doi:10.1016/j.jmb.2010.04.039 *J. Mol. Biol.* (2010) **399**, 541–546

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

COMMUNICATION

Consensus Protein Design without Phylogenetic Bias

Christian Jäckel[1], Jesse D. Bloom[2], Peter Kast[1], Frances H. Arnold[3] and Donald Hilvert[1]*

[1]*Laboratory of Organic Chemistry, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland*

[2]*Division of Biology, California Institute of Technology, MC 147-75, Pasadena, CA 91125, USA*

[3]*Division of Chemistry and Chemical Engineering, California Institute of Technology, MC 210-41, Pasadena, CA 91125, USA*

Received 31 January 2010; received in revised form 20 April 2010; accepted 22 April 2010
Available online
28 April 2010

Edited by F. Schmid

Consensus design is an appealing strategy for the stabilization of proteins. It exploits amino acid conservation in sets of homologous proteins to identify likely beneficial mutations. Nevertheless, its success depends on the phylogenetic diversity of the sequence set available. Here, we show that randomization of a single protein represents a reliable alternative source of sequence diversity that is essentially free of phylogenetic bias. A small number of functional protein sequences selected from binary-patterned libraries suffice as input for the consensus design of active enzymes that are easier to produce and substantially more stable than individual members of the starting data set. Although catalytic activity correlates less consistently with sequence conservation in these extensively randomized proteins, less extreme mutagenesis strategies might be adopted in practice to augment stability while maintaining function.

© 2010 Elsevier Ltd. All rights reserved.

Keywords: protein stabilization; multiple sequence alignments; consensus mutation; binary patterning; chorismate mutase

Utilization of proteins outside of their normal biological context (e.g., in diagnostic, medical, or industrial applications, or for the creation of novel receptors and catalysts)[1,2] often requires optimization of biophysical properties such as stability.[3,4] For this purpose, engineering methods that exploit statistical amino acid frequencies from multiple sequence alignments (MSAs) are widely used.[5,6] Data-driven consensus design is based on the simple assumption that the frequency of a given residue in an MSA of homologous proteins correlates with that amino acid's contribution to protein stability.[7] An artificial protein possessing the most frequent residue at each position should accordingly show maximum stability. Given the difficulty of predicting how individual residues contribute to overall stability,[8] this approach to protein stabilization is often preferable to classical rational design, particularly as it does not depend on the availability of structural information.

If an MSA were composed of fully independent protein sequences all selected for stable folding to the same structure, and if individual residues contributed additively to stability, then the stability contribution of a particular amino acid at a given position should be a roughly logarithmic function of its frequency in the MSA.[9] Indeed, approaches based on this idea have proven broadly successful at creating more stable proteins.[5,6,10] However, because the sequences of natural proteins generally derive from a common ancestor, they tend to be heavily biased by evolutionary relationships. This lack of statistical independence among different protein sequences violates one of the key assumptions underlying the idea of a logarithmic relationship between an amino acid's stability contribution and its frequency in an MSA. As a consequence, the purely statistical approach of simply replacing all

*Corresponding author. E-mail address: hilvert@org.chem.ethz.ch.

Abbreviation used: MSA, multiple sequence alignment.

0022-2836/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.

nonconsensus residues in conserved positions of a sequence or sequence motif with their consensus counterparts may often fail to yield a more stable protein,[6] unless large numbers of functionally and structurally similar proteins are available to minimize phylogenetic bias[11,12] or structural factors are considered in the design process.[13,14] Various mathematical algorithms, ranging from simply reducing the weighting of highly similar sequences[10] to complex likelihood-based methods that fully account for phylogeny,[15] have been employed to correct for such bias. Labor-intensive mutational analysis of conserved residues is another common strategy for determining individual contributions to stability,[16–19] although it does not necessarily guarantee success.[20]

Combinatorial methods have also been employed to identify stabilizing consensus mutations in natural homologs. For example, shuffling experiments with conserved residues in β-lactamases[21] or larger fragments of cytochrome P450 enzymes[22] have shown that *in-vitro*-generated sequence diversity can be a viable alternative to MSAs of natural proteins as input for consensus design. Extrapolating from these results, libraries created from a single protein via sequence randomization and selection should be an attractive source of fully unbiased sequence diversity. Since the resulting variants would be phylogenetically unrelated a priori, the frequency of a given residue in the randomized stretch should directly reflect its contribution to stability according to a standard Boltzmann-like relationship. To test this hypothesis, we have generated consensus enzymes based on functional clones isolated from libraries of individual extensively randomized proteins.

An all-helical homodimeric chorismate mutase from *Escherichia coli*, EcCM,[23] served as our starting point (Fig. 1). In a previous study, its N-terminal H1 helix, which is 42 amino acids long and forms a dimer-spanning coiled coil, was replaced with a module of randomized sequence.[24] The library maintained the polar/apolar binary pattern[25] of the original H1 helix, but replaced hydrophobic residues with mixtures of leucine (Leu), isoleucine (Ile), methionine (Met), and phenylalanine (Phe), as well as hydrophilic residues with mixtures of lysine (Lys), glutamate (Glu), aspartate (Asp), and asparagine (Asn). Only three highly conserved active-site residues in the helix and two amino acids needed for library construction purposes were held constant. Functional clones were isolated from the library by genetic complementation of chorismate-mutase-deficient bacteria. Here, we subjected the sequences of 26 catalytically active variants (Supplementary Table 1 and Supplementary Fig. 1a) to statistical analysis, calculating the conservation energy E for an amino acid at position i from its frequency f in the sequence alignment according to Eq. (1):

$$E_{\mathrm{aa},i} = -\ln f_{\mathrm{aa},i} \tag{1}$$

The consensus protein H1-E-cons (Fig. 2a) was then designed by choosing the amino acid that appeared most frequently in the selected proteins at each position of the H1 helix. In cases where no consensus could be determined due to equal occurrence of two or more residues, the residue encoded by the least frequent codon in the design set was chosen. Summing the individual contributions of all the consensus residues over the length of the helix affords the conservation energy for the entire binary-patterned module E_c (Eq. (2)), which serves as a measure of sequence conservation:

$$E_\mathrm{c} = \sum\nolimits_{1}^{i} E_{\mathrm{aa},i} \tag{2}$$

The gene for the consensus design was expressed in *E. coli*, and the resulting protein was biochemically characterized. For comparison, representative library members possessing low (H1-E-low), medium (H1-E-med), and high (H1-E-high) conservation energies were also produced (Fig. 2a).

While none of the characterized proteins from the original library of patterned EcCM variants was sufficiently stable for detailed biophysical studies, H1-E-cons was readily produced as a helical dimer, as judged by CD spectroscopy (Fig. 2b) and size-exclusion chromatography. Moreover, even though only eight different amino acids were used for randomization of the H1 helix, chemical denaturation experiments with guanidinium chloride

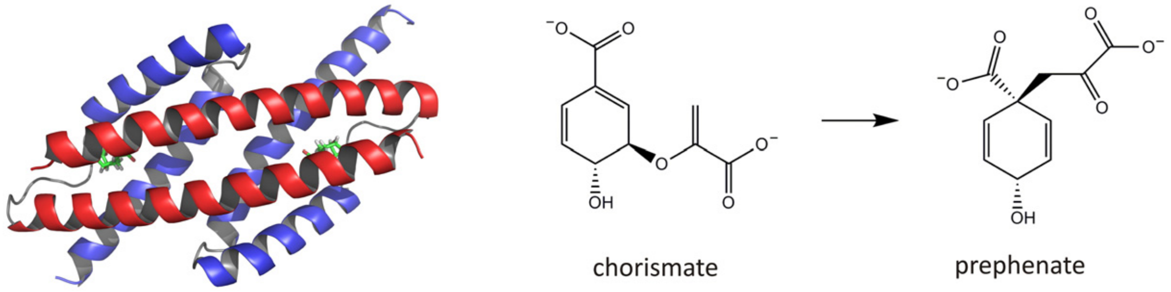
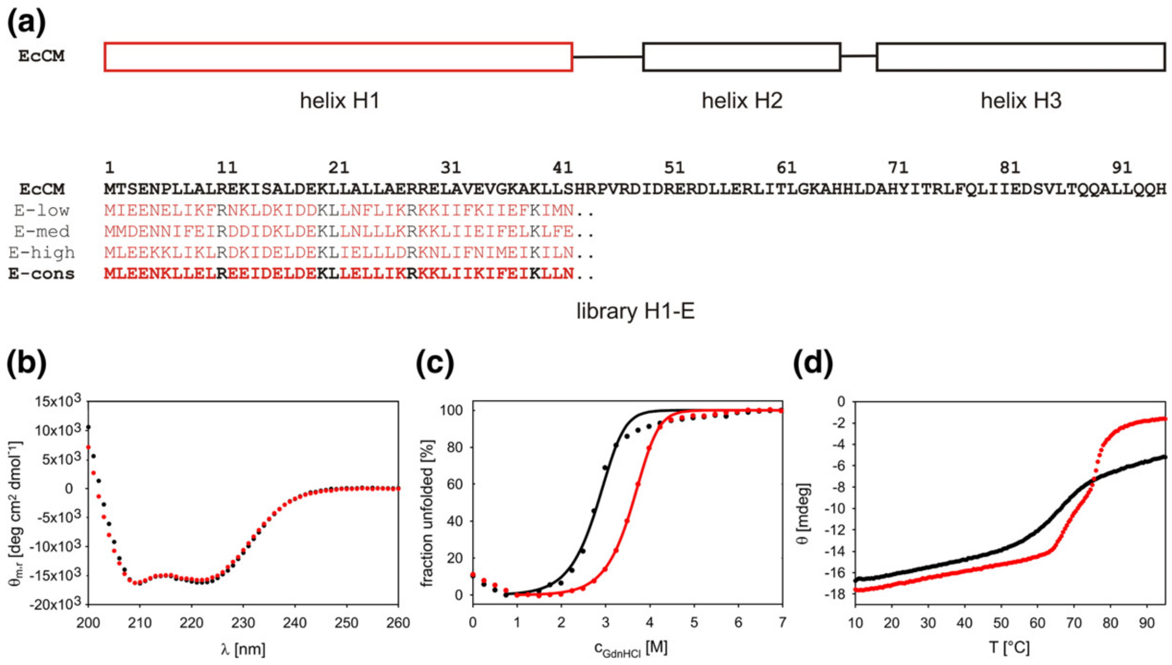


Fig. 1. Structural model of homodimeric helical bundle chorismate mutases (left) and the reaction they catalyze (right). The H1 and H2/3 helices of the protein are shown in red and blue, respectively. A transition-state analog inhibitor bound at the active sites is highlighted in green. The protein graphic was made with PyMOL (DeLano Scientific LLC) based on the crystal structure of the *E. coli* chorismate mutase EcCM (Protein Data Bank ID 1ECM).



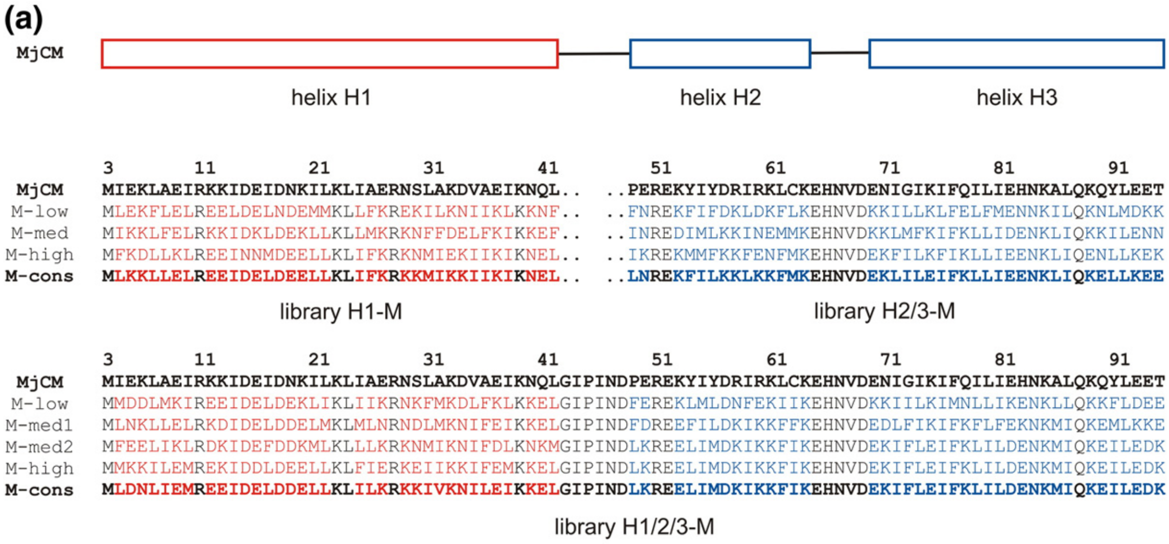
(e)

Protein	E_c	ΔG_u	m	k_{cat}	K_m	k_{cat} / K_m
		[kcal mol^{-1}]	[kcal mol^{-1} M^{-1}]	[s^{-1}]	[µM]	[M^{-1} s^{-1}]
EcCM	--	**15.3**	**-2.8**	**14.8**	**391**	**38000**
H1-E-low	40.5	--	--	--	--	--
H1-E-med	35.5	--	--	--	--	9700
H1-E-high	32.3	--	--	--	--	--
H1-E-cons	**23.4**	**17.9**	**-2.9**	**3.1**	**48**	**65000**

Fig. 2. Consensus design of a stable EcCM variant. (a) Protein sequences of representative H1-E library members with low (H1-E-low), medium (H1-E-med), and high (H1-E-high) conservation energies, as well as the consensus protein (H1-E-cons). Amino acid positions randomized and held constant are shown in red and black, respectively. The H1-E-low, H1-E-med, and H1-E-high variants could not be produced as soluble dimeric proteins at concentrations necessary for complete biochemical characterization. CD spectra (b) and chemical (c) and thermal (d) denaturation curves were thus only recorded for wild-type EcCM (black) and H1-E-cons (red) at a protein concentration of 5 µM in phosphate-buffered saline at pH 7.5. In contrast to EcCM, thermal melting of H1-E-cons is irreversible due to precipitation of the denatured protein, which may explain its quantitative denaturation above 90 °C. (e) Stability parameters (free energy of unfolding ΔG_u and cooperativity of unfolding m), as well as steady-state parameters for the EcCM-based proteins. Experimental details about gene construction and molecular cloning, protein production, and purification, as well as biophysical characterization, are described in Supporting Information.

revealed H1-E-cons to be 2.6 kcal mol^{-1} more stable than the parent EcCM protein (Fig. 2c and e). The thermal melting profile of the consensus protein exhibits two transitions (Fig. 2d)—the first matching that of wild-type EcCM at 67 °C and a second at 76 °C, suggesting that the newly designed H1/H1′ coiled coil unfolds at a significantly higher temperature than the rest of the protein. Notably, the improvement in stability was not achieved at the cost of catalytic efficiency. An 8-fold decrease in K_m compensates for a 5-fold lower k_{cat}, so the consensus design exhibits an apparent bimolecular rate constant k_{cat}/K_m that is two times larger than that of EcCM itself (Fig. 2e; Supplementary Fig. 5a).

These findings show that a phylogenetically unbiased consensus design can lead to substantial stabilization of secondary structural motifs in mesostable proteins. Analogous experiments with a chorismate mutase from *Methanococcus jannaschii* (MjCM),[26] an EcCM homolog, suggest that this strategy is general. Statistical analysis of sequences of 20 functional clones that were analogously selected from a binary-patterned library of the MjCM H1 helix (Supplementary Table 2 and Supplementary Fig. 1b)[24] afforded the consensus protein H1-M-cons, which was again produced, characterized, and compared to representative library proteins that varied in conservation energy (H1-M-low/H1-M-med/H1-M-high) (Fig. 3). In contrast to proteins from the EcCM library, all variants derived from the thermostable MjCM template yielded correctly folded homodimers



```
         3       11        21        31        41          51        61        71        81        91
MjCM     MIEKLAEIRKKIDEIDNKILKLIAERNSLAKDVAEIKNQL..  ..PEREKYIYDRIRKLCKEHNVDENIGIKIFQILIEHNKALQKQYLEET
M-low    MLEKFLELREELDELNDEMMKLLFKREKILKNIIKLKKNF..  ..FNREKFIFDKLDKFLKEHNVDKKILLKLFELFMENNKILQKNLMDKK
M-med    MIKKLFELRKKIDKLDEKLLKLLMKRKNFFDELFKIKKEF..  ..INREDIMLKKINEMMKEHNVDKKLMFKIFKLLIDENKLIQKKILENN
M-high   MFKDLLKLREEINNMDEELLKLIFKRKNMIEKIIKIKNEL..  ..IKREKMMFKKFENFMKEHNVDEKFILKFIKLLIEENKLIQENLLKEK
M-cons   MLKKLLELREEIDELDEELLKLIFKRKKMIKKIIKIKNEL..  ..LNREKFILKKLKKFMKEHNVDEKLILEIFKLLIEENKLIQKELLKEE

         3       11        21        31        41        51        61        71        81        91
MjCM     MIEKLAEIRKKIDEIDNKILKLIAERNSLAKDVAEIKNQLGIPINDPEREKYIYDRIRKLCKEHNVDENIGIKIFQILIEHNKALQKQYLEET
M-low    MMDDLMKIREEIDELDEKLIKLIIKRNKFMKDLFKLKKELGIPINDFEREKLMLDNFEKIIKEHNVDKKIILKIMNLLIKENKLLQKKFLDEE
M-med1   MLNKLLELRKDIDELDDELMKLMLNRNDLMKNIFEIKKELGIPINDFDREEFILDKIKKFFKEHNVDEDLFIKIFKFLFEKNKMIQKEMLKKE
M-med2   MFEELIKLRDKIDEFDDKMLKLLLKRKNMIKNIFDLKNKMGIPINDLKREELIMDKIKKFIKEHNVDEKIFLEIFKLILDENKMIQKEILEDK
M-high   MMKKILEMREKIDDLDEELLKLFIERKEIIKKIFEMKKELGIPINDLKREELIMDKIKKFIKEHNVDEKIFLEIFKLILDENKMIQKEILEDK
M-cons   MLDNLIEMREEIDELDDELLKLILKRKKIVKNILEIKKELGIPINDLKREELIMDKIKKFIKEHNVDEKIFLEIFKLILDENKMIQKEILEDK
```

(b)

Protein	E_c	ΔG_u	m	k_{cat}	K_m	k_{cat}/K_m
		[kcal mol^{-1}]	[kcal mol^{-1} M^{-1}]	[s^{-1}]	[µM]	[M^{-1} s^{-1}]
MjCM	--	**19.4**		**3.8**	**66**	**58000**
H1-M-low	37.4	13.5	-2.6	--	--	26
H1-M-med	34.7	12.6	-1.6	--	--	6
H1-M-high	29.1	13.0	-2.2	0.56	590	950
H1-M-cons	**22.0**	**16.3**	**-3.1**	**0.21**	**116**	**1850**
H2/3-M-low	46.2	15.9	-2.4	--	--	250
H2/3-M-med	41.2	15.5	-2.7	--	--	125
H2/3-M-high	35.9	17.3	-3.0	--	--	90
H2/3-M-cons	26.9	**18.5**	**-2.9**	**--**	**--**	**9**
H1/2/3-M-low	83.8	13.6	-2.2	0.12	275	430
H1/2/3-M-med1	73.1	11.9	-1.6	--	--	160
H1/2/3-M-med2	57.3	11.0	-1.5	0.12	218	560
H1/2/3-M-high	41.2	13.4	-2.3	0.04	1290	29
H1/2/3-M-cons	**30.1**	**17.1**	**-3.4**	**0.16**	**2230**	**70**

Fig. 3. Consensus design of stable MjCM variants. (a) Protein sequences of representative H1-M, H2/3-M, and H1/2/3-M library members with low (M-low), medium (M-med), and high (M-high) conservation energies, together with the derived consensus proteins (M-cons). Amino acid positions randomized in helices H1 and H2/3 are highlighted in red and blue, respectively. Residues held constant are shown in black. EcCM residue numbering was used for all MjCM-based proteins. The H1/2/3-M library was generated by shuffling the binary-patterned motifs of catalytically active selectants from libraries H1-M and H2/3-M, followed by a second round of genetic selection.[27] Because a single H2/3 fragment was highly abundant in the set of 25 H1/2/3-M selectants, two sequences of medium conservation, one with (H1/2/3-M-med1) and one without (H1/2/3-M-med2) the conserved motif, were chosen for characterization. The original H2/3-M consensus protein was toxic in *E. coli*. To reduce its unusually high net charge from +12 to +1, we mutated six consensus lysines to Glu (five) or Asn (one), which were nearly as equally abundant in the sequence alignment. The redesigned H2/3-M-cons could be produced *in vivo* as a correctly folded protein in high yield. The H1/2/3-M library contained an unprogrammed valine at position 32 in 14 of 25 sequences, which was consequently incorporated into the consensus protein. However, only reference proteins without this mutation were chosen for characterization to allow for unbiased frequency analysis and comparison of conservation energies. The protein originally chosen as H2/3-M-high could not be produced in *E. coli* and was therefore replaced by another variant from the same library that had a similar conservation energy. (b) Biochemical characterization of MjCM-based proteins. Denaturation experiments were performed at a protein concentration of 5 µM in phosphate-buffered saline at pH 7.5. Experimental details about gene construction and molecular cloning, protein production, and purification, as well as biophysical characterization, are described in Supporting Information.

(Supplementary Fig. 2b), enabling a direct comparison of consensus and reference proteins. Chemical denaturation experiments showed that the consensus protein is more stable than the three characterized library variants by 2.8–3.7 kcal mol^{-1} ($\Delta\Delta G_u$) (Fig. 3b; Supplementary Fig. 3a). Consistent with this result, H1-M-low/H1-M-med/H1-M-high begin to denature at temperatures lower than that for H1-M-cons, although meaningful T_m values could not be determined from the melting profiles because only partial unfolding was observed below 95 °C (Supplementary Fig. 4a). While H1-M-cons is not quite as robust as wild-type MjCM, its stability is comparable to that of H1-E-cons. H1-M-cons is also more active than the proteins selected directly from the patterned libraries and, in some cases, considerably so. Relative to wild-type MjCM, the catalytic efficiency of library members of low and medium conservation is reduced 2000-fold to 10,000-fold, whereas the consensus protein is only 30-fold less active (Fig. 3b; Supplementary Fig. 5b). The lower activity of H1-M-cons compared to the wild type may reflect a decreased tolerance of the thermostable scaffold for this particular simplified alphabet.

Because analogous selection experiments have also been reported for the H2/3 helices of MjCM alone and in combination with H1,[27] it was possible to extend the consensus design approach to the entire MjCM scaffold. Proteins H2/3-M-cons and H1/2/3-M-cons, as well as library members possessing low, medium, and high conservation energies, were generated by the procedures described above (input sequences are summarized in Supplementary Tables 3 and 4 and Supplementary Fig. 1c–e; CD spectra of the proteins are shown in Supplementary Fig. 2c and d). Like the H1-M variants, both consensus proteins were obtained as soluble homodimers in good yield. Indeed, in contrast to wild-type MjCM, most of the binary-patterned library proteins in this series exhibited favorable folding *in vivo*, suggesting that this property is actively selected for in the complementation experiments. While H2/3-M-cons and its progenitors show high thermostabilities comparable to that of wild-type MjCM (T_m>90 °C), H1/2/3-M-cons begins to denature at a higher temperature than the input proteins (Supplementary Fig. 4b and c). Chemical denaturation experiments showed that both H2/3-M-cons and H1/2/3-M-cons have significantly higher ΔG_u values than the original library variants, indicating considerable gains in stability (Fig. 3b; Supplementary Fig. 3b and c). Simultaneous optimization of all three helices gave a $\Delta\Delta G_u$ of 4.6±1.2 kcal mol^{-1}, which reflects roughly additive contributions of the H1-M (3.3±0.5 kcal mol^{-1}) and H2/3-M (2.3±0.9 kcal mol^{-1}) consensus designs. Additivity is also manifest in the stronger unfolding cooperativity m of H1/2/3-M-cons (−1.5 kcal mol^{-1} M^{-1}) relative to H1-M-cons (−1.0 kcal mol^{-1} M^{-1}) and H2/3-M-cons (−0.2 kcal mol^{-1} M^{-1}).

The success of library-based consensus design is notable given that the binary-patterned input proteins do not exhibit a consistent correlation between the calculated conservation energy E_c and the experimentally determined stability ΔG_u (Fig. 3b). This is presumably due to the fact that the contributions of individual sites to stability are not identical. As a consequence, a library protein with a low E_c can be less stable than a protein with a higher E_c if a few key residues that contribute disproportionately to stability are lacking in the former but are present in the latter. In contrast, the consensus sequence, which has the lowest E_c by definition, always contains the optimal (most frequent) residue at each variable position and can thus be expected to lead to a protein with enhanced stability. In the H2/3-M series, E_c is also a poor predictor of catalytic activity. In contrast to H1-E and H1-M, the k_{cat}/K_m values for the H2/3-M enzymes show a weak inverse correlation with sequence conservation (i.e., with $-E_c$), and this trend carries over to the fully remodeled H1/2/3-M proteins (Fig. 3b; Supplementary Fig. 5c and d). Given the severely restricted set of amino acids used to construct the binary-patterned libraries and the large fraction of the protein that was mutagenized, this reduction in catalytic activity is not terribly surprising. Some of the stabilizing mutations may subtly alter the placement of key catalytic residues or impair substrate access to the completely buried active site. Such explanations would be consistent with the common observation that activity is a local property, often resulting from highly synergistic interactions of a few residues,[28] whereas protein stability arises from small but additive effects distributed over the entire molecule.[8,29]

Overall, our results show that consensus design based on unbiased input sequences derived from binary-patterned libraries can reliably predict stabilized proteins. Any phylogenetic bias was precluded a priori by randomization of just a single parental sequence, followed by functional selection. The four libraries we tested were based on two different scaffolds and involved partial or complete diversification of secondary structural elements in the protein; each directly afforded consensus designs that were substantially more stable than any of the input proteins. Because side-chain diversity was greatly restricted through the use of a simplified alphabet, these effects could be achieved by statistical analysis of a relatively small number of binary-patterned sequences (≤30).

The fact that the consensus approach, which traditionally relies on many sequences of naturally evolved proteins, works efficiently with a small set of artificially randomized sequences generated by “synthetic evolution” experiments is notable. It indicates that simple stochastic sampling of residues selected from unbiased libraries composed of a restricted alphabet is a valid alternative to the consensus analysis of evolutionarily related proteins. Successful protein stabilization by a purely statistical method—independent of any specific mechanism operating during natural evolution—represents an intriguing, potentially general finding.

Because library-based consensus design does not depend on the availability of structurally and functionally related natural homologs, it represents a potentially powerful strategy for stabilizing proteins of industrial or therapeutic interest. For practical applications, the extreme mutagenesis strategy adopted in our proof-of-principle study is unlikely to be necessary. Instead, as suggested by our experiments with EcCM, the design of relatively small focused libraries should suffice to achieve significant increases in stability without loss of biological function. Optimizing the amino acid alphabet used for randomization with respect to structural propensities and functional diversity, and taking covariation into account in the design process[30] are additional strategies that might be exploited to maximize stabilization while preserving activity.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2010.04.039

References

1. Jäckel, C., Kast, P. & Hilvert, D. (2008). Protein design by directed evolution. *Annu. Rev. Biophys.* **37**, 153–173.
2. Toscano, M. D., Woycechowsky, K. J. & Hilvert, D. (2007). Minimalist active-site redesign: teaching old enzymes new tricks. *Angew. Chem. Int. Ed.* **46**, 3212–3236.
3. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M. & Witholt, B. (2001). Industrial biocatalysis today and tomorrow. *Nature*, **409**, 258–268.
4. Fasan, R., Chen, M. M., Crook, N. C. & Arnold, F. H. (2007). Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties. *Angew. Chem. Int. Ed.* **46**, 8414–8418.
5. Polizzi, K. M., Bommarius, A. S., Broering, J. M. & Chaparro-Riggers, J. F. (2007). Stability of biocatalysts. *Curr. Opin. Chem. Biol.* **11**, 220–225.
6. Lehmann, M. & Wyss, M. (2001). Engineering proteins for thermostability: the use of sequence alignments *versus* rational design and directed evolution. *Curr. Opin. Biotechnol.* **12**, 371–375.
7. Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. (1994). Sequence statistics reliably predict stabilizing mutations in a protein domain. *J. Mol. Biol.* **240**, 188–192.
8. Wintrode, P. L. & Arnold, F. H. (2001). Temperature adaptation of enzymes: lessons from laboratory evolution. *Adv. Protein Chem.* **55**, 161–225.
9. Ohage, E. C., Graml, W., Walter, M. M., Steinbacher, S. & Steipe, B. (1997). β-Turn propensities as paradigm for the analysis of structural motifs to engineer protein stability. *Protein Sci.* **6**, 233–241.
10. Lehmann, M., Kostrewa, D., Wyss, M., Brugger, R., D'Arcy, A., Pasamontes, L. & van Loon, A. P. G. M. (2000). From DNA sequences to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase. *Protein Eng.* **13**, 49–57.
11. Mosavi, L. K., Minor, D. L. & Peng, Z. Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl Acad. Sci. USA*, **99**, 16029–16034.
12. Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. (2003). Design of stable α-helical arrays from an idealized TPR motif. *Structure*, **11**, 497–508.
13. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503.
14. Vazquez-Figueroa, E., Chaparro-Riggers, J. & Bommarius, A. S. (2007). Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. *ChemBioChem*, **8**, 2295–2301.
15. Bloom, J. D. & Glassman, M. J. (2009). Inferring stabilizing mutations from protein phylogenies: application to influenza hemagglutinin. *PLoS Comput. Biol.* **5**, e1000349.
16. Nikolova, P. V., Henckel, J., Lane, D. P. & Fersht, A. R. (1998). Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. *Proc. Natl Acad. Sci. USA*, **95**, 14675–14680.
17. Rath, A. & Davidson, A. R. (2000). The design of a hyperstable mutant of the Abp1p SH3 domain by sequence alignment analysis. *Protein Sci.* **9**, 2457–2469.
18. Loening, A. M., Fenn, T. D., Wu, A. M. & Gambhir, S. S. (2006). Consensus guided mutagenesis of *Renilla* luciferase yields enhanced stability and light output. *Protein Eng. Des. Sel.* **19**, 391–400.
19. Lehmann, M., Loch, C., Middendorf, A., Studer, D., Lassen, S. F., Pasamontes, L. *et al.* (2002). The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.* **15**, 403–411.
20. Ryan, B. J., O'Connell, M. J. & O'Fagain, C. (2008). Consensus mutagenesis reveals that non-helical regions influence thermal stability of horseradish peroxidase. *Biochimie*, **90**, 1389–1396.
21. Amin, N., Liu, A. D., Ramer, S., Aehle, W., Meijer, D., Metin, M. *et al.* (2004). Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.* **17**, 787–793.
22. Li, Y. G., Drummond, D. A., Sawayama, A. M., Snow, C. D., Bloom, J. D. & Arnold, F. H. (2007). A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments. *Nat. Biotechnol.* **25**, 1051–1056.
23. Lee, A. Y., Karplus, P. A., Ganem, B. & Clardy, J. (1995). Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase. *J. Am. Chem. Soc.* **117**, 3627–3628.
24. Besenmatter, W., Kast, P. & Hilvert, D. (2007). Relative tolerance of mesostable and thermostable protein homologs to extensive mutation. *Proteins*, **66**, 500–506.
25. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M. & Hecht, M. H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. *Science*, **262**, 1680–1685.
26. MacBeath, G., Kast, P. & Hilvert, D. (1998). A small, thermostable, and monofunctional chorismate mutase from the archeon *Methanococcus jannaschii*. *Biochemistry*, **37**, 10062–10073.
27. Taylor, S. V., Walter, K. U., Kast, P. & Hilvert, D. (2001). Searching sequence space for protein catalysts. *Proc. Natl Acad. Sci. USA*, **98**, 10596–10601.
28. Holliday, G. L., Mitchell, J. B. O. & Thornton, J. M. (2009). Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577.
29. Wells, J. A. (1990). Additivity of mutational effects in proteins. *Biochemistry*, **29**, 8509–8517.
30. Magliery, T. J. & Regan, L. (2004). Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif. *J. Mol. Biol.* **343**, 731–745.