Protein stability promotes evolvability

Jesse D. Bloom*†, Sy T. Labthavikul*, Christopher R. Otey‡, and Frances H. Arnold*†

*Division of Chemistry and Chemical Engineering, ‡Biochemistry and Molecular Biophysics Option, Mail Code 210-41, California Institute of Technology, Pasadena, CA 91125

Edited by Michael Levitt, Stanford University School of Medicine, Stanford, CA, and approved February 14, 2006 (received for review November 21, 2005)

The biophysical properties that enable proteins to so readily evolve to perform diverse biochemical tasks are largely unknown. Here, we show that a protein's capacity to evolve is enhanced by the mutational robustness conferred by extra stability. We use simulations with model lattice proteins to demonstrate how extra stability increases evolvability by allowing a protein to accept a wider range of beneficial mutations while still folding to its native structure. We confirm this view experimentally by mutating marginally stable and thermostable variants of cytochrome P450 BM3. Mutants of the stabilized parent were more likely to exhibit new or improved functions. Only the stabilized P450 parent could tolerate the highly destabilizing mutations needed to confer novel activities such as hydroxylating the antiinflammatory drug naproxen. Our work establishes a crucial link between protein stability and evolution. We show that we can exploit this link to discover protein functions, and we suggest how natural evolution might do the same.

CYP102A1 | cytochrome P450 | directed evolution | lattice protein | mutational robustness

Biological systems are evolvable in the sense that mutation and selection are able to create new or improved phenotypes. A major biological question is how a system's physical properties influence its capacity to evolve (1, 2). In recent years, understanding the determinants of evolvability has also become an important practical concern as researchers increasingly use evolution to engineer everything from proteins (3) to designs for civil engineering structures (4).

Proteins are one of the simplest and best examples of evolvable biological systems, because they possess biochemical functions that can be altered with just a few mutations (5). One property that has been broadly hypothesized to contribute to evolvability is robustness to mutations (6), and proteins are often quite mutationally robust, with more than half of the single mutants of many proteins retaining their native functions (7–10). Because proteins usually must fold to function, and because mutated proteins generally adopt the original native structure if they fold at all (11, 12), retention of the basic native structure is normally a prerequisite for the acquisition of new functions. Extra thermodynamic stability makes a protein's native structure and function more robust to random mutations by increasing the fraction of mutants that continue to possess the minimal stability required to fold (10, 13).

Here, we investigate how stability affects a protein's evolvability by using controlled experiments to measure the fraction of a protein's mutants that exhibit new or improved function. We first use a simple computational model to establish a framework for understanding the relationships among protein stability, mutational robustness, and evolvability. We then validate this framework with experiments on members of the biochemically important cytochrome P450 enzyme family and describe specific examples that illustrate the biophysical basis of the connection between protein stability and evolvability. Finally, we discuss the implications of our work for understanding natural protein evolution and designing better protein engineering strategies.

Results

Simulations with Model Lattice Proteins. We used a simple conceptual framework (10) for understanding the relationship between protein stability and evolution. The premise is that evolution selects for a protein's biochemical function rather than its stability. However, because a protein's function typically depends on its ability to fold to a thermodynamically stable native structure (14), stability is still constrained during evolution. Specifically, we imagine that a protein must fold to its native structure with some minimal stability to remain folded at physiological conditions. If a protein fails to meet this minimal stability threshold, then it will neither fold nor function. If a protein does fold with at least the minimal required stability, then evolution selects for a protein's function and is indifferent to the amount of extra stability it possesses. Most proteins, however, will still be marginally stable, because highly stable sequences are rare (15).

This conceptual framework formed the basis for simulations with lattice proteins. Lattice proteins are highly simplified protein models that are useful tools for studying protein folding and evolution (16, 17). Our lattice proteins were chains of 20 amino acids that fold on a two-dimensional lattice, with the energy of each conformation equal to the sum of the pairwise interactions between nonbonded amino acids (18). Each lattice protein can occupy any of 41,889,578 possible conformations, and, by summing over all of these conformations, we could exactly determine the partition function and free energy of folding (ΔG_f). We set a minimal stability threshold for our lattice proteins by requiring them to fold to the original native structure with a stability of $\Delta G_f \leq 0$ (in no case did we observe a protein that stably folded to a new structure), which is equivalent to requiring the native structure to have a lower free energy than the ensemble of all nonnative conformations. For those proteins that stably folded, we measured function as the binding energy of the folded protein to a small rigid ligand (19), as shown in Fig. 1*A*. Our model, therefore, recapitulated the essential requirements imposed on real proteins of simultaneously folding and performing a biochemical task.

We first evolved a model protein to stably fold and strongly bind a ligand (Fig. 1*A*). This evolved protein had a stability of $\Delta G_f = -0.5$, meaning that it was only marginally stable, as is typical for real proteins (20). We then simulated the process of directed evolution with two rounds of random mutagenesis by error-prone PCR and screening to identify a stabilized variant of our model protein ($\Delta G_f = -1.5$) that contained three amino acid substitutions and exhibited the same ligand-binding energy as the original protein. To examine the evolvabilities of the original and stabilized model proteins, we computationally simulated screening libraries of 1,500 randomly mutated sequences for mutants that bound to new ligands with at least twice the affinity of the parent proteins. For all four new ligands we examined, the parent proteins bound the new ligand with equal affinity, yet, each time, the mutant library from the stabilized parent produced more than twice as many unique improved mutants (Fig. 1*B*).

Fig. 1*C* shows why the stabilized model protein was more

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

†To whom correspondence may be addressed. E-mail: jesse.bloom@gmail.com or frances@cheme.caltech.edu.

© 2006 by The National Academy of Sciences of the USA

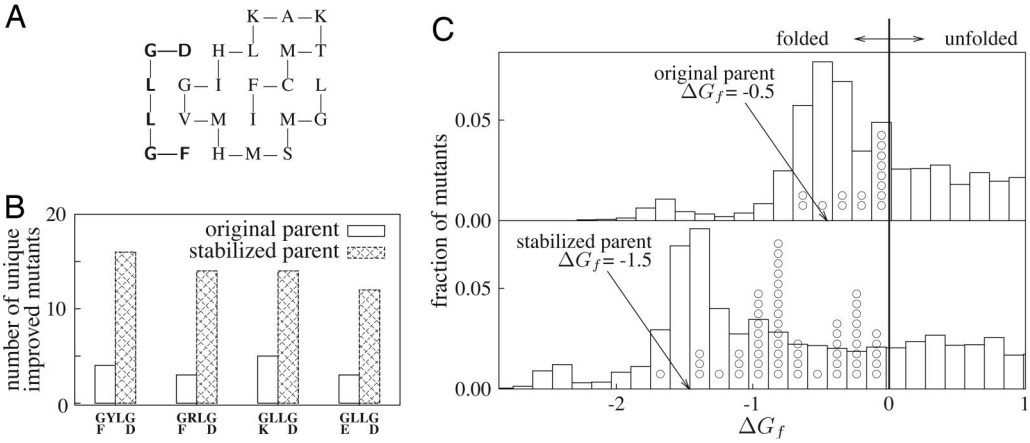


Fig. 1. Increased stability enhances evolvability of a model lattice protein. (*A*) The original model protein (right side) that had been evolved to bind to a rigid ligand (left side, in bold). (*B*) Mutants of a stabilized model protein were more likely than mutants of the original model protein to show improved binding to the four new ligands shown below the bars. The bars show the number of mutants of 1,500 screened that bound the new ligand with at least twice the affinity of the parent. (*C*) The stabilized model protein was more evolvable because more of its destabilized but improved mutants satisfied the minimal stability cutoff. The bars show the distribution of stabilities among all mutants in the libraries, and the circles show the stabilities of the improved mutants.

evolvable. The mutants in both libraries exhibited similar changes in stability ($\Delta\Delta G$ values), but the extra stability of the stabilized protein meant that a larger fraction of its mutants continued to fold (46% versus 35% among all mutants with at least one mutation), confirming previous findings that more stable lattice proteins are more robust to mutations (10, 21).

The improved mutants tended to be destabilized and so were more frequent in the library from the stabilized parent. Although the more stable parent had a <50% increase in the fraction of mutants that folded, it had nearly four times more improved mutants (56 versus 15). The fact that extra stability increases the number of improved mutants much more than it increases the number of mutants that retain parental function indicates that improved mutants tended to be more destabilized than the typical folded mutant.

Experiments on Cytochrome P450 BM3 Variants. To experimentally test the effect of stability on the evolvability of real proteins, we randomly mutated two variants of a cytochrome P450 BM3 (also known as CYP102A1) heme domain peroxygenase (22) and screened for mutants with new or improved activity on five substrates. The cytochrome P450 superfamily contains members involved in important biochemical processes such as drug metabolism and steroid biosynthesis (23, 24). P450 BM3 catalyzes subterminal hydroxylation of medium- and long-chain fatty acids (25). The 21B3 variant is a laboratory-evolved version of the P450 BM3 heme domain that efficiently hydroxylates 12-*p*-nitrophenoxycarboxylic acid (12-pNCA, structure shown in Fig. 2), using hydrogen peroxide in place of the NADPH cofactor and oxygen (22). The 5H6 variant was created by laboratory evolution of 21B3, selecting for mutants that were more thermostable while retaining activity on 12-pNCA (26). We quantified the stabilities of the enzymes by the temperature (T_{50}) at which half of the protein irreversibly denatured after a 10-min incubation. Because denaturation is irreversible, these T_{50} values are not equilibrium thermodynamic measurements and so cannot be directly related to ΔG_f. However, the T_{50} values were highly correlated with the stability to irreversible denaturation by urea, supporting the notion that they reflect universal aspects of protein stability rather than unique characteristics of the process of irreversible thermal denaturation (see *Supporting Materials and Methods* and Fig. 5, which are published as supporting information on the PNAS web site). As measured by the T_{50} values, the 21B3 enzyme is only marginally stable (T_{50} = 47°C), whereas 5H6 is much more stable (T_{50} = 62°C) (melting curves are shown in Fig. 6, which is published as supporting information on the PNAS web site). The 5H6 enzyme differs from 21B3 at 8 residues (of 464 total). Both variants displayed nearly the same activity (measured as total turnovers) on 12-pNCA and all other substrates examined in this work.

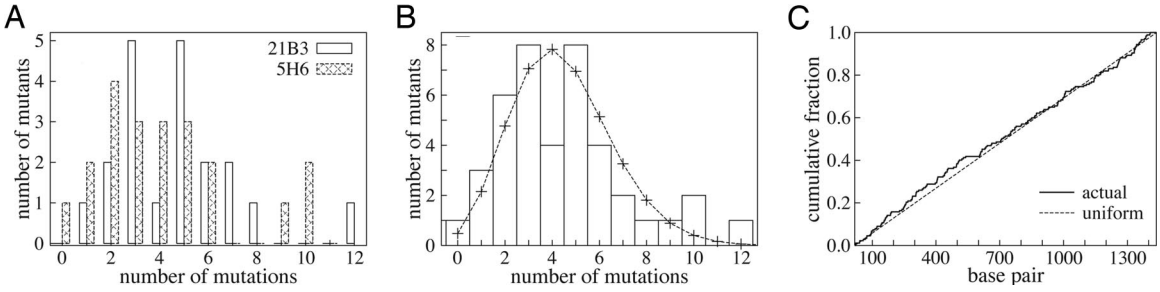


Fig. 2. Distribution of mutations in the two P450 error-prone PCR libraries. (*A*) The distribution of mutations among 20 randomly chosen 21B3 mutants and 21 randomly chosen 5H6 mutants. The distributions are statistically indistinguishable ($P = 0.84$). (*B*) The distribution of mutations among all 41 sequenced mutants is consistent with the theoretical prediction for an error-prone PCR library (lines) (49, 50) ($P = 0.11$). (*C*) The mutations are uniformly distributed along the gene ($P = 0.66$). The lines show the cumulative fraction of mutations that occur at or before that position in the gene. All *P* values are from Kolmogorov–Smirnov tests (53) and represent the probability that the samples or theoretical curves would differ by at least this much if they were generated by the same underlying distribution.

Table 1. Mutation frequencies in error-prone PCR libraries

Base pairs sequenced	58,719
Total mutations	182
Mutation frequency, %	0.31 ± 0.02
Avg. mutations per gene	4.5 ± 0.3
Synonymous mutations, %	28
Nonsynonymous mutations, %	63
Frameshift/nonsense mutations, %	9
Mutation types, %	
A → T, T → A	25
A → C, T → G	6
A → G, T → C	53
G → A, C → T	10
G → C, C → G	0
G → T, C → A	3
Frameshift	3

Statistics are for all 41 randomly chosen mutants. Standard errors are calculated assuming Poisson counting statistics.

We created mutant libraries of both 21B3 and 5H6 using error-prone PCR. The libraries were generated under identical conditions and had the same distributions of mutations (Fig. 2). The overall mutation rate was 4.5 ± 0.3 nucleotide mutations per gene (Table 1). We examined 522 mutants from each library for retention of folding, as assayed by the characteristic Soret band at 450 nm in the carbon monoxide-binding difference spectrum (27). As expected, mutants of the stabilized 5H6 protein were more likely than those of the 21B3 protein to fold (61% of 5H6 mutants contained at least half the folded protein of the parent versus 33% for 21B3, raw data are shown in Fig. 7 and Table 3, which are published as supporting information on the PNAS web site). Most of these folded mutants retained at least half the parental activity on 12-pNCA (94% and 96% for 5H6 and 21B3), indicating that mutations that disrupted parental function generally did so by preventing the formation of properly folded protein, confirming the experimental findings of ref. 10 that more stable proteins are more robust to mutations.

We examined the evolvability of the 21B3 and 5H6 enzymes by screening for mutants that hydroxylated any of five new substrates: the antiinflammatory drug naproxen, 3-phenoxytoluene, phenoxyacetic acid, the β-adrenergic receptor blocking agent propranolol, and 2-methylbenzofuran (structures shown in Fig. 9, which is published as supporting information on the PNAS web site). We screened for hydroxylation activity using the 4-aminoantipyrene (4-AAP) assay to measure the total amount of product after completion of the reaction (28) and determined that neither 21B3 nor 5H6 had detectable activity on the first three substrates, both had equal weak activity on propranolol, and 21B3 had trace activity on 2-methylbenzofuran (Table 2). We used consistent quantitative criteria to identify mutants that had either acquired new activity or improved by >50% over the parental level in the 4-AAP assay. We screened 8,160 mutant–substrate pairs for each parent. From these pairs, we identified 13 improved mutants of 5H6 and 4 improved mutants of 21B3 (Fig. 3*A* and Table 2). All of the improved mutants had unique protein sequences (given in Table 4, which is published as supporting information on the PNAS web site). Thus, we found

Table 2. Summary of improved P450 mutants

	21B3				5H6			
Substrate	Protein	Activity	m_{aa}	T_{50}, °C	Protein	Activity	m_{aa}	T_{50}, °C
Propranolol (1,190 screened)	Parent	0.07 ± 0.02	0	47	Parent	0.08 ± 0.02	0	62
	14C10	0.29 ± 0.05	3	44	27G8	0.15 ± 0.01	1	65
	27B2	0.15 ± 0.03	3	49	27G12	0.25 ± 0.03	7	55
	31B12	0.14 ± 0.02	2	48	30B10	0.19 ± 0.02	6	58
					32F7	0.15 ± 0.02	4	55
					36G11	0.22 ± 0.02	3	64
					37F4	0.21 ± 0.02	2	63
					38F11	0.16 ± 0.02	3	60
3-Phenoxytoluene (2,210 screened)	Parent	None	0	47	Parent	None	0	62
					20D1	0.05 ± 0.02	1	60
					23E4	0.30 ± 0.03	5	48
					28B5	0.05 ± 0.03	3	60
					29G8	0.05 ± 0.02	2	50
Naproxen (2,210 screened)	Parent	None	0	47	Parent	None	0	62
					13C9	0.16 ± 0.02	1	54
Phenoxyacetic acid (1,785 screened)	Parent	None	0	47	Parent	None	0	62
	38F7	0.30 ± 0.04	3	46				
2-Methylbenzofuran (765 screened)	Parent	0.07 ± 0.02	0	47	Parent	None	0	62
					32H1	0.11 ± 0.02	1	61

The leftmost column gives the total number of mutants of each parent screened on that substrate. Subsequent columns give the activity, number of amino acid substitutions (m_{aa}), and stabilities as measured by the T_{50} values. Mutants are named according to the plate and well in which they were found, and sequences are given in Table 5. Activities represent the median ± SD of the total product formed per well of 80 μl of 5 μM protein, as measured by the A_{500} in the 4-AAP assay (raw data are in Fig. 8, which is published as supporting information on the PNAS web site.) and are indicated as ''none'' when indistinguishable from the background. With an extinction coefficient of 4,800 M^{-1} cm^{-1} for the 4-AAP/phenol complex (52), each unit of A_{500} corresponds to ≈80 nmol of product from the ≈0.4 nmol of protein per well.

EVOLUTION

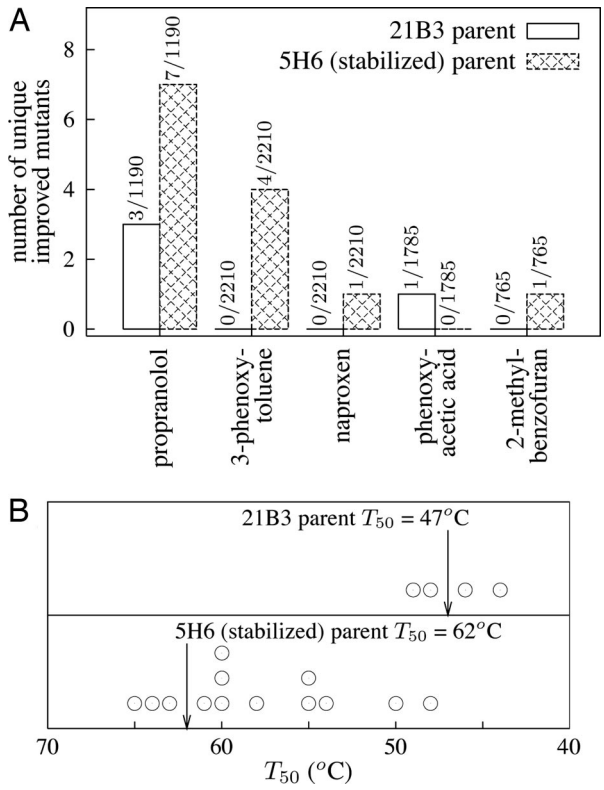


Fig. 3. Increased stability enhances evolvability of the P450 BM3 heme domain. (*A*) The stable 5H6 protein yielded more mutants with new or improved activity than the marginally stable 21B3 protein. The counts above the bars give the number of improved mutants of the total number of mutants screened. (*B*) Some of the improved 5H6 mutants were greatly destabilized relative to the parent protein, whereas the stabilities of the improved 21B3 mutants clustered around those of the parent protein (circles show T_{50} values for improved mutants).

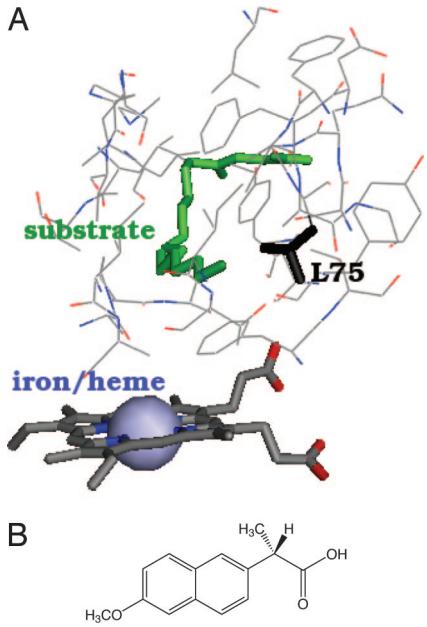


Fig. 4. The functionally innovative but destabilizing L75R mutation can only be tolerated by the stabilized parent. (*A*) Leucine 75 is positioned close to the substrate in the hydrophobic P450 BM3 substrate-binding pocket (30). Mutating L75 to a positively charged arginine conferred naproxen activity on the stabilized 5H6 parent but disrupted the proper folding of the marginally stable 21B3 parent. (*B*) The antiinflammatory drug naproxen, which contains a negatively charged carboxylic acid group.

more than three times more improved mutants in the 5H6 library than in the 21B3 library.

To assess the importance of stability in conferring enhanced evolvability on the 5H6 protein, we measured the stabilities of all improved mutants (melting curves are in Fig. 6). Fig. 3*B* shows that none of the improved 21B3 mutants was destabilized by >3°C but that the thermostable 5H6 parent produced improved mutants that were destabilized by as much as 14°C. We identified specific beneficial but destabilizing substitutions that could be made only in the stabilized parent. For example, neither 21B3 nor 5H6 exhibited activity on naproxen, presumably because the negatively charged naproxen molecule does not enter the hydrophobic P450 BM3 substrate-binding pocket. Mutating leucine 75 in the substrate-binding pocket to arginine allowed 5H6 to hydroxylate naproxen by providing a compensating positive charge for the naproxen molecule (Fig. 4). However, burial of this arginine residue in the hydrophobic binding pocket substantially destabilized the 5H6 mutant ($\Delta T_{50} = -8$°C). When we made the same substitution to 21B3, we could recover only inactive and improperly folded protein (as indicated by a carbon monoxide difference spectrum peak at 420 nm (29), as shown in Fig. 10, which is published as supporting information on the PNAS web site). The F275S substitution (located 12 Å from the substrate) (30) is another example of a beneficial substitution that could be made only in the stabilized parent. This substitution conferred 3-phenoxytoluene activity on the 5H6 parent, but decreased the T_{50} by 7°C. When we made this substitution in 21B3, we again could not recover any folded protein (see Fig. 11, which is published as supporting information on the PNAS web site). In contrast, the F205L substitution (located near the substrate-binding pocket) (30) found in a 21B3 mutant with improved activity on propranolol did not have a substantial effect on stability, and slightly improved the activity of both 21B3 and 5H6 when introduced into those sequences (see Fig. 12, which is published as supporting information on the PNAS web site).

Discussion

We have shown that more stable proteins are more evolvable because they are better able to tolerate functionally beneficial but destabilizing mutations. Our work touches on the relationship between protein stability and function, which has historically been a subject of considerable confusion. Despite repeated speculation to the contrary (31–33), high stability and function are not inherently incompatible, because a wealth of experiments have shown that proteins can be dramatically stabilized without sacrificing their biological functions (34–37). But protein stability and function often appear to trade off at the level of individual mutations. This apparent tradeoff is, at least partly, because of the simple fact that most randomly chosen mutations (functionally beneficial or not) are destabilizing (38–41). In addition, residues in a protein's active site often must satisfy functional constraints (such as maintaining buried charges or cavities in a protein's interior) that make them poorly optimized for stability (42–44). Therefore, mutating active-site residues often enhances stability at the cost of function (42–44), and, likewise, acquiring new functions can require destabilizing mutations (as is the case for our L75R mutation in P450, which confers activity on naproxen by burying a positive charge). However, it remains unclear whether active-site constraints intensify the tradeoff between stability and functional evolution, because a seemingly opposite argument can be made that mutations to an active site that is already poorly optimized for

stability should be less destabilizing than typical mutations (they could even enhance stability if, for example, they confer function on smaller substrates by reducing the size of a cavity in a protein's interior). If functionally innovative mutations tend to be more destabilizing than random mutations, then extra protein stability should enhance the rate of functional innovation more than it enhances the mutational robustness of the native function. In our lattice protein simulations, extra stability increased the rate of functional innovation by nearly 300%, whereas it increased mutational robustness by only 50%; however, we feel that our lattice model is too crude to confidently extrapolate conclusions involving residue-level properties to real proteins. In our P450 experiments, extra stability also increased the number of functionally improved mutants (from 4 to 13) more than it increased mutational robustness (by a factor of 1.8); however, here, the statistics are too poor to conclude that functional innovation is improved significantly more than mutational robustness. Therefore, in our minds, it remains unclear whether extra protein stability promotes evolvability simply by improving the tolerance to all mutations (some of which happen to be functionally beneficial) or whether the effect is further amplified by a tendency for functionally innovative mutations to be especially destabilizing.

In either case, our work argues, quite generally, that extra stability will enhance evolvability. Although it is clearly possible to stabilize proteins without interfering with their functions (34–37), proteins tend to be only marginally more stable than is required by their environment (20). This marginal stability is probably because of the fact that natural selection does not directly favor extra stability in the face of predominantly destabilizing mutations, causing stability to drift toward the minimum evolutionary requirement (15, 45). Naturally evolving proteins must, therefore, wait for functionally neutral mutations to stabilize the structure to counterbalance the effects of other destabilizing but functionally beneficial mutations (46). In this sense, a protein's stability represents a hidden dimension in evolution: Extra stability is neutral with respect to selection for protein function, but it can be crucial in allowing a protein to tolerate mutations that confer beneficial phenotypes. We have shown that protein engineering by directed evolution is more effective if direct selection for extra stability is used to increase a protein's evolvability. The extent to which natural evolution might also select for evolvability has been the subject of much recent theoretical speculation (2, 6, 47). We suggest that one possible mechanism by which natural evolution could increase evolvability would be to stabilize proteins undergoing adaptive evolution or provide systems to buffer the effects of destabilizing protein mutations.

Methods

P450 Mutant Libraries. We used error-prone PCR to create mutant libraries of the marginally stable 21B3 (22) and the thermostable 5H6 (26) variants of the cytochrome P450 BM3 heme domain. The template DNA was the appropriate gene cloned into the pCWori (48) plasmid as described in refs. 22 and 26. We confirmed the sequences of the 21B3 and 5H6 genes by sequencing them with the primers *pCWori_for* (5′-GAAACAGGATCCATCGATGCTTAGGAGGTCAT-3′, *pCWori_rev* (5′-GCGTATCACGAGGCCCTTTCGTCTTCAAGC-3′), and *pCWori_mid_rev* (5′-CCAGCTTGTGGCCAACCCGAC-3′). The sequences matched those that were reported (22, 26), with 21B3 containing 10 amino acid substitutions relative to the wild-type P450 BM3 heme domain (I58V, F87A, H100R, F107L, A135S, M145V, N239H, S274T, K434E, and V446I in the numbering scheme, where residue 1 is the threonine after the cleaved N-terminal methionine) and 5H6 containing 8 amino acid substitutions relative to 21B3 (L52I, S106R, V145M, A184V, L324I, V340M, I366V, and E442K) as well as the removal of one histidine from the C-terminal His tag.

The error-prone PCRs for the two parents were carried out in parallel by using identical conditions to ensure the same mutation rate for both. The reactions were 100 μl and contained 13 ng of template plasmid (corresponding to 3 ng of gene), 0.5 μM of the oligonucleotide primers (*pCWori_for* and *pCWori_rev_clone*, 5′-GCTCATGTTTGACAGCTTATCATCG-3′), 200 μM dATP and dGTP, 500 μM dTTP and dCTP, 7 mM $MgCl_2$, 200 μM $MnCl_2$, 1× Applied Biosystems PCR Buffer, and 5 units of *Taq*. PCR conditions were 95°C for 5 min, followed by 16 cycles of 30 s at 95°C, 30 s at 51°C, and 60 s at 72°C. Gel electrophoresis versus a known standard indicated that this yielded PCR product at a concentration of $\approx$12 ng/μl, for a PCR efficiency of $\lambda = 0.45$. The PCR products were cloned into pCWori with BamHI and EcoRI, electroporated into a catalase-free strain of *Escherichia coli* (48) and plated on LB plates containing 100 μg/ml ampicillin. Transformation of a control ligation with no insert indicated that the background rate of plasmid self-ligation was $<1\%$.

To measure the mutation rates, we randomly selected 20 21B3 clones and 21 5H6 clones for sequencing with primers *pCWori_for* and *pCWori_rev*, allowing us to read each gene from bp 18 to bp 1,436. The 21B3 clones contained a total of 95 nucleotide mutations in the 28,380 sequenced base pairs, with 28 synonymous mutations, 60 nonsynonymous mutations, and 7 mutations leading to premature truncation of the gene (frameshift or nonsense mutations). The 5H6 clones contained a total of 87 mutations in the 29,799 sequenced base pairs, with 23 synonymous mutations, 55 nonsynonymous mutations, and 9 mutations leading to premature truncation of the gene. The distributions of mutations in the two libraries were statistically indistinguishable (Fig. 2*A*). After confirming that the mutation rates in the two libraries were indistinguishable, we combined the sequencing results for further analysis (Table 1). Fig. 2*B* shows that the distribution of mutations is consistent with the theoretical distribution for error-prone PCR (49, 50), which gives the probability that a mutant in a library with an average of $\langle m_{\mathrm{nt}} \rangle$ mutations per gene has m_{nt} nucleotide mutations as

$$\Pr(m_{\mathrm{nt}}) = (1 + \lambda)^{-n} \sum_{k=0}^{n} \binom{n}{k} \lambda^k \frac{(kx)^{m_{\mathrm{nt}}} e^{-kx}}{m_{\mathrm{nt}}!}, \qquad [1]$$

where n is the number of PCR cycles, λ is the PCR efficiency, and $\langle m_{\mathrm{nt}} \rangle (1 + \lambda)/n\lambda$. We also confirmed that the mutations were distributed uniformly along the gene sequence (Fig. 2*C*). If each position in the gene is equally likely to be mutated, then among 41 sequenced clones, 156.3 positions should be mutated once, 9.7 positions should be mutated twice, and 0.4 positions should be mutated three times, in good agreement with the observed values of 148, 13, and 1.

Screening for Improved Mutants. Single mutants were grown in 96-well deep-well plates in Luria broth (LB) supplemented with 100 μg/ml ampicillin for 20–24 h at 30°C, 215 rpm, and 80% relative humidity. One-hundred μl of these LB cultures were used to inoculate deep-well plates containing 400 μl per well of terrific broth (TB) supplemented with 100 μg/ml ampicillin, 0.5 mM δ-aminolevulinic acid, and 0.2 mM isopropyl β-D-thiogalactoside (IPTG) and grown as before. The cells were pelleted and frozen overnight and then lysed by resuspending each well with 600 μl of 100 mM [4-(2-hydroxyethyl)- 1-piperazinepropanesulfonic acid] (EPPS) (pH 8.2) containing 0.5 mg/ml lysozyme and 2 units/ml DNase and incubating for 1 hour at 37°C. The lysis debris was pelleted by centrifugation, and 80 μl of clarified lysate was added to wells of microtiter plates and then combined with 20 μl of 6× substrate stock (1.5 mM 12-pNCA in 36% DMSO or a 6× stock containing 6% DMSO

and 6% acetone with substrate so that concentrations in the stocks were 60 mM 3-phenoxytoluene, 60 mM naproxen, 150 mM phenoxyacetic acid, 30 mM 2-methylbenzofuran, or 30 mM propranolol). Reactions were initiated by adding 20 μl of 24 mM hydrogen peroxide. Endpoints for the 12-pNCA assay were read after 20–30 min at 398 nm (22), whereas endpoints for all other substrate were taken after 1.5–2 h with the 4-amino antipyrene (4-AAP) assay as described in ref. 28. More details are in *Supporting Materials and Methods.*

The 21B3 and 5H6 mutant libraries were screened in parallel by using consistent quantitative criteria for identifying improved mutants. Briefly, all mutants with readings that were at least 50% greater than the larger of the parental or null vector internal standards on the 96-well plate were selected for rescreening. Rescreening was performed by growing and assaying a full row of a 96-well plate for each candidate mutant. Mutants that still appeared at least 50% improved were then expressed in flasks containing 200 ml of TB supplemented with 100 μg/ml ampicillin, 0.5 mM δ-aminolevulinic acid, and 0.4 mM IPTG. The protein concentrations were normalized to 5 μM, and verification assays were performed in microtiter plates. All mutants identified as improved were required to be at least 50% improved over the parental activity or at least 50% improved over the background reading if there was no parental activity. Summary statistics for this process are shown in Table 4, the final readings in the verification assay are shown in Fig. 8, and a comprehensive description of the methods is in the *Supporting Materials and Methods*.

The T_{50} values were measured as described in ref. 26, except that we measured retention of the peak in the carbon monoxide difference spectrum (27) rather than retention of activity (*Supporting Materials and Methods*). Site-directed mutants with the L75R, F275S, and F205L substitutions were constructed by using PCR overlap extension mutagenesis (51) and assayed for function as in the verification assays.

We thank M. Landwehr for bountiful advice and the synthesis of 12-pNCA and C. O. Wilke for helpful advice. J.D.B. is supported by a Howard Hughes Medical Institute predoctoral fellowship. This work was supported by National Institutes of Health Grant R01 GM068664-01. This work was inspired, in part, by a Santa Fe Institute workshop on Evolutionary Innovations supported by a grant from the Packard Foundation.

1. Kauffman, S. A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ. Press, Oxford).
2. Kirschner, M. & Gerhart, J. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8420–8427.
3. Bloom, J. D., Meyer, M. M., Meinhold, P., Otey, C. R., MacMillan, D. & Arnold, F. H. (2005) *Curr. Opin. Struct. Biol.* **15,** 447–452.
4. Kicinger, R., Arciszewski, T. & Jong, K. D. (2005) *Computers Struct.* **83,** 1943–1978.
5. Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C. & Tawfik, D. S. (2005) *Nat. Genet.* **37,** 73–76.
6. Wagner, A. (2005) *Robustness and Evolvability in Living Systems* (Princeton Univ. Press, Princeton).
7. Rennell, D., Bouvier, S. E., Hardy, L. W. & Poteete, A. R. (1991) *J. Mol. Biol.* **222,** 67–87.
8. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S. & Miller, J. H. (1994) *J. Mol. Biol.* **240,** 421–433.
9. Guo, H. H., Choe, J. & Loeb, L. A. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 9205–9210.
10. Bloom, J. D., Silberg, J. J., Wilke, C. O., Drummond, D. A., Adami, C. & Arnold, F. H. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 606–611.
11. Lesk, A. M. & Chothia, C. (1980) *J. Mol. Biol.* **136,** 225–270.
12. Chothia, C. & Lesk, A. M. (1986) *EMBO J.* **5,** 823–826.
13. Nikolova, P. V., Wong, K. B., DeDecker, B., Henckel, J. & Fersht, A. R. (2000) *EMBO J.* **19,** 370–378.
14. Anfinsen, C. B. (1973) *Science* **181**, 223–230.
15. Taverna, D. M. & Goldstein, R. A. (2002) *Proteins* **46,** 105–109.
16. Chan, H. S. & Bornberg-Bauer, E. (2002) *Appl. Bioinformatics* **1,** 121–144.
17. Xia, Y. & Levitt, M. (2004) *Curr. Opin. Struct. Biol.* **14,** 202–207.
18. Miyazawa, S. & Jernigan, R. L. (1985) *Macromolecules* **18,** 534–552.
19. Bloom, J. D., Wilke, C. O., Arnold, F. H. & Adami, C. (2004) *Biophys. J.* **86,** 1–7.
20. Fersht, A. R. (1999) *Structure and Mechanism in Protein Science.* (Freeman, New York).
21. Bornberg-Bauer, E. & Chan, H. S. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 10689–10694.
22. Cirino, P. C. & Arnold, F. H. (2003) *Angew. Chem. Int. Ed.* **42,** 3299–3301.
23. de Montellano, P. R. O. (1995) *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Plenum, New York).
24. Lewis, D. F. V. (2001) *Guide to Cytochromes P450: Structure and Function* (Taylor and Francis, London; New York).
25. Munro, A. W., Leys, D. G., McLean, K. J., Marshall, K. R., Ost, T. W. B., Daff, S, Miles, C. S., Chapman, S. K., Lysek, D. A., Moser, *et al.* (2002) *Trends Biochem. Sci.* **27,** 250–257.
26. Salazar, O., Cirino, P. C. & Arnold, F. H. (2003) *Chem. Biochem.* **4,** 891–893.
27. Otey, C. R. (2003) in *Methods in Molecular Biology*, eds. Arnold, F. H. & Georgiou, G. (Humana, Totowa, NJ), Vol. 230, pp. 137–139.
28. Otey, C. R. & Joern, J. M. (2003) in *Methods in Molecular Biology*, eds. Arnold, F. H. & Georgiou, G. (Humana, Totowa, NJ), Vol. 230, pp. 141–148.
29. Martinis, S. A., Blanke, S. R., Hager, L. P. & Sligar, S. G. (1996) *Biochemistry* **35,** 14530–14536.
30. Haines, D. C., Tomchick, D. R., Machius, M. & Peterson, J. A. (2001) *Biochemistry* **40,** 13456–13465.
31. Somero, G. N. (1995) *Annu. Rev. Physiol.* **57**, 431–468.
32. Fields, P. A. (2001) *Comp. Biochem. Physiol. A* **129**, 417–431.
33. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. (2005) *Nat. Rev. Genet.* **6,** 678–687.
34. Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 12809–12813.
35. van den Burg, B., Vriend, G., Veltman, O. R., Venema, G. & Eusink, V. G. H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2056–2060.
36. Zhao, H. & Arnold, F. H. (1999) *Protein Eng.* **12,** 47–53.
37. Serrano, L., Day, A. G. & Fersht, A. R. (1993) *J. Mol. Biol.* **223,** 305–312.
38. Godoy-Ruiz, R., Perez-Jimenez, R., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (2004) *J. Mol. Biol.* **336,** 313–318.
39. Pakula, A. A. & Sauer, R. T. (1989) *Annu. Rev. Genet.* **23,** 289–310.
40. Matthews, B. W. (1993) *Annu. Rev. Biochem.* **62**, 139–160.
41. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajimi, K. & Sarai, A. (2004) *Nucleic Acids Res.* **32,** D120–D121.
42. Meiering, E. M., Serrano, L. & Fersht, A. R. (1992) *J. Mol. Biol.* **225,** 585–589.
43. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 452–456.
44. Beadle, B. M. & Shoichet, B. K. (2002) *J. Mol. Biol.* **321,** 285–296.
45. Arnold, F. H., Wintrode, P. L., Miyazaki, K. & Gershenson, A. (2001) *Trends Biochem. Sci.* **26,** 100–107.
46. Wang, X., Minasov, G. & Shoichet, B. K. (2002) *J. Mol. Biol.* **320,** 85–95.
47. Earl, D. J. & Deem, M. W. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11531–11536.
48. Barnes, H. J., Arlotto, M. P. & Waterman, M. R. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 5597–5601.
49. Sun, F. (1995) *J. Comput. Biol.* **2**, 63–86.
50. Drummond, D. A., Iverson, B. L., Georgiou, F. & Arnold, F. H. (2005) *J. Mol. Biol.* **350,** 806–816.
51. Higuchi, R., Krummel, B. & Saiki, R. K. (1988) *Nucleic Acids Res.* **16,** 7351–7367.
52. Otey, C. R., Silberg, J. J., Voigt, C. A., Endelman, J. B., Bandara, G. & Arnold, F. H. (2004) *Chem. Biol.* **11,** 309–318.
53. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (2002) *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, U.K.), 2nd Ed., pp. 620–628.

Supporting Materials and Methods

Model Lattice Protein. A model lattice protein was represented as a chain of $N = 20$ monomers of 20 types corresponding to the natural amino acids. A model protein could occupy any of the 41,889,578 conformations (representing 910,972 unique contact sets) corresponding to all length 20 self-avoiding walks on a two-dimensional lattice. A conformation $\mathcal{C}$ had an energy of

$$E(\mathcal{C}) = \sum_{i=1}^{N} \sum_{j=1}^{i-2} C_{ij}(\mathcal{C}) \times \epsilon(\mathcal{A}_i, \mathcal{A}_j),$$

where $C_{ij}(\mathcal{C})$ is 1 if residues i and j are nearest neighbors in conformation $\mathcal{C}$ and 0 otherwise, and $\epsilon(\mathcal{A}_i, \mathcal{A}_j)$ is the interaction energy between residue types $\mathcal{A}_i$ and $\mathcal{A}_j$, given by table 5 of ref. 1. Energies are in units of $k_B T$, and $T = 1.0$ for all simulations.

A model protein folds to a target conformation $\mathcal{C}_t$ with stability

$$\Delta G_f(\mathcal{C}_t) = E(\mathcal{C}_t) + T \ln\{Q(T) - \exp[-E(\mathcal{C}_t)/T]\}$$

where $Q(T)$ is the partition sum

$$Q(T) = \sum_{\{\mathcal{C}_i\}} \exp[-E(\mathcal{C}_i)/T].$$

Proteins were considered stably folded to $\mathcal{C}_t$ if and only if $\Delta G_f(\mathcal{C}_t) \leq 0.0$.

A model protein's activity was represented by binding to a small rigid ligand, much as in ref. 2. If a model protein did not stably fold to a unique conformation, it was considered inactive. If it did stably fold to a unique conformation, then we computed the binding energy of the ligand as the sum of the pairwise interactions between the ligand and protein residues, searching over all possible rotational and translational positions of the ligand. The binding affinity of

the ligand to the folded conformation is given by the association constant K_a, calculated as the exponential of the negative binding energy.

To create the model protein shown in Fig. 1*A*, we first chose the target conformation of the protein, and then performed an adaptive walk from a random starting sequence until we found a sequence that folded to the conformation. We evolved this sequence for 10,000 generations with a population size of 10 and a mutation rate of 5×10^{-4} per residue per generation, with the fitness of a model protein proportional to its binding affinity to the ligand shown in Fig. 1*A*. To create a stable variant of this model protein, we simulated using error-prone PCR to make a library of 1,500 mutant sequences with the mutations distributed according to the error-prone PCR distribution (3,4) with an average of 1.5 mutations per protein, 16 PCR cycles, and a PCR efficiency of 0.45. We performed two rounds of this error-prone PCR, selecting the most stable sequence that still bound to the ligand with the same binding energy as the original model protein at the end of each round. This procedure yielded the stabilized model protein with the sequence IFFMTKIKFHIGVMHMSMGL. We then simulated creating error-prone PCR libraries of both the original and stabilized model proteins using the same procedure as above. We screened each of the four new ligands shown in the legend of Fig. 1*B* on one library from each parent, and recorded the number of mutants that bound to the new ligand with a binding affinity at least 2-fold higher than that for the parent. These counts are the data shown by the bars in Fig. 1*B*.

High-Throughput Screening of Mutants. Single mutant colonies were picked from transformation plates with sterile toothpicks and used to inoculate sterile 1-ml deep-well plates (Falcon), with each well containing 400 μl of liquid LB with 100 μg/ml ampicillin. As controls, wells A1, A2, A3, and A4 were always inoculated with the parent (21B3 or 5H6);

wells A5, A6, A7, and A8 were always inoculated with a negative control (the pCWori plasmid lacking a P450 gene); and well H12 was not inoculated. Each of the remaining 87 wells contained a different mutant. These deep-well plates were grown for 20 to 24 hours in a humidified shaker (Kuhner ISF-1-W) at 215 or 225 rpm, 30^{o}C, and 80% relative humidity. To express the proteins, a pipetting robot (Beckman Multimek 96) was used to transfer 100 μl per well from these LB deep-well plates to 2 ml deep-well plates (Falcon) containing 400 μl of terrific broth (TB) supplemented with 100 μg/ml ampicillin, 0.5 mM δ-aminolevulinic acid, and 0.2 mM isopropyl β-D-thiogalactoside (IPTG). These TB deep-well plates were also grown for 20 to 24 hours in the humidified shaker at 215 or 225 rpm at 30^{o}Cand 80% relative humidity. After this growth, the cells were harvested by centrifugation at 6,100$\times g$ for 10 min at 4^{o}C, and stored in the freezer. The LB deep-well plates were stored at 4^{o}C so that improved mutants could be streaked from the plates.

To perform assays, the pelleted cells were resuspended in 600 μl of lysis buffer (100 mM 4-(2-hydroxyethyl)-1-piperazinepropanesulfonic acid (EPPS), pH 8.2, with 0.5 mg/ml lysozyme and 2 units/ml deoxyribonuclease) per well with the pipetting robot. The plates were incubated for 1 hour at 37^{o}C and then centrifuged at 6,100$\times g$ for 10 min at 4^{o}C to pellet cell debris. The pipetting robot was used to add 80 μl per well of this clarified lysate to 96-well microtiter plates. To assay for folded protein, high-throughput carbon monoxide (CO)-binding difference spectra were measured as described in ref. 5 with the modifications that 80 μl of clarified lysate in the EPPS buffer was used and that the heme was reduced by the addition of 20 μl of 0.1 M sodium hydrosulfite in 1.3 M potassium phosphate, pH 8.0. Blank spectra were read prior to binding CO with a Spectra Max Plus 384 plate reader (Molecular Devices) at every 10 nm from 400 to 500 nm, as well as at the points 447 nm (the absorbance peak for both 21B3 and 5H6)

and 490 nm. After 5-10 minutes of incubation with CO, the absorbance readings were read at these points.

Mutants were screened for the retention of activity on 12-p-nitrophenoxycarboxcylic acid (12-pNCA, Fig. 9) by using a slightly modified version of the method described in ref. 6. A 6$\times$ stock of 12-pNCA was prepared by combining 3.6 ml of 4.17 mM 12-pNCA in DMSO with 6.4 ml of 100 mM EPPS (pH 8.2) immediately before use. Twenty μl of this 6$\times$ 12-pNCA stock was added to the 80 μl of clarified lysate in each well of the microtiter plate, and the plate reader was used to mix the plate and blanked at 398 nm. To initiate the reaction, 20 μl of a 6$\times$ stock of hydrogen peroxide (6$\times$ stock was 24 mM H_2O_2 in 100 mM EPPS, pH 8.2, made immediately before use) was added to each well and mixed with the plate reader. The amount of final product was quantified by reading the absorbance at 398 nm after 20 to 30 minutes.

Mutants were screened for the acquisition of new or improved activity by using the 4-aminoantipyrene (4-AAP) assay, which detects phenol-like compounds (7), produced by either direct hydroxylation of an aromatic ring or as dealkylation products after hydroxylation of a carbon that is ether bonded to an aromatic ring. This assay was used to screen for activity on five substrates (Fig. 9) at the following final concentrations: 3-phenoxytoluene (10 mM), naproxen (10 mM), phenoxyacetic acid (25 mM), 2-methylbenzofuran (5 mM), and propranolol (5 mM). Stocks with 6$\times$ concentrations of these substrates were made immediately before use by dissolving the substrate in equal volumes of DMSO and acetone and then adding 100 mM EPPS (pH 8.2) so that the final concentrations of DMSO and acetone in the 6$\times$ stocks were each 6%. Twenty microliters of 6$\times$ substrate stock was added to the 80 μl of clarified lysate in each well of the microtiter plate, and 20 μl of the 6X hydrogen peroxide stock was added to initiate the reaction. The microtiter plates were mixed with the plate reader, and the

reactions were allowed to run for 1.5 to 2 hours at room temperature before being quenched by the addition of 120 μl of 4 M urea in 0.1 M sodium hydroxide. A pipetting robot was used to add and mix 36 μl of 0.6% 4-AAP, and the plate reader was zeroed at 500 nm. The robot was then used to add and mix 36 μl of 0.6% potassium peroxodisulfate, and, after 30 min, the product was quantified by reading the absorbance at 500 nm.

Stability and Verification Assays. Protein was expressed for stability and verification assays by growing the cells in flasks with 200 to 300 ml of TB supplemented with 100 μg/ml of ampicillin at 30^{o}C and 215 rpm. When the cells reached an optical density between 0.8 and 1.2 at 600 nm, they were induced by adding IPTG to a final concentration of 0.4 mM, as well as δ-aminolevulinic acid to a final concentration of 0.5 mM. The cells were then grown at 30^{o}C and 215 rpm for an additional 20 to 24 hours. A few mutants did not express well in these conditions and so were grown at the milder conditions of 28^{o}C and 180 rpm. The cells were harvested by centrifugation at 6,000 to 8,000$\times g$ and 4^{o}C for 10 min and the pellets stored at -20 ^{o}C. Prior to use, pellets from 100 ml of culture were resuspended in 4 ml of 100 mM EPPS (pH 8.2) and lysed by sonication. Cell debris was pelleted by centrifugation at 6,000 to 8,000$\times g$ and 4^{o}C for 10 to 15 min. The clarified lysates were passed through PD-10 desalting columns (Amersham Biosciences) to remove small molecules that might appear as background in the 4-AAP assay. The P450 protein was quantified from the CO binding difference spectrum (extinction coefficient of 91 mM^{-1}. cm^{-1} for the A_{447}-A_{490}) and the concentration of P450 was adjusted to 5 μM by the addition of more 100 mM EPPS (pH 8.2).

For verification assays, the 5 μM clarified lysate was pipetted into 96-well microtiter plates (typically a full row of 12 wells for each mutant, although some samples had fewer wells because of air bubbles or a limited protein supply), and 12-pNCA and 4-AAP assays were

performed as described above.

For stability assays, 150 μl of the 5 μM clarified lysate was added to rows of 96-well PCR plates (Bio-Rad). The PCR plates were heated to different temperatures by using the gradient method of a PCR cycler (MJ Research, PTC-200) for 10 min then cooled to 4^{o}C. The PCR plates were centrifuged at 5,000 to 6,000$\times g$ for 5 to 10 minutes at 4^{o}C to pellet denatured debris, and 80 μl of the supernatant was used for CO-binding difference spectrum assays. The temperature at which half of the protein was denatured (T_{50}) was determined by fitting a sigmoidal curve to the percentage of remaining CO binding difference spectrum, as shown in Fig. 6.

Retention of Folding and Activity on 12-pNCA. Our high-throughput screening allowed us to determine whether mutants retained the CO-binding difference spectrum peak characteristic of folded P450 heme domains and whether they retained the high activity on 12-pNCA of both the 21B3 and 5H6 parents. We collected CO-binding difference spectra and 12-pNCA activity data for six plates each of 21B3 and 5H6 (522 mutants of each). The CO-binding difference spectra and the 12-pNCA readings for these plates are shown in Fig. 7. Because there was often some variation between plates, we classified the mutants as folded/unfolded and active/inactive relative to the parental controls on the same plate. These binary classifications require defining a threshold for the fraction of the parental reading the mutant must retain. Table 3 shows the fractions for different thresholds.

Identification of Mutants with New Activity We sought to measure the frequencies with which 21B3 and 5H6 mutants acquired activity on 3-phenoxytoluene, naproxen, propranolol, 2-methylbenzofuran, and phenoxyacetic acid. In order to compare the frequencies for the 21B3 and 5H6 libraries, we developed consistent quantitative methods for identifying

improved mutants to ensure that both libraries were treated identically. We first performed high-throughput screening of the mutants for activity using the 4-AAP assay. Overall we screened 8,160 mutants each of 21B3 and 5H6 (Table 4). Because of the variation between plates, each mutant was evaluated relative to the four parental and four null vector controls on its own plate. Because of problems with the pipetting robot, two wells in the microtiter plates (A12 and D10) consistently contained air bubbles and so were disregarded, meaning that 85 mutants were screened for each plate. We identified candidates for improved mutants and then rescreened these candidates to eliminate false positives and performed carefully controlled assays to verify improved mutants. The procedure for this process was as follows:

1. The A_{500} reading for each mutant was compared to the median parental reading and the median null control reading. If the A_{500} reading was both 1.5 times greater and 0.01 greater than the two medians, respectively, then the mutant was considered a candidate to be improved. In a few cases, candidates identified by these criteria were disregarded because they were obviously spurious (noticeable air bubble in the well, no folded protein as indicated by the CO binding difference spectrum for the well, or parental medians were abnormally low due to poor lysis in some wells). Candidates for improved mutants were streaked for single colones on LB plates supplemented with 100 μg/ml ampicillin from the 1-ml deep-well plates that had been stored at 4^{o}C.

2. All candidate improved mutants were rescreened by growing new plates and by using the same high-throughput screening procedure, but now growing an entire row (12 samples) in the 96-well plate for each mutant. The median reading for the candidate mutant was compared with the median readings for rows of parental and null vector samples. If

wells were obviously spurious (air bubbles or poor lysis as indicated by CO-binding difference spectrum), they were excluded from the calculations of the medians. If the mutant reading was 0.01 greater than both the parental and null vector samples, and if the difference between the mutant and null control reading was ¿1.5 times the difference between the mutant and parental reading, then the mutant was considered to have passed the rescreen and was analyzed on a verification plate.

3. Verification plates were used to gather high-quality data, since all protein samples in these plates were adjusted to the same concentration (5 μM). If the parental reading was significantly higher than the null control reading (indicating some parental activity), the median mutant reading minus the null vector reading was required to be 1.5 times greater than the median parent reading minus the null vector reading. If the parental reading was roughly equal to the null control reading, the median reading was required to be 1.5 times greater than both the parental and null control reading.

Fig. 8 shows the activities for the parents and all improved mutants as measured on the verification plates. Table 4 summarizes the statistics for the different steps of this process. In the end, we identified 13 improved 5H6 mutants and 4 improved 21B3 mutants. These mutants were sequenced by using the primers *pCWori_for*, *pCWori_rev*, and *pCWori_rev_clone* to identify the mutations. These mutations are summarized in Table 5.

Creation and Analysis of Site-Directed Mutants. We examined the effects of parental stability on some of the amino acid substitutions observed in our improved mutants by using site-directed mutagenesis to make these mutations in both the 21B3 and 5H6 parents. We constructed the mutants by PCR overlap extension mutagenesis (8) using the following primers

(the induced amino acid substitutions are indicated in the primer names):

mutL75R_for (5-CTTAAGTCAAGCGCGTAAATTTGTACGTG-3),

mutL75R_rev (5-CACGTACAAATTTACGCGCTTGACTTAAG-3),

mutF205L_for (5-CAAGCGCCAGCTTCAAGAAGATATCAAGG-3),

mutF205L_rev (5-CCTTGATATCTTCTTGAAGCTGGCGCTTG-3),

mutF275S_for (5-GCGGTCTTTTAACATCTGCGCTGTATTTCTTAG-3),

mutF275S_rev (5-CTAAGAAATACAGCGCAGATGTTAAAAGACCGC-3), *pCWori_for*, and *pCWori_rev*. All mutants were sequenced to ensure that the genes contained only the desired mutations. Protein was expressed in flasks as for the improved mutants, and clarified lysate was used for functional and thermostability assays as before.

The L75R substitution conferred naproxen activity on the 5H6 parent (mutant 5H6-13C9). This substitution is in the substrate-binding pocket and destabilized the mutant by 8^{o}C. When we made this substitution in the 21B3 parent, we were able to recover only inactive protein with a CO-binding difference spectrum peak at 420 nm (Fig. 10). A CO-binding difference spectrum peak at 420 nm is a characteristic of an improperly folded P450 protein (9,10), indicating that the L75R substitution is too destabilizing to be tolerated by the 21B3 parent.

The F275S substitution was one of two substitutions found in a 5H6 mutant improved on 3-phenoxytoluene (mutant 5H6-29G8). We made the F275S single mutants of both 5H6 and 21B3. The 5H6-F275S mutant is destabilized relative to the parent but is active on 3-phenoxytoluene (Fig. 11). We were unable to recover any folded protein for the 21B3-F275S mutant, suggesting that the substitution is too destabilizing to be tolerated by 21B3.

The F205L substitution was found in conjunction with other substitutions in 21B3 mutant improved on propranolol (21B3-14C10). We made this substitution in both 5H6 and 21B3.

Both mutants folded, and neither was substantially destabilized. Both mutants had slightly improved activity on propranolol, indicating that this beneficial substitution is tolerated by both parents (Fig. 12).

P450 stability: relationship of T_{50} values to other measures of stability. In this work, we have argued that more stable variants of a protein should be more tolerant to mutations and, therefore, more evolvable. Our biophysical arguments and our lattice protein simulations are both based on consideration of the equilibrium thermodynamic stability of the native structure ΔG_f. It is mutational changes to this true thermodynamic stability ($\Delta\Delta G$ values) that are expected to be largely additive (11), and so ΔG_f is the proper basis for the formulation of our arguments. However, our experiments with the P450 variants do not measure ΔG_fbut, instead, measure T_{50}, defined as the temperature at which half of the protein irreversibly denatures after 10 min incubation. If we were able to measure true melting temperatures (T_m) for reversible thermal denaturation, then we could compellingly argue that these T_m values closely correspond with ΔG_f, because both experimental and thermodynamic considerations (12) suggest that changes in T_m are linearly correlated with changes in ΔG_f for proteins with the same structure and high sequence identity (as is the case for all of our mutants). However, our T_{50} values measure the extent of irreversible thermal denaturation, which can be affected by factors other than ΔG_f (such as the kinetics of unfolding, aggregation, or chemical modification of side chains). In this section, we discuss why we used T_{50} values, and show that these T_{50} values are highly correlated with the stability to irreversible chemical denaturation.

In order to measure ΔG_f, it is necessary to find a reversible method for denaturing the protein. Thermal denaturation was irreversible for all of our P450 variants (at least in the buffer

conditions and at the level of protein purity of our assays), as can be seen from Fig. 6, which shows that none of the mutants refolded when exposed to sufficiently high temperatures. If chemical denaturation is reversible, ΔG_f scales linearly with the concentration of chemical denaturant, and the slope (m value) is the same for proteins of the same structure and high sequence identity (13). Therefore, for reversible chemical denaturation, ΔG_f is linearly related to the concentration of denaturant at which half of the protein unfolds. Other researchers have performed chemical denaturation studies on various cytochromes P450. They have generally found that denaturation by guanidinium chloride is irreversible even at low concentrations (14,15), but that urea denaturation is sometimes reversible at low urea concentrations before becoming irreversible at higher concentrations (14,16). In general, P450 denaturation appears to proceed through multiple intermediates and pathways (14-16), perhaps explaining why the reversibility of denaturation is so sensitive to the particular conditions and P450 variants used.

We tested the stability of 12 of our P450 BM3 heme domain variants to urea denaturation. We began by obtaining samples of the 12 variants as for the thermal denaturation studies. We then adjusted these samples to $\approx$5 μM in the buffer of 100 mM EPPS (pH 8.2). Half of each sample was diluted to 1.8 μM with buffer, whereas the other half was diluted in the same fashion with 8 M urea in 100 mM EPPS (pH 8.2), resulting in a final urea concentration of 5.1 M. Both samples were incubated overnight at 13^{o}C, and then CO difference spectra were measured. Fig. 5*A* shows that all 12 samples were completely or nearly completely denatured by the urea treatment, as indicated by the disappearance of the Soret peak at 450 nm. To see whether the denaturation was reversible, we removed the urea from six of the variants by dialysis against a 1000-fold excess of buffer. None of the variants exhibited signficant refolding (Fig. 5*A*), indicating that urea denaturation was irreversible for our P450 variants and assay

conditions.

To measure the stability of our variants to irreversible urea denaturation, we added 75 μl samples of the 5 μM P450 samples to rows of 96-well microtiter plates. We then used the pipetting robot to add and mix 200 μl of solutions of various concentrations of urea in our buffer so that the final urea concentrations were those shown on the X axes of the plots in Fig. 5*B*. The microtiter plates were incubated overnight at 13 ^{o}C. We then added 25 μl of 0.1 M sodium hydrosulfite in 1.3 M potassium phosphate (pH 8.0) and read the CO difference spectra. Figure 5B shows that the Soret peak diminished with increasing urea concentrations. We fit sigmoidal curves to these denaturation plots and quantified the stability to irreversible urea denaturation as the concentration of urea at which half of the protein had unfolded ([urea]$_{50}$). We then compared the [urea]$_{50}$ values for these 12 variants to the T_{50} values measured in Fig. 6. Fig. 5*C* shows that the [urea]$_{50}$ and T_{50} values are linearly correlated. Therefore, although we are unable to measure ΔG_f, our measures of stability to irreversible thermal and urea denaturation are nearly the same, supporting the view that the T_{50} values reflect universal aspects of protein stability rather than unique characteristics of the process of irreversible thermal denaturation.

1. Miyazawa, S. & Jernigan, R. L. (1985) *Macromolecules* **18**, 534–552.

2. Bloom, J. D., Wilke, C. O., Arnold, F. H. & Adami, C. (2004) *Biophys. J.* **86**, 1–7.

3. Sun, F. (1995) *J. Comput. Biol.* **2**, 63–86.

4. Drummond, D. A., Iverson, B. L., Georgiou, F. & Arnold, F. H. (2005) *J. Mol. Biol.* **350**, 806–816.

5. Otey, C. R. (2003) in *Methods in Molecular Biology*, eds. Arnold, F. H. & Georgiou, G. (Humana, Totawa, NJ) Vol. 230, pp. 137–139.

6. Cirino, P. C. & Arnold, F. H. (2003) *Angew. Chem. Int. Ed.* **42**, 3299–3301.

7. Otey, C. R & Joern, J. M. (2003) in *Methods in Molecular Biology*, eds. Arnold, F. H. & Georgiou, G. (Humana, Totawa, NJ) Vol. 230, pp. 141–148.

8. Higuchi, R., Krummel, B. & Saiki, R. K. (1988) *Nucleic Acids Res.* **16**, 7351–7367.

9. Martinis, S. A., Blanke, S. R., Hager, L. P. & Sligar, S. G. (1996) *Biochemistry* **35**, 14530–14536.

10. Wells, A. V., Li, P. & Champion, P. M. (1992) *Biochemistry* **31**, 4384–4393.

11. Wells, J. A. (1990) *Biochemistry* **29**, 8509–8517.

12. Rees, D. C. & Robertson, A. D. (2001) *Protein Sci.* **10**, 1187–1194.

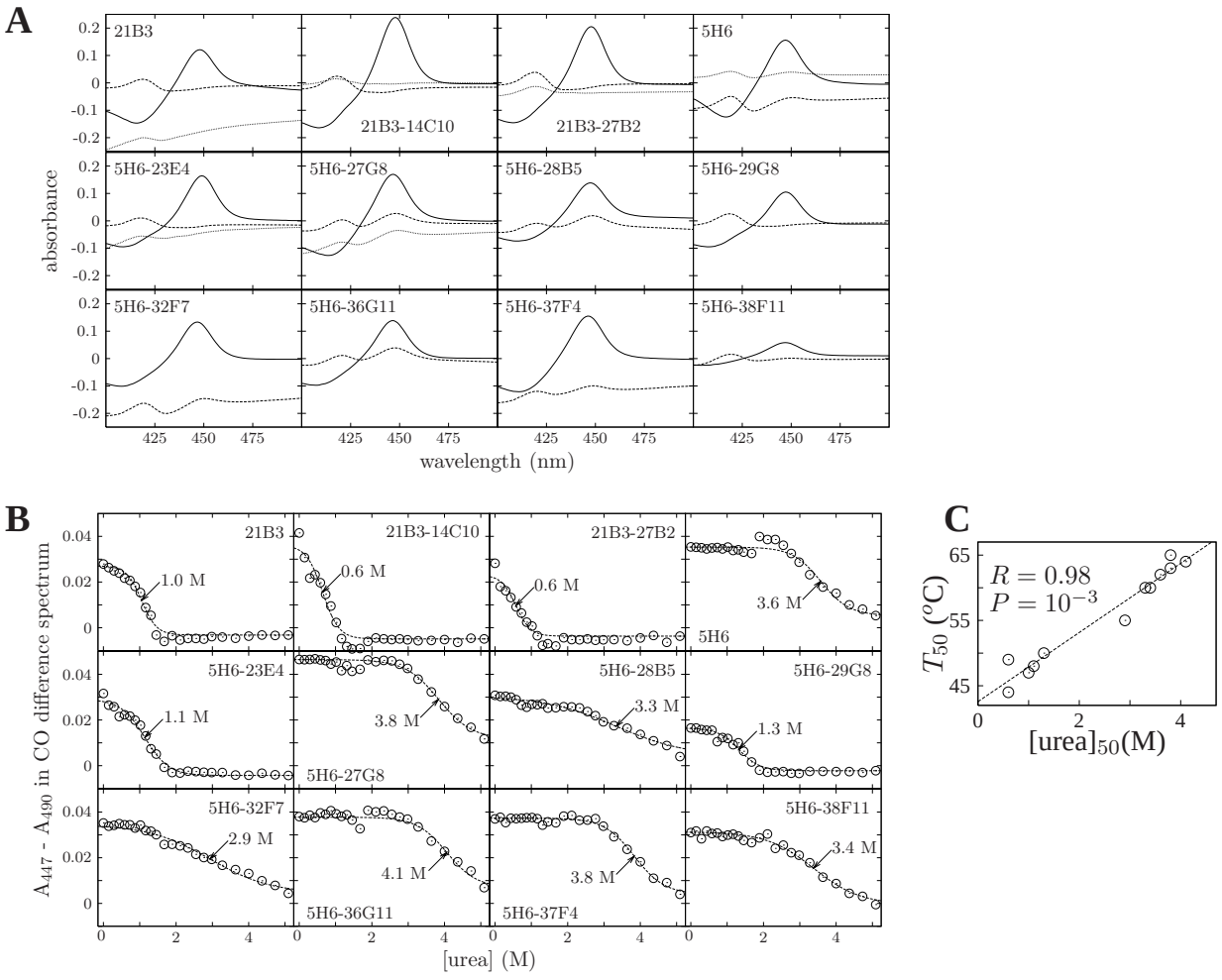
13. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995) *Protein Sci.* **4**, 2138–2148.

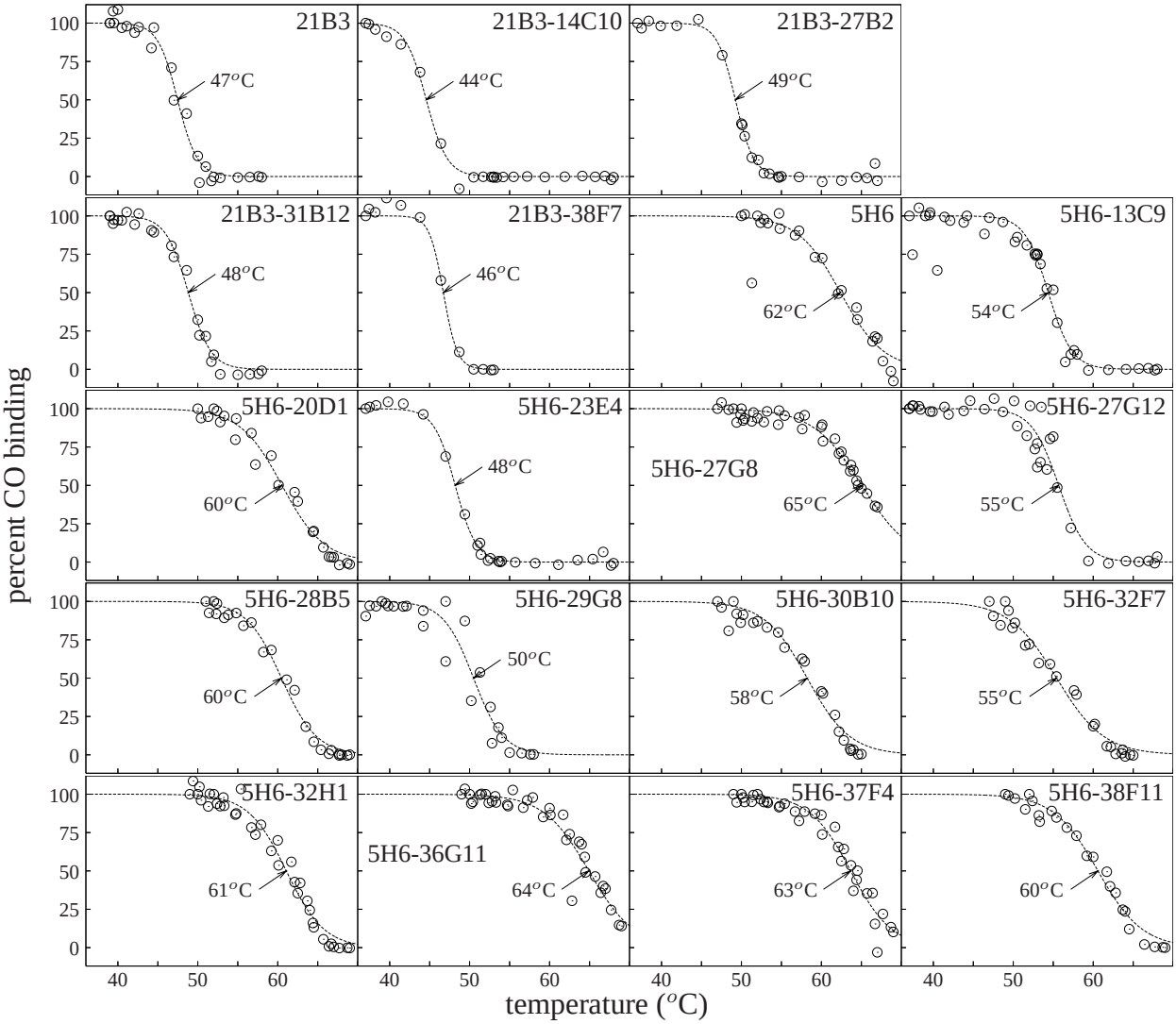
14. Yu, X. C., Shen, S. & Strobel, H. W. (1995) *Biochemistry* **34**, 5511–5517.

15. Munro, A. W., Lindsay, J. G., Coggins, J. R., Kelly, S. M. & Price, N. C. (1996) *Biochem. Biophys. Acta* **1296**, 127–137.

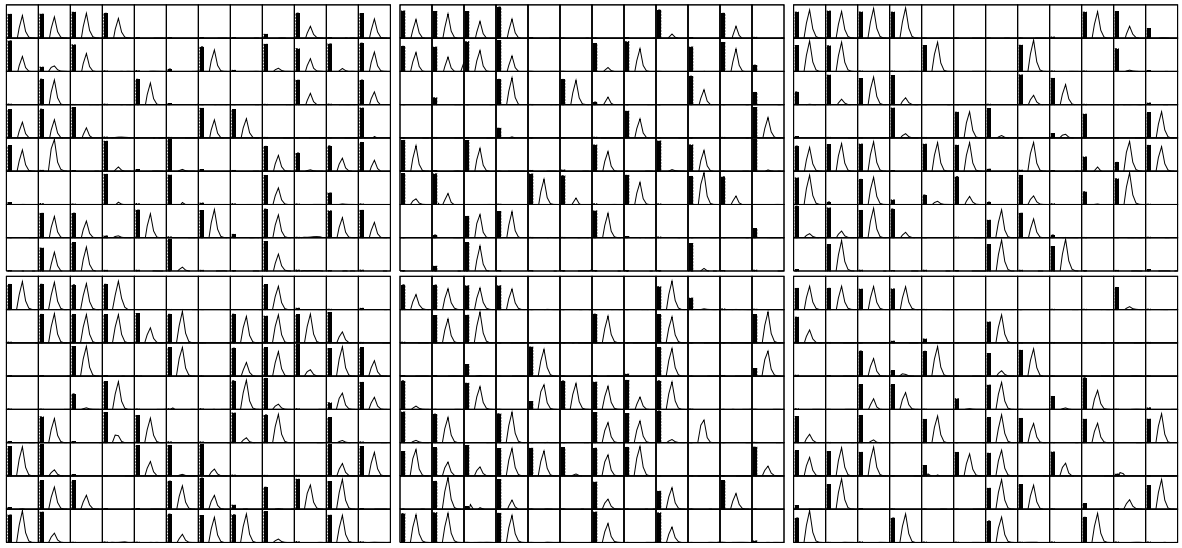
16. Murugan, R. & Mazumdar, S. (2004) *J. Biol. Inorg. Chem.* **9**, 477–488.

17. Haines, D. C., Tomchick, D. R., Machius, M. & Peterson, J. A. (2001) *Biochemistry* **40**, 13456–13465.

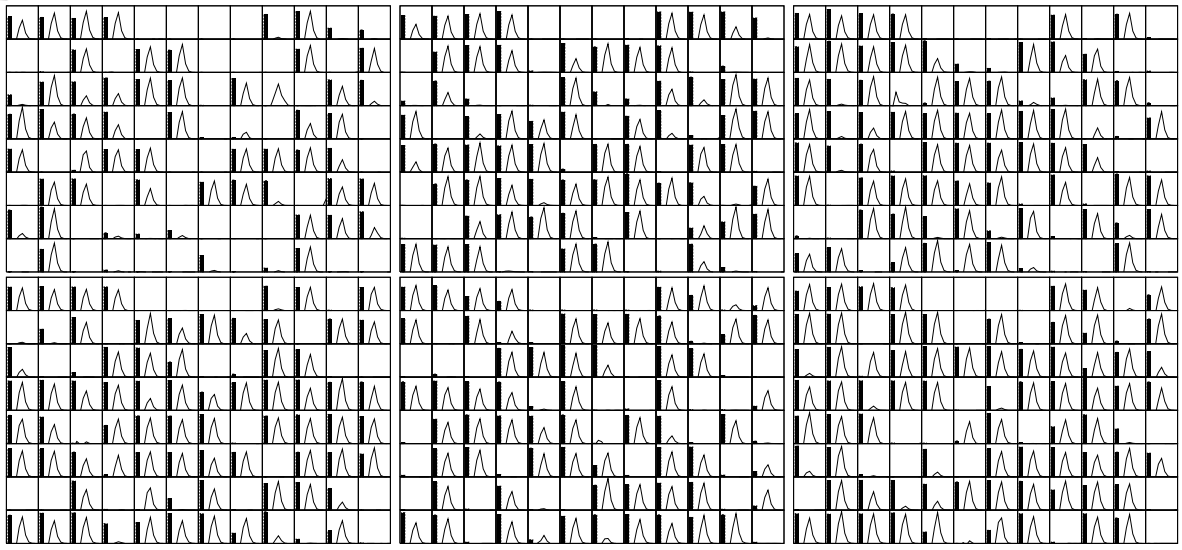

A
21B3
21B3-14C10
21B3-27B2
5H6
5H6-23E4
5H6-27G8
5H6-28B5
5H6-29G8
5H6-32F7
5H6-36G11
5H6-37F4
5H6-38F11
absorbance
wavelength (nm)
B
1.0 M
0.6 M
0.6 M
3.6 M
1.1 M
3.8 M
3.3 M
1.3 M
2.9 M
4.1 M
3.8 M
3.4 M
A447 - A490 in CO difference spectrum
[urea] (M)
C
R = 0.98
P = 10−3
T50 (oC)
[urea]50(M)

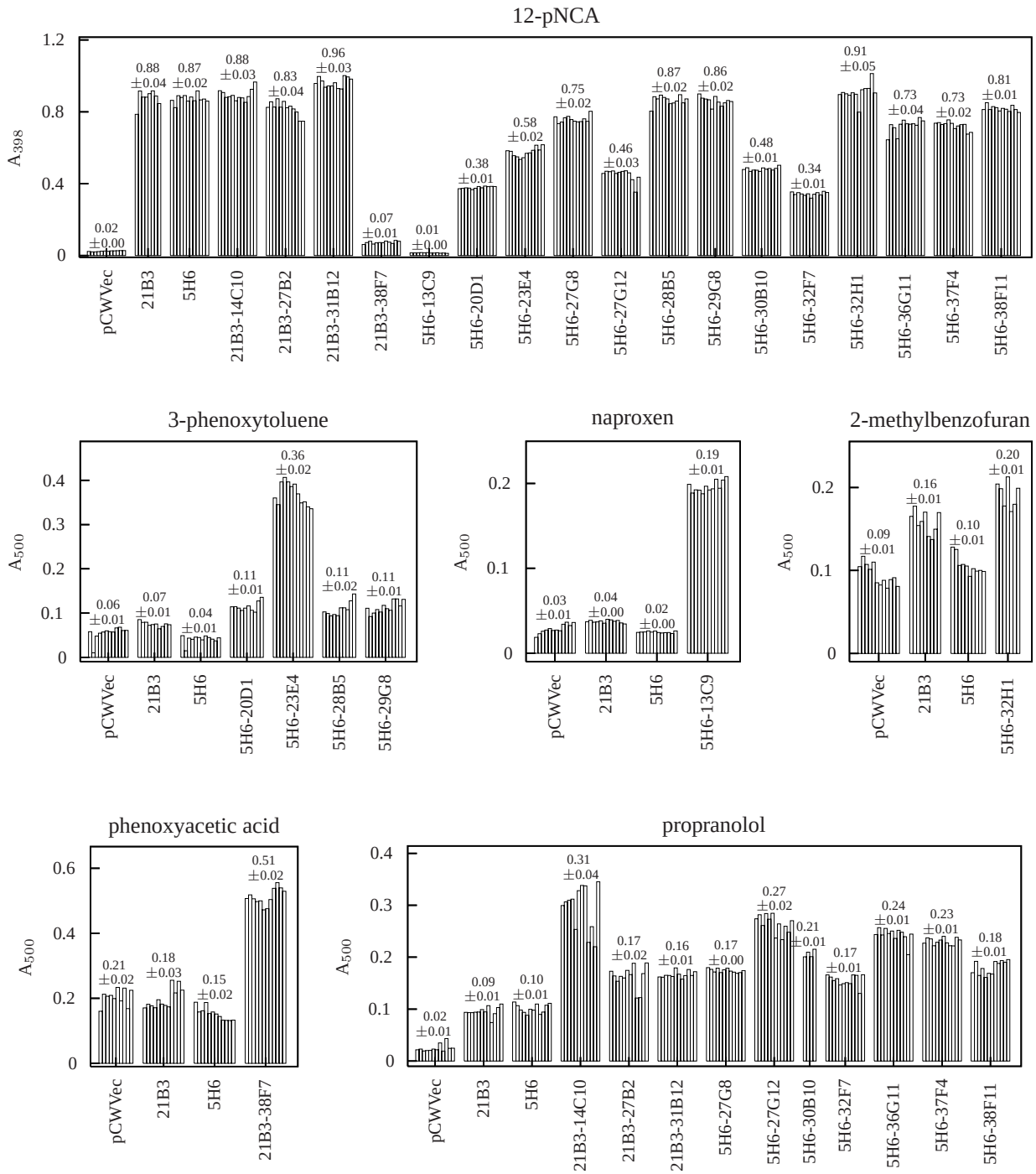

21B3
47°C
21B3-14C10
44°C
21B3-27B2
49°C
21B3-31B12
48°C
21B3-38F7
46°C
5H6
62°C
5H6-13C9
54°C
5H6-20D1
60°C
5H6-23E4
48°C
5H6-27G8
65°C
5H6-27G12
55°C
5H6-28B5
60°C
5H6-29G8
50°C
5H6-30B10
58°C
5H6-32F7
55°C
5H6-32H1
61°C
5H6-36G11
64°C
5H6-37F4
63°C
5H6-38F11
60°C
percent CO binding
temperature (°C)

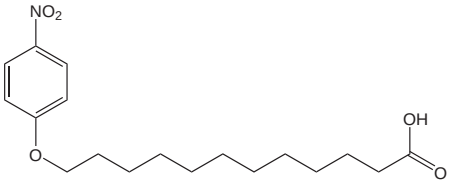
A



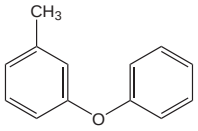
B



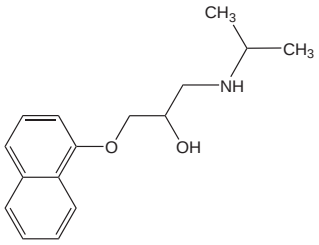

12-pNCA
A398
1.2
0.8
0.4
0
0.02 ±0.00
0.88 ±0.04
0.87 ±0.02
0.88 ±0.03
0.83 ±0.04
0.96 ±0.03
0.07 ±0.01
0.01 ±0.00
0.38 ±0.01
0.58 ±0.02
0.75 ±0.02
0.46 ±0.03
0.87 ±0.02
0.86 ±0.02
0.48 ±0.01
0.34 ±0.01
0.91 ±0.05
0.73 ±0.04
0.73 ±0.02
0.81 ±0.01
pCWVec
21B3
5H6
21B3-14C10
21B3-27B2
21B3-31B12
21B3-38F7
5H6-13C9
5H6-20D1
5H6-23E4
5H6-27G8
5H6-27G12
5H6-28B5
5H6-29G8
5H6-30B10
5H6-32F7
5H6-32H1
5H6-36G11
5H6-37F4
5H6-38F11
3-phenoxytoluene
A500
0.4
0.3
0.2
0.1
0
0.06 ±0.01
0.07 ±0.01
0.04 ±0.01
0.11 ±0.01
0.36 ±0.02
0.11 ±0.02
0.11 ±0.01
pCWVec
21B3
5H6
5H6-20D1
5H6-23E4
5H6-28B5
5H6-29G8
naproxen
A500
0.2
0.1
0
0.03 ±0.01
0.04 ±0.00
0.02 ±0.00
0.19 ±0.01
pCWVec
21B3
5H6
5H6-13C9
2-methylbenzofuran
A500
0.2
0.1
0
0.09 ±0.01
0.16 ±0.01
0.10 ±0.01
0.20 ±0.01
pCWVec
21B3
5H6
5H6-32H1
phenoxyacetic acid
A500
0.6
0.4
0.2
0
0.21 ±0.02
0.18 ±0.03
0.15 ±0.02
0.51 ±0.02
pCWVec
21B3
5H6
21B3-38F7
propranolol
A500
0.4
0.3
0.2
0.1
0
0.02 ±0.01
0.09 ±0.01
0.10 ±0.01
0.31 ±0.04
0.17 ±0.02
0.16 ±0.01
0.17 ±0.00
0.27 ±0.02
0.21 ±0.01
0.17 ±0.01
0.24 ±0.01
0.23 ±0.01
0.18 ±0.01
pCWVec
21B3
5H6
21B3-14C10
21B3-27B2
21B3-31B12
5H6-27G8
5H6-27G12
5H6-30B10
5H6-32F7
5H6-36G11
5H6-37F4
5H6-38F11



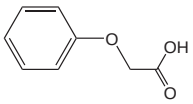
12-pNCA



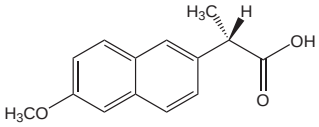
3-phenoxytoluene



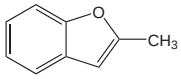
propranolol



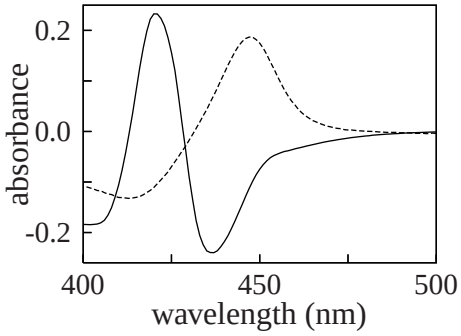
phenoxyacetic acid

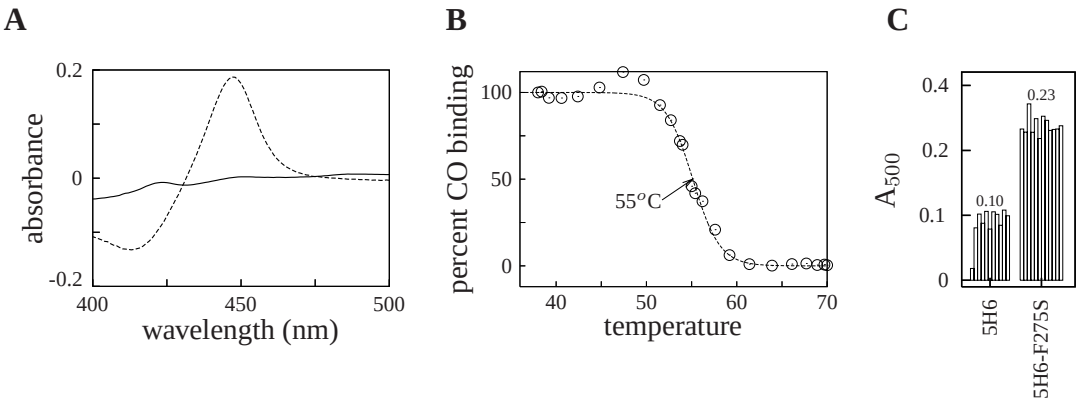


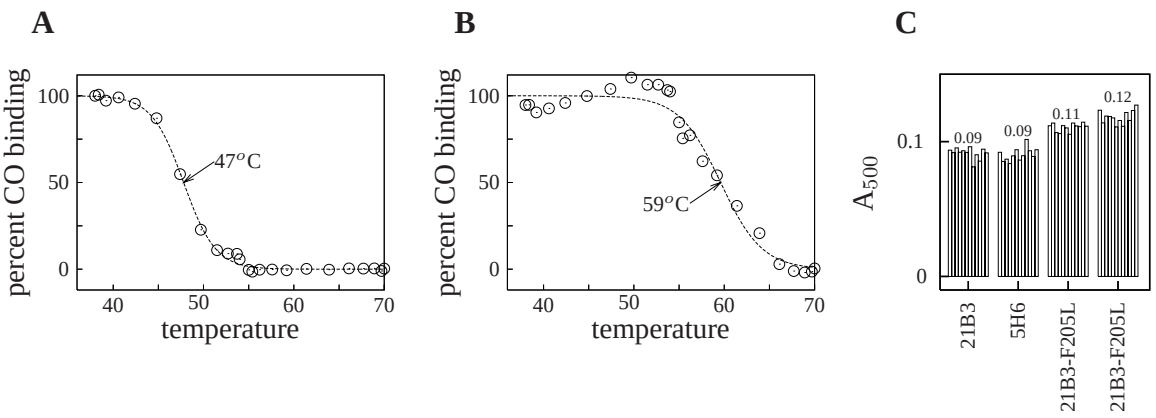
naproxen



2-methylbenzofuran


0.2
0.0
-0.2
absorbance
400
450
500
wavelength (nm)


A
0.2
0
-0.2
absorbance
400
450
500
wavelength (nm)
B
100
50
0
percent CO binding
55°C
40
50
60
70
temperature
C
0.4
0.2
0.1
0
A500
0.10
0.23
5H6
5H6-F275S


A
B
C
percent CO binding
100
50
0
40
50
60
70
temperature
47°C
59°C
A500
0.1
0
0.09
0.09
0.11
0.12
21B3
5H6
21B3-F205L
21B3-F205L