

The Biochemical Architecture of an Ancient Adaptive Landscape

Mark Lunzer,¹ Stephen P. Miller,¹ Roderick Felsheim,^{1,2}
Antony M. Dean^{1,3*}

Molecular evolution is moving from statistical descriptions of adaptive molecular changes toward predicting the fitness effects of mutations. Here, we characterize the fitness landscape of the six amino acids controlling coenzyme use in isopropylmalate dehydrogenase (IMDH). Although all natural IMDHs use nicotinamide adenine dinucleotide (NAD) as a coenzyme, they can be engineered to use nicotinamide adenine dinucleotide phosphate (NADP) instead. Intermediates between these two phenotypic extremes show that each amino acid contributes additively to enzyme function, with epistatic contributions confined to fitness. The genotype-phenotype-fitness map shows that NAD use is a global optimum.

The role of epistasis—interactions among mutations that produce nonadditive effects on phenotype and fitness—in evolution remains hotly debated (1–8). Although routinely detected in natural and in experimental populations (4, 9, 10), its presence need not imply the existence of multiple peaks in an adaptive landscape (11). Indeed, the question remains: Are adaptive landscapes rugged, or are they smooth?

Characterizing the adaptive landscape of an enzyme is conceptually simple. Mutations controlling a phenotype must be identified. Mutants of intermediate phenotype must be engineered so that the connections between genotype and phenotype (the genotype-phenotype map) can be explored. The fitness of each mutant must be determined so that the relationships between genotype and fitness (the genotype-fitness map) can be established. Finally, a model relating phenotype to fitness (the phenotype-fitness map) is needed to specify the mechanism of selection.

We characterized the adaptive landscape governing coenzyme use by isopropylmalate dehydrogenase (IMDH), an enzyme that catalyzes a step in the biosynthesis of leucine, an essential amino acid. All IMDHs use nicotinamide adenine dinucleotide (NAD) as a coenzyme, although some related isocitrate dehydrogenases (IDHs) lie at the other phenotypic extreme and use nicotinamide adenine dinucleotide phosphate (NADP) instead. Six amino acid residues critical to coenzyme use have been identified (12–15) (Fig. 1). Enzyme performance ($P = k_{\text{cat}}/K_m$) and preference ($P^{\text{NAD}}/P^{\text{NADP}}$ —the number of NADs turned over for each NADP turned over when both coenzymes are present in equimolar concentrations) are phenotypes relevant to fitness (16). The fitnesses of engineered mutants are estimated using the *Escherichia coli* chemostat competition assay (17). Finally, the physiological basis of fitness is described using a simple model of metabolism.

¹BioTechnology Institute, ²Department of Entomology, ³Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN 55108, USA.

*To whom correspondence should be addressed. E-mail: adean@biosci.umn.edu

Protein engineering (18) was used to switch the coenzyme specificity of *E. coli* IMDH from NAD to NADP. Unlike most IMDHs, *E. coli* IMDH already has the Arg-341 found in all NADP-dependent IDHs. The remaining five replacements (Asp236Arg, Asp289Lys, Ile290Tyr, Ala296Val and Gly337Tyr) were introduced into the coenzyme-binding pocket of *E. coli* IMDH by site-directed mutagenesis. Specificity was changed by a factor of 20,000, from a 100-fold preference for NAD ($P^{\text{NAD}} = k_{\text{cat}}^{\text{NAD}}/K_m^{\text{NAD}} = 82 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and $P^{\text{NADP}} = k_{\text{cat}}^{\text{NADP}}/K_m^{\text{NADP}} = 0.84 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$) to a 200-fold preference for NADP ($P^{\text{NAD}} = 0.18 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ and $P^{\text{NADP}} = 37 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$). The engineered “RKYVYR” enzyme is both as active and as specific toward NADP as the wild-type enzyme is toward NAD.

To characterize the genotype-phenotype map, we engineered various combinations of amino acids at the six sites (table S1) (18). The kinetic performances of 164 mutant enzymes toward NAD and NADP were estimated. Nested analyses of variance (NANOVA) (19) of \log_e -

transformed performances and preferences show that a simple linear additive model of the form

$$\log_e(y) = m + \sum_{j=1}^6 a_{i,j}, \text{ df} = 14 \quad (1)$$

explains most of the data (y is performance or preference, m is the sample mean and $a_{i,j}$ is the additive deviation caused by amino acid i at site j): $r^2 = 0.95$ for $\log_e(P^{\text{NAD}})$, $r^2 = 0.92$ for $\log_e(P^{\text{NADP}})$, and $r^2 = 0.97$ for $\log_e(P^{\text{NADP}}/P^{\text{NAD}})$. Performance and preference are dominated by additive effects (Table 1). There is no evidence for epistasis in these genotype-phenotype maps.

Statistical additivity implies thermodynamic additivity. Simple enzyme transition state theory (16) suggests

$$\Delta\Delta G_{i,j}^{\ddagger \text{mut}} = \sum_{j=1}^6 \Delta\Delta G_{i,j}^{\ddagger} = RT \sum_{j=1}^6 (a_{i,j} - a_{i,\text{wt}}) \quad (2)$$

where $\Delta\Delta G_{i,j}^{\ddagger \text{mut}} = \Delta G_{i,j}^{\ddagger \text{wt}} - \Delta G_{i,j}^{\ddagger \text{mut}}$ is the total difference in free energies between the enzyme transition states of the mutant (mut) and the wild type (wt), and $\Delta\Delta G_{i,j}^{\ddagger} = \Delta G_{i,j}^{\ddagger \text{wt}} - \Delta G_{i,j}^{\ddagger \text{mut}}$ represents the difference attributable to replacing a single wt amino acid at site j with an amino acid i . Thermodynamic additivity has been seen in a number of studies of protein folding (20), protein-protein interactions (21), and catalysis (22). The lack of epistasis in coenzyme performance by IMDH is typical of many molecular genotype-phenotype maps, although nonadditive effects arise in some (23–25).

No enzyme performs well with both coenzymes (Fig. 2A). Given thermodynamic additivity, the performances of each of the remaining 512 – 164 = 348 mutant intermediates can be predicted by summing the additive effects (Table 1). Again, the interior of the plot is

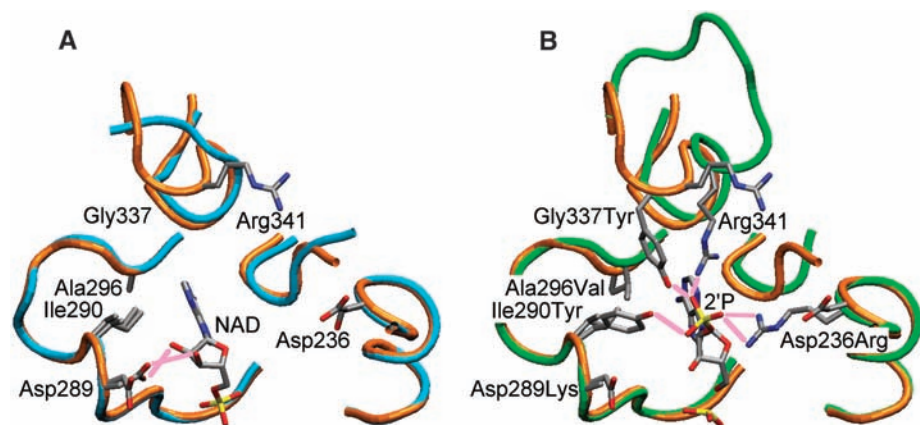


Fig. 1. Crystallographic structures identify amino acids determining coenzyme use. Only key residues are shown (gray, carbon; red, oxygen; blue, nitrogen; yellow phosphorus). (A) Structural alignment of *E. coli* IMDH (13) (brown main-chain; labels designate the amino acid followed by the site number) and *Thermus thermophilus* IMDH (14) (blue main-chain) showing the double H-bond (pink lines) critical to NAD use. (B) Structural alignment of *E. coli* IMDH and *E. coli* IDH (green main-chain) with NADP bound (15) showing IDH residues (following the IMDH site number) H-bonding to the 2'-phosphate (2'P) of bound NADP (H-bonds from the disordered 289Lys not shown).

empty (Fig. 2B). Evidently, a performance trade-off restricts severely the possible phenotypes upon which selection can act.

The genotype-fitness map reveals strong epistasis in fitness. Ninety IMDH mutants, representing a stratified sample of kinetic performances, were recombined individually into the *leu* operon on the *E. coli* chromosome and their fitnesses relative to those of the wild type determined in chemostat competition (17, 18). A NANOVA (19) (residues within sites) of fitness assuming only additive effects produced a poor fit ($r^2 = 0.85$). Interactions were not mod-

eled because they required many more degrees of freedom than our NANOVA design permitted. Eliminating “transitional” residues (table S1) (18) from the analysis allows pairwise interactions to be modeled. The resulting NANOVA included six significant pairwise interactions ($r^2 = 0.99$) and was a marked improvement over the strictly additive model ($r^2 = 0.87$). Hence, epistasis is present in the genotype-fitness map.

The phenotype-fitness map shows how epistasis, absent in the genotype-phenotype map, arises in the genotype-fitness map. Fitness is

commonly a concave function of enzyme performance (26, 27). Assume fitness (w) is a hyperbolic function of intracellular IMDH performance toward isopropylmalate (V_{\max}/K_m^{ipm})

$$w = w_{\max} \frac{V_{\max}}{K_m^{ipm}} / \left(K + \frac{V_{\max}}{K_m^{ipm}} \right) \quad (3)$$

where w_{\max} is maximum fitness when $V_{\max}/K_m^{ipm} \rightarrow \infty$, K is the performance necessary to produce $w_{\max}/2$, V_{\max} is the maximum intracellular rate when isopropylmalate is saturating, and K_m^{ipm} is the concentration of isopropylmalate necessary to produce $V_{\max}/2$. The concave nature of Eq. 3 is typical of the nonlinear responses in metabolic flux to changes in enzyme activities that produce genetic dominance, phenotypic robustness, and epistasis at higher levels of biological organization (28, 29). Epistasis in fitness arises because the same mutation producing the same proportional increase in activity in a wild type as in a mutant compromised by mutation will cause a smaller increase in fitness in the wild type (because $w \approx w_{\max}$ when $V_{\max}/K_m^{ipm} \gg K$) than it will in the mutant [because $w = (w_{\max}/K)V_{\max}/K_m^{ipm}$ when $V_{\max}/K_m^{ipm} \ll K$].

Substituting a kinetic model describing the IMDH random bi-ter kinetic mechanism for V_{\max}/K_m^{ipm} in Eq. 3 and collecting terms produces the phenotype-fitness map in terms of the coenzyme kinetics,

$$w = w_{\max} \left(\frac{k_{cat}^{NAD}}{K_m^{NAD}} + \frac{k_{cat}^{NADP}}{K_m^{NADP}} R \right) / \left(A \left(1 + \frac{B}{K_m^{NAD}} + \frac{C}{K_m^{NADP}} \right) + \left(\frac{k_{cat}^{NAD}}{K_m^{NAD}} + \frac{k_{cat}^{NADP}}{K_m^{NADP}} R \right) \right) \quad (4)$$

Table 1. Additive effects ($a_{i,j}$) of amino acid replacements on coenzyme use.

Site	Residue	Performance effects				Preference effect*	
		NAD	SE	NADP	SE	NAD-NADP	SE
236	Arg	-0.250	±0.039	0.735	±0.046	-0.985	±0.041
	Asp	0.250	±0.039	-0.735	±0.046	0.985	±0.041
289	Lys	-0.657	±0.060	1.547	±0.071	-2.204	±0.062
	Asp	0.850	±0.057	-0.722	±0.069	1.572	±0.062
	Asn†	-0.019	±0.059	0.506	±0.072	-0.509	±0.045
	Glu†	-0.183	±0.074	-1.332	±0.089	1.148	±0.079
290	Tyr	-0.680	±0.082	0.659	±0.092	-1.367	±0.102
	Ile	2.218	±0.078	0.255	±0.094	1.963	±0.083
	His†	-0.824	±0.099	0.684	±0.119	-1.508	±0.106
	Leu†	2.267	±0.076	1.058	±0.120	1.209	±0.109
	Phe†	-0.516	±0.099	-0.869	±0.119	0.353	±0.102
	Lys†	0.159‡	±0.109	0.911	±0.123	-0.729	±0.107
	Asn†	-1.981	±0.096	-1.713	±0.116	-0.268	±0.112
296	Gln†	-0.633	±0.094	-1.039	±0.112	0.406	±0.109
	Val	-0.672	±0.038	-0.577	±0.047	-0.095	±0.041
	Ala	0.672	±0.038	0.577	±0.047	0.095	±0.041
337	Tyr	-0.193	±0.094	0.058‡	±0.046	-0.139	±0.040
	Gly	0.193	±0.094	-0.058‡	±0.046	0.139	±0.040
341	Arg	0.291	±0.047	0.375	±0.058	-0.084‡	±0.050
	Ser	-0.291	±0.047	-0.375	±0.058	0.084‡	±0.050
	m	-0.095	±0.055	0.013	±0.065	-0.108	±0.056

*The preference effect is defined as $a_{w_{i,j}} - a_{i,j}$. †Possible transitional replacements attributable to multiple base substitutions needed to exchange Asp for Lys at site 289 and Ile for Tyr at site 290. ‡Not significantly different from zero.

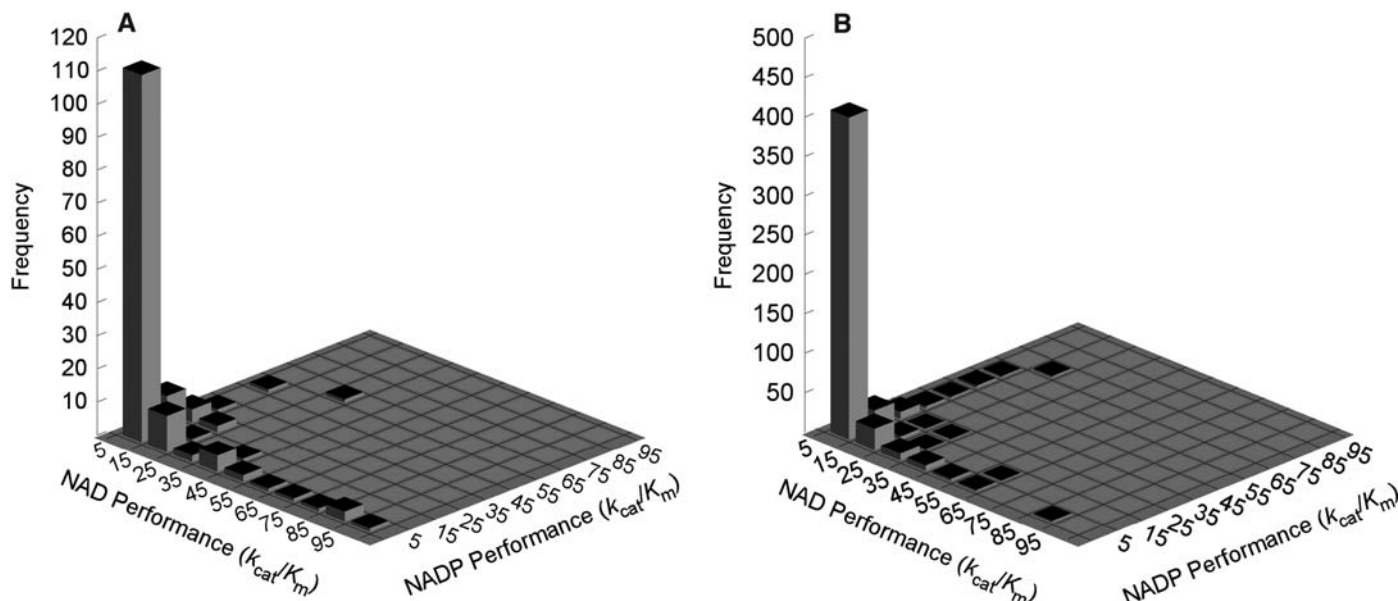
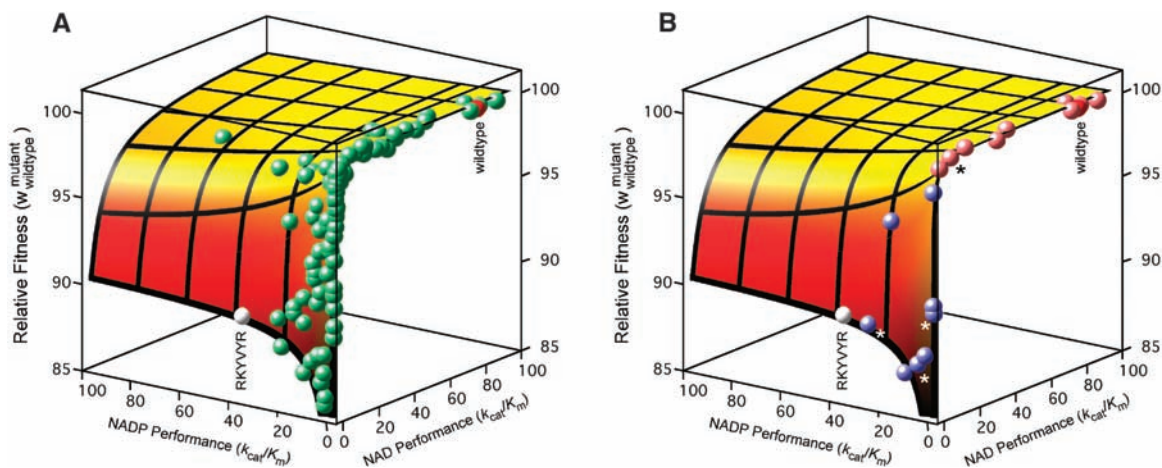


Fig. 2. Performances ($10^3 \text{ M}^{-1} \text{ s}^{-1}$) of engineered IMDH mutants toward NAD and NADP reveal a trade-off in enzyme function. (A) Distribution of performances for the 164 engineered enzymes constructed. (B) Distribu-

tion of performances for 512 genotypic intermediates predicted on the assumption of thermodynamic additivity. Symptomatic of a trade-off in performance, the interiors of both plots are devoid of mutants.

Fig. 3. The phenotype-fitness map of IMDH. (A) The fitnesses (green spheres of fitness radius $w = 0.5\% \approx 2 \text{ SE}$) of 90 engineered mutants plotted against their coenzyme performances (10³ M⁻¹s⁻¹). The fitted surface is the estimated phenotype-fitness map (Eq. 4). It reveals a single broad adaptive peak on which resides the NAD-specific wild-type enzyme (red sphere). NADP use is advantageous only in mutants with very poor NAD performance (e.g., the RKYVYR mutant, white sphere). (B) Escape from the lower NADP-use plateau to the higher NAD-use plateau is possible because some single amino acid replacements (blue spheres from the RKYVYR mutant, pink spheres from



the wild type; asterisks denote fitness values predicted from kinetic data) produce sufficiently large effects on performance and fitness that the maladaptive valley at the origin is bypassed.

where A , B , C , D , and R are constants associated with kinetic terms and coenzyme pools unaffected by our mutations (18). Equation 4 is a hypothesis that describes fitness in terms of the kinetic parameters obtained for each mutant enzyme. It fits the data well—nonlinear regression yields $r^2 = 0.97$.

Noting that the $1/K_m^{NAD(P)}$ values are necessarily correlated with—and hence can be collapsed into—the $k_{cat}^{NAD(P)}/K_m^{NAD(P)}$ values allows the phenotype-fitness map to be visualized (Fig. 3A). IMDH fitness is maximized exclusively by high performance with, and high preference for, NAD (the wild type on the right fitness plateau). NADP use is suboptimal.

Product inhibition by NADPH lowers the fitness of NADP users. Most intracellular NADP is in the reduced form, NADPH, which has a 30-fold higher affinity for IMDH ($K_m^{NADP}/K_m^{NADPH} \approx 30$). Thermodynamic additivity ensures that mutations in the coenzyme-binding pocket that improve performance with NADP also increase affinity for NADPH (fig. S1) (18). Consequently, any benefit gained by improved performance with NADP is offset by intensified product inhibition by abundant NADPH. A similar correlation for NAD use does not generate a measurable cost because so little intracellular NAD is in the reduced NADH form ($B \ll C$, Eq. 4) (18).

The phenotype-fitness map has a single peak (the broad NAD-use plateau in Fig. 3A). In principle, the trade-off in performance (Fig. 2) could combine with mutations of small functional effect to force all paths from the NADP-dependent RKYVYR mutant to the NAD-dependent wild type through the maladaptive valley at the origin. The result would be two peaks on the genotype-fitness map, with the higher NAD-use plateau inaccessible from the lower NADP-specific allele.

The genotype-fitness map has just one adaptive peak. The fitnesses of all 512 mutant genotypes were predicted using Eq. 4 with enzyme performances calculated assuming thermodynamic additivity (Table 1 and Fig. 2B). A

fitter genotype was mutationally accessible to every genotype—except for the NAD-dependent wild type, which was predicted to be fittest. Only one peak is expected because some mutations have sufficiently large phenotypic effects that the maladaptive valley at the origin of the phenotype-fitness map is bypassed (Fig. 3B).

Defining all mutational connectivities between the genotypes on the phenotype-fitness map completes the IMDH adaptive landscape. With its single peak, the landscape is far less rugged than those envisioned by Wright (1). With epistasis consigned to a minor role, this landscape lies closer to Fisher's conception (2) than to any other. Ironically, the landscape might have been more rugged, with two adaptive peaks, had another of Fisher's assumptions, that of many mutations each of small effect (2), proven correct.

Coenzyme use by IMDH is likely representative of a large class of adaptive landscapes in which thermodynamic additivity in molecular function (Table 1) (20–22) combines with concave fitness functions at the organismal level (Eq. 3) (26, 27). Nevertheless, landscapes are likely to be more rugged whenever epistasis in genotype-phenotype maps (23–25) combines with complex phenotype-fitness maps (26).

Our landscape provides a mechanism sufficient to explain why all IMDHs use NAD. Conservation of this phenotype implies that we have characterized an ancient adaptive landscape—unchanged in all lineages, in all habitats, since the last common ancestor. Such ancient landscapes can explain adaptive processes at the very dawn of life's diversity.

References and Notes

1. S. Wright, in *Proceedings of the Sixth International Congress of Genetics*, D. F. Jones, Ed. (Brooklyn Botanic Garden, Menasha, WI, 1932), pp. 356–366.
2. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
3. J. Maynard Smith, *Nature* **225**, 563 (1970).
4. M. C. Whitlock, P. C. Phillips, F. B.-G. Moore, J. Tonsor, *Annu. Rev. Ecol. Syst.* **26**, 601 (1995).

5. C. L. Burch, L. Chao, *Nature* **406**, 625 (2000).
6. J. A. Coyne, N. H. Barton, M. Turelli, *J. Org. Evol.* **51**, 643 (1997).
7. M. J. Wade, C. J. Goodnight, *Evol. Int. J. Org. Evol.* **52**, 1537 (1998).
8. D. M. Weinreich, L. Chao, *Evol. Int. J. Org. Evol.* **59**, 1175 (2005).
9. S. J. Schrag, V. Perrot, B. R. Levin, *Proc. R. Soc. Lond. B. Biol. Sci.* **264**, 1287 (1997).
10. S. Maisnier-Patin, O. G. Berg, L. Lijas, D. I. Andersson, *Mol. Microbiol.* **46**, 355 (2002).
11. D. M. Weinreich, R. A. Watson, L. Chao, *Evol. Int. J. Org. Evol.* **59**, 1165 (2005).
12. G. Zhu, G. B. Golding, A. M. Dean, *Science* **307**, 1279 (2005).
13. G. Wallon et al., *J. Mol. Biol.* **266**, 1016 (1997).
14. J. H. Hurley, A. M. Dean, *Structure* **2**, 1007 (1984).
15. J. H. Hurley, A. M. Dean, D. E. Koshland Jr., R. M. Stroud, *Biochemistry* **30**, 8671 (1991).
16. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (Freeman, New York, 1999).
17. M. Lunzer, A. Natarajan, D. E. Dykhuizen, A. M. Dean, *Genetics* **162**, 485 (2002).
18. Materials and methods are available as supporting material on Science Online.
19. R. R. Sokal, F. J. Rohlf, *Biometry* (Freeman, New York, ed. 3, 1995).
20. J. A. Wells, *Biochemistry* **29**, 8509 (1990).
21. W. S. Sandberg, T. C. Terwilliger, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8367 (1993).
22. T. Aita, M. Iwakura, Y. Huseimi, *Protein Eng.* **14**, 633 (2001).
23. M. A. Qasim et al., *Biochemistry* **42**, 6460 (2003).
24. X. Wang, G. Minasov, B. K. Shoichet, *J. Mol. Biol.* **320**, 85 (2002).
25. M. A. DePristo, D. M. Weinreich, D. L. Hartl, *Nat. Rev. Genet.* **6**, 678 (2005).
26. D. L. Hartl, D. E. Dykhuizen, A. M. Dean, *Genetics* **111**, 655 (1985).
27. A. M. Dean, *Genetics* **139**, 19 (1995).
28. H. Kacser, J. Burns, *Genetics* **97**, 639 (1981).
29. U. Alon, M. G. Surette, N. Barkai, S. Leibler, *Nature* **397**, 168 (1999).
30. We thank B. Kerr, L. Merlo and four anonymous reviewers for constructive criticism of the manuscript. Supported by NIH grant GM060611 (A.M.D.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/310/5747/499/DC1

SOM Text
Materials and Methods
Figs. S1 and S2
Table S1
References and Notes

2 June 2005; accepted 15 September 2005
10.1126/science.1115649