An Active Enzyme Constructed from a 9-Amino Acid Alphabet*5

Kai U. Walter, Katherina Vamvaca, and Donald Hilvert

From the Laboratory of Organic Chemistry, Swiss Federal Institute of Technology, ETH Hönggerberg, CH-8093 Zürich, Switzerland

Nature employs a set of 20 amino acids to produce a repertoire of protein structures endowed with sophisticated functions. Here, we combined design and selection to create an enzyme composed entirely from a set of only 9 amino acids that can rescue auxotrophic cells lacking chorismate mutase. The simplified protein captures key structural features of its natural counterpart but appears to be somewhat less stable and more flexible. The potential of a dramatically reduced amino acid alphabet to produce an active catalyst supports the notion that primordial enzymes may have possessed low amino acid diversity and suggests that combinatorial engineering strategies, such as the one used here, may be generally applied to create enzymes with novel structures and functions.

MATERIALS AND METHODS

Reagents—Restriction enzymes were from New England Biolabs. T4 DNA ligase was from Fermentas. Pfu Turbo polymerase was from Stratagene. Protein concentration was determined with the Coomassie Plus protein assay reagent (Pierce), using bovine serum albumin as the calibration standard.

Library Construction—Libraries were constructed by combinatorial site-directed mutagenesis. N- and C-terminal fragments were amplified from plasmid pKT-A3 containing the parent CM gene (9) using outside primers sspREX03 (21 bp, 5’-CATCCGGCTGTATAATGTGT-3’) and TMUTN (19 bp, 5’-ATCAAGCGGCCGACTAGT-3’) in combination with an overlapping mutagenic oligonucleotide pair for each library: 01KW9RQX (21 bp, 5’-ACCAAAATGATCRAMAAAGAAA-3’)/02KW9RXQ (20 bp, 5’-AGCAATTTCTTCTTYGATCAT-3’) for the Gln88Xaa library (Xaa: Glu, Asn, Asp); 01KW9L2X (36 bp, 5’-CTACAAAGGAGASAAACWTS6GACGAAGACCTCCTCAT-3’)/02KW-9L2X (36 bp, 5’-GAGGCTTCTGTC5AWGTTSYCTTCTTGAGA-5’/AACTTT-3’) for the L2 library; and 01KW9L1X4 (56 bp, 5’-CTCTAACAGGGAAGAAGACTGDGDGWSDSGSACAGACCTCGA-CGGTGAAGATGT-3’)/02KW9L1X4 (56 bp, 5’-AACTCTTCACGGTGGAATCTGTSSWSSHHWSHWHCAGGTCTTTGAGT-3’) for the L1 library, where R = A, G; M = A, C; K = G, T; Y = C, T; S = C, G; W = A, T; D = A, G, T and H = A, C, T. Randomized codons are underlined. All oligonucleotides were obtained from Microsynth (Balgach, Switzerland). PCR products were purified by agarose gel electrophoresis. To obtain full-length genes, complementary fragment pairs were annealed and extended in a second PCR using the outside primers sspREX03 and TMUTN. Genes were treated with restriction enzymes (NdeI/XhoI), purified, and ligated with the 2800-bp NdeI/XhoI fragment of vector pKT-A3 (9). The L1 library was constructed from the active members of the L2 library, which was constructed in turn from the Gln88Xaa library. Final constructs were confirmed by DNA sequencing.

Selection—Library plasmid pools were transformed into the KA12/pKIMP-UAUC selection strain with efficiencies greater than 106 clones/μg of library DNA (10). Transformed cells were washed in minimal medium and plated onto M9c minimal medium plates (lacking Tyr and Phe). Plates were typically incubated for 3 days at 30 °C or 6 days at 25 °C.

Protein Production and Purification—The gene encoding the simplified 9-amino acid CM was subcloned as a 282-bp NdeI/XhoI fragment into the T7 promoter-based expression vector pET-22b-pATCH, which appends a His12 tag at the C terminus, and overexpressed in Escherichia coli strain KA13 (11). Expression from the T7 promoter was induced with isopropyl-1-thio-β-d-galactopyranoside. The enzyme was purified by affinity chromatography on nickel-nitrilotriacetic acid-agarose resin (Qiagen) followed by preparative fast protein liquid chromatography on a Superdex 75 (26/60) gel filtration column (Amersham Biosciences).

* This work was supported by the Swiss Federal Institute of Technology and the Schweizerischer Nationalfonds. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
* The on-line version of this article (available at http://www.jbc.org) contains a supplemental text and three supplemental figures.
1 To whom correspondence should be addressed: Swiss Federal Institute of Technology, ETH Hönggerberg, CH-8093 Zurich, Switzerland. Tel.: 41-1-632-3176; Fax: 41-1-632-1486; E-mail: hilvert@org.chem.ethz.ch.
2 The abbreviations used are: CM, chorismate mutase; ANS, 1-anilinonaphthalene 8-sulfonate; EcCM, E. coli CM; MjCM, M. jannashii CM.
MjCM was produced as described previously (12). Protein integrity was confirmed by mass spectrometry.

Kinetic Assays—CM activity measurements were performed at 20 °C in acetate buffer (pH 5) as reported previously (12). The inhibition assay was carried out under the same conditions in the presence of a transition state analog compound 1 (red), based on the x-ray structure of EcCM (15). Residues Gln69 and Ser94 in EcCM are substituted with Glu88 and Asn84 in the 9-amino acid enzyme. Residue numbers are referenced to EcCM.

RESULTS

CMs catalyze the rearrangement of chorismate to prephenate and are essential for the biosynthesis of the aromatic amino acids Tyr and Phe in bacteria, fungi, and plants. The binary patterned mutase that served as the natural mutase (AroQ CMs of E. coli) and the simplified enzyme and the transition state analog (compound 1, 0–100 μM) (13).

ANS Fluorescence—Fluorescence measurements were performed as described previously (14). The number of protected labile H atoms was calculated by subtracting the observed mass from the mass of the fully deuterated protein.

CD Spectroscopy—Thermal unfolding curves of a 16 μM protein solution (in the presence and absence of 320 μM inhibitor) were monitored by CD at 222 nm, increasing the temperature from 10 to 95 °C in 1 °C steps.

Despite dramatic simplification relative to natural mutases, the selected 9-amino acid enzymes are fully functional in vivo, replacing wild-type CM in bacterial metabolism (Fig. 2B). Bacterial cells encoding the simplified catalysts grew only somewhat slower than those with wild-type enzymes from E. coli (EcCM) and Methanococcus jannashii (MjCM), presumably due to the Q88E active site mutation. In natural AroQ CMs, Gln98 serves to stabilize the transition state during catalysis, acting as a hydrogen bond donor (15), and replacing this residue with glutamic acid shifts the pH optimum to approximately pH 5 (17). Under physiological conditions (pH 7), glutamic acid should be largely deprotonated, which would be deleterious for substrate binding and catalysis (Fig. 1B).

One of the simplified CM variants displaying high activity in vivo (Fig. 2A) was overproduced for detailed characterization. A His6 tag was

FIGURE 1. AroQ structure and active site (15). A, the homodimeric EcCM is shown with a transition state analog inhibitor (13) bound at its active sites; the two identical polypeptide chains are colored blue and pink for clarity. B, proposed interactions between residues in the evolved active site of the simplified enzyme and the transition state analog inhibitor, compound 1 (red), based on the x-ray structure of EcCM (15). Residues Gln88 and Ser84 in EcCM are substituted with Glu88 and Asn84 in the 9-amino acid enzyme. Residue numbers are referenced to EcCM.
9-Amino Acid Enzyme

FIGURE 2. Strategy for engineering simplified CMs. A, amino acid sequences of the binary patterned parent protein (14 building blocks) and simplified variants (13, 11, and 9 building blocks). Polar and apolar residues are shown in red and blue, respectively. Residues that do not belong to the 8-amino acid alphabet (black) were replaced by selection in three steps. Arginine residues were held constant. B, in vivo complementation. The ability of the minimized enzyme to complement the CM deficiency of KA12/pKIMP-UAUC cells (10) was evaluated by streaking clones on minimal medium (M9c) in the absence (left) and presence (right) of Tyr and Phe. Cells bearing the simplified enzyme without (i) and with (ii) a His6 tag grow similarly. Cells containing wild-type EcCM (iii) and MjCM (iv) or the vector (v) are included as positive and negative controls.

attached at its C terminus to facilitate purification. The addition of this tag does not alter the behavior of the protein in vivo (Fig. 2B). The overproduced enzyme was purified by affinity chromatography (nickel-nitritoltriacetic acid) and gel filtration, typically yielding 6 mg of soluble protein/liter of liquid culture. Although the activity of the wild type remains constant over a broad pH range (pH 5–9) (12), the simplified enzyme displays optimal activity at pH 5 as a result of the Q88E mutation (Fig. 1B, Supplemental Fig. 1). Under these conditions, its $k_{cat}$ value (0.9 s$^{-1}$) is only 3-fold lower than that of MjCM (3.1 s$^{-1}$), whereas its $K_m$ value (830 $\mu$M) is 40-fold higher than that of the wild type (20 $\mu$M). In line with its lower specific activity, the simplified CM binds the transition state analog inhibitor (compound 1, Fig. 1B) 50-fold less tightly ($K_i = 10$ $\mu$M) than the wild-type enzyme ($K_i = 0.21$ $\mu$M).

The simplified enzyme is highly helical as judged by far UV CD spectroscopy (Supplemental Fig. 2A) and elutes as a dimer from a size-exclusion chromatography column (Supplemental Fig. 2B) like typical AroQ mutases (12). Its $^1$H-NMR spectrum exhibits similar peak dispersion to wild-type MjCM (Supplemental Fig. 3, A and B) but generally contains fewer peaks due to the reduced amino acid diversity. The addition of the transition state analog inhibitor causes only a slight increase in NMR peak dispersion (Supplemental Fig. 3, C and D) and in helical content of the enzyme (Supplemental Fig. 2A). However, some of the biophysical properties differ somewhat from wild-type CMs. For instance, the minimized mutase undergoes slightly faster H/D exchange than its natural counterpart (Fig. 3A). An additional 10–14 backbone amides (out of ~100) appear to exchange more rapidly in the simplified enzyme, suggesting greater fraying of its helices (18). Addition of the transition state analog tightens both structures to a similar extent ($\approx 3$–5%) (14). Chemical denaturation studies showed that the minimized protein is less stable ($\Delta G_U$ (H$_2$O) = 9.5 kcal mol$^{-1}$) than MjCM ($\Delta G_U$ (H$_2$O) = 24.0 kcal mol$^{-1}$) (12). Its thermal unfolding is reversible but noncooperative, although the addition of inhibitor induces modest cooperativity ($T_m = 55$ °C) (Fig. 3B). For comparison, the wild type unfolds with a high degree of cooperativity, even in the absence of ligand ($T_m = 88$ °C) (12). Finally, the protein binds the hydrophobic dye ANS (19) to a greater extent than MjCM, leading to enhancement of fluorescence and a blue shift of the emission maximum (Fig. 3C). Addition of the transition state analog impairs ANS binding, probably by tightening the packing of the hydrophobic core.

DISCUSSION

The fact that less than half of the proteinogenic amino acids are sufficient to construct a chorismate mutase that is metabolically competent, conferring near wild-type levels of cell growth to an auxotrophic host, supports the hypothesis that primitive enzymes may have contained a reduced set of building blocks (1). Nevertheless, dispensing with 11 of the 20 standard amino acids significantly reduces the diversity of favorable internal packing interactions, leading to destabilization of the overall protein structure. As a consequence, the simplified enzyme displays certain properties (noncooperative thermal unfolding, ANS binding) reminiscent of the molten globule state (20, 21). Interestingly, it appears to be less molten than a previously engineered CM monomer (14, 22), probably due to additional stabilizing interactions in the hydrophobic region located between its two active sites. In terms of enzyme evolution, these two proteins can be viewed as models for primitive catalysts in that both are less complex than a modern day enzyme such as MjCM with respect to primary sequence or quaternary structure.

The reduced amino acid alphabet we used to create the minimized enzyme was arbitrarily chosen to suit a design strategy based on two
Some of these residues (especially Ala and Gly) have been generally considered indispensable building blocks for the formation of helices and turns in severely simplified proteins (23, 24). Given the restriction (no small amino acids) and redundancy (Asp/Glu and Ile/Leu) of the 9-amino acid alphabet, more extensive simplification of the mutase may well be feasible.

The CM selection system, like selection systems more generally (25), is ideally suited to explore alternative amino acid alphabets to identify the best minimal set, as well as to pinpoint residues that are critical for folding and function by re-expanding these alphabets via additional rounds of directed evolution. The optimal building blocks for producing active α-helical and β-sheet proteins are unlikely to be the same, and analogous experiments on structurally distinct CMs, which include all-α-helical (15), pseudo-α/β-barrel (26), and all-β-sheet architectures (27), promise valuable insights into the unique structural requirements of different scaffolds.

In this context, recent advances in computation, which have made possible the design of novel proteins with atomic level accuracy (28–30) as well as the redesign of natural proteins to create novel binding (31) or catalytic sites (32), can be expected to reinforce experimental efforts to identify viable alternative alphabets for specific folds. In turn, simplified alphabets may facilitate computational searching by reducing the number of sequences that need to be sampled and should bring us closer to the still unrealized goal of designing new enzymes from scratch.

Acknowledgments—We are grateful to Sean V. Taylor and Kenneth J. Woyceckowsky for helpful discussions and critical reading of the manuscript. We thank Alexander Eletsky for assistance with NMR measurements.

REFERENCES

1. Davis, B. K. (2002) Prog. Biophys. Mol. Biol. 79, 77–133
2. Regan, L., and DeGrado, W. F. (1988) Science 241, 976–978
3. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) Science 262, 1680–1685
4. Davidson, A. R., Lumb, K. J., and Sauer, R. T. (1995) Nat. Struct. Biol. 2, 856–863
5. Schafmeister, C. E., LaPorte, S. L., Miercke, L. J. W., and Stroud, R. M. (1997) Nat. Struct. Biol. 4, 1039–1046
6. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantharova, V. P., Yi, Q., and Baker, D. (1997) Nat. Struct. Biol. 4, 805–809
7. Silverman, J. A., Balasubramanian, R., and Harbury, P. B. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 3092–3097
8. Akanuma, S., Kigawa, T., and Yokoyama, S. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 13549–13553
9. Taylor, S. V., Walter, K. U., Kast, P., and Hilvert, D. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 10596–10601
10. Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 5043–5048
11. MacBeath, G., and Kast, P. (1998) BioTechniques 24, 789–794
12. MacBeath, G., Kast, P., and Hilvert, D. (1998) Biochemistry 37, 10062–10073
13. Bartlett, P. A., and Johnson, C. R. (1985) J. Am. Chem. Soc. 107, 7792–7793
14. Vamvaca, K., Vögeli, B., Kast, P., Pervushin, K., and Hilvert, D. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 12860–12864
15. Lee, A. Y., Karpus, P. A., Ganem, B., and Clardy, J. (1995) J. Am. Chem. Soc. 117, 3627–3628
16. Firth, A. E., and Patrick, W. M. (2005) Bioinformatics (Oxf.) 21, 3314–3315
17. Zhang, S., Kongsaeree, P., Clardy, J., Wilson, D. B., and Ganem, B. (1996) Bioorg. Med. Chem. 4, 1015–1020
18. Smith, D. L., Deng, Y., and Zhang, Z. (1997) J. Mass Spectrom. 32, 135–146
19. Stryer, L. (1968) Science 162, 526–533
20. Pitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E., and Razgulyaev, O. I. (1990) FEBS Lett. 262, 20–24
21. Betz, S. F., Raleigh, D. P., and DeGrado, W. F. (1993) Carr. Opin. Struct. Biol. 3, 601–610

FIGURE 3. Biochemical characterization. A, H/D exchange of the fully minimized protein (green circles) and the wild-type MjCM (black diamonds) at pH 6.5. MjCM contains 229 labile H atoms, of which 99 (101 amino acids – 2 Pro residues) are backbone amides, whereas the minimized protein has a total of 230 labile H atoms and 101 backbone amides. In general, backbone amides involved in secondary structure formation generally undergo slower exchange than labile side-chain H atoms (18). B, thermal denaturation of the simplified enzyme in the absence (green circles) and presence of inhibitor (red squares). The published curve (12) for the wild type is provided for comparison. C, ANS fluorescence. Emission spectra of ANS alone (gray) with either the free (black) and ligand-bound (blue) MjCM or the free (green) and ligand-bound (red) minimized mutase are shown.
22. MacBeath, G., Kast, P., and Hilvert, D. (1998) *Science* **279**, 1958–1961
23. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties*, 2nd Ed., pp. 171–260, W. H. Freeman, New York
24. Wang, J., and Wang, W. (1999) *Nat. Struct. Biol.* **6**, 1033–1038
25. Taylor, S. V., Kast, P., and Hilvert, D. (2001) *Angew. Chem. Int. Ed. Engl.* **40**, 3311–3325
26. Chook, Y. M., Ke, H., and Lipscomb, W. N. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 8600–8603
27. Haynes, M. R., Stura, E. A., Hilvert, D., and Wilson, I. A. (1994) *Science* **263**, 646–652
28. Dahiyat, B. I., and Mayo, S. L. (1997) *Science* **278**, 82–87
29. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) *Science* **302**, 1364–1368
30. Harbury, P. B., Plecs, J. J., Tidor, B., Alber, T., and Kim, P. S. (1998) *Science* **282**, 1462–1467
31. Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) *Nature* **185–190
32. Dwyer, M. A., Looger, L. L., and Hellinga, H. W. (2004) *Science* **304**, 1967–1971