tations may have implications for drug therapy, because computer simulations suggest that under these conditions, recombination impairs rather then facilitates the evolution of drug resistance in HIV (*24*). Third, it is unclear whether the beneficial effects of recombination for negative epistasis are negated by detrimental effects reflecting an equal degree of positive epistasis. However, approximate calculations assuming weak linkage between the locus coding for recombination rate and the loci under selection suggest that variation in the epistatic interactions weakens the selection for recombination and that recombination is selected against for positive mean epistasis (*25*). Finally, it remains to be shown whether the pattern of epistasis found in HIV-1 is representative of that in other organisms (in particular of that in eukaryotes). However, in contrast to studies measuring epistasis in *Escherichia coli*, yeast, or *Caenorhabditis elegans* (*10*, *13*, *15*), we have measured epistasis in an organism in which recombination occurs frequently. Therefore, it may in fact be more appropriate to extrapolate from retroviruses than from organisms that are effectively asexual.

The predominance of interactions with positive epistasis in HIV-1 raises the question of why retroviruses have evolved the capacity to recombine (*24*). Whether drift-based explanations for the benefits of recombination are applicable to HIV-1 remains to be investigated. Recent studies suggest that drift can favor the evolution of recombination even for very large populations, as long as there are a sufficient number of loci under selection (*26*). An alternative explanation is the repair of single-strand breaks (*27*). Retroviruses have single-stranded RNA genomes that are susceptible to ribonucleases and other agents during the viral life cycle. Template switching by RT in retroviruses could extend transcription beyond breakage points and could thus explain why retroviruses evolved to carry two complete genomes. According to this hypothesis, however, recombination in retroviruses would be the consequence but not the cause of the evolution of template switching.

Note added in proof: After this paper had been accepted, a related paper or epistasis in the vesicular stomatitis virus by Sanjuán *et al.* appeared (*29*).

References and Notes

1. J. Maynard Smith, Ed., *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1976).
2. R. E. Michod, B. R. Levin, Eds., *The Evolution of Sex* (Sinauer Press, Sunderland, MA, 1988).
3. N. H. Barton, B. Charlesworth, *Science* **281**, 1986 (1998).
4. S. P. Otto, T. Lenormand, *Nature Rev. Genet.* **3**, 252 (2002).
5. M. W. Feldman, F. B. Christiansen, L. D. Brooks, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 4838 (1980).
6. A. S. Kondrashov, *Nature* **336**, 435 (1988).
7. N. H. Barton, *Genet. Res.* **65**, 123 (1995).
8. A. S. Kondrashov, *J. Hered.* **84**, 372 (1993).
9. W. R. Rice, *Nature Rev. Genet.* **3**, 241 (2002).
10. S. F. Elena, R. E. Lenski, *Nature* **390**, 395 (1997).
11. S. F. Elena, *J. Mol. Evol.* **49**, 703 (1999).
12. M. de la Pena, S. F. Elena, A. Moya, *Evolution* **54**, 686 (2000).
13. A. D. Peters, P. D. Keightley, *Genetics* **156**, 1635 (2000).
14. M. C. Whitlock, D. Bourguet, *Evolution* **54**, 1654 (2000).
15. D. M. Wloch, K. Szafraniec, R. H. Borts, R. Korona, *Genetics* **159**, 441 (2001).
16. C. L. Burch, P. E. Turner, K. A. Hanley, *J. Evol. Biol.* **16**, 1223 (2003).
17. A. Rivero, F. Balloux, S. A. West, *Evolution* **57**, 1698 (2003).
18. C. L. Burch, L. Chao, *Genetics* **167**, 559 (2004).
19. H. M. Temin, *Trends Genet.* **7**, 71 (1991).
20. A. Jung *et al.*, *Nature* **418**, 144 (2002).
21. D. N. Levy, G. M. Aldrovandi, O. Kutsch, G. M. Shaw, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4204 (2004).
22. Materials and methods are available as supporting material on *Science* Online.
23. E. Szathmary, *Genetics* **133**, 127 (1993).
24. M. T. Bretscher, C. L. Althaus, V. Müller, S. Bonhoeffer, *BioEssays* **26**, 180 (2004).
25. S. P. Otto, M. W. Feldman, *Theor. Popul. Biol.* **51**, 134 (1997).
26. M. M. Iles, K. Walters, C. Cannings, *Genetics* **165**, 2249 (2003).
27. J. M. Coffin, *J. Gen. Virol.* **42**, 1 (1979).
28. The R Project for Statistical Computing (www.r-project.org).
29. R. Sanjuán, A. Moya, S. F. Elena, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15376 (2004).

30. We acknowledge the ViroLogic Clinical Reference Laboratory for their efforts in generating the replication capacity and PR/RT sequence data and C. Althaus, V. Müller, S. Otto, T. Pfeiffer, and M. Salathé for valuable comments. S.B. acknowledges the Swiss National Science Foundation for financial support. The development of the drug resistance database and replication capacity assay used in this study was funded in part by Small Business Innovative Research–Advanced Technology Grants from the National Institute of Allergy and Infectious Diseases, NIH (R43 AI057068 and R43 AI050321). S.B. is an equity holder in ViroLogic, Inc.

Supporting Online Material

www.sciencemag.org/cgi/content/full/306/5701/1547/DC1
Materials and Methods
References

22 June 2004; accepted 14 October 2004

Femtomolar Sensitivity of a NO Sensor from *Clostridium botulinum*

Pierre Nioche,[1] Vladimir Berka,[2] Julia Vipond,[3] Nigel Minton,[4] Ah-Lim Tsai,[2] C. S. Raman[1]*

Nitric oxide (NO) is extremely toxic to *Clostridium botulinum*, but its molecular targets are unknown. Here, we identify a heme protein sensor (SONO) that displays femtomolar affinity for NO. The crystal structure of the SONO heme domain reveals a previously undescribed fold and a strategically placed tyrosine residue that modulates heme-nitrosyl coordination. Furthermore, the domain architecture of a SONO ortholog cloned from *Chlamydomonas reinhardtii* indicates that NO signaling through cyclic guanosine monophosphate arose before the origin of multicellular eukaryotes. Our findings have broad implications for understanding bacterial responses to NO, as well as for the activation of mammalian NO-sensitive guanylyl cyclase.

Nitric oxide (NO) is a small, short-lived, and highly reactive gaseous molecule. In mammals, NO is biosynthesized from L-arginine by nitric oxide synthases, and it plays a key role in many and disparate cellular responses including host defense against microbial pathogens (*1*). Denitrifying bacteria generate NO (1 to 70 nM) by a unique mode of respiration in which nitrogen oxides (NO_3^-, NO_2^-, NO, and N_2O) are reduced to N_2 (*2*). Nondenitrifiers, like *Escherichia coli*, can produce NO during anaerobic nitrate respiration ($NO_3^- \rightarrow NO_2^- \rightarrow NH_4^+$), although at much reduced levels (*3*). *Clostridium botulinum*, a strict anaerobe, is neither a denitrifier nor capable of making NO via other mechanisms. It is the etiological agent of botulism and produces the most toxic substance known to humans [median lethal dose (LD_{50}) $\approx$ 0.2 ng per kilogram of body weight]. Since the late 1920s sodium nitrite, with NO as the antimicrobial principle, has been used to inhibit the growth of heat-resistant *C. botulinum* spores and toxin production in cured meats (*4*–*6*). However, the molecular strategies used by this bacterium to recognize and avoid NO in its native and host environments have remained elusive.

To identify candidate NO sensors in *C. botulinum*, we investigated the hypothesis (*7*, *8*) that a prokaryotic counterpart to the mammalian NO receptor, soluble guanylyl cyclase (sGC) (*9*, *10*), exists and that a bacterial NO sensor may communicate with the chemotaxis machinery. Thus, we screened the genome of *C. botulinum* (www.sanger.ac.uk/Projects/C_botulinum/) for an ortholog

[1]Structural Biology Research Center and Department of Biochemistry and Molecular Biology, [2]Division of Hematology, Internal Medicine, University of Texas Medical School, Houston, TX 77030, USA. [3]Health Protection Agency, Porton Down, Salisbury, Wiltshire SP4 0JG, UK. [4]Center of Biomolecular Sciences and Institute of Infection, Immunity and Inflammation, University of Nottingham, Nottingham NG7 2RD, UK.

*To whom correspondence should be addressed: E-mail: c.s.raman@uth.tmc.edu

of the NO-binding heme domain (*11*) of human sGC-β_1. We found a gene that encodes a two-domain protein (598 amino acids in length) in which the N-terminal domain shares 15% sequence identity with the first 186 amino acids of sGC-β_1 and the C-terminal domain is similar to that of a methyl-accepting chemotaxis protein (MCP). Because the N-terminal region of mammalian sGC is indispensable for NO binding, we named the *C. botulinum* protein SONO (sensor of NO). Reverse transcription–polymerase chain reaction analysis (fig. S1) (*12*) showed a transcript of the predicted size and confirmed that SONO was expressed in actively growing vegetative *C. botulinum* cells.

We overproduced SONO N-terminal heme domain (residues 1 to 186, CB-SONO$_{HD}$) in *E. coli*. Purified CB-SONO$_{HD}$ is red in color, and electronic spectroscopy revealed the presence of a b-type heme (*12*). Mass spectrometry confirmed the identity of the heme prosthetic group (iron protoporphyrin IX) (*13*). A heme protein has not been previously identified in *clostridia*. To assess the coordination structure of the heme in CB-SONO$_{HD}$, we recorded its electron paramagnetic resonance (EPR) spectra (*12*). The protein yielded a classic three-line spectrum (*14*) (Fig. 1A) diagnostic of a five-coordinate (5c) low-spin heme-nitrosyl, Fe(II)(NO), complex that is unaffected by exposure to air. The main spectral feature is the three-line hyperfine splitting centered at $g_z = 2.007$ with a separation of 1.7 mT. The splitting arises from the strong coupling between the unpaired electron spin and the nuclear spin ($I = 1$) of the ^{14}N atom of NO. It further indicates that the proximal Fe–histidine bond in the protein is readily broken. Such an air-stable heme-nitrosyl is unprecedented (*15–17*) [supporting online material (SOM) note 1]. We followed the aerobic decay of the three-line EPR signal of CB-SONO$_{HD}$ as a function of time. A half-life ($t_{1/2}$) of 70 ± 16 hours ($n = 3$) was obtained (fig. S2) and is equivalent to a dissociation rate constant (k_{off}) of $\approx 3 \times 10^{-6}$ s^{-1}. NO binding to SONO$_{HD}$ is diffusion limited [$k_{on} \approx 10^8$ M^{-1} s^{-1} (*18*)], and we obtain an equilibrium dissociation constant ($K_D = k_{off}/k_{on}$) of $\approx 30 \times 10^{-15}$ M. This value approaches that suggested for NO binding to four-coordinate model heme compounds (*15*), but has not been previously reported for Fe(II)-heme proteins (SOM note 2). The presence of an air-stable Fe(II)(NO) species indicates that molecular oxygen cannot react with this complex.

We used CB-SONO$_{HD}$ to quantify the amount of NO generated during aerobic growth of *E. coli*. *E. coli* are thought to generate NO only during anaerobic nitrate respiration (*3*) because NO production is undetectable during aerobic growth [as we confirmed with chemiluminescence-based NO analyzer (Eco Physics CLD88sp) with a sensitivity of <1 part per billion)]. However, the EPR spectra of intact *E. coli* cells aerobically expressing CB-SONO$_{HD}$ (Fig. 1B) resembled those of purified CB-SONO$_{HD}$, suggesting that the bound NO originated in the *E. coli* cytosol (SOM note 3). Cells that did not express CB-SONO$_{HD}$ are devoid of this signal. Quantitative EPR measurements (*12*) indicated that the *E. coli* cells produced (assuming a linear rate of formation) approximately one molecule of NO cell^{-1} min^{-1} (SOM note 4). The number of NO molecules bound to highly purified CB-SONO$_{HD}$ was in the same range as that measured from intact cells. Note, however, that one free molecule of NO per *E. coli*, with a cell volume of 10^{-15} liter (*19*), translates to a concentration of $\sim 10^{-9}$ M. It is conceivable that one or more NO molecules may persist within the cytoplasm of a denitrifying cell, but this would not be true for *E. coli*, which lacks the denitrification machinery (SOM note 5). Therefore, our ability to detect NO production in aerobically grown *E. coli* (*12*) is a direct consequence of CB-SONO$_{HD}$ functioning as a high-affinity sensor of NO.

We determined the crystal structure of a SONO ortholog (table S1) from the extremely thermophilic organism *Thermoanaerobacter tengcongensis*, strain MB4^T (*20*) (SOM note 6) [TT-SONO; 39% sequence identity with CB-SONO (fig. S3)] due to difficulties in growing crystals of CB-SONO$_{HD}$. The tertiary topology of TT-SONO heme domain (TT-SONO$_{HD}$) is characterized by a proximal domain with an αββαββ motif and an α-helical distal domain (fig. S4). This fold has not been described previously. The heme is sandwiched between the two domains (Fig. 2), and its iron atom is coordinated by His^{102}. Heme recognition is mediated by hydrogen bonds between the propionates and the strictly conserved side chains of Tyr^{131}, Ser^{133}, and Arg^{135} (figs. S3 and S5; SOM note 7). Numerous nonbonded contacts exist between the tetrapyrrole and the residues that form the binding pocket. With the exception of Tyr^{140}, the distal heme pocket is apolar and there are no polar atoms within 7 Å of the heme iron. During structure refinement, it became clear that a distal ligand was bound to the heme iron. The protein has a Soret maximum at 416 nm, which suggests the presence of a hexacoordinate Fe(II)(O_2) complex (SOM note 8). As a consequence, we modeled dioxygen into the density (Fig. 3). The chemical makeup of the TT-SONO$_{HD}$ distal heme pocket resembles that of the Thr-E11→Val variant (*21*) of *Cerebratulus lacteus* minihemoglobin (fig. S6). This protein also purifies as a high-affinity Fe(II)(O_2) complex and, like TT-SONO$_{HD}$, has a Tyr residue within hydrogen-bonding distance to the terminal oxygen (SOM note 9).

Although both TT-SONO$_{HD}$ and CB-SONO$_{HD}$ have a similar constellation of side chains in the distal heme pocket (fig. S6), we have been unable to generate a stable Fe(II)(O_2) complex with CB-SONO$_{HD}$. NO-stripped CB-SONO$_{HD}$ also readily oxidizes to its Fe(III) state. Like *C. botulinum*, *T. tengcongensis* is a strict anaerobe; however, it has an optimal growth temperature of 75°C (*20*). At 70°C, TT-SONO$_{HD}$ did not form a stable Fe(II)(O_2) complex, but instead readily oxidized to the Fe(III) state. Upon addition of exogenous NO, TT-SONO$_{HD}$ also

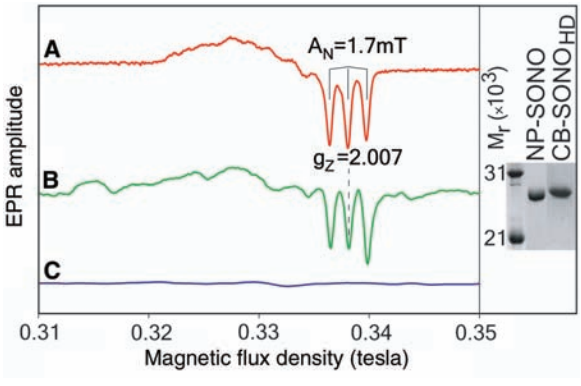

Fig. 1. EPR spectra of SONO heme domain (SONO$_{HD}$). (**A**) CB-SONO$_{HD}$ purifies as a heme-nitrosyl complex. (**B**) *E. coli* cells overproducing CB-SONO$_{HD}$. (**C**) Purified NP-SONO. (Sidebar) Coomassie-stained SDS-polyacrylamide gel. The relative molecular sizes of NP- and CB-SONO$_{HD}$ are 22,482 and 22,630, respectively.

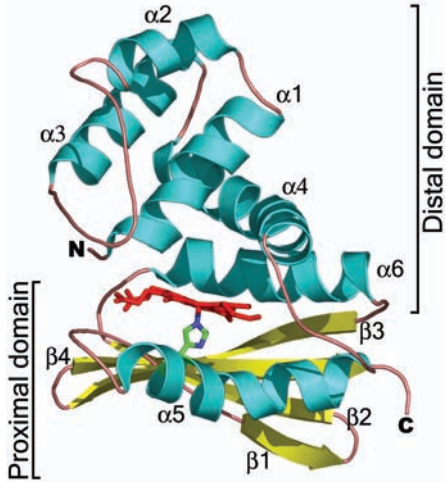

Fig. 2. Three-dimensional structure of TT-SONO$_{HD}$. Ribbon diagram depicts residues 1 to 181 of the protein. The heme prosthetic group (red) and the histidine coordinating its iron are also shown.

generated a five-coordinate Fe(II)(NO) complex at 70°C. In contrast, a six-coordinate (6c) Fe(II)(NO) complex was formed at 25°C, and we found no evidence for a heme-nitrosyl complex in purified TT-$SONO_{HD}$.

To assess the role of electrostatic interactions (*22*, *23*) in stabilizing the bound ligand, we replaced the distal Tyr^{139} with a Phe in CB-$SONO_{HD}$ (fig. S6). This substitution resulted in a protein that was unable to generate a 5c-heme-nitrosyl, but instead produced a stable 6c-Fe(II)(NO) complex (fig. S7). The ability of an amino acid side chain in the distal pocket to modulate the bond strength of the proximal Fe-His linkage is unparalleled in heme protein research. Altogether, these results identify that NO binds on the distal side in CB-$SONO_{HD}$ and suggest that electrostatic interaction with Tyr^{139} is necessary for generating a neet 5c-Fe(II)(NO) complex and subsequently breaking the Fe-His bond (SOM note 10).

There appears to be a strong evolutionary pressure to modulate the NO sensitivity and specificity for SONO. For example, the rate of NO dissociation from CB-$SONO_{HD}$ is at least two orders of magnitude slower than that of sGC (*24*). To evaluate the structural basis for such variation, we used TT-$SONO_{HD}$ as a template to generate homology models of $SONO_{HD}$ from human sGC-β_1 and *Nostoc punctiforme* (filamentous cyanobacterium) [(NP-SONO; the full-length protein, comprising amino acids 1 to 183, represents the heme domain and shares 40% sequence identity with human sGC-β_1 (fig. S3)]. Unlike CB-$SONO_{HD}$ and TT-$SONO_{HD}$, neither human sGC-β_1 nor NP-SONO incorporates a Tyr side chain for interacting with the bound ligand (fig. S6). Furthermore, the distal heme pocket lacks a polar atom within 7 Å of the heme iron. Consistent with this, purified NP-SONO is characterized by a Soret maximum at 431 nm and shows no apparent EPR signal (Fig. 1C) because it contains a Fe(II) five-coordinate high-spin heme (SOM note 11). Most heme proteins, except sGC (*25*, *26*), can maintain this state only under anaerobic conditions, because oxygen will readily bind to the sixth coordination position. However, the Fe(II)-heme of NP-SONO remains five-coordinate even after several weeks of aerobic storage at 4°C. Because fossil data show that the *Nostoc* branch of the cyanobacterial tree diverged by 2.1 billion years ago (*27*), total exclusion of oxygen from the coordination sphere of SONO, concomitant with a diminution of NO sensitivity, may have occurred before the origin of unicellular eukaryotes.

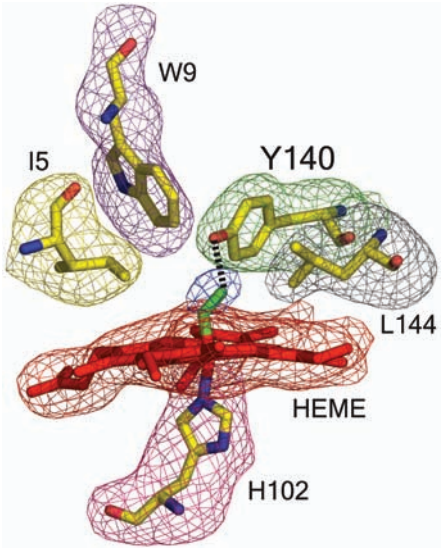

Fig. 3. Omit difference electron density map ($F_{obs} - F_{calc}$) of the heme pocket. Maps are shown at 3σ level. Oxygen ligand density is in blue. Dotted line indicates that the phenolic –OH of Tyr^{140} is within hydrogen-bonding distance of the oxygen ligand.

To better delimit when during evolution the functional transition from SONO to sGC may have taken place, we examined the genome of *Chlamydomonas reinhardtii*—the eukaryotic unicellular green alga whose chloroplasts are derived from cyanobacteria (*28*). We found a *Chlamydomonas* expressed sequence tag whose product shares similarity with NP-SONO. We cloned the gene, and its product represents a full-length sGC-β with close resemblance to human sGC-β_1 (fig. S8; SOM note 12). A sGC ortholog has not been previously reported in a chloroplast-containing organism. The two proteins share similar nonpolar distal heme pockets, and both contain a Phe side chain in close proximity to the heme iron. Thus, eukaryotic sGC-β_1 may have evolved by the fusion of bacterial $SONO_{HD}$ and a cyclase, thereby coupling NO binding to cyclic guanosine monophosphate production before the origin of animals.

In summary, we have uncovered a heme protein sensor that may be used by *C. botulinum* to recognize the presence of NO and mediate a phobic response as a protective mechanism (SOM note 13). Although *C. botulinum* is a motile organism, we have no direct evidence that it moves away from a source of noxious NO. However, two independent observations suggest that CB-SONO can sense and transduce signals to the chemotaxis machinery. First, the two-domain architecture of CB-SONO is reminiscent of the soluble aerotransducer HemAT (*29*) in which a myoglobin-like heme domain is used to sense O_2 and the methyl-accepting chemotaxis protein domain helps with initiating bacterial aerotaxis. Second, the gene encoding SONO in *Vibrio cholerae* is repressed by about eightfold alongside other chemotaxis genes in vibrios shed by cholera-infected individuals (*30*, *31*). Furthermore, we draw a parallel between the suggested taxis role for SONO in bacteria and the ability of sGC to confer chemoattractive responses to dendrites and axons (*32*). Our structural findings also provide mechanistic insights into how NO activates sGC in mammalian systems (figs. S9 and S10; SOM note 14). Taken together, our studies indicate that $SONO_{HD}$, with alterations to its NO sensitivity, may have been coopted to serve as a NO-activated molecular switch for regulating distinct cellular processes.

Note added in proof: During revision of this manuscript, structural characterization (*33*, *34*) of a similar protein was reported.

References and Notes

1. L. J. Ignarro, in *Nitric Oxide: Biology and Pathobiology*, L. Ignarro, Ed. (Academic Press, San Diego, CA, 2000), pp. 3–19.
2. W. G. Zumft, *J. Mol. Microbiol. Biotechnol.* **4**, 277 (2002).
3. H. Corker, R. K. Poole, *J. Biol. Chem.* **278**, 31584 (2003).
4. D. Reddy, J. R. Lancaster Jr., D. P. Cornforth, *Science* **221**, 769 (1983).
5. R. Cammack *et al.*, *Biochim. Biophys. Acta* **1411**, 475 (1999).
6. P. M. Davidson, in *Food Microbiology: Fundamentals and Frontiers*, M. P. Doyle, L. R. Beuchat, T. J. Montville, Eds. (American Society for Microbiology, Washington, DC, 2001), pp. 593–627.
7. R. J. P. Williams, *Biochim. Biophys. Acta* **1058**, 71 (1991).
8. S. J. Ferguson, *Biochim. Biophys. Acta* **1058**, 17 (1991).
9. F. Murad, *Angew. Chem. Int. Ed. Engl.* **38**, 1856 (1999).
10. D. Koesling, M. Russwurm, E. Mergia, F. Mullershausen, A. Friebe, *Neurochem. Int.* **45**, 813 (2004).
11. B. Wedel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2592 (1994).
12. Materials and methods are available as supporting material on *Science* Online.
13. R. Timkovich, personal communication.
14. T. Yonetani, A. Tsuneshige, Y. Zhou, X. Chen, *J. Biol. Chem.* **273**, 20323 (1998).
15. T. G. Traylor, V. S. Sharma, *Biochemistry* **31**, 2847 (1992).
16. P. C. Ford, I. M. Lorkovic, *Chem. Rev.* **102**, 993 (2002).
17. G. R. Wyllie, W. R. Scheidt, *Chem. Rev.* **102**, 1067 (2002).
18. J. S. Olson, personal communication.
19. C. E. Outten, T. V. O'Halloran, *Science* **292**, 2488 (2001).
20. Y. Xue, Y. Xu, Y. Liu, Y. Ma, P. Zhou, *Int. J. Syst. Evol. Microbiol.* **51**, 1335 (2001).
21. A. Pesce *et al.*, *J. Biol. Chem.* **279**, 33662 (2004).
22. G. N. Phillips Jr., M. L. Teodoro, T. Li, B. Smith, J. S. Olson, *J. Phys. Chem.* **103**, 8817 (1999).
23. M. R. Thomas, D. Brown, S. Franzen, S. G. Boxer, *Biochemistry* **40**, 15047 (2001).
24. V. G. Kharitonov, V. S. Sharma, D. Magde, D. Koesling, *Biochemistry* **36**, 6814 (1997).
25. R. Gerzer, E. Böhme, F. Hofmann, G. Schultz, *FEBS Lett.* **132**, 71 (1981).
26. J. R. Stone, M. A. Marletta, *Biochemistry* **33**, 5636 (1994).
27. A. H. Knoll, *Science* **285**, 1025 (1999).
28. W. Martin *et al.*, *Nature* **393**, 162 (1998).
29. S. Hou *et al.*, *Nature* **403**, 540 (2000).
30. D. S. Merrell *et al.*, *Nature* **417**, 642 (2002).
31. A. Camilli, personal communication.
32. F. Polleux, T. Morrow, A. Ghosh, *Nature* **404**, 567 (2000).
33. P. Pellicena *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12854 (2004).
34. D. S. Karow *et al.*, *Biochemistry* **43**, 10203 (2004).
35. C.S.R. dedicates this paper to M. Perutz and T. Traylor. C.S.R. is grateful to F. Rudolph and Rice University for providing laboratory space during the 2001 Houston flood; S. Boxer, J. Eraso, P. Ford, A. Grossman, J. Groves, M. Ikeda-Saito, J. Linder, W. Martin, J. Meeks, J. Olson, T. Poulos, F. Robb, J. Schrager, J. Schultz, J. Shapleigh, T. Spiro, S. Veeraraghavan, F.A. Walker, and W. Zumft for stimulating discussions; A. Camilli, J. Kaper, J. Olson, N. Stokes, R. Timkovich, T. Tomita, and B. Weiss for communicating unpublished results; C. Churcher, E. Harris, J. Kropat, T. McKarns, J. Meeks, S. Merchant, F. Murad, J. Parkhill, I. Scott, Eco Physics, Japanese Collection of Microorganisms, and Kazusa DNA Research Institute for access to strains, genomic DNA, RNA, protocols, and instrumentation; G. Bricogne for providing BUSTER and C. Vonrhein for assistance with using the program; and the Advanced Light Source (beamline 8.3.1, J. Holton, J. Tanamachi), Stanford

Synchrotron Radiation Laboratories (beamline 9.2, M. Soltis), and European Synchrotron Radiation Facility (beamline ID29, W. Shepard) for beam time and help with data collection. This work is supported by the Pew Charitable Trusts through a Pew Scholar Award (C.S.R.), The Robert A. Welch Foundation grant AU-1524 (C.S.R.), NIH grant R01 AI054444 (C.S.R.), and Association pour la Recherche sur le Cancer (P.N.).

The GenBank accession number for *C. reinhardtii* sGC-β cloned in this work is AY343540. Coordinates and structure factors are available from the RCSB Protein Data Bank under accession code 1XBN.

Supporting Online Material
www.sciencemag.org/cgi/content/full/1103596/DC1
Materials and Methods
Figs S1 to S10
Table S1
References and Notes

3 August 2004; accepted 27 September 2004
Published online 7 October 2004;
10.1126/science.1103596
Include this information when citing this paper.

Compensated Deleterious Mutations in Insect Genomes

Rob J. Kulathinal,[1] Brian R. Bettencourt,[2] Daniel L. Hartl[1]*

Relatively little is known about the importance of amino acid interactions in protein and phenotypic evolution. Here we examine whether mutations that are pathogenic in *Drosophila melanogaster* become fixed via epistasis in other Dipteran genomes. Overall divergence at pathogenic amino acid sites is reduced. However, ~10% of the substitutions at these sites carry the exact same pathogenic amino acid found in *D. melanogaster* mutants. Hence compensatory mutation(s) must have evolved. Surprisingly, the fraction 10% is not affected by phylogenetic distance. These results support a selection-driven process that allows compensated amino acid substitutions to become rapidly fixed in taxa with large populations.

By mapping sequence space onto a fitness surface, the "fitness landscape" provides a powerful metaphor to understand how proteins evolve in populations. Sequence evolution may be visualized to traverse fitness peaks and valleys as certain sequence combinations are deleterious, advantageous, or neutral to an organism's overall reproductive success. Using this landscape, Sewall Wright (*1*) championed the view that evolution occurs via epistatic or "coadapted" genetic interactions (*2*). In contrast, R.A. Fisher envisioned selection primarily acting on additive effects of individual loci (*3*). Whether epistatic interactions play an important role in the evolutionary trajectory of proteins—particularly in species with large effective population sizes—remains an open question.

The recently completed genomic sequence of *D. pseudoobscura* (*4*) offers a unique opportunity to approach this question by integrating the extensive mutant phenotypic information from *D. melanogaster* with a full set of orthologous gene sequences. These two fly species diverged between 40 and 50 million years ago (mya) and contain ample divergence information for comparative analyses. For nearly a century, *D. melanogaster* has been the target of extensive mutagenic screens, and a curated database of characterized mutant phenotypes and their corresponding molecular etiologies is freely accessible in the public domain (*5*).

We analyzed single amino acid residues that, when mutated in *D. melanogaster*, cause a drastic fitness loss yet appear as the wild-type amino acid at its homologous site (termed "index site") in the *D. pseudoobscura* protein. For such pathogenic substitutions to become fixed in another species, second-site or compensatory mutations must have coevolved (*6*). The number of compensated pathogenic deviations—or CPDs, using the parlance of Kondrashov *et al*. (*6*)—was compared to the number of index-site substitutions in *D. pseudoobscura* that contain an amino acid other than the one known to be pathogenic in *D. melanogaster* (Fig. 1). "%CPD" refers to the fraction of CPDs among substituted index sites. We sorted through all phenotypic mutants available from FlyBase (version 3.1) with the assumption that none of these mutants could persist in a natural population. From 35,311 "Gene" entries in FlyBase, we found 2245 single-site amino acid mutations that lead to a defined mutant phenotype, representing 525 unique genes. Most genes have between one and three different mutations; thus, our sample is not biased toward particular loci.

Of these genes, 475 had unambiguously aligned orthologs in *D. pseudoobscura* (*4*) representing 328,060 aligned amino acid sites. Overall, 77.8% of all amino acid sites were conserved between the two species. All instances in which a pathogenic amino acid site in *D. melanogaster* was present in the wild-type *D. pseudoobscura* protein were tabulated. To ensure site-specific orthology, we counted the number of conserved sites among 10 flanking amino acids on each side of the index site. Single-site insertion/deletion mutations were counted as one amino acid change. When a 50% identity criterion among flanking sites was used, we found that among 1527 amino acid sites causing a phenotypic mutation in *D. melanogaster*, only 64 had substitutions in the *D. pseudoobscura* ortholog. Thus, substitutions among index sites were significantly less frequent than were random substitutions (95.8% versus 77.8%, $P < 0.0001$), an observation also reported in humans and mice (*7*), which suggests that selective constraints are maintained over time. Surprisingly, six of these index sites, or 1 in 10 substitutions at pathogenic sites, contained exactly the same amino acid that causes a deleterious phenotypic change in *D. melanogaster* [%CPD ~ 10% (Table 1)]. Similarly, using a 75% identity criterion (close to the average protein divergence estimated between these species), we found four CPDs among 31 index sites.

The same analysis was then applied to the more distantly related *Anopheles gambiae* genome (divergence ~250 mya). We aligned 210,778 orthologous amino acid sites (insertions and deletions ignored) from 317 proteins, with an overall 52.4% identity. When a 50% flanking region cutoff was used (corresponding approximately to the average amino acid divergence between these species), there were 77 substitutions among 784

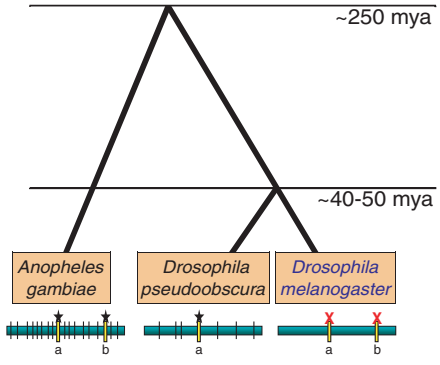

Fig. 1. Phylogenetic relationship of *D. melanogaster*, *D. pseudoobscura*, and *A. gambiae* and identification of compensated deleterious mutations. *D. melanogaster* is the reference species with site-specific mutational data on phenotypic mutants. Sites possessing amino acid residues that cause phenotypic deviations in *D. melanogaster* are called index sites (yellow bars). Other substituted sites are not informative (black vertical bars). We concentrate on index sites that contain substitutions (indicated as a and b) in at least one of the other species (denoted by a star). Index-site substitutions are of two types: the exact same pathogenic amino acid (CPDs) or other, non-exact amino acid substitutions.

[1]Department of Organismic and Evolutionary Biology, [2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: dhartl@oeb.harvard.edu