

Letter to the Editor

Homology Modeling Suggests a Functional Role for Parallel Amino Acid Substitutions Between Bee and Butterfly Red- and Green-Sensitive Opsins

Adriana D. Briscoe

Department of Cellular and Structural Biology, University of Colorado Health Sciences Center, Denver, Colorado

Homology modeling is the process by which a protein sequence is threaded through the known atomic coordinates of a three-dimensional (3D) crystal structure, and the potential energy of the modeled molecule is minimized by adjusting the angle, orientation, and distances between constituent amino acids. Tools that automate this process are now available on the World Wide Web (Guex and Peitsch 1997), along with a database of empirically derived structural models (i.e., Protein Data Bank). These resources allow the theoretical modeling of many more structures than are currently available from experimental studies. Opsins are good candidate genes for homology modeling because of the availability of a high-resolution crystal structure of bovine rhodopsin (Palczewski et al. 2000) and the existence of physiological data for a number of naturally occurring spectral variants. Opsins are the protein component of the visual pigments and are members of the G protein-coupled receptor family, characterized by seven membrane-spanning alpha helices. The opsin protein is covalently attached to a light-sensitive chromophore, which in bovine rhodopsin is 11-*cis*-retinal and in bees and butterflies is either 11-*cis*-retinal or 11-*cis*-3-hydroxyretinal (Smith and Goldsmith 1990). The transmembrane domains of opsin form a binding pocket around the chromophore, and specific amino acids in the binding pocket modulate the absorption spectrum maximum that is most likely to trigger the chromophore's isomerization and activation of the opsin protein.

Vertebrate red- and green-sensitive cone pigments have been extensively studied by site-directed mutagenesis (Asenjo, Rim, and Oprian 1994), and the molecular basis of spectral tuning is well understood. Spectral tuning is determined by five to seven amino acids that are additive in effect (Yokoyama and Radlwimmer 1999). In contrast, very little is known about spectral tuning of invertebrate pigments (Britt et al. 1993). Comparison of known tuning sites of vertebrate opsins with homologous sites in butterfly revealed that the butterfly opsins are invariant at most of these positions, despite having a diversity of absorption spectrum maxima in butterflies that is comparable to that found in mammals (Briscoe 2000). Instead, a different set of candidate spectral tuning sites was found in an analysis of an expanded collection of butterfly and moth opsins (Briscoe 2001). At three of these sites, 115, 139, and 142, parallel/conver-

gent changes occurred along two red-shifted branches, *Heliconius sara* and the duplicated opsin *Papilio Rh3*. As an example, reconstructed parallel evolutionary changes at site 139 are shown in figure 1A.

To test whether the opsin protein is evolving in a similar way in other insect species, I analyzed the bee opsin data set of Ascher, Danforth, and Ji (2001). The data set of Ascher, Danforth, and Ji (2001) was downloaded from GenBank, the introns removed, and the remaining sequence data were aligned and checked for the region that included amino acids 139 and 142 (amino acid 115 was outside the region they sequenced). Of the 69 sequences in the data set, 64 were long enough to include these sites in a reconstruction of ancestral amino acid sequences. A total of 154 amino acid sites were included in the analysis. The maximum likelihood tree shown in figure 10 of their paper was used as the topology (minus the five sequences that were too short, *Tetraloniella* sp., *Exomalopsis completa*, *E. rufiventris*, *Apis nigrocincta*, and *Ceratina calcarata*) assumed for ancestral state reconstruction in MacClade using the maximum parsimony principle. A simplified version of their tree is shown in figure 1B. The ancestral states were reconstructed unambiguously at site 139; there was only a single most-parsimonious alternative at all nodes in the tree. Ancestral states for amino acid 142 were unambiguous everywhere in the tree, except at the node that leads to a clade composed of *Holcopasites calliopsisidis*, *H. ruthae*, *Leiopodus singularis*, *Xeromelecta californica*, and *Zacoscymia maculata* (see node marked by an asterisk in fig. 1B). All individual branches of the tree were examined for instances of parallel or convergent amino acid substitution at amino acids 139 and 142 and parallel changes at both sites were observed along two branches, *Diadasia diminuta* and *Euglossa imperialis*. A statistical test (Zhang 1997) showed that this number is larger than that expected by chance alone ($P < 0.00007$). Like the red-shifted butterfly opsins, a Phe to Tyr substitution was observed to occur at amino acid 139. Parallel substitutions at amino acid 142, however, involved a different kind of substitution, Gly to Ser, than the convergent Ala to Val or Ile to Val substitutions observed along the red-shifted butterfly lineages.

Substitutions at amino acid 139 among the bee opsins are relatively infrequent. Out of the 64 bee sequences, substitutions at this site occur only along the two branches already mentioned and along a third branch leading to a clade that includes *Paranthidium jugatorium* and *Anthidiellum notatum*. Like the species mentioned above, a Phe to Tyr substitution at amino acid 139 is observed to have occurred along this third branch, and both species have an Ser at amino acid 142 (acquired along an ancestral branch leading to this pair of species). In general, substitutions at amino acid 142 oc-

Key words: compensatory mutations, correlated evolution, spectral tuning, adaptation, visual pigments, structural biology.

Address for correspondence and reprints: Department of Ecology and Evolutionary Biology, University of California, Irvine, 321 Steinhaus Hall, Irvine, CA 92697. E-mail: abriscoe@uci.edu.

Mol. Biol. Evol. 19(6):983–986. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

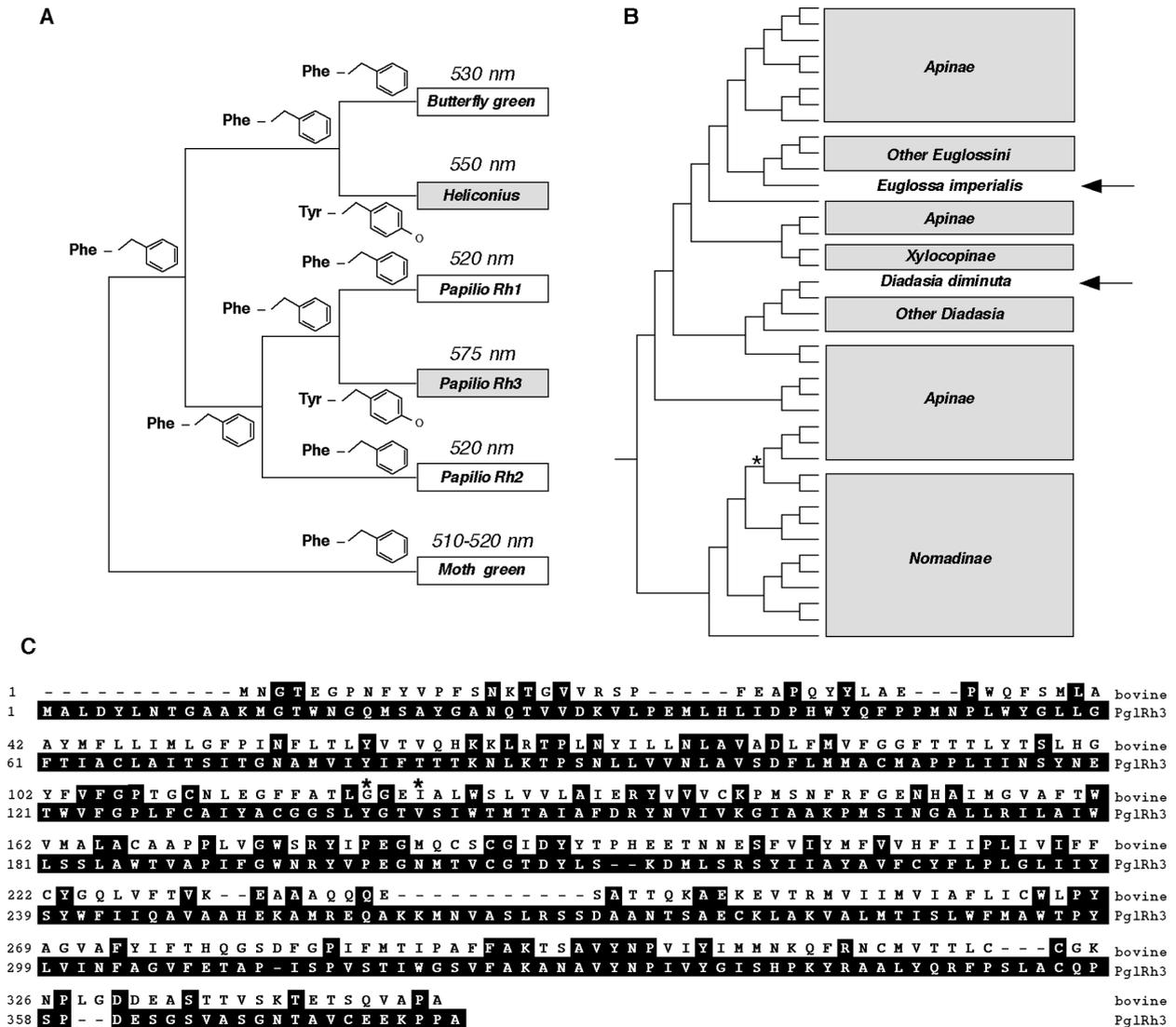


FIG. 1.—A, Reconstructed ancestral amino acids at site 139 among a group of red- and green-sensitive butterfly visual pigments. At least twice within this clade of opsins, red-shifted pigments have evolved: once, after a gene duplication event giving rise to the *Papilio Rh3* gene and once after a speciation event leading to the *Heliconius* butterfly lineage. B, Simplified phylogeny of bee LWS opsins as shown in figure 10 of Ascher, Danforth, and Ji (2001). Arrows indicate the positions of sequences undergoing parallel amino acid evolution at sites 139 and 142. Asterisk indicates node where ancestral state for amino acid 142 is ambiguous. C, Alignment of bovine rhodopsin and *P. glaucus* Rh3 (PglRh3) used in homology modeling. Location of amino acids 139 and 142 are indicated with an asterisk. Note: numbering is relative to *Papilio Rh3* and differs from numbering reported in Briscoe (2001), where amino acids 139 and 142 correspond to amino acids 94 and 97, respectively. GenBank Accession number of PglRh3: AF098283.

cur more frequently. In total, substitutions of the ancestral Gly at amino acid 142 include Cys (once), Thr (once), and Ser (seven times). No reversions were observed at either of these sites.

Compensating mutations at adjacent buried sites of proteins are thought to be rare (Baldwin et al. 1996). To determine whether there is a potential for physical interaction between these sites that might explain their tendency for correlated change, a homology model of *Papilio glaucus* Rh3 was constructed. In the sequence alignment shown, the lengths of the seven transmembrane helices and of two of the cytoplasmic and one of the extracellular loops of bovine rhodopsin are predicted to be nearly the same for both bee and butterfly long wavelength-sensitive (LWS) opsins based upon a visual

inspection of conserved amino acids flanking these domains (fig. 1C). Length differences between bovine and bee and butterfly LWS opsins exist as follows: extracellular loop II (E-II) contains a 2-amino acid indel, cytosolic loop III (C-III) contains a 13-amino acid indel, and E-III contains a 1-amino acid indel. Differences also exist in the length of the C- and N-termini, which are not involved in regulating the absorption properties of the visual pigment (fig. 1C).

The amino acid sequence of *P. glaucus* Rh3 (Briscoe 1999) was submitted to the SwissModel server using the default First Approach mode. This resulted in a truncated model of Rh3, fitted against the high-resolution (2.8 Å) bovine rhodopsin template. The coordinates of the bovine rhodopsin structure (1F88.pdb) were then

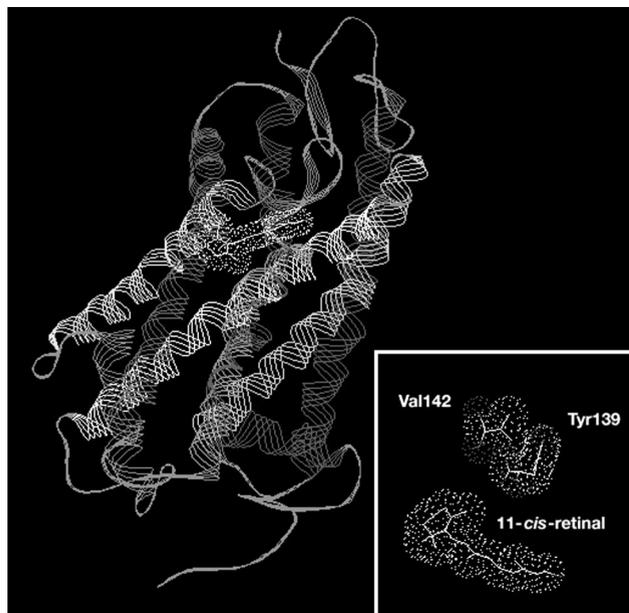


FIG. 2.—Ribbon diagram of 3D structure of *Papilio* Rh3 based upon homology modeling using bovine rhodopsin as a template. Seven transmembrane domains are indicated by helices. Inset: 11-*cis*-retinal chromophore and equivalent amino acids at sites 139 and 142 are shown with their Van der Waals forces.

downloaded from the Protein Data Bank. A manual alignment of *P. glaucus* Rh3 and bovine rhodopsin was performed in SwissPDBViewer, using 76 identical amino acids scattered throughout the protein as anchoring points and submitted directly to the SwissModel server. A homology model was generated from this procedure and inspected for the location of amino acids 139 and 142 using SwissPDBViewer (Guex and Peitsch 1997) (fig. 2).

Comparison of amino acids 139 and 142 of *Papilio* Rh3 with homologous sites in bovine rhodopsin reveals a correspondence to Gly120 and Ile123, respectively. Vertebrate visual pigments, including bovine rhodopsin, have a carboxylic acid residue, Glu113, in the third transmembrane domain that serves as a counterion to the protonated Schiff base nitrogen of the chromophore (Sakmar, Franke, and Khorana 1989; Zhukovsky and Oprian 1989). In a series of site-directed mutagenesis experiments, Zhukovsky, Robinson, and Oprian (1992) moved the location of the counterion from amino acid 113 to 120 and showed that the wild-type spectral properties could be partially recovered at position 120. From these experiments it can be concluded that the homologous site in *Papilio* Rh3, amino acid 139, is likely to be close enough to interact with the chromophore and potentially affect its absorption spectrum maximum. The type of amino acid substitution observed at this site in bee and butterfly opsins, a Phe to Tyr substitution, involves the same amino acids that are known to modulate spectral tuning of vertebrate opsins in other positions within the chromophore-binding pocket (Asenjo et al. 1994). Amino acid 142, located one helical turn above 139 in transmembrane domain III, is close enough in the structural model of *Papilio* Rh3 to have interacting Van

der Waals fields with amino acid 139. Whereas amino acid 142 is unlikely to have a direct effect on spectral tuning, it may have a steric effect on adjacent sites, potentially affecting protein stability (Baldwin et al. 1996). Both of these sites would be interesting to test for functional effects using site-directed mutagenesis.

Structural models are useful because they can provide detailed information about the nature of molecular interactions giving rise to the functional properties of a molecule. The relative volume, packing, and chemical nature of interacting components are important clues for determining which sites are likely to be under diversifying selection. Crystal structures are time-consuming to produce, and because of this, they are not likely to be available for many of the naturally occurring variants potentially of interest to biologists. In the absence of highly resolved structures, homology models are important tools for filling in the knowledge gaps. In the present study, a homology model of *Papilio* Rh3 provides a context for interpreting the correlated changes between two sites detected through the analysis of sequence data alone.

Amino acid 139 was also found to be within 4.5 Å distance of the polyene chain of the chromophore, as estimated from the bovine rhodopsin 3D structure (see fig. 6 of Palczewski et al. 2000). Mutagenesis of the homologous site in bovine rhodopsin (Gly120) has previously confirmed its proximity to and potential for spectral tuning effects on the chromophore (Zhukovsky, Robinson, and Oprian 1992). This suggests that the substitutions observed at amino acid 139 may in fact have an effect on the absorption spectrum maximum of the visual pigment, which in turn might explain why this site was observed to change in parallel along red-shifted butterfly opsin lineages (Briscoe 2001).

Natural selection leaves behind a trail of mutations that may be useful for inferring the functional domains or active sites of proteins. One approach for inferring which amino acid sites are functionally important is to map onto a gene tree instances of parallel or convergent amino acid substitution that can be correlated with convergent phenotypic evolution. Such an approach has previously been used to identify adaptive amino acid changes in stomach lysozymes of foregut fermenting mammals (Stewart, Schilling, and Wilson 1987; Swanson, Irwin, and Wilson 1991). Because parallel/convergent changes may be caused, either by chance or by selection, tests have been developed to determine whether such changes, if found, can be attributed to chance alone (Zhang and Kumar 1997). An excess of convergent or parallel change may be the result of natural selection. Such an approach was used to analyze parallel/convergent changes detected along two red-shifted butterfly opsin lineages (Briscoe 2001). An excess of parallel/convergent changes was observed at three sites, 115, 139, and 142. The present study extends the finding of correlated parallel changes at two of these sites to other insect LWS opsins.

Acknowledgments

My thanks to Jack Neff for intriguing discussions about the pollination ecology of *Diadasis*, Sudhir Ku-

mar for statistical advice, Steven Britt for discussions about spectral tuning, and Xiaojiang Chen for comments on the structural model. This work was supported by an NSF grant IBN-0082700 to A.B. and Steven G. Britt.

LITERATURE CITED

- ASCHER, J. S., B. N. DANFORTH, and S. JI. 2001. Phylogenetic utility of the major opsin in bees (Hymenoptera: Apoidea): a reassessment. *Mol. Phylogenet. Evol.* **19**:76–93.
- ASENJO, A. B., J. RIM, and D. D. OPRIAN. 1994. Molecular determinants of red/green color discrimination. *Neuron* **12**:1131–1138.
- BALDWIN, E., J. XU, O. HAJISEYEDJAVADI, W. A. BAASE, and B. W. MATTHEWS. 1996. Thermodynamic and structural compensation in “size-switch” core repacking variants of bacteriophage T4 lysozyme. *J. Mol. Biol.* **259**:542–559.
- BRISCOE, A. D., 1999. Intron splice sites of *Papilio glaucus PglRh3* corroborate insect opsin phylogeny. *Gene* **230**:101–109.
- . 2000. Six opsins from the butterfly *Papilio glaucus*: molecular phylogenetic evidence for paralogous origins of red-sensitive visual pigments in insects. *J. Mol. Evol.* **51**:110–121.
- . 2001. Functional diversification of lepidopteran opsins following gene duplication. *Mol. Biol. Evol.* **18**:2270–2279.
- BRITT, S. G., R. FEILER, K. KIRSCHFELD, and C. S. ZUKER. 1993. Spectral tuning of rhodopsin and metarhodopsin in vivo. *Neuron* **11**:29–39.
- GUEx, N., and M. C. PEITSCH. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714–2723 [<http://www.expasy.ch/swissmod/SWISS-MODEL.html>].
- PALCZEWSKI, K., T. KUMASAKA, T. HORI ET AL. (12 co-authors). 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289**:739–745.
- SAKMAR, T. P., R. R. FRANKE, and H. G. KHORANA. 1989. Glutamic acid-113 serves as the retinylidene Schiff-base counterion in bovine rhodopsin. *Proc. Natl. Acad. Sci. USA* **86**:8309–8313.
- SMITH, W. C., and T. H. GOLDSMITH. 1990. Phyletic aspects of the distribution of 3-hydroxyretinal in the Class Insecta. *J. Mol. Evol.* **30**:72–84.
- STEWART, C.-B., J. W. SCHILLING, and A. C. WILSON. 1987. Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* **330**:401–404.
- SWANSON, K. W., D. M. IRWIN, and A. C. WILSON. 1991. Stomach lysozyme gene of the langur monkey—tests for convergence and positive selection. *J. Mol. Evol.* **33**(5):418–425.
- YOKOYAMA, S., and F. B. RADLWIMMER. 1999. The molecular genetics of red and green color vision in mammals. *Genetics* **153**:919–932.
- ZHANG, J. 1997. CAPE: convergent and parallel evolution at the amino acid level. Pennsylvania State University, University Park, Pennsylvania.
- ZHANG, J., and S. KUMAR. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol. Biol. Evol.* **14**:527–536.
- ZHUKOVSKY, E. A., and D. D. OPRIAN. 1989. Effect of carboxylic-acid side chains on the absorption maximum of visual pigments. *Science* **246**:928–930.
- ZHUKOVSKY, E. A., P. R. ROBINSON, and D. D. OPRIAN. 1992. Changing the location of the Schiff base counterion in rhodopsin. *Biochemistry* **31**:10400–10405.

DAVID IRWIN, reviewing editor

Accepted February 25, 2002