods* (Sinauer, New York), Ver. 4.
- Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998) *Trends Biochem. Sci.* **23**, 403–405.
- Hall, B. G. (2001) *Phylogenetic Trees Made Easy* (Sinauer, New York).
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002) *Bioinformatics* **18**, 502–504.
- 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.
- Petrov, D. A. (2002) *Genetica* **115**, 81–91.
- Ranz, J. M., Casals, F. & Ruiz, A. (2001) *Genome Res.* **11**, 230–239.
- Glusman, G., Yanai, I., Rubin, I. & Lancet, D. (2001) *Genome Res.* **11**, 685–702.
- Zhang, X. & Firestein, S. (2002) *Nat. Neurosci.* **5**, 124–133.
- Robertson, H. M. (2000) *Genome Res.* **10**, 192–203.
- Robertson, H. M. (2001) *Chem. Senses* **26**, 151–159.
- Krieger, J., Raming, K., Dewey, Y. M., Bette, S., Conzelmann, S. & Breer, H. (2002) *Eur. J. Neurosci.* **16**, 619–628.
- Krieger, J., Klink, O., Mohl, C., Raming, K. & Breer, H. (2003) *J. Comp. Physiol. A* **189**, 519–526.
- Todres, E. Z., Nardi, J. B. & Robertson, H. M. (2000) *Insect Mol. Biol.* **9**, 581–590.
- Troemel, E. R., Chou, J. H., Dwyer, N. D., Colbert, H. A. & Bargmann, C. I. (1995) *Cell* **83**, 207–218.
- Sengupta, P., Chou, J. H. & Bargmann, C. I. (1996) *Cell* **84**, 899–909.
- Bray, S. & Amrein, H. (2003) *Neuron* **39**, 1019–1029.

# A genomic perspective on nutrient provisioning by bacterial symbionts of insects

Nancy A. Moran<sup>\*\*</sup>, Gordon R. Plague<sup>\*†</sup>, Jonas P. Sandström<sup>§</sup>, and Jennifer L. Wilcox<sup>\*</sup>

<sup>\*</sup>Department of Ecology and Evolutionary Biology and <sup>†</sup>Center for Insect Science, University of Arizona, Tucson, AZ 85721; and <sup>§</sup>Department of Entomology, Swedish University of Agricultural Sciences, SE-75598 Uppsala, Sweden

Many animals show intimate interactions with bacterial symbionts that provision hosts with limiting nutrients. The best studied such association is that between aphids and *Buchnera aphidicola*, which produces essential amino acids that are rare in the phloem sap diet. Genomic studies of *Buchnera* have provided a new means for inferring metabolic capabilities of the symbionts and their likely contributions to hosts. Despite evolutionary reduction of genome size, involving loss of most ancestral genes, *Buchnera* retains capabilities for biosynthesis of all essential amino acids. In contrast, most genes duplicating amino acid biosynthetic capabilities of hosts have been eliminated. In *Buchnera* of many aphids, genes for biosynthesis of leucine and tryptophan have been transferred from the chromosome to distinctive plasmids, a feature interpreted as a mechanism for overproducing these amino acids through gene amplification. However, the extent of plasmid-associated amplification varies between and within species, and plasmid-borne genes are sometimes fewer in number than single copy genes on the (polyploid) main chromosome. This supports the broader interpretation of the plasmid location as a means of achieving regulatory control of gene copy number and/or transcription. *Buchnera* genomes have eliminated most regulatory sequences, raising the question of the extent to which gene expression is moderated in response to changing demands imposed by host nutrition or other factors. Microarray analyses of the *Buchnera* transcriptome reveal only slight changes in expression of nutrition-related genes in response to shifts in host diet, with responses less dramatic than those observed for the related nonsymbiotic species, *Escherichia coli*.

Symbioses between bacteria and eukaryotes, i.e., chronic infections that are part of the normal life history of the host and are often beneficial, are ubiquitous in nature but have historically received little attention from experimental biologists or ecologists. This situation has been reversed in the last few years, which have seen a surge of interest and progress in this field (1). One useful view of these symbioses is as persistent intimate associations in which partners interact through the transfer of molecules, particularly small molecules that are essential to the growth requirements of each organism. However, the interdependence of partners that makes symbiosis so intriguing also has made it difficult to study. In particular, the inability to culture most symbionts has thwarted characterization of their chemical interactions with hosts. Recently, however, the acquisition of DNA sequences, including whole genomic sequences, of symbionts has enabled major progress in defining the biosynthetic capabilities that underlie contributions to hosts.

Invertebrates generally, and insects especially, show a particularly striking variety of symbioses with bacteria. Many of these associations are conspicuous because of the presence of a “bacteriome” (or “mycetome”), a specialized structure in the host body that houses the symbionts and that can occupy a substantial proportion of the host biomass. The symbionts are often intracellular but frequently enclosed within the cytoplasm

by a host-derived membrane. Bacteriome-associated symbioses of animals, with emphasis on arthropods, were studied extensively by light microscopy during the first half of the 20th century and summarized in a book by the prominent symbiosis researcher Paul Buchner (2). The volume describes a kind of fairyland of intimate and highly specialized biological interactions, many of which have not been further studied. Despite the anatomical and histological diversity of these symbiotic arrangements, Buchner proposed a unified functional role for bacteriome occupants, hypothesizing that their *raison d'être* is the provisioning of nutrients that animals are unable to synthesize themselves and that are absent or limiting in the specialized diets exploited by particular animal groups. Extensive sections of his book are devoted to insects with highly restricted diets, such as plant sap or vertebrate blood. Although nutritional provisioning has been better documented for some associations than others, the overall evidence supports the view that symbiosis has often enabled the expansion of animal niches and has contributed substantially to evolutionary diversification. The macroevolutionary role of nutrition-based symbiosis can be appreciated by recognizing that insects feeding on phloem or xylem sap, including aphids, psyllids, whiteflies, scale insects, planthoppers, cicadas, spittlebugs, and most leafhoppers, are dependent for their way of life on obligate bacterial endosymbionts.

Recently, molecular approaches have enabled substantial progress in understanding the evolution and functioning of the obligate symbionts of several insect hosts. Based on DNA sequence data, we now know that intracellular, bacteriome-associated symbiotic associations of insects, involving vertical transmission through eggs, typically descend from ancient infections of ancestors dating back 100 million years or more (1). Phylogenetic studies have also revealed that many of the intracellular symbionts within insects are closely related to the well studied experimental organism, *Escherichia coli*, for which functional information is available for many gene products. This fact has provided an ideal basis for using gene sequences to infer symbiont metabolic capabilities, including their contributions to host nutrition.

Here we consider the question of how a bacterial genome has been modified in the context of the new demands of symbiosis, especially nutrient provisioning to host tissues. We focus on the best-studied insect symbiont, *Buchnera aphidicola*, which has coevolved with its aphid hosts for >100 million years (3). We have divided the question into two main aspects: (i) modifications of genome contents and organization and (ii) modification of controls on gene expression. Because different hosts encoun-

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Chemical Communication in a Post-Genomic World,” held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

<sup>\*</sup>To whom correspondence should be addressed. E-mail: nmoran@u.arizona.edu.

© 2003 by The National Academy of Sciences of the USA

ter varying nutritional demands and varying dietary conditions, we expect some heritable modifications manifested as differences among species or populations. In addition, individual host species must deal with a variety of nutritional limitations during the course of a season or while using different resources or habitats; thus, we might expect that symbiont productivity can be modified as a result of environmental factors.

### Nutritional Provisioning as a Primary Task of Symbionts in Sap-Feeders

Animals generally are lacking in pathways underlying amino acid production, typically requiring 10 “essential” amino acids in the diet; this limitation must stem from ancient gene loss in an ancestor of Metazoa. Animals also have a large number of dietary requirements for enzymatic cofactors (or “vitamins”). Studies of phloem sap composition, based on collections from the severed mouthparts of aphids, show that nitrogen is almost exclusively present as amino acids, and that nonessential amino acids predominate (4). In general, the essential amino acids are present in relative concentrations considerably lower than those found in aphid proteins or in dietary requirements determined for insects. This imbalance implies that insect growth potential can be enhanced by symbionts, which absorb abundant amino acids and sugars from the host and use these to generate limiting amino acids through biosynthetic capabilities lacking in insects. However, both the total concentrations of pooled amino acids as well as the relative concentrations of individual essential amino acids can differ dramatically among plant species and phenological stages (e.g., refs. 5 and 6). Amino acid levels and profile also fluctuate, depending on the plant response to aphid feeding (7). Experimental studies have provided some evidence that the nonessential amino acid glutamate is the major nitrogenous compound exported from the aphid and used by *Buchnera* for production of the essential, limiting amino acids (8–10).

### Reading the Genome Sequences of Insect Symbionts: Old Tools for New Tasks

Complete genome sequences are available for *Buchnera* of three distantly related aphid species [*Acyrtosiphon pisum* (11), *Schizaphis graminum* (12), and *Baizongia pistaciae* (13)], for *Wigglesworthia brevipalpis*, the obligate symbiont of tsetse fly (14), and for *Blochmannia camponotii*, the obligate symbiont of carpenter ants (15). The genome sequences of several other insect symbionts are expected to be available soon. Genomic sequences allow us to identify which pathways symbionts possess, and, even more uniquely, which pathways they lack. This information has greatly extended our knowledge of the nutritional role of symbionts within hosts and provides a foundation for further studies of chemical exchange between host and symbiont.

Based on the systems studied so far, bacteriome associates in insects show extensive genomic modifications, including massive loss of ancestral genes and many rearrangements affecting order and orientation of genes on the chromosome. For *Buchnera*, *Wigglesworthia*, and *Baumannia*, genomes are reduced to only ~20% of the genes present in ancestors and modern relatives, and each contains many unique gene arrangements, yielding only small regions of synteny with related bacteria (15–17). Furthermore, in contrast to most other bacterial genomes including those of its closest relatives, these symbiont genomes show little evidence of gene uptake. Indeed, *Buchnera* lineages appear to lack any novel genes not present in related free-living bacteria, and comparisons of *Buchnera* genomes, that diverged with their ancestors >100 million years ago, show lower levels of gene acquisition or genomic rearrangements than do any other fully sequenced bacteria (11).

The combination of genome reduction and absence of gene acquisition implies that, in these insect symbioses, the basis for the symbiotic contributions lies in ancestral genes and pathways,

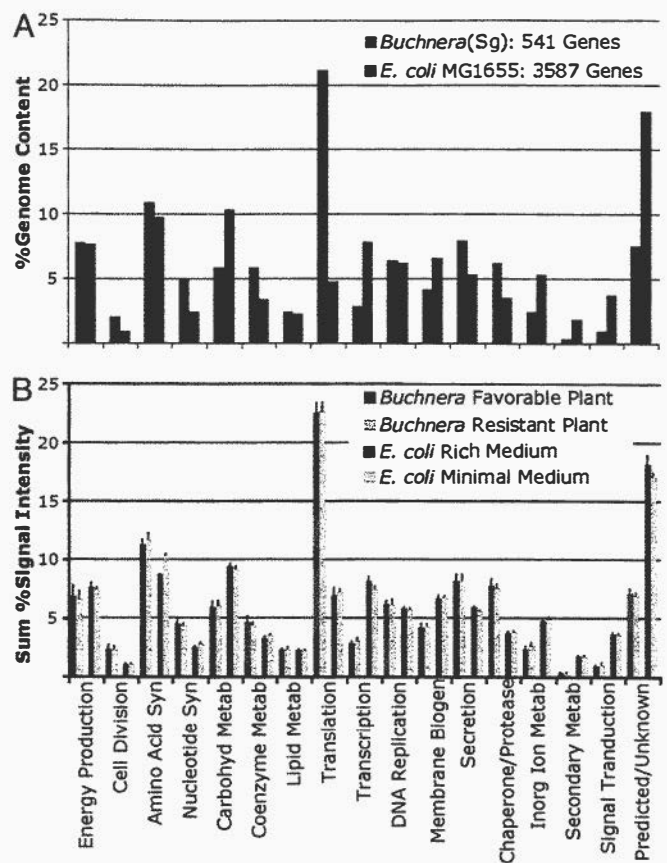


Fig. 1. (A) Proportional representation of genes grouped by functional categories within the genomes of *Buchnera* (*S. graminum*) and *E. coli*, based on functional categorizations established for the Clusters of Orthologous Groups (COGs) (52). (B) Representation of COG functional categories in the transcriptomes of *Buchnera* (*S. graminum*) and *E. coli*. *Buchnera* were from mixed-age host aphids grown at 20°C on plants differing in phloem concentrations of total and essential amino acids (barley versus resistant wheat cultivar; plants as in ref. 7). *E. coli* data are from the public ASAP database (53) and depict transcriptomes of early log-phase cultures grown in amino acid “Rich” LB medium versus “Minimal” Mops medium plus 0.2% glucose (J. Glasner, personal communication). Microarray attributes, RNA extraction procedures, and data normalization steps were similar for both bacterial species (42, 49). Bar height is the sum of gene signals in a given COG divided by the number of genes in that COG; bars are standard deviations based on experimental replicates (five replicates per *Buchnera* treatment; two and six replicates for *E. coli* LB and Minimal Media treatments).

present in free-living relatives, that were coopted for the purpose of provisioning hosts after the initiation of symbiosis. For both *Buchnera* and *Wigglesworthia*, the maintenance of ancestral genes for nutrient biosynthesis provides a clear indication of their mutualistic relationship to hosts and sharply contrasts with the loss of these same pathways from genomes of obligately parasitic bacteria, which acquire metabolites from host tissues (1). *Buchnera* of *A. pisum* has genes for all of the essential amino acid pathways (total of 54 loci), plus genes for fixation of inorganic sulfur and for synthesis of riboflavin (13). Because unneeded genes are quickly eliminated from these genomes, the retention of these biosynthetic pathways gives a clear indication of a continued contribution of the corresponding nutrients to hosts.

The prominence of essential amino acid production as an activity of *Buchnera* is evident from the set of genes retained (Fig. 1A). The *E. coli* genome encodes ~60 genes for the core biosynthetic pathways underlying production of essential amino acids, ~20 for the core pathways for nonessential amino acids

**Table 1. The ratios of copies of plasmid-borne amino acid biosynthetic genes (*leuABCD*, *trpEG*) to chromosomal gene copies for *Buchnera* of different aphid species and strains**

Aphid host	Aphid clone/population	<i>leuABCD</i>	<i>trpEG</i>
<i>A. pisum</i>	12 United Kingdom clones	–	2.4–16.2*
	N. A. Moran lab clone 5A (Madison, WI)	0.6	4.8
<i>Diuraphis noxia</i>	P. Baumann lab clone (Lincoln, NE)	0.9 <sup>†</sup>	1.8 <sup>‡</sup>
	South Africa population	0.3	0.4
<i>Rhopalosiphum maidis</i>	N. A. Moran lab clone (Tucson)	–	0.3
<i>S. graminum</i>	Biotype B (K. A. Shufran lab clone)	–	0.5
	Biotype E (T. Mittler lab clone)	23.5 <sup>†</sup>	14.5 <sup>§</sup>
	Biotype E (P. Baumann lab clone)	1.4	2.1
	Biotype E (N. A. Moran lab clone)	1.9	1.5
	Biotype E (K. A. Shufran lab clone)	1.6	2.6
	Biotype G (K. A. Shufran lab clone)	0.5	2.4
	Biotype SC (K. A. Shufran lab clone)	–	0.5
<i>Uroleucon ambrosiae</i>	86 individuals, 15 U.S. populations	0.5–2.8 <sup>¶</sup>	0.3–1.9 <sup>¶</sup>

Unless otherwise noted, values are previously unpublished and were estimated by using real-time quantitative PCR as in ref. 25 (primer sequences are available upon request). The reason for the discrepancy among *Buchnera* (*S. graminum*) estimates is uncertain, although the concordance of the quantitative PCR estimates across biotypes provides support for their accuracy.

\*Ref. 27; calculated by using quantitative DNA hybridization.

<sup>†</sup>Ref. 26; calculated by using quantitative DNA hybridization.

<sup>‡</sup>Ref. 30; calculated by using quantitative DNA hybridization.

<sup>§</sup>Ref. 18; calculated by using quantitative DNA hybridization.

<sup>¶</sup>Ref. 25; calculated by using real-time quantitative PCR.

plus >50 genes involved in transport of amino acids. Of these, *Buchnera* retains almost all of the genes for production of essential amino acids and only two (or four, depending on how amino acids are categorized) for synthesis of nonessential amino acids or amino acid transport (13). Overall, the core genes for biosynthesis of essential amino acids comprise a larger proportion ( $\approx 10\%$ ) of the *Buchnera* genome than of the *E. coli* genome (<2%).

#### How Are Genomes of Obligate Symbionts That Provision Hosts Modified Relative to Those of Free-Living Relatives?

The large majority of genes underlying amino acid biosynthetic pathways are confined to the main *Buchnera* chromosome. However, one of the early striking discoveries about the *Buchnera* genome, seemingly linked to the nutrient-provisioning role, was the finding that rate-limiting genes for biosynthesis of tryptophan (*trpEG*, ref. 18) and the genes for biosynthesis of leucine (*leuABCD*, ref. 19) are encoded on two distinct types of multicopy plasmids. Each of these plasmids has apparently evolved only one or very few times (20–22), and some early-branching *Buchnera* lineages retain these genes on the main chromosome in an ancestral position (e.g., refs. 12 and 23).

The initial interpretation of the plasmid position of *trpEG* and *leuABCD* was that it enabled overproduction of the end products for host nutrition, because the plasmid location allowed amplification of gene copy number relative to single copy chromosomal genes. The subsequent discovery that each *Buchnera* cell contains multiple chromosomal copies (24) raised the more general interpretation of the plasmid location of *trpEG* and *leuABCD* as an arrangement allowing either increase or decrease in copy number relative to chromosomal genes (25), and perhaps allowing regulation of the level of amplification in response to environmental conditions. Ratios of plasmid-borne *trpEG* and *leuABCD* copies to chromosomal gene copies do vary among species, possibly in a direction that is related to host needs (26). Table 1 summarizes previously published and newly estimated ratios of plasmid-borne amino acid biosynthetic genes to single copy chromosomal genes in *Buchnera* of different aphid strains and species. The ratios of plasmid-borne genes to single-copy chromosomal genes range from much less than 1-fold (implying fewer plasmid copies than chromosome copies within a cell) to

>10-fold, with rather similar ranges for relative amplification of both *leu* and *trp* genes (ref. 26 and Table 1). Values vary both within and between species, with at least part of the variation heritable (e.g., ref. 27). Thus, the ratios of copies of *trpE* to *trpB* (a single copy chromosomal gene) appear stable among different aphids of the same maternal line (27). Also, the amplification level may be quite constant for some species and more variable for others, based on analyses of geographically dispersed isolates of several species (refs. 25, 27, and 28 and Table 1).

In the case of *trpEG*, the variation in ratio of plasmid-borne to chromosomal genes is due both to differences in numbers of repeats per plasmid (21) and to differences in relative plasmid copy number. But another variable affecting the number of functional *trpEG* copies is the frequent inactivation of some gene copies. Early stop codons, large deletions, and frameshift mutations have been found in at least some of the plasmid-borne *trpEG* copies of several aphid species (13, 28–31) and can be geographically widespread within a species (28).

The rarity of origins of plasmid amplification, the apparent heritability and stability of amplification levels, and the frequent inactivation of *trpEG* all suggest that levels of plasmid amplification do not provide a route for quick responses to changes in demands for leucine or tryptophan production. Thus, although the plasmids are likely to be involved in adjusting production of amino acids, the exact mechanism is not clear. This picture is supported by data suggesting that tryptophan production is not correlated with the ratio of *trpEG* to single copy chromosomal gene copies among different *A. pisum* strains (27), although these ratios could have included inactivated *trpEG* copies (see 32). Observations so far do not exclude the possibility of some feedback control of plasmid copy number by amino acid concentrations, as a major heritable component does not rule out plasticity in the ratio of plasmid-borne to chromosomal genes.

#### Degradation of Symbiont Provisioning Through Gene Inactivation and Loss

In contrast to the situation in *A. pisum* in which all of the essential amino acids pathways are conserved, several lines of evidence indicate that individual pathways are sometimes inactivated or lost in particular *Buchnera* strains. The absence of a gene can only be definitively concluded when whole genome sequences

are determined. Among the three *Buchnera* genomes now available for distantly related aphid species with very different patterns of host plant use, loss of the pathway for fixation of inorganic sulfur occurred independently in two lineages; this loss precludes synthesis of cysteine and methionine by using inorganic sulfur sources (11, 12). In addition, *Buchnera* of *Baizongia pistaciae* has lost capacity for arginine biosynthesis, with the relevant genes deleted from positions at several locations in the genome (12). The fact that two of the three sequenced *Buchnera* genomes have lost functionality of some nutrition-related pathways strongly suggests that such losses are widespread evolutionary events. This hypothesis is reinforced by nutritional studies suggesting that certain amino acids are dietary requirements for some aphid/*Buchnera* isolates. Results of nutritional studies on *A. pisum* and several other aphid species have been interpreted as indicating species-specific or strain-specific requirements for most of the 10 essential amino acids (including arginine, histidine, isoleucine, leucine, lysine, phenylalanine, threonine, tryptophan, and valine) (e.g., refs. 33 and 34). Nutritional studies also have shown variation among species in whether inorganic sulfate could replace either or both of the sulfur-containing amino acids (e.g., ref. 35).

Genomic sequence data show that the essential amino acid pathways are generally present in *Buchnera* but sometimes inactivated or lost (11, 12), providing striking support for the following interpretation of nutritional results. First, the usual absence of dietary requirement for essential amino acids, which contrasts with animals generally, reflects provisioning by *Buchnera*. Second, individual pathways are repeatedly lost in individual *Buchnera* strains, resulting in reversion to host dependence on dietary sources. Because the *Buchnera* of most aphids appear to have retained a large complement of amino acid biosynthetic pathways, a corollary is that the lineages undergoing these inactivation events may be evolutionary dead ends. The report of substantially smaller genomes in the *Buchnera* of some aphid species (36) raises the possibility that even more erosion of biosynthetic capabilities has occurred in some lineages, which may encounter enriched diets or rely on additional, novel symbionts (or “secondary” symbionts) (e.g., refs. 37–40) for some amino acids.

The emerging picture is that provisioning services of obligate symbionts can be eroded by losses of gene function. The lack of recombination or gene uptake by *Buchnera* implies that such losses are irreversible. Inactivation of genes underlying biosynthetic abilities may occur under circumstances in which dietary supplies are sufficient, and data on phloem sap contents indicates that particular amino acids may be adequate in some plants (4, 41). Nonetheless, these losses are expected to impose permanent limitations on future abilities of aphid lineages to colonize novel foodplants. Possibly, associations with secondary symbionts may be driven by the need to replace functions formerly provided by *Buchnera*.

### Integration of Symbiont Gene Expression with the Nutritional Economy of Hosts

Experiments based on full genome microarrays provide a general overview of patterns of gene expression in a symbiont genome, allowing us to ask questions of how expression diverges from that of related free-living bacteria and to address the issue of whether symbionts are able to alter gene expression to fit host needs. Laboratory and statistical methods underlying the microarray results presented here have been described (42), and we limit ourselves to a generalized overview of symbiont gene expression patterns. Fig. 1B contrasts transcript abundances, grouped by gene functional categories, between *Buchnera* of *Schizaphis graminum* and *E. coli*. *Buchnera* is represented as samples derived from aphids feeding on nutritionally favorable and unfavorable plants (7); *E. coli* is represented as cultures from growth media with and without preformed amino acids (J. D.

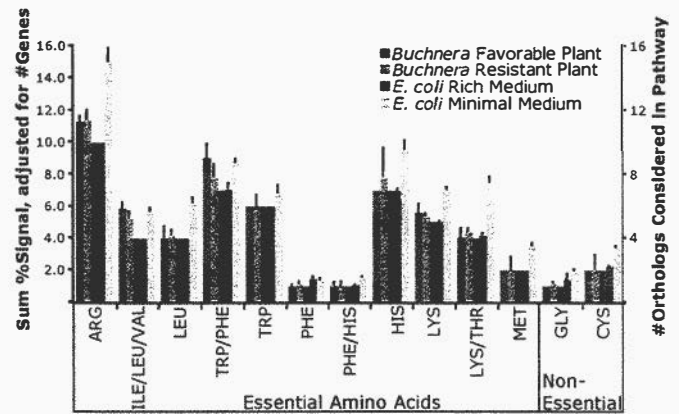


Fig. 2. Transcript abundances of genes in essential and nonessential amino acid biosynthetic pathways. Colored bars and left y axis are average transcript signals summed for a given pathway and averaged over experimental replicates. Gray bars and right y axis show numbers of genes per pathway grouping and represent expected signal intensity based on numbers of genes. All data are for orthologs shared between *Buchnera*(Sg) and *E. coli*. ARG, *argA*, *argB*, *argC*, *argD*, *argE*, *carA*, *carB*, *argF*, *argG*, and *argH*; ILE/LEU/VAL, *ilvH*, *ilvI*, *ilvC*, and *ilvD*; LEU, *leuA*, *leuC*, *leuD*, and *leuB*; TRP/PHE, *aroH*, *aroB*, *aroD*, *aroE*, *aroK*, *aroA*, and *aroC*; TRP, *trpE*, *trpD*, *trpC*, *trpA*, and *trpB*; PHE, *pheA*; PHE/HIS, *hisC*; HIS, *hisG*, *hisI*, *hisA*, *hisH*, *hisF*, *hisB*, and *hisD*; LYS, *dapA*, *dapB*, *dapD*, *dapF*, and *lysA*; LYS/THR, *thrA*, *asd*, *thrB*, and *thrC*; MET, *metE*, and *metF*; GLY, *glyA*; CYS, *cysE*, and *cysK*. Block shading below the graph groups amino acid pathways as branched, aromatic, etc.

Glasner, personal communication). For both organisms, the profiles across the functional categories are strikingly similar for genome content (Fig. 1A) versus transcriptome content (Fig. 1B). The fraction of genes in a particular clusters of orthologous groups (COG) category shows a highly significant linear relationship to the pooled expression level for that category, both for *Buchnera* (for both treatments,  $R^2 = 0.98$ ,  $P < 0.0001$ ) and *E. coli* (for both treatments,  $R^2 = 0.93$ ,  $P < 0.0001$ ). As noted above, the *Buchnera* genome has differentially retained genes for essential amino acid biosynthesis (a subset of the amino acid metabolism category found in *E. coli* as shown in Fig. 1), and this greater representation extends to the transcriptome, suggesting greater investment in functions that underlie host provisioning. Another notable feature is that chaperonins are relatively highly expressed in *Buchnera*, even under nonstress conditions, as previously noted (Fig. 1B and ref. 42).

The parallel between number of genes and transcript abundance extends to the level of individual amino acids pathways for both *Buchnera* and *E. coli*, although, as expected, the growth media has a major influence on the expression levels in *E. coli* (Fig. 2). The main amino acids that are deficient in the *S. graminum* diet (the phloem sap of grasses) are arginine and lysine, which are two of the most common amino acids in the insect proteins (4). Transcript abundances for genes in these pathways tend to be high (Fig. 2). In contrast, genes underlying production of methionine and cysteine show relatively low transcript levels, possibly reflecting the fact that these pathways are of no use for amino acid provisioning in this *Buchnera* species, because the upstream pathway for sulfur fixation has been inactivated (11).

As noted above, aphids feed on heterogeneous populations of plants that impose varying demands on *Buchnera*'s biosynthetic capabilities, raising the question of how the symbiont genome responds. In free-living bacteria, availability of a particular amino acid has a dramatic effect on its rate of production, and this feedback control arises from a myriad of mechanisms, involving transcript production as well as inhibition of enzymes. For the relatively long and energetically expensive pathways

underlying production of the essential amino acids, transcript production is typically governed by multiple regulatory mechanisms in *E. coli* and other bacteria. These same pathways underlie the central contribution of *Buchnera* to host nutrition. In view of the varied diets of hosts and the apparent cost of excess production of biosynthetic enzymes and end products, we might expect mechanisms for transcriptional regulation of these genes in *Buchnera*. This expectation is reinforced by the fact that amino acid provisioning comprises a relatively large proportion of the genome and transcriptome of *Buchnera*.

But, if transcriptional control mechanisms are present at all, they are altered dramatically in *Buchnera*. The first evidence for loss of the usual regulatory mechanisms came from examination of regions upstream to structural genes for the biosynthetic enzymes. These lack leader peptides and show altered sequences that indicate degraded or altered binding properties, as noted, for example, for aromatic amino acid pathways (43, 44). The first complete genome of *Buchnera* revealed the lack of recognizable orthologs of any of the regulatory proteins that bind with free amino acids to block or activate transcriptional promoter sites (13), although *metR* was identified as an intact ORF in the second genome (11).

The loss of recognizable gene regulatory features in sequence analyses contrasts with findings of physiological studies. Several of these have provided evidence that aphids with an intact symbiosis can down-regulate their production of specific amino acids when abundant. This suggests that *Buchnera* has one or more negative feedback mechanisms governing amino acid production (45–47). For example, in studies on *Aphis fabae* provided with radioactively labeled glutamic acid, a greater proportion of label was incorporated into the essential amino acid isoleucine on diets lacking isoleucine than on diets containing it (48). Similarly, in *A. pisum*, neosynthesis of all essential amino acids was depressed on a diet with plentiful essential amino acids relative to diets with limiting amounts, with synthesis of histidine and arginine completely suppressed (45). Although it is plausible that the apparent mechanisms for regulating amino acid production involve only feedback inhibition of the enzymes rather than transcriptional regulation, this would provide a stark contrast to the situation in other bacteria, in which regulation of transcription is a prominent means of control.

As reported in Wei *et al.* (49), the fraction of the *E. coli* transcriptome devoted to amino acid biosynthetic genes increases markedly for colonies grown on minimal media (no exogenous amino acids provided). For *Buchnera*, in contrast, pooled transcript abundance for genes underlying amino acid biosynthesis differs little from aphids confined to plants with dramatically different nutritional qualities (Fig. 1B). This relative constancy of transcript abundances extends to individual

biosynthetic pathways (Fig. 2). The low response to changes in plant quality could reflect a lack of symbiont ability to adaptively alter transcription rates. However, other explanations are possible, including experimental limitations preventing detection of fine-tuned feedback mechanisms; for example, the pooling of RNA from maternal and embryonic symbionts may obscure biologically meaningful changes. Regulation of amino acid biosynthesis could be achieved, in part, through regulation of amino acid transport from host to symbiont. The *Buchnera* genome lacks apparent importers, but each *Buchnera* cell is enclosed within a host membrane that is capable of active transport of glutamate and aspartate, which are substrates for biosynthesis of the other amino acids (10). Such host control could govern the overall levels of amino acid biosynthesis within *Buchnera* cells but would not provide a way to fine-tune the production of individual amino acids according to host needs. The aphids seem not to alter their assimilation of essential amino acids in response to different dietary concentrations of these compounds (50), indicating that their regulation within aphids must involve differential symbiont productivity.

## Conclusions

The genomes of the obligate symbionts of insects provide a clear picture of modification of an ancestral genome for a particular role in exchanging nutrients with the host. The picture emerging so far tells us that the essential genetic capabilities are ancient and consist of pathways that are widely distributed among nonsymbiotic bacteria. It appears that the maternally inherited, obligate bacteriome-associates of insects are derived through reduction and minor modifications of ancestral genomes rather than through acquisition or invention of novel “symbiosis” genes. Nonetheless, some of the critical elements for maintenance of symbiosis remain unidentified. We do not yet know why the  $\gamma 3$  Proteobacteria have given rise to a large proportion of the symbiotic lineages living in insects. Perhaps this group possesses some yet-to-be-discovered key genes that enable symbiosis. One possibility is the frequent presence of a type III secretion apparatus that might serve as the initiator of the intracellular association, as in the secondary symbionts of tsetse flies and the related symbionts of grain weevils (51). In addition, the control of gene expression may involve novel mechanisms, a possibility that is supported by the apparent plasticity of *Buchnera* contributions depending on host diet.

We thank H. McLaughlin and H. Dunbar for laboratory assistance, K. A. Shufan for providing aphid clones, J. Glasner for access to unpublished data, and H. Ochman for comments on the manuscript. Funding was from National Science Foundation Grant 9978518 (to N.A.M.), a National Science Foundation postdoctoral fellowship (to J.L.W.), and a National Institutes of Health postdoctoral fellowship (to G.R.P.).

- Moran, N. A. (2002) *Cell* **108**, 583–586.
- Buchner, P. (1965) *Endosymbiosis of Animals with Plant Microorganisms* (Wiley, New York).
- Munson, M. A., Baumann, P., Clark, M. A., Baumann, L., Moran, N. A., Voegtlin, D. J. & Campbell, B. C. (1991) *J. Bacteriol.* **173**, 6321–6324.
- Sandström, J. P. & Moran, N. A. (2001) *Physiol. Entomol.* **26**, 202–211.
- Bernays, E. A. & Klein, B. A. (2002) *Physiol. Entomol.* **27**, 275–284.
- Sandström, J. & Pettersson, J. (1994) *J. Insect Physiol.* **40**, 947–955.
- Telang, A., Sandström, J., Dyreson, E. & Moran, N. A. (1999) *Entomol. Exp. Appl.* **91**, 402–412.
- Febvay, G., Liadouze, I., Guillaud, J. & Bonnot, G. (1995) *Arch. Insect Biochem. Physiol.* **29**, 45–69.
- Sasaki, T. & Ishikawa, H. (1995) *J. Insect Physiol.* **41**, 41–46.
- Whitehead, L. F. & Douglas, A. E. (1993) *J. Gen. Microbiol.* **139**, 821–826.
- Tamas, I., Klasson, L., Canback, B., Naslund, A. K., Eriksson, A. S., Wernegreen, J. J., Sandström, J. P., Moran, N. A. & Andersson, S. G. (2002) *Science* **296**, 2376–2379.
- van Ham, R. C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J. M., Jiménez, L., Postigo, M., Silva, F. J., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100**, 581–586.
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. (2000) *Nature* **407**, 81–86.
- Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M. & Aksoy, S. (2002) *Nat. Genet.* **32**, 402–407.
- Gil, R., Silva, F. J., Zientz, E., Delmotte, F., Gonzalez-Candelas, F., Latorre, A., Rausell, C., Kamerbeek, J., Gadau, J., Holldobler, B., *et al.* (2003) *Proc. Natl. Acad. Sci. USA* **100**, 9388–9393.
- Moran, N. A. & Mira, A. (2001) *Genome Biol.* **2**, RESEARCH0054.
- Silva, F. J., Latorre, A. & Moya, A. (2001) *Trends Genet.* **17**, 615–618.
- Lai, C. Y., Baumann, L. & Baumann, P. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 3819–3823.
- Bracho, A. M., Martinez-Torres, D., Moya, A. & Latorre, A. (1995) *J. Mol. Evol.* **41**, 67–73.
- Baumann, L., Baumann, P., Moran, N. A., Sandström, J. & Thao, M. L. (1999) *J. Mol. Evol.* **48**, 77–85.
- Rouhbakhsh, D., Clark, M. A., Baumann, L., Moran, N. A. & Baumann, P. (1997) *Mol. Phylogenet. Evol.* **8**, 167–176.
- van Ham, R. C., González-Candelas, F., Silva, F. J., Sabater, B., Moya, A. & Latorre, A. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10855–10860.
- Lai, C. Y., Baumann, P. & Moran, N. A. (1995) *Insect Mol. Biol.* **4**, 47–59.
- Komaki, K. & Ishikawa, H. (1999) *J. Mol. Evol.* **48**, 717–722.



25. Plague, G. R., Dale, C. & Moran, N. A. (2003) *Mol. Ecol.* **12**, 1095–1100.
26. Thao, M. L., Baumann, L., Baumann, P. & Moran, N. A. (1998) *Curr. Microbiol.* **36**, 238–240.
27. Birkle, L. M., Minto, L. B. & Douglas, A. E. (2002) *Physiol. Entomol.* **27**, 302–306.
28. Wernegreen, J. J. & Moran, N. A. (2000) *Proc. R. Soc. London Ser. B* **267**, 1423–1431.
29. Baumann, L., Clark, M. A., Rouhbakhsh, D., Baumann, P., Moran, N. A. & Voegtlin, D. J. (1997) *Curr. Microbiol.* **35**, 18–21.
30. Lai, C. Y., Baumann, P. & Moran, N. A. (1996) *Appl. Environ. Microbiol.* **62**, 332–339.
31. Wernegreen, J. J., Richardson, A. O. & Moran, N. A. (2001) *Mol. Phylogenet. Evol.* **19**, 479–485.
32. Birkle, L. M. & Douglas, A. E. (1999) *Heredity* **82**, 605–612.
33. Srivastava, P. N. & Auclair, J. L. (1975) *J. Insect Physiol.* **21**, 1865–1871.
34. Srivastava, P. N. (1987) in *World Crop Pests*, eds. Mink, A. K. & Harrewijn, P. (Elsevier, Amsterdam), Vol. 2A, pp. 99–122.
35. Turner, R. B. (1971) *J. Insect Physiol.* **17**, 2451–2456.
36. Gil, R., Sabater-Muñoz, B., Latorre, A., Silva, F. J. & Moya, A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 4454–4458.
37. Fukatsu, T. (2001) *Appl. Environ. Microbiol.* **67**, 5315–5320.
38. Montllor, C. B., Maxmen, A. & Purcell, A. H. (2002) *Ecol. Entomol.* **27**, 189–195.
39. Sandström, J. P., Russell, J. A., White, J. P. & Moran, N. A. (2001) *Mol. Ecol.* **10**, 217–228.
40. Tsuchida, T., Koga, R., Shibao, H., Matsumoto, T. & Fukatsu, T. (2002) *Mol. Ecol.* **11**, 2123–2135.
41. Sandström, J. & Moran, N. A. (1999) *Ent. Exp. Appl.* **91**, 203–210.
42. Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D. & Moran, N. A. (2003) *Mol. Microbiol.* **48**, 1491–1500.
43. Jimenez, N., González-Candelas, F. & Silva, F. J. (2000) *J. Bacteriol.* **182**, 2967–2969.
44. Kolibachuk, D., Rouhbakhsh, D. & Baumann, P. (1995) *Curr. Microbiol.* **30**, 313–316.
45. Febvay, G., Rahbé, Y., Rynkiewicz, M., Guillard, J. & Bonnot, G. (1999) *J. Exp. Biol.* **202**, 2639–2652.
46. Leckstein, P. M. & Llewellyn, M. (1973) *J. Insect Physiol.* **19**, 973–980.
47. Liadouze, I., Febvay, G., Guillaud, J. & Bonnot, G. (1995) *J. Insect Physiol.* **41**, 33–40.
48. Douglas, A. E., Minto, L. B. & Wilkinson, T. L. (2001) *J. Exp. Biol.* **204**, 349–358.
49. Wei, Y., Lee, Y. M., Richmond, C., Blattner, F. R., Rafalski, J. A. & LaRossa, R. A. (2001) *J. Bacteriol.* **183**, 545–556.
50. Wilkinson, T. L. & Ishikawa, H. (1999) *Entomol. Exp. Appl.* **91**, 195–201.
51. Dale, C., Plague, G. R., Wang, B., Ochman, H. & Moran, N. A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 12397–12402.
52. Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V. (2001) *Nucleic Acids Res.* **29**, 22–28.
53. Glasner, J. D., Liss, P., Plunkett, G., III, Darling, A., Prasad, T., Rusch, M., Byrnes, A., Gilson, M., Biehl, B., Blattner, F. R. & Perna, N. T. (2003) *Nucleic Acids Res.* **31**, 147–151.

# Chemical communication among bacteria

Michiko E. Taga and Bonnie L. Bassler\*

Department of Molecular Biology, Princeton University, Princeton, NJ 08544-1014

Cell–cell communication in bacteria is accomplished through the exchange of chemical signal molecules called autoinducers. This process, called quorum sensing, allows bacteria to monitor their environment for the presence of other bacteria and to respond to fluctuations in the number and/or species present by altering particular behaviors. Most quorum-sensing systems are species- or group-specific, which presumably prevents confusion in mixed-species environments. However, some quorum-sensing circuits control behaviors that involve interactions among bacterial species. These quorum-sensing circuits can involve both intra- and interspecies communication mechanisms. Finally, anti-quorum-sensing strategies are present in both bacteria and eukaryotes, and these are apparently designed to combat bacteria that rely on cell–cell communication for the successful adaptation to particular niches.

quorum sensing | autoinducer | cell–cell communication

In a process called quorum sensing, bacteria monitor the presence of other bacteria in their surroundings by producing and responding to signaling molecules known as autoinducers. The concentration of autoinducer in a given environment is proportional to the number of bacteria present; therefore, detecting autoinducers gives bacteria a mechanism for “counting” one another. Responding to autoinducers by altering gene expression gives bacteria a means to perform particular behaviors only when living in a community but not when living in isolation. Most quorum-sensing controlled behaviors are productive only when a group of bacteria carries them out in synchrony; they include bioluminescence, secretion of virulence factors, biofilm formation, sporulation, conjugation, and pigment production (1–3). Because quorum sensing allows bacteria to coordinate the behavior of the group, it enables them to take on some of the characteristics of multicellular organisms.

There are two general types of bacterial quorum-sensing systems: Gram-negative LuxIR circuits and Gram-positive oligopeptide two-component circuits. Gram-negative quorum-sensing bacteria typically possess proteins homologous to the LuxI and LuxR proteins of *Vibrio fischeri*, the bacterium in which they were initially discovered (Fig. 1A) (4). The LuxI-type proteins catalyze the formation of a specific acyl-homoserine lactone (AHL) autoinducer that freely diffuses into and out of the cell and increases in concentration in proportion to cell population density. The LuxR-type proteins each bind a specific AHL autoinducer when the concentration of autoinducer reaches a threshold level. The LuxR–AHL complexes activate transcription of target genes by recognizing and binding specific DNA sequences at quorum-sensing-regulated promoters (4–6). Some functions controlled by LuxIR-type quorum-sensing systems include plasmid conjugation in *Agrobacterium tumefaciens* (7), antibiotic production in *Erwinia carotovora* (8), biofilm production and virulence gene expression in *Pseudomonas aeruginosa* (9, 10), and expression of factors necessary for symbiosis in *Sinorhizobium meliloti* (11). Currently, there are >70 known LuxIR quorum-sensing systems in Gram-negative bacteria (1, 12–14).

Low G+C Gram-positive bacteria typically use modified oligopeptides as autoinducers (15–17). These signals are generally referred to as autoinducing polypeptides (AIPs) (Fig. 1B). AIPs are produced in the cytoplasm as precursor peptides and are subsequently cleaved, modified, and exported. AIPs specifically interact with the external domains of membrane-bound two-component sensor kinase proteins. Interaction of the autoinducer with its cognate sensor stimulates the kinase activity of the sensor kinase protein, resulting in the phosphorylation of its partner response regulator protein. The phosphorylated response regulator protein binds DNA and alters the transcription of target genes. Some examples of behaviors controlled by AIP quorum-sensing systems include genetic competence and sporulation in *Bacillus subtilis* (18, 19), competence for DNA uptake in *Streptococcus pneumoniae* (20), and virulence factor expression in *Staphylococcus aureus* (21) and *Enterococcus faecalis* (22).

Despite the fact that the autoinducer signals and detection apparatuses can be highly similar, both LuxIR and oligopeptide two-component-type quorum-sensing systems function such that a response is elicited only to the autoinducer of the bacterial species that produced it. This signaling specificity stems in large part from subtle differences in the autoinducer molecules and their receptors. Here we discuss what is known about the requirements for species-specific cell–cell communication. Quorum-sensing circuits also exist that are species nonspecific; their properties are discussed below.

## Specificity in AHL Communication Systems

AHL autoinducers all share a common homoserine lactone moiety and differ only in their acyl side chain moieties (Fig. 2A) (3, 12). However, an autoinducer of one bacterial species rarely influences expression of target genes in another species. Two primary factors control this exquisite signaling specificity: first, the substrate specificity of the LuxI-like proteins, and second, specificity in the binding of the LuxR-type proteins to their cognate AHLs.

LuxI proteins link the side chain group of specific acyl-acyl carrier proteins to the homocysteine moiety of *S*-adenosylmethionine (23–25). These reactions are very precise, because LuxI proteins recognize only the ACP containing a specific acyl chain moiety. The reliability of this reaction is vital to maintaining species-specific communication within each quorum-sensing system, because it ensures that each LuxI protein generates only the correct AHL signal molecule.

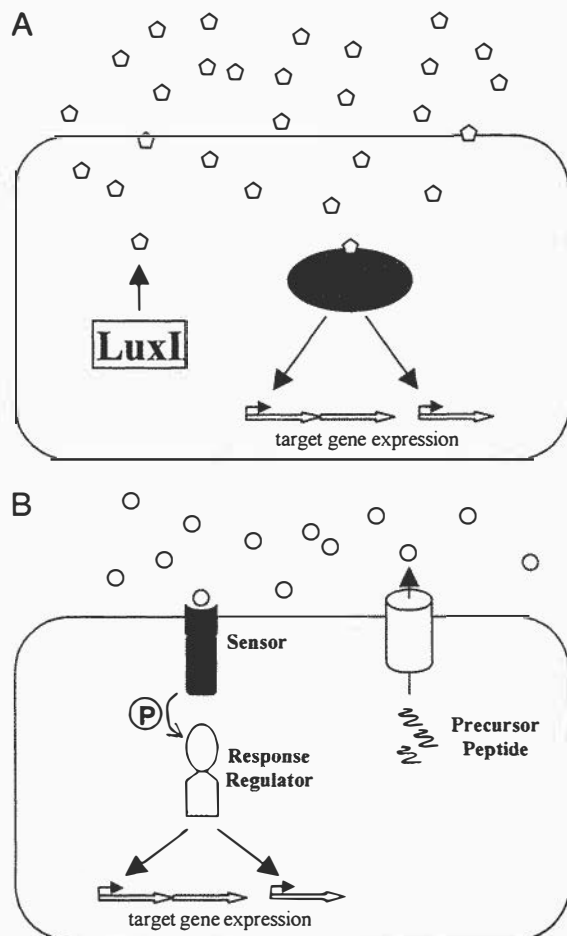
The molecular interactions underlying the specificity in LuxI-directed signal production have been explored by structural analyses of EsaI, the LuxI homologue of *Pantoea stewartii* (26).

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Chemical Communication in a Post-Genomic World,” held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: AHL, acyl-homoserine lactone autoinducer; AIP, autoinducing peptide; Al-1, 3-hydroxybutanoyl-homoserine lactone; Al-2, 3a-methyl-5,6-dihydrofuro[2,3-d][1,3,2]dioxaborole-2,2,6,6a-tetraol.

\*To whom correspondence should be addressed. E-mail: bbassler@molbio.princeton.edu.

© 2003 by The National Academy of Sciences of the USA



**Fig. 1.** Canonical quorum-sensing circuits. (A) In typical Gram-negative LuxI/LuxR circuits, the LuxI-type protein catalyzes the synthesis of an AHL autoinducer (pentagons). The LuxR-type protein binds the AHL and controls the expression of target genes. (B) In typical Gram-positive AIP two-component quorum-sensing circuits, precursor peptides are cleaved, modified, and exported by dedicated transporters. The resulting AIPs (circles) are detected by two-component sensor-histidine kinases, and sensory information is relayed by phosphorylation (P) of cognate response regulators that, in turn, control target gene expression. The proteins responsible for autoinducer binding are shaded in gray.

EsaI catalyzes the formation of the AHL 3-oxohexanoyl-homoserine lactone. The x-ray crystal structure of EsaI shows that a hydrophobic cavity in the protein likely encapsulates the acyl moiety of the acyl-acyl carrier proteins. The extreme preference of EsaI for a six-carbon acyl side chain is due to the size of the binding pocket. In addition, preference for a 3-oxo AHL results from a favorable hydrogen bond between the C3 carbonyl of the AHL and the hydroxyl group of a threonine residue on EsaI. Modeling studies suggest that the corresponding pocket in LasI of *P. aeruginosa* is larger, and consistent with this, LasI produces a longer autoinducer, 3-oxododecanoyl-homoserine lactone. This study suggests that the length and derivatization of the acyl side chains of AHLs are determined by differences in the structures of the binding cavities of the LuxI-type proteins.

Specificity in LuxR–AHL interaction is critical for bacteria to distinguish AHLs produced by their own species from AHLs of other species. The structural basis for ligand–receptor interaction has been most thoroughly examined with the TraR protein of *A. tumefaciens* and its cognate AHL autoinducer, 3-oxooctanoyl-homoserine lactone (27). The x-ray structure of

TraR in complex with its cognate AHL and DNA shows a precise interaction between TraR and the acyl moiety on the AHL. In this case, the C3 keto group of the AHL is stabilized by hydrogen bonding to a water molecule present in an autoinducer-binding cavity. Presumably, alterations in the size and shape of the AHL-binding pocket in other LuxR proteins will correspond to side chain lengths and substitutions of their cognate ligands.

The specificity of the AHL-LuxR pair does not extend to DNA binding and transcriptional activation, because the LuxR proteins all bind to similar DNA regulatory elements termed “lux boxes” (28, 29). Although interchangeability between AHLs and their LuxR partners is not tolerated, together the pairs can control gene expression at noncognate lux boxes. For example, the AHL molecule produced by *V. fischeri* (3-oxohexanoyl-homoserine lactone) together with LuxR, expressed in *Escherichia coli*, can induce transcription of *lasB*, a gene normally regulated by LasR and 3-oxododecanoyl-homoserine lactone in *P. aeruginosa*. Similarly, 3-oxododecanoyl-homoserine lactone, coupled with LasR and expressed in *E. coli*, will activate the *V. fischeri* quorum-sensing controlled target *luxCDABE* (29). Again, these findings support the hypothesis that target specificity inherent in LuxR circuits stems exclusively from the selection of a particular AHL by its cognate LuxR-type protein.

### Specificity in AIP Communication Systems

As in LuxIR circuits, similarities exist between oligopeptide autoinducers and cognate two-component sensors from different groups or species of Gram-positive bacteria, yet each signal-response system is remarkably specific. AIP synthesis is inherently accurate, because precise signal generation is guaranteed by the specific DNA sequence encoding the precursor protein. All AIPs are cleaved from longer precursor peptides, and many AIPs are also subject to posttranslational modifications (Fig. 2B). Signaling accuracy is achieved through the highly sensitive nature of the AIP receptors to alterations in AIP structure (16).

AIP signaling specificity has been analyzed extensively in *S. aureus* (30–32). Several clinically important groups within the *S. aureus* species have been characterized, and they all have slightly different AIPs, processing enzymes, and receptors (33, 34). The AIPs are 8 to 9 aa long and contain thiolactone rings involving invariant cysteine residues situated 5 aa from the C terminus (Fig. 2B) (35). Specificity in AIP signaling, at least in the staphylococci, is entirely determined by the AIP–sensor kinase interaction. This hypothesis is supported by findings that the cognate response regulators in all of the *S. aureus* groups are completely conserved, as are the target promoter regions (30, 33). Interestingly, the AIP produced by one group of *S. aureus* not only activates its own virulence cascade but also inhibits virulence in all other groups. Inhibition of virulence requires far less specificity in AIP structure, because the thiolactone ring portion common to all of the *S. aureus* AIPs acts as a universal inhibitor of AIP signaling (30–32). The group-specific AIP signaling in *S. aureus* indicates that quorum sensing in *S. aureus* occurs at the subspecies level. Inhibition by one *S. aureus* group of cell–cell communication in other *S. aureus* groups is presumed to benefit the group that first establishes its quorum-sensing cascade, because it facilitates the formation of a single-group infection.

### Quorum Sensing Using Multiple Autoinducers and Sensors

Quorum-sensing systems consisting of single autoinducer-sensor pairs are sufficient to control gene expression in response to changes in cell density. However, many bacteria have two or more quorum-sensing systems. Use of multiple quorum-sensing systems allows bacteria to integrate pieces of sensory information, which presumably confers plasticity to the genetic network. The hierarchies can be set up in series or in parallel, the former enabling regulation of genes in a temporally defined manner, and

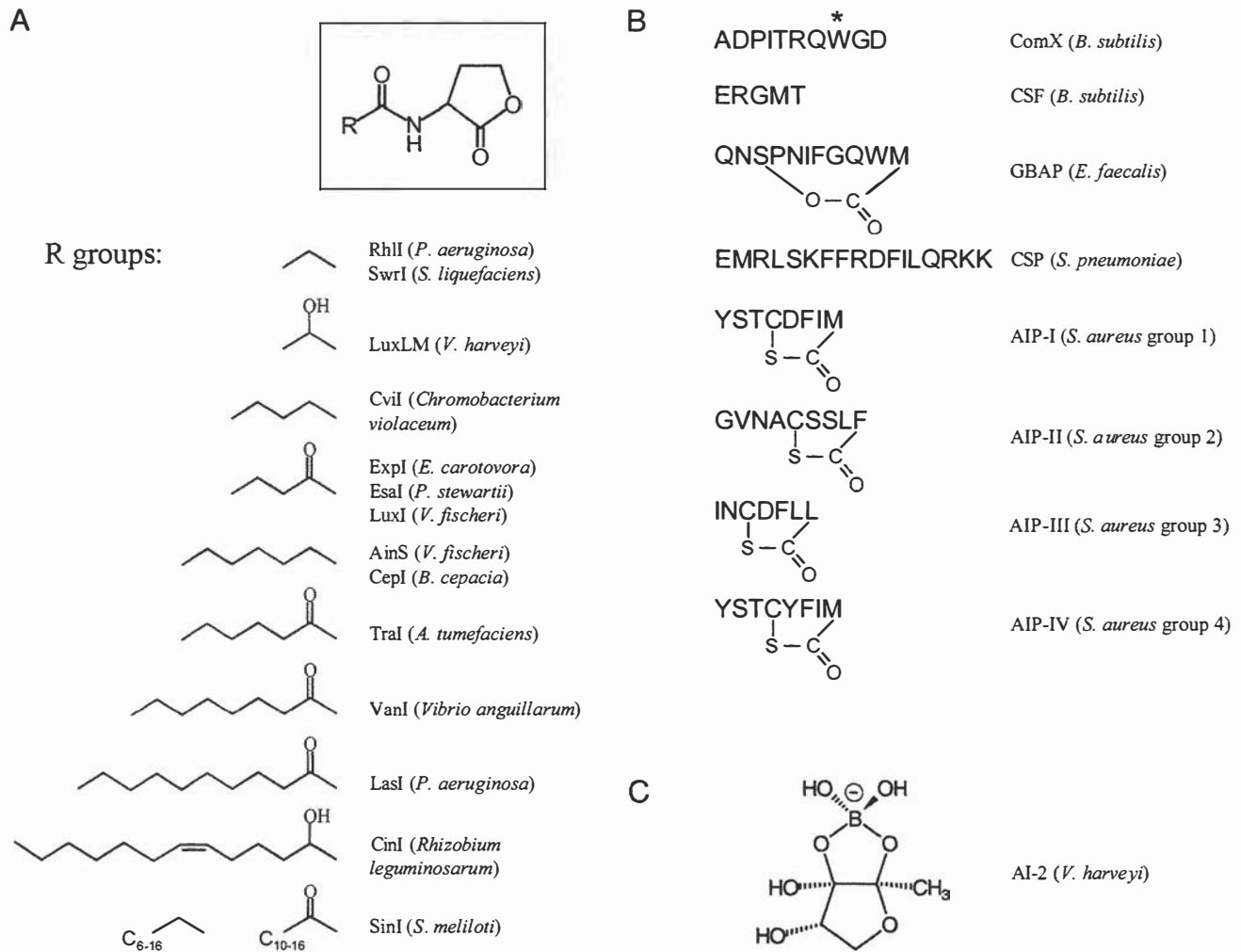


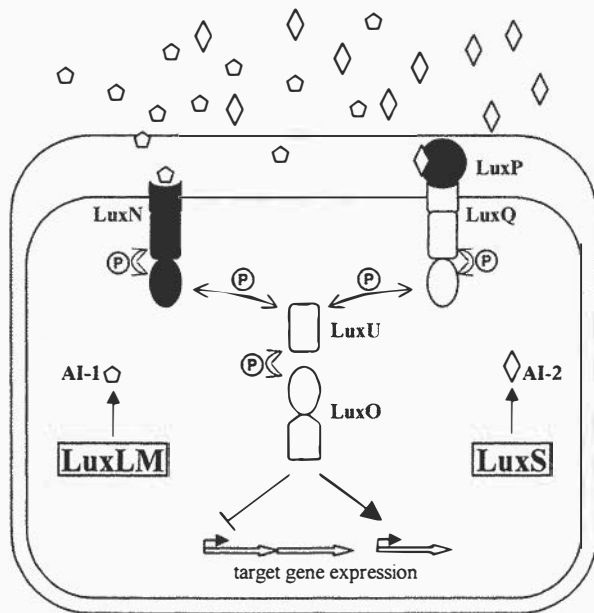
Fig. 2. Structures of different autoinducers. (A Upper) AHL autoinducers share a common homoserine lactone moiety. (Lower) Side chains of some different AHLs are shown (R groups). The synthases and the organisms that produce them are listed. (B) AIP peptide sequences, their designations, and the organisms that produce them are shown. The asterisk above the tryptophan residue in ComX represents an isoprenyl group. (C) AI-2 of *V. harveyi*.

the latter enabling regulation of discrete groups of genes or converging to regulate an identical set of genes.

In *P. aeruginosa*, two AHL quorum-sensing systems, LasIR and RhlIR, act in series. At high cell density, the concentration of both AHLs is high, and LasR binds its specific AHL to activate the expression of particular target genes. One of the genes activated by the AHL–LasR complex is *rhlR*, which encodes a second AHL receptor, RhlR. RhlR, in turn, binds its cognate AHL autoinducer and induces expression of its own target genes. Thus, genes controlled by the LasIR system are expressed before those controlled by the RhlIR system. This temporal pattern of gene regulation allows *P. aeruginosa* to express different virulence factors at various stages in the infection process (36, 37).

The two *P. aeruginosa* AHL signals promote virulence in a variety of different hosts, the most important human example being the cystic fibrosis (CF) lung. Not only is the CF lung susceptible to damage caused by AHL-controlled virulence factors, but it is also vulnerable to direct effects of the AHL molecules themselves. Specifically, the *P. aeruginosa* autoinducers induce the production of the chemokine IL-8 in the CF lung. As a result, neutrophils are recruited to the lung, and this action facilitates the formation of a potent *P. aeruginosa* infection. Thus, in this system, bacterial AHLs direct the expression of bacterial and host factors that enhance the infection process (38).

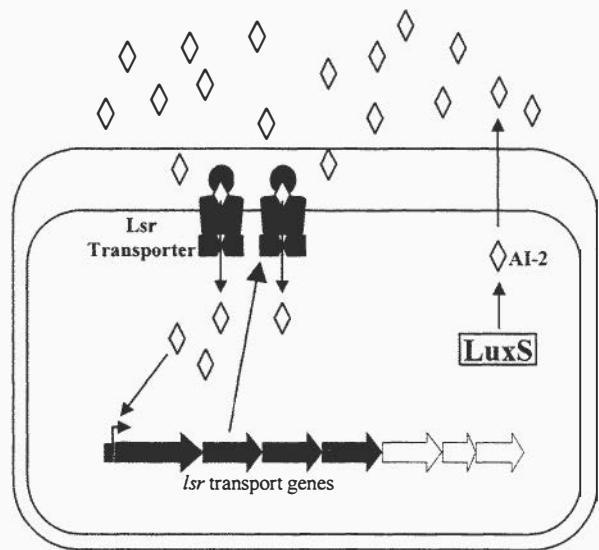
In two *Vibrio* species, *Vibrio harveyi* and *Vibrio cholerae*, multiple quorum-sensing systems converge to regulate a single group of genes. *V. harveyi* uses two parallel systems to regulate the expression of target genes, including those required for bioluminescence, and *V. cholerae* has three systems that jointly control the virulence regulon (1; 39). In *V. harveyi* (Fig. 3), the two autoinducer synthases, LuxLM and LuxS, each catalyze the synthesis of a specific autoinducer: the AHL 3-hydroxybutanoyl-homoserine lactone (denoted AI-1) in the case of LuxLM, and a unique furanosyl borate diester, 3a-methyl-5,6-dihydrofuro-[2,3-*d*][1,3,2]dioxaborole-2,2,6,6a-tetraol (AI-2), in the case of LuxS (Fig. 2 A and C) (40–42). Unlike other Gram-negative quorum-sensing systems in which AHL autoinducers are detected by a cytoplasmic LuxR-type protein, detection of the two autoinducers in *V. harveyi* occurs in the periplasm via cognate two-component sensor kinase proteins. AI-1 initiates signal transduction in the sensor kinase protein LuxN (43), and AI-2 binds the periplasmic binding protein LuxP, which in turn initiates signaling from the sensor kinase protein LuxQ (44). Information from both LuxN and LuxPQ converges at the phosphorelay protein LuxU (45, 46), and LuxU transmits the phosphorylation signal to the response regulator LuxO. LuxO controls transcription of target genes including those encoding luciferase (*luxCDABE*) (47, 48).



**Fig. 3.** The *V. harveyi* quorum-sensing circuit. Two autoinducers, the AHL AI-1 (pentagons) and AI-2 (diamonds), are produced by their cognate synthases LuxLM and LuxS, respectively. AI-1 and AI-2 bind cognate sensors (LuxN and LuxPQ, respectively), initiating a phosphorylation cascade (P) that travels through LuxU and alters the phosphorylation state of the response regulator protein LuxO. LuxO controls the expression of target genes. The proteins responsible for autoinducer binding are shaded in gray.

It seems paradoxical that *V. harveyi* uses two quorum-sensing systems to regulate the same set of target genes, because either system alone should be sufficient. One possible explanation for this molecular setup is that the autoinducers play different roles in cell–cell communication. Support for this idea stems from the finding that AI-1, like other AHL autoinducers, is species-specific, whereas many diverse species of bacteria possess a conserved LuxS homologue and produce AI-2, suggesting that AI-2 may function in interspecies cell–cell communication. Additionally, unlike AHL and AIP signaling, which is restricted to Gram-negatives and -positives, respectively, LuxS and AI-2 exist in both Gram-negative and -positive bacteria, suggesting that AI-2-mediated communication arose before AHL and AIP signaling (2, 42, 49).

The *V. harveyi* quorum-sensing circuit can distinguish among the presence of none, one, or both autoinducers. However, maximal expression of Lux occurs only when both autoinducers are present, suggesting that the circuit could act as a coincidence detector for both autoinducers (50). Because the *V. harveyi* quorum-sensing regulon has not yet been fully defined, it remains possible that additional classes of target genes exist that are regulated exclusively by either AI-1 or AI-2. If so, *V. harveyi* could differentially modify its behavior depending on whether it makes up the majority or the minority of a given mixed-species population. If, on the other hand, the *V. harveyi* quorum-sensing circuit functions primarily as a coincidence detector that regulates gene expression in response to the presence of both autoinducers, the advantage of this scheme could be that the simultaneous detection of two signals reduces the vulnerability of the circuit to noise or to “trickery.” A system that works by a combinatorial scheme could protect the cell–cell communication circuit from molecules made by other organisms in the environment that are similar in structure to the autoinducers (50).



**Fig. 4.** Lsr-mediated transport of AI-2 in *S. typhimurium*. AI-2 activates transcription of the *lsr* genes, four of which (shaded in gray) encode the Lsr transporter apparatus that functions to internalize AI-2.

### Interspecies Cell–Cell Communication: LuxS and AI-2

AI-2 is the only species-nonspecific autoinducer known. Because of its widespread occurrence, AI-2 is proposed to act as a universal quorum-sensing signal for interaction between species of bacteria (42, 49). Besides controlling light production in *V. harveyi*, the various roles that AI-2 plays in other species are beginning to be defined. Among other things, AI-2 controls the expression of genes required for virulence in *E. coli*, *V. cholerae*, *Clostridium perfringens*, and *Streptococcus pyogenes*; iron acquisition in *Porphyromonas gingivalis* and *Actinobacillus actinomycetemcomitans*; antibiotic production in *Photobacterium luminescens*; motility in *Campylobacter jejuni*; and mixed-species biofilm formation between *P. gingivalis* and *Streptococcus gordonii* (51). In *Salmonella typhimurium*, AI-2 was recently shown to control a seven-gene operon, called the *lsr* operon (for LuxS Regulated) (Fig. 4) (52). Four of the *lsr* operon genes encode an ABC transporter whose function is to promote internalization of AI-2. No additional AI-2-regulated genes have been identified in *S. typhimurium*, suggesting that AI-2 may have a role that is different from a classic quorum-sensing autoinducer in some bacteria, including *S. typhimurium*.

In *S. typhimurium*, and presumably in other bacteria that possess the *lsr* operon, AI-2 is produced in the cytoplasm and is released into the extracellular environment, where it accumulates. AI-2 is subsequently transported back inside the cell through the Lsr transporter (Fig. 4) (52). The benefit that bacteria gain from this cyclic process is unknown, although many possibilities exist. For example, AI-2 could be a quorum-sensing signal that impacts as-yet-unidentified genes in *S. typhimurium*. Other quorum-sensing regulated genes may exist that were not identified in the analysis revealing the *lsr* genes due to redundancy in the quorum-sensing circuitry or a requirement for a specific growth condition that was not met in the laboratory. If so, import of AI-2 could be used to terminate the quorum-sensing cascade or to convey AI-2 to an internal detection apparatus. An alternative hypothesis is that internalization of AI-2 could be a mechanism that bacteria such as *S. typhimurium* have evolved to interfere with the AI-2 quorum-sensing systems of competing species. This action could serve to confound other bacteria that regulate specific functions using AI-2, thereby giving *S. typhimurium* an advantage in particular environments.

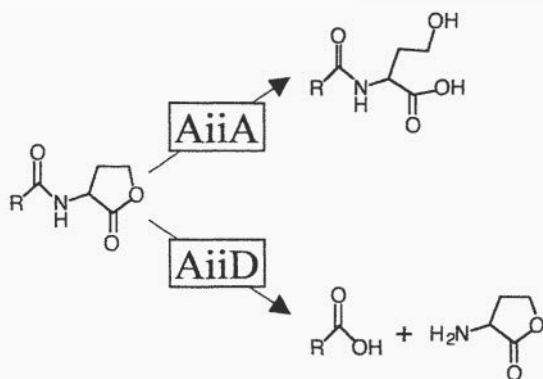


Fig. 5. Enzymatic inactivation of AHLs. (Upper) AiiA of *Bacillus* species hydrolyzes the lactone ring of AHLs. (Lower) AiiD of a *Ralstonia* strain hydrolyzes AHLs to release the acyl side chain moiety and homoserine lactone.

### Interference with Cell-Cell Communication: Biological Battles and Conspiracies

Quorum sensing often promotes behaviors that are detrimental to other organisms in the vicinity. For example, pathogens such as *P. aeruginosa* and *E. carotovora* use AHL-mediated quorum sensing to activate their virulence genes in specific animal and plant hosts, respectively (12). Likewise, in *Serratia liquefaciens*, antibiotic production and swarming motility, two behaviors required for virulence, are controlled by AHL-mediated quorum sensing (53). To protect themselves, some susceptible organisms have developed defenses that interfere with quorum sensing. Several anti-AHL-mediated quorum-sensing strategies have recently been discovered, and we present a few examples here. Additionally, in mixed-species environments, different species of bacteria can “conspire” with one another by responding to heterologous AHLs.

The macroalga *Delisea pulchra* produces a collection of halogenated furanones, many of which inhibit growth of pathogenic microorganisms on its surface. The furanones have structural similarity to AHLs and can prevent the AHL signal of *S. liquefaciens* from interacting with its cognate LuxR-type protein. In this model system, inhibition of AHL signaling prevents AHL-mediated motility in *S. liquefaciens*, which precludes bacterial colonization of the alga (54–57). Similarly, several organisms produce enzymes that modify AHLs, rendering them inactive. Marine algae such as *Laminaria digitata* prevent the formation of bacterial biofilms on their surfaces by producing haloperoxidases that generate oxidized halogens with microbicidal activity, such as hypochlorous acid (HOCl). Oxidized halogens react specifically with C3-oxo-AHLs and destroy their signaling capability. This finding is particularly significant because of the ability of the oxidized halogens to penetrate into biofilms (58).

In addition to eukaryotes, some bacteria produce enzymes that interfere with AHL signaling, presumably to gain an advantage over AHL-producing bacteria in specific biological niches. For example, an AHL lactonase (AiiA) isolated from *Bacillus* species has been shown to hydrolyze the lactone ring of AHLs to form acyl-homoserine, which does not function as a quorum-sensing signal (Fig. 5 Upper). Expression of AiiA in either the quorum-sensing pathogen *E. carotovora* or its susceptible plant host eliminates AHL production and infection, suggesting that AiiA can be used to prevent infection by pathogens that use quorum sensing to control virulence (59–62). What *in vivo* role AiiA plays in *Bacillus* species is not known, but it could be used to interfere with AHL-mediated quorum-sensing behaviors of soil bacteria that compete with *Bacillus* for the same niche. A different AHL inactivating enzyme, AiiD, was recently discovered in a *Ralstonia* strain isolated from a bacterial biofilm. Like AiiA, AiiD hydrolyzes AHLs, but in this case

the side chain is released from the intact homoserine lactone ring (Fig. 5 Lower). Again, the final result is elimination of quorum sensing. Introduction of *aiiD* into *P. aeruginosa* caused reduced accumulation of extracellular AHLs and inhibited AHL regulated behaviors such as swarming motility, virulence factor production, and paralysis of *Caenorhabditis elegans* (63). In another study, the bacterium *Variovorax paradoxus* was shown to be capable of using AHLs as a sole carbon and nitrogen source, suggesting that it too has AHL degrading activity (64). It appears that *V. paradoxus* has evolved the ability to degrade AHLs either as a defense against other bacteria that use AHLs as signaling molecules or as a metabolic scavenging method for using molecules released by neighboring bacteria. These recent discoveries suggest that the inhibition of AHL-mediated quorum sensing might be a widespread mechanism that bacteria use to compete with one another.

Instead of interspecies inhibition of AHL signaling, some bacteria use AHLs for interspecies communication during infection. For example, in cystic fibrosis lungs colonized by both *P. aeruginosa* and *Burkholderia cepacia*, in addition to producing and responding to its own AHL, *B. cepacia* responds to AHLs produced by *P. aeruginosa*, which promotes formation of a mixed-species biofilm (65). Because most bacteria reside in mixed-species environments, other examples of bacteria “conspiring” to carry out different behaviors could be discovered in examinations of quorum sensing in mixed populations.

Few cases of interference with AIP-mediated quorum-sensing systems are currently known, with the best-studied example being the cross-group inhibition of AIP signaling in *S. aureus*, discussed above. However, other systems for interfering with AIP-mediated quorum-sensing systems may exist. Although the autoinducer AI-2 is used for interspecies communication, no cases of interference with AI-2 quorum sensing are yet documented. However, as noted above, internalization of AI-2 by enterics such as *S. typhimurium* is considered one such possible anti-AI-2 communication mechanism.

### Conclusion

Recent research shows that chemical communication among bacteria is widespread and involves complex interconnected regulatory networks that serve to fine-tune the expression of diverse group behaviors. Specificity in autoinducer production and recognition is a key component of quorum sensing, because bacteria must preserve the fidelity of their communication circuits while existing in communities containing other organisms that produce molecules similar to their own autoinducers, either as quorum-sensing autoinducers or as autoinducer antagonists. The use of quorum sensing to control behaviors such as biofilm formation, symbiosis, and virulence factor expression indicates that quorum sensing is frequently used to regulate traits involving interactions between different organisms, both mutualistic and antagonistic. Thus, the ability of bacteria to distinguish self from other could also be a fundamental property of quorum-sensing systems. In support of this idea, it appears that some bacteria have evolved species-specific as well as “generic” signaling molecules. The recent discovery of autoinducer degrading enzymes demonstrates that interference with quorum sensing is an effective antibacterial strategy used in the wild. Synthetic anti-quorum-sensing strategies could be developed in the future as possible alternatives to antibiotics, because blocking cell-cell communication within or among bacterial species could prevent pathogenicity. Likewise, biotechnological approaches that promote beneficial quorum-sensing behaviors could be exploited for the production of industrial-scale natural products in bacteria.

This work was supported by Office of Naval Research Grant N00014-99-1-0767, National Science Foundation Grant MCB-0094447, and National Institute of General Medical Sciences Grant GM65859.

1. Miller, M. B. & Bassler, B. L. (2001) *Annu. Rev. Microbiol.* **55**, 165–199.
2. Bassler, B. L. (1999) *Curr. Opin. Microbiol.* **2**, 582–587.
3. Fuqua, C., Winans, S. C. & Greenberg, E. P. (1996) *Annu. Rev. Microbiol.* **50**, 727–751.
4. Engebrecht, J., Nealon, K. & Silverman, M. (1983) *Cell* **32**, 773–781.
5. Engebrecht, J. & Silverman, M. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4154–4158.
6. Engebrecht, J. & Silverman, M. (1987) *Nucleic Acids Res.* **15**, 10455–10467.
7. Piper, K. R., Beck von Bodman, S. & Farrand, S. K. (1993) *Nature* **362**, 448–450.
8. Bainton, N. J., Stead, P., Chhabra, S. R., Bycroft, B. W., Salmond, G. P., Stewart, G. S. & Williams, P. (1992) *Biochem. J.* **288**, 997–1004.
9. Davies, D. G., Parsek, M. R., Pearson, J. P., Iglewski, B. H., Costerton, J. W. & Greenberg, E. P. (1998) *Science* **280**, 295–298.
10. Passador, L., Cook, J. M., Gambello, M. J., Rust, L. & Iglewski, B. H. (1993) *Science* **260**, 1127–1130.
11. Marketon, M. M., Gronquist, M. R., Eberhard, A. & Gonzalez, J. E. (2002) *J. Bacteriol.* **184**, 5686–5695.
12. de Kievit, T. R. & Iglewski, B. H. (2000) *Infect. Immun.* **68**, 4839–4849.
13. Fuqua, C., Parsek, M. R. & Greenberg, E. P. (2001) *Annu. Rev. Genet.* **35**, 439–468.
14. Parsek, M. R. & Greenberg, E. P. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8789–8793.
15. Lazazzera, B. A. & Grossman, A. D. (1998) *Trends Microbiol.* **6**, 288–294.
16. Kleerebezem, M., Quadri, L. E., Kuipers, O. P. & de Vos, W. M. (1997) *Mol. Microbiol.* **24**, 895–904.
17. Sturme, M. H., Kleerebezem, M., Nakayama, J., Akkermans, A. D., Vaughn, E. E. & de Vos, W. M. (2002) *Antonie Leeuwenhoek* **81**, 233–243.
18. Magnuson, R., Solomon, J. & Grossman, A. D. (1994) *Cell* **77**, 207–216.
19. Solomon, J. M., Magnuson, R., Srivastava, A. & Grossman, A. D. (1995) *Genes Dev.* **9**, 547–558.
20. Haverstein, L. S. & Morrison, D. A. (1999) in *Cell-Cell Signaling in Bacteria*, eds. Dunney, G. M. & Winans, S. C. (Am. Soc. Microbiol., Washington, DC), pp. 9–26.
21. Novick, R. P. (1999) in *Cell-Cell Signaling in Bacteria*, eds. Dunney, G. M. & Winans, S. C. (Am. Soc. Microbiol., Washington, DC), pp. 129–146.
22. Qin, X., Singh, K. V., Weinstock, G. M. & Murray, B. E. (2000) *Infect. Immun.* **68**, 2579–2586.
23. Hanzelka, B. L. & Greenberg, E. P. (1996) *J. Bacteriol.* **178**, 5291–5294.
24. More, M. I., Finger, L. D., Stryker, J. L., Fuqua, C., Eberhard, A. & Winans, S. C. (1996) *Science* **272**, 1655–1658.
25. Val, D. L. & Cronan, J. E., Jr. (1998) *J. Bacteriol.* **180**, 2644–2651.
26. Watson, W. T., Minogue, T. D., Val, D. L., von Bodman, S. B. & Churchill, M. E. (2002) *Mol. Cell* **9**, 685–694.
27. Zhang, R. G., Pappas, T., Brace, J. L., Miller, P. C., Oulmassov, T., Molyneaux, J. M., Anderson, J. C., Bashkin, J. K., Winans, S. C. & Joachimiak, A. (2002) *Nature* **417**, 971–974.
28. Devine, J. H., Shadel, G. S. & Baldwin, T. O. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5688–5692.
29. Gray, K. M., Passador, L., Iglewski, B. H. & Greenberg, E. P. (1994) *J. Bacteriol.* **176**, 3076–3080.
30. Lyon, G. J., Mayville, P., Muir, T. W. & Novick, R. P. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 13330–13335.
31. Lyon, G. J., Wright, J. S., Christopoulos, A., Novick, R. P. & Muir, T. W. (2002) *J. Biol. Chem.* **277**, 6247–6253.
32. Lyon, G. J., Wright, J. S., Muir, T. W. & Novick, R. P. (2002) *Biochemistry* **41**, 10095–10104.
33. Ji, G., Beavis, R. & Novick, R. P. (1997) *Science* **276**, 2027–2030.
34. Jarraud, S., Lyon, G. J., Figueiredo, A. M., Gerard, L., Vandenesch, F., Etienne, J., Muir, T. W. & Novick, R. P. (2000) *J. Bacteriol.* **182**, 6517–6522.
35. Mayville, P., Ji, G., Beavis, R., Yang, H., Goger, M., Novick, R. P. & Muir, T. W. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1218–1223.
36. Brint, J. M. & Ohman, D. E. (1995) *J. Bacteriol.* **177**, 7155–7163.
37. Pesci, E. C. & Iglewski, B. H. (1997) *Trends Microbiol.* **5**, 132–134; discussion, 134–135.
38. Smith, R. S., Fedyk, E. R., Springer, T. A., Mukaida, N., Iglewski, B. H. & Phipps, R. P. (2001) *J. Immunol.* **167**, 366–374.
39. Miller, M. B., Skorupski, K., Lenz, D. H., Taylor, R. K. & Bassler, B. L. (2002) *Cell* **110**, 303–314.
40. Cao, J. G. & Meighen, E. A. (1989) *J. Biol. Chem.* **264**, 21670–21676.
41. Chen, X., Schauder, S., Potier, N., Van Dorsselaer, A., Pelczar, I., Bassler, B. L. & Hughson, F. M. (2002) *Nature* **415**, 545–549.
42. Surette, M. G., Miller, M. B. & Bassler, B. L. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1639–1644.
43. Bassler, B. L., Wright, M., Showalter, R. E. & Silverman, M. R. (1993) *Mol. Microbiol.* **9**, 773–786.
44. Bassler, B. L., Wright, M. & Silverman, M. R. (1994) *Mol. Microbiol.* **13**, 273–286.
45. Freeman, J. A. & Bassler, B. L. (1999) *J. Bacteriol.* **181**, 899–906.
46. Freeman, J. A., Lilley, B. N. & Bassler, B. L. (2000) *Mol. Microbiol.* **35**, 139–149.
47. Freeman, J. A. & Bassler, B. L. (1999) *Mol. Microbiol.* **31**, 665–677.
48. Lilley, B. N. & Bassler, B. L. (2000) *Mol. Microbiol.* **36**, 940–954.
49. Bassler, B. L., Greenberg, E. P. & Stevens, A. M. (1997) *J. Bacteriol.* **179**, 4043–4045.
50. Mok, K. C., Wingreen, N. S. & Bassler, B. L. (2003) *EMBO J.* **22**, 870–881.
51. Xavier, K. B. & Bassler, B. L. (2003) *Curr. Opin. Microbiol.* **6**, 191–197.
52. Taga, M. E., Semmelhack, J. L. & Bassler, B. L. (2001) *Mol. Microbiol.* **42**, 777–793.
53. Eberl, L., Winson, M. K., Sternberg, C., Stewart, G. S., Christiansen, G., Chhabra, S. R., Bycroft, B., Williams, P., Molin, S. & Givskov, M. (1996) *Mol. Microbiol.* **20**, 127–136.
54. Givskov, M., de Nys, R., Manefield, M., Gram, L., Maximilien, R., Eberl, L., Molin, S., Steinberg, P. D. & Kjelleberg, S. (1996) *J. Bacteriol.* **178**, 6618–6622.
55. Givskov, M., Eberl, L. & Molin, S. (1997) *FEMS Microbiol. Lett.* **148**, 115–122.
56. Manefield, M., de Nys, R., Kumar, N., Read, R., Givskov, M., Steinberg, P. & Kjelleberg, S. (1999) *Microbiology* **145**, 283–291.
57. Rasmussen, T. B., Manefield, M., Andersen, J. B., Eberl, L., Anthoni, U., Christophersen, C., Steinberg, P., Kjelleberg, S. & Givskov, M. (2000) *Microbiology* **146**, 3237–3244.
58. Borchardt, S. A., Allain, E. J., Michels, J. J., Stearns, G. W., Kelly, R. F. & McCoy, W. F. (2001) *Appl. Environ. Microbiol.* **67**, 3174–3179.
59. Dong, Y. H., Gusti, A. R., Zhang, Q., Xu, J. L. & Zhang, L. H. (2002) *Appl. Environ. Microbiol.* **68**, 1754–1759.
60. Dong, Y. H., Wang, L. H., Xu, J. L., Zhang, H. B., Zhang, X. F. & Zhang, L. H. (2001) *Nature* **411**, 813–817.
61. Dong, Y. H., Xu, J. L., Li, X. Z. & Zhang, L. H. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3526–3531.
62. Lee, S. J., Park, S. Y., Lee, J. J., Yum, D. Y., Koo, B. T. & Lee, J. K. (2002) *Appl. Environ. Microbiol.* **68**, 3919–3924.
63. Lin, Y. H., Xu, J. L., Hu, J., Wang, L. H., Ong, S. L., Leadbetter, J. R. & Zhang, L. H. (2003) *Mol. Microbiol.* **47**, 849–860.
64. Leadbetter, J. R. & Greenberg, E. P. (2000) *J. Bacteriol.* **182**, 6921–6926.
65. Riedel, K., Hentzer, M., Geisenberger, O., Huber, B., Steidle, A., Wu, H., Hoiby, N., Givskov, M., Molin, S. & Eberl, L. (2001) *Microbiology* **147**, 3249–3262.

# Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species

Gregory L. Challis\*<sup>†</sup> and David A. Hopwood<sup>‡</sup>

\*Department of Chemistry, University of Warwick, Coventry CV4 7AL, United Kingdom; and <sup>‡</sup>John Innes Centre, Norwich Research Park, Colney Norwich NR4 7UH, United Kingdom

**In this article we briefly review theories about the ecological roles of microbial secondary metabolites and discuss the prevalence of multiple secondary metabolite production by strains of *Streptomyces*, highlighting results from analysis of the recently sequenced *Streptomyces coelicolor* and *Streptomyces avermitilis* genomes. We address this question: Why is multiple secondary metabolite production in *Streptomyces* species so commonplace? We argue that synergy or contingency in the action of individual metabolites against biological competitors may, in some cases, be a powerful driving force for the evolution of multiple secondary metabolite production. This argument is illustrated with examples of the coproduction of synergistically acting antibiotics and contingently acting siderophores: two well-known classes of secondary metabolite. We focus, in particular, on the coproduction of  $\beta$ -lactam antibiotics and  $\beta$ -lactamase inhibitors, the coproduction of type A and type B streptogramins, and the coregulated production and independent uptake of structurally distinct siderophores by species of *Streptomyces*. Possible mechanisms for the evolution of multiple synergistic and contingent metabolite production in *Streptomyces* species are discussed. It is concluded that the production by *Streptomyces* species of two or more secondary metabolites that act synergistically or contingently against biological competitors may be far more common than has previously been recognized, and that synergy and contingency may be common driving forces for the evolution of multiple secondary metabolite production by these sessile saprophytes.**

Since the discovery of actinomycin in Selman Waksman's laboratory at Rutgers University in 1940, followed in 1943 by streptomycin, the first really effective drug to treat tuberculosis, the actinomycetes have been famous as producers of antibiotics and other "secondary metabolites" with biological activity. During the Golden Age of antibiotic discovery, in the 1950s and 1960s, such well-known antibacterial drugs as tetracycline, erythromycin, and kanamycin, antifungal agents like candicidin and nystatin, and anticancer drugs such as adriamycin were discovered through the efforts of academic and industrial researchers. After 1970 the rate of discovery of useful compounds declined progressively, although several important agents nevertheless came to light, including the antihelminthic avermectin, the immunosuppressants rapamycin and tacrolimus (FK506), and the natural herbicide bialaphos. In fact, the total number of known biologically active molecules continued to grow steadily after the end of the Golden Age. By the mid to late 1990s, thousands of antibiotics and compounds with other biological activities had been described. Estimates of the numbers vary. For example, Demain and Fang (1) gave the total number of antibiotics produced by bacteria and fungi as 5,000, whereas Berdy (2) had more than twice that number. Nevertheless there is agreement that the actinomycetes are responsible for more than two-thirds of the total. Within the actinomycetes, members of the genus *Streptomyces* account for 70–80% of secondary metabolites, with smaller contributions from genera such as *Saccharopolyspora*,

*Amycolatopsis*, *Micromonospora*, and *Actinoplanes*. What are secondary metabolites, what is their evolutionary significance, and why should the actinomycetes, those soil-dwelling, sporulating members of the high G + C branch of the Gram-positive bacteria, be such prolific producers of them?

There is no pithy one-line definition of the term secondary metabolite, but it nevertheless remains an indispensable epithet in discussions about microbial (and plant) biochemistry and ecology. It embraces the ideas that such compounds are characteristic of narrow taxonomic groups of organisms, such as strains within species, and have diverse, unusual, and often complex chemical structures. They are nonessential for growth of the producing organism, at least under the conditions studied, and are indeed typically made after the phase of most active vegetative growth when the producer is entering a dormant or reproductive stage (1). Their range of biological activities is wide, including the inhibition or killing of other microorganisms (the narrow definition of an antibiotic), but also toxic effects against multicellular organisms like invertebrates and plants. Then there are hormone-like roles in microbial differentiation, and roles in metal transport, a function that blurs the distinction between primary and secondary metabolism. More problematic are the many compounds that either have (so far) no demonstrated biological activity or an activity that is hard to relate to any competitive advantage to the producer, such as a specific effect on the vertebrate immune system exerted by a compound made by a saprophytic soil inhabitant.

The last category was in particular responsible for a widespread view that secondary metabolites were either neutral in evolutionary terms or significant merely as waste products or to keep metabolism ticking over. The implausibility of such general explanations was admirably discussed by Williams *et al.* (3), who emphasized two powerful arguments for the adaptive significance of secondary metabolites: the complex genetic determination of their biosynthesis, and the exquisite adaptation of many classes of compounds (six were described in detail) to interact with their targets. The former point has been further reinforced by innumerable genetic studies of the biosynthesis of natural products since 1989. Take, for example, erythromycin (4). The producer, *Saccharopolyspora erythraea*, devotes some 60 kb of its DNA to making this macrolide from propionate units by an amazing assembly-line process involving no fewer than 28 active sites arranged along three giant proteins,

---

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: PKS, polyketide synthase; NRPS, nonribosomal peptide synthetase; Ider, iron-dependent repressor.

<sup>†</sup>To whom correspondence should be addressed. E-mail: G.L.Challis@warwick.ac.uk.

© 2003 by The National Academy of Sciences of the USA



**Table 1. Gene clusters potentially directing the production of secondary metabolites in *S. coelicolor***

Biosynthetic system	Metabolite	Size, kb	Location
Type II PKS	Actinorhodin	22	5071–5092
Type II PKS	Gray spore pigment	8	5314–5320
Mixed	Methylenomycin	20	SCP1 plasmid
NRPS; type I modular PKS	Prodiginines	33	5877–5898
NRPS	CDA	80	3210–3249
NRPS	Coelichelin	20	0489–0499
NRPS	Coelibactin	26	7681–7691
NRPS	Unknown	14	6429–6438
Type I modular PKS	Unknown	70	6273–6288
Type I modular PKS	Unknown	10	6826–6827
Type I iterative PKS	Polyunsaturated fatty acid?	19	0124–0129
Chalcone synthase	Tetrahydroxynaphthalene	1	1206–1208
Chalcone synthase	Unknown	3.5	7669–7671
Chalcone synthase	Unknown	1	7222
Sesquiterpene synthase	Geosmin	2	6073
Sesquiterpene synthase	Unknown	2.5	5222–5223
Squalene-Hopene cyclase	Hopanoids	15	6759–6771
Phytoene synthase	Isorenieratine		0185–0191
Siderophore synthetase	Desferrioxamines	5	2782–2785
Siderophore synthetase	Unknown	4	5799–5801
Type II fatty acid synthase	Unknown	10	1265–1273
Butyrolactone synthase	Butyrolactones?	1	6266
Deoxysugar	Unknown	20	0381–0401

followed by hydroxylation and glycosylation involving 18 further proteins, and then has to protect its ribosomes from the highly specific toxicity of the antibiotic by an equally specific methylation of a site on the ribosomal RNA. This is not merely a means of dealing with any excess propionate the cell may produce, nor an idling of the metabolic machinery!

Firn and Jones (5) have made an important recent contribution to the debate. They point out that, whereas some secondary metabolites are indeed highly potent at the concentrations produced in nature and, in the case of antibiotics, against target organisms that actually interact with the producer in the wild, many have only moderate activity. They proposed a unifying model in which it is acknowledged that high-affinity, reversible, noncovalent interactions between a ligand and a protein only occur when the ligand has exactly the right molecular configuration to interact with the complex 3D structure of the protein (what they call “biomolecular activity”) and that this is a rare property. On the other hand, many more molecules have biological activity when tested against a whole organism containing thousands of potential targets that might be inhibited relatively inefficiently. There will therefore be a selective advantage in the evolution of traits that optimize the production and retention of chemical diversity at minimal cost, to allow for the gradual emergence of true biomolecular activity. As long as we accept a selective advantage for the ability to produce secondary metabolites, then, it is legitimate to ask why the streptomycetes and their relatives are preeminent in this ability. The answer almost certainly reflects both the habitat of the organisms and their lifestyle.

The classical habitat of *Streptomyces* species is as free-living saprophytes in terrestrial soils. Although this is doubtless correct, there is now abundant evidence that some species colonize the rhizosphere of plant roots and even plant tissues; in some cases antibiotic production by the streptomycete may protect the host plant against potential pathogens; the symbiont in turn acquires nutrients from the plant (6, 7). There is now good evidence also for the growth of actinomycetes in marine soils (8).

The soil is a proverbially complex environment in which innumerable stresses (chemical, physical, and biological) occur in a

temporally and spatially variable manner. Moreover, streptomycetes are nonmotile, so stresses cannot be avoided but have to be met. The need to combat stress was the explanation invoked by Bentley *et al.* (9) for the enormous numbers of genes that would encode regulators, transport proteins, and nutritional enzymes identified in the *Streptomyces coelicolor* genome sequence. A striking feature of this 8.7-megabase (Mb) genome is its notional division into a “core” region of  $\approx 4.9$  Mb and left and right “arms” of  $\approx 1.5$  and 2.3 Mb, respectively. Classes of genes that would encode unconditionally essential functions such as the machinery of DNA replication, transcription, and translation were found in the core, whereas examples of conditionally adaptive functions, such as the ability to grow on complex carbohydrates like cellulose, chitin, and xylan, occurred predominately in the arms. The genome sequence also revealed  $\approx 23$  clusters of genes, representing  $\approx 4.5\%$  of the genome, that were predicted to encode biosynthetic enzymes for a wide range of secondary metabolites (Table 1), only half a dozen of which had previously been identified. Interestingly in the present context, many of the clusters reside in the arms or close to their boundary with the core of the genome, as pointed out by Piepersberg (10) in a review that emphasizes the roles of secondary metabolites in chemical communication, the topic of this Sackler symposium. An even larger number of secondary metabolic gene clusters was found in the recently sequenced *Streptomyces avermitilis* genome (30 clusters covering  $\approx 6\%$  of the genome), again many of them in the arm regions (11). Thus, assuming a selective advantage for secondary metabolite production, such an advantage for many of the compounds is likely to be conditional, or sporadic.

As pointed out by Chater and Merrick (12), *Streptomyces* antibiotics are typically produced in small amounts at the transition phase in colonial development when the growth of the vegetative mycelium is slowing as a result of nutrient exhaustion and the aerial mycelium is about to develop at the expense of nutrients released by breakdown of the vegetative hyphae (13). Such antibiotics are proposed to defend the food source when other soil microorganisms threaten it. The hypothesis is strengthened by the finding of large numbers of antibiotics in other groups of differentiating microbes such as filamentous fungi and myxobacteria. The concept

has been extended to the possibility that competitors might be attracted to the amino acids, sugars, and other small molecules arising from the degraded vegetative mycelium and be killed and recycled by the developing *Streptomyces* colony, a concept dubbed “fatal attraction” in relation to *Myxococcus* by Shi and Zusman (14).

### Synergy and Contingency<sup>5</sup> in Secondary Metabolite Action

Against this background of likely selective advantages, then, actinomycetes have evolved some amazingly potent agents with biomolecular activity and numerous others with a more generalized biological effect. Perhaps more strikingly, there are now numerous examples of the production by an actinomycete of two chemically different metabolites that act either synergistically against a target microorganism, as recently emphasized by McCafferty and coworkers (15), or contingently to overcome competition with other microorganisms for nutrients. In this context, synergistic metabolites have a greater antibiotic activity against competitors in combination than the sum of their individual antibiotic activities (see below), whereas contingently acting metabolites possess similar biological activity (e.g., iron sequestration as below), but are independently recognized and used by the producing organism and competitors in its environment. In the rest of this article we discuss some examples of these phenomena and attempt to relate them to the ideas about the roles and evolution of secondary metabolism outlined above.

**$\beta$ -Lactam Antibiotics and Clavulanic Acid.** Clavulanic acid (1) (Fig. 1) is a natural inhibitor of  $\beta$ -lactamases (enzymes that confer resistance to  $\beta$ -lactam antibiotics in many microorganisms). The genetics and biochemistry of its production have been intensively studied in *Streptomyces clavuligerus* for >25 years (16). This actinomycete also produces several other  $\beta$ -lactam compounds, including a number of structurally related clavams (2) (Fig. 1), which are antipodal to clavulanic acid and are not  $\beta$ -lactamase inhibitors (although they do exhibit antifungal activity), and the cephalosporin antibiotic cephamycin C (3) (Fig. 1) (along with other biologically active intermediates on its biosynthetic pathway), which inhibits the transpeptidation reaction in cell wall biosynthesis. The early steps of clavulanic acid and clavam biosynthesis are common, but the pathways to these two structurally related metabolites diverge at clavaminic acid (16). In contrast, cephamycin is biosynthesized by a pathway that is mechanistically distinct from that for the clavams and clavulanic acid (16, 17). Combinations of  $\beta$ -lactam antibiotics and  $\beta$ -lactamase inhibitors are well known to be effective against  $\beta$ -lactam-resistant bacteria in comparison with  $\beta$ -lactams antibiotics alone. This is because of the synergistic action of these metabolites, which is reflected by the name chosen for the clinically used combination of clavulanic acid with methicillin: augmentin.

At first sight the coproduction of cephamycin C and clavulanic acid by *S. clavuligerus* might seem like an unusual coincidence. Closer inspection of metabolite production patterns among other producers of clavulanic acid, clavams, and cephamycin C, however, suggests that a strong selective pressure, rather than mere chance, has created actinomycetes that coproduce clavulanic acid and a  $\beta$ -lactam antibiotic such as cephamycin C. Thus, *S. clavuligerus*, *Streptomyces jumonjinensis*, *Streptomyces katsurahamanus*, and an unclassified *Streptomyces* sp. all produce clavulanic acid (16). Strikingly, all of these streptomycetes also produce cephamycin C (16). Indeed there are no known producers of clavulanic acid that do not also produce cephamycins (16). In contrast, several *Streptomyces* species produce clavams, but not clavulanic acid or cephamycins (16), and some actinomycetes, such as *Nocardia lactamdurans* and

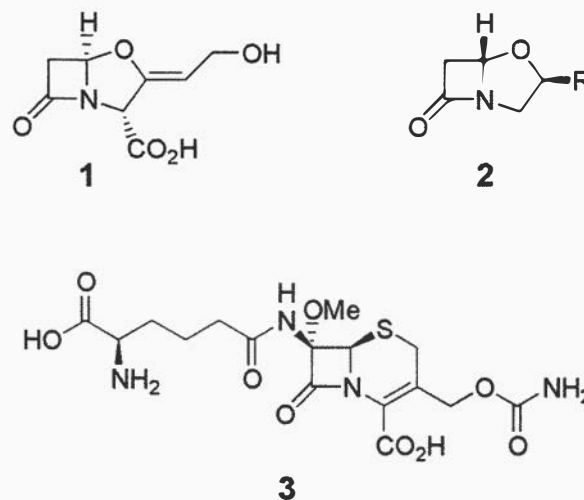


Fig. 1. Structures of clavulanic acid (1), clavams (R = variable group) (2), and cephamycin C (3), coproduced by several *Streptomyces* species. 1 and 3 act synergistically to inhibit cell wall biosynthesis in  $\beta$ -lactam-resistant bacteria.

*Streptomyces griseus* NRRL 3851, produce cephamycin C but not clavams or clavulanic acid (17). The fact that no known actinomycetes produce clavulanic acid alone, but there are actinomycetes that produce just cephamycin C or clavams, suggests that the production of clavulanic acid evolved in an ancestral clavam and cephamycin producer as a response to the acquisition of  $\beta$ -lactamase-mediated resistance in bacteria inhabiting the same environmental niche and thus posing biological competition. The existence of several clavam-only-producing strains suggests that one of these may initially have acquired the cephamycin pathway by horizontal transfer, which conferred a selective advantage against  $\beta$ -lactam-sensitive bacteria. These sensitive bacteria then acquired  $\beta$ -lactamase-mediated resistance to cephamycin, and eventually the production of clavulanic acid by a modification of the clavam biosynthetic pathway in the clavam/cephamycin producer was selected for by biological competition from the  $\beta$ -lactam-resistant bacteria. The production of clavulanic acid by the cephamycin/clavam producer would restore the effectiveness of cephamycin as an antibiotic against the  $\beta$ -lactam-resistant competition.

Support for the above hypothesis for the evolution of clavulanic acid production derives from analysis of the genes that direct clavam and clavulanic acid production. Thus, the clavulanic acid gene cluster is directly adjacent to the cephamycin cluster on the chromosomes of *S. clavuligerus*, *S. jumonjinensis*, and *S. katsurahamanus* (18). Such “superclustering” would be expected for gene clusters that direct the production of metabolites that act synergistically to benefit the producing organism (19, 20). In addition, the production of both cephamycin and clavulanic acid in *S. clavuligerus* is controlled principally by the *ccaR* (*dclX*) gene, which codes for an OmpR-like transcriptional regulator, and is located within the cephamycin cluster (21, 22). In contrast, the clavam cluster is at least 20–30 kb away from the cephamycin/clavulanic acid clusters on the *S. clavuligerus* chromosome, and the regulation of clavam production is distinct from the coregulated production of cephamycins and clavulanic acid (23–25). These observations suggest that clavulanic acid production in a clavam-producing ancestor, which acquired the ability to produce cephamycin, could have arisen via chromosomal duplication of the clavam cluster followed by subsequent acquisition of the late steps in the clavulanic acid pathway, which involve the stereochemical inversions that are key to  $\beta$ -lactamase activity. The chromosomal linkage of the cephamycin and the clavulanic acid clusters would facilitate simultaneous horizontal transfer of both clusters to other organisms, which would clearly be beneficial to recipients. Also, there would be pressure to evolve coregulation of

<sup>5</sup>Note that our use of “contingency” in this article relates to multiple metabolites acting on the same biological target to provide an organism with a contingency plan to combat unforeseeable biological competition. Moxon and coworkers (50) have used contingency to describe hypermutable loci coding for variable surface proteins in *Haemophilus influenzae* and *Neseria meningitidis*. The two uses of the word should not be confused.

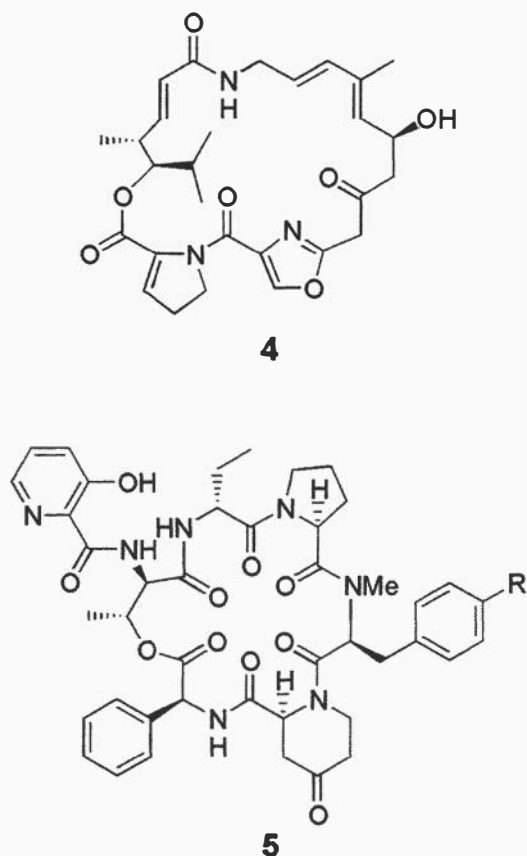


Fig. 2. Structures of pristinamycin I (4) and II (R = variable group) (5) components, which inhibit protein synthesis in bacteria by binding synergistically to the ribosome.

clavulanic acid and cephamycin production, as the coordinated production of a  $\beta$ -lactam antibiotic and a  $\beta$ -lactamase inhibitor is clearly beneficial for survival. Thus, there appears to be clear selective pressure and a plausible genetic mechanism for the evolution of the coproduction of clavulanic acid and cephamycin in *Streptomyces* species.

**Streptogramins.** Streptogramins are pairs of structurally unrelated antibiotics that synergistically inhibit bacterial ribosomal protein synthesis at the peptidyl transfer step (15). The type A streptogramins [e.g., pristinamycin IIA/virginiamycin M<sub>1</sub> (4) (Fig. 2)] are assembled through a mixed polyketide/nonribosomal peptide pathway, whereas the type B streptogramins [e.g., pristinamycin IA (5) (Fig. 2) (R = NMe<sub>2</sub>) and virginiamycin S<sub>1</sub> (6) (R = H)] are nonribosomally synthesized depsipeptides that contain several unusual nonproteinogenic amino acids (26–33). 4 is coproduced with 5 by *Streptomyces pristinaespiralis* NRRL 2958 and with 6 by *Streptomyces virginiae* (34). Several other streptomycetes, such as *Streptomyces mitakaensis*, *Streptomyces graminofaciens*, and *Streptomyces loidensis*, have been reported to coproduce similar type A and type B streptogramins. No species are known that produce only a type A or a type B streptogramin. Whereas type A or type B streptogramins alone are bacteriostatic, the combination of a type A and type B streptogramin is bacteriocidal. This increase in antibacterial efficacy has been shown to originate from synergistic binding of the type A and type B streptogramins to distinct sites on the ribosome (35, 36). Thus, binding of streptogramin A antibiotics to the ribosome increases the affinity of streptogramin B antibiotics for the ribosome by up to 40-fold, rendering dual antibiotic binding essentially irreversible (37). The binding of the type A antibiotics is

thought to change the conformation of the ribosomal RNA to expose a high-affinity binding site for the type B compounds (37). The synergistic antibiotic effect of the streptogramins has been exploited clinically in the form of a combination of semisynthetic pristinamycin IA and IIA derivatives known as synercid, which is used for the treatment of multidrug-resistant bacterial infections.

Evidence to support the idea that coproduction of streptogramin A and B antibiotics in several bacteria has evolved under strong selective pressure stems from analysis of the chromosomal arrangement of genes that direct streptogramin production. Thus, Bamas-Jacques *et al.* (38) demonstrated that genes directing the production of pristinamycins I and II are clustered in a 200-kb region of the chromosome of *S. pristinaespiralis*. The finding that genes for the biosynthesis of the pristinamycin I and II components are interspersed in this cluster led Bamas-Jacques *et al.* to suggest that the pathways to the type A and type B streptogramins coevolved in the same organism, perhaps from a common origin. The structural dissimilarity of the type A and type B streptogramins, however, implies that they are biosynthesized by quite different catalytic machinery. Indeed analysis of the genes that direct production of the pristinamycin type I and type II components indicates that they are biosynthesized by nonribosomal peptide and predominantly polyketide pathways, respectively (30–33). It is therefore hard to imagine how the pathways to the type I and type II metabolites could have evolved from a common origin. An alternative, stepwise model for the acquisition of the type A and type B streptogramin pathways cannot be ruled out on the basis of available evidence. Thus, an ancestral streptomycete may have originally produced only one of the streptogramin components, which was initially effective at inhibiting ribosomal protein synthesis in competing organisms. Over time the competing organisms would acquire resistance to the single streptogramin component, rendering it less effective. The pathway to the second streptogramin component could then have been acquired by chance horizontal transfer. The synergistic effect of the pair of streptogramins would give the recipient a renewed advantage over competing organisms, which would lead to retention of both pathways. Initially, the two pathways may have been located in distinct regions of the chromosomes. Subsequent chromosomal rearrangements (which are well known in *Streptomyces* species) could intersperse the two clusters with each other, which could facilitate the coregulated production of both streptogramin components and the horizontal transfer of both pathways into other organisms. The stepwise model for evolution of streptogramin production is similar to that outlined above for evolution of cephamycin and clavulanic acid production and might prove to be a useful general model for evolution of the production of multiple secondary metabolites that function synergistically.

**Other Potentially Synergistic, Coproduced Pairs or Groups of Antibiotics.** Streptomycetes are famed for producing multiple antibiotics, and it seems likely that there are many other examples of individual *Streptomyces* species that produce two or more antibiotics that act synergistically against a competing organism. One interesting potential example has arisen from sequencing of the *Streptomyces avermitilis* genome (11). Analysis of gene clusters that code for polyketide synthase (PKS) multienzyme systems, commonly associated with antibiotic biosynthesis, revealed that *S. avermitilis* has the capability to produce two structurally distinct antifungal compounds: oligomycin and a polyene macrolide (11). These antibiotics act on distinct molecular targets in eukaryotes. Thus, oligomycin inhibits mitochondrial F<sub>0</sub>F<sub>1</sub>-ATP synthase, whereas polyene macrolides bind irreversibly to fungal cell membranes, altering their permeability (39, 40). It seems possible that oligomycin and polyene macrolides could act synergistically against fungi and that a persistent fungal competitor in the natural environment of *S. avermitilis* has selected for the coproduction of these compounds. It is noteworthy, however, that the gene clusters directing oligomycin and polyene macrolide biosynthesis are separated by >2,500 kb on the

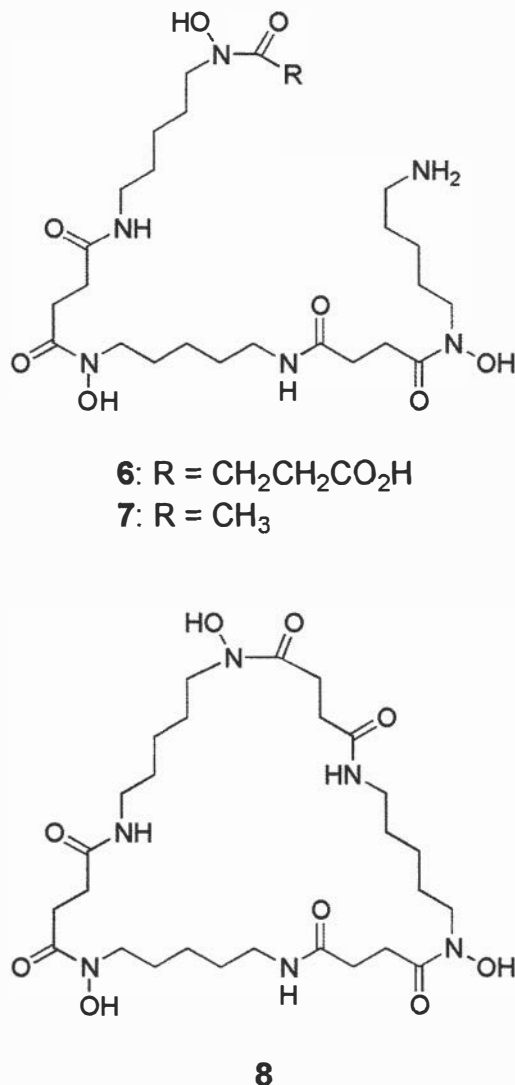


Fig. 3. Structures of desferrioxamine siderophores typically produced by *Streptomyces* species.

*S. avermitilis* chromosome, and it will be interesting to see whether the production of these two antifungal compounds is coregulated (41).

**Production of Multiple Siderophores.** Siderophores are small diffusible molecules excreted by many microorganisms that form very stable complexes with ferric iron. In Gram-positive bacteria, the iron-siderophore complexes are selectively recognized by membrane-associated receptors and actively transported into the cell by ATP-dependent transmembrane transporters (42). Once inside the cell, the iron-siderophore complex is dissociated, often by hydrolysis of the multidentate siderophore ligand and/or reduction of ferric iron to ferrous iron, which is stored in bacterioferritin and used as a cofactor in several vital cellular processes (43). Siderophores are produced by saprophytes to overcome the inherent aqueous insolubility of ferric iron, which limits its availability in soil. More than 10 distinct species of *Streptomyces* have been reported to produce characteristic desferrioxamine siderophores such as desferrioxamines G<sub>1</sub> (6), B (7), and E (8) (44) (Fig. 3). Until very recently, however, siderophores belonging to other structural classes such as peptide hydroxamates, catecholates,  $\alpha$ -hydroxycarboxylates, and thiazolines/oxazolines have not been reported as metabolic products of *Streptomyces* species.

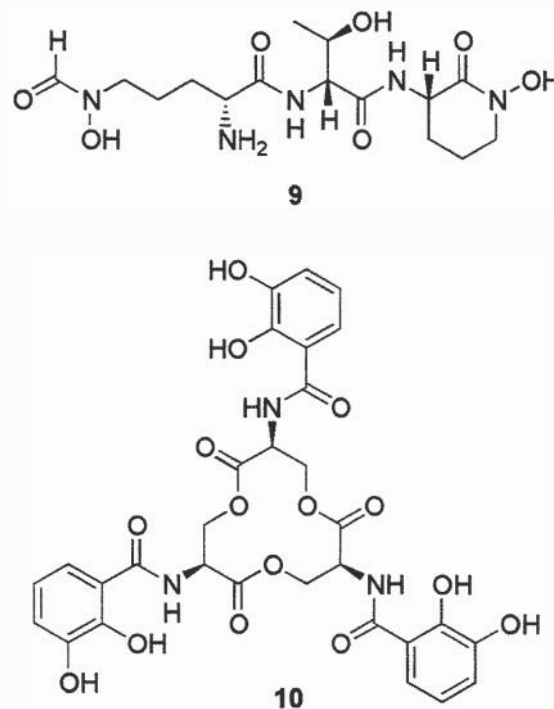


Fig. 4. Predicted structure of coelichelin (9) and structure of enterobactin (10), siderophores of diverse structure coproduced with desferrioxamines by some *Streptomyces* species that are thought to provide a contingency plan for iron uptake in the event of desferrioxamine piracy.

*In silico* analysis of the *S. coelicolor* genome sequence suggested that, in addition to a pathway for desferrioxamine biosynthesis, multiple nonribosomal peptide synthetase (NRPS) pathways for the assembly of structurally diverse siderophores exist (9). One of these pathways has been predicted to culminate in the peptide hydroxamate siderophore coelichelin (9) (Fig. 4) (45). Recently, Challis and coworkers have confirmed, using a combination of gene inactivation and metabolic profiling experiments, that two independent pathways exist in *S. coelicolor* for the production of hydroxamate siderophores (S. Lautru, F. Barona-Gomez, U. Wong, and G.L.C., unpublished work). Thus, inactivation of *desD*, which codes for a siderophore synthetase believed to catalyze the key step in desferrioxamine biosynthesis, causes abrogation of the production of 6 and 8, whereas inactivation of *cchH*, which codes for the NRPS predicted to assemble 9, leads to loss of production of a different hydroxamate siderophore (F. Barona-Gomez, U. Wong, A. Giannakopoulos, P. J. Derrick, and G.L.C., unpublished work).

The production of both the desferrioxamines and the other hydroxamate siderophore is maximal under iron-deficient conditions and completely suppressed under iron-sufficient conditions. Consistent with this observation, analysis of the *cch* and *des* clusters indicates that, although they are not closely linked on the *S. coelicolor* chromosome, their transcription is very likely to be coregulated, because intergenic regions in both clusters contain similar inverted repeat sequences that match the consensus sequence 5'-TTAGGTTAGGCTCACCTAA-3' for iron-dependent repressor (IdeR) binding. IdeR is known to regulate the transcription of siderophore biosynthesis genes in other Gram-positive bacteria, e.g., *Mycobacterium* species (46).

It is not immediately obvious what selective advantage *S. coelicolor* gains through the coregulated production of two structurally distinct hydroxamate siderophores. Yet this is not an isolated example of multiple siderophore production by *Streptomyces* species. Recently, Fiedler and coworkers (47) reported that *Streptomyces tendae* Tü 901/8c and *Streptomyces* sp. Tü 6125 produce

enterobactin 10 (the characteristic siderophore of *Enterobacteriaceae*; Fig. 4) in addition to the characteristic *Streptomyces* siderophores 7 and 8. It seems probable that enterobactin is biosynthesized in streptomycetes via a NRPS pathway, similar to the well-characterized *Escherichia coli* pathway, and that this pathway is distinct from that for desferrioxamine biosynthesis (48). A third potential example of multiple siderophore production by *Streptomyces* species has been uncovered through preliminary examination of the *S. avermitilis* genome sequence, which, in addition to a cluster of genes virtually identical to the *des* cluster of *S. coelicolor*, contains a gene cluster encoding a NRPS system predicted to produce a siderophore of similar structure to the mycobacterial siderophore myxochelin (11, 49). Thus, the production of structurally diverse secondary nonribosomally synthesized peptide siderophores, in addition to the characteristic desferrioxamine siderophores, may be common in streptomycetes.

An appealing explanation for the coregulated production of two or more structurally distinct siderophores by *Streptomyces* species stems from the observation that many organisms that neither biosynthesize nor excrete desferrioxamine-like siderophores are nevertheless able to specifically take up ferrioxamine complexes and use the iron associated with them. These organisms would pose a serious biological challenge to sessile streptomycetes that possess only a desferrioxamine pathway for scavenging iron from the environment. This would exert strong selective pressure on such streptomycetes for acquisition of a second cluster of genes directing the production of another siderophore (e.g., by horizontal transfer), whose ferric complex could be selectively recognized and taken up by the cell through a separate transport system from that for ferrioxamine uptake. Consistent with this model, analysis of the *des* and *cch* clusters of *S. coelicolor* reveals that a gene coding for a distinct iron-siderophore-binding lipoprotein is present in each cluster. Iron-siderophore-binding lipoproteins are receptors associated, via a covalently attached lipid, with the extracellular membrane of Gram-positive bacteria that selectively recognize iron-siderophore complexes and initiate their ATP-driven transport into the cell. Thus, the finding that iron-siderophore-binding lipoproteins are coded for in both the *des* and *cch* clusters suggests that the ferric complexes of each of these siderophores can be selectively and independently taken up by *S. coelicolor*. The biological competition

from other organisms able to take up ferrioxamines would be lessened by coregulated production of coelichelin and desferrioxamines, because the ferric coelichelin complex can be selectively absorbed into *S. coelicolor* cells through an independent uptake pathway. In this scenario, coelichelin and desferrioxamine are acting contingently rather than synergistically, because they allow *S. coelicolor* to survive in environments inhabited by unforeseen competitors whose ability to use ferrioxamine (or indeed other ferric siderophore complexes) cannot be readily anticipated.

The notion that a second cluster of genes directing siderophore production in primordial desferrioxamine-producing streptomycetes was acquired by horizontal transfer is supported by comparison of the structures of enterobactin, and that predicted for coelichelin (produced by *S. tendae* and *S. coelicolor*, respectively). Whereas enterobactin is the characteristic siderophore of *E. coli*, the predicted structure of coelichelin is similar to several hydroxamate siderophores produced by mycobacteria, suggesting that the "second" siderophores of streptomycetes derive from diverse origins. It will be interesting to see whether other, structurally diverse siderophores are isolated, along with the characteristic desferrioxamines, from further *Streptomyces* species in the future.

## Conclusions

The well known property of many *Streptomyces* species to produce multiple antibiotics or other secondary metabolites has attracted much recent attention, not least because analysis of the recently completed *S. coelicolor* and *S. avermitilis* genome sequences has suggested that this ability may be far greater than was previously thought. Clear evidence emerging from the recent literature suggests that two driving forces for the evolution of this phenomenon may be synergistic and contingent action against biological competitors, which *Streptomyces* species are not easily able to evade in their saprophytic lifestyle because of a lack of motility. Indeed, as the study of *Streptomyces* secondary metabolism continues, it is anticipated that many more cases of two or more structurally distinct metabolites that act synergistically or contingently against biological competition will be discovered.

Francisco Barona-Gomez and Sylvie Lautru are gratefully acknowledged for helpful discussions and suggestions on the manuscript.

- Demain, A. & Fang, A. (2000) in *History of Modern Biotechnology*, ed. Fichter, A. (Springer, Berlin), Vol. 1, pp. 2–39.
- Berdy, J. (1995) in *Proceedings of the Ninth International Symposium on the Biology of the Actinomycetes*, eds. Debabov, V. G., Dudnik, Y. V. & Danilenko, V. N. (All-Russia Research Institute for Genetics and Selection of Industrial Microorganisms, Moscow), pp. 13–34.
- Williams, D. H., Stone, M. J., Hauck, P. R. & Rahman, S. K. (1989) *J. Nat. Products* **52**, 1189–1208.
- Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. (1991) *Proc. Natl. Acad. Sci. USA* **252**, 675–679.
- Firm, R. D. & Jones, C. G. (2000) *Mol. Microbiol.* **37**, 989–994.
- Tokala, R. K., Strap, J. L., Jung, C. M., Crawford, D. L., Salove, M. H., Deobald, L. A., Bailey, J. F. & Morra, M. J. (2002) *Appl. Environ. Microbiol.* **68**, 2161–2171.
- Castillo, U. F., Strobel, G. A., Ford, E. J., Hess, W. M., Porter, H., Jensen, J. B., Albert, H., Robison, R., Condrón, M. A. M., Teplow, D. B., et al. (2002) *Microbiology* **148**, 2675–2685.
- Mincer, T. J., Jensen, P. R., Kauffman, C. A. & Fenical, W. (2002) *Appl. Environ. Microbiol.* **68**, 5005–5011.
- Bentley, S. D., Chater, K. F., Cerdano-Tarraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., et al. (2002) *Nature* **417**, 141–147.
- Piepersberg, W. (2002) in *Molecular Medical Microbiology*, ed. Sussman, M. (Academic, San Diego), Vol. 1, pp. 561–594.
- Ikedo, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003) *Nat. Biotechnol.* **21**, 526–531.
- Chater, K. F. & Merrick, M. J. (1979) in *Developmental Biology of Prokaryotes*, ed. Parish, J. H. (Blackwell, Oxford), pp. 93–114.
- Miguez, E. M., Hardisson, C. & Manzanal, M. B. (2000) *Int. Microbiol.* **3**, 153–158.
- Shi, W. & Zusman, D. R. (1993) *Nature* **366**, 414–415.
- McCafferty, D. G., Cudic, P., Yu, Y. K., Behena, D. C. & Kruger, R. (1999) *Curr. Opin. Chem. Biol.* **3**, 672–680.
- Jensen, S. E. & Paradkar, A. S. (1999) *Antonie van Leeuwenhoek* **75**, 125–133.
- Liras, P. (1999) *Antonie van Leeuwenhoek* **75**, 109–124.
- Ward, J. M. & Hodgson, J. E. (1993) *FEMS Microbiol. Lett.* **110**, 239–242.
- Jensen, S. E., Alexander, D. C., Paradkar, A. S. & Aido, K. A. (1993) in *Industrial Microorganisms: Basic and Applied Molecular Genetics*, eds. Baltz, R. H., Hegeman, G. D. & Skatrud, P. L. (Am. Soc. Microbiol., Washington, DC), pp. 169–176.
- Lawrence, G. J. (1997) *Trends Microbiol.* **5**, 355–359.
- Walters, N. J., Barton, B. & Earl, A. J. (1994) International Patent WO94/18326-A1.
- Perez-Llarena, F., Liras, P., Rodriguez-Garcia, A. & Martin, J. F. (1997) *J. Bacteriol.* **179**, 2053–2059.
- Mosher, R. H., Paradkar, A. S., Anders, C., Barton, B. & Jensen, S. E. (1999) *Antimicrob. Agents Chemother.* **43**, 1215–1224.
- Paradkar, A. & Jensen, S. E. (1995) *J. Bacteriol.* **177**, 1307–1314.
- Marsh, E. N., Chang, M. D.-T. & Townsend, C. A. (1992) *Biochemistry* **31**, 12648–12657.
- Blanc, V., Blanche, F., Crouzet, J., Jacques, N., Lacroix, P., Thibaut, D. & Zagorec, M. (1994) International Patent WO 9408014.
- Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M., Schleuniger, J., Bisch, D., Blanche, F., Debussche, L., Crouzet, J., et al. (1997) *Mol. Microbiol.* **23**, 191–202.
- Blanc, V., Lagneaux, D., Didier, P., Gil, P., Lacroix, P. & Crouzet, J. (1995) *J. Bacteriol.* **177**, 5206–5214.
- Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J., Barriere, J.-C., Debussche, L., Famechon, A., Paris, J.-M., Dutruc-Rosset, G., et al. (1996) International Patent WO 9601901.

30. de Crécy-Lagard, V., Blanc, V., Gil, P., Naudin, L., Lorenzon, S., Famechohn, A., Bamas-Jacques, N., Crouzet, J., Thiebaut, D., *et al.* (1997) *J. Bacteriol.* **179**, 705–713.
31. de Crécy-Lagard, V., Saurin, W., Thibaut, D., Gil, P., Naudin, L., Crouzet, J. & Blanc, V. (1997) *Antimicrob. Agents Chemother.* **41**, 1904–2009.
32. Thibaut, D., Bisch, D., Ratet, N., Maton, L., Couder, M., Debussche, L. & Blanche, F. (1997) *J. Bacteriol.* **179**, 697–704.
33. Thibaut, D., Ratet, N., Bisch, D., Faucher, D., Debussche, L. & Blanche, F. (1995) *J. Bacteriol.* **177**, 5199–5205.
34. Cocito, C. G. (1979) *Microbiol. Rev.* **43**, 145–198.
35. Porse, B. T. & Garrett, R. A. (1999) *J. Mol. Biol.* **286**, 375–387.
36. Contreras, A. & Vasquez, D. (1977) *Eur. J. Biochem.* **74**, 549–551.
37. Di Giambattista, M., Chinali, G. & Cocito, C. (1989) *J. Antimicrob. Chemother.* **24**, 485–507.
38. Bamas-Jacques, N., Lorenzon, S., Lacroix, P., de Swetschin, C. & Crouzet, J. (1999) *J. Appl. Microbiol.* **87**, 939–948.
39. Zotchev, S. B. (2003) *Curr. Med. Chem.* **10**, 211–223.
40. Corran, A. J., Renwick, A. & Dunbar, S. J. (1998) *Pestic. Sci.* **54**, 338–344.
41. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98**, 12215–12220.
42. Schneider, R. & Hantke, K. (1993) *Mol. Microbiol.* **8**, 111–121.
43. Winkelman, G. & Drechsel, H. (1997) in *Biotechnology*, eds. Rehm, H.-J. & Reed, G. (VCH, Weinheim, Germany), 2nd Ed., Vol. 7, pp. 200–246.
44. Müller, A. & Zähler, H. (1968) *Arch. Mikrobiol.* **62**, 257–263.
45. Challis, G. L. & Ravel, J. (2001) *FEMS Microbiol. Lett.* **187**, 111–114.
46. Schmitt, M. P., Predich, M., Doukhan, L., Smith, I. & Holmes, R. K. (1995) *Infect. Immun.* **63**, 4284–4289.
47. Fiedler, H.-P., Krastel, P., Müller, J., Gebhardt, K. & Zeeck, A. (2001) *FEMS Microbiol. Lett.* **196**, 147–151.
48. Crosa, J. H. & Walsh, C. T. (2002) *Microbiol. Mol. Biol. Rev.* **66**, 223–249.
49. Kunze, B., Bedorf, N., Kohl, W., Hoefle, G. & Reichenbach, H. (1989) *J. Antibiot.* **42**, 14–17.
50. Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. (1994) *Curr. Biol.* **4**, 24–33.

# Efficient oxidative folding of conotoxins and the radiation of venomous cone snails

Grzegorz Bulaj<sup>\*†</sup>, Olga Buczek<sup>\*</sup>, Ian Goodsell<sup>\*</sup>, Elsie C. Jimenez<sup>\*\*</sup>, Jessica Kranski<sup>†</sup>, Jacob S. Nielsen<sup>†</sup>, James E. Garrett<sup>†</sup>, and Baldomero M. Olivera<sup>\*5</sup>

<sup>\*</sup>Department of Biology, University of Utah, Salt Lake City, UT 84112; <sup>†</sup>Cognetix, Inc., 421 Wakara Way, Salt Lake City, UT 84108; and <sup>\*\*</sup>Department of Physical Sciences, College of Science, University of the Philippines Baguio, Baguio City, Philippines

The 500 different species of venomous cone snails (genus *Conus*) use small, highly structured peptides (conotoxins) for interacting with prey, predators, and competitors. These peptides are produced by translating mRNA from many genes belonging to only a few gene superfamilies. Each translation product is processed to yield a great diversity of different mature toxin peptides ( $\approx 50,000$ – $100,000$ ), most of which are 12–30 aa in length with two to three disulfide crosslinks. *In vitro*, forming the biologically relevant disulfide configuration is often problematic, suggesting that *in vivo* mechanisms for efficiently folding the diversity of conotoxins have been evolved by the cone snails. We demonstrate here that the correct folding of a *Conus* peptide is facilitated by a posttranslationally modified amino acid,  $\gamma$ -carboxyglutamate. In addition, we show that multiple isoforms of protein disulfide isomerase are major soluble proteins in *Conus* venom duct extracts. The results provide evidence for the type of adaptations required before cone snails could systematically explore the specialized biochemical world of “microproteins” that other organisms have not been able to systematically access. Almost certainly, additional specialized adaptations for efficient microprotein folding are required.

Chemical interactions between organisms, the major theme of this symposium, are usually mediated through organic molecules that elicit physiological effects in the targeted organisms that are of benefit to the producer. In this respect, a group of predatory molluscs, the cone snails (see Fig. 1), use an idiosyncratic strategy; small structured peptides, instead of conventional natural products, are the primary agents used for interacting with other animals. The cone snails belong to the genus *Conus*, comprising 500 different species, all of which capture prey by injecting a peptide-rich venom (1). Some, if not most, species of *Conus* also use venom to defend against predators and for competitive interactions (2). Thus, the “front-line” genes for the biotic interactions of *Conus* are those encoding their venom peptides. The peptidic nature of the functional gene products means that these can be directly elucidated and chemically synthesized from the gene sequence. Consequently, for the cone snails, the link between chemical ecology and genomics is unusually straightforward.

Each *Conus* species has a distinct set of biotic interactions characteristic of that species; this helps to rationalize why each one has a different repertoire of 100–200 venom peptides (only a subset of which are probably expressed at any one time). Thus, the 500 different living species of cone snails (3) can potentially produce  $\approx 50,000$ – $100,000$  different conopeptides in their venom ducts, an enormous diversity of gene products reflecting the complex chemical ecology of the genus as a whole (for a review, see ref. 2). Rapid advances in DNA sequence analysis have made these chemical agents much more amenable to systematic interspecific analysis than any comparably diverse set of natural products produced by any genus or family of animals or plants.

The analysis carried out on *Conus* venom peptides suggests that a majority of the estimated  $>50,000$  peptides are encoded by only  $\approx 12$  conotoxin gene superfamilies. These superfamilies have undergone rapid amplification and divergence, accompanying the parallel radiation and diversification of *Conus* species at a macroevolutionary level (*Conus* is arguably the most species-rich genus of living marine invertebrates). Each major *Conus* peptide gene superfamily comprises thousands of genes, encoding different peptides. This leads to the remarkable functional diversity seen among the  $\approx 50,000$  different peptides. A majority of all *Conus* peptides exert a powerful effect on some specific ion channel or receptor target (4). It is fair to say that the snails likely have evolved a greater diversity of ion channel-targeted pharmacological agents than even the largest of pharmaceutical companies (this diverse array includes peptides that are being developed for use as human pharmaceuticals). These venom peptides have allowed different cone snail species to specialize on at least five different phyla of prey and defend themselves against a spectrum of predators that might be even more diverse.

Most protein genes initially produce a translation product that is  $>100$  amino acids in length, a size sufficient to allow conventional folding by multiple intramolecular interactions. Toxin proteins found in venoms are generally smaller (50–100 aa), with additional stability provided by disulfide bonds. In effect, the cone snails have extended this tendency one step further, with some venom peptide superfamilies being the smallest highly structured but functionally diverse classes of gene products known (12–20 aa with two to three disulfide bonds). Conopeptide evolution has resulted in a large diversity of biological function being generated in each conotoxin superfamily. Thus, *Conus* peptide superfamilies are like other major classes of proteins produced through gene translation: structural and functional novelty can evolve, and thus, conotoxins are in many respects “microproteins” and differ from more conventional unstructured peptides.

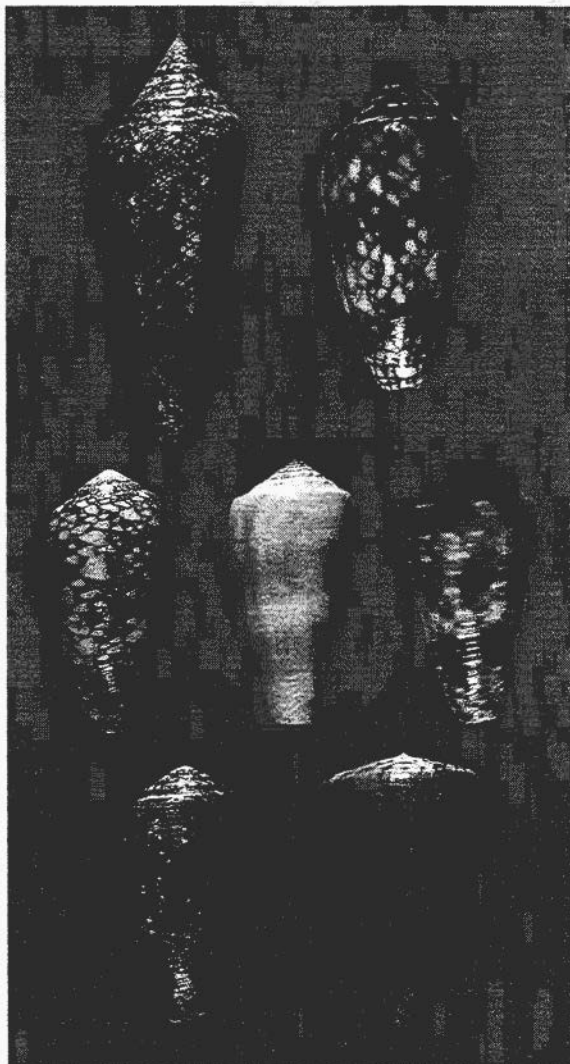
Each conotoxin superfamily has a characteristic arrangement of cysteine residues, which is assembled into a particular disulfide bonding configuration (the “disulfide framework”). The latter is the primary determinant of polypeptide backbone structure. Despite hypermutation of the amino acids between Cys residues, the disulfide framework generally remains conserved within a superfamily, generating a characteristic scaffold. In principle, for peptides with six Cys residues (characteristic of at least four conotoxin superfamilies) there are 15 different

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Chemical Communication in a Post-Genomic World,” held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: PDI, protein disulfide isomerase; ER, endoplasmic reticulum.

<sup>5</sup>To whom correspondence should be addressed. E-mail: olivera@biology.utah.edu.

© 2003 by The National Academy of Sciences of the USA



**Fig. 1.** Shells of cone snails. Seven different *Conus* species (of the  $\approx 500$  total) are illustrated: the examples shown represent the three major feeding types of *Conus*. Experimental data that are described in the text were obtained from these species. (Top) The glory-of-the-sea cone, *C. gloriamaris* (Left); the cloth-of-gold cone, *C. textile* (Right). (Middle) *C. omaria* (Left); *C. consors* (Center); *C. aurisiacus* (Right). (Bottom) The fly-speck cone, *C. stercusmuscarum* (Left); *C. betulinus* (Right). *C. betulinus* is worm-hunting, whereas *C. consors*, *C. aurisiacus*, and *C. stercusmuscarum* are piscivorous (fish hunting). *C. textile*, *C. gloriamaris*, and *C. omaria* are molluscivorous (mollusc hunting). A PCR screen of PDI was carried out on most of these species (see text). A comparison of the spasmodic peptide of *C. textile* and *C. gloriamaris* provided the basis for the work on the role of posttranslational modification in *Conus* peptide folding. *C. textile* was the species used for most of the experimental studies described in this article.

disulfide-bonded arrangements possible, with only one being biologically relevant. How does folding to this single structural framework (instead of the 14 other possibilities) occur? This is a specialized version of the more general protein folding problem, one of the central unsolved questions in biology.

Efficient solutions to this specialized folding problem must have been evolved by cone snails before they could effectively explore (in an evolutionary sense) the alternative world of microproteins that they have so efficiently exploited for their diverse physiological purposes. In this article, data are presented that are relevant to this previously unaddressed issue. There are two levels at which solutions to the specialized folding problem

of *Conus* peptides could have evolved: (i) specialized intramolecular interactions may stabilize conformation, and (ii) intermolecular interactions with extrinsic factors, perhaps within the endoplasmic reticulum (ER), may promote appropriate folding pathways within the ER. We explore both possibilities.

An unusual feature of *Conus* venom peptides is the high degree of posttranslational modification observed (5). We demonstrate that posttranslational modification can generate intramolecular interactions that facilitate conopeptide folding. For potential intermolecular interactions, we show that multiple isoforms of protein disulfide isomerase (PDI), which catalyze protein thiol disulfide exchange reactions, are present within *Conus* venom ducts, and we show that, apart from the conotoxins themselves, these isoforms are the major soluble protein components of the ducts.

## Materials and Methods

**Cloning of *Conus* PDI.** Full-length *Conus textile* PDI cDNAs were isolated by RT-PCR of venom duct RNA, using degenerate primers based on the amino acid sequence of the highly conserved thioredoxin-like active site motif found in PDI proteins isolated from other organisms. PCR products were gel-purified and cloned into a plasmid vector, and several cloned isolates were sequenced. The DNA sequence of the cloned PCR product contained a long ORF with significant homology to PDI proteins from other organisms. The DNA sequence of this internal PCR product was used to design nested PCR primers for 5' and 3' RACE procedures to isolate the full-length cDNA. *C. textile* venom duct cDNA was synthesized with 5' and 3' RACE adapters (Ambion, Austin, TX) and used for RACE amplifications. Specific 5' and 3' RACE products were gel-purified, cloned into a plasmid vector, and sequenced. The sequences of each of these RACE products overlapped with the previously isolated central portion of the *C. textile* PDI cDNA and together these three PCR-generated cDNAs could be merged to give the full-length cDNA sequence encoding the complete PDI protein. To detect PDI expression in venom ducts of other *Conus* species, RT-PCR was performed in the following species: *C. textile*, *Conus stercusmuscarum*, *Conus aurisiacus*, *Conus consors*, *Conus betulinus*, and *Conus omaria*. The PCR amplification was performed by using primers based on the thioredoxin active site sequence. Approximately 20 ng of venom duct cDNA and *Taq* polymerase was used for the PCRs.

**Protein Analysis of Venom Ducts.** The venom duct from *C. textile* was dissected and immediately divided into four equal parts. Each part of the venom duct was ground under liquid nitrogen. Extraction was performed in 1 ml of 10 mM Tris-HCl, pH 7.8 containing 0.25 M sucrose and 5 mM EDTA at 4°C. After homogenization, the solution was centrifuged. The 50  $\mu$ l of resulted supernatant was lyophilized and dissolved in 30  $\mu$ l of SDS-electrophoresis buffer, boiled for 5 min, and applied on 4–20% Tris-glycine gel. Proteins were then electroblotted onto Immobilon poly(vinylidene difluoride) membrane (0.45  $\mu$ m) (Millipore) for 1 h at 50 V. Proteins were visualized by using Coomassie blue staining, and the protein band of 55 kDa was cut out from the membrane. Amino acid sequencing was performed by the Edman degradation method. After protein separation, polyacrylamide gel was also stained with Coomassie blue.

**Peptide Synthesis and Folding.** Peptides were synthesized on solid support by using standard Fmoc [*N*-(9-fluorenyl)methoxycarbonyl] chemistry. All cysteines were protected with S-trityl groups. The peptides were removed from the resin and purified by using a semipreparative reversed-phase  $C_{18}$  HPLC column in a linear gradient of acetonitrile. After purification, the linear peptide was lyophilized. Folding reactions were initiated by injecting a resuspended linear peptide into a 200- $\mu$ l solution





Fig. 2. PDI from *C. textile* venom duct. (A) SDS/PAGE of extracts from different parts of a venom duct. Lanes 1–4 represent equal fragments from distal to proximal parts of the venom duct. Lane 5 is a bovine PDI. The ~55-kDa band from lane 3 was extracted, and its N-terminal sequence was determined (see text). (B) Structures of PDI-1 and PDI-2 from *C. textile*. The signal sequence, thio redoxin active sites (Trx), and the ER retention signal are marked. Domains a, b, b', a', and c are based on the assignments for human PDI (9, 10). The shaded residues represent differences between PDI-1 and PDI-2.

containing 0.1 M Tris-HCl (pH 8.7), 1 mM oxidized glutathione, and 2 mM reduced glutathione. The folding reactions contained CaCl<sub>2</sub> or MgCl<sub>2</sub> at concentrations indicated in the text or 1 mM EDTA. The final peptide concentration was 20 μM. After an appropriate folding time, the reactions were quenched by acidification with formic acid (10% final concentration). The samples were separated by reversed-phase C<sub>18</sub> analytical HPLC in the gradient of acetonitrile (from 9% to 31.5% acetonitrile in 0.1% trifluoroacetic acid). The accumulation of the native form was calculated relative to other folding species from integration of the HPLC peaks.

**Assays for Biological Activity of Synthetic Peptides.** The activities of synthetic peptides were evaluated by using 13- to 14-day-old mice, weighing 5.9–6.4 g; the intracranial injections were carried out as described (6).

## Results

**PDI Is the Major Soluble High Molecular Weight Polypeptidic Component of *Conus* Venom Ducts.** The oxidative folding of secreted proteins with disulfide bonds is thought to be promoted by PDIs (7, 8). These enzymes contain four thio redoxin-like domains, two of which have the sequence –CXXC–, that are believed to catalyze formation and isomerization of protein disulfide bonds. The unprecedented density of disulfide linkages in conotoxins (up to 50% of all amino acids can be Cys residues involved in disulfide bonding) led us to initiate the characterization of PDI from cone snail venom ducts. Protein extracts from *C. textile* venom ducts were prepared (see *Materials and Methods*), and the crude extracts were analyzed by SDS/PAGE. As shown in Fig. 2A, the expected low molecular mass gene products consistent with conotoxins are prominently stained on the gel, but in addition, a major protein band with a molecular mass of ~55 kDa stands out. This molecular mass is consistent with that of PDIs characterized from other organisms (see marker band in Fig. 2A).

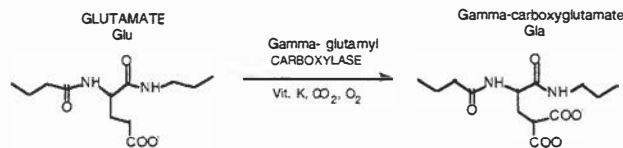
The identity of the major polypeptide in the 55-kDa band was investigated by microsequencing: a clear N-terminal sequence, –EEVEEQENVY–, was obtained; as will be discussed below, this sequence is consistent with the conclusion that the ~55-kDa band is PDI, possibly a mixture of different isoforms. Thus, apart from the conotoxins themselves, PDI appears to be the major soluble protein in *Conus* venom ducts. Such a high level of PDI along the whole length of the duct is consistent with conotoxin

synthesis in both distal and proximal portions of the duct and with a major role for PDI in the oxidative folding of conotoxins.

PDI expression was further evaluated in a variety of *Conus* species by using a PCR screen, with primers matching conserved sequences in PDIs (see *Materials and Methods*). The PCR fragment obtained from cDNA prepared from a variety of different *Conus* species representing the major feeding types of the genus, including *C. textile*, *C. stercusmuscarum*, *C. aurisiacus*, *C. consors*, *C. betulinus*, and *C. omaria* (see Fig. 1) was similar in size and intensity to that obtained from *C. textile*. To characterize *Conus* PDIs further, we analyzed a cDNA library from *C. textile*, the cloth-of-gold cone. The sequences of two full-length cDNA clones encoding two different PDI isoforms, PDI-1 and PDI-2, were determined (Fig. 2B). There is a high degree of sequence identity between the two isoforms, but there are two regions of high divergence: amino acids 4–19 (40% sequence divergence) and amino acids 71–96 (32% divergence). Both proteins are ~75% homologous to the well characterized human PDI sequence shown in Fig. 2.

The cleavage site for the removal of the signal sequence was identified from the microsequencing results for the mature protein detected in the SDS gel. Based on homology to the human PDI, five domains can be distinguished: a, b, b', a', and c (9, 10). Two of these (a, a') are domains that have sequence similarity to thio redoxin (a small protein cofactor in ribonucleotide reduction), whereas two (b, b') show little sequence similarity but have a thio redoxin-like fold. The two putative active-site –CGHC– sequences are located in domains a and a'; thus, the two regions of high divergence sandwich the –CGHC– site of the a domain. The C-terminal domain is characterized by a high degree of negatively charged Asp and Glu residues, postulated to serve as the Ca<sup>2+</sup> binding domain (11, 12). The C-terminal sequence RDEL– is a standard ER retention signal. Thus, the *Conus* PDI isoforms cloned contain all characteristic features of PDIs from other organisms. The results with the directly sequenced protein band in Fig. 2A suggest that there may be additional venom duct PDI isoforms. Preliminary molecular evidence for additional PDI isoforms has been independently obtained (I.G. and J.E.G., unpublished results). We are presently defining all *Conus* PDI isoforms and have expressed the two shown in Fig. 2B. Enzymatic properties of the expressed *Conus* PDIs need to be assessed to determine whether these may exhibit specificity for particular conotoxin substrates, relative to other Cys-rich polypeptides.

$\gamma$ -Carboxylation of Glutamate to  $\gamma$ -Carboxyglutamate (Gla or  $\gamma$ )



Peptide                      Source                      Sequence

"Gla-domain" region/peptides

h-FIX	Human	KLγγFVQGNLYRγCMγKCSFYAR	γVFYNTYKTTY FW
Conantokin-G (27)	<i>C. geographus</i>	GEY	γLQYNQYLIRY KSN#

Multiply disulfide-bonded, non-Gla domain *Conus* peptides

tx5a (28)	<i>C. textile</i>	γCCγDGWCCT' AAO
Bromosleeper (29)	<i>C. radiatus</i>	WATIDγCγYTTCNVTFKTC <del>C</del> CGOOGDWQC <del>V</del> YACP <del>V</del>
Spasmodic-Tx	<i>C. textile</i>	GCNNSCQYHSDCγSHCICTFRGCGAVN#
Spasmodic-Gm	<i>C. gloriamaris</i>	SCNNSCQSHSDCASHCICTFRGCGAVN#

(non- $\gamma$ -containing)

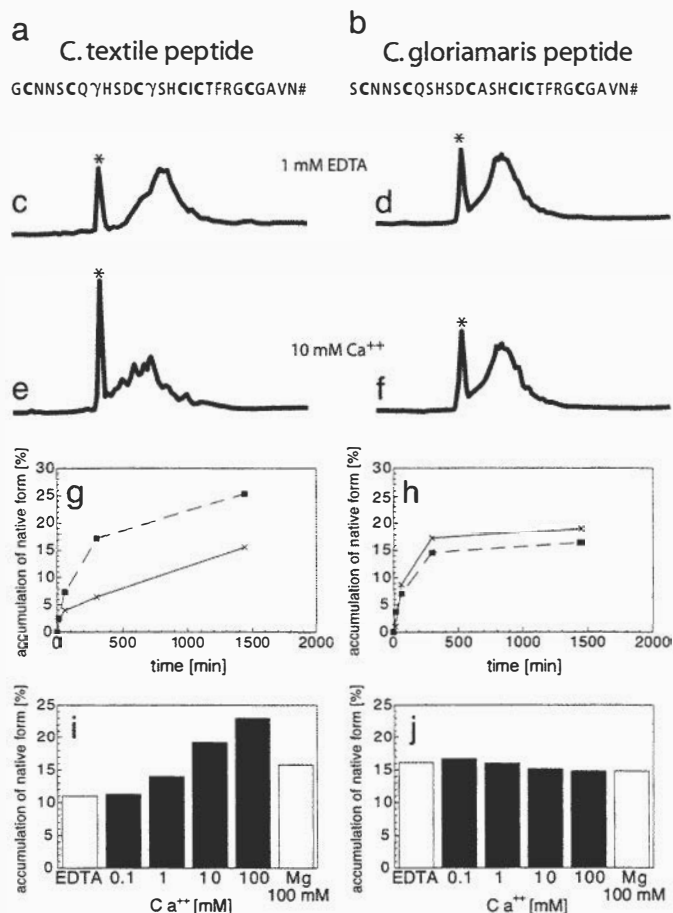
**Fig. 3.** Comparison of Gla-containing peptides from *Conus* species. The posttranslational modification of glutamate to  $\gamma$ -carboxyglutamate (Gla or  $\gamma$ ) is shown, and a comparison of the sequences of four Gla-containing *Conus* venom peptides to the N-terminal sequence of human blood clotting factor IX is shown. Note that one *Conus* peptide, conantokin-G from *C. geographus*, has a Gla domain motif that has similar spacing to factor IX (see boxed sequences), whereas the three other Gla-containing peptides do not have an arrangement of Gla residues characteristic of a Gla domain. #, amidated C terminus; W, 6-bromotryptophan; T', glycosylated threonine;  $\gamma$ ,  $\gamma$ -carboxyglutamate; h-FIX, human factor IX.

**A Role for  $\gamma$ -Carboxyglutamate in Conotoxin Folding.** A *Conus* peptide recently characterized from *C. textile*, the "spasmodic peptide" (6), has serendipitously presented an opportunity to evaluate the function of an unusual posttranslational modification found in *Conus* peptides, the conversion of glutamate to  $\gamma$ -carboxyglutamate (see Fig. 3). A closely homologous peptide identified by RT-PCR from a closely related species, the glory-of-the-sea cone, *Conus gloriamaris* (see Fig. 1) exhibited considerable sequence identity (24/27 amino acids identical) to the *C. textile* peptide (13). However, no  $\gamma$ -carboxyglutamate residues are present in the *C. gloriamaris* peptide. Thus, two very closely related natural peptides from different *Conus* species are largely identical in sequence, but one has two  $\gamma$ -carboxyglutamate residues, whereas the other has none (see Fig. 4 a and b).

Both the *C. gloriamaris* and *C. textile* peptides were chemically synthesized on solid support and folded in the presence of glutathione as described in *Materials and Methods*. The biological activities of the two purified synthetic peptides were evaluated by intracerebroventricular injection into mice; despite the differences in  $\gamma$ -carboxylation, no difference in biological activity could be detected. When injected into mice, both peptides elicited the response characteristic of the native spasmodic peptide, hypersensitivity to touch at a dose of 2 nmol and

hyperactivity and convulsions at 5 nmol. These results indicate that the Gla residues do not play a major role in the molecular interactions between the mature folded peptide and the physiological target of the *C. textile* peptide.

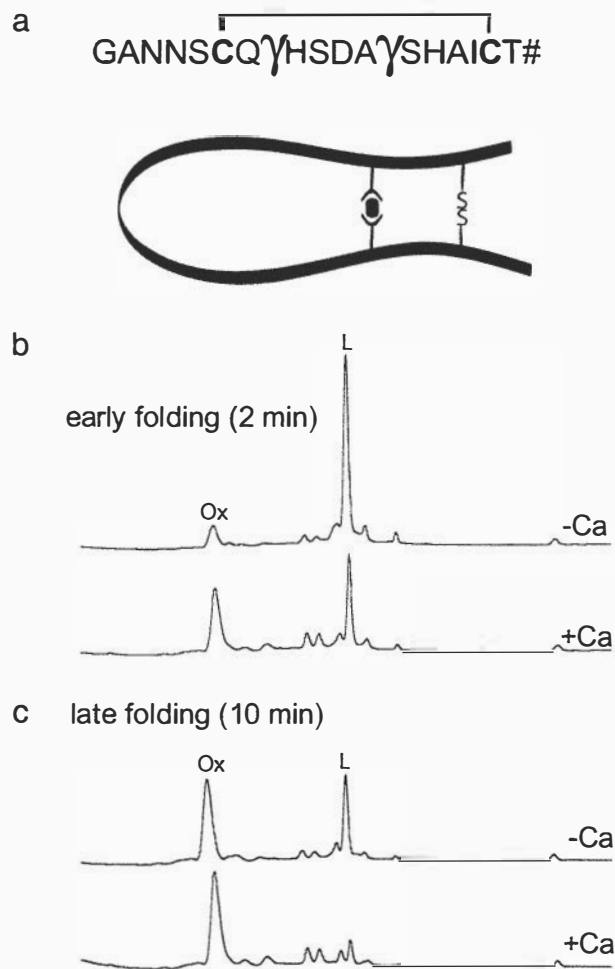
We then assessed whether the Gla residues might affect peptide folding. Under our standard Ca-free folding conditions, yields of the properly folded *C. gloriamaris* peptide were slightly higher than for *C. textile* peptide (Fig. 4 c and d). Because  $\gamma$ -carboxyglutamate residues chelate  $\text{Ca}^{2+}$ , we evaluated the effects of  $\text{Ca}^{2+}$  on the folding of the two peptides. As shown in Fig. 4e, the addition of  $\text{Ca}^{2+}$  results in a significant increase in the yield of properly folded *C. textile* peptide; strikingly,  $\text{Ca}^{2+}$  elicited no increase in the yield of the *C. gloriamaris* peptide, which lacks  $\gamma$ -carboxyglutamate (Fig. 4f). Above 10 mM  $\text{Ca}^{2+}$ , the *C. textile* (Gla<sup>+</sup>) peptide is more efficiently folded than the *C. gloriamaris* (Gla<sup>-</sup>) peptide, a reversal of the situation in the absence of  $\text{Ca}^{2+}$  (Fig. 4 g and h). Furthermore, the magnitude of the increase in yield of correctly folded peptide is not observed when  $\text{Mg}^{2+}$  is added over the same concentration range (Fig. 4 i and j). Moreover,  $\text{Ca}^{2+}$  at 10 mM concentration did not significantly change the reductive unfolding kinetics of *C. textile* (Gla<sup>+</sup>) peptide (data not shown). These results are consistent with the conclusion that once the peptides are correctly folded,



**Fig. 4.** Effect of Ca<sup>2+</sup> on the oxidative folding of peptides from *C. textile* and *C. gloriamaris*. (a and b) Sequences of peptides from *C. textile* and *C. gloriamaris*. (c–f) HPLC analysis of oxidative folding of the peptides. The folding mixtures contained reduced and oxidized glutathione, and the reactions were carried out in the presence of 10 mM CaCl<sub>2</sub> or 1 mM EDTA. Experimental details are described in *Materials and Methods*. The correctly folded species (\*) has the shortest retention time. (c) Folding of *C. textile* peptide in the absence of Ca. (d) Folding of *C. gloriamaris* peptide in the absence of Ca. (e) Folding of *C. textile* peptide in the presence of Ca. (f) Folding of *C. gloriamaris* peptide in the presence of Ca. (g and h) Kinetics of forming the native peptides from *C. textile* (g) and *C. gloriamaris* (h). The folding reactions were performed as described in *Materials and Methods*. Dashed lines, +10 mM Ca<sup>2+</sup>; solid lines +1 mM EDTA. (i and j) Accumulation of the native peptides from *C. textile* (i) and *C. gloriamaris* (j) in the folding reactions at 24 h, with increasing concentration of calcium ions. Two other bars represent accumulation of the native peptides in the presence of 1 mM EDTA or 100 mM Mg<sup>2+</sup>. All experimental points are averages from duplicate experiments.

$\gamma$ -carboxyglutamate residues make no significant contribution to biological activity or stability; however, the  $\gamma$ -carboxyglutamate residues clearly facilitate oxidative folding when Ca<sup>2+</sup> is present, as would be the case in the ER under *in vivo* folding conditions.

To further investigate a mechanism of the Ca<sup>2+</sup>-assisted oxidative folding of the Gla-containing peptide, we synthesized a model peptide, which represents a fragment of the conotoxin from *C. textile*, but containing only one pair of cysteines and two Gla residues (Fig. 5a). The second pair of Cys was replaced by Ala. The rationale for using this model peptide is to assess whether the presence of the Gla residues accelerates formation of a disulfide bond in the presence of Ca<sup>2+</sup>. As shown in Fig. 5b and c, the presence of 10 mM Ca<sup>2+</sup> in the reaction mixture increased the accumulation of the oxidation product both at early (2 min) and later (10 min) time points. These results provide a mechanistic rationale, at least in part, for the observed



**Fig. 5.** Ca<sup>2+</sup>-assisted oxidation of the Gla-containing model peptide. The peptide (a) was designed based on the sequence of peptide from *C. textile*. Cys-2, Cys-12, and Cys-16 were replaced by Ala residues, and Thr-19 had an amidated C terminus (labeled #). The disulfide bond between Cys-6 and Cys-18 is also shown. A hypothetical model illustrating how a coordination of Ca<sup>2+</sup> by the two Gla residues in the peptide may favor formation of a disulfide bridge (-SS-) by bringing two Cys thiols in close proximity. (b and c) HPLC analysis of oxidation of the model peptide with a mixture of oxidized (1 mM) and reduced (2 mM) glutathione and in the presence of either 10 mM CaCl<sub>2</sub> or 1 mM EDTA. The oxidation mixtures were quenched by acidification after 2 or 10 min and analyzed by reversed-phase C<sub>18</sub> analytical HPLC. L, linear (reduced) peptide; Ox, oxidized form.

effects of Ca<sup>2+</sup> on folding of the conotoxin from *C. textile* (see *Discussion*).

## Discussion

In the course of their evolutionary history, the venomous cone snails have developed an unprecedented number of diverse, highly structured peptides that are expressed in their venom ducts. These peptides as a class are the smallest ribosomally translated functional gene products (microproteins). Most of these peptides potentially interfere with the function of a specific ion channel target. To have the potent and selective pharmacological activity required, the peptides have to be properly folded with the correct disulfide connectivity. Clearly, cone snails have generated mechanisms to facilitate small-peptide folding; this article begins to address such mechanisms. There are at least two motivations for embarking on this general research direction: the first (rather pedestrian) reason is that understanding factors that facilitate proper conotoxin folding *in*

*vivo* should help in the routine production of properly folded *Conus* peptides *in vitro*. However, our long-term hope is that by elucidating how cone snails are able to fold these small peptides properly, we will also gain insights into the more general problem of protein folding.

All *Conus* peptide precursors have highly conserved N-terminal signal sequences that direct the newly translated precursor to the ER membrane, transferring it from a highly reducing, low  $\text{Ca}^{2+}$  environment (the cytosol) to an oxidizing, high  $\text{Ca}^{2+}$  environment (the ER matrix). In the oxidizing environment of the ER (14), the high density of cysteine residues in most *Conus* peptides would be expected to lead to rapid formation of disulfide bonds. We provide evidence for high levels of multiple PDI isoforms in *Conus* venom ducts; these enzymes generally regulate the formation of disulfide bonds and prevent them from being formed randomly. PDI appears to be the most highly expressed high molecular weight polypeptide component of *Conus* venom ducts. PCR amplification of PDI is only intermittently successful when other *Conus* tissues are similarly analyzed, presumably because of a much lower frequency of PDI transcripts. These observations are consistent with a key role for PDI in the folding of the major secreted products of these cells, the conotoxins. An obvious possibility to be explored is whether different *Conus* PDI isoforms have evolved specialized features to facilitate proper folding of the specific disulfide frameworks in the various conotoxin superfamilies. An additional question is whether *Conus* PDIs have synergistic interactions with other specialized chaperones in the ER that could lead to even greater facilitation of conotoxin folding.

The many posttranslational modifications observed in *Conus* peptides are believed to enhance the affinity and/or pharmacological specificity of modified conopeptides; however, the results presented above suggest an additional function for some *Conus* peptide posttranslational modifications: to facilitate proper folding. In the example presented ( $\gamma$ -carboxylation of glutamate residues), formation of the correctly disulfide-bonded conopeptide was enhanced by the Gla residues in the presence of physiological concentrations of  $\text{Ca}^{2+}$ , but did not appear to be important for activity of the folded peptide. The results presented suggest that the presence of  $\gamma$ -carboxylglutamate residues facilitate folding of the *C. textile* peptide *in vivo*.

A role for  $\text{Ca}^{2+}$  in oxidative folding was previously suggested for secreted proteins such as the low-density lipoprotein (LDL) receptor (15, 16), Notch1 receptors (17) or  $\alpha$ -lactalbumin (18). For the LDL receptor, a "calcium box" rich in Asp or Glu was postulated to direct folding by chelating  $\text{Ca}^{2+}$  and nucleating early folding intermediates as the LDL receptor entered the ER. Similarly, as disulfide-rich  $\gamma$ -carboxylated *Conus* peptides are translocated through the ER membrane, we suggest that a compact calcium mini-box with  $\gamma$ -carboxylglutamate residues may either facilitate a favorable spatial orientation of Cys residues for correct disulfide bond formation at early stages of folding, leading to the final biologically active conformation, or alternatively, inhibit formation of undesirable disulfide bonds. The experiment on a model peptide (Fig. 5) provides support for the former, but does not eliminate the possibility that the latter may contribute as well.  $\gamma$ -Carboxylation does not appear to be required for the biological activity of the mature, properly folded peptide gene product, nor is it required for maintaining the structure of peptide once correct disulfide bonds are formed.

$\gamma$ -Glutamyl carboxylases, the enzymes that convert glutamate to  $\gamma$ -carboxylglutamate (Gla), were previously shown to be highly conserved, clearly predating the divergence of vertebrates, arthropods, and molluscs (19). Thus, an ancestral role (more general than the specialized uses of the modification in blood clotting in mammals and venom peptides in *Conus*) that provided selective pressure for enzyme conservation as animal phyla

diverged is strongly indicated. The present results raise the possibility that the proper folding of proteins, directed by the presence of Gla residues at strategic loci in the polypeptide, may have been that ancestral role. We suggest that this putative ancient role of Gla was recapitulated to facilitate folding of some of the very rapidly evolving, unusually small *Conus* peptides.<sup>5</sup> Such a folding mechanism for proteins may have been more generally important earlier in evolution, but it was probably largely supplanted later by other mechanisms for facilitating folding of larger polypeptides, such as specialized molecular chaperones (20, 21).

The role of Gla suggested above provides an attractive general mechanism for folding small polypeptides, perhaps even including the primordial proteins. Based on structural work on signal recognition particle peptides, it was recently suggested (22) that the first proteins evolved as membranes formed, when RNA still dominated biochemistry. Specifically, the first functional polypeptide-like chains in incipient life forms were created to deal with a membrane surrounding the catalytic/informational RNA. If this view is correct, then the possibility is raised that  $\gamma$ -carboxylglutamate, with its capacity both for interacting with membranes and directing folding may have been present in the earliest functional polypeptides, which were presumably much smaller than present-day proteins. Once a  $\text{Ca}^{2+}$ -free cytosol evolved, however, a doubly negatively charged amino acid might become a liability for intracellular protein function, and in most taxa at the present time,  $\gamma$ -carboxylglutamate is probably largely a relict amino acid in a few secreted proteins. This modification remains prominent only in those present-day phylogenetic systems where more specialized uses have evolved (such as mammalian blood clotting and *Conus* venom peptides) (Fig. 3).

The canonical organization of conopeptide precursors, into an N-terminal signal sequence, an intervening precursor, and a single copy of the mature toxin region at the C terminus was first established a decade ago (23). The striking contrast found between the conserved signal sequence and propeptide regions versus the hypermutable mature toxin region led to the suggestion that the conserved regions of the precursor were essential for proper folding "such that the single biologically active disulfide-bonded configuration can be specifically formed." It was pointed out in the same paper that *in vivo*, additional factors to facilitate the folding process might be required, such as PDI; in this work, we provided the characterization of PDI from *Conus* venom ducts. Evidence that posttranslational modification enzymes may also be extrinsic factors that facilitate proper folding was presented above. For both PDI and posttranslational modification enzymes, the propeptide region may play a key role by providing anchor sites, so that these factors can act efficiently on the mature toxin region. This has been directly demonstrated for the posttranslational modification enzyme that carboxylates glutamate to  $\gamma$ -carboxylglutamate, a specific sequence present in the conotoxin precursor propeptide region is recognized by the enzyme (24). Evidence that PDIs facilitate proper folding of a conopeptide precursor was also recently obtained (O.B. and G.B., unpublished results). Thus, the relatively conserved propeptide regions may have a function in establishing the specific disulfide bonding conformation of *Conus* peptides.

<sup>5</sup>An obvious question is why  $\gamma$ -carboxylglutamate in *C. textile* peptide is not present in *C. gloriamaris* peptide. *C. textile* peptide is from a tropical *Conus* species from a warm, shallow-water habitat. *C. gloriamaris* peptide is from a deep-water species (~100 fathoms); the significantly cooler habitat may minimize the pressure to facilitate folding by the presence of Gla. An alternative/additional rationale is that *C. textile* peptide is a major venom component, whereas *C. gloriamaris* peptide was never directly detected in venom, but only from a cDNA clone (implying that it may be a quantitatively minor venom component). A higher production rate would provide additional selective pressure for mechanisms that facilitate folding.

The high expression levels and multiple isoforms of PDI in *Conus* venom ducts and the involvement of a posttranslational modification enzyme are by no means the only adaptation that cone snails have made to efficiently fold the small, highly structured, rapidly evolving peptides in their venoms. The studies described here only serve to identify candidates in what is likely a long list of molecular adaptations necessary for the genus *Conus* to implement their unusual biochemical/pharmacological strategy for interacting with the prey, predators, and competitors in their environment.

In closing, we address the link between the amplification and diversification of conotoxins and macroevolutionary trends in *Conus*. The fossil record indicates that the genus had its origins after the Cretaceous extinction, with a modest initial radiation in the Eocene (25); molecular data are consistent with several early lineages of *Conus* diverging from each other at this time (2, 26). Together, the fossil record and molecular data suggest that one of these early *Conus* lineages expanded rapidly at the beginning of the Miocene, a radiation that basically continues to

the present day. This gave rise to the great majority of the shallow-water species, including all of the fish-hunting and mollusc-hunting *Conus*.

The peptidic nature of conotoxins will make it possible to assess general evolutionary patterns in the conotoxin genes as these radiations of *Conus* lineages took place. However, we can begin to evaluate a different level of molecular adaptations as well. As we understand more about the oxidative folding process, we should in principle be able to determine at which point the various components that facilitate oxidative folding of conotoxins appear. Thus, one question that can be evaluated through a comparison of different early diverging lineages is whether most of these adaptations took place early in the history of the genus, or whether a breakthrough in folding smaller conotoxins could have been a factor that made the major radiation of cone snails in the Miocene possible.

This work was supported by National Institute of General Medical Sciences Grant GM 48677.

1. Olivera, B. M., Rivier, J., Clark, C., Ramilo, C. A., Corpuz, G. P., Abogadie, F. C., Mena, E. E., Woodward, S. R., Hillyard, D. R. & Cruz, L. J. (1990) *Science* **249**, 257–263.
2. Olivera, B. M. (2002) *Annu. Rev. Ecol. Syst.* **33**, 25–42.
3. Röckel, D., Korn, W. & Kohn, A. J. (1995) *Manual of the Living Conidae* (Verlag Christa Hemmen, Wiesbaden, Germany).
4. Terlau, H. & Olivera, B. M. (2004) *Physiol. Rev.* **84**, 40–67.
5. Craig, A. G., Bandyopadhyay, P. & Olivera, B. M. (1999) *Eur. J. Biochem.* **264**, 271–275.
6. Lirazán, M. B., Hooper, D., Corpuz, G. P., Ramilo, C. A., Bandyopadhyay, P., Cruz, L. J. & Olivera, B. M. (2000) *Biochemistry* **39**, 1583–1588.
7. Goldenberger, R. F., Epstein, C. J. & Anfinsen, C. B. (1963) *J. Biol. Chem.* **238**, 628–635.
8. Tu, B. P., Ho-Schleyer, S. C., Travers, K. J. & Weissman, J. S. (2000) *Science* **290**, 1571–1574.
9. Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. & Creighton, T. E. (1997) *Curr. Biol.* **7**, 239–245.
10. Darby, N. J., Penka, E. & Vincentelli, R. (1998) *J. Mol. Biol.* **276**, 239–247.
11. Macer, D. R. & Kock, G. L. (1988) *J. Cell Sci.* **91**, 61–70.
12. Lucero, H. A. & Kaminer, B. (1999) *J. Biol. Chem.* **274**, 3243–3251.
13. Miles, L. A., Dy, C. Y., Nielsen, J., Barnham, K. J., Hinds, M. G., Olivera, B. M., Bulaj, G. & Norton, R. S. (2002) *J. Biol. Chem.* **277**, 43033–43040.
14. Gilbert, H. F. (1990) *Adv. Enzymol. Relat. Areas Mol. Biol.* **63**, 69–172.
15. Bieri, S., Atkins, A. R., Lee, H. T., Winzor, D. J., Smith, R. & Kroon, P. A. (1998) *Biochemistry* **37**, 10994–11002.
16. Koduri, V. & Blacklow, S. C. (2001) *Biochemistry* **40**, 12801–12807.
17. Aster, J. C., Simms, W. B., Zavala-Ruiz, Z., Patriub, V., North, C. L. & Blacklow, S. C. (1999) *Biochemistry* **38**, 4736–4742.
18. Rao, K. R. & Brew, K. (1989) *Biochem. Biophys. Res. Commun.* **163**, 1390–1396.
19. Bandyopadhyay, P. K., Garrett, J. E., Shetty, R. P., Keate, T., Walker, C. S. & Olivera, B. M. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 1264–1269.
20. Gupta, R. S., Aitken, K., Falah, M. & Singh, B. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2895–2899.
21. Gupta, R. S. (1995) *Mol. Biol. Evol.* **12**, 1063–1073.
22. Walter, P., Keenan, R. & Schmitz, U. (2000) *Science* **287**, 1212–1213.
23. Woodward, S. R., Cruz, L. J., Olivera, B. M. & Hillyard, D. R. (1990) *EMBO J.* **1**, 1015–1020.
24. Bandyopadhyay, P. K., Colledge, C. J., Walker, C. S., Zhou, L.-M., Hillyard, D. R. & Olivera, B. M. (1998) *J. Biol. Chem.* **273**, 5447–5450.
25. Kohn, A. J. (1990) *Malacologia* **32**, 55–67.
26. Espiritu, D. J. D., Watkins, M., Dia-Monje, V., Cartier, G. E., Cruz, L. J. & Olivera, B. M. (2001) *Toxicon* **39**, 1899–1916.
27. McIntosh, J. M., Olivera, B. M., Cruz, L. J. & Gray, W. R. (1984) *J. Biol. Chem.* **259**, 14343–14346.
28. Walker, C., Steel, D., Jacobsen, R. B., Lirazán, M. B., Cruz, L. J., Hooper, D., Shetty, R., Delacruz, R. C., Nielsen, J. S., Zhou, L., *et al.* (1999) *J. Biol. Chem.* **274**, 30664–30671.
29. Craig, A. G., Jimenez, E. C., Dykert, J., Nielsen, D. B., Gulyas, J., Abogadie, F. C., Porter, J., Rivier, J. E., Cruz, L. J., Olivera, B. M. & McIntosh, J. M. (1997) *J. Biol. Chem.* **272**, 4689–4698.

# Non-self recognition, transcriptional reprogramming, and secondary metabolite accumulation during plant/pathogen interactions

Klaus Hahlbrock<sup>\*†</sup>, Pawel Bednarek<sup>\*</sup>, Ingo Ciolkowski<sup>\*</sup>, Björn Hamberger<sup>\*</sup>, Andreas Heise<sup>\*</sup>, Hiltrud Liedgens<sup>‡</sup>, Elke Logemann<sup>\*</sup>, Thorsten Nürnberger<sup>§</sup>, Elmon Schmelzer<sup>\*</sup>, Imre E. Somssich<sup>\*</sup>, and Jianwen Tan<sup>\*</sup>

<sup>\*</sup>Max-Planck-Institut für Züchtungsforschung, Carl-von-Linne-Weg 10, D-50829 Köln, Germany; <sup>†</sup>von Leerodt Strasse 6a, D-52445 Titz, Germany; and <sup>§</sup>Institute of Plant Biochemistry, Weinberg 3, D-06120 Halle, Germany

Disease resistance of plants involves two distinct forms of chemical communication with the pathogen: recognition and defense. Both are essential components of a highly complex, multifaceted defense response, which begins with non-self recognition through the perception of pathogen-derived signal molecules and results in the production, *inter alia*, of antibioticly active compounds (phytoalexins) and cell wall-reinforcing material around the infection site. To elucidate the molecular details and the genomic basis of the underlying chains of events, we used two different experimental systems: suspension-cultured cells of *Petroselinum crispum* (parsley) and wild-type as well as mutant plants of *Arabidopsis thaliana*. Particular emphasis was placed on the structural and functional identification of signal and defense molecules, and on the mechanisms of signal perception, intracellular signal transduction and transcriptional reprogramming, including the structural and functional characterization of the responsible cis-acting gene promoter elements and transacting regulatory proteins. Comparing *P. crispum* and *A. thaliana* allows us to distinguish species-specific defense mechanisms from more universal responses, and furthermore provides general insights into the nature of the interactions. Despite the complexity of the pathogen defense response, it is experimentally tractable, and knowledge gained so far has opened up a new realm of gene technology-assisted strategies for resistance breeding of crop plants.

Most plant/pathogen interactions are fierce battles of attack and counterattack. These battles are fought with highly sophisticated means for the survival of the individual and, in the end, of the entire population or species. On the plant side, the most immediate defense response includes the reprogramming of cellular metabolism and highly dynamic, structural rearrangements within and around the attacked cells. In the cases of locally invading, fungal or fungus-like pathogens, the counterstroke of the plant commences in a highly localized fashion with the perception of chemical and physical signals from the intruder and ends with the accumulation of soluble, antibioticly active compounds and wall-bound, barrier-forming substances. The initiating event, attempted penetration of a potential pathogen, immediately activates an elaborate safe-guard system of non-self recognition based on specifically adapted plant receptors. These receptors recognize characteristic pathogen-borne surface molecules and transduce that information to numerous genes through a network of intracellular signaling cascades that orchestrate an extensive, defense-oriented transcriptional reprogramming of the affected cell. Among the major changes in cellular metabolism is the rapid accumulation of various secondary metabolites, some of which are likely to be integral to the complex, multicomponent defense response.

This network of events, from the initial stage of recognition by the plant to the successful confinement or death of the pathogen,

is far too fine-meshed to be elucidated by using one single experimental system. For our investigations of a few selected key events, we have used two complementary model systems: suspension-cultured *Petroselinum crispum* (parsley) cells and *Arabidopsis thaliana* plants. Cells and protoplasts of *P. crispum* proved to be ideal tools for analyzing the cell biology, biochemistry, and molecular biology of the defense response. Wild-type and mutant *A. thaliana* plants were particularly suited for transgenic studies and for investigating defined host plant/pathogen interactions in combination with the plant genetic background. Overlaps at certain focal points enabled us to directly compare these two systems and to infer both species-specific and universal defense mechanisms. In both experimental systems, our analyses of secondary metabolites encompassed aromatic phenylpropanoid as well as indolic compounds, among which antibioticly active phytoalexins and physicochemically active barrier-forming substances were of particular interest.

Here, we combine an overview of earlier findings with data obtained from recent experiments specifically designed to facilitate an interspecies comparison, a summary of the general principles observable so far, and a brief outline of one possible strategy for practical application of the results in crop plant breeding.

## From Elicitor Perception to Secondary Metabolite Accumulation

**Introductory Overview.** The use of cultured *P. crispum* cells as our preferred experimental system for molecular and cell biological analyses offers several major advantages. First, the exogenously applied, pathogen-derived signal molecule is a chemically defined, small, and highly specific peptide elicitor that enables detailed structural and functional analyses based on specially designed synthetic modifications. Second, protoplasts derived from these cells retain full elicitor responsiveness and hence are a particularly valuable tool for analyzing gene promoter elements and their transcriptional regulators by using a simple and highly reproducible transfection and transient expression assay. Third, the synchronous elicitation of cultured cells, in sharp contrast to the largely asynchronous infections of whole-plant tissue, enables the precise determination of temporal gene expression characteristics. Furthermore, most, if not all, of the readily identifiable, elicitor-induced, soluble and wall-bound aromatic metabolites in *P. crispum*

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: ACE, ACGT-containing element; bZIP, basic leucine zipper; 4CL, 4-coumarate:CoA ligase; ROS, reactive oxygen species; CMPG, cys/met/pro/gly; PAL, phenylalanine ammonia-lyase; WRKY, trp/arg/lys/tyr.

<sup>†</sup>To whom correspondence should be addressed. E-mail: hahlbroc@mpiz-koeln.mpg.de.

© 2003 by The National Academy of Sciences of the USA

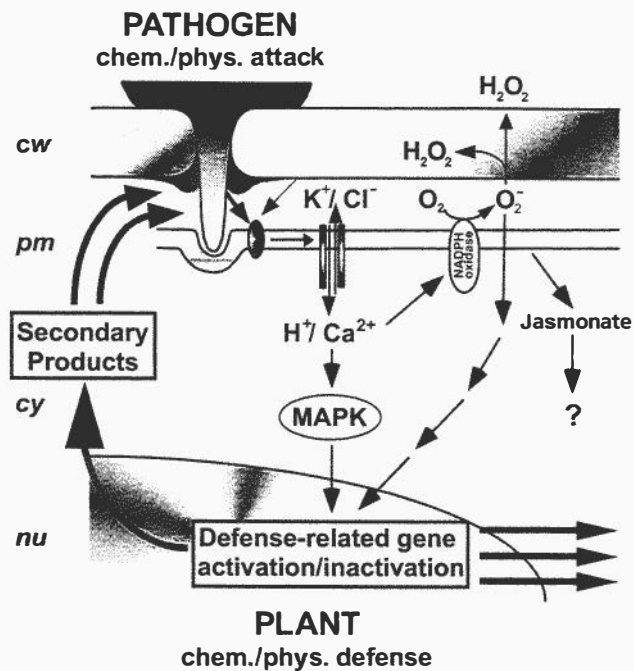


Fig. 1. Schematic outline of early molecular responses of *P. crispum* cells to attempted penetration of a *Phytophthora sojae* hypha. cw, cell wall; pm, plasma membrane; cy, cytoplasm; nu, nucleus.

are phenylpropanoid derivatives; that is, they are members of a single and well characterized class of compounds, in contrast to more heterogeneous chemical responses observed in many other systems, including *A. thaliana*.

Fig. 1 schematically outlines the various extra- and intracellular events elucidated so far for the *P. crispum* system. These events extend from physical perception of the pathogen's infection hypha and chemical recognition of one highly active component of a putative mixture of pathogen-derived elicitors to transcriptional reprogramming of the cell's metabolic state and the consequential accumulation of defense-related, aromatic secondary compounds. All of the underlying cell biological experiments were conducted by using either *Phytophthora sojae* or *P. infestans*, two closely related pathogenic oomycetes to which *P. crispum* is nonhost resistant. For most of the molecular analyses, the live pathogen was replaced either with a chemical derivative (elicitor) or with a physical mimic (a sharp needle). As far as applicable to the largely undifferentiated, cultured cells or protoplasts, and as far as individual components have been analyzed, the response to elicitor was essentially the same as that observed with infected, whole-plant tissue, except for the additional occurrence of hypersensitive cell death at true infection sites. It should be noted in this connection that hypersensitive cell death, though not further investigated in *P. crispum*, is probably among the most efficient defense responses in this and many other systems.

**Signal Perception.** Even the earliest steps in the interaction between *P. crispum* and *Phytophthora* spp. are highly complex, but a few key events have been identified. After the formation of an appressorium as a tight physical holdfast, the pathogen develops an infection hypha in an attempt to invade the cell below (Fig. 1). This hypha is likely to exert mechanical force during further growth and penetration by virtue of strong physical support from the appressorium, but at the same time exposes its surface to the plant's surveillance mechanism for non-self recognition. The mechanical interaction has been experimentally decoupled from chemical signaling by replacing the hypha with a needle mimic. Even a gentle touch with

Table 1. Competitor activities of Pep13 analogs

Peptide sequence	Competitor activity, %
VWNQPVRGFKVYE (Pep13)	100
VWLQPVRGFKVYE	8
VWLLPVRGFKVYE	0
VWN <u>E</u> PVRGFKVYE	1
VWN <u>A</u> QPVRGFKVYE	2
VWN <u>AAA</u> QPVRGFKVYE	5
VWQPVRGFKVYE	0

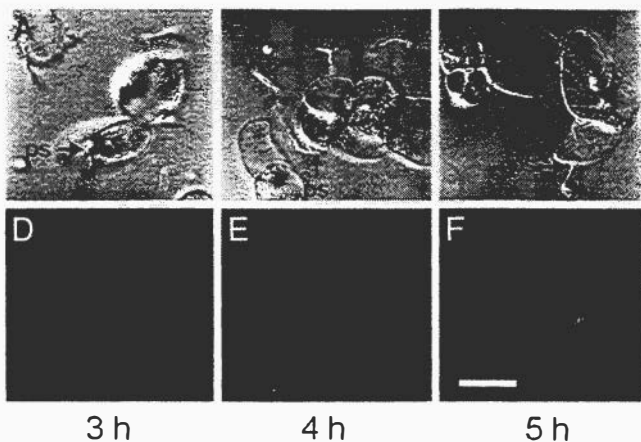
Competitor activities using the standard binding assay (2) are given in percent of Pep13. Bold letters mark the essential W2 and P5 residues (2); underlining indicates amino acid substitutions or insertions.

a tungsten needle induces several, though not all, of the reactions observed after elicitor treatment or true infections, including the activation of some defense-related genes and the migration of nucleus and cytoplasm toward the site of physical contact (1).

In addition to pathogen-borne, exogenous elicitors, attacked plant cells are exposed to numerous breakdown products from their own damaged cell wall, some of which can also have significant roles in signal perception by synergistically acting as so-called endogenous elicitors. We speculate that the overall composition of the respective mixture of elicitor molecules enables the plant to activate the most appropriate defense response against a particular type of attacking pathogen. However, this facet has not been conclusively demonstrated in any pathosystem.

Isolation and purification to homogeneity of the most elicitor-active component from the culture filtrate of *Phytophthora sojae* yielded a 42-kDa glycoprotein containing an oligopeptide of 13 aa (Pep13) that in pure form is necessary and sufficient to stimulate the same complex defense response as observed either with a crude elicitor preparation or with the live pathogen (2). Elicitor activity of Pep13 is not restricted to *P. crispum* and leads, for example, to PR gene activation in cultured cells or leaves of potato (*Solanum tuberosum*) (3). Binding studies using radioiodinated Pep13 as a ligand and *P. crispum* microsomal membranes or protoplasts as receptor preparations in combination with chemical cross-linking identified a plasma membrane-associated protein complex containing a 91-kDa protein with high binding affinity for Pep13. Studies on structure-activity relationships using various chemically synthesized Pep13 derivatives revealed the particular importance of amino acid residues W2 and P5 for both receptor binding and phytoalexin induction (2, 3). Moreover, Table 1 shows that the nature and the spacing of amino acids N3 and Q4 between these two essential residues are equally important for Pep13-receptor binding. Together, these findings indicate a high degree of structural specificity within this small signal peptide, despite the previous observation that all amino acid residues N- and C-terminal of W2 and P5 can be individually replaced with an alanine residue without significant loss of elicitor activity (2).

These results were corroborated and extended by recent data showing a high degree of structural conservation, along with an important functional role, of Pep13 as a surface-exposed loop structure within its parent glycoprotein, a Ca<sup>2+</sup>-dependent transglutaminase (3). This type of transglutaminase occurred in all *Phytophthora* species analyzed, with the Pep13 motif almost fully conserved in all cases, but absent in all known plant protein sequences. Importantly, the same structural requirements were essential for both elicitor activity of Pep13 (ref. 2; Table 1) and transglutaminase activity of the intact protein (3). Thus, a strong selective pressure for conservation of the Pep13 motif as an essential component of transglutaminase activity, combined with its unique occurrence in pathogens (as opposed to plants) reveals this oligopeptide region as the target against which the highly specific



**Fig. 2.** Progressive elevation of the  $\text{Ca}^{2+}$  level in *Phytophthora sojae*-infected *P. crispum* cells. Hyphal penetration sites (ps) were identified under white light (A–C); Fluo-4 fluorescence was visualized under blue light (D–F). Time points are given in hours postinoculation. (Bar = 100  $\mu\text{m}$ .)

plant receptor for non-self recognition arose during evolution of this defense response.

Some oomycete species, including *Phytophthora sojae*, possess not only the Pep13-containing transglutaminase, but also a 24-kDa cell-wall protein, necrosis-inducing *Phytophthora* protein 1 (NPP1), that elicits a defense response in *P. crispum* cells very similar to that observed with Pep13 (4). Unlike Pep13, however, NPP1 additionally induces the formation of hypersensitive cell death-like lesions in various dicotyledonous plants, including *P. crispum* and *A. thaliana*. In *P. crispum*, the NPP1-mediated defense response does not involve the Pep13 receptor, but employs all of the other signaling components involved in the Pep13-triggered response (Fig. 1), suggesting an early convergence of the two elicitation pathways. It is presently open whether simultaneous recognition of both elicitors leads to synergistic effects.

**Signal Transduction.** Receptor-mediated influx of extracellular  $\text{Ca}^{2+}$  is among the first detectable responses of *P. crispum* cells to treatment with either Pep13 or elicitor-active derivatives thereof (2). This  $\text{Ca}^{2+}$  influx is conducted in part by a novel type of Pep13-responsive, plasma membrane-associated ion channel, the activation of which is a prerequisite for the triggering of all subsequent responses (5, 6). Previous results demonstrating the rapid elevation of the cytoplasmic free  $\text{Ca}^{2+}$  level in elicitor-stimulated cells by using the bioluminescent  $\text{Ca}^{2+}$  indicator, apoaequorin (6), have now been extended by monitoring the spatial progression of intracellular  $\text{Ca}^{2+}$  accumulation in the course of the infection process. By using the indicator dye Fluo-4, we could demonstrate a rapid, strong elevation of the  $\text{Ca}^{2+}$  level that progressed steadily within a few hours from the initial site of hyphal penetration throughout the entire cell (Fig. 2) and remained high for a prolonged period, as with the measurements using apoaequorin (6). Besides concomitant, functionally unresolved increases in several other ion fluxes (Fig. 1), two additional plasma membrane-associated,  $\text{Ca}^{2+}$  influx-dependent, defense-related events occur more or less simultaneously with the onset of  $\text{Ca}^{2+}$  accumulation: the generation of reactive oxygen species (ROS) and jasmonate (2, 7).

Among these multiple elicitor-induced events within the plasma membrane, the elevated  $\text{Ca}^{2+}$  and ROS levels are of particular relevance for the subsequent intracellular signal transduction to the nucleus. However, although both of them are potent mediators of defense-related gene activation, they affect, at least in part, different sets of target genes. One cytoplasmic signaling pathway leads from  $\text{Ca}^{2+}$  via the activation of at least three mitogen-activated

protein kinases to strong increases in *PR* gene expression, whereas another ROS-related pathway triggers the activation of phenylpropanoid-biosynthetic genes and thus the induction of the various aromatic compounds to be discussed below (5, 7). These observations indicate the involvement of at least two distinct signaling cascades, one ROS-dependent and the other ROS-independent, in defense-related gene activation in *P. crispum*. By contrast, the molecular target(s) of a third Pep13-stimulated cascade, the jasmonate pathway (Fig. 1), are still elusive. In cultured *P. crispum* cells, pharmacological inhibitors of Pep13-induced jasmonate accumulation did not impair phytoalexin production or *PR* gene activation (7).

**Targets of Intracellular Signaling.** The cell culture system has revealed at least three major targets of the intracellular signaling pathways: extensive transcriptional reprogramming of the affected cell from “normal” to defense-oriented metabolism (8, 9); reorganization of the cytoskeleton and translocation of the nucleus, together with a sizable portion of the cytoplasm, to the penetration site (10); and extracellular conversion and extension of the signaling both to locally confined areas around the infection site (10, 11) and systemically throughout the entire affected organ or even the whole organism. In this latter regard, however, only circumstantial evidence exists so far in *P. crispum*. Among these multiple targets, the two focal points of our studies are the phenomenon of metabolic reprogramming, particularly the mechanisms of defense-related gene activation and inactivation, and the nature and function of the subsequently accumulating aromatic metabolites.

Treatment of cultured *P. crispum* cells with the *Phytophthora sojae*-derived peptide elicitor (either Pep13 or Pep25, an equally effective, slightly longer version; ref. 2) leads to rapid transcriptional activation (8) or repression (12) of at least several dozens of genes. Although gene repression is mechanistically as well as metabolically as interesting a phenomenon as is gene activation, our investigations have been focused mainly on the latter, largely because most of the genes analyzed so far in relation to aromatic secondary metabolism are strongly activated by elicitor. Three major outcomes of these studies are particularly noteworthy: the identification of several distinct classes of elicitor-responsive, cis-acting gene promoter elements; the discovery of two new families of regulatory proteins, the trp/arg/lys/tyr (WRKY) (13) and cys/met/pro/gly (CMPG) (14) protein families; and the demonstration of a few cases of exceptionally rapid, immediate-early gene activation, commencing within minutes after elicitor application and thereby possibly revealing immediate target genes of the elicitor-induced, intracellular signaling (13, 14).

In the work summarized up to this point, our studies had been conducted with suspension-cultured *P. crispum* cells or protoplasts. By contrast, most of the analyses discussed in the following paragraphs were a combination of initiating studies using the cell culture system and subsequent, often more extensive, genome- and mutant-based investigations with *A. thaliana* plants. Because of this sequential approach, most of the basic new discoveries were made in *P. crispum*, whereas more general insight was gained by including *A. thaliana*.

**Cis-Acting Promoter Elements.** Several elicitor-responsive, cis-acting elements were first identified on *P. crispum* gene promoters and then served as starting points for comparative, more extensive analyses using the fully sequenced genome of *A. thaliana*. Equally important was the essential role of these elements in the identification of cognate binding proteins, again initially in *P. crispum* and then in *A. thaliana*. The first major finding was the discovery of a set of three almost invariably cooccurring elements specifically on the promoters of phenylpropanoid-biosynthetic genes. This set (boxes P, A, and L) was initially identified on the *PcPAL1* gene by “*in vivo* footprinting” and has since been shown to occur on every newly discovered phenylalanine ammonia-lyase (*PAL*) gene (with



**A** *P. crispum*

**PcPAL1**  
**P** CTCCAACAAACCC...  
**A** ...CCGTCC...  
**L** ...TTCACCTACC

**PcELI7**  
**S** CAGCCACCAAAGAGGACCCAGAAT

**PcPR2**  
**D** TACAATTCAAACATTGTTCAAACAAGGAACC

**PcPR1**  
**W1** CTTAATTTGACCGAGTA  
**W2** TCAAAGTTGACCAATAA  
**W3** TTTATTATGACTAAATAGTCAG

**PcWRKY1**  
**W<sub>BC</sub>** TGCTTTGACTGAGAAAATATTTTCCATGGTCAAAT

**PcCMPG1**  
**E17** TGTCAAATGGTCAACATTCAAC

**PcACO**  
**2×ACE** TCCTCCACGTATCATCCCCCATCCAATCCACGTATCAA

**B** *A. thaliana*

**AtCMPG1**  
**F** TTGTCAAATGCATTAAAATTCAAACATTCAACGGTCAAAT

**At4CL4**  
**W<sub>-1113</sub>** TTGTTTTGCAATAGTCAAAACTATACATCC  
**W<sub>-673</sub>** CAGCAAATATTATTGACCAAAGAAATGCAAA  
**W<sub>-496</sub>** AGAATTTTGGAGGGTCAACTGCGGAATGTA

Fig. 3. List of functionally identified, elicitor-responsive cis-acting elements in *P. crispum* and *A. thaliana*. See text for references and further explanations.

the occasional exception of the A box), as well as on all other genes constituting the three central steps of “general phenylpropanoid metabolism” in plants (15). In addition to this P, A, and L box combination, several other elicitor-responsive elements (Fig. 3) were subsequently identified on numerous *P. crispum* genes (box W and the D, S, and E17 elements; refs. 14 and 16). Of these, certain W box-containing elements, including E17, are particularly interesting because of their apparent involvement in immediate-early gene activation (see below).

Notwithstanding that the D and S elements have not been precisely defined, all of the positively or negatively elicitor-responsive elements from *P. crispum* (Fig. 3) occurred in structurally and functionally related forms in *A. thaliana* as well. Moreover, whenever elements derived from *P. crispum* gene promoters were tested in transgenic *A. thaliana* plants, or whenever *A. thaliana*-derived elements were tested in transfected *P. crispum* protoplasts, the results were essentially the same as those obtained within the species of origin.

**Regulatory Proteins.** Besides the identification of a P-box-binding factor (not sought in *A. thaliana*) and apart from the recognition of a close structural similarity of boxes P and L both to one another and to the independently defined myeloblastosis recognition element, two more basic discoveries resulting from these studies were the WRKY and CMPG protein families. Three *PcWRKY* isoforms with particularly interesting, distinct properties were initially identified in *P. crispum* (13). Their most characteristic features are the presence of the WRKY domain within otherwise poorly conserved proteins, and high affinities as well as high binding specificities for the W box. The most extensively analyzed representative, *PcWRKY1*, shows exceptionally rapid, immediate-early transcrip-

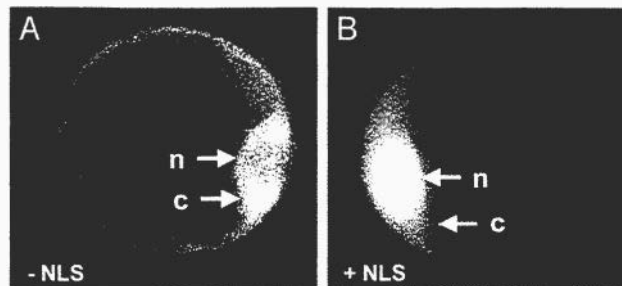


Fig. 4. Intracellular localization of a *PcCMPG1*::GFP fusion protein without (A) and with (B) a putative nuclear localization signal (NLS) within *PcCMPG1*. n, nucleus; c, cytoplasm.

tional induction by elicitor. Functional, tandemly arranged, elicitor-responsive W-box combinations within the *PcWRKY1* gene promoter itself suggest autoregulation of this gene (17).

Of the subsequently identified CMPG gene family, only one representative, *PcCMPG1*, has been analyzed in *P. crispum*. *PcCMPG1* was originally discovered as the fastest responding of all elicitor-responsive genes tested in this species, with an onset of immediate-early mRNA induction even faster (within < 5 min) than that recorded for *PcWRKY1* (14). Several lines of circumstantial evidence suggest that CMPG proteins also have regulatory functions, though not as DNA-binding transcriptional regulators, but more likely by exerting regulatory activities involving protein-protein interactions. The occurrence of a putative nuclear localization signal (NLS) within the *PcCMPG1* protein and at least partial, NLS-dependent translocation of a fusion product with green fluorescent protein into the nucleus (Fig. 4) suggest that the putative regulatory function is exerted within this cellular compartment.

Simultaneous with the rapid transcriptional activation of defense-related pathways by elicitor or infection, several unrelated metabolic activities, which are probably dispensable under these conditions, are equally rapidly repressed. In *P. crispum*, a striking example of such putatively compensatory repression is the flavonoid branch of phenylpropanoid metabolism, the products of which serve as UV-protective agents in this and many other species. The responsible genes as well as some of their transcriptional regulators are both strongly activated by UV light and strongly repressed by elicitor (12). Both stimuli act through the same promoter unit, possibly by inducing inversely acting isoform combinations of the differentially regulated common plant regulatory factor (CPRF) family of basic leucine zipper (bZIP) transcriptional regulators (12). Several members of this family had been identified as ACGT-containing element (ACE)-binding proteins with UV-light induction or repression patterns indicative of such a dual regulatory role (18).

These recently discovered or extended protein families were the next obvious targets for comparison between the two species. Similar to the cis-acting elements, all families of regulatory proteins analyzed so far in *P. crispum*, whether with or without DNA-binding properties, proved to have structurally and, as far as tested, functionally related counterparts in *A. thaliana*, including the large *AtWRKY* and *AtCMPG* protein families, as well as the two independently investigated, similarly large families of *AtMYB* and *AtbZIP* proteins. All four of these families can be subdivided into structurally distinct subclasses. In each case, several, but not all, of the family members are positively or negatively responsive to elicitor (or infection), to UV light, or to yet other stimuli that lie outside the scope of this article. A comparative overview of all available data on the intensely investigated WRKY, CMPG and bZIP protein families in *A. thaliana* and on all presently known family members in *P. crispum* is given in Fig. 5. Particularly noteworthy in this connection is the striking similarity between

WRKY		CMPG		CPRF/ bZIP	
At2g04880	At4g24240	At5g65500	At2g31370	At2g31370	At2g31370
At5g56270	At4g31550	At2g19410	At1g06070	At1g06070	At1g06070
At2g03340	At2g23320	At5g61550	At1g06850	At1g06850	At1g06850
At1g13960	At2g24570	At4g25160	At2g40620	At2g40620	At2g40620
At1g55600	At2g30590	At5g51270	At1g43700	At1g43700	At1g43700
At4g12020	At3g04670	At2g45910	At4g38900	At4g38900	At4g38900
At4g26640	At5g28650	At3g61390	At2g21230	At2g21230	At2g21230
At2g30250	PcWRKY3	At2g45920	At2g42380	At2g42380	At2g42380
At5g07100	At1g30650	At2g45920	At3g58120	At3g58120	At3g58120
At4g30930	At5g45050	At1g01660	At2g13150	At2g13150	At2g13150
At2g38470	At4g01250	At1g01670	At2g12980	At2g12980	At2g12980
At4g26440	At5g52830	At3g49055	At2g12900	At2g12900	At2g12900
At2g37260	At4g23550	At4g36550	At2g13130	At2g13130	At2g13130
At3g01970	At2g34820	At5g18330	At2g24340	At2g24340	At2g24340
At3g01080	At1g29280	At5g18320	At5g08150	At5g08150	At5g08150
AV526908	At3g58710	At5g18340	At2g21235	At2g21235	At2g21235
PcWRKY1	At5g24110	At5g18340	At2g40950	At2g40950	At2g40950
PcWRKY2	At5g22570	At5g1840	At3g56660	At3g56660	At3g56660
At4g31800	At4g11070	At3g52450	At3g10800	At3g10800	At3g10800
At1g80840	At2g46400	At2g35930	At1g42990	At1g42990	At1g42990
At2g25000	At5g45270	At3g19390	At5g11260	At5g11260	At5g11260
PcWRKY4	At4g23810	At1g49750	At3g17610	At3g17610	At3g17610
At1g62300	At2g40750	At1g49750	At5g07160	At5g07160	At5g07160
At1g68150	At2g40740	AtCMPG5	At5g06950	At5g06950	At5g06950
At4g22070	At5g01900	AtCMPG1	At5g06960	At5g06960	At5g06960
At1g69810	At1g66600	PcCMPG1	At3g02840	At3g02840	At3g02840
At4g04450	At1g66560	At3g02840	At1g08320	At1g08320	At1g08320
At4g01720	At1g80590	At1g20815	At5g06840	At5g06840	At5g06840
At1g18860	At1g66550	AtCMPG4	At1g68840	At1g68840	At1g68840
At5g15130	At3g56400	At3g49810	At5g65210	At5g65210	At5g65210
PcWRKY5	At5g46350	At3g07360	At5g10030	At5g10030	At5g10030
At5g46350	AC003672	At5g01830	At1g22070	At1g22070	At1g22070
At4g39410	At4g39410	At5g42340	At1g77920	At1g77920	At1g77920
At2g47260	At2g47260	At1g29340	At4g35040	At4g35040	At4g35040
At5g41570	At5g41570	At2g28830	At2g16770	At2g16770	At2g16770
At4g18170	At4g18170	At3g46510	At3g51960	At3g51960	At3g51960
At2g46130	At2g46130	At3g54850	At4g34000	At4g34000	At4g34000
At5g49520	At5g49520	At2g23140	At3g19290	At3g19290	At3g19290
At5g43290	At5g43290	At1g71020	At1g00000	At1g00000	At1g00000
At5g26170	At5g26170	At1g49720	At1g49720	At1g49720	At1g49720
At5g64810	At5g64810	At5g67340	At2g36270	At2g36270	At2g36270
At1g64000	At1g64000	At3g54790	At3g44460	At3g44460	At3g44460
At1g69310	At1g69310	At1g10580	At5g44080	At5g44080	At5g44080
At2g21900	At2g21900	At1g10580	At1g03970	At1g03970	At1g03970
At3g62340	At3g62340	At1g60190	At2g41070	At2g41070	At2g41070
At1g29860	At1g29860	At1g67530	At3g56850	At3g56850	At3g56850
At5g13080	At5g13080	At1g24330	At4g35900	At4g35900	At4g35900
		At1g27910	At2g17770	At2g17770	At2g17770
		AtCMPG3	At5g42910	At5g42910	At5g42910
		At3g18710			
		AtCMPG2			
		At5g09800			
		At1g56040			
		At1g56030			

Fig. 5. Schematic representation of the complete WRKY (20), CMPG (19), and common plant regulatory factor/bZIP (21) protein families from *A. thaliana* and the initially identified members from *P. crispum* (gray shading). Subdivisions within each family (19–21) are indicated by boxes, and gene codes are given where existing. Filled circles indicate known responsiveness to infection, elicitor, or other functionally related stimuli (19–22).

*PcCMPG1* and *AtCMPG1* not only in structural terms, but also with regard to the exceptionally rapid, immediate-early induction by elicitor (14, 19).

**Secondary Products.** Among the most abundant aromatic compounds accumulating in the culture fluid of elicitor-treated *P. crispum* cells or protoplasts, or at infection sites proper, are several structurally related furanocoumarins and various likewise closely related butylidene phthalide aglycones and glycosides (23). Both classes of compounds are biosynthetically derived from phenylalanine by combinations of the three common steps of general phenylpropanoid metabolism with two of the many subsequently diverging branch pathways. Furanocoumarins and butylidene phthalides possess antimicrobial activity and hence are assumed to act as phytoalexins in this species.

However, in contrast to the far-reaching similarities throughout all cis-acting elements and protein families investigated, *P. crispum* and *A. thaliana* differ considerably at the secondary metabolite

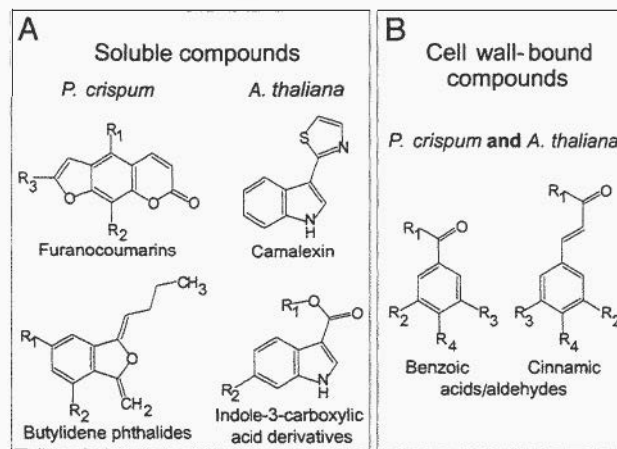


Fig. 6. Chemical structures of major, characteristic classes of compounds accumulating on elicitor treatment or infection in *P. crispum* and *A. thaliana*. Soluble (A) and cell wall-bound (B) material was extracted and analyzed by using established protocols (23, 24) with slight modifications.

level. In fact, all major aromatic metabolites accumulating in soluble form in infected leaves (24) or roots (P.B., unpublished results) of *A. thaliana* are indole derivatives, in contrast to their phenylpropanoid or betaketide origin in *P. crispum*, whereas all major compounds constitutively present in *A. thaliana* are, notably, phenylpropanoid derivatives. Fig. 6A gives a comparative overview of the elicitor- or infection-induced classes of aromatic compounds identified in the two systems. In parallel with the induction of indolic metabolites, some of the constitutively present phenolics in *A. thaliana* decline strongly during the infection process, again in leaves as well as in roots. Furthermore, in addition to the various metabolites identified, many less abundant aromatic as well as aliphatic compounds are probably also induced or repressed, but were not monitored in our studies.

Remarkably, this striking difference in defense-related, phenylpropanoid and indolic aromatic metabolism between *P. crispum* and *A. thaliana* appears to be largely restricted to soluble compounds. In the fraction of wall-bound, alkali-labile substances, only one or two indolic compounds, indole-3-carboxylic acid and possibly the corresponding aldehyde, accumulated strongly in *A. thaliana*, but not in *P. crispum*. All other induced wall-bound aromatic compounds are phenylpropanoid (benzoate and cinnamate) derivatives with similar or identical substitution patterns in both *A. thaliana* and *P. crispum* (Fig. 6B). Although so far unproven, it seems probable that these wall-bound compounds accumulate, at least in part, locally around the site of attempted penetration of the pathogen, where the papilla (Fig. 1) is formed by apposition of callose and other, largely unidentified materials onto the cell wall.

**Causal Connections.** Two closely related long-term goals of these studies have been to identify those metabolic junctions that causally connect the numerous individual components of the overall defense response, as well as those final products of the long chains of metabolic events that directly contribute to, or are decisive for, the actual impairment or killing of the pathogen. Three different experimental approaches have recently yielded insights in this direction.

The first approach involves the use of Actinomycin D as a general gene transcription inhibitor. Earlier studies had revealed the total inhibition of PAL induction in UV-irradiated *P. crispum* cells by low doses of Actinomycin D. Now, we used elicitor (1  $\mu\text{g/ml}$  Pep25) instead of UV light for induction, and observed strong inhibitory effects at three different metabolic levels on addition of Actinomycin D (50  $\mu\text{g/ml}$ ) 30 min before the onset of elicitor treatment.

E17-27 TCAATATG**TCAATGGTCAA**CATTCAAC  
 E17-21m4 TGT**CA**ATGGTCAA**C**ATTCAAC  
 E17-21m11 TGTCAATGGT**CA**ACATTCAAC

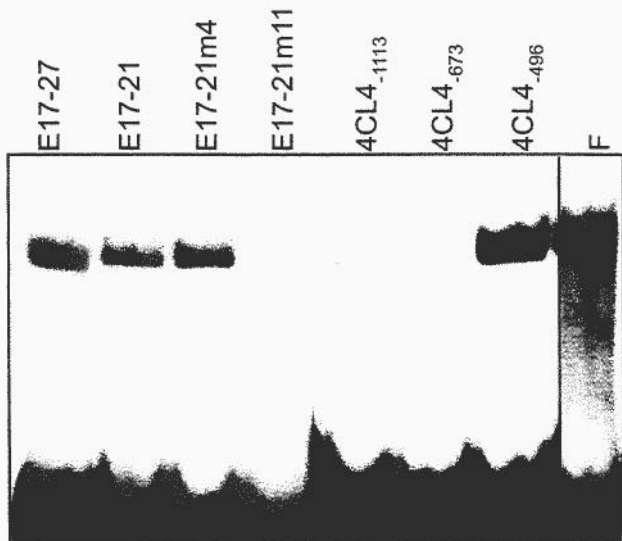


Fig. 7. Gel-shift assay indicating widely differing affinities of a few selected W box-containing elements for WRKY11, a selected member of the WRKY family of cognate DNA-binding proteins. Procedures for heterologous expression and purification of the protein (25) and for gel-shift analysis (13) were followed essentially as described. DNA probes E17-21, 4CL4, and F were those shown in Fig. 3. For probes E17-27 and E17-21m, see ref. 14.

Under these conditions, neither CMPG1 or WRKY1 mRNA (as measured 30 min after elicitor addition) nor PAL activity (24-h time point), nor any of the soluble or wall-bound elicitor-inducible compounds listed in Fig. 6 (24-h time point), accumulated significantly above the 0-h control level. Although these results do not demonstrate direct causal connections between any two of the items analyzed, they clearly indicate that in each case induction is the result of transcriptional activation of the responsible gene(s), for the immediate-early induced mRNAs as well as for PAL activity (the first committed step in phenylpropanoid biosynthesis) and for the various metabolic end products. Thus, we can exclude the theoretical possibility that at least some of the monitored effects were caused by transcription-independent mechanisms, an alternative that might, for example, occur in the form of secondary product synthesis by liberation and/or conversion of preformed intermediates. However, we cannot exclude relatively minor but possibly relevant contributions from such sources.

The second approach concerns the extent to which the numerous structural variants of cis-acting elements interact with different members of the large families of cognate binding proteins. This is a difficult question to answer conclusively for the situation *in vivo*. However, some of the available *in vitro* assays can give at least valuable hints in this direction. A sensitive tool for determining DNA-protein binding affinities is the analysis of band shifts caused by protein binding of DNA probes on electrophoretic mobility gels. Fig. 7 shows that a few structurally widely different, elicitor-responsive, W box-containing elements, including E17 from *PcCMPG1*, F from *AtCMPG1*, and three less precisely defined promoter regions from *At4CL4* exhibit distinct binding affinities for one selected WRKY protein, *AtWRKY11*. Importantly, however, nearly all of these elements are bound more or less efficiently by this particular family member. Considering the large size of the WRKY family of W box-binding proteins and the large number of potential W box-containing sequences and sequence arrangements, this result indicates a vast regulatory potential founded in an almost

#### Hierarchical structure of pathogen defense in plants

Triggered by non-self recognition, executed by interconnected signalling cascades and superimposed by highly dynamic intracellular rearrangements

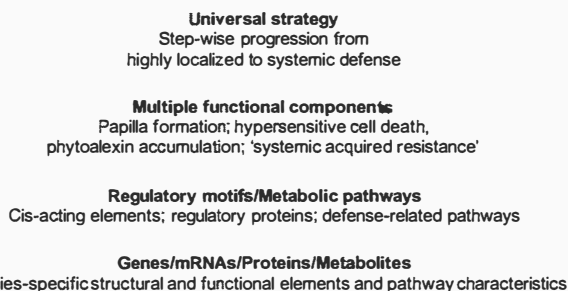


Fig. 8. Complexity pyramid illustrating the hierarchical organization and modular structure of the pathogen defense response in higher plants. Note that the examples used to explain the pivotal role of the two central levels of the pyramid in the gradual upward progression from the particular to the universal (26) are an incomplete selection from the modules discussed in the text.

infinite number of possible combinatorial permutations. Particularly relevant with regard to our goal to identify possible causal connections, it is apparent that any given W box-containing sequence has a high potential to bind with a certain affinity to any given WRKY protein, implicating a similarly wide range of functional relationships *in vivo*.

The third case exemplifies one of the most promising and powerful techniques available for the analysis of causal connections: the employment of defined mutants. Among the large and rapidly growing number of *A. thaliana* knockout mutants, many have been functionally associated with defined steps in defense-related signaling or secondary product formation, or more generally, with the disease resistance phenotype. A presently incomplete analysis employing several such mutants is expected to add substantially to our understanding of the network of defense-related metabolic interconnections.

#### Pathogen Defense in Plants: Paradigm of Structural and Functional Complexity

At each of the levels investigated, the plant's response to pathogen attack represents a paradigm of biological complexity (9). In addition to the complexity within each individual metabolic level, a hierarchically superimposed complexity pyramid emerges as the contours of the overall picture come into focus (Fig. 8). In accord with the notion of a complexity pyramid "from the particular to the universal" (26), our results point to at least three, possibly four distinct levels of hierarchical organization. At the top of the pyramid, one universal, multicomponent defense strategy governs several more or less universal, functional modules. Together with the various regulatory motifs and defense-related pathways at the upper and lower central levels of the pyramid, respectively, these modules mediate, coordinate, and execute the strategic response, whereas the bottom harbors the numerous individual, species-specific, defense-related genes and their products.

**Universal Defense Strategy.** In *P. crispum*, *A. thaliana*, and all other systems analyzed to date, the overall defense response consists of a series of sequentially activated defense measures. These proceed from the highly localized to the less highly localized and finally to the systemic, with an optional termination of the process at certain stages when pathogen confinement has been achieved. Among the most intensely investigated components of this response are the papilla formed around the site of attempted penetration; the highly localized, hypersensitive death of the immediately affected cell; the

local accumulation of antibiologically active substances, including H<sub>2</sub>O<sub>2</sub> and phytoalexins; and the systemic accumulation of hydrolytic enzymes, such as glucanase and chitinase. This strictly ordered, temporally, spatially, and functionally modular defense strategy entails dramatic intracellular rearrangements (10) concomitant with the onset of transcriptional reprogramming at the infection site.

However, the individual functional modules are probably not fully autonomous, but rather interconnected, and in some cases they may even partially overlap. Obvious examples of such metabolic or regulatory overlaps are the various interrelated responses to elicitor stimulation within the plasma membrane; the dual or even manifold roles of Ca<sup>2+</sup> in intracellular regulation; the role of H<sub>2</sub>O<sub>2</sub> both as a toxic agent and in cell-wall cross-linking; and the initial, common steps in the biosynthesis of phenylalanine-derived phytoalexins and cell wall-bound compounds. Even though the mechanistic details of such interconnections are presently obscure, the overall response of the plant reveals that they exist.

**Regulatory Motifs and Transcription Factor Families.** Both the perception of pathogen-derived elicitors and its subsequent intracellular signal transduction employ basic mechanisms that appear to be universal not only in higher plants, but also, at least to a considerable extent, in higher animals (27), suggesting a common evolutionary origin of the general defense strategy throughout all higher eukaryotes.

Numerous, more or less directly defense-related genes are major targets of two of the intracellular signal transduction chains in *P. crispum* (Fig. 1). Probably all of them contain at least one of a small number of basic types of elicitor-response element, two of which (containing either the W box or the P/A/L set of boxes) occur most frequently on these genes throughout all plants examined. Remarkably, the P/A/L box-containing element has so far been found on all genes encoding functionally identified or putative phenylpropanoid-biosynthetic enzymes. Moreover, these genes share, at least in *P. crispum*, two additional, remarkable features: they are activated by the Ca<sup>2+</sup>/O<sub>2</sub><sup>-</sup>-dependent signaling pathway, and the mRNA accumulation patterns follow identical time courses for all P/A/L-containing genes, in sharp contrast to the uncoordinated behavior of nearly all other elicitor-responsive genes analyzed (8). Most probably, these three common features are causally connected. Whether this extends to the cognate DNA-binding proteins remains to be seen.

It presently appears doubtful whether a similar close metabolic relationship exists among the large number of genes bearing W box-containing promoters. The W box is not only by far the most frequently occurring structural element of elicitor-response elements in plants, but is also notorious for its challenging diversity of closely spaced repetitions and/or combinations with other sequences. In such cases, at least one W box is usually essential for conveying the elicitor response, and certain types of repeat structure have been associated with immediate-early gene activation (14, 17, 19). Fig. 7 exemplifies the broad range of binding affinities exerted by just a few structurally distinct representatives of such W-box arrangements toward one selected member of the WRKY family of transcriptional regulators. This sample includes three W box-containing fragments from the *At4CL4* gene promoter, the only presently known case in *A. thaliana* where W boxes cooccur with the characteristic P/A/L-box set on phenylpropanoid-biosynthetic genes. The recently identified *At4CL4* gene encodes a rare isoform of 4-coumarate:CoA ligase (4CL) with unusual substrate specificity and may therefore have an exceptional metabolic function and atypical expression mode. Here, the W-box regions from the *At4CL4* promoter were used to test binding strength toward WRKY proteins, as predicted from sequence relationships with previously analyzed elements. The results substantiate these predictions and furthermore reveal a large influence of the sur-

rounding sequence on the binding affinity of W box-containing elements.

Although much less is known about the S and D elements, their identification as strong elicitor-responsive elements indicates that the diversity of such elements is by no means confined to those containing P/A/L and W boxes. A particularly interesting recent observation in this context was the repression by elicitor of previously UV light-activated genes through a positively UV light-responsive and negatively elicitor-responsive ACE/ACE element combination, as manifested in the *PcACO* promoter fragment shown in Fig. 3A. This inverse response of one promoter element to different kinds of stress is yet another example of the complexity and connectivity of regulatory circuits, including the possible involvement of the same family of DNA-binding proteins in both up and down-regulation of genes (12).

Taken together, these results reveal a modularity even at the level of elicitor-responsive gene promoters, with W box- and (metabolically more confined) P/A/L box-containing elements being the most abundant, universally occurring representatives. Recurrent modularity in the fine structure of these basic building blocks at the species and gene levels generates uniqueness and biological specificity. This principle also applies to the cognate DNA-binding proteins and to the peripheral transcriptional regulators. Just as with cis-acting elements, a few distinct classes of transacting factors, each occurring as structurally diversified families, yield discrete but universal modules. Thus, the number of functional combinations for these modules, amplified by homo- and heterodimeric forms, may parallel the number of physiological challenges to which the plant must respond.

Binding of a regulatory protein complex is required for inactivation as well as for activation of a gene. Although the mechanisms of gene inactivation have been studied much less extensively than those of gene activation, it is unlikely that they fundamentally differ. Some of our results require us to presume that it may be the combination of isoforms of a given transcription factor family that changes with up- or down-regulation of a particular gene, whereas the basic type remains unchanged (12).

**Metabolic Pathways and Aromatic Secondary Products.** A ramified dimension of complexity is reached at the level of secondary plant metabolism. In contrast to the cis-acting elements and transacting factors, many secondary metabolic pathways are unique to certain plant genera, families or even species, and most of them are species-specific at least with regard to the specified composition of the respective product bouquet. Accordingly, all major soluble, elicitor- or pathogen-induced aromatic compounds in *P. crispum* are species-specific mixtures of differently substituted phenylpropanoid derivatives, whereas in *A. thaliana* they are exclusively indolic intermediates or end products.

It is therefore all the more surprising that, with the exception of one or two indolic compounds in *A. thaliana*, all major induced cell-wall constituents are similar or identical phenylpropanoids in these two and several other species. Two alternatives seem equally plausible: either there is greater evolutionary pressure on the conservation of the cell wall-bound compounds or this branch of phenylpropanoid metabolism, whether it has a common evolutionary origin in all plants or has converged from multiple origins, has limited degrees of chemical freedom, in contrast to the various pathways generating phytoalexins that have evolved a greater chemical diversity. Although phytoalexin production and cell-wall reinforcement with phenylpropanoid derivatives can be regarded as two independent functional modules occupying parallel hierarchical positions within the overall defense strategy, their species-specific patterns of diversity could not differ more.

### Practical Applications

Our motivation for these investigations has been two-fold: fascination with an exciting combination of molecular, cellular, and

ecologically oriented research and the coupling of this excitement with the expectation that scientific discoveries would reveal new approaches to breeding disease resistance in crop plants. To reach this level of practical application, two requirements had to be fulfilled. First, the basic principles of pathogen defense in plants had to be understood. We suppose that this stage has been reached with sufficient clarity, despite many remaining gaps in our knowledge about the underlying mechanisms. Second, a biologically meaningful, heritable, crop plant-adapted strategy must be feasible in every respect, including physiological tolerance by the plant and acceptance by the public. Comparing the largely unexplored efficiency and overwhelming diversity of chemical defense, particularly the difficulty in pinpointing individual, universally applicable (physiologically tolerable as well as consumer-friendly) defense-related compounds, with the proven high efficiency of hypersensitive cell death as a defense mechanism, we decided on probing a new gene technology-based strategy in the latter direction.

Our strategy employs the recently identified, rapidly, strongly, locally, and specifically elicitor-responsive promoter elements (Fig. 3), either alone or in combinations, for example in conjunction with a gene encoding a cell death-conferring principle, such as a broadly acting ribonuclease (28). Transcriptional activation of such a construct on infection of a transgenic plant would cause or intensify hypersensitive suicidal death of the affected cell, and thus confer or augment a particularly efficient defense mechanism. A first successful proof of principle has demonstrated the feasibility of this strategy (28). However, further improvements will be necessary to adapt various details to the special needs of crop plant breeding. For example, strict avoidance of unspecific transgene expression in response to exogenous stimuli other than infections, or to endogenous effectors during plant development, is absolutely essential. This is by no means a trivial obstacle, because most infection-responsive genes also respond to wounding and other stresses, and,

in addition, are expressed during certain stages of development, such as flowering, senescence or root-tip growth.

Some of the promoter elements listed in Fig. 3, notably E17 and F, are exceptionally specific in their response to infections (14, 19) and hence may be particularly well suited for the design of artificial defense mechanisms. Recently published (16) as well as unpublished data (A.H.) on the specifics of promoter design and expression modes are aimed at broadening the basis for further improvements of this strategy, including its sophistication in detail. In the long term, for this or any other strategy for breeding disease resistance, multigenic traits will have to be introduced that are not easily overcome by mutations in the pathogens.

## Conclusions

The combination of a universal strategy with an almost infinite number of species-specific variations of the pathogen defense response in plants ideally illustrates the interplay of conceptual richness, metabolic options, physiological constraints and ecological demands within which evolution takes place. Hierarchical subdivision of the overall response into multiple functional modules is executed by interconnected, partly universal, partly species-specific signaling pathways that together mediate an extensive, rapid transcriptional reprogramming of exclusively species-specific genes. A few large, highly diversified families of cis-acting elements and transacting factors are the pivotal links between the particular and the universal. The potential for fine tuning through combinatorial permutations accounts for the remarkable interspecies functionality of these elements and their cognate factors in transgenic plants and impels their use in new strategies of crop plant breeding for an environmentally safe disease management.

We thank Dr. Bill Martin, Düsseldorf, Germany, and Dr. Dierk Scheel, Halle, Germany, for valuable comments on the manuscript.

- Gus-Mayer, S., Naton, B., Hahlbrock, K. & Schmelzer, E. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8398–8403.
- Nürnberger, T., Nennstiel, D., Jabs, T., Sacks, W. R., Hahlbrock, K. & Scheel, D. (1994) *Cell* **78**, 449–460.
- Brunner, F., Rosahl, S., Lee, J., Rudd, J. J., Geiler, C., Kauppinen, S., Rasmussen, G., Scheel, D. & Nürnberger, T. (2002) *EMBO J.* **21**, 6681–6688.
- Fellbrich, G., Romanski, A., Varet, A., Blume, B., Brunner, F., Engelhardt, S., Felix, G., Kemmerling, B., Krzymowska, M. & Nürnberger, T. (2002) *Plant J.* **32**, 1–16.
- Scheel, D. (1998) *Curr. Opin. Plant Biol.* **1**, 305–310.
- Blume, B., Nürnberger, T., Nass, N. & Scheel, D. (2000) *Plant Cell* **12**, 1425–1440.
- Kroj, T., Rudd, J. J., Nürnberger, T., Gäbler, Y., Lee, J. & Scheel, D. (2003) *J. Biol. Chem.* **278**, 2256–2264.
- Batz, O., Logemann, E., Reinold, S. & Hahlbrock, K. (1998) *Biol. Chem.* **379**, 1127–1135.
- Somssich, I. E. & Hahlbrock, K. (1998) *Trends Plant Sci.* **3**, 86–90.
- Gross, P., Julius, C., Schmelzer, E. & Hahlbrock, K. (1993) *EMBO J.* **12**, 1735–1744.
- Schmelzer, E., Krüger-Lebus, S. & Hahlbrock, K. (1989) *Plant Cell* **1**, 993–1001.
- Logemann, E. & Hahlbrock, K. (2001) *Proc. Natl. Acad. Sci. USA* **99**, 2428–2432.
- Rushton, P. J., Tovar Torres, J. T., Parniske, M., Wernert, P., Hahlbrock, K. & Somssich, I. E. (1996) *EMBO J.* **15**, 5690–5700.
- Kirsch, C., Logemann, E., Lippok, B., Schmelzer, E. & Hahlbrock, K. (2001) *Plant J.* **26**, 217–227.
- Hahlbrock, K., Scheel, D., Logemann, E., Nürnberger, T., Parniske, M., Reinold, S., Sacks, W. R. & Schmelzer, E. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 4150–4157.
- Rushton, P. J., Rheinstädler, A., Lipka, V., Lippok, B. & Somssich, I. E. (2002) *Plant Cell* **14**, 749–762.
- Eulgem, T., Rushton, P. J., Schmelzer, E., Hahlbrock, K. & Somssich, I. E. (1999) *EMBO J.* **18**, 4689–4699.
- Weisshaar, B., Armstrong, G. A., Block, A., da Costa e Silva, O. & Hahlbrock, K. (1991) *EMBO J.* **10**, 1777–1786.
- Heise, A., Lippok, B., Kirsch, C. & Hahlbrock, K. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9049–9054.
- Eulgem, T., Rushton, P. J., Robatzek, S. & Somssich, I. E. (2000) *Trends Plant Sci.* **5**, 199–206.
- Jakoby, M., Droege-Laser, W., Kroj, T., Tiedemann, J., Vicente-Carbajosa, J., Weisshaar, B. & Parcy, F. (2002) *Trends Plant Sci.* **7**, 106–111.
- Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H.-S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S. A., et al. (2002) *Plant Cell* **14**, 559–574.
- Hagemeyer, J., Batz, O., Schmidt, J., Wray, V., Hahlbrock, K. & Strack, D. (1999) *Phytochemistry* **51**, 629–635.
- Hagemeyer, J., Schneider, B., Oldham, N. J. & Hahlbrock, K. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 753–758.
- Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A. & Nakamura, K. (2001) *Biosci. Biotechnol. Biochem.* **65**, 2428–2436.
- Oltvai, Z. N. & Barabasi, A.-L. (2002) *Science* **298**, 763–764.
- Nürnberger, T. & Brunner, F. (2002) *Curr. Opin. Plant Biol.* **5**, 318–324.
- Strittmatter, G., Janssens, J., Opsomer, C. & Botterman, J. (1995) *Bio/Technology* **13**, 1085–1089.

# Systemins: A functionally defined family of peptide signals that regulate defensive genes in Solanaceae species

Clarence A. Ryan\* and Gregory Pearce

Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340

Numerous plant species have been known for decades that respond to herbivore attacks by systemically synthesizing defensive chemicals to protect themselves from predators. The nature of systemic wound signals remained obscure until 1991, when an 18-aa peptide called systemin was isolated from tomato leaves and shown to be a primary signal for systemic defense. More recently, two new hydroxyproline-rich, glycosylated peptide defense signals have been isolated from tobacco leaves, and three from tomato leaves. Because of their origins in plants, small sizes, hydroxyproline contents (tomato systemin is proline-rich), and defense-signaling activities, the new peptides are included in a functionally defined family of signals collectively called systemins. Here, we review structural and biological properties of the systemin family, and discuss their possible roles in systemic wound signaling.

prosystemin | systemin receptor | plant defense

**S**ystemin, the initial peptide signal found in plants, is an intracellular signaling molecule that is synthesized within the amino acid sequence of a 200-aa precursor, called prosystemin (1, 2). Systemin induces proteinase inhibitor protein synthesis in leaves of young tomato plants when supplied for a few minutes through their cut stems at nanomolar concentrations (1). Radioactively labeled systemin, when placed on wound sites on leaves, is found in the phloem (1, 3). A key role for systemin in systemic signaling was established by showing that tomato plants expressing an antisense prosystemin gene become deficient in long-distance wound signaling and are more susceptible to insect attacks than wild-type plants (4).

In contrast to animal peptide hormones, the systemin precursor protein lacks a leader or signal sequence that is required for synthesis and processing through the secretory pathway (2). Immunolocalization techniques revealed that prosystemin is localized in parenchyma cells of vascular bundles (5). This localization in the vicinity of the sieve tubes of the phloem may facilitate transport of systemin and the oxylipins it induces in response to wounding to distal cells. Systemin activates defensive genes by interacting with a cell-surface receptor, called SR160, a 160-kDa transmembrane protein with an extracellular leucine-rich domain, and an intracellular receptor kinase domain (6, 7). The interaction of systemin with the receptor is the first step of a complex intracellular signaling pathway that involves the activation of a mitogen-activated protein kinase (MAPK) (8), the rapid alkalization of the extracellular medium (7, 9), the activation of a phospholipase (10, 11), and the release of linolenic acid that is converted into oxylipins such as phytodienoic acid and jasmonic acid that are powerful signals for defense genes (Fig. 1) (12, 13). The pathway exhibits analogies to the inflammatory response in animals (14) in which wounding activates MAPKs, phospholipases, the release of arachidonic

acid from membranes, and its conversion to prostaglandins, which are analogs of phytodienoic acid and jasmonic acid

The early alkalization in response to systemin in tomato suspension cultures was the basis for the development of an assay system that led to the identification and characterization of the systemin receptor, SR160 (7). SR160 is homologous to the BRI1 receptor from *Arabidopsis* (15), with a high percentage of amino acid identity. This was the first indication that the systemin receptor may be a close relative of the BRI1 receptor. This possibility was confirmed by the identification and cloning of the tomato brassinolide receptor, BRI1 (16), which was found to be identical to the tomato SR160 receptor. The identity of a receptor with two functions, i.e., defense and development, was unique in plants, but examples are known in the animal kingdom. The dual function of the SR160/BRI1 receptor was supported by experiments in which the tomato SR160/BRI1 receptor cDNA was expressed in tobacco, which does not express a prosystemin gene and therefore does not produce systemin as a defense signal (17). Transformed tobacco suspension-cultured cells synthesized the receptor and targeted it to the cell surface membranes of tobacco, where it displayed the identical binding characteristics with systemin as SR160 in tomato cells. The systemin–receptor interaction in tobacco cells induced the alkalization response, indicating that signaling components for the early steps in the systemin signaling pathway were present in tobacco and could be activated by the tomato SR160 receptor when it interacted with systemin. Additionally, a tomato mutant *cu-3*, which was caused by a mutation in the BRI1 receptor and led to the isolation of the BRI1 gene (16), is severely impaired in systemin signaling (17).

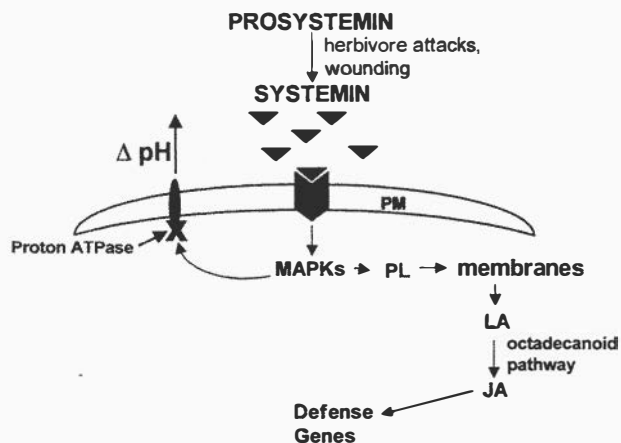
Because tobacco does not produce systemin, the presence of components in tobacco cells that react to the systemin–receptor interaction indicated that the BRI1 receptor may have, or may have had in the past, a defensive role in plants that was co-opted by systemin as the prosystemin gene evolved in species of the Solanaceae subtribe of the Solanaceae family. Tobacco does exhibit a fairly strong systemic defense response to wounding in young plants, but it is much weaker in older plants. Wounded tobacco plants synthesize a trypsin inhibitor (TTI) that is a paralog of tomato inhibitor II (18), which is induced in tomato leaves in response to wounding. The induction of TTI in tobacco

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: MAPK, mitogen-activated protein kinase; TobHypSys, tobacco hydroxyproline-rich systemin; TomHypSys, tomato hydroxyproline-rich systemin.

\*To whom correspondence should be addressed. E-mail: cabudryan@hotmail.com.

© 2003 by The National Academy of Sciences of the USA



**Fig. 1.** A simplified diagram of the systemin signaling pathway. The pathway shows several key steps of the signaling pathway, and in particular the steps leading to the blockage of a proton ATPase that leads to the alkalization of the extracellular medium, which is the basis of the assay developed to identify signaling peptides.

leaves in response to wounding indicates a genetic link between the wound-signaling systems of tomato and tobacco, despite the absence of systemin in tobacco. The synthesis of TTI in young tobacco plants is strongly induced by jasmonic acid (18), indicative of the early steps of signaling that result in the release of linolenic acid from membranes, similar to tomato plants. The roles of both systemin and jasmonate in systemic signaling have been the subject of considerable speculation (19, 20). We hypothesize here that the evolution of the prosystemin gene in species of the Solanaceae subtribe resulted in the production of systemin, a strong systemic signal that is not found in other plants, that amplifies the jasmonate signaling pathway. Prosystemin released from cells at the wound site is likely processed to systemin by proteinases also released from damaged cells. This would allow diffusion of systemin to the apoplast of nearby unwounded vascular cells to interact with its receptor and induce the synthesis of jasmonates. As jasmonates move through the plant, it would induce more prosystemin along with proteolytic enzymes that are known to be induced by jasmonate (14) that could process the nascent prosystemin to systemin to continue to amplify the jasmonate signal in nearby cells.

A major source of jasmonates at wound sites is from linolenic acid that is produced by the degradation of membrane lipids within the cellular debris. This source of jasmonates would likely provide an important “kick-start” for defense signaling, as the oxylipins diffuse into the vascular system and are transported to parenchyma cells (5) to up-regulate signaling pathway genes, including the prosystemin gene (in Solanaceae species). This hypothetical scenario led us to suspect that other peptide signals that were not systemic may be present in tobacco and tomato plants that might help amplify wound signaling. Such peptides in tobacco, which lacks a systemic peptide signal, might contribute to a localized amplification of the synthesis of jasmonates in response to wounding, and to amplification of jasmonate synthesis in the absence of systemin.

### Tobacco Systemins I and II

The search for peptide signals in tobacco was facilitated by the development of a biological assay that is based on the alkalization of the medium of tomato suspension-cultured cells in response to systemin, which is characterized by an increase in pH (up to 1 pH unit per 10 min) in the culture medium (9). Suspension cultured tobacco cells do not exhibit an alkalization

Peptide	Amino Acid Sequence	Pentose Units
TobHypSys I	<sup>1</sup> RGANLPOOSOASSOOSKE <sup>18</sup>	9
TobHypSys II	<sup>1</sup> NRKPLSOOSORPADGQR <sup>18</sup>	6
Systemin	<sup>1</sup> AVQSKPPSKRDPPKMTD <sup>18</sup>	0

**Fig. 2.** The amino acid sequences of tobacco hydroxyproline-rich systemins, TobHypSys I and TobHypSys II, are shown compared with the sequence of tomato systemin. Hydroxyproline (O), proline (P), threonine (T), and serine (S) residues are in red, charged amino acids are in black, and neutral amino acids are in blue. The ranges of pentose units attached to each peptide, determined by mass spectrometry, are shown at the right.

response when supplied with tomato systemin, but do so in response to a crude peptide fraction obtained from tobacco leaves, suggesting that a peptide–receptor interaction may be occurring that is coupled to an intracellular response. The alkalization of 1 ml of suspension cultured tobacco cells in response to 1- $\mu$ l aliquots from fractions eluting from HPLC or other columns revealed the presence of two peptides that, when purified and characterized, were found to be 18-aa glycopeptides that contained multiple hydroxyproline residues (21). The two peptides are active in the alkalization assay with tobacco suspension cultures at nM concentrations, and both cause a rapid activation of a 48-kDa MAPK, similar to the 48-kDa MAPK activated by tomato systemin in tomato cells (8). These peptides, supplied to young excised tobacco plants through their cut stems at nM concentrations, induce the synthesis of TTI in leaves.

Because of their similarities to tomato systemin in signaling properties, the two peptides were called tobacco systemin I and II (21). However, because of their hydroxyproline (O) contents, they are now named tobacco hydroxyproline-rich systemin (TobHypSys) I and II to identify them as members of a functionally related systemin family (22). Their amino acid sequences are shown in Fig. 2. Neither peptide exhibits homology with tomato systemin, but -OOS- motifs found in the tobacco peptides are posttranslational modifications of the primary translation motif -PPS- that is found in tomato systemin (1). The two new peptides are rich in P/O residues, and in S and T residues as well. These three amino acids make up 50% of each peptide and are likely involved in their recognition as defense signals.

Mass spectroscopy of the two peptides revealed that the attached carbohydrate moieties consist of pentose residues; nine in TobHypSys I and six in TobHypSys II. The structural properties of TobHypSys I and II (leader sequence, hydroxylation of -P- residues, and carbohydrate decorations) indicate that they are synthesized through the secretory system, unlike tomato systemin, which is not glycosylated, whose prolines are not hydroxylated, and whose precursor has no signal sequence (1, 2). Both TobHypSys peptides originate from a single 165-aa-long preproprotein, including a signal sequence, with the TobHypSys I sequence near the N terminus and the TobHypSys II sequence near the C terminus (21). The presence of multiple signaling peptides contained in a single preproprecursor is a characteristic of many animal peptide hormones, but the two tobacco systemins provide the first example in plants of a peptide hormone precursor harboring multiple peptide signals.

Although tobacco does not use a tomato systemin homolog for systemic wound signaling, TobHypSys I and II appear to serve roles in defense signaling. Because proTobHypSys is hydroxylated and glycosylated, like well characterized hydroxyproline-rich glycoproteins (23), it may be associated with cell walls, and may be processed from the precursor at wound sites to provide signals to amplify the synthesis of oxylipins

Peptides	Amino Acid Sequences	Pentose Units
TomHypSys I	<sup>1</sup> RTOYKTOOOOTSSSOT <sup>18</sup> Q	8-17
TomHypSys II	<sup>1</sup> GRDYVASOOOORPQDEQR <sup>20</sup> Q	12-16
TomHypSys III	<sup>1</sup> GRDSVLPPOOSOK <sup>15</sup> D	10

Fig. 3. The amino acid sequences of tomato hydroxyproline-rich systemins, TomHypSys I, proline (P), TomHypSys II, and TomHypSys III, are shown. Hydroxyproline (O), threonine (T), and serine (S) residues are colored as in Fig. 2. The number of pentose units associated to each peptide are shown in the right column.

during long distance wound signaling. Zhang and Baldwin (24) have elegantly shown that wounding of tobacco causes the synthesis of jasmonic acid that acts as a systemic signal from leaves to roots. It may be that the TobHypSys peptides help generate jasmonic acid that is targeted to the roots of the plant in response to wounds.

### Tomato Leaf Hydroxyproline-Containing Systemin Peptides

The isolation of 18-aa, glycosylated, hydroxyproline-containing tobacco systemins led to an investigation of the possibility that tomato plants may also have peptide defense signals similar to the tobacco systemins. The alkalization assay used to identify and isolate the two tobacco systemins was used to analyze tomato leaf extracts for peptide signals in addition to systemin. The assay identified several components from tomato leaf extracts that caused an alkalization response. Purification and characterization of these components confirmed that one peptide was tomato systemin and identified three new peptides (22). The novel peptides exhibit several properties similar to TobHypSys I and II, being hydroxyproline-rich glycopeptides, and ranging in size from 15 to 20 aa. Each of the peptides contains an internal continuous sequence of from 5 to 11 aa variously composed of O, P, S, or T residues, and all are flanked by various charged residues (Fig. 3). The peptides are decorated with variable numbers of pentose residues, but their identities and locations on the peptides have not been determined. The amino acid sequences of tomato hydroxyproline-rich systemin (TomHypSys) II and III indicated that they shared limited amino acid sequence homology and were likely products of gene duplication-elongation events. The three tomato peptides exhibit similar biological activities as tomato systemin, indicating that they are defense signals (22). They all exhibit similar specific activities in the alkalization assays, and all are effective inducers of proteinase inhibitors I and II synthesis when supplied to young tomato plants. The tomato peptides were therefore included in the functionally defined systemin family and named tomato hydroxyproline-rich systemins, i.e., TomHypSys I (18 amino acid residues), TomHypSys II (20 amino acid residues), and TomHypSys III (15 amino acid residues). Although the three TomHypSys peptides are powerful inducers of defense genes when supplied to excised tomato plants, they do not serve as primary systemic signals, because tomato plants transformed with an antisense prosystemin gene were incapable of systemic signaling in response to wounding (4).

Isolation and characterization of cDNAs coding for the tomato peptides revealed that all three were derived from the same 146-aa preproprotein precursor that includes a signal sequence. This precursor, along with the precursor of the TobHypSys peptides, provides the only examples in plants of polyprotein hormone precursors. Box diagrams of the three precursor proteins that harbor the six members of the systemin



Fig. 4. Box diagrams of the precursors of tobacco and tomato systemin peptides. The open boxes represent the leader sequences of the newly translated proteins. The HypSys peptides are shown as hatched boxes, with their numeral identities above each peptide. The systemin peptide is shown as Sys. The length of each preproprecursor is shown in parentheses.

family are compared in Fig. 4. A comparison of the amino acid sequences of the TomHypSys precursor with the TobHypSys precursor revealed a 10-aa sequence at their N termini that were identical at eight residues. The nucleotide sequence identity of this sequence was 90%. The significance of this identity is not clear, but does suggest that the two precursor genes may have a common ancient precursor, and that this sequence may have an important function that has been conserved. No homology was evident between prosystemin and either of the two preproprecursor proteins. However, it is of interest that the sequence of tomato systemin contains 7 of 18 residues that are P, S, and T (1). Because prosystemin has been found only in species of the Solanaceae subtribe of the Solanaceae family, we speculate that prosystemin may have been a member of the TomHypSys family and that some mutational event may have caused the loss of the leader sequence that resulted in the synthesis of the nascent precursor peptide to shift from a secretory pathway origin to a cytoplasmic origin, providing a powerful systemic defense signal (systemin) that was retained in the evolving species of the Solanaceae subtribe.

### Perspectives

The multiple P, O, S, and T residues in all six members of the systemin family in tomato and tobacco plants suggest that these residues have important structural roles for interacting with receptors. The P residues confer polyproline II structures (PP II) that have distinct kinks that may be the key to receptor recognition (25, 26). PP II structures are commonly found in peptide ligands of animals, where they appear to be important for recognition by receptors (27). In all five HypSys peptides, the central P and O residues are flanked by basic or acidic amino acids, either internally or near both the N and C termini.

The discovery of the HypSys defense signals in tomato and tobacco raise many questions about wound signaling in these and other plant species. The relationship of the systemins to local and systemic signaling and whether the HypSys peptides interact with homologs of the systemin receptor or have entirely different receptors for each peptide remain to be determined. Also of interest is whether the different peptides in tomato plants can activate the same complement of defense genes as systemin in response to wounding. The presence of a family of functionally similar HypSys defense signaling peptides in tomato and tobacco that are derived from paralogous precursors introduces the possibility that, in other plant families, related defense signaling peptides may be present that share a common ancestral origin. A search for such signals is now possible by using the same strategies that led to the discovery of the defense signaling peptides and their genes in tomato and tobacco that are described herein.

We thank S. Vogtman for plant growth and maintenance of plants, G. Munske for amino acid sequence analyses, and W. Seims for matrix-assisted laser desorption ionization mass spectrometric analyses. This research was supported by Washington State University College of Agriculture and Home Economics Project 1791 and by National Science Foundation Grant IBN 0090766.



1. Pearce, G., Strydom, D., Johnson, S. & Ryan, C. A. (1991) *Science* **253**, 895–897.
2. McGurl, B., Pearce, G., Orozco-Cardenas, M. & Ryan, C. A. (1992) *Science* **255**, 1570–1573.
3. Narvaez-Vasquez, J., Pearce, G., Orozco-Cardenas, M. L., Franceschi, V. R. & Ryan, C. A. (1995) *Planta* **195**, 593–600.
4. Orozco-Cardenas, M., McGurl, B. & Ryan, C. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 8273–8276.
5. Narvaez-Vasquez, J. & Ryan, C. A. (2003) *Planta*, in press.
6. Scheer, J. M. & Ryan, C. A. (1999) *Plant Cell* **11**, 1525–1535.
7. Scheer, J. M. & Ryan, C. A. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9585–9590.
8. Stratmann, J., Scheer, J. & Ryan, C. A. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 8862–8867.
9. Meindl, T., Boller, T. & Felix, G. (1998) *Plant Cell* **10**, 1561–1570.
10. Narvaez-Vasquez, J., Florin-Christensen, J. & Ryan, C. A. (1999) *Plant Cell* **11**, 1–13.
11. Ishiguro, S., Kawai-Oda, A., Weda, K., Nishida, L. & Okada, K. (2001) *Plant Cell* **13**, 2191–2209.
12. Farmer, E. E. & Ryan, C. A. (1991) *Plant Cell* **4**, 129–134.
13. Wasternack, C. & Hause, B. (2002) *Prog. Nucleic Acid Res. Mol. Biol.* **72**, 165–221.
14. Bergey, D., Howe, G. & Ryan, C. A. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 12053–12058.
15. Li, J. & Chory, J. (1997) *Cell* **90**, 926–938.
16. Montoya, T., Nomura, T., Farrar, K., Kaneta, T., Yokota, T. & Bishop, G. (2002) *Plant Cell* **14**, 3163–3176.
17. Scheer, J. M., Pearce, G. & Ryan, C. A. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 10114–10117.
18. Pearce, G., Johnson, S. & Ryan, C. A. (1993) *Plant Physiol.* **102**, 639–644.
19. Li, L., Li, C., Gyu, I. L. & Howe, G. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 6416–6421.
20. Stratmann, J. (2003) *Trends Plant Sci.* **8**, 247–250.
21. Pearce, G., Moura, D. S., Stratmann, J. & Ryan, C. A. (2001) *Nature* **411**, 817–820.
22. Pearce, G. & Ryan, C. A. (2003) *J. Biol. Chem.* **278**, 30044–30050.
23. Sommer-Knudson, J., Bacic, A. & Clarke, A. E. (1998) *Phytochemistry* **47**, 483–497.
24. Zhang, Z.-P. & Baldwin, I. T. (1997) *Planta* **203**, 436–441.
25. Toumadje, A. & Johnson, C., Jr. (1995) *J. Am. Chem. Soc.* **117**, 7023–7024.
26. Vanhoof, G., Goossens, F., De Meester, I., Hendriks, D. & Scharpe, S. (1995) *FASEB J.* **9**, 736–744.
27. Ferriss, P. J., Woessner, J. P., Waffenschmidt, S., Kilz, S., Drees, J. & Goodenough, U. W. (2001) *Biochemistry* **40**, 2978–2987.

# *Manduca sexta* recognition and resistance among allopolyploid *Nicotiana* host plants

Yonggen Lou\* and Ian T. Baldwin†

Department of Molecular Ecology, Max Planck Institute for Chemical Ecology, Jena 07745, Germany

**Allopolyploid speciation occurs instantly when the genomes of different species combine to produce self-fertile offspring and has played a central role in the evolution of higher plants, but its consequences for adaptive responses are unknown. We compare herbivore-recognition and -resistance responses of the diploid species and putative ancestral parent *Nicotiana attenuata* with those of the two derived allopolyploid species *Nicotiana clevelandii* and *Nicotiana bigelovii*. *Manduca sexta* larvae attack all three species, and in *N. attenuata* attack is recognized when larval oral secretions are introduced to wounds during feeding, resulting in a jasmonate burst, a systemic amplification of trypsin inhibitor accumulation, and a release of volatile organic compounds, which function as a coordinated defense response that slows caterpillar growth and increases the probability of their being attacked. Most aspects of this recognition response are retained with modifications in one allotetraploid (*N. bigelovii*) but lost in the other (*N. clevelandii*). Differences between diploid and tetraploid species were apparent in delays (maximum 1 and 0.5 h, respectively) in the jasmonate burst, the elicitation of trypsin inhibitors and release of volatile organic compounds, and the constitutive levels of nicotine, trypsin inhibitors, diterpene glycosides, rutin, and caffeoylputrescine in the leaves. Resistance to *M. sexta* larvae attack was most strongly associated with diterpene glycosides, which were higher in the diploid than in the two allotetraploid species. Because *M. sexta* elicitors differentially regulate a large proportion of the *N. attenuata* transcriptome, we propose that these species are suited for the study of the evolution of adaptive responses requiring trans-activation mechanisms.**

**P**olyploid speciation plays a central role in the evolution of plants; as much as 70% of all angiosperm species are thought to have had polyploidization in their lineages (1). Allopolyploid speciation occurs instantly when the genomes of different species combine to produce self-fertile offspring and is a common speciation mechanism in particular taxa. Polyploid speciations are frequently associated with adaptive radiations, with the polyploid taxa exhibiting greater ability to survive under unfavorable conditions, perhaps because of increased heterozygosity (2). However the rapid changes in genomic architecture that result from the combination of new genomes are likely to wreak havoc with adaptations that rely on the trans-activation of many genes.

Evidence is rapidly accumulating that plants “recognize” attack from particular herbivore species and tailor their induced responses accordingly, and that a large fraction of a plant’s transcriptome is involved (3–11). However, little to nothing is known about how these complex responses are preserved or altered during polyploid speciation events. Some polyploids are more resistant to herbivore and pathogen attack than are closely related diploids (12), but in autopolyploid complexes of *Heuchera grossulariifolia* (Saxifragaceae) no clear associations of resistance with ploidy levels have been found (13–15). Consequently, plant polyploidy has not yet been integrated into our understanding of the evolution of insect–plant chemical inter-

actions, although for some plant–herbivore systems, the mechanisms of herbivore recognition and the adaptive tailoring of defense responses are understood in sufficient detail to compare responses across polyploid taxa.

In the genus *Nicotiana*, allopolyploidy has played an important role in speciation (16). *Nicotiana attenuata* (*Na*), a North American species that has been involved in two allopolyploid speciation events in the formation of two other North American species, *Nicotiana bigelovii* (*Nb*) and *Nicotiana clevelandii* (*Nc*), is also one of the best-studied species with regard to herbivore-recognition mechanisms (17). Cytologically, *Na* is a 12-paired species ( $n = 12$ ) and is thought to be the common ancestor during the amphidiploid speciation of *Nb* and *Nc* (both 24-paired species; Fig. 1), which Goodspeed (16) deduced from their close similarity in habit, leaf, inflorescence, and trichome morphology, and approximation of “*Drosera* scheme” pairing in  $F_1 Nb \times Na$  and  $Nc \times Na$  hybrids. The other parental line involved in the amphiploid origin of *Nb* and *Nc* is an early 12-paired *alatoid* race, which became the progenitor of the section *alatae* and is thought to be extinct (16). Goodspeed’s phylogenetic hypothesis has recently been tested with molecular techniques and found to be consistent with the available data (18).

Plant–herbivore interactions have been intensively investigated with ecological, chemical, and molecular approaches in the *Na*–*Manduca sexta* system (17). From this work, it is clear that *Na* recognizes feeding by the larvae of its specialist sphingid herbivore, *M. sexta*, as evidenced by *Manduca*-induced specific patterns of hormone signaling (JA, ethylene), secondary metabolite accumulation (responsible for both direct and indirect defenses), and gene transcript accumulation. These herbivore-induced responses are different from those induced by mechanical damage or exogenous applications of methyl jasmonate (MeJA), and are elicited by FACs in the OS of the larvae (8, 11, 19) (Fig. 1).

We compare these herbivore recognition and resistance responses observed in the diploid species *Na* with those observed in the two allopolyploid species *Nc* and *Nb*. While *Na* is found throughout the Great Basin Desert after fires in pinyon–juniper–sage habitats and north along the Sierras into California and Oregon, *Nb* is found in sandy washes along the California coast, and *Nc* grows in drier habitats throughout Baja California and

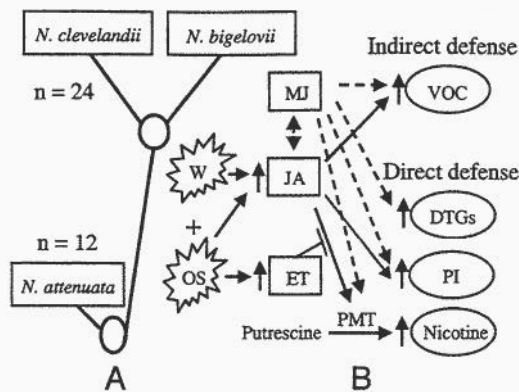
This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Chemical Communication in a Post-Genomic World,” held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: *Na*, *Nicotiana attenuata*; *Nb*, *Nicotiana bigelovii*; *Nc*, *Nicotiana clevelandii*; FACs, fatty acid–amino acid conjugates; OS, oral secretions and regurgitants; JA, jasmonate; MeJA, methyl jasmonate; DTG, diterpene glycoside; VOCs, volatile organic compounds; TrypPI, trypsin inhibitor.

\*Present address: Institute of Applied Entomology, Zhejiang University, Hangzhou 310029, China.

†To whom correspondence should be addressed. E-mail: baldwin@ice.mpg.de.

© 2003 by The National Academy of Sciences of the USA



**Fig. 1.** (A) The two tetraploid ( $n = 24$ ) species *Nb* and *Nc* are thought to have arisen from an allopolyploid speciation of an ancestral diploid ( $n = 12$ ) *Na* that hybridized with an extinct diploid *alatooid* species as proposed by Goodspeed (16) and confirmed by Chase et al. (18). (B) Attack by *Manduca sexta* larvae is recognized by *Na* when fatty acid–amino acid conjugates (FACs) in larval oral secretions and regurgitants (OS) are introduced into plant wounds (W) during feeding, resulting in signal crosstalk between jasmonate (JA)- and ethylene (ET)-mediated pathways and the elicitation of direct [nicotine, protease inhibitors (PI), and diterpene glycosides (DTGs)] and indirect [volatile organic compounds (VOCs)] defenses. MJ, methyl jasmonate; PMT, putrescine *N*-methyltransferase.

southern California (16). Two decades of fieldwork has established that *M. sexta* larvae are one of the three most abundant and damaging herbivores found on *Na*. Much less has been published on the herbivore communities of the two tetraploids; however, *M. sexta* larvae have been found on both species in nature (I.T.B., unpublished observations). Given that the JA burst in response to *M. sexta* OS applications to mechanical wounds is the signature of herbivore recognition that likely organizes much of the tailoring of the defense responses, we determine whether the three species differ in the timing and magnitude of the JA burst. Because the JA burst occurs locally and is not strongly influenced by leaf ontogeny (20), we measure it in a single, standardized leaf node in all three species. In addition to wounding, OS applications, and herbivore attack, JA applications are also known to elicit the accumulation of herbivore-induced defense metabolites in *Na*, such as nicotine (9), protease inhibitors (21, 22), VOCs (19), caffeoylputrescine, chlorogenic acid, and DTGs (23). These defense responses exhibit a mixture of systemic and local responses, and we measured them in three phyllotactically adjacent leaves after elicitation, to characterize both systemic and localized elicitation in all three species. Because the indirect defense (VOC release) is a whole-plant response, we measure whole-plant releases. Exogenous application of the methyl ester of JA, MeJA, provides a convenient and reproducible elicitor of insect resistance, and we examine MeJA-elicited resistance to *M. sexta* attack.

## Materials and Methods

**Plant Growth.** *Na* seeds originated from a population in Utah (24); *Nc* and *Nb* v. *bigelovii* seeds were kindly supplied by Verne A. Sisson (Oxford Tobacco Research Station, Oxford, NC) and originated from collections made by H. Goodspeed (16). Seeds were sterilized and germinated on agar (for smoke treatment of *Na* seeds see ref. 25) and after 10 days of growth, planted into soil in Teku pots (Waalwijk, The Netherlands) and once established, transferred to 1-liter pots in soil and grown in the glasshouse at 26–28°C, under 16 h of light supplemented by Philips Sun-T Agro 400- or 600-W sodium lights. Plants in the same stage of rosette growth, 2 weeks after their transfer to 1-liter pots were used in all experiments.

**Plant Treatment.** Plants were treated with 150  $\mu\text{g}$  of MeJA in 20  $\mu\text{l}$  of lanolin paste (19) applied to two leaves: at nodes 0 (source–sink transition leaf) and one node older (node 1). Controls (lanolin) were similarly treated with 20  $\mu\text{l}$  of pure lanolin. For *M. sexta* OS-treated plants, leaves at the same two nodal positions per plant were damaged by rolling a fabric pattern wheel over the leaf surface to create a standardized mechanical wound, and 20  $\mu\text{l}$  of OS [diluted 1:10 (vol/vol) with water] from fourth–fifth instar larvae was added to the puncture wounds on each leaf. Controls (water) were wounded and treated with 20  $\mu\text{l}$  of deionized water. Unmanipulated plants (controls) were included in each experiment.

**Comparison of Induced Secondary Metabolites. JA burst.** Thirty-six plants of each species were selected and randomly assigned to two treatment groups: OS (20  $\mu\text{l}$  of OS) or water (20  $\mu\text{l}$  of deionized  $\text{H}_2\text{O}$ ) were added to the lamina of node 1 leaves immediately after three rows of puncture wounds were created with a fabric pattern wheel. The treated leaves were harvested at 0, 0.5, 1.0, 1.5, 3, and 8 h after wounding and treatment, and JA was extracted (three plants per treatment per harvest) for analysis by GC–MS with a doubly labeled internal standard ([1,2- $^{13}\text{C}$ ]JA) as described in ref. 26.

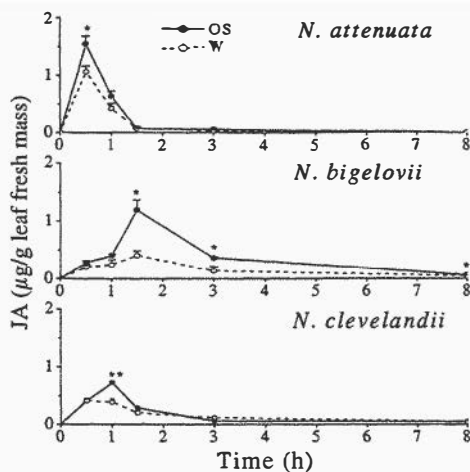
**Trypsin Inhibitor (TrypPI).** Plants of each species were randomly assigned to the five treatments (three to five plants per treatment and harvest). Leaves at nodes –1, 1, and 2 were harvested (at 1300 hours) at 1, 2, 4, 6, and 8 days after treatment for the lanolin and MeJA treatments. For OS, water, and control treatments, plants were harvested 4 days after treatment. TrypPI concentration was measured as described in ref. 21 and expressed as nmol per mg of protein.

**Nonvolatile secondary metabolites.** Plants of each species were randomly assigned to MeJA, lanolin, OS, water, and control treatments, each with three to six replicates, and harvested as described for TrypPI measures. Leaf extracts were prepared for analysis of nicotine, rutin, caffeoylputrescine, chlorogenic acid, and DTGs by HPLC as described in ref. 23.

**Volatiles.** Sixteen plants of each species were randomly assigned to four treatment groups (four replicates each): MeJA, lanolin, OS, and water. Plants were individually placed in 50-liter glass chambers, and the VOCs released were trapped on super Q (Alltech Associates) traps for 6 h, 24 h after elicitation (the time of maximum release after a single elicitation) and measured by GC–MS as described in ref. 19. Germacrene A was confirmed by mass spectra, a diagnostic thermal Cope-rearrangement to  $\beta$ -elemene (27) and retention time of an authentic standard from a liverwort (*Frullania macrocephalum*) extract (kindly supplied by Jan-Willem de Kraker, Wageningen University, Wageningen, The Netherlands). VOCs were expressed as percentages of peak areas relative to the internal standard, tetralin, per 6 h of trapping per plant.

**Herbivory Experiment.** Forty plants were randomly assigned to two treatment groups: MeJA and lanolin. Freshly hatched *M. sexta* L. (Lepidoptera: Sphingidae) larvae (eggs from North Carolina State University Insectary, Raleigh) were placed individually on the node –1 leaf of each plant that had been treated with MeJA or pure lanolin 4 days earlier. Larval mass was measured (to 0.1 mg) on the second, fourth, and sixth day after the start of the experiment.

**Statistical Analysis.** TrypPI activity and nicotine and rutin data were log transformed, DTG and caffeoylputrescine values were square root transformed, and chlorogenic acid values were inversion transformed before analysis to meet requirements of normality. Differences in JA burst, OS-induced TrypPI, and VOCs were determined by *t* tests. All other data were analyzed by multivariate ANOVA (MANOVA). If the MANOVA anal-



**Fig. 2.** JA concentrations ( $\pm 1$  SE) in *Na*, *Nb*, and *Nc* leaves that were wounded with a fabric pattern wheel and the resulting puncture wounds were immediately treated with 20  $\mu$ l of either *M. sexta* OS (●) or deionized water (W, ○) at time 0. Asterisks indicate level of significant differences between members of a pair (\*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ).

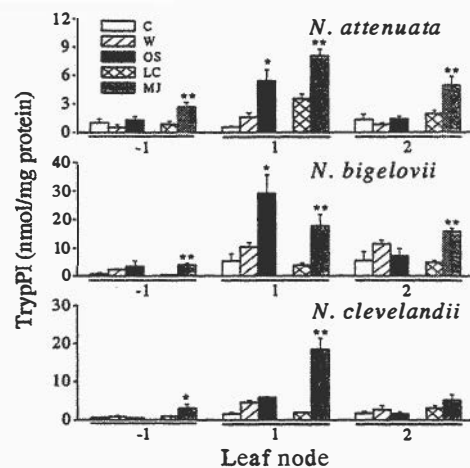
ysis was significant ( $P < 0.05$ ), univariate ANOVAs for the individual effects and Fisher least significant difference posthoc tests to detect significant differences between groups were conducted. Data were analyzed with STATVIEW (SAS Institute, Cary, NC), and the results are available in *Statistical Analysis*, which is published as supporting information on the PNAS web site, www.pnas.org.

## Results

**OS-Elicited JA Burst.** Application of *M. sexta* OS to puncture wounds in *Na* leaves resulted in a transient JA burst (attaining values that were 30% higher than those of wounded and water-treated leaves), which reached maximum values at 30 min and waned to control levels by 90 min (Fig. 2), as previously described (8, 28). Surprisingly, the JA burst was significantly attenuated in *Nc*, in which maximum values measured (at 1 h) were half of those observed in *Na* and *Nb*. The JA burst observed in *Nb*, although comparable in magnitude to that of *Na*, was delayed even longer (by 1 h) than that observed in *Nc* (delay of 0.5 h in comparison to that of *Na*). Moreover, the JA burst in *Nb*, in comparison to those of *Na* or *Nc*, did not wane rapidly, and significant differences were still found at the 8-h harvest (Fig. 2).

**TrypPI.** Whereas constitutive levels of TrypPI in untreated control plants (C) did not differ among the three species, the elicited levels did, and averaged across all elicitation treatments, *Nb* had the highest TrypPI concentrations, followed by *Nc*, and the lowest in *Na*. (Fig. 3). OS significantly amplified wound-induced increases of TrypPI concentrations in the treated leaf in both *Na* (3.4-fold) and *Nb* (2.8-fold), but not in *Nc* (1.3-fold) (Fig. 3). In contrast, treatment with MeJA resulted in long-lasting (>8 days; Fig. 8, which is published as supporting information on the PNAS web site) significant local and systemic increases in TrypPI activity in all three species in which the absolute increase in the treated leaf (node 1) was always greater than the systemic response and the systemic response in older sink leaves (node 2) was greater than those in younger source leaves (node -1) (Fig. 3).

**Nicotine.** MeJA treatment elicited significant long-lasting increases in nicotine levels in all three species, with the largest increase observed in leaves at node -1 in *Na* and *Nb*, but in *Nc* leaves, no consistent ranking was found (Figs. 9 and 10, which are

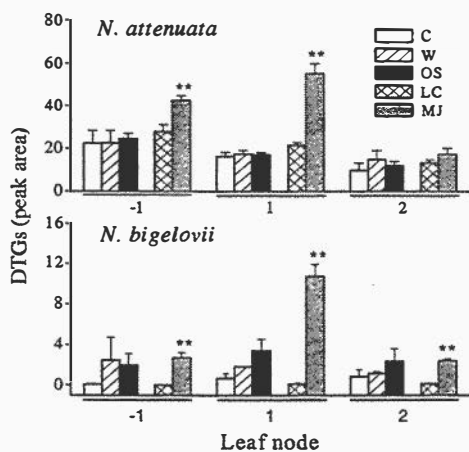


**Fig. 3.** Mean ( $\pm 1$  SE) TrypPI concentrations of *Na*, *Nb*, and *Nc* leaves growing at nodes -1, 1, and 2, 4 days after leaves at node 0 and 1 were treated with 20  $\mu$ l of lanolin containing 150  $\mu$ g of MeJA (MJ) or with 20  $\mu$ l of pure lanolin (LC), or were wounded and treated with 40  $\mu$ l of *M. sexta* OS or 40  $\mu$ l of deionized water (W) or left unwounded and untreated (C). Asterisks indicate level of significant differences between members of a treatment and control pair (LC vs. MJ and OS vs. W: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ).

published as supporting information on the PNAS web site). Quantities measured in the three species followed a pattern similar to that observed for TrypPIs: *Nb* had the highest (5.7- and 2.5-fold higher than those in *Na* and *Nc*, respectively) and *Nc* had intermediate values (2.3-fold higher than those of *Na*) as averaged across the five harvests from all treatments. OS treatment of puncture wounds did not elicit increases that were significantly higher than those elicited by wounding and water treatments in all three species (Fig. 9), despite the higher elicitation of JA by OS treatment (Fig. 2). As previously described (9, 29), the suppression of wound-induced nicotine increases reflects an OS-elicited ethylene burst that suppresses transcripts of the rate-limiting enzyme in nicotine biosynthesis (putrescine *N*-methyltransferase; Fig. 1).

**DTGs.** DTG levels were highest in *Na*, intermediate in *Nb*, and not detectable in *Nc* (Fig. 4). DTGs levels decreased with increasing leaf age in *Na* but not in *Nb*. MeJA elicitation significantly increased DTG levels in both *Na* and *Nb* both locally and systemically, but the strongest elicitation was observed in the treated leaf (Fig. 4). Wounding and treatment with OS or water did not elicit DTG increases in either *Na* or *Nb*.

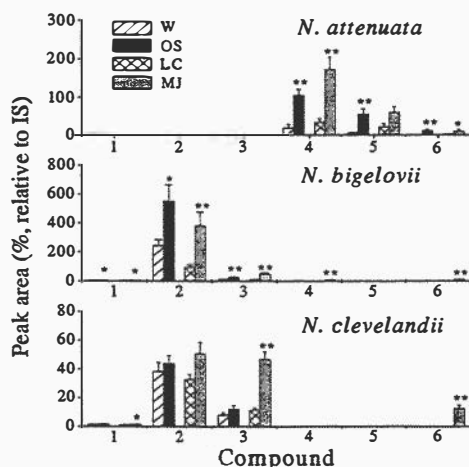
**VOCs.** Six compounds dominated the headspace of MeJA- or OS-treated plants: *cis*- $\beta$ -ocimene, *trans*- $\beta$ -ocimene, linalool,  $\alpha$ -bergamotene, germacrene A, and *cis*-jasmane. Qualitative and quantitative differences were found among the three species. Four compounds (not *trans*- $\beta$ -ocimene and linalool) were found in the headspace of *Na*, five (not germacrene A) in *Nb*, and four (not  $\alpha$ -bergamotene and germacrene A) in *Nc* (Fig. 5). In general, the VOC profile composition was more similar between the two tetraploid species than between the tetraploid and diploid species. MeJA elicitation dramatically increased the release of  $\alpha$ -bergamotene and *cis*-jasmane but not germacrene A and *cis*- $\beta$ -ocimene in *Na*; *cis*- $\beta$ -ocimene, *trans*- $\beta$ -ocimene, linalool, *cis*-jasmane (which was not detectable before elicitation), and  $\alpha$ -bergamotene in *Nb*; and *cis*- $\beta$ -ocimene (but not *trans*- $\beta$ -ocimene), linalool, and *cis*-jasmane in *Nc*. OS treatments also increased the amounts of VOCs released in both *Na* and *Nb*, but surprisingly, not in *Nc*. In *Na* and *Nb*, OS treatment increased the release of three, albeit different, VOCs:  $\alpha$ -bergamotene, ger-



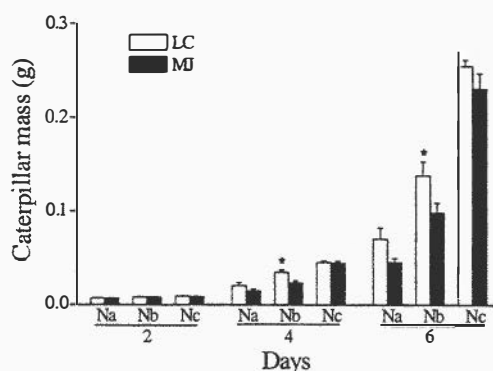
**Fig. 4.** Mean (+1 SE) DTG peak areas of *Na* and *Nb* leaves growing at nodes -1, 1, and 2, 4 days after leaves at node 0 and 1 were treated with either 20  $\mu$ l of lanolin containing 150  $\mu$ g of MeJA (MJ) or 20  $\mu$ l of pure lanolin (LC), wounded and treated with 40  $\mu$ l of *M. sexta* OS or 40  $\mu$ l of deionized water (W), or left unwounded and untreated (C). DTGs were not detected in extracts of *Nc* leaves. Asterisks indicate level of significant differences between members of a treatment and control pair (LC vs. MJ and OS vs. W: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ).

macrene A, and *cis*-jasmone; and *cis*- $\beta$ -ocimene, *trans*- $\beta$ -ocimene, and linalool, respectively (Fig. 5).

**Phenolics.** MeJA treatment (but not OS treatment) significantly increased caffeoylputrescine in all species, but the increases (25.6-fold) were larger in *Nc* than those in *Na* (10.1-fold) and *Nb* (2.5-fold). Elicited changes in chlorogenic acid and rutin were complex, with both increases and decreases observed after the different treatments (Table 1, which is published as supporting information on the PNAS web site). Chlorogenic acid concentrations were significantly reduced by OS and W treatments in *Na* and *Nb*, but in *Nc*, only W treatment resulted in significant decreases. MeJA treatment significantly increased levels in older leaves at node 2 in *Na* and *Nc* but decreased levels in *Nb* in all



**Fig. 5.** Mean (+1 SE) peak areas of VOCs in the headspace of *Na*, *Nb*, and *Nc* plants sampled for 6 h, starting 24 h after treatment with 20  $\mu$ l of lanolin containing 150  $\mu$ g of MeJA (MJ), 20  $\mu$ l of pure lanolin (LC), or wounding and treatment with 40  $\mu$ l of *M. sexta* OS or 40  $\mu$ l of water (W) to leaves at nodes 0 and 1. Numbers identify compounds: 1, *cis*- $\beta$ -ocimene; 2, *trans*- $\beta$ -ocimene; 3, linalool; 4, *cis*- $\alpha$ -bergamotene; 5, germacrene A; and 6, *cis*-jasmone. Asterisks indicate level of significant differences between members of a treatment and control pair (LC vs. MJ and OS vs. W: \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ).



**Fig. 6.** Mean (+1 SE) mass of 12–20 replicate *M. sexta* larvae fed on individual *Na*, *Nb*, and *Nc* plants, 2, 4, and 6 days after plants were treated with 20  $\mu$ l of lanolin containing 150  $\mu$ g of MeJA (MJ) or 20  $\mu$ l of pure lanolin (LC) to leaves at nodes 0 and 1. Asterisks indicate significant differences between members of a pair (\*,  $P < 0.05$ ).

leaves. MeJA treatment decreased rutin levels in all species but in different leaves (at nodes -1 in *Na* and *Nb*, but 1 in *Nc*). OS treatment decreased rutin levels only in the treated leaf of *Nc*.

**Herbivory Experiment.** Weight gain was greatest when larvae fed on *Nc*, followed by those fed *Nb*, and the least on *Na* (Fig. 6). By day 6, the masses of caterpillars fed on *Na* and *Nb* were only 23.62% and 48.75% of those feeding on *Nc*, respectively. Interestingly, differences in larval mass between caterpillars fed on MeJA- and lanolin-treated plants were significantly reduced only in *Nb*-fed larvae at days 4 and 6, although the masses of caterpillars fed both MeJA-treated *Na* and *Nc* plants were consistently lower than those fed on lanolin-treated plants from the respective species (Fig. 6).

## Discussion

Adaptive phenotypic responses increases the “fit” between organisms and their environment by altering the expression of a large number of genes in response to environmental signals. Different stress factors, such as insect herbivores, pathogens, and abiotic factors, elicit physiological, biochemical, and morphological changes in plants, likely as a result of crosstalk among a large number of signal transduction pathways, which include JA, ethylene, abscisic acid, and salicylates (30). Specific combinations of signals are thought to provide “signature” sets thought to activate an appropriate response to a specific stress. However, little is known whether crosstalk among signaling pathways results in adaptive responses, but work on plant–herbivore interactions provides some of the best examples to date.

When herbivores attack plants, they cause wounding, but a plant’s response to herbivore attack in many cases cannot be mimicked by mechanical wounding or simple JA applications, which are thought to mediate many wound responses (30–32). Several different types of elicitors in the OS of herbivorous insects have been reported to alter a plant’s wound response, including enzymatic elicitors such as  $\beta$ -glucosidase (33) and glucose oxidase (34) as well as FACs (8, 35). *Manduca* larvae contain at least 8 FACs in their OS that are necessary and sufficient for the JA burst and VOC release (8) as well as the amplification of TrypPIs (A. Roda, A. Steppuhn, and I.T.B., unpublished results) observed in *Manduca*-attacked *Na* plants. Moreover, the two most abundant FACs in *M. sexta* OS are responsible for 64% of the up-regulated (of 67) and 49% of the down-regulated (of 78) genes that are differentially regulated when *M. sexta* OS is added to plant wounds (11). *Manduca* OS is also known to elicit an ethylene burst, which reduces wound-induced nicotine accumulation by down-regulating transcripts of

putrescine *N*-methyltransferase, which catalyzes the key regulatory step in nicotine biosynthesis (9, 29).

Although much remains unknown about the functional significance of these complex alterations, evidence is accumulating that the herbivore-specific increases in TrypPIs and VOCs, as well as the down-regulation of the wound-induced nicotine production, represent adaptive tailoring of the plant's defense response against *Manduca* attack. The VOC release functions as a potent indirect defense in nature, by attracting the generalist predator *Geocoris pallens* to feeding larvae (36). This voracious predator is size-selective, preferentially attacking eggs and larvae in the first three instars (37), and the up-regulation of TrypPIs by *Manduca* attack slows the growth of larvae (22), keeping them in stages that are more vulnerable to the predator. In addition, nicotine, which is sequestered by *M. sexta* larvae through dietary intake, negatively affects the performance of the parasitoids of *M. sexta* (38) and hence is coopted for the defense of the herbivore. Hence when *Na* is attacked by this nicotine-tolerant herbivore, it will likely realize a fitness benefit by the coordinated up-regulation of the predator-attracting VOC release and the amplification of wound-induced TrypPI production, while suppressing wound-induced nicotine production. A recently discovered natural mutant of *Na* that lacks the ability to produce TrypPIs is also deficient in herbivore-induced VOC release (22). Moreover, nicotine production is costly, requiring 8% of whole-plant nitrogen, an investment that cannot be recouped by metabolism, and is associated with diminished intraspecific competitive abilities for soil nitrogen (17). Hence, when *Na* is attacked by this nicotine-tolerant herbivore, it will likely realize a fitness benefit from suppressing its induced nicotine production when growing in competition with conspecifics, as it commonly does as a result of its germination behavior that synchronizes growth with the postfire environment (17).

To compare the ability of the diploid species, *Na*, to recognize attack from *M. sexta* larvae with that of the two allopolyploid species, *Nb* and *Nc*, we examined the timing of the JA burst and the subsequently elicited changes in secondary metabolites when *M. sexta* OS was applied to mechanically generated wounds on leaves. From these experiments, it was clear that whereas a statistically significant JA burst occurred in all three species, the *Nc*'s JA increase was only half that observed in *Na* and *Nb* (Fig. 2). *Nc*'s attenuated JA burst was also associated with a lack of OS-elicited VOC release (Fig. 5), TrypPI increase (Fig. 3), and reduction in chlorogenic acid contents (Table 1), responses that were clearly preserved in *Nb*. The lack of OS-elicited responses in *Nc* was not due to an inability of this species to respond because MeJA elicitation resulted in VOC, TrypPI, and chlorogenic acid increases. These results demonstrate that *Na*'s herbivore recognition mechanism has partly been conserved during allopolyploid speciation in *Nb*, but lost in *Nc*, a result consistent with Goodspeed's phylogenetic hypothesis that *Nb* is more closely related to *Na* than *Nc* is (16).

Not all OS-elicited responses appear to be lost in *Nc*, however. The down-regulation of wound-induced nicotine production by OS appears to be retained in all species. The observation that *M. sexta* OS elicitation did not result in higher nicotine levels compared with water treatments of mechanical wounds (Fig. 9) despite a higher JA pool induced (Fig. 2), as has been observed in another diploid *Nicotiana* species [*Nicotiana sylvestris* (39)] suggests that OS-elicited ethylene signaling remains intact in all three species (9). Similarly, some JA-elicited responses are not elicited by OS treatment in all three species. DTGs (Fig. 4), caffeoylputrescine, and rutin (Table 1) were elicited by MeJA treatment but not by OS treatment. For these metabolites, JA elicitation clearly recruits a signal cascade that is not activated by OS treatment.

In addition to the differences in response to OS, the tetraploids also differed in the amount and timing of certain elicited

metabolites. The rapid activation and waning of the JA burst observed in *Na* was both delayed and lasted significantly longer in *Nb*. As such, *Nb*'s wound-induced JA dynamics are more similar to those observed in *Arabidopsis*, tomato, and potato (26). In nature, *Na* grows in close association with sagebrush, *Artemisia tridentata* Nuttall subsp. *tridentata* (Asteraceae), which releases MeJA in high, allelopathically active quantities (40, 41), potentially sufficient to influence the defense responses in the neighboring *Na* plants (42). In contrast, neither *Nb* nor *Nc* is commonly found growing in close association with sagebrush, and the rapid endogenous JA dynamics observed in *Na* may allow it to distinguish endogenous from exogenously derived JA.

The three species also differed in constitutive levels of secondary metabolites in a manner consistent with a gene-dose effect. For example, both tetraploid species had significantly higher nicotine (8.31- and 2.72-fold higher than *Na*: Fig. 9) and TrypPI (9.18- and 2.63-fold higher than those in *Na*: Fig. 3) levels, and lower DTGs (Fig. 4), rutin, and caffeoylputrescine (Table 1) levels compared with the level observed in *Na*. Comparing the two tetraploid species, all of these metabolites, with the exception of caffeoylputrescine, were significantly higher in *Nb* than in *Nc*. Differences in ecological habit between the tetraploid and diploid species may select for high nicotine and TrypPI levels. Of the three species, only *Na* "chases" fires in ecological time by mass-germinating from long-lived seed banks with smoke-related germination cues. By timing its growth with the ephemeral but high-resource postfire environment, *Na* is commonly exposed to very strong intraspecific competition, which in turn likely selects for rapid growth (17). Because both constitutive and inducible protease inhibitor production (22) and nicotine production (24) are associated with growth and fitness reductions, it is possible that selecting habitats that place a premium on fast growth also selects for low constitutive defense levels. Neither *Nb* nor *Nc* times their germination with the postfire environment and therefore they may not be under similarly strong selection for rapid growth and competitive ability.

MeJA-elicitation altered the secondary metabolite profiles in all three species, and by comparing these changes with MeJA-elicited changes in *M. sexta* larval performance, we could infer their relative influence on larval performance. *M. sexta* larval growth was highest in *Nc*, followed by in *Nb*, and the lowest in *Na* (Fig. 6). Significant difference in larval mass between

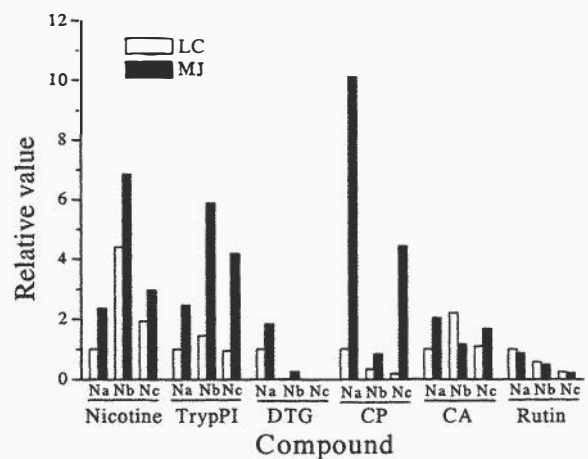


Fig. 7. Relative values of nicotine, TrypPI, DTGs, caffeoylputrescine (CP), chlorogenic acid (CA), and rutin in MeJA-treated (solid bars) and lanolin-treated (open bars) *Na*, *Nb*, and *Nc* plants compared with the corresponding chemical concentrations in lanolin-treated *Na*, which were arbitrarily assigned a value of 1.0.

caterpillars fed on MeJA- and lanolin-treated plants was found only in *Nb*, although caterpillars fed on MeJA-treated plants tended to have lower masses than those fed on lanolin-treated plants in *Na* and *Nc*. We standardized the MeJA-elicited metabolite changes to the uninduced levels found in *Na*, which was arbitrarily assigned a value of 1 (Fig. 7) to identify metabolites that correlated with the observed changes in *M. sexta* larval growth (Fig. 6). This analysis suggested that DTGs, which were highest in *Na*, intermediate in *Nb*, and not detectable in *Nc*, were most strongly associated with larval performance (Fig. 7). However, larval mass did not significantly differ between MeJA- and lanolin-treated *Na* plants, whereas DTGs did differ. Complex nonlinear interactions between other induced metabolites (nicotine and TrypPIs) and DTGs may influence the relationship with larval mass gain and DTG content. DTGs have been reported to inhibit the larval growth of tobacco budworm larvae, *Heliothis virescens* (43); however, conclusive evidence that DTGs are directly responsible for the observed effects will require direct manipulations of DTG production in plants. Interestingly, although nicotine (44) and TrypPI (22) have been reported to inhibit larval growth in *Na* and *N. sylvestris*, this analysis suggests that neither is as strongly correlated as DTGs are with *M. sexta* performance.

Although the ecological significance of the loss and retention of herbivore recognition abilities in the two allopolyploid species

remains unknown, this species complex may be an ideal system in which to study the retention and modification of trans-activated adaptive responses. The changes in secondary metabolites elicited by *Manduca* OS in *Na* are accompanied by a large-scale transcriptional change (6, 8, 10, 11), the majority of which can be elicited by only two FACs in *Manduca* OS (11). Given that these differentially regulated genes are likely dispersed throughout *Na*'s 12 pairs of chromosomes, a small number of FAC-regulated trans-active elements are likely responsible for herbivore recognition in this species. How these putative trans-active elements have retained their ability to respond to OS in *Nb*, while recruiting different cis-elements [as is suggested by the different spectrum of VOCs released by OS elicitation (Fig. 5)], but losing the ability in *Nc*, will likely provide important insights into the maintenance or modification of polygenic adaptive traits during allopolyploid speciation. Moreover, given that it is possible to hybridize the different North American *Nicotiana* species to create artificial tetraploid lines (16), these evolutionary hypotheses are eminently falsifiable.

We thank B. Krock for identification of germacrene A; B. Krock, C. von Dahl, J. Zavala, R. Halitschke, T. Kruegel, and C. McInerney for invaluable analytical support; two anonymous reviewers for substantially improving the manuscript; and the Max Planck Society and the Deutsche Forschungsgemeinschaft (SPP 1152) for funding.

- Masterson, J. (1994) *Science* **264**, 421–424.
- Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7719–7723.
- Korth, K. L. & Dixon, R. A. (1997) *Plant Physiol.* **115**, 1299–1305.
- Walling, L. L. (2000) *J. Plant Growth Regul.* **19**, 195–216.
- Reymond, P., Weber, H., Damond, M. & Farmer, E. E. (2000) *Plant Cell* **12**, 707–719.
- Hermesmeier, D., Schittko, U. & Baldwin, I. T. (2001) *Plant Physiol.* **125**, 683–700.
- Schittko, U., Hermesmeier, D. & Baldwin, I. T. (2001) *Plant Physiol.* **125**, 701–710.
- Halitschke, R., Schittko, U., Pohnert, G., Boland, W. & Baldwin, I. T. (2001) *Plant Physiol.* **125**, 711–717.
- Winz, R. A. & Baldwin, I. T. (2001) *Plant Physiol.* **125**, 2189–2202.
- Hui, D., Iqbal, J., Lehmann, K., Gase, K., Saluz, H. P. & Baldwin, I. T. (2003) *Plant Physiol.* **131**, 1877–1893.
- Halitschke, R., Gase, K., Hui, D., Schmidt, D. & Baldwin, I. T. (2003) *Plant Physiol.* **131**, 1894–1902.
- Schoen, D. J., Burdon, J. J. & Brown, A. H. D. (1992) *Theor. Appl. Genet.* **83**, 827–832.
- Thompson, J. N., Cunningham, B. M., Segraves, K. A., Althoff, D. M. & Wagner, D. (1997) *Am. Nat.* **150**, 731–743.
- Nuismer, S. L. & Thompson, J. N. (2001) *Proc. R. Soc. London Ser. B* **268**, 1937–1940.
- Janz, N. & Thompson, J. N. (2002) *Oecologia* **130**, 570–575.
- Goodspeed, T. H. (1954) *The Genus Nicotiana* (Chronica Botanica, Waltham, MA).
- Baldwin, I. T. (2001) *Plant Physiol.* **127**, 1449–1458.
- Chase, M. W., Knapp, S., Cox, A. V., Clarkson, J. J., Butsko, Y., Joseph, J., Savolainen, V. & Parokonny, A. S. (2003) *Ann. Bot.* **92**, 107–127.
- Halitschke, R., Kessler, A., Kahl, J., Lorenz, A. & Baldwin, I. T. (2000) *Oecologia* **124**, 408–417.
- Ohnmeiss, T. E., McCloud, E. S., Lynds, G. Y. & Baldwin, I. T. (1997) *New Phytol.* **137**, 441–452.
- Van Dam, N. M., Horn, M., Mares, M. & Baldwin, I. T. (2001) *J. Chem. Ecol.* **27**, 547–568.
- Glawe, G. A., Zavala, J. A., Kessler, A., Van Dam, N. M. & Baldwin, I. T. (2003) *Ecology* **84**, 79–90.
- Keinanen, M., Oldham, N. J. & Baldwin, I. T. (2001) *J. Agric. Food Chem.* **49**, 3553–3558.
- Baldwin, I. T. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8113–8118.
- Kruegel, T., Lim, M., Gase, K., Halitschke, R. & Baldwin, I. T. (2002) *Chemoecology* **12**.
- Park, J.-H., Halitschke, R., Kim, H. B., Baldwin, I. T., Feldmann, K. A. & Feyereisen, R. (2002) *Plant J.* **31**, 1–12.
- de Kraker, J.-W., Franssen, M. C. R., de Groot, A., König, W. A. & Bouwmeester, H. J. (1998) *Plant Physiol.* **117**, 1381–1392.
- Schittko, U., Preston, C. A. & Baldwin, I. T. (2000) *Planta* **210**, 343–346.
- Kahl, J., Siemens, D. H., Aerts, R. J., Gabler, R., Kuhnemann, F., Preston, C. A. & Baldwin, I. T. (2000) *Planta* **210**, 336–342.
- Kessler, A. & Baldwin, I. T. (2002) *Ann. Rev. Plant Biol.* **53**, 299–328.
- Baldwin, I. T. (1988) *Oecologia* **378**–381.
- Baldwin, I. T., Halitschke, R., Kessler, A. & Schittko, U. (2001) *Curr. Opin. Plant Biol.* **4**, 351–358.
- Mattiacci, L., Dicke, M. & Posthumus, M. A. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 2036–2040.
- Musser, R. O., Hum-Musser, S. M., Eichenseer, H., Peiffer, M., Ervin, G., Murphy, J. B. & Felton, G. W. (2002) *Nature* **416**, 599–600.
- Turlings, T. C. J., Alborn, H. T., Loughrin, J. H. & Tumlinson, J. H. (2000) *J. Chem. Ecol.* **26**, 189–202.
- Kessler, A. & Baldwin, I. T. (2001) *Science* **291**, 2141–2144.
- Kessler, A. & Baldwin, I. T. (2002) *Ecology* **83**, 2346–2354.
- Barbosa, P., Gross, P. & Kemper, J. (1991) *Ecology* **72**, 1567–1575.
- McCloud, E. S. & Baldwin, I. T. (1997) *Planta* **203**, 430–435.
- Preston, C. A., Laue, G. & Baldwin, I. T. (2001) *Biochem. Syst. Ecol.* **29**, 1007–1023.
- Preston, C. A., Betts, H. & Baldwin, I. T. (2002) *J. Chem. Ecol.* **28**, 2343–2369.
- Karban, R., Baldwin, I. T., Baxter, K. J., Laue, G. & Felton, G. W. (2000) *Oecologia* **125**, 66–71.
- Snook, M. E., Johnson, A. W., Severson, R. F., Teng, Q., White, R. A., Jr., Sisson, V. A. & Jackson, D. M. (1997) *J. Agric. Food Chem.* **45**, 2299–2308.
- Voelckel, C., Kruegel, T., Gase, K., Heidrich, N., van Dam, N. M., Winz, R. & Baldwin, I. T. (2001) *Chemoecology* **11**, 121–126.

# Evolutionary dynamics of an *Arabidopsis* insect resistance quantitative trait locus

Juergen Kroymann, Susanne Donnerhacke, Domenica Schnabelrauch, and Thomas Mitchell-Olds\*

Department of Genetics and Evolution, Max Planck Institute for Chemical Ecology, Winzerlaer Strasse 10, 07745 Jena, Germany

**Glucosinolate profiles differ among *Arabidopsis thaliana* ecotypes, caused by the composition of alleles at several glucosinolate biosynthetic loci. One of these, *GS-Elong*, harbors a family of methylthioalkylmalate synthase (MAM) genes that determine the side chain length of aliphatic glucosinolate structures. Fine mapping reveals that *GS-Elong* constitutes an insect resistance quantitative trait locus, caused by variation in glucosinolate profiles conferred by polymorphism of *MAM* alleles in this region. A sequence survey of randomly chosen ecotypes indicates that *GS-Elong* is highly variable among *A. thaliana* ecotypes: indel polymorphisms are frequent, as well as gene conversion events between gene copies arranged in tandem. Furthermore, statistical methods of molecular population genetics suggest that one of the genes, *MAM2*, is subject to balancing selection. This may be caused by ecological tradeoffs, i.e., by contrasting physiological effects of glucosinolates on generalist vs. specialist insects.**

**R**esistance to insect herbivores is genetically variable in many plant populations (1). Ecological and evolutionary interactions between host plants and their insect enemies are often mediated by secondary metabolites. Molecular genetics allows us to clone and characterize genes controlling secondary metabolism and insect resistance. When these genes are identified, molecular population genetics provides statistical tests for neutral evolution or natural selection, thus elucidating the evolutionary forces responsible for genetic variation in ecologically important traits. Here we show that a gene family involved in the synthesis of secondary metabolites controls insect resistance, shows complex molecular variation, and maintains excess amino acid polymorphism because of balancing natural selection.

Glucosinolates are amino acid-derived plant secondary compounds present in the Capparales (2, 3). Their biosynthesis occurs in three independent stages, chain elongation of the amino acid, formation of the core structure (consisting of a  $\beta$ -thioglucose moiety and a sulfonated oxime), and side chain modifications. Both side chain elongation and modification contribute to the variation of glucosinolate structures, and >30 different glucosinolates have been identified in the model plant *Arabidopsis thaliana* (4, 5). However, *A. thaliana* ecotypes vary extensively both in their glucosinolate composition and quantity (4). Genetically, most of this natural variation can be explained by the combination of alleles at five genetic loci within the *A. thaliana* genome (4). Among these, *GS-Elong* has a central role as it controls side chain length of methionine-derived glucosinolates, thereby determining potential for subsequent modification steps (6, 7). *GS-Elong* consists of a small gene family encoding methylthioalkylmalate synthase (MAM) enzymes responsible for carbon chain elongation in glucosinolate biosynthesis. Leaves of the Columbia (Col-0) ecotype of *A. thaliana* contain predominantly glucosinolates with four methylene groups ( $C_4$ ) in their basic side chain. In contrast, Landsberg *erecta* (Ler-0) leaves accumulate primarily glucosinolates with three methylene groups ( $C_3$ ) in their basic side chain. Ecologically, diversification in glucosinolate profiles may represent an

adaptation to challenges by microbial pathogens and herbivorous insects.

## Materials and Methods

**Plant Material.** *A. thaliana* seeds were obtained from the Nottingham *Arabidopsis* Stock Centre. Ecotypes were chosen without regard to glucosinolate phenotype from throughout the species' native range. Ecotypes were grown as single plants in  $5 \times 5$  cm<sup>2</sup> pots filled with a 1:3 vermiculite/standard soil (Einheitserdenwerk, Fröndenber, Germany) mix under 11.5 h/12.5 h light/dark cycles. For glucosinolate analysis and insect feeding screens, seeds from near-isogenic lines were planted in damp Scotts Redi Earth at a density of 337 plants per m<sup>2</sup> in 96-celled flats, covered with clear plastic grow domes and stratified for 48 h at 4°C in the dark. Afterward, flats were moved to ventilated growth rooms with constant air flow and  $\approx 40\%$  humidity and 23°C. Plants were grown at a distance of 30 cm from fluorescent light banks with four bulbs of cool white and four bulbs of wide spectrum lights at a 14 h light/10 h dark photoperiod. Seeds germinated in 2–3 days. Grow domes were removed after 5 days under lights and plants were fertilized once with 1 ml of Scotts Peters Professional Peat Lite Special 20N:10P:20K with trace elements and 1 liter per flat, added to the bottom of the tray. All flats were daily rotated within shelves, between shelves, and end to end to compensate for slight variations in temperature and light intensity in the growth room. For growth rate experiments, five seeds per line were planted in  $7.5 \times 7.5$  cm<sup>2</sup> pots, and 18 pots per flat in a random arrangement. Pots were daily rotated between flats. Otherwise, conditions were identical to those for glucosinolate analysis and insect feeding screens.

**Herbivory Screens.** Diamondback moth (*Plutella xylostella*) eggs were obtained from New York State Agricultural Experiment Station (Geneva), and a colony was maintained at University of Montana (Missoula). Insects were raised on artificial diet according to published procedures (8). Beet armyworm (*Spodoptera exigua*) eggs were obtained from Benzon Research (Carlisle, PA), and larvae were raised on an artificial diet from Southland Products (Lake Village, AR). Plants were infested 18 days after transfer to the growth room with one 72-h-old *P. xylostella* or one 94-h-old *S. exigua* larva per plant. Plant damage was scored after 48 h.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: MAM, methylthioalkylmalate synthase; Col, Columbia ecotype; Ler, Landsberg *erecta*; QTL, quantitative trait locus.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AJ486882–AJ486953).

\*To whom correspondence should be addressed. E-mail: tmo@ice.mpg.de.

© 2003 by The National Academy of Sciences of the USA



**Growth Rate Measurement.** Plants were harvested 16 days after germination. Plants for each pot (i.e., for each line) were pooled, dried in a hot air oven at 45°C, and then weighed. For each line, there were 14 or 15 replicates.

**Glucosinolate Analysis.** Leaves were harvested 18 days after plants were moved to the growth rooms, i.e., 15–16 days after germination. Glucosinolate extraction and high-pressure liquid chromatography were carried out as described (7).

**Molecular Methods.** Total leaf DNA was extracted with Qiagen (Hilden, Germany) genomic-tips 100/G following the manufacturer's instructions. Total leaf RNA was isolated with Trizol (Life Technologies). First-strand cDNA was synthesized from 1 µg total RNA following ref. 9. Details on PCRs for *MYB37* (=at5g23000), *MAM1* (=at5g23010), *MAM2*, and *MAM-L* (=at5g23020) genes are given in the supporting information, which is published on the PNAS web site, www.pnas.org.

PCR products were gel purified with QIAquick (Qiagen), and cloned into pCRII TOPO TA, pCR2.1 TOPO TA, or pCR-XL-TOPO vectors (all from Invitrogen). Plasmids were isolated according to standard procedures.

Sequences were obtained either directly from gel-purified PCR products, or from recombinant plasmids. In the latter case, several independent clones were analyzed with vector- and insert-specific primers. Sequencing was done on ABI 377 or 3700 DNA sequencers with Big Dye Terminators (Applied Biosystems). Sequence of the entire *Ler-0* *GS-Elong* region was assembled from two overlapping *Ler-0* bacterial artificial chromosomes according to standard methods. Assembly and comparison of DNA sequence data were carried out with DNASTAR (DNASTAR, Madison, WI).

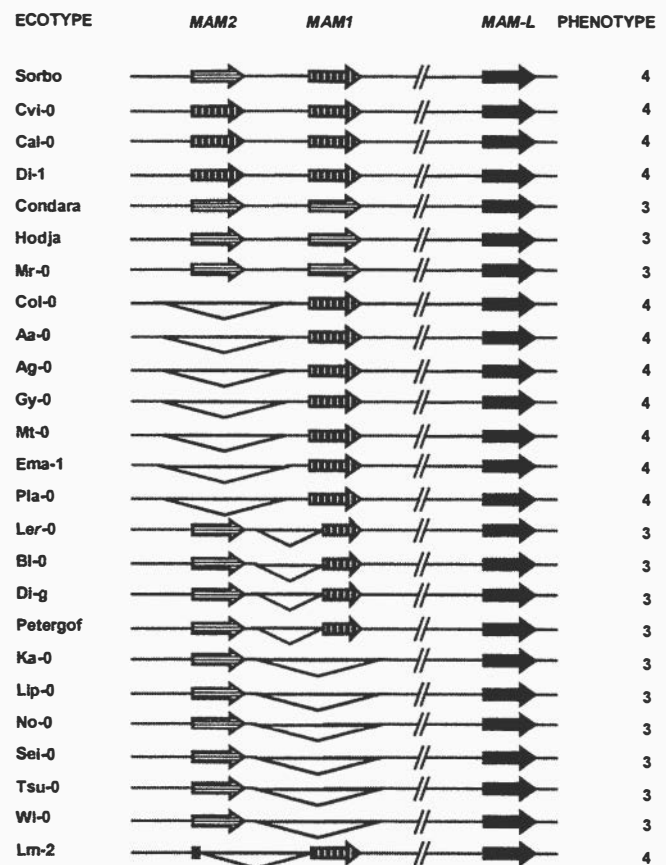
Genotyping of recombinant inbred lines followed ref. 7, including additional markers (see supporting information). Resulting PCR products were purified and analyzed by sequencing, except for those amplified with recMS3f/recMS3r, which were separated on 6% Metaphor (Biozym, Germany) agarose gels.

**Statistical Methods.** Calculation of Tajima's *D*, coalescent simulations, and the McDonald–Kreitman test were carried out with DNASP3.84 (10). A neighbor-joining tree (11) of *MAM1* and *MAM2* alleles was constructed with TREECON (12) following ref. 13. Reliability of the branching order was estimated by bootstrapping (100 replicates; ref. 14).

For measurements of glucosinolates and resistance to *S. exigua* and *P. xylostella*, we obtained least square means for each near isogenic line, analyzing a randomized complete blocks design with SYSTAT (SPSS, Chicago). Quantitative trait locus (QTL) mapping with line means used a fixed-effects ANOVA with markers in the *GS-Elong* interval cross-classified with the *AlkOhp* marker (15), which was also examined in this experiment (data not shown). Previous studies have shown a QTL influencing glucosinolate concentration in the *GS-Elong* region (16); hence, in these new experiments, we used a standard *P* = 0.05 significance threshold based on this *a priori* expectation.

## Results

**Complex Organization of *GS-Elong* in *A. thaliana* Ecotypes.** The organization of *GS-Elong* is highly variable in *A. thaliana*, and indel polymorphisms of large regions are common (Fig. 1). In addition to a *MAM-L* (*MAM*-like) gene present in all ecotypes investigated, *GS-Elong* may harbor two further loci, *MAM1* and *MAM2*. Some ecotypes contain both loci, whereas others possess either *MAM2* or *MAM1*, or, as the *Ler-0* ecotype, a *MAM2* in addition to a truncated, nonfunctional *MAM1* locus. In the case of the *Lm-2* ecotype, the 5' part of a *MAM2*-like sequence is fused to the 3' part of a *MAM1*-like sequence, which may have been caused by deletion of the intervening region. Reciprocal



**Fig. 1.** *GS-Elong* region in *A. thaliana* ecotypes. Maximal levels of divergence between *MAM1* and *MAM2* (nearly 5%) occur in the Sorbo ecotype, which likely represents the ancestral gene arrangement. Genes are patterned as most similar to Sorbo *MAM1* (vertically patterned) or Sorbo *MAM2* (horizontally patterned). Triangles indicate large deletions. Predominant glucosinolate side chain length is indicated on the right: glucosinolates with four methylene groups (4, e.g., Col-0) or three methylene groups (3, e.g., Ler-0). Notice that synthesis of  $C_4$  glucosinolates is completely associated with occurrence of a full-length Sorbo-like *MAM1* gene. Otherwise,  $C_3$  glucosinolates are synthesized.

deletions of *MAM2* in Col-0 and *MAM1* in Ler-0 have the consequence that the remaining *MAM1* and *MAM2* genes segregate as alleles of each other, although they are phylogenetically paralogs resulting from a gene duplication event (Fig. 1).

**Gene Conversion Between *MAM1* and *MAM2* Loci.** An additional layer of complexity is added by apparent interlocus gene conversion between *MAM1* and *MAM2*. Pairwise comparisons of sequence similarity between *MAM* genes of ecotypes containing both a *MAM2* and a *MAM1* locus reveal a maximal divergence between the Sorbo *MAM2* and *MAM1* genes; both nucleotide and protein sequences differ by  $\approx 5.0\%$  (Table 1). Therefore, the Sorbo ecotype likely represents the basal configuration of *MAM* genes at *GS-Elong* in *A. thaliana*. In contrast, in other ecotypes *MAM2* and *MAM1* show regions of much greater similarity, with a minimum value of  $\approx 1\%$  divergence between paralogous loci in the Conlara and Hodja ecotypes, indicating that genetic information was exchanged between *MAM2* and *MAM1* in these haplotypes.

Sliding window analyses of nucleotide polymorphism between *MAM1* and *MAM2* sequences were performed for those ecotypes harboring both loci, as well as for all *MAM1* and *MAM2*

**Table 1. Pairwise comparisons of amino acids (upper triangle) and nucleotides (lower triangle)**

		MAM1													MAM2									
		Aa-0, Ag-0, Col-0, Mt-0	Gy-0	Ema-1	Pla-0	Ler-0 trunc	Sorbo	Cvi-0	Cal-0	Di-1	Lm-2	Condara, Hodja	Mr-0	Mr-0	Cal-0	Ka-0, No-0, Sei-0, Tsu-0, Wl-0	Lip-0	Sorbo	Bl-1	Di-g, Ler-0, Petergof	Cvi-0	Condara, Hodja		
MAM1	Aa-0	***	0.0	0.2	0.2	-	0.6	0.4	1.2	1.2	1.8	3.8	2.0	4.1	4.3	3.8	3.8	4.7	4.5	4.7	4.1	4.3		
	Ag-0	0.0	***	0.2	0.2	-	0.6	0.4	1.2	1.2	0.0	3.8	2.0	4.1	4.3	3.8	3.8	4.7	4.5	4.7	4.1	4.3		
	Col-0	0.4	0.4	***	0.0	-	0.8	0.6	1.4	1.4	2.0	4.1	2.2	4.3	4.5	4.1	4.1	4.9	4.7	4.9	4.3	4.5		
	Mt-0	0.4	0.4	0.0	***	-	0.8	0.6	1.4	1.4	2.0	4.1	2.2	4.3	4.5	4.1	4.1	4.9	4.7	4.9	4.3	4.5		
	Gy-0	1.0	1.0	0.9	1.0	***	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
	Ema-1	1.6	1.6	1.6	1.6	1.6	***	1.0	1.8	1.8	2.0	3.6	2.6	4.7	4.5	4.1	4.1	4.9	4.7	4.9	4.3	4.5		
	Pla-0	1.3	1.4	1.2	1.3	0.4	1.7	***	1.2	1.2	1.8	3.8	2.4	4.5	4.3	3.8	3.8	4.7	4.5	4.7	3.6	4.3		
	Ler-0 trunc	1.5	1.5	1.4	1.5	0.5	1.9	1.0	***	0.0	1.0	4.5	3.2	4.7	4.1	3.6	3.6	4.9	4.7	4.9	4.7	4.9		
	Sorbo	1.3	1.4	1.3	1.3	0.4	1.8	0.9	0.2	***	1.0	4.5	3.2	4.7	4.1	3.6	3.6	4.9	4.7	4.9	4.7	4.9		
	Cvi-0	2.3	2.4	2.3	2.4	0.5	2.7	1.7	1.4	1.5	***	3.4	2.6	4.1	3.0	3.4	3.4	3.8	3.6	3.8	3.6	3.8		
	Cal-0	3.9	3.9	3.7	3.8	2.6	3.0	3.3	3.5	3.4	3.7	***	1.8	1.4	0.8	1.6	1.6	1.2	1.0	0.4	1.0	0.8		
	Di-1	2.6	2.7	2.6	2.6	0.9	3.3	2.8	3.3	3.2	2.0	2.7	***	2.0	2.2	2.6	2.6	2.6	2.4	2.6	2.0	2.2		
	Lm-2	4.3	4.3	4.2	4.3	3.0	4.8	3.9	4.5	4.3	3.1	2.1	1.5	***	1.4	1.8	1.8	0.6	0.4	0.6	1.6	0.6		
Condara, Hodja	4.1	4.1	4.0	4.2	1.8	4.4	3.5	3.7	3.5	2.3	1.4	1.4	1.0	***	0.8	0.8	1.2	1.0	1.2	1.0	1.2			
Mr-0	4.0	4.0	3.9	4.0	1.7	4.4	3.4	3.6	3.5	2.3	1.5	1.4	1.1	0.1	***	0.0	1.6	1.4	1.6	1.4	1.6			
MAM2	Ka-0	4.0	4.1	4.0	4.1	1.8	4.4	3.4	3.6	3.5	2.3	1.5	1.5	1.1	0.2	0.0	***	1.6	1.4	1.6	1.4	1.6		
	No-0	4.0	4.1	4.0	4.1	1.8	4.4	3.4	3.6	3.5	2.3	1.5	1.5	1.1	0.2	0.0	***	1.6	1.4	1.6	1.4	1.6		
	Sei-0	4.0	4.1	4.0	4.1	1.8	4.4	3.4	3.6	3.5	2.3	1.5	1.5	1.1	0.2	0.0	***	1.6	1.4	1.6	1.4	1.6		
	Tsu-0	4.0	4.1	4.0	4.1	1.8	4.4	3.4	3.6	3.5	2.3	1.5	1.5	1.1	0.2	0.0	***	1.6	1.4	1.6	1.4	1.6		
	Wl-0	4.0	4.1	4.0	4.1	1.8	4.4	3.4	3.6	3.5	2.3	1.5	1.5	1.1	0.2	0.0	***	1.6	1.4	1.6	1.4	1.6		
	Lip-0	4.7	4.7	4.6	4.7	2.6	5.0	3.8	4.4	4.2	2.9	1.9	2.0	0.7	0.7	0.8	0.8	***	0.2	0.4	1.4	0.4		
	Sorbo	4.5	4.5	4.4	4.6	2.4	4.8	3.7	4.3	4.1	2.8	1.8	2.0	0.7	0.6	0.6	0.7	0.4	***	0.2	1.2	0.2		
	Bl-1	4.5	4.6	4.5	4.6	2.4	4.9	3.7	4.3	4.2	2.8	1.8	2.0	0.7	0.6	0.6	0.7	0.4	0.0	***	1.4	0.4		
	Di-g, Ler-0, Petergof	3.7	3.7	3.6	3.8	2.0	4.0	2.6	3.4	3.3	2.0	2.0	2.0	1.4	0.9	1.0	1.0	1.2	1.2	1.3	***	1.0		
	Cvi-0	5.0	5.0	4.9	5.1	3.1	4.5	4.3	4.8	4.6	3.3	1.1	2.4	1.1	1.2	1.2	1.3	1.0	0.9	0.9	1.7	***		

Values indicate percent divergence. Nucleotide divergence was calculated for the entire genes, including 1,098 positions 5' of the start codons, i.e., including 5' untranslated and promoter sequences (except for the truncated Ler-0 MAM1, where only the remaining portion of the gene was used) and 60 positions 3' of the stop codons, i.e., including 3' untranslated sequences.

\*The Lm-2 sequence is a fusion between MAM2 and MAM1.

sequences surveyed (Fig. 2). The nucleotide polymorphism pattern between MAM1 and MAM2 in Sorbo largely reflects the general pattern seen among all MAM1 and MAM2 genes (Fig. 2 A and C), further strengthening the hypothesis that Sorbo reflects the basal configuration of MAM loci at *GS-Elong* in *A. thaliana*. In contrast, all other ecotypes harboring two MAM loci, i.e., Cal-0, Condara, Hodja, Cvi-0, and Mr-0, show at least partial sequence identity between the respective genes. However, observed regions of sequence identity differ between ecotypes, suggesting that multiple gene conversion events occurred independently of each other. Genetic information has evidently been transferred in both directions, i.e., in some cases from the MAM1 to the MAM2 locus, and in others from MAM2 to MAM1. For example, in the Condara and Hodja ecotypes most of the MAM1 locus is very similar to the Sorbo MAM2 sequence (Fig. 2B). In contrast, in Cal-0, most of the 3' part of MAM1 resembles a MAM2-like sequence (Fig. 2E). Finally, in Cvi-0, one central part of MAM2 was converted into a MAM1-like sequence, whereas the 3' part of MAM1 turned into a MAM2-like sequence (Fig. 2D).

**Qualitative Effects Caused by Variation in the Organization of *GS-Elong*.** Ecotypes with a functional MAM1 sequence, like Col-0, accumulate methionine-derived glucosinolates with four methylene groups (C<sub>4</sub>) in their side chain, whereas those with only a single MAM2 sequence, like Ler-0, sequester predominantly C<sub>3</sub> aliphatic glucosinolates (4, 7). Plants with both a MAM1- and a

MAM2-like sequence produce primarily C<sub>4</sub> glucosinolates, indicating that MAM1 function is dominant over MAM2 function with respect to short-chain glucosinolates. As can be seen from Fig. 1, presence or absence of a full-length Sorbo-like MAM1 gene gives complete prediction of glucosinolate phenotype. Therefore, loss of the MAM1 function, either by deletion of the MAM1 allele or by gene conversion, results in a switch from a C<sub>4</sub> to a C<sub>3</sub>-producing ecotype, provided that MAM2 function is retained.

**Quantitative Effects Caused by MAM1/MAM2 Polymorphism.** To investigate the impact on plant performance of natural variation between MAM genes, Col-0 was crossed to a recombinant inbred line (17), CL5, which has a Ler-0 MAM2 allele and is predominantly Col-0 elsewhere in the genome (7). A total of 5,000 F<sub>2</sub> progeny were screened for recombination in a 210-kb interval containing the *GS-Elong* region. Prereproductive rosette plants from 58 recombinant, homozygous, near-isogenic F<sub>4</sub> lines were assayed for glucosinolates, growth rate, and resistance to larvae of two herbivorous lepidopteran insects, *P. xylostella* and *S. exigua*, a crucifer specialist and generalist, respectively (Fig. 3).

QTLs for total leaf aliphatic glucosinolate concentration and resistance to *S. exigua* both center at a 15-kb nonrecombinant region containing Col-0 MAM1 or Ler-0 MAM2, but not MAM-L (Fig. 4;  $F = 92.34$ ,  $df = 1, 54$ ,  $P = 0.00001$ ,  $n = 292$  for aliphatic glucosinolates;  $F = 13.70$ ,  $df = 1, 54$ ,  $P = 0.0005$ ,  $n = 1,212$  for *S. exigua*). These data show that the Ler-0 MAM2 allele at

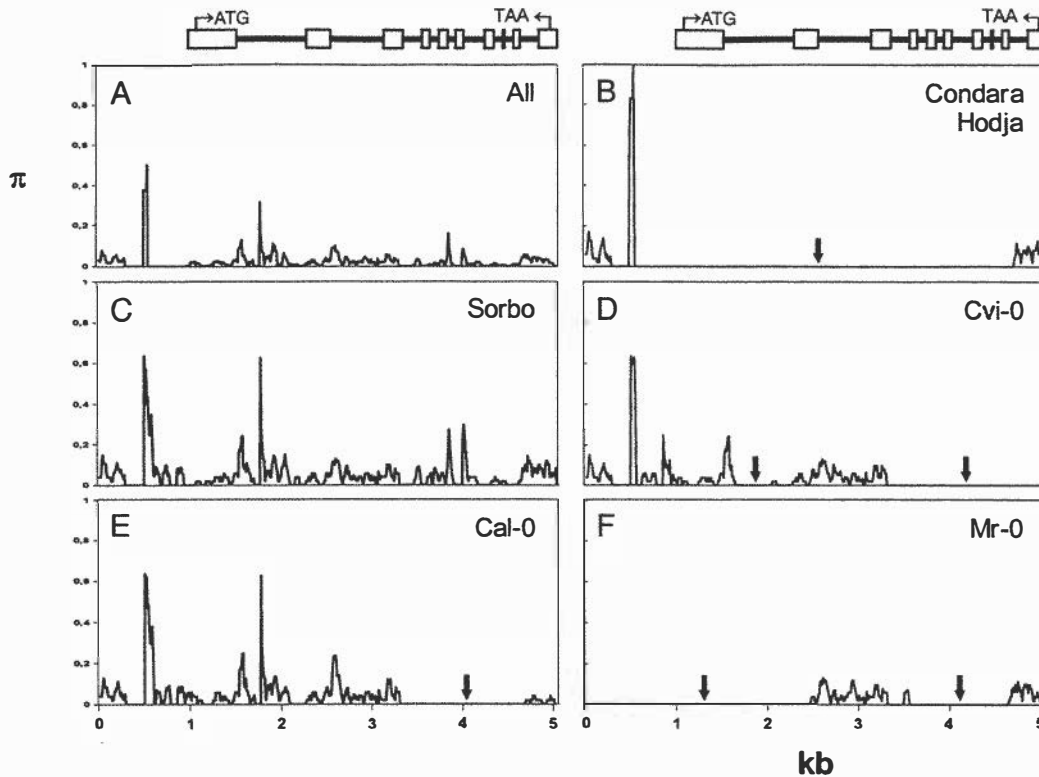


Fig. 2. Sliding window analysis of nucleotide polymorphism ( $\pi$ ) among *MAM1* and *MAM2* genes. Arrows indicate regions where compared sequences are identical because of one or more presumptive gene conversion events between paralogous genes. The peaks between 500 and 590 and between 1,820 and 1,830 nucleotides stem from complex changes resulting in poor sequence alignment. (A) All sampled *MAM1* and *MAM2* alleles (except truncated alleles). (B) *MAM1* versus *MAM2* in the Condara ecotype. Differences between *MAM1* and *MAM2* are very similar in Hodja. (C) *MAM1* versus *MAM2* in the Sorbo ecotype. (D) *MAM1* versus *MAM2* in the Cvi-0 ecotype. (E) *MAM1* versus *MAM2* in the Cal-0 ecotype. (F) *MAM1* versus *MAM2* in the Mr-0 ecotype. Window size, 50 nt; step width: 10 nt. Alignment gaps are included in scaling of the horizontal axis. The *MAM* gene structure is depicted above the panels.

*GS-Elong* causes increased aliphatic glucosinolate concentration and greater resistance to the generalist herbivore, *S. exigua*, in comparison to the Col-0 *MAM1* allele (Figs. 3 and 4). Herbivory

by the specialist insect, *P. xylostella*, was unaffected by the allelic state at *GS-Elong* ( $F = 2.50$ ,  $df = 1, 54$ ,  $P = 0.12$ ,  $n = 1,239$ ). Finally, above-ground biomass was measured in a separate experiment on 3,130 plants in the same 58 mapping lines at 16

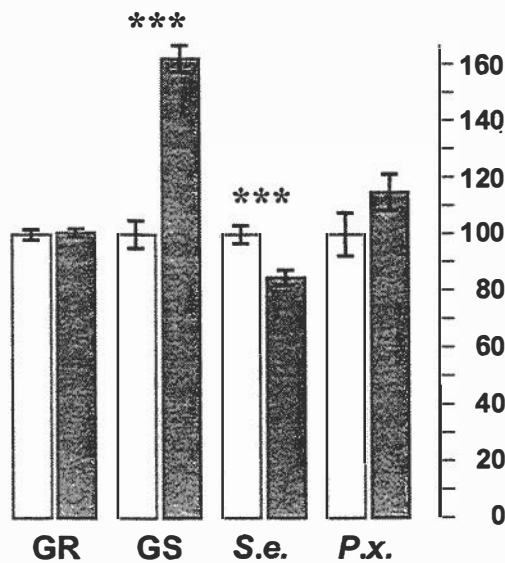


Fig. 3. Quantitative effects in near-isogenic lines carrying the Col-0 *MAM1* (white bars) or the Ler-0 *MAM2* allele (gray bars) at *GS-Elong*. Data are least square means ( $\pm$  standard error) from ANOVA (Col-0 = 100) for growth rate (GR), total aliphatic glucosinolates (GS), damage by *S. exigua* (S.e.), and by *P. xylostella* (P.x.). \*\*\*,  $P < 0.001$ .

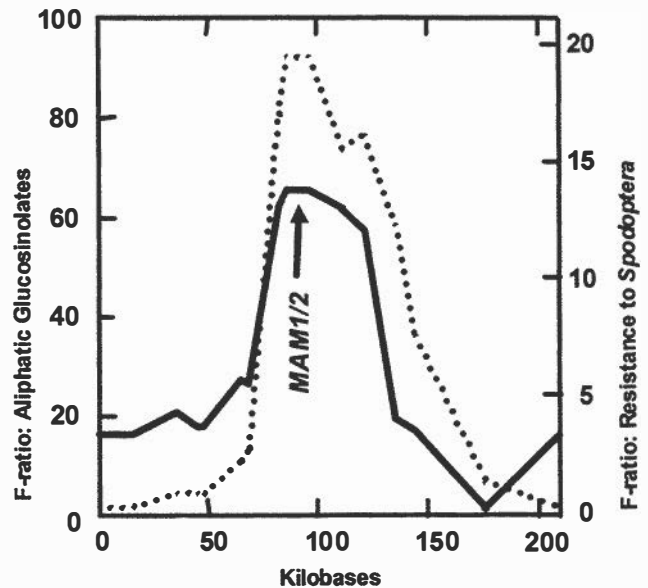


Fig. 4. QTL mapping of total leaf glucosinolate content (dashed line) and resistance to *S. exigua* (solid line). Statistical significance is indicated by F ratios.

**Table 2. Population genetic parameters in the *A. thaliana* GS-Elong region**

Gene	<i>MYB37</i>	<i>MAM2</i>	<i>MAM1</i>	<i>MAM-L</i>
Coding positions	721	1,518	1,518	1,509
No. of sequences	18	11	8	18
No. of haplotypes	4	4	3	7
$\pi_{\text{total}}$	0.0014	0.0045	0.0017	0.0034
$\pi_{\text{synonymous}}$	0.0030	0.0065	0.0039	0.0083
$\pi_{\text{nonsynonymous}}$	0.0009	0.0038	0.0010	0.0019
Tajima's <i>D</i>	0.43 (NS)	1.86*	-1.22 (NS)	0.40 (NS)
McDonald-Kreitman, <i>G</i> ; <i>P</i>	0.003; 0.954 (NS)	6.098; 0.014*	0.108; 0.743 (NS)	0.092; 0.762 (NS)

All calculations are based on coding regions of genes. Only the third exon of *MYB37* was amplified, which comprises >70% of the entire ORF. Statistical significance of *D* was investigated by coalescent simulations. Nucleotide diversity ( $\pi_{\text{total}}$ ) at *MAM2* is 0.0072 if we include alleles that have experienced gene conversion. NS, not significant.

\* $P < 0.05$ .

days after germination. There was no trace of a significant growth rate QTL at *GS-Elong* ( $F = 0.128$ ,  $df = 1, 54$ ,  $P = 0.73$ ).

**Nucleotide Polymorphism Patterns.** Statistical methods of molecular population genetics were used to test for nonneutral nucleotide polymorphism at *GS-Elong* (18, 19). Haplotypes for which evidence for gene conversion was found (e.g., Fig. 2) were excluded from significance testing to conform to the statistical assumption of identically and independently distributed point mutations. With the remaining alleles, two independent statistical tests reject the equilibrium neutral hypothesis at *MAM2*. First, a positive Tajima's *D* indicates too many intermediate frequency polymorphisms ( $D = 1.86$ ,  $P < 0.05$  or  $P < 0.01$ , based on coalescent simulations without recombination or with free recombination, respectively; Table 2). Second, we found too many amino acid polymorphisms segregating in *A. thaliana*. When compared with the corresponding gene from closely related *Arabidopsis lyrata*, the McDonald-Kreitman test (19) shows the ratio of replacement to synonymous nucleotide polymorphisms within *A. thaliana* is significantly higher than between species ( $G = 6.098$ ;  $P = 0.014$ ).

## Discussion

**Molecular Basis of an Insect Resistance QTL.** Quantitative analyses demonstrate that *GS-Elong* constitutes an insect resistance QTL, caused by variation in glucosinolate quantity, quality, or both. Reciprocal deletions of *MAM1* and *MAM2* have occurred in *Ler-0* and Col-0 ecotypes, respectively, so these paralogous loci segregate as alleles in our mapping population. A *Ler-0* *MAM2* allele at *GS-Elong* confers higher resistance to the generalist insect herbivore *S. exigua* than a Col-0 *MAM1* allele. Herbivory by a specialist insect, *P. xylostella*, was unaffected by the allelic state at *GS-Elong*. This is consistent with studies demonstrating that generalist insects are sensitive toward glucosinolate-based plant defenses, whereas specialists may be able to cope with these compounds (16, 20, 21). Moreover, feeding and oviposition of crucifer specialist insects may be stimulated by glucosinolates and their degradation products (21, 22). Indeed, *P. xylostella* has evolved a counteradaptation that enables it to circumvent hydrolysis of glucosinolates by myrosinase, and, thus, avoids the formation of toxic breakdown products (23).

This insect resistance QTL displayed complex molecular variation (Fig. 1) with multiple independent mutations causing production of C<sub>3</sub> glucosinolates, extensive gene conversion, deletion of large genomic regions, and functionally important genes appearing in *Ler-0* that are absent from the *A. thaliana* Col-0 sequence. Similarly, previous studies in model organisms have shown that gene conversion has important effects on sequence variation in *Hsp70* genes in *Drosophila* (24), and multiple independent deletions at the *FRI* locus influence life history variation in *A. thaliana* (25, 26). Conceptually, the

*GS-Elong* region exemplifies complex dynamics predicted for the evolution of gene families (27). In practice, the difficulty of such studies should not be underestimated. Our sustained attempts to characterize allelic variation in the *GS-Elong* region by using PCR ultimately failed, and we finally sequenced 300 kb of *Ler-0* bacterial artificial chromosomes to identify genes which were absent from Col-0.

**Evidence for Natural Selection.** Two independent statistical tests reject a neutral evolutionary hypothesis at *MAM2*. We found too many intermediate frequency nucleotide polymorphisms and too many amino acid changes segregating in *A. thaliana* ( $P < 0.05$  by Tajima's *D* and McDonald-Kreitman tests; Table 2). Although these findings reject a standard equilibrium neutral model, could these patterns be attributable to nonstandard or nonequilibrium demographic processes rather than to nonneutral evolution? For example, metapopulation structure and population decline can cause positive values of Tajima's *D* (28, 29). Likewise, relaxation of selection caused by bottlenecks or fixation of deleterious mutations in small populations can cause elevated levels of nonsynonymous polymorphism (30). Because these population processes affect multiple loci throughout the genome, we examined nucleotide polymorphism at three genes immediately flanking *MAM2* (Table 2). These results show that sequence polymorphisms adjacent to *MAM2* (*MYB37*, *MAM1*, and *MAM-L*) are compatible with an equilibrium neutral model, based on Tajima's and McDonald-Kreitman analyses (all  $P > 0.05$ ). Moreover, Haubold *et al.* (31) examined nucleotide polymorphism at 14 loci in a 170-kb genomic region containing the *MAM* gene family. They found that genes with contrasting patterns of variation ( $\pi$  and *D*) are located within a few kilobases of one another. That result corroborates our current finding, where nonneutral variation at *MAM2* contrasts sharply with neutral patterns of polymorphism at adjacent loci. Furthermore, this local scale of nucleotide variation appears to be typical for *A. thaliana* (32).

Molecular variation at *MAM2* also contrasts with results from other genes in *A. thaliana*. Although excess nonsynonymous polymorphism has been observed in comparisons of several *A. thaliana* genes with congeneric relatives (e.g., refs. 26, 33, and 34), this reflects locus-specific effects that are not found in most genes (34–39). Furthermore, *MAM2* displays an excess of intermediate-frequency nucleotide polymorphisms, in contrast to most other *A. thaliana* genes (26, 33, 34, 37, 39), including data from  $\approx 500$  loci sampled throughout the *A. thaliana* genome (K. Schmid and T.M.-O., unpublished data).

Natural genetic variation at *MAM2* shows too much intermediate-frequency nucleotide polymorphism and too many amino acid variants, relative to neutral predictions, suggesting that balancing selection maintains functional diversity at this ecologically important gene. Balancing selection refers to evolutionary mechanisms

that maintain more genetic variation than expected under neutrality (40), such as genotype-by-environment interaction, frequency-dependent selection, or trench warfare models of host-enemy coevolution (41). However, *MAM2* polymorphism has no impact on glucosinolate identity (Fig. 1), suggesting that nonneutrality at *MAM2* is likely caused by selection on glucosinolate quantity and not quality. In support of this interpretation, ecological analyses of natural selection on *A. thaliana* in the field find stabilizing selection for intermediate glucosinolate concentrations (42).

**Allocation Costs or Ecological Tradeoffs?** Balancing selection implies that different selective factors favor contrasting phenotypes. For example, genetic variation for insect resistance could be maintained by tradeoffs among components of fitness, if gains in one aspect of fitness were balanced by losses in other fitness components (43). One form of tradeoff, allocation costs, has been proposed to explain genetic polymorphism in plant resistance to insect herbivores. Allocation costs occur when defense mechanisms are energetically expensive, so that genotypes with strong defenses have fewer resources to invest in growth and reproduction (44). To test for possible allocation costs of glucosinolate production, we quantified growth rate (biomass) before bolting. Juvenile biomass is positively correlated with individual fitness in *Arabidopsis* and *Brassica* (45, 46). Furthermore, prereproductive vegetative growth rate measures resource availability in the exact environment and growth stage where the *Ler-0 MAM2* allele causes increased aliphatic glucosinolate concentration and greater resistance to the generalist herbivore, and allows us to test for allocation costs independently of tolerance (1). However, highly replicated quantification of juvenile growth rate in the fine-scale mapping lines gave no evidence for the existence of allocation costs. In this growth

environment and developmental stage, *MAM2* provides resistance to generalist insect herbivores without physiological costs. Further experiments will be required to measure fitness consequences in other developmental stages and environments.

An alternative mechanism of tradeoffs is well documented in the *Brassica* literature. Ecological costs occur when defensive metabolites are toxic to some herbivores but stimulate feeding or oviposition by adapted, specialist insects (43). Indeed, our insect-feeding assays detected enhanced resistance against the generalist herbivore *S. exigua*, conferred by the *MAM2* allele at *GS-Elong*, whereas herbivory by the crucifer specialist *P. xylostella* was unaffected by the allelic state of *MAM* genes at *GS-Elong*. On the other hand, there is compelling evidence that exposure to glucosinolates, and their degradation products does stimulate feeding, reproduction, or oviposition of a variety of crucifer specialists including butterflies, moths, aphids, beetles, and others, but not by generalists (ref. 22, reviewed in ref. 21). Thus, balancing selection at *MAM2* may be explained by ecological tradeoffs caused by contrasting biological effects of glucosinolates on specialist versus generalist herbivores, although we cannot rule out that allocation costs might have additional impact in some environments.

We thank Deana Pedersen (University of Montana, Missoula) for measuring plant growth rate and Dr. Scott A. McCuine (Department of Soil and Crop Sciences, Texas A&M University, College Station) for providing *Ler-0* bacterial artificial chromosomes. T.M.-O. was supported by European Union Contract No. QLRT-2000-01097, the Bundesministerium für Bildung und Forschung, U.S. National Science Foundation Grant DEB-9527725, and the Max-Planck-Gesellschaft. J.K. was supported by the Deutsche Forschungsgemeinschaft and the Max-Planck-Gesellschaft.

- Rausher, M. D. (2001) *Nature* **411**, 857–864.
- Halkier, B. A. (1999) *Trends Plant Sci.* **11**, 425–431.
- Rask, L., Andreasson, E., Ekblom, B., Eriksson, S., Pontoppidan, B. & Meijer, J. (2000) *Plant Mol. Biol.* **42**, 93–113.
- Kliebenstein, D. J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J. & Mitchell-Olds, T. (2001) *Plant Physiol.* **126**, 811–825.
- Reichelt, M., Brown, P. D., Schneider, B., Oldham, N. J., Stauber, E., Tokuhisa, J., Kliebenstein, D. J., Mitchell-Olds, T. & Gershenzon, J. (2002) *Phytochemistry* **59**, 663–671.
- Campos de Quiros, H., Magrath, R., McCallum, D., Kroymann, J., Schnabelrauch, D., Mitchell-Olds, T. & Mithen, R. (2000) *Theor. Appl. Genet.* **101**, 429–437.
- Kroymann, J., Textor, S., Tokuhisa, J. G., Falk, K. L., Bertram, S., Gershenzon, J. & Mitchell-Olds, T. (2001) *Plant Physiol.* **127**, 1077–1088.
- Shelton, A. M., Cooley, R. J., Kroening, M. K., Wilsey, W. T. & Eigenbrode, S. D. (1991) *J. Entomol. Sci.* **26**, 17–26.
- Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
- Rozas, J. & Rozas, R. (1999) *Bioinformatics* **15**, 174–175.
- Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
- Van de Peer, Y. & De Wachter, R. (1997) *Comput. Appl. Biosci.* **13**, 227–230.
- Tajima, F. & Nei, M. (1984) *Mol. Biol. Evol.* **1**, 269–285.
- Felsenstein, J. (1985) *Evolution (Lawrence, Kans.)* **39**, 783–791.
- Kliebenstein, D. J., Lambrix, V. M., Reichelt, M., Gershenzon, J. & Mitchell-Olds, T. (2001) *Plant Cell* **13**, 681–693.
- Kliebenstein, D. J., Pedersen, D., Barker, B. & Mitchell-Olds, T. (2002) *Genetics* **161**, 325–332.
- Lister, C. & Dean, C. (1993) *Plant J.* **4**, 745–750.
- Tajima, F. (1989) *Genetics* **123**, 585–595.
- McDonald, J. H. & Kreitman, M. (1991) *Nature* **351**, 652–654.
- Blau, P. A., Feeny, P., Contardo, L. & Robson, D. S. (1978) *Science* **200**, 1296–1298.
- Raybold, A. F. & Moyes, C. L. (2001) *Heredity* **87**, 383–391.
- Pivnick, K. A., Jarvis, B. J. & Slater, G. P. (1994) *J. Chem. Ecol.* **20**, 1407–1427.
- Ratzka, A., Vogel, H., Kliebenstein, D. J., Mitchell-Olds, T. & Kroymann, J. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 11223–11228.
- Bettencourt, B. R. & Feder, M. E. (2002) *J. Mol. Evol.* **54**, 569–586.
- Hagenblad, J. & Nordborg, M. (2002) *Genetics* **161**, 289–298.
- Le Corre, V., Roux, F. & Reboud, X. (2002) *Mol. Biol. Evol.* **19**, 1261–1271.
- Lynch, M., O’Hely, M., Walsh, B. & Force, A. (2001) *Genetics* **159**, 1789–1804.
- Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993) *Genetics* **134**, 1289–1303.
- Wakeley, J. & Aliacar, N. (2001) *Genetics* **159**, 893–905.
- Eyre-Walker, A., Keightley, P. D., Smith N. G. & Gaffney, D. (2002) *Mol. Biol. Evol.* **19**, 2142–2149.
- Haubold, B., Kroymann, J., Ratzka, A., Mitchell-Olds, T. & Wiehe, T. (2002) *Genetics* **161**, 1269–1278.
- Tian, D., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 11525–11530.
- Kawabe, A., Yamane, K. & Miyashita, N. T. (2000) *Genetics* **156**, 1339–1347.
- Olsen, K. M., Womack, A., Garrett, A. R., Suddith, J. I. & Purugganan, M. D. (2002) *Genetics* **160**, 1641–1650.
- Aguadé, M. (2001) *Mol. Biol. Evol.* **18**, 1–9.
- Hauser, M. T., Harr, B. & Schlötterer, C. (2001) *Mol. Biol. Evol.* **18**, 1754–1763.
- Kawabe, A., Innan, H., Terauchi, R. & Miyashita, N. T. (1997) *Mol. Biol. Evol.* **14**, 1303–1313.
- Kawabe, A. & Miyashita, N. T. (1999) *Genetics* **153**, 1445–1453.
- Kuittinen, H. & Aguadé, M. (2000) *Genetics* **155**, 863–872.
- Nordborg, M. & Innan, H. (2002) *Curr. Opin. Plant Biol.* **5**, 69–73.
- Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. (2001) *Science* **292**, 2281–2285.
- Mauricio, R. & Rausher, M. D. (1997) *Evolution (Lawrence, Kans.)* **51**, 1435–1444.
- Purrrington, C. B. (2000) *Curr. Opin. Plant Biol.* **3**, 305–308.
- Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. (2003) *Nature* **423**, 74–77.
- Mitchell-Olds, T. (1996) *Evolution (Lawrence, Kans.)* **50**, 140–145.
- Mitchell-Olds, T. & Bradley, R. D. (1996) *Evolution (Lawrence, Kans.)* **50**, 1859–1865.

# Diversification of furanocoumarin-metabolizing cytochrome P450 monooxygenases in two papilionids: Specificity and substrate encounter rate

Weimin Li\*, Mary A. Schuler†, and May R. Berenbaum\*\*

Departments of \*Entomology and †Cell and Structural Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Diversification of cytochrome P450 monooxygenases (P450s) is thought to result from antagonistic interactions between plants and their herbivorous enemies. However, little direct evidence demonstrates the relationship between selection by plant toxins and adaptive changes in herbivore P450s. Here we show that the furanocoumarin-metabolic activity of CYP6B proteins in two species of swallowtail caterpillars is associated with the probability of encountering host plant furanocoumarins. Catalytic activity was compared in two closely related CYP6B4 and CYP6B17 groups in the polyphagous congeners *Papilio glaucus* and *Papilio canadensis*. Generally, P450s from *P. glaucus*, which feeds occasionally on furanocoumarin-containing host plants, display higher activities against furanocoumarins than those from *P. canadensis*, which normally does not encounter furanocoumarins. These P450s in turn catalyze a larger range of furanocoumarins at lower efficiency than CYP6B1, a P450 from *Papilio polyxenes*, which feeds exclusively on furanocoumarin-containing host plants. Reconstruction of the ancestral CYP6B sequences using maximum likelihood predictions and comparisons of the sequence and geometry of their active sites to those of contemporary CYP6B proteins indicate that host plant diversity is directly related to P450 activity and inversely related to substrate specificity. These predictions suggest that, along the lineage leading to *Papilio* P450s, the ancestral, highly versatile CYP6B protein presumed to exist in a polyphagous species evolved through time into a more efficient and specialized CYP6B1-like protein in *Papilio* species with continual exposure to furanocoumarins. Further diversification of *Papilio* CYP6Bs has likely involved interspersed events of positive selection in oligophagous species and relaxation of functional constraints in polyphagous species.

insects | metabolism | molecular modeling

Cytochrome P450 monooxygenases (P450s) comprise a vast superfamily of heme–thiolate enzymes that catalyze the NADPH-associated reductive cleavage of oxygen to produce a functionalized product and water (1). The genes encoding these enzymes constitute one of the largest known gene superfamilies (<http://drnelson.utmem.edu/CytochromeP450.html>), with the enormous proliferation reflecting the functional versatility of their encoded proteins. Studies in a wide variety of organisms have demonstrated that P450-catalyzed reactions are important for detoxification of exogenous compounds, such as drugs, toxic pollutants, pesticides, and plant allelochemicals, as well as biosynthesis of endogenous agents, such as steroid hormones, pheromones, and defense compounds (2–5).

It has been suggested that, although the earliest P450s in eukaryotes were important in the metabolism of endogenous substrates, coevolution between plants and herbivorous animals, including insects, expedited the diversification of P450 families (6, 7). The earliest eukaryotic P450s in both plants and animals used reactive oxygen to metabolize endogenous compounds,

such as steroids and fatty acids. Subsequent reciprocal adaptive selection between plants and herbivorous animals was associated with the rapid diversification of P450s initiating 400 million years ago, concomitant with the colonization of terrestrial habitats by plants and animals. Plants have used P450s to produce defense compounds (allelochemicals), and herbivorous animals, including insects, have used P450s to metabolize the toxins produced by plants. Multiple duplication and divergence events are thought to have allowed xenobiotic-metabolizing P450s, such as CYP2 and CYP3 in mammals and CYP6 in insects, to diversify and acquire new functions. Insect genome projects have revealed tremendous diversity in putative xenobiotic-metabolizing P450 families, with approximately half of the 90 P450s in the *Drosophila melanogaster* genome belonging to families CYP6 and CYP4 (8). In the evolution of these large gene families, Hughes and Nei (9) and Ota and Nei (10) have proposed that duplication events may be followed by a winnowing process whereby some duplicate genes “die out” because of accumulation of deleterious mutations. This “evolution by the birth-and-death process” allows the number of functional genes within a family to remain stable. The birth-and-death model may be particularly applicable to the diversification of P450s in herbivorous insects, which, during host shifts, encounter different selective forces associated with the biochemical defense profiles of their host plants.

CYP6B family genes and proteins have been characterized in two groups of lepidopterans: the *Helicoverpa/Heliothis* complex and *Papilio* species (11). Within the genus *Papilio*, proliferation of CYP6B genes has occurred within the context of dietary furanocoumarins, a class of secondary compounds that confer protection against herbivores, because, on activation by UV light, they bind covalently to DNA and protein (12, 13). Furanocoumarins occur in two structural configurations, linear and angular, in over a dozen plant families, and are most widely distributed and diverse in Rutaceae and Apiaceae, the preferred hosts for most *Papilio* species (14).

Despite the toxicity of furanocoumarins, the oligophagous *Papilio polyxenes* specifically feeds on furanocoumarin-containing Apiaceae (11). Transcripts of at least one gene, CYP6B1, are expressed at elevated levels in response to supplemental furanocoumarins (15). The CYP6B1 protein encoded by this gene displays very high activity against linear furanocoumarins and considerably less activity against angular furanocoumarins. Surprisingly, two closely related polyphagous *Papilio*

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Chemical Communication in a Post-Genomic World,” held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviation: P450s, cytochrome P450 monooxygenases.

\*To whom correspondence should be addressed. E-mail: maybe@uiuc.edu.

© 2003 by The National Academy of Sciences of the USA

**Table 1. Specific activities of CYP6B proteins coexpressed with house fly NADPH P450 reductase in baculovirus expression system**

P450	CO-diff	Specific activity (nmol/min/nmol P450). <sup>*</sup> means ± SD				
		Angelicin	Trioxsalen	Psoralen	Xanthotoxin	Bergapten
CYP6B4 ( <i>Pg</i> )	450	1.906 ± 0.180 <sup>a</sup>	1.412 ± 0.090 <sup>a</sup>	2.208 ± 0.115 <sup>a</sup>	3.214 ± 0.174 <sup>a</sup>	3.541 ± 0.126 <sup>a</sup>
CYP6B17 ( <i>Pg</i> )	450	0.060 ± 0.023 <sup>c</sup> (97%↓)	0.557 ± 0.081 <sup>b</sup> (61%↓)	0.381 ± 0.150 <sup>b</sup> (83%↓)	0.800 ± 0.251 <sup>d</sup> (75%↓)	1.122 ± 0.141 <sup>d</sup> (68%↓)
CYP6B21 ( <i>Pg</i> )	450	0.518 ± 0.130 <sup>b</sup> (73%↓)	0.773 ± 0.170 <sup>b</sup> (45%↓)	1.800 ± 0.211 <sup>a</sup> (18%↓)	2.547 ± 0.087 <sup>b</sup> (21%↓)	2.150 ± 0.266 <sup>b</sup> (39%↓)
CYP6B25 ( <i>Pc</i> )	450/420	0.372 ± 0.210 <sup>b</sup> (80%↓)	0.179 ± 0.190 <sup>d</sup> (87%↓)	0.627 ± 0.346 <sup>b</sup> (72%↓)	1.212 ± 0.160 <sup>c</sup> (62%↓)	1.486 ± 0.124 <sup>c</sup> (58%↓)
CYP6B26 ( <i>Pc</i> )	420	ND (100%↓)	ND (100%↓)	ND (100%↓)	ND (100%↓)	ND (100%↓)
CYP6B1 <sup>†</sup>	450	0.640	— <sup>‡</sup>	2.560	6.980	— <sup>‡</sup>

<sup>\*</sup>Specific activities for each furanocoumarin were compared, and significant differences are represented with superscript letters ( $P < 0.05$ ) (ANOVA test).

<sup>†</sup>CYP6B1 activities were measured at different ratio of MOIs at 4:1 of P450 vs. P450-NADPH reductase recombinant viruses (28).

<sup>‡</sup>Not tested.

species, *Papilio glaucus*, which occasionally encounters furanocoumarins, and *Papilio canadensis*, which is unlikely to encounter furanocoumarins because of the absence of furanocoumarins in its available host plants, also have inducible metabolisms of furanocoumarins (16). Sixteen highly conserved genes (92–99% protein identity) belonging to two groups, designated the *CYP6B4* and *CYP6B17* groups, have been isolated from these two species (16, 17). Although all of these P450 transcripts are inducible by furanocoumarins, the induced level of transcripts achieved in *P. glaucus* is generally higher than in *P. canadensis* (16). The initial member of this group to be defined functionally, CYP6B4 from *P. glaucus*, has the demonstrated capacity to metabolize linear furanocoumarins (ref. 18 and Table 1). Whether P450s paralogous or orthologous to CYP6B4 also possess these activities was unknown.

To study the evolution of structure and function of P450s within the context of shifts in host plant utilization and concomitant changes in the chemical milieu experienced by the insect, we compared the *CYP6B* induction profiles and protein functionalities in these two species in the context of host plant furanocoumarin chemistry. For this study, representative P450s from each of the CYP6B4 and CYP6B17 groups in these species were expressed in baculovirus expression systems in conjunction with the insect NADPH P450 reductase needed for full functional activity. Enzymes were tested for their activity and specificity with respect to host plant furanocoumarins and related compounds to determine whether reduced probability of encounter leads to loss-of-function sequence changes consistent with a birth-and-death process of gene diversification. Also, ancestral P450 genes were reconstructed by maximum likelihood methods to chart a theoretical course for this process.

## Methods

**Construction and Expression of Recombinant Baculovirus.** Three *P. glaucus* sequences, including the *CYP6B4* cDNA and the *CYP6B17* and *CYP6B21* genomic DNAs, and two *P. canadensis* sequences, including the *CYP6B25* and *CYP6B26* cDNAs were expressed by using the baculovirus expression system. For ex-

pression of the *CYP6B17* and *CYP6B21* sequences, their 5' UTR and introns were removed by using a PCR-based strategy before subcloning them into the pFASTBac1 baculovirus expression vector (Invitrogen). Briefly, this strategy involved amplifying the 5' end 417 bp of the *CYP6B17* or *CYP6B21* coding sequences with a forward N1 primer (5'-CCGCTCGAGATCATGTTAACGATATTAT-3') that contains the start codon and a reverse C5 primer (5'-CGGCTTAAGTTTTCTGACGTG-3'), and inserting the resulting PCR product into pBluescript SK vector. The remainder of the coding sequence was generated without its intron by amplifying the *CYP6B17* and *CYP6B21* genomic sequences with the INTFOR1 primer (5'-CTGGCCAGAGAAAATGCTTAGGAATGCGGTTTGGACAA-3') that spans the sequences flanking the intron and a reverse C2 primer (5'-CGGAAGCTTCAATAATTCGTGGTAAAA-3') and ligating this PCR product with the 5' coding sequence and the pBluescript SK vector. The resulting cDNA sequences, which contain an *Xho*I site upstream from the translation start, a *Hind*III site downstream from the translation stop and an *Eae*I site at the junction between the two fragments, were checked for amplification errors by sequencing.

All of these *CYP6B* cDNAs were subcloned into suitable restriction sites of the pFASTBac1 baculovirus expression vector for construction of recombinant viruses and expressed in the *Bac-to-Bac* expression system (Invitrogen). All procedures of construction and expression of recombinant virus in insect Sf9 cells were performed as described by the manufacturer. *P. polyxenes* CYP6B1 cDNA (L. Pan, Z. Wen, J. Baudry, M.R.B., and M.A.S., unpublished data) was expressed as a positive control to monitor P450 expression quality. Sf9 cell cultures were grown to a density of  $0.8 \times 10^6$  cells per ml in SF-900 serum-free medium supplemented with 8–10% FBS, 50  $\mu$ g/ml streptomycin sulfate, and 50 units/ml penicillin and cotransfected with recombinant P450 virus at a multiplicity of infection (MOI) of 2 and recombinant house fly NADPH P450 virus at an MOI of 20 (ref. 19 and L. Pan, Z. Wen, J. Baudry, M.R.B., and M.A.S., unpublished data). This MOI ratio of 1:10 was intended to supplement the limited electron transfer capacities of Sf9 cells.

Hemin was added to 2  $\mu\text{g}/\text{ml}$  final concentration 24 h after infection. Insect cells were harvested from  $\approx 6$  plates for each P450 and lysed as described by Chen *et al.* (20), and protein aliquots were frozen in liquid nitrogen before analysis. Carbon monoxide (CO) difference spectra were measured as in Omura and Sato (21) by using an extinction coefficient for the reduced CO complex of  $91 \text{ mM}^{-1}\text{cm}^{-1}$ .

**P450 Metabolic Activity Assays.** *In vitro* metabolism assays of furanocoumarins were conducted as described in Li *et al.* (16) except that the substrate concentration in each reaction was lowered to 1  $\mu\text{g}/\text{ml}$  (corresponding to final concentrations of 4.6  $\mu\text{M}$  xanthotoxin, 4.6  $\mu\text{M}$  bergapten, 5.4  $\mu\text{M}$  angelicin, or 5.4  $\mu\text{M}$  psoralen) and a different furanocoumarin was added as an internal control after reactions were terminated. Rates of *O*-deethylation of 7-ethoxycoumarin were determined by measuring the fluorescence of the 7-hydroxycoumarin product (22). All metabolic assays were performed with four replicates and repeated at least twice by using proteins prepared from independent infections. Metabolic activities were compared by analysis of variance as described (16).

**Sequence Alignments, Phylogenetic Analyses, and Reconstruction of Ancestral Sequences.** Amino acid sequences of 22 CYP6B proteins, 12 other CYP6 proteins, and the divergent CYP321A1 protein were aligned to rabbit CYP2C5 (PDB ID 1DT6) and *Bacillus megaterium* CYP102 (PDB ID 2HPD), for which crystal structures are available (23, 24). Multiple sequence alignments were performed by using CLUSTALW or multialignment modules within the MOE program (Chemical Computing Group, Montreal) with the Gonnet weight model and structural alignment enabled. Alignments generated by both methods were compared and modified according to secondary structures of the CYP2C5 and CYP102 proteins. The finalized alignment was analyzed by the MEGA program to construct the phylogeny of the CYP6 family. A maximum parsimony phylogeny of CYP6 sequences was generated by using max-mini branch-and-bound method, and the inferred phylogeny was tested by 500 bootstrap tests. A maximum likelihood method was used to construct ancestral CYP6B sequences from the above multiple alignment of protein sequences on the most parsimonious phylogeny by using PAML (25) with the implementation of Jones amino acid transformation model and iteratively estimated  $\gamma$  shape parameter. To subsequently determine the relationship of ancestral sequences to their descendant P450s, relative amino acid distances between the ancestral sequences and their descendant branches were computed by using the MEGA program based on either complete protein sequences or active site amino acid residues. Active site residues were defined, based on our recent furanocoumarin docking study, as amino acids that are  $<4.5 \text{ \AA}$  from the oxo-heme moiety or furanocoumarin substrates docked into 3D models of CYP6B proteins (26).

## Results and Discussion

**Catalytic Activity of CYP6B Enzymes.** For comparative purposes, five *Papilio* CYP6B enzymes, including CYP6B4 (*P. glaucus*), CYP6B17 (*P. glaucus*), CYP6B21 (*P. glaucus*), CYP6B25 (*P. canadensis*), and CYP6B26 (*P. canadensis*), were coexpressed with house fly P450 reductase in Sf9 cells as described in *Methods*. The quality and quantity of the expressed P450 proteins were determined by reduced CO difference analysis (21) using Sf9 cell lysates prepared from cells cotransfected with recombinant P450 and P450 reductase viruses. In these assays, all of the *P. glaucus* enzymes generated CO-difference maxima at 450 nm (P450 form) with no significant absorbance at 420 nm, indicating that all of the P450 proteins fold and incorporate heme correctly into their apoproteins. Of the two *P. canadensis* proteins, CYP6B25, an in-group homologue to CYP6B4 (17), generated

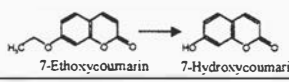
a major CO-difference peak at 450 nm and a minor peak at 420 nm and CYP6B26, an in-group homologue to CYP6B17 and seven amino acids shorter at its C terminus (16), generated a CO-difference peak only at 420 nm (P420 form), indicating that this enzyme is incorrectly folded. Molecular modeling and site-directed mutagenesis presented in Chen *et al.* (20) have indicated that aromatic residues F116, H117, and F484 positions are required for correct folding of the *P. polyxenes* CYP6B1 enzyme, and that two of these modulate the range of furanocoumarins metabolized. This seems not to be the case for *P. glaucus*/*P. canadensis* enzymes that contain a nonaromatic residue (L484) replacing F484 in the *P. polyxenes* enzyme and nonetheless generate only functional P450. The low stability of the CYP6B25 protein, which also contains L484, is apparently caused by substitutions at other positions because F116, H117, and L484 are conserved between CYP6B4 and CYP6B25. The inability of the CYP6B26 protein to fold into stable P450 indicates the importance of C-terminal residues in determining overall protein folding or configuration of the heme-binding domain.

To determine the influence of furanocoumarin exposure on P450 catalytic activity and selectivity, the metabolic activities of the expressed CYP6B proteins against five linear and angular furanocoumarins were assayed and compared with the metabolic activity of the *P. polyxenes* CYP6B1 protein coexpressed with P450 reductase in Sf9 cells (Table 1) (27). The substrates tested include methoxylated (bergapten, xanthotoxin), trimethylated (trioxsalen), and unsubstituted (psoralen) linear furanocoumarins and an unsubstituted (angelicin) angular furanocoumarin. These metabolic data indicate that all of the *Papilio* CYP6B enzymes are able to turn over furanocoumarin substrates to some extent, except for the denatured CYP6B26 protein, which did not metabolize any of the tested substrates (Table 1).

The relative abilities of the CYP6B proteins to metabolize linear furanocoumarins were closely associated with the probability of encountering furanocoumarins in host plants. CYP6B1, characterized from the specialist *P. polyxenes*, which encounters high levels of furanocoumarins in all of its host plants, turns over linear furanocoumarins at the highest rate (6980 pmol/min/nmol P450 for xanthotoxin) among all of CYP6B enzymes tested. CYP6B4 and CYP6B17, characterized from the generalist *P. glaucus*, which occasionally encounters furanocoumarins, turn over linear furanocoumarin substrates at rates lower than does CYP6B1 but significantly higher than do CYP6B25 and CYP6B26, characterized from the generalist *P. canadensis*, which never encounters furanocoumarins naturally. Further comparisons of P450 substrate specificities have indicated that CYP6B1 is more selective than the CYP6B proteins in polyphagous *P. glaucus*/*P. canadensis*. CYP6B1 exhibits very high activity toward the methoxylated linear furanocoumarins xanthotoxin and bergapten, lower activity toward unsubstituted and other linear furanocoumarins, and much lower activity toward the angular furanocoumarin, angelicin (ratio of activity for xanthotoxin/psoralen/angelicin is 1.0/0.4/0.1) (Table 1) (18, 27). This finding suggests that linear and angular furanocoumarins in *P. polyxenes* are metabolized by distinct P450s that possess relatively high substrate specificity. The in-group homologues CYP6B4 and CYP6B25 display more uniform activities toward a number of furanocoumarin substrates. Whereas CYP6B4 and CYP6B25 turn over methoxylated furanocoumarins at higher rates than other furanocoumarins, they metabolize psoralen and angelicin at rates higher than those of CYP6B1 (for CYP6B4, xanthotoxin/psoralen/angelicin is 1.0/0.7/0.6; for CYP6B25, it is 1.0/0.5/0.3). Importantly, these ratios demonstrate that these P450s exhibit the same preference for unsubstituted linear and angular furanocoumarins in that they turn over psoralen and angelicin at approximately the same rate. In comparison, CYP6B17, the paralog of CYP6B4 in *P. glaucus*, turns over linear



**Table 2. Ethoxycoumarin 7-O-deethylation activities of *P. glaucus*/*P. canadensis* CYP6B enzymes**

P450s					
	CYP6B4	CYP6B17	CYP6B21	CYP6B25	CYP6B26
O-deethylation* (pmol/min/nmol P450)	102 ± 6	143 ± 4	518 ± 4	ND <sup>†</sup>	ND <sup>†</sup>

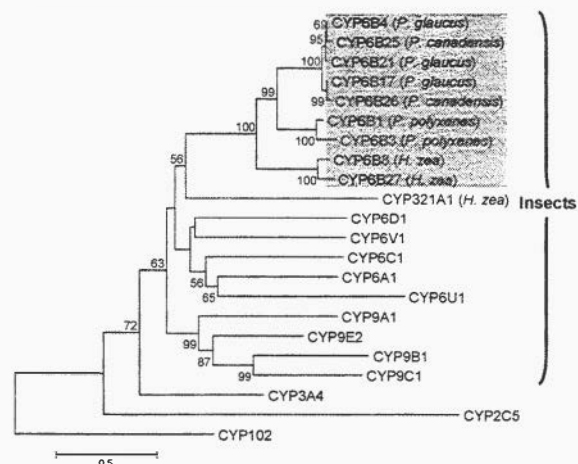
\*Values listed are mean ± SD of at least two independent determinations each with four replicates.

<sup>†</sup>No detectable O-deethylation activity found for CYP6B25 and CYP6B26.

furanocoumarins at rates lower than CYP6B4 and exhibits no activity against angular furanocoumarin. Together, these data indicate that, among the enzymes tested, furanocoumarin metabolism in *P. glaucus* is mediated primarily by CYP6B4.

Additional assays performed to determine the range of substrates other than furanocoumarins metabolized by CYP6B enzymes in *P. glaucus* and *P. canadensis* demonstrate that at least one substrate, 7-ethoxycoumarin (EC), representing alkoxyated coumarins that are widely present in the host plants of swallowtails (12), is metabolized by all three *P. glaucus* P450s at levels lower than those defined for furanocoumarin substrates. Interestingly, despite its higher activities against furanocoumarins, the 7-ethoxycoumarin O-deethylase (ECOD) activity of CYP6B4 is the lowest of all of the *P. glaucus* P450s tested, with an activity that is 5.1- and 1.4-fold lower than those of CYP6B21 and CYP6B17, respectively (Table 2). In contrast, CYP6B25 and CYP6B26 from *P. canadensis* have no detectable ECOD activity. Recent studies have indicated that the more specialized CYP6B1 from *P. polyxenes* also has no ECOD activity (Z. Wen, personal communication).

The catalytic efficiencies and substrate ranges of these CYP6B proteins map closely onto the range of host plants encountered by the insect species producing these P450s. In the specialist *P. polyxenes*, which has the highest frequency of encountering furanocoumarins, the CYP6B enzyme examined is more specialized and capable of metabolizing furanocoumarins at higher rates than the CYP6B proteins expressed in other *Papilio* species. In the generalist *P. glaucus*/*P. canadensis*, the CYP6B proteins have broader substrate ranges than the CYP6B1 proteins but metabolize this broader range of substrates at lower rates. The *P. canadensis* enzymes have even lower metabolic activities for all of the substrates tested than those of their *P. glaucus* orthologues. These relatively low activities may relate to the unpredictability of the chemical milieu encountered by *P. canadensis* and *P. glaucus* compared with the specialist *P. polyxenes* and their need to maintain P450 catalytic sites capable of metabolizing this broad range of substrates. For *P. glaucus*, which infrequently encounters these toxins, the low activities probably result from the lack of selection pressure to maintain furanocoumarin metabolism. This is particularly evident in *P. canadensis*, which has no natural exposure to furanocoumarins, and the lower activities of its CYP6B enzymes probably result from the accumulation of many deleterious mutations that have no purifying selection driving their elimination. Within *P. glaucus*, differences in substrate specificities between the CYP6B4 and CYP6B17 paralogs are likely to be the consequence of positive selection for duplicated genes. CYP6B7, CYP6B8, and CYP6B27, which are derived from *Helicoverpa zea*/*Helicoverpa armigera*, two closely related agricultural pests with extremely wide host ranges, are inducible by a wide range of chemicals, including insecticides, plant hormones, and plant allelochemicals, such as xanthotoxin (28–30). Among these, the CYP6B8 protein has a demonstrated ability to metabolize a range of plant allelochemicals and insecticides (31), as befits its highly diverse



**Fig. 1.** Reduced representation of the most parsimonious phylogeny of insect and mammalian P450s. The CYP subfamily was highlighted with gray background. The non-CYP6B sequences were retrieved from GenBank. The phylogenetic tree has been drawn proportional to branch lengths calculated by using maximum likelihood method. The numbers below internal branches are bootstrap values. The species origin of each CYP6B protein is indicated in parentheses after each sequence name.

chemical environment. This finding further confirms the association between P450 versatility and host plant range.

**Phylogeny of CYP6B Genes and Reconstruction of Ancestral CYP6B Sequences.** Given the close relationship between CYP6B catalytic efficiency and substrate specificity and the complexity of the chemical environment experienced by the insect, it is of great interest to explore how this association was established and how it evolved into the present-day CYP6B enzymes. For this analysis, a phylogeny of CYP6B genes was constructed based on the multiple sequence alignments described in *Methods* and resulting in the most parsimonious phylogeny shown in Fig. 1. In this, CYP6B sequences form a distinct clade with the *H. zea* CYP321A1 sequence that is phylogenetically separate from other CYP6 clades. The close relationship of the CYP321A1 protein to CYP6B proteins is confirmed by similarities in their function; CYP321A1 is the only non-CYP6B insect P450 known to be capable of metabolizing furanocoumarins (32). Within the CYP6B branch, *Papilio* and *Helicoverpa* CYP6B genes form distinct clades, suggesting no major duplication event before separation of these species. However, serial duplications occurred in both *P. glaucus*/*P. canadensis* and *H. zea*/*H. armigera* branches (Fig. 1), yielding at least 17 closely related CYP6B genes in the *P. glaucus*/*P. canadensis* genomes that do not appear to be pseudogenes (16, 17). After the first duplication, two paralogous groups, designated the CYP6B4-group and the CYP6B17-group (Fig. 1), formed. Although early studies demonstrated distinct induction profiles for each of these groups (16), the studies presented here indicate that members of these groups also display distinct substrate preferences, indicating functional divergence of these P450s after their initial duplication. The “redundancy” of CYP6B genes in these polyphagous species is likely a mechanism for adapting to complex chemical environments, maintaining one functional gene while another evolves with different metabolic capabilities. Duplicated copies of P450s may well provide candidates for subsequent functional divergence to adopt new functions as host plant patterns change.

To characterize the process of functional divergence among CYP6Bs, three ancestral sequences, the ancestor of the whole CYP6B subfamily (Anc1), the ancestor of the *Papilio* branch (Anc2), and the ancestor of the *P. glaucus*/*P. canadensis* branch

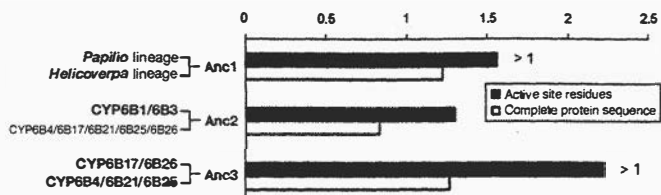


Fig. 2. Relative distance of ancestral P450s to CYP6Bs. Grouped bars represent the ratio of average amino acid distance between ancestral sequence and the upper branch to that of the lower branch. The distance ratio was computed either based on active site residues (filled bars) or on the complete protein sequence (open bars). The ratios indicated as >1 are significantly different from unity ( $P < 0.05$ ).

(Anc3), were reconstructed and compared with current CYP6B proteins (Figs. 2 and 3). The relative distances (ratios of distance between ancestral sequences and respective descendant branches) of all three ancestral sequences are not greater from unity than those computed based on the complete protein sequences, suggesting that these genes evolve at approximately the same rate from their common ancestors (Fig. 2). Similar computations were based on functionally more important active site residues, which directly interact with substrates; the relative distances of Anc1 to *Papilio* or *Helicoverpa* sequences, and Anc3 to CYP6B4 group or CYP6B17 group of sequences, respectively, are statistically different from unity ( $P < 0.05$ ) (Fig. 2). Anc1

displayed a significantly closer relationship to *Helicoverpa* P450s. The amino acid substitution rate leading to the *Papilio* lineage is 1.5-fold higher than the rate leading to the *Helicoverpa* lineage. It is conceivable that Anc1 may have similar functionality to that of *Helicoverpa* P450s and very likely catalyzed a broader range of substrates with lower metabolic efficiency against furanocoumarins compared with most of the *Papilio* P450s. Anc2 is proposed to have existed in a specialist species with constant exposure to furanocoumarins and to have had relatively higher catalytic activities against furanocoumarins compared with most *Papilio* P450s (17). This is likely to be the case, because the amino acid substitution rate of the CYP6B1/CYP6B3 lineage appears to be 1.3-fold that of the CYP6B4/CYP6B17 lineage, although this difference is not statistically significant because of higher sequence variation between CYP6B1/CYP6B3 pair. Anc3 is more closely related to the CYP6B4 group of P450s than the CYP6B17 group, as indicated by comparing the postduplication amino acid substitution rates leading to these lineages. Among all CYP6B lineages, the CYP6B17 and CYP6B4 branches displayed the most dramatic difference (2.1-fold) in the amino acid replacement rate from the immediate ancestor (Fig. 2). After the duplication that led to diversification of the CYP6B4 and CYP6B17 groups, the CYP6B4 group evolved relatively more slowly, and thus very likely resembles Anc3 in both sequence and function; the CYP6B17 group, however, as the duplicate presumably relieved of purifying selection, evolved faster and adapted to novel functionality.

Further insight into the evolutionary lineage of these P450 proteins has been gathered by comparing the predicted structures of these proteins with those predicted for the more efficient CYP6B1 protein. In this latter protein, an aromatic network that involves residues Phe-116, His-117, Phe-484, and Phe-371 is critical for substrate binding affinity to the CYP6B1 active site (20, 33). Three-dimensional models of the present day CYP6B4 and the ancestral Anc1 and Anc2 proteins that we have con-

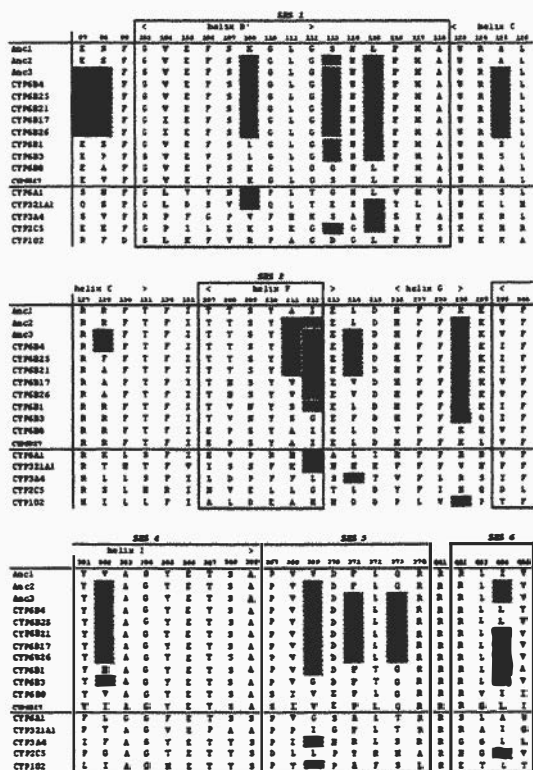


Fig. 3. A multiple sequence alignment of the amino acid residues in the active sites of CYP6B and their ancestral P450s. Three ancestral sequences are shown, the ancestral CYP6B sequence (Anc1), the ancestral *Papilio* CYP6B (Anc2) and the ancestral *P. glaucus/P. canadensis* CYP6B (Anc3). Residue positions are labeled according to their position in CYP6B4 and CYP6B1. The residues inside the black boxes represent SRS domains (35). Amino acids identical to Anc1 were highlighted in yellow; amino acids identical to Anc2 but not Anc1 are shown in blue; amino acids identical to Anc3, but neither Anc1 nor Anc2, are shown in red. Amino acids that were proposed to be critical to P450 substrate specificity (Fig. 4) are indicated in bold.

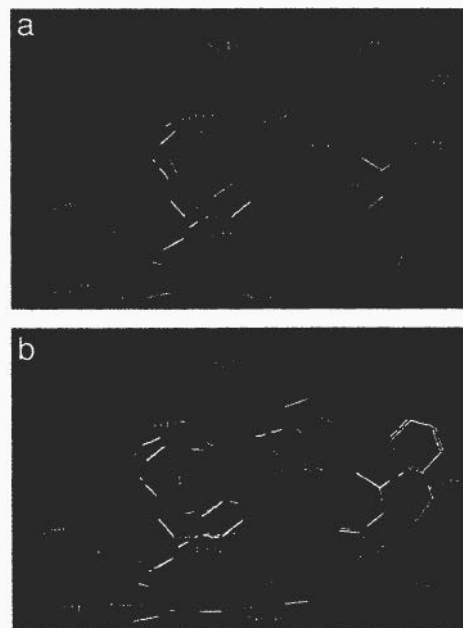


Fig. 4. Superposition of the active sites of ancestral CYP6B proteins with CYP6B1 and CYP6B4. (a) Anc1. (b) Anc2. The amino acid residues of CYP6B1 and CYP6B4 and the ancestral sequences are shown with the residues in the CYP6B1 sequence or progenitor to it labeled in red and amino acids in the CYP6B4 sequence or progenitor to it labeled in yellow. Residues involved in the aromatic interactive network in the CYP6B1 model are displayed as balls and sticks (33).

structured (26) were compared to evaluate the existence and importance of this aromatic network in defining the stabilities and substrate specificities of these proteins. The catalytic pocket of Anc1 displayed substantial differences from that of the CYP6B1 protein (Fig. 4). Among the four networked residues in the CYP6B1 model, Phe-116, His-117, and Phe-371 are conserved in other *Papilio* CYP6B proteins. In the Anc1 model, the aromatic Phe-116 and His-117 residues are not oriented perpendicular to one another in positions typical for aromatic-aromatic interactions, and the side chain of Phe-371 projects away from the catalytic pocket in a configuration that may not allow it to directly interact with other residues in the catalytic pocket (Fig. 4a). In Anc1, the catalytic pocket is probably larger in volume and with more flexibility than that defined within the CYP6B1 protein because a nonaromatic Ile-484 replaces the larger Phe-484 found in the CYP6B1 protein. In this regard, Anc1 is very similar to *H. zea* CYP6B8, which is also predicted to contain a large and flexible catalytic pocket able to accommodate a wide range of substrates (31). The absence and/or weakening of the  $\pi$ - $\pi$  stacking interactions between Phe-484 and furanocoumarin substrates that facilitate binding in the CYP6B1 catalytic pocket predicts that Anc1 probably catalyzed metabolism of furanocoumarins significantly less efficiently than the CYP6B1 protein.

In contrast, Anc2 is very likely a CYP6B1-like protein that maintains the ability to metabolize furanocoumarins. Residues involved in formation of the aromatic network, Phe-116, His-117, Phe-371, and Phe-484, are conserved and occupy approximately identical spatial locations to those in the CYP6B1 protein (Fig. 4b), indicating that the Anc2 enzyme has an aromatic network that is similar to that in CYP6B1 protein. This might be presumed to make the catalytic pocket relatively narrow and rigid, indicating similar structural constraints to that in the

CYP6B1 protein, presumably resulting from strong selection by constant exposure to furanocoumarins. This finding provides important supporting evidence for our hypothesis that the common ancestor of these *Papilio* species encountered furanocoumarins with high frequency in its host plants. Two residues (Phe-371 and Phe-484) contributing to this aromatic network are not preserved in the *P. glaucus*/*P. canadensis* branches, as suggested in our recent studies (26). The lack of important aromatic interactions in the CYP6B4 active site probably causes lower binding affinity to furanocoumarins, which in turn leads to lower turnover rates for linear furanocoumarins than those in *P. polyxenes*. The less efficient but more versatile CYP6B4 protein is well suited to the toxicological needs of the polyphagous *P. glaucus*.

In summary, our studies demonstrated for the first time how shifts in herbivore utilization of host plants are associated with the evolution of the sequence, structure and function of their P450s. It is highly likely that insects possessing CYP6B enzymes originated from a broadly polyphagous species with limited capacity for specialized metabolism. As the specialist papilionids diverged from this generalist ancestor in association with their furanocoumarin-containing host plants, the ability to metabolize a broad range of substrates declined as the ability to metabolize the principal host plant defense compounds, the furanocoumarins, reached a level of refinement rarely equaled by other herbivores. This tradeoff between P450 breadth and specificity may explain in part the ubiquity of oligophagy that characterizes lepidopterans today (34).

We thank Dr. Jerome Baudry for extensive training in modeling programs and Dr. Liping Pan for providing recombinant baculovirus containing NADPH P450 reductase. This work has been supported by National Science Foundation Grant IBN 02 12242 (to M.R.B.) and National Institutes of Health Grant R01 GM50007 (to M.A.S.).

- Lewis, D. F. V., Watson, E. & Lake, B. G. (1998) *Mutat. Res.* **410**, 245–270.
- Schuler, M. A. (1996) *Crit. Rev. Plant Sci.* **15**, 235–284.
- Chapple, C. (1998) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 311–343.
- Feyereisen, R. (1999) *Annu. Rev. Entomol.* **44**, 507–533.
- Scott, J. G. (1999) *Insect Biochem. Mol. Biol.* **29**, 757–777.
- Gonzalez, F. J. & Nebert, D. W. (1990) *Trends Genet.* **6**, 182–186.
- Nelson, D. R. (1998) *Comp. Biochem. Physiol.* **121**, 15–22.
- Tijet, N., Helvig, C. & Feyereisen, R. (2001) *Gene* **262**, 189–198.
- Hughes, A. L. & Nei, M. (1989) *Mol. Biol. Evol.* **6**, 559–579.
- Ota, T. & Nei, M. (1994) *Mol. Biol. Evol.* **11**, 469–482.
- Berenbaum, M. R. (2002) *J. Chem. Ecol.* **28**, 873–896.
- Berenbaum, M. R. (1991) in *Herbivores: Their Interactions with Secondary Plant Metabolites*, eds Rosenthal, G. & Berenbaum, M. R. (Academic, San Diego), pp. 221–249.
- Berenbaum, M. R. (1995) *Arch. Insect Biochem. Physiol.* **29**, 119–134.
- Murray, R. D. H., Mendez, J. & Brown, S. A. (1982) *The Natural Coumarins* (Wiley, Chichester, U.K.).
- Cohen, M. B., Schuler, M. A. & Berenbaum, M. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10920–10924.
- Li, W., Berenbaum, M. R. & Schuler, M. A. (2001) *Insect Biochem. Mol. Biol.* **31**, 999–1011.
- Li, W., Petersen, R. A., Berenbaum, M. R. & Schuler, M. A. (2002) *Insect Mol. Biol.* **11**, 543–551.
- Hung, C.-F., Berenbaum, M. R. & Schuler, M. A. (1997) *Insect Biochem. Mol. Biol.* **27**, 377–385.
- Koener, J. F., Carino, F. A. & Feyereisen, R. (1993) *Insect Biochem. Mol. Biol.* **23**, 439–447.
- Chen, J.-S., Berenbaum, M. R. & Schuler, M. A. (2002) *Insect Mol. Biol.* **11**, 175–186.
- Omura, T. & Sato, R. (1964) *J. Biol. Chem.* **239**, 2379–2387.
- Waxman, D. J. & Chang, T. K. H. (1998) in *Cytochrome P450 Protocols*, eds Phillips, I. R. & Shephard, E. A. (Humana, Totawa, NJ), pp. 175–179.
- Ravichandran, K. G., Boddupalli, S. S., Hasemann, C. A., Peterson, J. A. & Deisenhofer, J. (1993) *Science* **261**, 731–734.
- Williams, P. A., Cosme, J., Sridhar, V., Johnson, E. F. & McRee, D. E. (2000) *Mol. Cell.* **5**, 121–131.
- Yang, Z. (1997) *Comput. Appl. Biosci.* **13**, 555–556.
- Li, W. (2003) Ph.D. thesis. (University of Illinois at Urbana-Champaign, Urbana).
- Wen, Z., Pan, L., Berenbaum, M. R. & Schuler, M. A. (2003) *Insect Biochem. Mol. Biol.*, in press.
- Li, X., Berenbaum, M. R. & Schuler, M. A. (2002) *Insect Mol. Biol.* **11**, 343–351.
- Li, X., Schuler, M. A. & Berenbaum, M. R. (2002) *Nature* **17**, 712–715.
- Ranasinghe, C. & Hobbs, A. A. (1999) *Insect Mol. Biol.* **8**, 443–447.
- Li, X. (2003) Ph.D. thesis. (University of Illinois at Urbana-Champaign, Urbana).
- Sasabe, M. (2002) M.S. thesis (University of Illinois at Urbana-Champaign, Urbana).
- Baudry, J., Li, W., Pan, L., Berenbaum, M. R. & Schuler, M. A. (2003) *Protein Eng.*, in press.
- Bernays, E. A. & Graham, M. (1988) *Ecology* **69**, 886–892.
- Gotoh, O. (1992) *J. Biol. Chem.* **267**, 83–90.

# Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera

Wendell L. Roelofs\*<sup>†</sup> and Alejandro P. Rooney\*

\*New York State Agricultural Experiment Station, Cornell University, Geneva, NY 14456; and <sup>†</sup>National Center for Agricultural Utilization Research, Agricultural Research Service, U.S. Department of Agriculture, Peoria, IL 61604

Contributed by Wendell L. Roelofs, June 18, 2003

**A great diversity of pheromone structures are used by moth species (Insecta: Lepidoptera) for long-distance mating signals. The signal/response channel seems to be narrow for each species, and a major conundrum is how signal divergence has occurred in the face of strong selection pressures against small changes in the signal. Observations of various closely related and morphologically similar species that use pheromone components biosynthesized by different enzymes and biosynthetic routes underscore the question as to how major jumps in the biosynthetic routes could have evolved with a mate recognition system that is based on responses to a specific blend of chemicals. Research on the desaturases used in the pheromone biosynthetic pathway for various moth species has revealed that one way to make a major shift in the pheromone blend is by activation of a different desaturase from mRNA that already exists in the pheromone gland. Data will be presented to support the hypothesis that this process was used in the evolution of the Asian corn borer, *Ostrinia furnacalis* species. In that context, moth sex-pheromone desaturase genes seem to be evolving under a birth-and-death process. According to this model of multigene family evolution, some genes are maintained in the genome for long periods of time, whereas others become deleted or lose their functionality, and new genes are created through gene duplication. This mode of evolution seems to play a role in moth speciation, as exemplified by the case of the Asian corn borer and European corn borer, *Ostrinia nubilalis* species.**

Chemical communication systems in insects have provided exciting challenges to researchers in chemistry, biochemistry, physiology, ecology, genetics, and behavior for over four decades. Much of this research has been focused on moths in the order Lepidoptera, which is the second largest insect order with well over a hundred thousand described species. Most of the hundreds of species studied have been found to use a long-distance chemical communication system for attracting mates ([www.nysaes.cornell.edu/fst/faculty/acree/pheronet/index.html](http://www.nysaes.cornell.edu/fst/faculty/acree/pheronet/index.html)). Initially, pheromone components were characterized, and behavior mediated by these chemical cues was studied. However, increased knowledge of the precise blends used by different species only raised more questions on many aspects of the communication system. Underlying these questions was the fundamental curiosity about the extensive radiation seen in moths and what role pheromones played in the speciation process. How did sex pheromones evolve and could changes in this mating system give rise to isolated populations that become new species? Studies on pheromone biosynthetic pathways provided basic information on this issue, but the use of molecular techniques in the postgenomics era has become essential in addressing these questions.

## Pheromone Biosynthetic Pathways

By the 1980s, many moth pheromone components had been characterized, and a majority were acetates, alcohols, or alde-

hydes with long hydrocarbon chains (10–18C) containing 1–3 double bonds with variable positions and geometric configurations (1). Studies with labeled precursors were carried out to determine whether the pheromone components were produced via fatty acid precursors and what regulated the chain length and double bond positions. Data on the red-banded leafroller moth, *Argyrotaenia velutinana*, showed that the two major pheromone components, (Z)-11-tetradecenyl acetate (Z11-14:OAc) and (E)-11-tetradecenyl acetate (E11-14:OAc) were produced with  $\Delta$ 11 desaturation of myristic acid (14:Acid), followed by reduction and acetylation (2). Data on the cabbage looper, *Trichoplusia ni*, showed that the main pheromone component, (Z)-7-dodecenyl acetate (Z7-12:OAc), was produced by  $\Delta$ 11 desaturation of palmitic acid (16:Acid) followed by chain shortening to Z9-14:Acid and then to Z7-12:Acid, with subsequent reduction and acetylation (3).

As biosynthetic pathways were defined for more moth species, it became obvious that biochemical pathways involving two key enzymatic steps, desaturation and chain shortening, had evolved in the terminal segments (pheromone gland) of female moths. The limited chain-shortening steps added diversity by generating cascades of compounds of different chain lengths with double bond positions two carbons closer to the functional group with each round of chain shortening. It was soon found, however, that the integral membrane desaturases were able to add diversity by evolving unique substrate specificities, as well as regio- and stereospecificities. Thus, although many species were found to use rare  $\Delta$ 11 desaturases that produced Z11-16:Acid, mixtures of Z11/E11-14:Acid, or only E11-14:Acid, other species were found to use different desaturases. Primitive moth species in New Zealand are particularly interesting because the complex of morphologically similar species in the *Planotortrix* and *Ctenopseutis* genera produce pheromone components that involve at least  $\Delta$ 5,  $\Delta$ 7,  $\Delta$ 9, and  $\Delta$ 11 desaturases, either alone or in combination (4). How did these closely related species evolve such diverse desaturases, and how was it possible to mutate from one strongly stabilized pheromone system to another that used pheromone components with different double bond positions?

## Desaturase Genes

The biosynthetic enzymes in the moth pheromone glands could have evolved from genes involved in normal fatty acid metabolism, but it was soon recognized (5) that the resolution powers

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Chemical Communication in a Post-Genomic World," held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviations: ACB, Asian corn borer; ECB, European corn borer; Z14-16:Acid, (Z)-14-hexadecenoic acid (fatty acids and acetates of other configurations, chain lengths, and position of unsaturation are named similarly).

<sup>†</sup>To whom correspondence should be addressed. E-mail: WLR1@cornell.edu.

of data on biosynthetic pathways for phylogenetic reconstruction are limited and do not necessarily impart evidence of evolutionary direction. A move to the molecular level was dictated in an attempt to determine homologies among the metabolic  $\Delta 9$  desaturases and sex-pheromone desaturases.

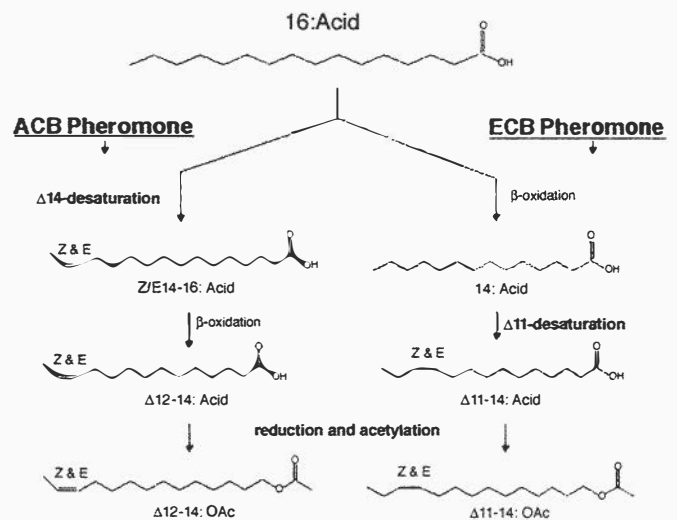
A collaborative effort in the early 1990s was initiated on the gene that encodes a  $\Delta 11$  desaturase in the sex-pheromone gland of female *T. ni*. Pheromone mRNA was isolated and reverse transcribed to make cDNA, which was used as template in PCR reactions with degenerate primers designed from conserved areas of rat and mouse acyl-CoA  $\Delta 9$  desaturases (6, 7). Candidate desaturase clones were obtained but had to be assayed for functionality. This proved to be difficult with an *in vitro* reconstitutive biochemical assay because integral membrane desaturases were shown to use a complex consisting of the desaturase protein, NADH-cytochrome  $b_5$  reductase (a flavoprotein), and cytochrome  $b_5$  (a hemoprotein). However, a discovery that the yeast *OLE1* desaturase can be functionally replaced *in vivo* by the rat desaturase (7) led the way to the development of yeast expression systems that could be used for the functional assay of pheromone desaturases (8).

The structures of  $\Delta 9$  and  $\Delta 11$  desaturase-encoding cDNAs from *T. ni* and the corn earworm moth, *Helicoverpa zea*, were first characterized and reported (8–10). Research on other selected species resulted in the characterization of structures of other pheromone desaturases with different regio- and stereospecificities, including Z/E11, E11, Z10, Z9, and Z/E14 desaturases (8–15). Cloning studies using mRNA isolated from fat bodies and pheromone glands revealed that there were at least two classes of  $\Delta 9$  desaturases in the Lepidoptera. One produced a mixture containing palmitoleic acid (Z9–16:Acid) > oleic acid (Z9–18:Acid), and the other with a reverse ratio of Z9–18 > Z9–16. It also became obvious that there were desaturase genes present in the pheromone gland that were not functioning to produce unsaturated fatty acid product and some desaturase clones that showed no activity in any of the available functional assays. These results provided some insights into how this multigene family evolved and how these genes could play a role in the speciation process.

### Corn Borer Speciation

Research on the European corn borer (ECB), *Ostrinia nubilalis*, has shown that pheromone production (female) and response (male) are not genetically linked (16), which provides the basis for models on the evolution of new pheromones based on asymmetric tracking (17, 18) and includes examples wherein a large mutational effect in female pheromone production is subsequently tracked by male response. Desaturase cloning studies with ECB surprisingly revealed a way that the large change in female pheromone production could be effected. The studies (15) showed that the ECB pheromone gland contained mRNA for three different desaturases, but unsaturated products were only present in the gland from one of them, the  $\Delta 11$  that makes precursors for the pheromone components, Z11-/E11–14:OAc (19) (Fig. 1). Another one was a  $\Delta 9$  desaturase, which is commonly present in pheromone glands, and the third was found to be a  $\Delta 14$  desaturase. The presence of the  $\Delta 14$  desaturase is significant because it is the desaturase used by the Asian corn borer (ACB), *O. furnacalis* (20), to produce its unusual mixture of Z/E12–14:OAc pheromone components (21) (Fig. 1).

The ACB is the only *Ostrinia* species in the world known to use the  $\Delta 14$  desaturase, with all others using a  $\Delta 11$  desaturase (22). Research on mRNA from ACB pheromone glands showed that they had the same three desaturase clones as ECB, only in this case the only unsaturated product was that from the  $\Delta 14$  desaturase. A sudden switch in pheromone components from those produced by  $\Delta 11$  desaturation to the products of a different biosynthetic pathway involving  $\Delta 14$  desaturation could help



**Fig. 1.** Pheromone biosynthetic pathways for ACB and ECB from hexadecanoic acid (16:Acid) and proceeding through different routes to the 14-carbon acetate pheromone components.

explain how the ACB population was derived about a million years ago. Studies of several other genes involved in the chemical communication system of ACB and ECB also seem to be unchanged, supporting the recent derivation of ACB. A single pheromone binding protein in male antennae has an identical structure in ACB and ECB males (23), and the reductase enzyme in ACB and ECB has similar specificities with a preference for Z11–14:Acid, even though ACB needs to reduce Z/E12–14:Acids (24).

The activation of a nonfunctional  $\Delta 14$  desaturase in the pheromone gland of an ancestral *Ostrinia* species would provide the mutational shift in pheromone production needed to initiate evolution of this chemical communication system, but would there be any males to respond to this pheromone blend? A screening of ECB males in flight-tunnel assays (25) showed that there are rare males that exhibit a broad range of responses that include flight to the ACB pheromone blend. This was surprising because ECB males would not be expected to respond to  $\Delta 12$  acetates and also because the ACB blend of  $\Delta 12$  acetates contains 33% of the E isomer (33% of E11 acetate is antagonistic to this race of ECB males) (25). However, the rare males exhibited complete upwind flights to both the ECB (97:3 mix of  $\Delta 11$  acetates) and ACB blend (2:1 mix of  $\Delta 12$  acetates). The existence of these rare males supports the possibility for a major shift in pheromone blend being tracked by male response and fits well with simulation models (26), showing that sudden major switches in pheromone blend and male response seem more likely than accumulation of small changes.

Discovery of the  $\Delta 14$  desaturase in ECB gives rise to many more questions on the origin of this gene and how long it has been carried in moths. Research in the postgenomic era is required to answer these questions.

### Evolutionary Mechanisms

In a previous study (15), we identified several unique aspects of insect desaturase multigene family evolution. The first was the discovery that this family originated before the split among Lepidoptera, Diptera, and Orthoptera, which is estimated to have occurred  $\approx 350$  million years ago (27). Second, we found that the family is composed of at least four gene clusters, evolving at disparate rates that are correlated with the function of each group (15). For instance, the  $\Delta Z9$  (16 > 18) group has evolved the slowest and contains metabolic desaturases, which

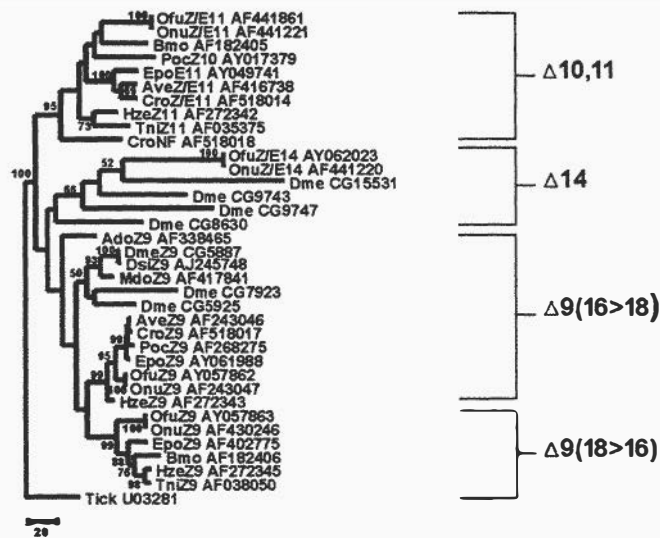


Fig. 2. Phylogeny of the desaturase genes of various insect species. Only species for which complete sequences were available were used. The tree was reconstructed from JTT amino acid distances (28) using the maximum likelihood method as implemented in the PROML computer program in the PHYLIP software package (<http://evolution.genetics.washington.edu/phylip.html>). Numbers along branches indicate bootstrap support from 1,500 replicates. The accession numbers for sequences are given after the species abbreviation, as per ref. 15. In the case of *D. melanogaster*, the gene names are given as they are listed in the *Drosophila* genome database ([www.fruitfly.org](http://www.fruitfly.org)). The tree is rooted with the desaturase gene from the tick (*Amblyomma americanum*), and the log likelihood was  $-14,878.90361$ .

presumably represent the ancestral function of this gene family. The  $\Delta Z9(18 > 16)$  and  $\Delta 10,11$  groups are composed of sex-pheromone desaturase genes and evolved at faster rates, which is coincident with a change in function from metabolism to reproduction. The  $\Delta 14$  group is made up of the fastest-evolving sequences (15) that are nonfunctional (e.g.,  $\Delta 14$  of *O. nubilalis*), function in sex-pheromone biosynthesis (e.g.,  $\Delta 14$  of *O. furnacalis*), or are yet to be functionally characterized. Interestingly, we found (15) that several dipteran sequences cluster within this group (Fig. 2). However, until we know the function of the proteins these genes encode, we cannot be certain on the basis of phylogenetic data alone, whether or not these represent new groups distinct from the sequences classified in the  $\Delta 14$  group.

Perhaps the most interesting finding of these studies is that reciprocal desaturase gene nonfunctionalization in *O. furnacalis* and *O. nubilalis* has resulted in a change in pheromone composition between these species. In this case, the  $\Delta 14$  gene of *O. furnacalis* makes a functional desaturase gene product but the  $\Delta Z/E11$  does not, whereas the opposite is true in *O. nubilalis*. Because the ORFs of the  $\Delta Z/E11$  sequences are identical and the ORFs of the  $\Delta 14$  differ by two amino acids in the *Ostrinia* species, the reciprocal nonfunctionalization event must have been relatively recent [i.e., at the time of their speciation, which is inferred to have occurred  $\approx 1$  million years ago (R. Harrison, personal communication), assuming a lepidopteran mtDNA substitution rate of 2% per million years (29)]. There are at least two possible mechanisms to account for the nonfunctionality of these genes. First, they may have recently become pseudogenes. The paradigm pseudogene is one that is no longer transcribed, possesses numerous nucleotide insertions or deletions resulting in frameshift mutations, and is highly divergent from its functional counterpart(s) (30). What is unusual about the *Ostrinia* nonfunctional genes is that they are well conserved between the two species (Fig. 2). However, recent pseudogenization could

account for this. In addition, there are several other cases in which pseudogenes have been shown to be conserved and even possess intact ORFs. For example, the bacteria species *Haemophilus aegyptius* and *Haemophilus influenzae* are sister taxa that diverged very recently, and they each possess *hap* pseudogenes that are identical in sequence (31). In this case, not enough time has elapsed for substitutions to have occurred in either species' pseudogene. Highly conserved pseudogene sequences may also result among paralogues that duplicated recently, as has been shown to occur among the three aquaporin pseudogenes of humans (99% sequence similarity) (32) and the human desaturase functional/pseudogene pair (95% sequence similarity) (33). There are even numerous examples in which a pseudogene produces an mRNA transcript but not a translated protein (e.g., refs. 34–37).

Another possible explanation for the nonfunctionality of certain *Ostrinia* genes is that there is an epigenetic mechanism repressing their transcription and/or expression (38). For example, the genes *nanos* and *pumilio* interact to control the translational repression of genes involved in embryonic patterning and spermatogenesis–oogenesis switching (39). However, in cells in which the *nanos* protein is not required, it is translationally repressed through the action of at least two proteins known as Smaug and Bicaudal (reviewed in ref. 40). Such mechanisms for translational repression are found throughout eukaryotes, suggesting that a similar one might be involved in the control of moth sex-pheromone desaturase genes, although whether or not such a mechanism can account for the situation with *Ostrinia* remains to be shown.

Nevertheless, sequence conservation and/or preservation of an intact ORF do not necessarily ensure the functionality of a gene. The only reliable way to determine whether a protein-coding gene with an intact ORF is functional is to determine whether a functional protein product is made. Our work on the *Ostrinia* genes shows that in no instance has a protein product or predicted sex pheromone been detected in an experimental assay (15). The next step, which we are currently pursuing, is to determine how the genes became nonfunctional. In this regard, we draw attention to a study on the *desat2* gene of *Drosophila melanogaster*, which showed that this gene has been nonfunctionalized in the Canton-S strain (41). In this case, a mutation in the promoter has inactivated the *desat2* gene, rendering it a recently formed pseudogene. Follow-up studies showed that this nonfunctionalization of the *desat2* gene has occurred in other *Drosophila* races and has led to an incipient speciation event among those with a functional *desat2* versus those with a nonfunctional *desat2* (42). This is because the functional *desat2* races produce a different pheromone complement than the nonfunctional *desat2* races, leading to a change in mating response. In addition to explaining how sex-pheromone desaturase gene nonfunctionalization has occurred in *Drosophila*, these studies reinforce our contention that desaturase gene nonfunctionalization can lead to speciation in insects (15). One question that remains is how gene duplication has played into the picture.

It is clear from the phylogenetic tree shown in Fig. 2 that there have been various points of desaturase gene duplication during the course of insect evolutionary history. The first duplication event gave rise to the  $\Delta 10,11$  group followed by subsequent duplications that gave rise to the  $\Delta 14$  and  $\Delta 9(18 > 16)$  groups. As we stated previously (15), we believe that the ancestral gene that gave rise to these groups was a Z9 gene from the  $\Delta 9(16 > 18)$  group based on the fact that (i) these genes function in metabolism in much the same way as in other animals and (ii) the involvement of desaturases in the production of sex pheromones is known to be a derived function found only in insects thus far. On further examination, it can be inferred that the ancestral moth Z9 gene duplicated in their common ancestor subsequent to the divergence from flies, roughly 330 million years ago (27),

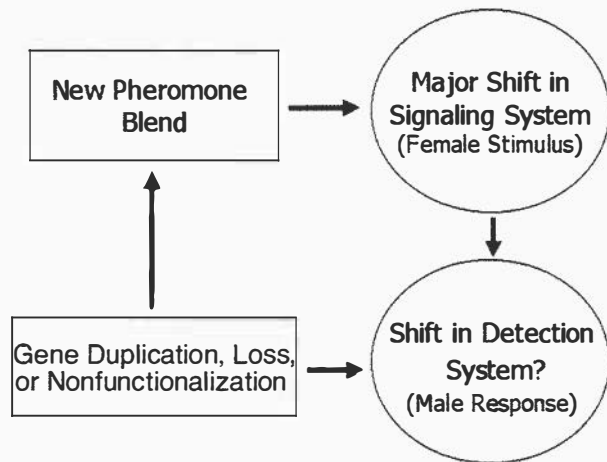


Fig. 3. Proposed birth-and-death mechanism for speciation in *O. furnacalis* and *nubilalis*. Here, gene duplication, loss, or nonfunctionalization leads to a shift in pheromone blend (female stimulus), which, if it leads to speciation, is accompanied by a shift in the detection system (male response), which is also brought about through gene duplication, loss, or nonfunctionalization.

and gave rise to the lepidopteran  $\Delta 9$  ( $18 > 16$ ) group (Fig. 1). Similarly, the ancestral dipteran Z9 gene seems to have duplicated sometime after the dipteran–lepidopteran divergence followed by another duplication event, thus resulting in three dipteran paralogs that cluster within the  $\Delta 9$  ( $16 > 18$ ) group (Fig. 1). Likewise, a nonfunctional gene (CroNF) seems to have duplicated from the ancestral moth Z11 gene, long before *Choristoneura rosaceana* diverged from the other moth species shown in the  $\Delta 10,11$  group.

The above results, along with those of previous studies (15, 41, 42), indicate that gene duplication, gene loss, and pseudogene formation have influenced the evolution of desaturase genes in moths and flies. These processes are characteristic of the birth-and-death model of multigene family evolution (43–46). According to the birth-and-death model, multigene families are created through gene duplication, which gives rise to new member genes, but some genes become deleted from the genome or degenerate into pseudogenes, whereas others are maintained. As a result, the members of a multigene family subject to birth-and-death evolution will evolve more or less independently, and different paralogs will be shared by different species or lineages. Thus, in a phylogenetic analysis of a multigene family including member genes from relatively closely related species, we should see the following observations: (i) sequences will cluster by gene or duplication order and not by species; (ii) low levels of sequence homogeneity between genes will be observed (especially at noncoding sites), except in the case of recent duplicates; and (iii) evidence of gene loss/deletion or pseudogene formation will be apparent. As we have already shown, all of these occur in the insect desaturase multigene family. In addition, we believe that the male moth pheromone-receptor system, which is composed of olfactory receptor loci (47, 48), may also be subject to a birth-and-death model of evolution, perhaps as a result of coevolution with the desaturase multigene family. A large and diverse olfactory receptor multigene family would provide an adaptive advantage for male moths, allowing for the rapid evolution of male response to female pheromone blends (Fig. 3). Although such studies have not yet been conducted on moths, we predict that high levels of olfactory receptor gene duplication, gene loss, and pseudogenization will be found for these species on the basis of what has been found in other animals. For instance, a family of at least 60 olfactory receptor loci have been identified in *D. melanogaster* (49), and in verte-

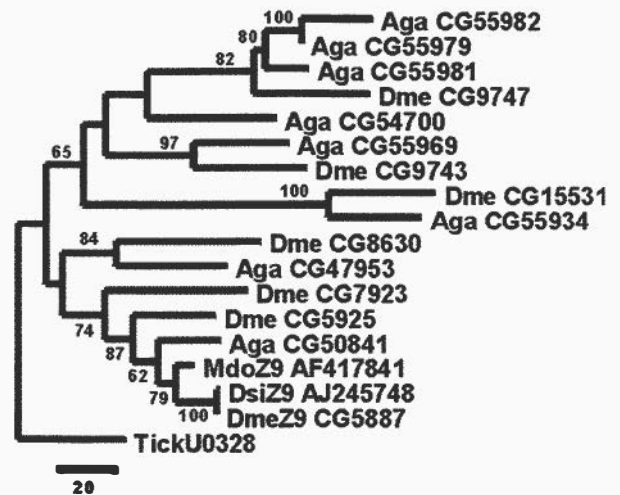


Fig. 4. Phylogeny of the desaturase genes of flies. The tree was reconstructed as in Fig. 1. Species and gene abbreviations are as per Fig. 2. The tree is rooted with the desaturase gene from the tick (*A. americanum*), and the log likelihood was  $-9,899.33096$ . Note that only partial sequences for the *A. gambiae* genes were available at the time this study was completed.

brates and *Caenorhabditis elegans* hundreds to thousands of olfactory receptor loci, including many pseudogenes, have been found (50, 51).

#### Future Directions: Genomics and Beyond

One particularly interesting avenue for future research involves studying how desaturase genes diversify once they duplicate. One possibility is that the genes originated under a subfunctionalization model (52–55) of gene family diversification (56). Under the subfunctionalization model, a generalized multifunctional gene/protein duplicate gives rise to one or more paralogs. The paralogous genes subsequently diverge and specialize in different functions that were previously all carried out by the ancestral, generalized multifunctional gene/protein. The problem is that direct evidence confirming multifunctionality of the ancestral/progenitor moth desaturase gene is lacking. An alternative to the subfunctionalization model is one in which sex-pheromone genes acquired their functions through rapid evolution subsequent to gene duplication via positive Darwinian selection (e.g., ref. 57). The ancient origin of insect desaturase genes and their increased substitution rates (15), which are highly pronounced in some cases (e.g., the  $\Delta 14$  and  $\Delta 10,11$  group genes in Fig. 2), are highly suggestive that nucleotide substitutions have become saturated. This obviously confounds attempts to detect positive Darwinian selection. For example, the estimate of synonymous substitution (57) between the *O. furnacalis* Z9 ( $16 > 18$ ) and Z/E14 genes is  $1.44 \pm 0.44$  and is  $2.96 \pm 1.20$  between the Z/E11 and Z/E14 genes. Clearly, these values are well above the saturation level (58, 59). On the basis of our current state of knowledge, it is not possible to determine which of the above models best fits the insect desaturases. Further studies aimed at determining whether the desaturases of basal insects (e.g., Collembola) are multifunctional may help to answer this question.

Another promising area of research is on the genomics of desaturase genes. Such studies are important for understanding the underlying basis of insect reproduction in terms of behavior as well as biochemistry. Consequently, until we obtain information on the number of desaturase genes and their genomic organization in various insect lineages, our understanding of the core elements of insect reproduction will be incomplete. Furthermore, such studies will provide insights into the evolution of insect desaturases. For example, now that their complete ge-

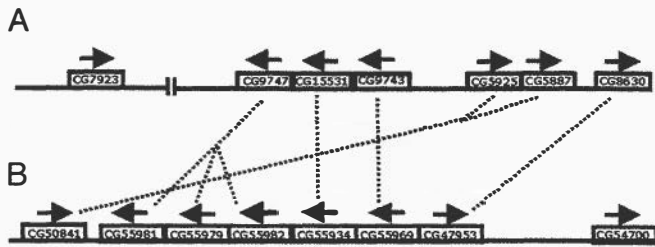


Fig. 5. Genomic map of desaturase genes in *D. melanogaster* chromosome 3 (A) and *A. gambiae* chromosome 2 (B). The vertical lines bisecting the *D. melanogaster* chromosome represent the division between the left and right arms of chromosome 3. Arrows above genes represent the direction of transcription. Dotted lines connect orthologous counterparts between species, on the basis of clustering patterns depicted in Fig. 4.

nome sequences are available, we know that the *D. melanogaster* and *Anopheles gambiae* genomes possess seven and eight desaturase genes, respectively (Figs. 4 and 5). By combining information from phylogeny (Fig. 4) and genomic organization (Fig. 5), we can make inferences regarding several aspects of desaturase evolution, genomics, and even gene function in dipterans. For instance, CG50841 of *A. gambiae* clusters significantly with the *D. melanogaster* genes CG5925 (*desat2*) and CG5887 (*desat1*). Because both of these genes are Z9 desaturases (Fig. 2), we predict that the *A. gambiae* CG50841 is also a Z9 desaturase.

The above approach can also be used to deduce the history of desaturase gene duplication and genomic organization. For example, the dotted lines connecting genes in Fig. 5 represent inferred orthologous relationships. Thus, *D. melanogaster* CG9747 and the ancestor of *A. gambiae* genes CG55981, CG55979, and CG55982 share an orthologous relationship. One piece of evidence for this comes from our phylogenetic analysis (Fig. 4), in which these genes cluster together with relatively high statistical support. Another piece of evidence comes from the conserved genomic organization of the fruit fly and mosquito desaturase genes (Fig. 5). In this case, the genomic location and direction of transcription of inferred orthologous gene pairs is conserved relative to other gene pairs. By using this line of reasoning in light of our results, we can also infer that there were four rounds of gene duplication that took place before the divergence of mosquitoes and flies. These events gave rise to (i) *D. melanogaster* CG8630 and *A. gambiae* CG47953, (ii) the ancestor of *D. melanogaster* CG5925/CG5887 and *A. gambiae*

CG50841, (iii) *D. melanogaster* CG9743 and *A. gambiae* CG55969, (iv) *D. melanogaster* CG15531 and *A. gambiae* CG55934, and (v) *D. melanogaster* CG9747 and the ancestor of *A. gambiae* CG55981/CG55979/CG55982. Once the *A. gambiae* and *D. melanogaster* lineages diverged, subsequent gene duplications and, in at least one instance, gene loss can account for an inability to identify an orthologous counterpart of some genes. For instance, *D. melanogaster* CG7923 (*fad2*) seems to have duplicated from the ancestral gene that gave rise to CG5925 and CG5887 subsequent to the divergence from *A. gambiae*, thus accounting for the lack of an orthologous counterpart in that species. On the other hand, it seems that *A. gambiae* CG54700 has been lost from the *D. melanogaster* lineage, because the orthologous relationships of all other genes can be determined except for this one, and its phylogenetic position is uncertain as well, because of lack of statistical support (Fig. 4). Incidentally, these results show that dipteran desaturase genes undergo birth-and-death evolution, as discussed in the previous section.

The above discussion illustrates the power of combining phylogenetic and genomic analyses. This approach has enhanced our ability to reconstruct the evolutionary history of desaturase evolution in flies and allows us to make inferences that could not have been made by using either approach alone. In the same way, knowledge of the genomic organization of desaturase genes from moths would improve our understanding of this multigene family in moths and expand our opportunities to conduct comparative genomic studies across even deeper evolutionary divides. In that context, it will be interesting to see what sorts of evolutionary patterns characterize the desaturase genes of other insect species in which the reproductive biology of pheromones is different but no less complicated. Among these are the cockroaches, in which each species makes a single pheromone so chemically complex that it is not duplicated by other species (60). This stands in contrast to moths, which make blends of several different pheromones that are relatively simple chemical structures. Whether or not similar modes of evolution characterize the genes controlling these disparate, yet related, systems remains to be seen, but we anticipate that the evolutionary complexity of whatever model(s) are uncovered will reflect this functional and chemical disparity.

We thank C. Linn for contributions to the research and manuscript, W. Liu for conducting the molecular studies on desaturases, R. Harrison and T. J. Ward for providing helpful comments/discussion, and A. L. Hughes and J. Zhang for reviews of the manuscript.

- Roelofs, W. L. & Bjostad, L. B. (1984) *Bioorg. Chem.* **12**, 279–298.
- Bjostad, L. B. & Roelofs, W. L. (1981) *J. Biol. Chem.* **256**, 7936–7940.
- Bjostad, L. B. & Roelofs, W. L. (1983) *Science* **220**, 1387–1389.
- Foster, S. P. & Roelofs, W. L. (1996) *Arch. Insect Biochem. Physiol.* **32**, 135–147.
- Dugdale, J. S. (1997) in *Insect Pheromone Research*, eds. Cardé, R. T. & Minks, A. K. (Chapman & Hall, New York), pp. 463–472.
- Thiede, M. A., Ozols, J. & Strittmatter, P. (1986) *J. Biol. Chem.* **261**, 13230–13235.
- Stukey, J. E., McDonough, V. M. & Martin, C. E. (1990) *J. Biol. Chem.* **265**, 20144–20149.
- Knipple, D. C., Miller, S. J., Rosenfield, C.-L., Liu, W., Tang, J., Ma, P. W. K. & Roelofs, W. L. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 15287–15292.
- Liu, W., Ma, P. W. K., Marsella-Herrick, P., Rosenfield, C.-L., Knipple, D. C. & Roelofs, W. L. (1999) *Insect Biochem. Mol. Biol.* **29**, 435–443.
- Rosenfield, C.-L., You, K.-M., Marsella-Herrick, P., Roelofs, W. L. & Knipple, D. C. (2001) *Insect Biochem. Mol. Biol.* **31**, 949–964.
- Liu, W., Jiao, H., Murray, N., O'Connor, M. & Roelofs, W. L. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 620–624.
- Hao, G., Liu, W., O'Connor, M. & Roelofs, W. L. (2002) *Insect Biochem. Mol. Biol.* **32**, 961–966.
- Liu, W., Jiao, H., O'Connor, M. & Roelofs, W. L. (2002) *Insect Biochem. Mol. Biol.* **32**, 1489–1495.
- Hao, G., O'Connor, M., Liu, W. & Roelofs, W. L. (2002) *J. Insect Sci.* **2**, www.insectscience.org/2.26/.
- Roelofs, W., Liu, W., Hao, G., Jiao, H., Rooney, A. P. & Linn, C. E., Jr. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 13621–13626.
- Phelan, P. L. (1997) in *Insect Pheromone Research*, eds. Cardé, R. T. & Minks, A. K. (Chapman & Hall, New York), pp. 563–579.
- Phelan, P. L. (1992) in *Insect Chemical Ecology*, eds. Roitberg, B. D. & Isman, M. B. (Chapman & Hall, New York), pp. 265–314.
- Löfstedt, C. (1993) *Philos. Trans. R. Soc. London B* **240**, 167–177.
- Ma, P. W. K. & Roelofs, W. L. (2002) *Zool. Sci.* **19**, 501–511.
- Zhao, C.-H., Löfstedt, C. & Wang, X. (1990) *Arch. Insect Biochem. Physiol.* **15**, 57–65.
- Klun, J. A., Bierl-Leonhardt, B. A., Schwarz, M., Litsinger, L. A., Barrion, A. T., Chiang, H. C. & Zhungxie, J. (1980) *Life Sci.* **27**, 1603–1606.
- Ishikawa, Y., Takanashi, T., Kim, C., Hoshizaki, S., Tatsuki, S. & Huang, Y. (1999) *Entomol. Exp. Appl.* **91**, 237–244.
- Willett, C. S. & Harrison, R. G. (1999) *Insect Biochem. Mol. Biol.* **29**, 277–284.
- Zhao, C.-H., Fang, L., Bengtsson, M. & Löfstedt, C. (1995) *J. Chem. Ecol.* **21**, 1795–1810.
- Linn, C. E., Jr., Young, M. S., Gendle, M., Glover, T. J. & Roelofs, W. L. (1997) *Physiol. Entomol.* **22**, 212–223.
- Butlin, R. K. & Trickett, A. J. (1997) in *Insect Pheromone Research*, eds. Cardé, R. T. & Minks, A. K. (Chapman & Hall, New York), pp. 548–562.
- Gaunt, M. W. & Miles, M. A. (2002) *Mol. Biol. Evol.* **19**, 748–761.
- Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992) *Comput. Appl. Biosci.* **8**, 275–282.
- Brower, A. V. (1994) *Proc. Natl. Acad. Sci. USA* **95**, 6491–6495.



30. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
31. Kilian, M., Poulsen, K. & Lomholt, H. (2002) *Mol. Microbiol.* **46**, 1367–1380.
32. Kondo, H., Shimomura, I., Kishida, K., Kuriyama, H., Makino, Y., Nishizawa, H., Matsuda, M., Maeda, N., Nagaretani, H., Kihara, S., *et al.* (2002) *Eur. J. Biochem.* **269**, 1814–1826.
33. Zhang, L., Lan, G. E., Parimoo, S., Stenn, K. & Prouty, S. M. (1999) *Biochem. J.* **340**, 255–264.
34. Kimura, M., Tokai, T., Matsumoto, G., Fujimura, M., Hamamoto, H., Yoneyama, K., Shibata, T. & Yamaguchi, I. (2003) *Genetics*, in press.
35. Lanteri, M., Giordanengo, V., Vidal, F., Gaudray, P. & Lefebvre, J. C. (2002) *Glycobiology* **12**, 785–792.
36. Moller, L. B., Petersen, C., Lund, C. & Horn, N. (2000) *Gene* **257**, 13–22.
37. Triglia, T., Thompson, J. K. & Cowman, A. F. (2001) *Mol. Biochem. Parasitol.* **116**, 55–63.
38. Wolffe, A. P. & Matzke, M. A. (1999) *Science* **286**, 481–486.
39. Parisi, M. & Lin, H. (2000) *Curr. Biol.* **10**, R81–R83.
40. Lipschitz, H. D. & Smibert, C. A. (2000) *Genet. Dev.* **10**, 476–488.
41. Dallerac, R., Labeur, C., Jallon, J. M., Knipple, D. C., Roelofs, W. L. & Wicker-Thomas, C. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 9449–9454.
42. Fang, S., Takahashi, A. & Wu, C. I. (2002) *Genetics* **162**, 781–784.
43. Hughes, A. L. & Nei, M. (1989) *Mol. Biol. Evol.* **6**, 559–579.
44. Nei, M. & Hughes, A. L. (1992) in *Proceedings of the 11th Histocompatibility Workshop and Conference*, eds. Tsuji, K., Aizawa, M. & Sasazuki, T. (Oxford Univ. Press, Oxford), pp. 27–38.
45. Nei, M., Gu, X. & Simikova, T. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 7799–7806.
46. Rooney, A. P., Piontkivska, H. & Nei, M. (2002) *Mol. Biol. Evol.* **19**, 68–75.
47. Laue, M., Maida, R. & Redkozubov, A. (1997) *Cell Tissue Res.* **288**, 149–158.
48. Krieger, J., Raming, K., Dewey, Y. M., Bette, S., Conzelmann, S. & Breer, H. (2002) *Eur. J. Neurosci.* **16**, 619–628.
49. Vosshall, L. B. (2001) *Chem. Senses* **26**, 207–213.
50. Glusman, G., Yanai, I., Rubin, I. & Lancet, D. (2001) *Genome Res.* **11**, 685–702.
51. Kratz, E., Dugas, J. C. & Ngai, J. (2002) *Trends Genet.* **18**, 29–34.
52. Hughes, A. L. (1994) *Proc. R. Soc. London Ser. B* **256**, 119–124.
53. Hughes, A. L. (1999) *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, Oxford).
54. Jensen, R. A. (1976) *Annu. Rev. Microbiol.* **30**, 409–425.
55. Lynch, M. & Force, A. (2000) *Genetics* **154**, 459–473.
56. Knipple, D. C., Rosenfield, C. L., Nielsen, R., You, K. M. & Jeong, S. E. (2002) *Genetics* **162**, 1737–1752.
57. Zhang, J., Rosenberg, H. F. & Nei, M. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 3708–3713.
58. Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, Oxford).
59. Tanaka, T. & Nei, M. (1989) *Mol. Biol. Evol.* **6**, 447–459.
60. Roelofs, W. L. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 44–49.



Arthur M. Sackler

COLLOQUIA

OF THE NATIONAL ACADEMY OF SCIENCES

## Chemical Communication in a Post-Genomic World

January 17–19, 2003

Arnold and Mabel Beckman Center, Irvine, CA

Organized by May R. Berenbaum and Gene E. Robinson

### Program

Friday, January 17

Welcome Reception and Registration

Saturday, January 18

#### Genomics of Chemical Attraction: Genomics of Olfaction

Cori Bargmann, University of California, San Francisco

Genomics of Olfaction in *Caenorhabditis elegans*

John Carlson, Yale University

Genomics of Olfaction in *Drosophila melanogaster*

Linda Buck, Fred Hutchinson Cancer Research Center

Genomics of Olfactory Receptors in Rodents

Gene Robinson, University of Illinois, Urbana–Champaign

Pheromone Regulation of Division of Labor in Honey Bee Colonies: From Behavior to Gene Expression Profiles in the Brain

#### Genomics of Chemical Attractions: Molecular Genetics, Evolution, and Biochemistry of Attractant Signaling

Hugh Robertson, University of Illinois, Urbana–Champaign

Molecular Evolution of the Insect and Nematode Chemoreceptor Superfamilies

Wendell Roelofs, Cornell University

Molecular Genetics of Pheromone Biosynthesis in Lepidoptera

John Hildebrand, University of Arizona

Central Processing of Chemosensory Signals: Anticipating Post-Genomic Opportunities

Nancy Moran, University of Arizona

Interactions of Insects with Bacterial Symbionts

Thomas Eisner, Cornell University

Can We Survive Without Natural Products Chemistry?

Jerrold Meinwald, Cornell University

The Role of Natural Products Chemistry in a Post-Genomic World

Sunday, January 19

**Molecular Genetics/Genomics of Chemical Repulsion and Defense:  
Predator/Prey and Pathogen/Host Interactions**

Baldomero Olivera, University of Utah

**Venomous Cone Snails, Specialists in Neuropharmacology: Reconstructing an  
Evolutionary History of Drug Development**

Bonnie Bassler, Princeton University

**How Bacteria Talk to Each Other**

Sir David Hopwood, John Innes Centre, Norwich Research Park

***Streptomyces* Secondary Metabolites: Much More than Weapons of Mass Destruction**

Klaus Hahlbrock, Max Planck Institute

**Transcriptional Reprogramming and Secondary Metabolite Accumulation During  
Plant/Pathogen Interactions**

**Molecular Genetics/Genomics of Chemical Repulsion and Defense: Plant/Insect  
Interactions**

Ian Baldwin, Max Planck Institute

**Allopolyploid Speciation and Its Effects on Ecological Adaptations: Plant–Herbivore  
Interactions in *Nicotiana* Native to North America**

Clarence Ryan, Washington State University

**Signaling for Herbivore Resistance in Solanaceae Species: Systemins, Prosystemins, and  
the Systemin Receptor**

Thomas Mitchell-Olds, Max Planck Institute

**Genomics of Plant/Insect Interactions in *Arabidopsis***

May Berenbaum, University of Illinois, Urbana–Champaign

**Cytochrome P450s in Insects—An Embarrassment of Riches**