

# The Structural Basis of Molecular Adaptation

G. Brian Golding\* and Antony M. Dean†

\*Department of Biology, McMaster University, Hamilton, Ontario, Canada; and †Department of Biological Chemistry, Finch University of Health Sciences/The Chicago Medical School, North Chicago

The study of molecular adaptation has long been fraught with difficulties, not the least of which is identifying out of hundreds of amino acid replacements those few directly responsible for major adaptations. Six studies are used to illustrate how phylogenies, site-directed mutagenesis, and a knowledge of protein structure combine to provide much deeper insights into the adaptive process than has hitherto been possible. Ancient genes can be reconstructed, and the phenotypes can be compared to modern proteins. Out of hundreds of amino acid replacements accumulated over billions of years those few responsible for discriminating between alternative substrates are identified. An amino acid replacement of modest effect at the molecular level causes a dramatic expansion in an ecological niche. These and other topics are creating the emerging field of “paleomolecular biochemistry.”

## Introduction

The neutral theory of molecular evolution (Kimura 1968a, 1968b, 1983) proposes that most sequence changes in nucleic acids and proteins are selectively equivalent. Although still controversial, this theory nevertheless highlights the need to convincingly demonstrate the action of natural selection at the molecular level. Yet this was to prove so challenging that a decade later, Lewontin (1979) lamented, “it has proved remarkably difficult to get compelling evidence for changes in enzymes brought about by selection, not to speak of adaptive changes . . .”

More recently, we have witnessed the arrival of new and powerful molecular tools. These have provided us with an unprecedented ability to determine the nucleotide sequence of any given stretch of DNA from any given individual from any given species. The resulting surveys of molecular variation have revolutionized fields as diverse as taxonomy and systematics, origins, biogeography, population structure, anthropology, and behavioral ecology. Our understanding of the interplay between various evolutionary forces and constraints has improved immeasurably. However, while the telltale signatures of selection have been detected in this cataloging of genic variation (reviewed in Golding 1994), any detailed understanding of adaptive change requires more information than raw sequence and phylogeny alone provide. It requires phenotypes.

Physiological population geneticists (e.g., Koehn and Hilbish 1987; Powers et al. 1991; Watt 1991) have largely eschewed phylogenetics in favor of stressing the biochemical basis of molecular adaptation. Their approach emphasizes the importance of phenotypes with a strong genetic component. These undertakings have contributed greatly to our understanding of selection in natural populations. If they tend to be limited to studying balanced polymorphisms, they are at least complemented by laboratory studies of microbial populations, in which strong directional selection generally prevails.

Key words: molecular adaptation, amino acid replacements, protein structure, phylogeny.

Address for correspondence and reprints: Antony M. Dean, Department of Biological Chemistry, FUHS/CMS, 3333 Green Bay Road, North Chicago, Illinois 60064-3095. E-mail: deana@mis.finchcms.edu.

*Mol. Biol. Evol.* 15(4):355–369. 1998

© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Here, simplified reproducible environments provide the necessary control to tease apart underlying molecular mechanisms (e.g., Dykhuizen and Dean 1994; Rosenzweig et al. 1994; Krishnan, Hall, and Sinnott 1995). Both approaches, in the field and on the petri dish, are concerned with current selection. Neither addresses the problem of studying ancient adaptations. Phylogenies are needed for that.

And so, three decades later, the field of molecular adaptation emerges cleaved between phylogenetics and physiological genetics, between history and mechanism, between pattern and process. That this is the case is hardly surprising, as a brief reflection quickly exposes the difficulty in their unification. A large number of sequence differences accumulate over evolutionary time, but not all need be adaptive. Even with relatively recent selective events, hitchhiking often ensures that additional replacements tag along with the selective sweep, making it difficult, even impossible, to identify the adaptive replacement by phylogenetic means alone. Together, phylogenetic and phenotypic evidence are insufficient for understanding molecular adaptation—we still need to identify which of many replacements are directly responsible for adaptive changes.

In the inaugural article for *Molecular Biology and Evolution*, “Species Adaptation in a Protein Molecule,” Perutz (1983) brought a fundamentally different perspective to the matter. He described how the function of hemoglobin relates to its three-dimensional structure. Comparing hemoglobins from various species, and using his intimate knowledge of structure–function relations, he chose from a myriad of amino acid replacements those few most likely to be responsible for observed functional differences. In so doing he had attempted to unite form and function and phylogeny to glean insight into the process of molecular adaptation. The one limitation Perutz had was in testing his deductions. In the early to mid-1980s the only tool available to him was the comparative method. Now, in the mid- to late 1990s, site-directed mutagenesis can be used to engineer proteins. The *in vitro* functional effects of each and every amino acid replacement within a phylogeny can now be determined with exquisite precision.

Here, we expound the view that molecular anatomy is just as key to understanding molecular adaptation as

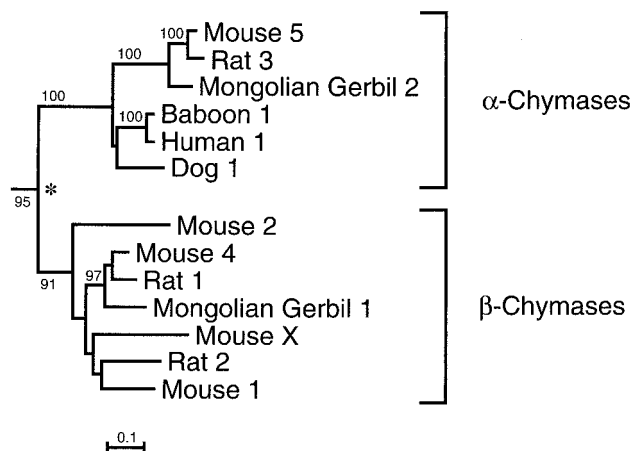


FIG. 1.—Maximum-parsimony phylogeny of the chymases rooted using granzymes and cathepsins (not shown). The asterisk denotes the position of the reconstructed ancestral chymase sequence (modified with additional taxa from Chandrasekharan et al. 1996). Numbers refer to the percentage support from 1,000 bootstrapped trees. The tree was rooted using a large number of related serine proteases: kallikreins, granzymes, killer cell proteases, and cathepsins (not shown).

phylogeny and physiological ecology (after all, fossil anatomy is key to understanding ancient morphological adaptations). We avoid summarizing the results of an exhaustive literature search in favor of a didactic approach, choosing six studies that we feel illustrate the range of evolutionary questions that can be addressed using protein engineering and comparative molecular anatomy. Not only are the techniques available to dissect the very nature of selective changes, but their history can be explored as well. As these examples illustrate, different molecules respond in different ways to selective pressures—a wealth of evolutionary pathways that is only beginning to be uncovered.

## Six Studies

### Chymase: An Ancient Phenotype Reconstructed

Chymases (mast cell proteases) are a class of serine proteases related to chymotrypsin that hydrolyze the Phe8-His9 bond of angiotensin I to produce angiotensin II, a potent vasoconstrictor hormone. Chymases fall into two families,  $\alpha$  and  $\beta$  (fig. 1). Primate  $\alpha$ -chymases are

**Table 1**  
Kinetic Parameters of Modern and Ancestral Chymases

ENZYME	KINETIC PERFORMANCE ( $k_{cat}/K_m$ ) ( $\mu\text{M}^{-1}\text{s}^{-1}$ )	
	Ang I	Ang II
Human $\alpha$ -chymase . . . . .	3.6	ND <sup>b</sup>
Rat $\beta$ -chymase-1 . . . . .	ND <sup>a</sup>	0.085
Ancestral chymase . . . . .	4.3	ND <sup>b</sup>

<sup>a</sup> Not detectable because both peptide bonds are hydrolyzed simultaneously.

<sup>b</sup> Not detectable because the  $k_{cat} < 0.01\%$  of that of human  $\alpha$ -chymase.

highly specific and only hydrolyze the Phe8-His9 bond (table 1). Rat  $\beta$ -chymase-1, like chymotrypsin, is less specific and further hydrolyzes the hormone by attacking its Tyr4-Ile5 bond. Thus, angiotensin II is formed by  $\alpha$ -chymase and degraded by  $\beta$ -chymase.

Chandrasekharan et al. (1996) constructed a phylogenetic tree to determine whether the narrow specificity of primate  $\alpha$ -chymase is a derived or an ancestral state. Maximum parsimony was used to construct a phylogeny of four  $\alpha$ -chymases (from human, baboon, dog, and mouse) and six  $\beta$ -chymases (from mouse and rat) which was rooted with a large number of related serine proteases of diverse function (fig. 1). Unfortunately, the phylogeny gave no clue as to the specificity of the ancestral chymase. The sequence of the ancestral chymase protein was inferred by maximum parsimony. Assignments at 15 sites in the ancestral sequence were ambiguous: 8 residues were assigned to adjust the net charge to +18 and preserve the two charge clusters characteristic of chymases, while the remaining 7 residues were determined arbitrarily using PAUP. Molecular modeling suggests that these ambiguous replacements are unlikely to have a marked influence on specificity because none are found in the active site cleft of the protease (fig. 2). Molecular modeling also reveals that the active site cleft of the ancestral enzyme is composed of a mosaic of  $\alpha$ - and  $\beta$ -chymase residues. Hence, phylogenetic analysis and molecular modeling were insufficient to infer the range of specificity of the ancestral chymase.

Chandrasekharan et al. (1996) reconstructed the ancestral enzyme. So different from modern sequences (between 52 and 77 out of 226 residues) was the inferred

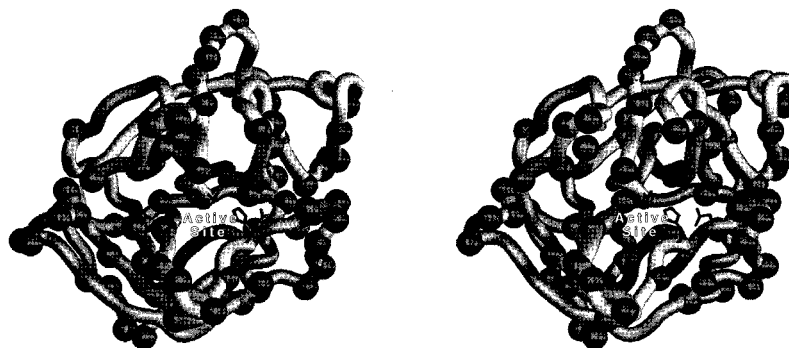


FIG. 2.—Stereo view of the structure of rat chymase-2 showing the positions of the 66 amino acid replacements (spheres) that occurred during divergence from the ancestral enzyme. Many of the replacements in the active-site cleft close to the catalytic triad (bonds) undoubtedly influence specificity, while more distant replacements are expected to have little or no effect.

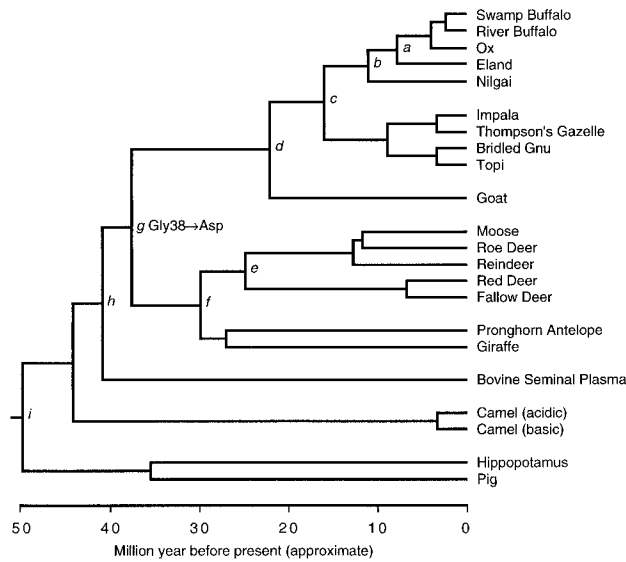


FIG. 3.—Phylogeny of artiodactyl RNase superfamily (after Jermann et al. 1995) showing the position of the Gly38→Asp replacement that decreases activity toward double-stranded RNA. Italicized letters denote nodes used for reconstructing ancient enzymes.

ancestral sequence that its entire gene was synthesized chemically. Nevertheless, the reconstructed chymase is highly active, efficiently cleaving angiotensin I to form angiotensin II (table 1). It does not cleave angiotensin II at the Tyr4-Ile5 bond, however. This experiment demonstrates that the narrow specificity of primate  $\alpha$ -chymase is the ancestral state, and the broader specificity of the rat  $\beta$ -chymase is the derived state.

The probability that the exact ancestral sequence was reconstructed is rather small because of errors accumulated across so many sites. On the other hand, and as we shall illustrate in later examples, only a small number of replacements need confer a change in specificity. Hence, the likelihood of reconstructing an ancestral phenotype is greater than the likelihood of accurately reconstructing an ancestral sequence. Exactly when the loss of angiotensin-II-forming activity occurred and which replacements were responsible have yet to be determined. Nevertheless, these results demonstrate the power of combining phylogenetic inference in reconstructing ancient phenotypes with protein engineering, and provide an interesting example of evolutionary degeneration—a specialized enzyme evolving a broader substrate specificity.

#### RNase A: Replacements with Functional Effects Identified

Ribonucleases hydrolyze the phosphodiester bonds of RNA. Encoded by an extensive multigene family that arose through gene duplication and divergence, they are involved in diverse cellular functions, from neurotoxicity to endothelial-cell-stimulatory activity. Indeed, repaired pseudogenes derived from this family appear to be rapidly evolving new functions (Trabesinger-Ruef et al. 1996).

Jermann et al. (1995) analyzed the evolutionary history of RNase A, a digestive enzyme secreted by the

**Table 2**  
Kinetic Parameters and Thermal Transition Temperatures of Bovine and Reconstructed Ancestral Ribonucleases (after Jermann et al. 1995)

RNASE	ANCESTOR OF:	KINETIC PERFORMANCE		
		$(k_{cat}/K_m)_{Poly(A)}$ (relative to bovine)	Poly(A)·Poly(U) (relative to bovine) ( $\pm 0.5^\circ\text{C}$ )	$T_m$
Bovine		1.0	1.0	59.3
<i>a</i> . . . . .	Ox, buffalo, eland	1.2	1.4	60.6
<i>b</i> . . . . .	<i>a</i> and nilgai	1.2	1.0	61.0
<i>c</i> . . . . .	<i>b</i> and gazelles	0.9	0.8	60.7
<i>d</i> . . . . .	Bovids	0.8	0.9	58.4
<i>e</i> . . . . .	Deer	0.7	1.0	61.1
<i>f</i> . . . . .	Deer, pronghorn, giraffe	0.7	1.0	58.6
<i>g</i> . . . . .	Pecora	0.9	1.0	59.1
<i>h</i> . . . . .	<i>g</i> and seminal RNase	1.1	5.2	58.9
<i>i</i> . . . . .	Ruminata	0.9	5.0	58.2
<i>j</i> . . . . .	Artiodactyla	0.7	4.6	56.5

pancreas, and which is particularly abundant in the guts of a number of mammalian taxa. They used a parsimony algorithm to infer the ancestral sequences in a phylogeny of 21 species of artiodactyls (fig. 3) determined by Beintema et al. (1986). Site-directed mutagenesis was used to construct 13 of the ancestral sequences, each of which was expressed in *Escherichia coli*, the enzyme was purified, and its catalytic properties were determined. Benner et al. (1996) named this approach “paleomolecular biochemistry.”

The kinetic properties of the reconstructed ancestral enzymes are similar to those of extant RNases (table 2). This is not surprising from a structural standpoint, because all of the replacements lie on the surface of the enzyme at least 5 Å from the active site—positions that are expected to least influence function (fig. 4). Nevertheless, least influence does not mean no influence: the evolved enzymes of ruminant artiodactyls are more stable to thermal denaturation, are less susceptible to proteolysis, and, while they remain active against single-stranded RNA, are fivefold less active toward double-stranded RNA. Further experiments established that a single amino acid replacement, Gly38→Asp, accounts for most of the change in activity toward double-stranded RNA.

The reconstructed ancestral sequences reveal that the functional changes in RNase A occurred 40 MYA, around the time foregut rumination evolved. That brain and seminal plasma RNases also diverged at this time suggests an ancient gene duplication event followed by divergence and functional specialization. However, whether any of the replacements in pancreatic RNase A were subject to natural selection, or, for that matter, whether any were selectively neutral, is not known. The Gly38→Asp replacement might be neutral if ruminants no longer need the double-stranded RNA activity in an enzyme specialized for the foregut environment. Alternatively, a possible adaptive role is suggested by this same replacement occurring independently in the hip-

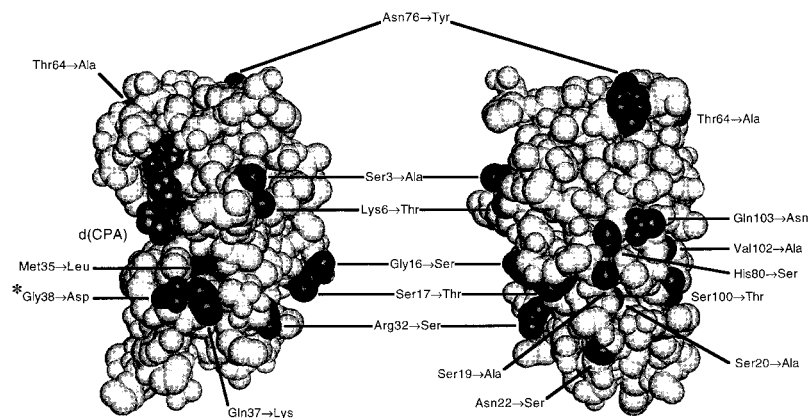
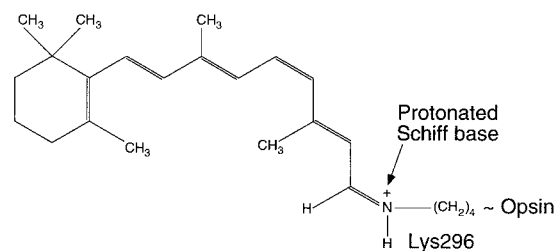


FIG. 4.—The van der Waals surface of bovine RNase A showing the positions of the amino acid replacements (black), from the most ancient RNase (node *i* in fig. 3) reconstructed by Jermann et al. (1995) to the modern ox, with respect to d(CPA) (gray) bound in the active site. All amino acid replacements are at the enzyme surface with the exception of Met35→Leu, which is partially buried. All amino acid replacements are at least 5 Å from the active site, including the Gly38→Asp replacement (asterisk) that causes the decrease in activity toward double-stranded RNA.

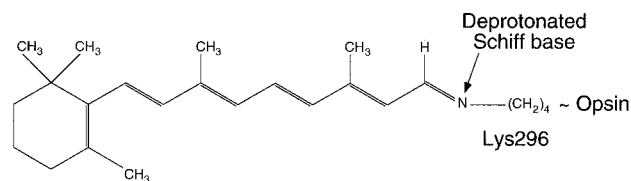
popotamus, which, although it lacks true foregut rumination, does have a complex forestomach.

#### Opsins: Eyeing Ancient Adaptations

The retina of the eye contains the visual pigments necessary for sight. These consist of a chromophore, usually 11-*cis*-retinal (fig. 5), which lies in a pocket at the center of a transmembrane protein called an opsin. Human rhodopsin absorbs light around a  $\lambda_{\max}$  of 495



11-*cis*-retinal in an unenergized opsin



all-*trans*-retinal in metarhodopsin II

FIG. 5.—The retinal chromophore of visual pigments. The first step in vision is a photon energized isomerization of the 11-*cis*-retinal prosthetic group (attached via a protonated Schiff base to Lys296 of the opsin) into the all-*trans* configuration. This produces mechanical work in the form of a 5-Å movement in the visual pigment. Converting mechanical work into an electrical impulse is initiated when the Schiff base linkage deprotonates, forming photoactivated metarhodopsin II, which, in turn, triggers an enzymatic cascade resulting in hyperpolarization of the plasma membrane and transmission of a nerve impulse to the visual cortex of the brain. A series of steps then returns the visual pigment to its original state while the membrane depolarizes. The whole cycle takes but a few seconds to complete (Stryer 1995, pp. 332–339).

nm to confer vision in dim light. Human color vision in bright light is conferred by three types of visual pigment with  $\lambda_{\max}$  values of 420 nm (blue), 530 nm (green), and 560 nm (red) (Nathans 1987). Amino acid replacements among the opsins, which are encoded by a multigene family that arose through gene duplication and divergence, modulate the  $\lambda_{\max}$  values of visual pigments by influencing the physical environment around the protonated Schiff base (fig. 5). Hence, the evolution of color vision is characterized by spectral tuning of visual pigments through amino acid replacements in related opsin proteins.

Phylogenetic analysis reveals that red-like opsins arose independently in fish and in reptiles and mammals following duplication of an ancestral opsin (fig. 6; Yokoyama 1997). The replacements most likely responsible for this spectral shift were identified as Ala180→Ser, Phe227→Tyr, and Ala285→Thr (Yokoyama and Yokoyama 1990): all three are near the chromophore and all three occurred independently in lineages leading to the red-like opsins of the Mexican cavefish (*Astyanax fasciatus*) and man. Site-directed mutagenesis has been used to replace the equivalent residues in bovine rhodopsin, causing increased  $\lambda_{\max}$  values of 2, 10, and 14 nm respectively (Chan, Lee, and Sakmar 1992). While these replacements explain the majority of the shift from green to red, engineering human green-like and red-like opsins reveals that four additional replacements (Tyr116→Ser, Thr230→Ile, Ser233→Ala, and Phe309→Tyr) make up the minor contribution necessary to obtain the full 30-nm shift (Asenjo, Rim, and Orian 1994).

Reconstructing ancestral sequences from a diverse range of opsin sequences indicates that the vertebrate ancestor had a single visual pigment absorbing around 530 nm (green) and that the first functional replacements to occur in land animals were Phe227→Tyr and Ala285→Thr (fig. 6; Yokoyama 1998). These two replacements account for much of the spectral shift. Ancestral animals had only two opsins (blue and red) and so, like many of their modern descendants, including

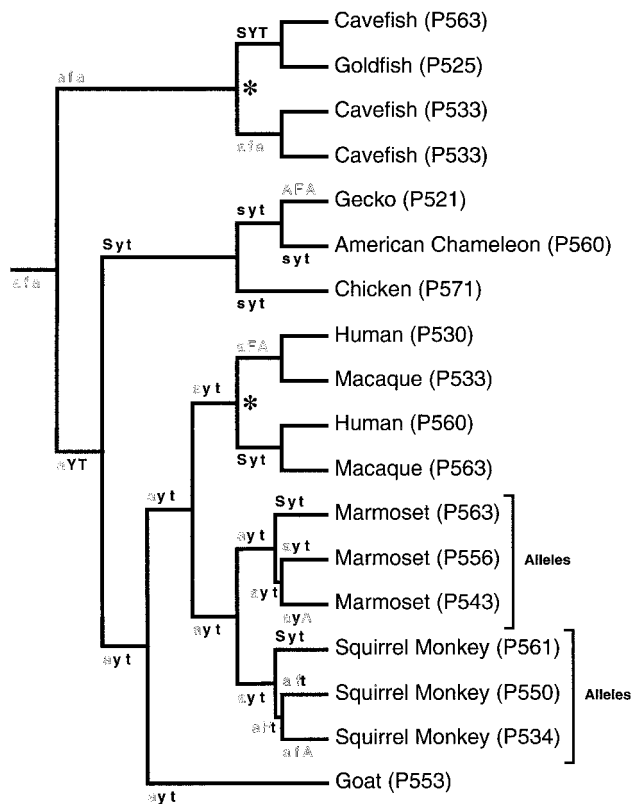


FIG. 6.—Nearest-neighbor-joining tree of the green-red opsin family.  $\lambda_{\max}$  values are given in parentheses. Amino acids conferring sensitivity to red (solid) and green (outlined) light are capitalized to denote new replacements. The asterisks mark gene duplications. After Yokoyama (1998) and Shyue et al. (1995).

most mammals, had dichromatic vision. However, Old World monkeys and primates evolved trichromatic color vision (blue, green, and red). Figure 6 reveals that this was achieved through duplication of the red-shifted opsin gene followed by reversion of one copy to the functionally ancestral state. The green-like opsin of the gecko is also functionally atavistic and represents an independent series of reversions.

That the same parallel amino acid replacements (Ala180→Ser, Phe227→Tyr, and Ala285→Thr) generating red-sensitive opsins have occurred independently in fish, reptiles, and mammals strongly implies an adaptive role in their evolution. One plausible scenario suggests that red-sensitive opsins arose as a response to the transition from blue water environments to the more reddish photic environments of shallow water (cavefish) (Yokoyama, Knox and Yokoyama 1995) and land (animals) (Yokoyama 1998). The trichromatic vision of Old World monkeys and primates may represent an adaptive response to facilitate the detection of red and yellow fruits against dappled foliage (Mollon 1991).

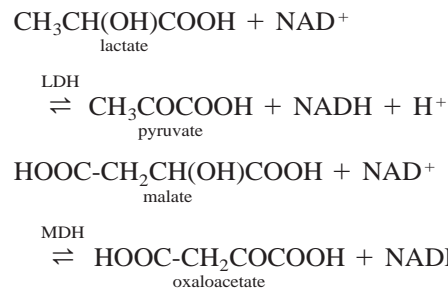
Primates and Old World monkeys have trichromatic vision because they possess one autosomal blue-sensitive opsin gene and two X-linked opsin genes, one red-sensitive and the other green-sensitive. Many New World monkeys possess one autosomal blue-sensitive opsin gene and only one X-linked opsin gene (however, see Jacobs et al. 1996). The latter is polymorphic in

squirrel monkeys (*Saimiri sciureus*), where the alleles have  $\lambda_{\max}$  values of 534, 550, and 561 nm, and in marmosets (*Callithrix jacchus jacchus*), where the alleles have  $\lambda_{\max}$  values of 543, 556, and 563 nm (Neitz, Neitz, and Jacobs 1991). Consequently, the males and homozygous females of some species of New World monkeys have dichromatic vision, whereas the heterozygous females have trichromatic vision. Evolutionary analysis (Shyue et al. 1995) reveals that the allelic lineages of squirrel monkeys and marmosets arose independently after the species diverged (fig. 6) some 16.4–19.0 MYA.

Dichromatic individuals (males and homozygous females) detect camouflaged objects more readily (Morgan, Adam, and Mollon 1992), while the differing  $\lambda_{\max}$  values of the three alleles may permit individuals to explore different photic environments (Mollon, Bowmaker, and Jacobs 1984). Whether or not such scenarios are responsible for maintaining trichromatic/dichromatic vision in New World monkeys, and whether these polymorphisms aid foraging by gregarious fruit-eating species remains debatable. One thing is certain, however: the allelic lineages of the New World monkeys have been retained for millions of years (5.1–5.9 Myr in squirrel monkeys and 9.8–11.4 Myr in marmosets), implicating strong balancing selection and suggesting that the trichromatic vision of Old World monkeys and primates is also adaptive (Shyue et al. 1995).

#### Lactate Dehydrogenase and Malate Dehydrogenase: Functional Lability Versus Functional Stability

Lactate dehydrogenase (LDH) and malate dehydrogenase (MDH) both utilize NAD as a coenzyme and catalyze, respectively, the interconversion of lactate to pyruvate and malate to oxaloacetate, viz.:



Both reactions are chemically similar, with malate merely having an additional carboxyl moiety attached to the  $\beta$ -methyl of lactate. Both activities are important (LDH in glycolysis and MDH in Krebs' cycle and photosynthesis), and most cells have distinct functional genes for each enzyme.

How many amino acid replacements would it take to completely convert an LDH into an MDH, given that these enzymes differ at roughly 230 out of 320 sites (excluding insertions and deletions)? A hundred, perhaps? Fifty, maybe? Amazingly, the answer is one! Replacing Gln102 by Arg in the LDH of *Bacillus stearothermophilus* converts the enzyme into an efficient, highly specific MDH (Wilks et al. 1988) (table 3).

The three-dimensional structure of these proteins hints that such an interchange might be possible. Even

**Table 3**  
**Kinetic Parameters of Wild-Type and Engineered LDHs and MDHs**

ENZYME	PERFORMANCE ( $k_{cat}/K_m$ ) ( $\mu\text{M}^{-1}\text{s}^{-1}$ )		PREFERENCE	
	Pyruvate	Oxaloacetate	Pyruvate/Oxaloacetate	Oxaloacetate/Pyruvate
<i>Bacillus stearothermophilus</i> LDH <sup>a</sup> . . . . .	4.2	0.004	1,050	0.00095
Engineered LDH (Gln102→Arg). . . . .	0.0005	4.2	0.00012	8,400
<i>Haloarcula marismortui</i> MDH <sup>b</sup> . . . . .	NM <sup>d</sup>	0.2	0	—
Engineered MDH (Arg102→Gln) . . . . .	0.0056	$6.8 \cdot 10^{-5}$	82.3	0.12
<i>Escherichia coli</i> MDH <sup>c</sup> . . . . .	NM	26	0	—
Engineered MDH (Arg102→Gln) . . . . .	$1.2 \cdot 10^{-4}$	NM	—	0

<sup>a</sup> Wilks et al. (1988).

<sup>b</sup> Cendrin et al. (1993).

<sup>c</sup> Boernke et al. (1995).

<sup>d</sup> No measurable activity.

with amino acid sequence identities as low as 25%, and the various insertions and deletions that have accrued over billions of years of evolution, these enzymes retain a common characteristic three-dimensional fold (fig. 7). Moreover, both enzymes share a common catalytic machinery of conserved residues preserved in three-dimensional space. In fact, their active sites only differ at one critical location: the uncharged Gln102 of LDH is replaced in MDH by a positively charged Arg that forms a double ionic H-bond to the additional  $\beta$ -carboxyl of malate (fig. 8). In contrast, attempts to convert MDHs into efficient LDHs have met with less success (Wilks et al. 1992; Nicholls et al. 1994; Cendrin et al. 1993; Boernke et al. 1995). Although Arg102→Gln replacements produce enzymes specific for pyruvate, they have lower catalytic efficiencies (table 3). Engineering MDH to LDH eliminates the double ionic H-bond to the substrate. With less energy available for substrate binding, less energy is available to stabilize the transition state. The result is a loss of catalytic power.

How does substrate specificity map onto phylogeny? Do MDHs occasionally arise in the LDH lineages and vice versa? After all, it takes only one replacement to convert an LDH into an MDH, and perhaps no more than a few to convert an MDH into an efficient LDH. A total of 124 sequences of lactate dehydrogenase and malate dehydrogenase genes were collected from public databases and aligned using the program CLUSTAL W (Thompson, Higgins, and Gibson 1994). The result was adjusted by hand to incorporate known three-dimension-

al structural information. Phylogenies were reconstructed using neighbor-joining (Saitou and Nei 1987) and maximum likelihood (Adachi and Hasegawa 1992).

The resulting phylogeny clearly separates the sequences into three distinct groups (fig. 9). The majority of LDH and MDH sequences separate into two large clusters, their dual presence in most organisms indicating that their genes duplicated and diverged in a single common ancestor. Between these groups lie a collection of intermediate forms (some of which may be MDHs, judging from the presence of Arg102), the existence of which is not surprising, given that a single replacement is sufficient to change the substrate specificity in these dehydrogenases. What is surprising, and in marked contrast to the opsins, is that the phylogeny so closely reflects functional differences. Even though a single replacement is sufficient to change substrate specificity, there is no evidence in figure 9 that any such switching occurred for billions of years. Rather, there must be enough of a selective advantage due to supplementary substitutions to prevent a duplicate copy of MDH from replacing a native LDH gene (or vice versa).

#### Hemoglobin: Different Species, Different Genes, Different Replacements—Same Mechanism, Same Effect

Hemoglobin delivers oxygen from lungs and gills to tissues and has long been a subject of intense study by structural biochemists, comparative physiologists, and evolutionary biologists. The role of the Glu6<sup>B</sup>→Val



FIG. 7.— $C\alpha$  traces of monomers of *Bacillus stearothermophilus* LDH (fat gray line) on *Escherichia coli* MDH (thin black line), with black dots marking the active sites. Both protein folds are remarkably similar despite minor variations in secondary structure (e.g., bottom right), yet their amino acid sequences are only 25% identical.

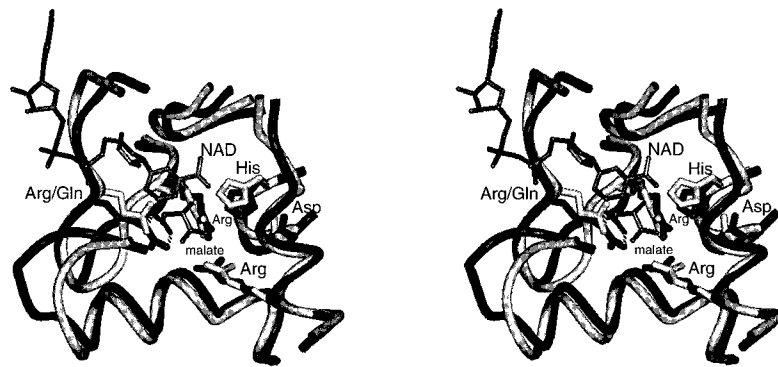


FIG. 8.—Detail of the active sites of LDH (gray) and MDH (black). The active sites differ only in one critical residue: at left the Gln in lactate dehydrogenase is replaced by the Arg in malate dehydrogenase that forms H-bonds (dashed lines) to the  $\beta$ -carboxylate of 2S-malate.

mutation in affording heterozygotes some protection against the ravages of malaria while causing homozygotes to suffer a debilitating anemia remains a classic tale of microadaptation. Yet there is far more to the evolution of hemoglobins than this rightfully celebrated example. Indeed, a vast literature exists, with many comparative studies (reviewed by Perutz 1983; Clementi et al. 1994) providing a wealth of hypotheses for experimental and evolutionary investigation. We shall describe just one such example.

Adult hemoglobins of higher vertebrates are tetramers formed from pairs of homologous subunits (two  $\alpha$  and two  $\beta$ ) that each bind  $O_2$  at their hemes. These

tetramers exist in an equilibrium between two states, a high-affinity R state and a low-affinity T state (fig. 10). Various effectors (e.g.,  $H^+$ ,  $Cl^-$ ,  $CO_2$ , diphosphoglycerate in man, inositol pentaphosphate in birds) exert important physiological effects by preferentially binding to the deoxygenated T state, thereby lowering affinity for  $O_2$ . For example, in respiring tissues, the buildup of lactic acid and bicarbonate reduces pH so that additional protons bind and stabilize the deoxygenated T state, thereby facilitating the release of  $O_2$  for aerobic metabolism.

The bar-headed goose (*Anser indicus*) migrates over Mount Everest at altitudes exceeding 9 km, where

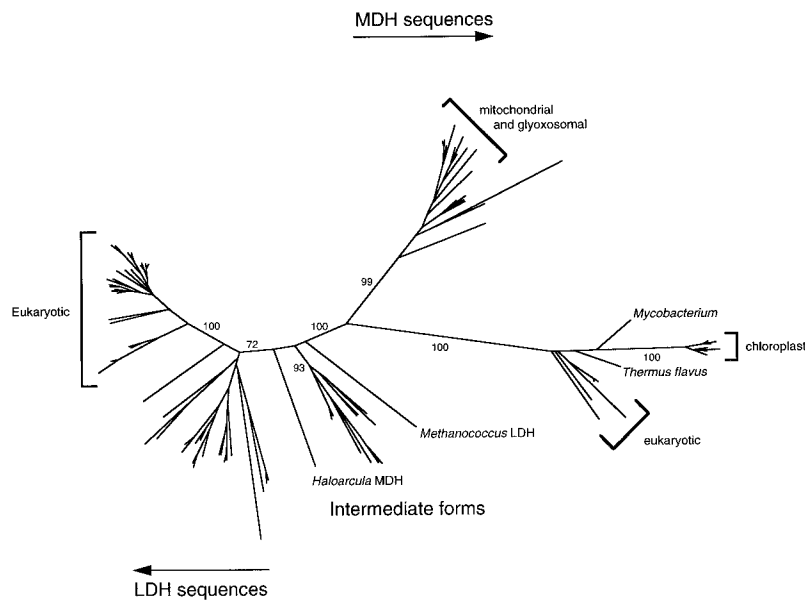


FIG. 9.—The neighbor-joining consensus phylogeny for LDH and MDH. Selected percentages are shown based on 200 bootstrapped sequences. Although the majority of LDH sequences cluster together (at left), only eukaryotic LDHs form a significant group (100% of 200 trees). The intermediate forms (middle) consist of 10 sequences (*Bacillus* MDHs [three species], *Chloroflexus* MDHs [three species], a *Synechocystis* sequence, *Toxoplasma* LDHs [two species], and a *Plasmodium* LDH) that branch together in 93% of 200 trees and cannot be split into separate clusters according to likelihood tests. In contrast, the two archaeobacterial sequences may be monophyletic or polyphyletic and can branch closer to LDHs or MDHs. The claim (Synstad, Emmerhoff, and Sirevag 1996) that the *Chloroflexus* MDHs are unusual because they are more similar to LDHs is not supported. The MDHs clearly fall into two groups. In the first (at top), consistent with the endosymbiotic origin of organelles, proteobacterial MDHs branch just proximal to the mitochondrial and glyoxosomal MDH sequences. The mitochondrial and cytosolic MDH genes of *Saccharomyces cerevisiae* branch at the base of this cluster, suggesting transfer or exchange of sequences in yeast. In the second (at right), chloroplast MDH sequences cluster together, branching near *Thermus flavus* and *Mycobacterium*. The remainder of the eukaryotic cytosolic MDH sequences (seven species) branch at the base of this group. The deep divergence between the proteobacteria and *Thermus/Mycobacterium* indicates an ancient duplication (McAlister-Henn 1988).

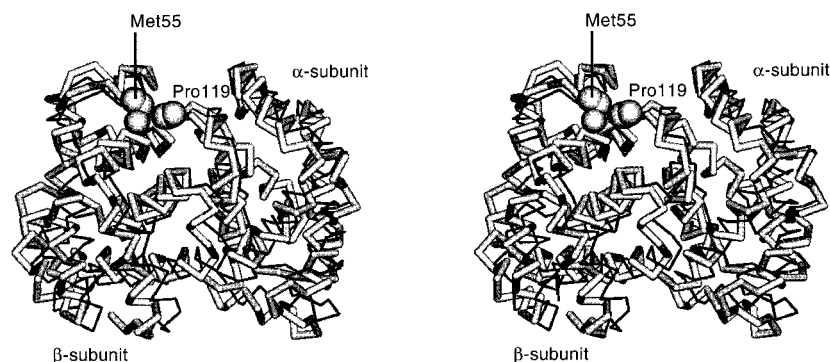


FIG. 10.—The backbones ( $C\alpha$  traces) of oxygenated (R state: thin black lines) and deoxygenated (T state: fat gray lines) human hemoglobin. The evident shift in position is caused by one  $\alpha 1\beta 1$  dimer turning as a rigid body with respect to the second  $\alpha 1\beta 1$  dimer (not shown) of the tetramer. Met55 $\beta$  and Pro119 $\alpha$  (van der Waals surfaces) are shown to be in contact at the subunit interface of the  $\alpha 1\beta 1$  dimer.

the partial pressure of  $O_2$  is only 30% of that at sea level. The high affinity of its hemoglobin for  $O_2$  in the presence of  $Cl^-$  and inositol pentaphosphate is undoubtedly one among many adaptations to vigorous exercise in such a rarefied atmosphere (table 4). The related low-flying greylag goose (*Anser anser*) has a hemoglobin with a normal  $O_2$  affinity. The  $\alpha$  chains differ by only three amino acid replacements and the  $\beta$  chains by just one.

On examining the X-ray structure of human hemoglobin, Perutz (1983) suggested that the Pro119 $\alpha$ →Ala replacement, which is unique among bird sequences, might be responsible for the high  $O_2$  affinity of the bar-headed goose hemoglobin. The Ala replacement removes an important van der Waals (fig. 11) contact between the  $\alpha 1$  and  $\beta 1$  subunits, and should shift the equilibrium from the low- $O_2$ -affinity T state toward the high- $O_2$ -affinity R state (several studies reveal that weakening contacts across this interface shifts hemoglobins toward the high- $O_2$ -affinity R state: Asakura et al. 1976; Amiconi et al. 1989). A recent X-ray analysis confirms that the Pro119 $\alpha$ →Ala replacement in bar-headed goose hemoglobin eliminates this critical intersubunit contact (Zhang et al. 1996). The Andean goose (*Chloephaga melanoptera*), which lives 6 km high in the Andes of South America and isn't a goose at all (it's a duck: fig. 12), also has a high- $O_2$ -affinity hemoglobin. Comparisons of Andean goose sequences with other avian sequences led Heibl, Braunitzer, and Schneegans (1987) to suggest that its high  $O_2$  affinity arises as a consequence of the Leu55 $\beta$ →Ser replacement. This removes

the very same intersubunit contact as Pro119 $\alpha$ →Ala in the bar-headed goose, but this time from the opposite subunit (fig. 11).

Jessen et al. (1991) tested the hypothesis that replacements at the Pro119 $\alpha$ -Leu55 $\beta$  contact alone can shift the equilibrium from the low- $O_2$ -affinity T state toward the high- $O_2$ -affinity R state. Following-site directed mutagenesis to introduce the Pro119 $\alpha$ →Ala replacement into human globin (70% identical to goose globins), reconstituted tetramers were found to have oxygen affinities that, in the presence of  $Cl^-$  and diphosphoglycerate (the human equivalent of inositol pentaphosphate), exceeded normal human hemoglobin by a factor greater than that observed between bar-headed and greylag geese (table 4). Engineering the Met55 $\beta$ →Ser into human globin also resulted in a reconstituted hemoglobin with higher affinity for  $O_2$ . Importantly, these replacements had no detectable effect on other properties of human hemoglobin, such as the Bohr effect (Weber et al. 1993). X-ray crystallography was used to show that the Met55 $\beta$ →Ser replacement had no effect on hemoglobin structure, save the gap introduced by replacing a larger amino acid by a smaller one.

#### Isocitrate Dehydrogenase: From Catabolism to Anabolism, 3.5 Billion Years Ago

Isocitrate dehydrogenases (IDHs) catalyze the oxidation of isocitrate to  $\alpha$ -ketoglutarate, an important intermediate in the energy-generating Krebs' cycle and a precursor for ammonia fixation and glutamate biosynthesis. Together with  $\beta$ -isopropylmalate dehydrogenase (IMDH), which catalyses a chemically similar reaction in leucine biosynthesis, IDHs form an ancient and highly divergent family whose sequences and structures are wholly unrelated to those of other enzymes (fig. 13).

IDHs utilize NADP or NAD as coenzymes (cosubstrates). Although NADP and NAD are chemically equivalent, they play very different metabolic roles: NADPH provides the reducing power for biosynthesis, while NADH provides the electrons for energy production in the form of ATP. A switch from utilizing NAD to utilizing NADP represents a major shift in metabolic role, from energy production to biosynthesis.

Such a switch evolved in eubacteria. Phylogenetic analysis (fig. 14) indicates that the NADP-dependence

**Table 4**  
 **$O_2$  Affinity of Natural and Engineered Hemoglobins (after Jessen et al. 1991)**

HEMOGLOBIN	RESIDUE		AFFINITY <sup>a</sup> $P_{50}$
	119 $\alpha$	55 $\beta$	
Bar-headed goose . . . . .	Ala	Leu	2.0
Greylag goose . . . . .	Pro	Leu	2.8
Human . . . . .	Pro	Met	5.8
Human (mutant 1) . . . . .	Ala	Met	3.3
Human (mutant 2) . . . . .	Pro	Ser	3.4

<sup>a</sup> Affinity is defined as the partial pressure of  $O_2$  (in mm Hg) necessary for half-saturation.



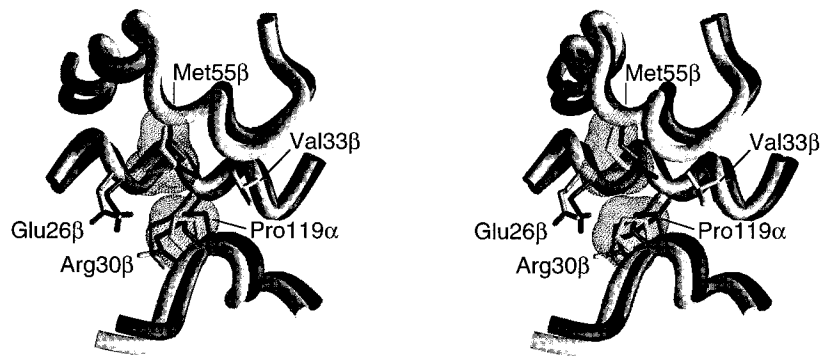


FIG. 11.—A close-up of Met55 $\beta$  contacting Pro119 $\alpha$  (dotted van der Waals surfaces) in the deoxygenated T state (gray C $\alpha$  worm and side chains). The contact is maintained in the oxygenated R state (black C $\alpha$  worm and side chains), although the tip of the Met55 $\beta$  side chain has flipped 180°. The side chains of other amino acids in the vicinity are shown, including Arg30 $\beta$ , which also forms an intersubunit contact through an H-bond.

of certain eubacterial IDHs is a shared derived character that evolved on or around the time eukaryotes first appeared (Dean and Golding 1997). The pentose phosphate shunt, the usual source of NADPH, is inoperative during growth on acetate, all of which enters Krebs' cycle. Here, IDH provides 90% of the NADPH for biosynthesis (Walsh and Koshland 1985). Indeed, bacteria with NAD-dependent IDHs lack either a respiratory chain or a complete Krebs' cycle and so are incapable of growth on acetate. Evidently, 3.5 billion years ago an ancestral eubacterium evolved an NADP-dependent IDH in response to expanding its niche to growth on acetate.

Based on a knowledge of high-resolution X-ray structures (fig. 15) of the binary complexes of *E. coli* IDH with NADP (Hurley et al. 1991) and *Thermus thermophilus* IMDH with NAD (Hurley and Dean 1994), Chen, Greer and Dean (1995) replaced six amino acids (Lys344 $\rightarrow$ Asp, Tyr345 $\rightarrow$ Ile, Val351 $\rightarrow$ Ala, Tyr391 $\rightarrow$ Lys, Arg395 $\rightarrow$ Ser, Arg292' $\rightarrow$ Asp) in the coenzyme-binding pocket of wild-type *E. coli* IDH to

cause a shift in preference from NADP to NAD by a factor exceeding five million. However, the overall activity of the engineered IDH toward NAD was poor. By retaining Arg292' (Arg292' is present in several NAD-dependent enzymes) and by introducing two additional "haphazard" substitutions at sites remote to the nucleotide-binding pocket (Cys332 $\rightarrow$ Tyr and Cys201 $\rightarrow$ Met), the overall performance with NAD was improved to a level comparable to the eukaryotic NAD-dependent mitochondrial IDH from yeast, while a comparable preference for NAD was maintained (table 5). The X-ray structure of the engineered IDH was determined (Hurley, Chen and Dean 1996): NAD occupies precisely the same position as seen in IMDH.

The coenzyme specificity of *T. thermophilus* IMDH has also been inverted (Chen, Greer and Dean 1996). This feat of engineering required more than merely replacing amino acids in the nucleotide-binding pocket with those of IDH. In IMDH a  $\beta$ -turn replaces the  $\alpha$ -helix and loop of IDH, thereby eliminating a key residue, Arg395, that H-bonds to the 2'-phosphate of NADP (fig. 15). The seven residues of the  $\beta$ -turn in IMDH were replaced by a 13-residue sequence modeled on the  $\alpha$ -helix and loop of *E. coli* IDH, but containing additional substitutions to ensure correct packing against the remaining hydrophobic core. Together with four direct replacements (Ser292' $\rightarrow$ Arg, Asp344 $\rightarrow$ Lys, Ile345 $\rightarrow$ Tyr, and Ala351 $\rightarrow$ Val), a shift in preference from NAD to NADP by a factor 100,000 was generated. The resulting mutant has a 1,000-fold preference for NADP and is twice as active as the wild-type enzyme (table 5).

These results demonstrate that the coenzyme specificities of the decarboxylating dehydrogenases are determined by residues lining the nucleotide-binding pocket and that the many differences outside the nucleotide-binding pockets contribute relatively little to discrimination between the coenzymes.

The availability of high-resolution X-ray structures of the binary complexes of wild-type IDH and IMDH proved critical to identifying the determinants of specificity. Sequence alignments alone cannot identify changes in local secondary structures, and the critical

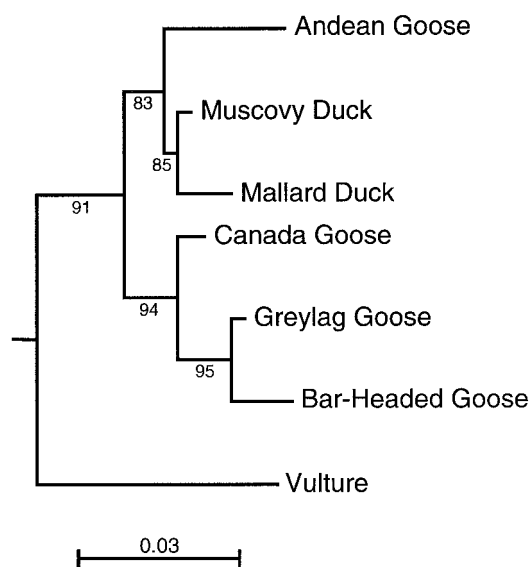


FIG. 12.—A neighbor-joining tree based on concatenated  $\alpha$  and  $\beta$  bird hemoglobin amino acid sequences. Bootstrap values are percentages from 1,000 trees.



FIG. 13.—Superposition of *E. coli* IDH (gray worm) on *T. thermophilus* IMDH (black worm). Superimposing X-ray structures greatly facilitates alignment of amino acid sequences when identities are low (~20% in this case, and only 16 residues are conserved among >600 sites in the family as a whole) and identifies gaps unambiguously (e.g., the  $\alpha$ -helix and loop which form the hook at the bottom right of the IDH structure are missing in IMDH). The gray dot denotes the position of the coenzyme-binding site.

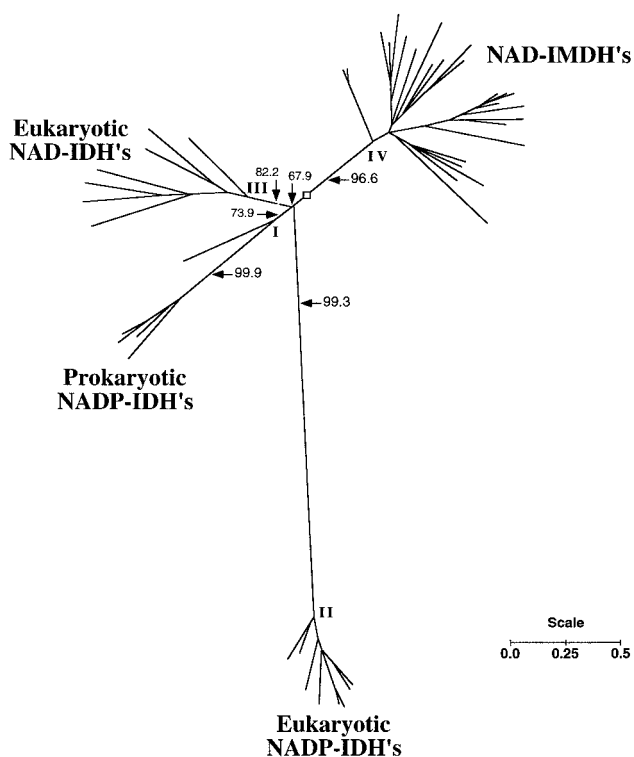


FIG. 14.—A maximum-likelihood phylogeny of the 2-hydroxy acid  $\beta$ -decarboxylating dehydrogenases based on amino acid sequences. Bootstrap values of 1,000 maximum-parsimony and 1,000 nearest-neighbor-joining trees support monophyly for the four major groups (eubacterial IDHs, eukaryotic NAD-IDHs, eukaryotic NADP-IDHs, and NAD-IMDHs). The phylogeny is rooted along the branch linking the IDHs to the IMDHs, because the common ancestor was undoubtedly prototrophic, capable of synthesizing glutamate and leucine, and hence must have had both enzymatic functions. This position implies that an ancestral gene duplication was followed by functional specialization prior to the divergence of eukaryotes and eubacteria. Furthermore, eubacterial NADP-IDHs include both Gram +ve and Gram -ve species, which supports very ancient divergence, on or about the time that eukaryotes first appeared, some 3.5 billion years ago.

Val351 $\rightarrow$ Ala substitution would have remained unnoticed because eukaryotic IMDHs retain the Val. Even the experimental demonstration that changes at the sites and secondary structures identified by X-ray crystallography are sufficient to generate changes in specificity is crucial. Parallel work on the substrate specificities of these enzymes has yet to yield significant changes while retaining catalytic efficiency.

## Discussion

Determining which characters of an organism are selected, which result from correlated responses, and which are selectively neutral presents an extraordinary problem for evolutionists. Even the adaptive nature of such celebrated examples as the giraffe's neck and the peacock's tail remain matters of dispute. This difficulty is brought into sharp relief at the biochemical level where the neutral theory, without ever invoking positive selection, has proven extraordinarily robust in explaining the broad patterns of molecular evolution.

Most often, we examine sequences that diverged thousands or millions of years ago, in which a substantial number of substitutions have accumulated. Many may be neutral. Others, although selected, could simply be treadmill adaptations (Dean and Golding 1997) accrued as populations track ever-changing environments. Although important, they do not alter protein function in a major way. A few represent major adaptations of large effect. Identifying these among so many others is exceedingly difficult.

## Approaches

Three of the studies described began with some obvious ecological, physiological, or metabolic clue—differences in ambient light (opsins), flight at extreme altitudes (hemoglobins), growth on acetate (IDHs). In each of these cases, a substantial amount of information was available indicating that selection had been at work. The LDH/MDH study started as an exercise in protein engineering, and the chymase and RNase studies sought to understand how function evolved, rather than why any changes might be adaptive. Whether any of the

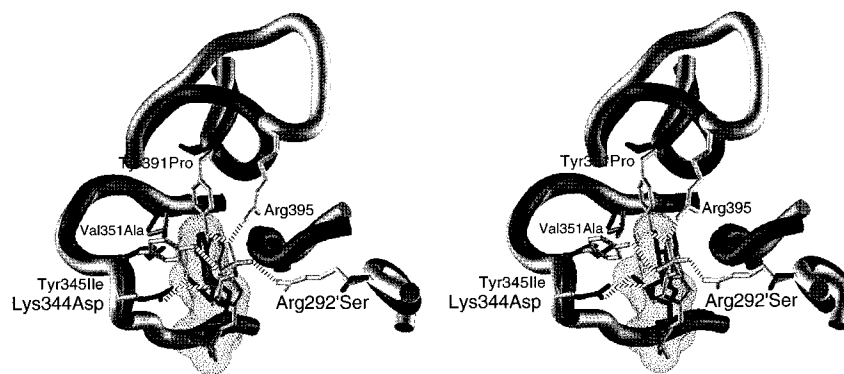


FIG. 15.—Superposition of the coenzyme binding sites of *E. coli* IDH with bound NADP (gray) and *T. thermophilus* IMDH with bound NAD (black, with NAD surfaced), showing side chains (IDH numbering) and H-bonds (dashed lines) critical to specificity. In IDH, H-bonds form between Tyr345, Tyr391, Arg395, Arg292', and the 2'-phosphate of NADP (Lys344 is disordered in this structure). In IMDH, amino acid replacements remove all H-bonds to the 2'-phosphate. Tyr345 is replaced by Ile, and Val351 is replaced by Ala. These smaller amino acids allow the coenzyme to tilt to the left, enabling a double H-bond to form between an Asp (that replaces Lys344) and the ribose hydroxyls of NAD. The introduction of the negatively charged Asp side chain also disrupts NADP binding through electrostatic repulsion of the 2'-phosphate. The nicotinamide mononucleotide moieties of both coenzymes are not shown.

changes in chymase and RNase are adaptive remains unknown.

Phylogenies provide historical context, allow function to be mapped onto genealogy, and help identify likely replacements of adaptive significance. Studies of avian hemoglobins compared sequences from closely related species to focus attention on just a few replacements, one or more of which were responsible for functional differences. The opsin sequences are far more divergent than the avian hemoglobins, and pairs of sequences differ at many more sites. Yet, in this rather “bushy” phylogeny, several functional parallelisms and reversals greatly aided in identifying key amino acids. The LDH/MDH and IDH/IMDH phylogenies are so divergent that X-ray structures provide the only reliable means to align sequences and to identify candidate residues. Yet, even here, phylogenies provided the only means to determine when adaptive events occurred.

Sequence comparisons alone are insufficient to identify replacements of functional consequence. Site-directed mutagenesis, particularly when guided by phylogeny, can be used to search among a limited number of replacements. This approach successfully identified replacements conferring functional differences in RNases and opsins. However, the best way to identify likely replacements is by comparing X-ray structures. Protein structures enabled Perutz (1983) to predict which one of four replacements was responsible for increasing the ox-

xygen affinity in bar-headed goose hemoglobin. Protein structures led Chen, Greer, and Dean (1995, 1996) to correctly identify a handful of residues critical to coenzyme specificity in IDH and IMDH and Wilks et al. (1988) to correctly identify the active-site residue responsible for substrate specificity in LDH.

Site-directed mutagenesis experiments provide rigorous tests of structural, functional, and, by implication, adaptive hypotheses. These studies also demonstrate, as does a substantial body of the biochemistry literature, that replacements generally act independently. Protein evolution is not the horrendously nonlinear problem that many have imagined, and although nonlinearities can and do occur, such complications are rare. Investigating the structural basis of molecular adaptation is a tractable proposition.

#### Major Adaptive Shifts Usually Require Just a Few Replacements

Examples of major adaptive shifts requiring just a few replacements include the conversion of LDH into MDH (Wilks et al. 1988), the evolution of an organophosphate hydrolase from a carboxylesterase to confer insecticide resistance on blowflies (Newcomb et al. 1997), and the acquisition of lactase activity by *E. coli* evolved  $\beta$ -galactosidase (Hall 1984). Even the dramatic changes in coenzyme specificity engineered into IDH

**Table 5**  
**Kinetic Parameters of Wild-Type and Engineered IDH and IMDH**

ENZYME	PERFORMANCE ( $k_{cat}/K_m$ ) ( $\mu\text{M}^{-1}\text{s}^{-1}$ )		PREFERENCE	
	NADP	NAD	NADP/NAD	NAD/NADP
<i>Escherichia coli</i> NADP-IDH . . . . .	4.7	0.00069	6,900	0.00015
<i>Saccharomyces cerevisiae</i> NAD-IDH . . . . .		0.19		
Engineered NAD-IDH . . . . .	0.00081	0.164	0.005	200
<i>Thermus thermophilus</i> NAD-IMDH . . . . .	0.00015	0.0125	0.012	80
Engineered NADP-IMDH . . . . .	0.02	0.00002	1,000	0.001

and IMDH require no more than a handful of replacements (Chen, Greer, and Dean 1995, 1996).

#### Adaptive Replacements Are Not Solely Confined to Active Sites

The assumption that amino acid replacements far from active sites must be selectively neutral because they inevitably lack functional consequences is wrong. The Gly38→Asp replacement in RNase lies 5 Å from the active site yet produces a fourfold change in specificity toward double-stranded RNA (Jermann et al. 1995). Similarly, two replacements outside the active site of NAD-IDH produce a 16-fold improvement in activity (Chen, Greer, and Dean 1995). In neither case are we certain how these replacements affect function. The side chain introduced into RNase should force the main chain to adopt another conformation, the effects of which might be transmitted into the active site (S. Benner, personal communication). The two replacements engineered into NAD-IDH produce subtle conformational changes that affect the positioning of several catalytic residues (Hurley, Chen, and Dean 1996). While such replacements do not produce dramatic functional changes, they can be adaptive in fine-tuning function. The Gly38→Asp replacement in RNase may be one such example. Replacements away from the heme-binding site in bar-headed and Andean geese hemoglobins produce the modest twofold increases in oxygen affinity so essential for flight at high altitude (Jessen et al. 1991).

#### Ecological Consequences

The changes in hemoglobin oxygen affinity and the shifts in the  $\lambda_{\max}$  values of opsins may be modest, but their ecological consequences are marked. Few birds could possibly migrate over the Himalayas. A modest molecular change results in a dramatic expansion of an ecological niche. The shift of  $\lambda_{\max}$  of a cavefish visual pigment toward the red end of the spectrum is undoubtedly a response to the shift in niche from deep to shallow water. Equally incorrect, however, is the assumption that every subtle change in function is of adaptive importance. In the intense competition imposed by the chemostat, a 5% increase in the activity of *E. coli*  $\beta$ -galactosidase would produce a selection coefficient of only 0.02% (Dean 1995). This would be selectively neutral in any population with an effective size smaller than 2,500.

#### Different Solutions to the Same Evolutionary Challenge

Independent attempts to solve an identical evolutionary challenge frequently produce different solutions. Both adaptive replacements to flight at high altitude eliminate the same van der Waals contact between the  $\alpha$  and  $\beta$  subunits of hemoglobin. Yet, two different species solve this problem by different mutations in different genes. The bar-headed goose eliminates the contact from the  $\alpha$  side, while the Andean goose eliminates the same contact from the  $\beta$  side. In vultures that fly at high altitudes, a different suite of replacements at the  $\alpha 1\beta 1$  and  $\alpha 1\beta 2$  hemoglobin interfaces confer high oxygen af-

finity (Hiebl et al. 1987, 1988, 1989). Old World monkeys and New World monkeys each face the problem of visually detecting fruits in a forest canopy. Old World monkeys and some New World monkeys evolved trichromatic vision through gene duplication and functional divergence: other New World monkeys evolved a balanced polymorphism, maintaining both dichromatic and trichromatic vision.

#### Evolutionary Reversals and Identical Solutions

The LDH/MDH phylogeny could hardly provide a more conventional view of molecular evolution—gene duplication followed by functional specialization is a common enough theme. What is stunning is the fact that, even after billions of years accumulating hundreds of replacements, a single amino acid substitution is sufficient to interchange function.

The stability of functional specialization displayed by LDH and MDH contrasts with chymase. Here, the ancestral molecular phenotype is far more specific than its evolutionary descendant. Evidently, an ancient serine protease evolved into a chymase with marked substrate specificity, only to change course and evolve into an enzyme with broad specificity again. Similarly, evolution in opsins is characterized by several reversals.

Evolution in the opsins also contrasts with that in hemoglobins. Following gene duplication ( $\lambda_{\max} \sim 550$  nm) in Old World primates, one opsin evolved sensitivity to longer wavelengths ( $\lambda_{\max} \sim 560$  nm), while the other reverted to its ancestral phenotype ( $\lambda_{\max} \sim 530$  nm) (Nei, Zhang, and Yokoyama 1997). A similar reversal at a single locus in New World monkeys produced an allele with the ancestral phenotype. Another reversal evolved independently in the geckos. And, amazingly, each of these reversals is associated with precisely the same suite of replacements (fig. 6). Furthermore, the evolution of sensitivity to longer wavelengths ( $\lambda_{\max} \sim 560$  nm) in fish has produced precisely the same replacements as in land animals. Evidently, adaptive evolution is more constrained in opsins than in the high-altitude avian globins.

#### Radical and Conservative Replacements

Adaptive changes need not be the “radical” substitutions emphasized in amino acid substitution tables such as PAM matrices. Gln↔Arg replacements occur relatively frequently (PAM250  $\log_e$  odds of 0 for Gln↔Arg vs.  $\log_e$  odds of 2 for Gln↔Gln) and might be considered “conservative.” Yet by any stretch of the imagination the Gln102→Arg replacement in *Bacillus stearothermophilus* LDH is radical—it causes a  $10^7$ -fold change in specificity. Comparing *Bacillus stearothermophilus* LDH with any MDH sequence reveals a substantial number of “radical” replacements that, collectively at least, exert minimal effect. The practice of using the terms “radical” and “conservative” with regard to amino acid replacements based on measures of how frequently they are interchanged during evolution should be abandoned. The connection between frequency and function is tenuous at best.

## Conclusions

There is something unique about molecular structure. Contained within the linear array of amino acids of a peptide is an element of genetics that, upon comparison with related sequences, provides a record of evolutionary history. The three-dimensional structure is a morphology that directly relates to function, phenotypes upon which selection can act. By bringing together phylogeny, form, and function, protein structures have much to offer the field of molecular evolution in general and the study of molecular adaptation in particular. These structures, together with protein engineering, allow evolutionary hypotheses to be tested far more rigorously than previously imagined.

While structural biology contributes to evolutionary biology, evolutionary biology contributes to structural biology. An evolutionary approach identified the key amino acid replacements responsible for spectral tuning in opsins (Yokoyama 1998), amino acid replacements that were overlooked with other approaches. An evolutionary approach identified the amino acid replacement responsible for the change in specificity in RNase (Jermann et al. 1995), an amino acid replacement that might be judged wholly innocuous from an inspection of the structure alone.

To really understand past adaptations, one ideally needs to study ancient organisms in their ancient habitats. In lieu of this, a great deal of progress can still be made. The examples discussed here demonstrate that no single method alone is sufficient. A reconstructed phylogeny is necessary to build an ancient chymase. Protein engineering is necessary to test the functional effects of replacements in RNase and opsins. A knowledge of globin structure reveals that two different replacements, at two different positions, in two different genes, from two different species, remove the same intersubunit contact to produce the same functional consequence, via the same functional mechanism, in response to the same selective pressure. High-resolution X-ray structures provide a ready means to align the highly divergent sequences of LDH and MDH, and those of IDH and IMDH, while the structures of their binary complexes provide the only means to identify functionally important replacements. It is the mixture of evolutionary theory, phylogenetic reconstruction, structural information, and protein engineering, along with contributions from metabolism, physiology, and ecology, that is so critical to understanding adaptation at the molecular level. Darwin needed to be broadly knowledgeable. So, too, do molecular evolutionists.

## Acknowledgments

We humbly apologize to those whose excellent work we omitted. We gratefully thank Dan Dykhuizen, Ward Watt, and Shozo Yokoyama for critically reviewing earlier drafts (especially Boddington's Ales) and Barry Hall for his encouragement and support. This work was supported by an NSERC grant to G.B.G. and NIH and NSF grants to A.M.D.

## LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1992. Molphy: programs for molecular phylogenetics, I. Protml: maximum likelihood inference of protein phylogeny. *Computer Science Monograph* 27, Japanese Institute of Statistical Mathematics, Tokyo.
- AMICONI, G., F. ASCOLI, D. BARRA, A. BERTOLLINI, R. M. MATARESE, D. VERZILI, and M. BRUNORI. 1989. Selective oxidation of methionine  $\beta$ 55D6 at the  $\alpha$ 1 $\beta$ 1 interface in hemoglobin completely destabilizes the T state. *J. Mol. Biol.* **264**:17745–17749.
- ASAKURA, T., K. ADACHI, J. S. WILEY, L. FUNG, C. HO, J. V. KILMARTIN, and M. F. PERUTZ. 1976. Structure and function of haemoglobin Philly (Tyr C1 (35) $\beta$ →Phe). *J. Mol. Biol.* **104**:185–195.
- ASENJO, A. B., J. RIM, and D. D. OPRIAN. 1994. Molecular determination of human red/green color discrimination. *Neuron* **12**:1131–1138.
- BEINTEMA, J. J., W. M. FITCH, and A. CARSANA. 1986. Molecular evolution of pancreatic-type ribonucleases. *Mol. Biol. Evol.* **3**:262–275.
- BENNER, S. A., T. M. JERMANN, J. G. OPITZ et al. (11 co-authors). 1996. Developing new synthetic catalysts. How nature does it. *Acta Chem. Scand.* **50**:243–248.
- BOERNKE, W. E., C. S. MILLARD, P. W. STEVENS, S. N. KAKAR, F. J. STEVENS, and M. I. DONNELLY. 1995. Stringency of substrate specificity of *Escherichia coli* malate dehydrogenase. *Arch. Biochem. Biophys.* **322**:43–52.
- CENDRIN, F., J. CHROBOCZEK, G. ZACCAI, H. EISENBERG, and M. MEVARECH. 1993. Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaeobacterium *Halococcus marismortui*. *Biochemistry* **32**:4308–4313.
- CHAN, T., M. LEE, and T. P. SAKMAR. 1992. Introduction of hydroxyl-bearing amino acids causes bathochromic spectral shifts in rhodopsin: amino acid substitutions responsible for red-green pigment spectral tuning. *J. Biol. Chem.* **267**:9478–9480.
- CHANDRASEKHARAN, U. M., S. SANKER, M. J. GLYNIAS, S. S. KARNIK, and A. HUSAIN. 1996. Angiotensin II-forming activity in a reconstructed ancestral chymase. *Science* **271**:502–505.
- CHEN, R., A. GREER, and A. M. DEAN. 1995. A highly active decarboxylating dehydrogenase with rationally inverted coenzyme specificity. *Proc. Natl. Acad. Sci. USA* **92**:11666–11670.
- . 1996. Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc. Natl. Acad. Sci. USA* **93**:12171–12176.
- CLEMENTI, M. E., S. G. CONDO, M. CASTAGNOLA, and B. GIARDINA. 1994. Hemoglobin function under extreme life conditions. *Eur. J. Biochem.* **233**:309–317.
- DEAN, A. M. 1995. A molecular investigation of genotype by environment interactions. *Genetics* **139**:19–33.
- DEAN, A. M., and G. B. GOLDING. 1997. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl. Acad. Sci. USA* **94**:3104–3109.
- DYKHUIZEN, D. E., and A. M. DEAN. 1994. Predicted fitness changes along an environmental gradient. *Evol. Ecol.* **8**:524–541.
- GOLDING, G. B. 1994. Non-neutral evolution: theories and molecular data. Chapman and Hall, New York.
- HALL, B. G. 1984. The evolved  $\beta$ -galactosidase system of *Escherichia coli*. Pp. 165–185 in R. P. MORTLOCK, ed. *Microorganisms as model systems for studying evolution*. Plenum Press, New York.

- HEIBL, I., G. BRAUNITZER, and D. SCHNEEGANSS. 1987. The primary structures of the major and minor hemoglobin-components of adult Andean goose (*Cloephaga melanoptera*, Anatidae): the mutation Leu→Ser in position 55 of the  $\beta$ -chains. *Biol. Chem. Hoppe Seyler* **368**:1559–1569.
- HEIBL, I., D. SCHNEEGANSS, F. GRIMM, J. K. ÖSTERS, and G. BRAUNITZER. 1987. High altitude respiration of birds. The primary structures of the major and minor hemoglobin component of adult European black vulture (*Aegypius monachus*, *Aegypiinae*). *Biol. Chem. Hoppe Seyler* **368**:11–18.
- HEIBL, I., R. WEBER, D. SCHNEEGANSS, and G. BRAUNITZER. 1989. The primary structure and functional properties of the major and minor hemoglobin component of adult white-headed vulture (*Trigonoceps occipitalis*, *Aegypiinae*). *Biol. Chem. Hoppe Seyler* **370**:699–706.
- HEIBL, I., R. WEBER, D. SCHNEEGANSS, J. K. ÖSTERS, and G. BRAUNITZER. 1988. Structural adaptations in the major and minor hemoglobin components of adult Rüppell's Griffon (*Gyps rueppelli*, *Aegypiinae*): a new molecular pattern for hypoxic tolerance. *Biol. Chem. Hoppe Seyler* **369**:217–232.
- HURLEY, J. H., R. CHEN, and A. M. DEAN. 1996. Determinants of cofactor specificity in isocitrate dehydrogenase: structure of an engineered NADP<sup>+</sup>→NAD<sup>+</sup> specificity-reversal mutant. *Biochemistry* **35**:5670–5678.
- HURLEY, J. H., and A. M. DEAN. 1994. Structure of 3-isopropylmalate dehydrogenase in complex with NAD<sup>+</sup>: ligand-induced loop closing and mechanism for cofactor specificity. *Structure* **2**:1007–1016.
- HURLEY, J. H., A. M. DEAN, D. E. KOSHLAND JR., and R. M. STROUD. 1991. Catalytic mechanism of NADP<sup>+</sup>-dependent isocitrate dehydrogenase: implications from the structures of magnesium-isocitrate and NADP<sup>+</sup> complexes. *Biochemistry* **30**:8671–8678.
- JACOBS, G. H., M. NEITZ, J. F. DEEGAN, and J. NEITZ. 1996. Trichromatic color vision in New World monkeys. *Nature* **382**:156–158.
- JERMANN, T. M., J. G. OPITZ, J. STACKHOUSE, and S. A. BENNER. 1995. Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* **374**:57–59.
- JESSEN, T. H., R. E. WEBER, G. FERMI, J. TAME, and G. BRAUNITZER. 1991. Adaptation of bird hemoglobins to high altitudes: demonstration of molecular mechanism by protein engineering. *Proc. Natl. Acad. Sci. USA* **88**:6519–6522.
- KIMURA, M. 1968a. Evolutionary rate at the molecular level. *Nature* **217**:624–626.
- . 1968b. Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res. Camb.* **11**:247–269.
- . 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge, England.
- KOEHN, R. K., and T. J. HILBISH. 1987. The adaptive importance of genetic variation. *Am. Sci.* **75**:134–141.
- KRISHNAN, S., B. G. HALL, and M. L. SINNOTT. 1995. Catalytic consequences of experimental evolution: catalysis by a 'third-generation' evolvant of the second  $\beta$ -galactosidase of *Escherichia coli*, *ebg<sup>abcde</sup>*, and by *ebg<sup>abcd</sup>*, a 'second-generation' evolvant containing two supposedly 'kinetically silent' mutations. *Biochem. J.* **312**:971–977.
- LEWONTIN, R. C. 1979. Adaptation. *Sci. Am.* **239**:156–169.
- MCALISTER-HENN, L. 1988. Evolutionary relationships among the malate dehydrogenases. *Trends Biochem. Sci.* **13**:178–181.
- MOLLON, J. D. 1991. The uses and evolutionary origins of primate color vision. Pp. 306–319 in J. R. CRONLY-DILLON and R. L. GREGORY, eds. *Evolution of the eye and visual pigments*. CRC Press, Boca Raton, Fla.
- MOLLON, J. D., J. K. BOWMAKER, and G. H. JACOBS. 1984. Variations of colour vision in a New World primate can be explained by a polymorphism of retinal photoreceptors. *Proc. R. Soc. Lond. B* **222**:373–399.
- MORGAN, M. J., A. ADAM, and J. D. MOLLON. 1992. Dichromats detect colour-camouflaged objects that are not detected by trichromats. *Proc. R. Soc. Lond. B* **248**:291–295.
- NATHANS, J. 1987. Molecular biology of visual pigments. *Annu. Rev. Neurosci.* **10**:163–194.
- NEI, M., J. ZHANG, and S. YOKOYAMA. 1997. Color vision of ancestral organisms of higher primates. *Mol. Biol. Evol.* **14**:611–618.
- NEITZ, M., J. NEITZ, and G. H. JACOBS. 1991. Spectral tuning of pigments underlying red-green color vision. *Science* **252**:971–974.
- NEWCOMB, R. D., P. M. CAMPBELL, D. L. OLLIS, E. CHEAH, R. J. RUSSEL, and J. G. OAKSHOTT. 1997. A single amino acid substitution converts a carboxylesterase to an organophosphorous hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl. Acad. Sci. USA* **94**:7464–7468.
- NICHOLLS, D. J., M. DAVEY, S. E. JONES, J. MILLER, J. J. HOLBROOK, A. R. CLARKE, M. D. SCAWEN, T. ATKINSON, and C. R. GOWARD. 1994. Substitution of the amino acid at position 102 with polar and aromatic residues influences substrate specificity of lactate dehydrogenase. *J. Protein Chem.* **13**:129–133.
- PERUTZ, M. F. 1983. Species adaptation in a protein molecule. *Mol. Biol. Evol.* **1**:1–28.
- POWERS, D. A., T. LAUERMAN, D. CRAWFORD, and L. DI MICHELE. 1991. Genetic mechanisms for adapting to a changing environment. *Annu. Rev. Genet.* **25**:629–659.
- ROSENZWEIG, R. F., R. R. SHARP, D. S. TREVES, and J. ADAMS. 1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* **137**:903–917.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SHYUE, S. K., D. HEWETT-EMMETT, H. G. SPERLING, D. M. HUNT, J. K. BOWMAKER, J. D. MOLLON, and W. H. LI. 1995. Adaptive evolution of color vision genes in higher primates. *Science* **269**:1265–1267.
- STRYER, L. 1995. *Biochemistry*. W. H. Freeman and Co., New York.
- SYNSTAD, B., O. EMMERHOFF, and R. SIREVAG. 1996. Malate dehydrogenase from the green gliding bacterium *Chloroflexus aurantiacus* is phylogenetically related to lactic dehydrogenases. *Arch. Microbiol.* **165**:346–353.
- THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- TRABESINGER-RUEF, N., T. JERMANN, T. ZANKEL, B. DURRANT, G. FRANK, and S. A. BENNER. 1996. Pseudogenes in ribonuclease evolution: a source of new biomacromolecular function? *FEBS Lett.* **382**:319–322.
- WALSH, K., and D. E. KOSHLAND JR. 1985. Branch point control by the phosphorylation state of isocitrate dehydrogenase. *J. Biol. Chem.* **260**:8430–8437.
- WATT, W. B. 1991. *Biochemistry, physiological ecology, and population genetics—the mechanistic tools of evolutionary biology*. *Funct. Ecol.* **5**:145–154.
- WEBER, R. E., T. H. JESSEN, H. MALTE, and J. TAME. 1993. Mutant hemoglobins ( $\alpha^{119}$ -Ala and  $\beta^{55}$ -Ser) functions related to high-altitude respiration in geese. *J. Appl. Physiol.* **75**:2646–2655.

- WILKS, H. M., A. CORTES, D. C. EMERY, D. J. HALSALL, A. R. CLARKE, and J. J. HOLBROOK. 1992. Opportunities and limits in creating new enzymes. Experiences with the NAD-dependent lactate dehydrogenase frameworks of humans and bacteria. *Ann. N.Y. Acad. Sci.* **672**:80–93.
- WILKS, H. M., K. W. HART, R. FEENEY, C. R. DUNN, H. MUIRHEAD, W. N. CHIA, D. A. BARSTOW, T. ATKINSON, A. R. CLARKE, and J. J. HOLBROOK. 1988. A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science* **242**:1541–1544.
- YOKOYAMA, R., B. E. KNOX, and S. YOKOYAMA. 1995. Rhodopsin from fish, *Astyanax*: role of tyrosine 261 in the red shift. *Invest. Ophthalmol. Vis. Res.* **36**:939–945.
- YOKOYAMA, R., and S. YOKOYAMA. 1990. Convergent evolution of the red- and green-like visual pigment genes in fish, *Astyanax fasciatus*, and human. *Proc. Natl. Acad. Sci. USA* **87**:9315–9318.
- YOKOYAMA, S. 1998. Molecular genetic basis of adaptive selection: examples from color vision in vertebrates. *Annu. Rev. Genet.* (in press).
- ZHANG, J., H. ZIQIAN, J. R. H. TAME, G. LU, R. ZHANG, and X. GU. 1996. The crystal structure of a high oxygen affinity species of hemoglobin (bar-headed goose haemoglobin in the oxy form). *J. Mol. Biol.* **255**:484–493.

SHOZO YOKOYAMA, reviewing editor

Accepted November 10, 1997