A *de Novo* Designed Protein with Properties That Characterize Natural Hyperthermophilic Proteins

Xin Jiang, Edmund J. Bishop, and Ramy S. Farid*

Department of Chemistry
Rutgers, The State University of New Jersey
Newark, New Jersey 07102

Received September 23, 1996

Enzymes that evolved from thermophiles and hyperthermophiles, organisms that thrive at temperatures above 60 and 80 °C, respectively, have attracted enormous attention because they serve as paradigms for delineating factors responsible for protein stability as well as exhibit considerable potential for a number of biotechnological applications. This paper describes a *de novo* designed protein that exhibits key features that characterize natural hyperthermophilic proteins (HYP). The hydrophobic interior of the synthetic protein was designed using a new computer program called CORE. Demonstrated successful design and synthesis of proteins that are stable at elevated temperatures lays the ground work for the development of artificial enzymes that function at temperatures approaching 100 °C.

Despite limited data, it is clear that HYPs are stabilized at high temperature by a unique combination of a number of optimized structural properties that include high rigidity, increased hydrophobic interactions, enhanced packing efficiency, larger number of H-bonds and salt-bridges, decreased conformational strain, and lower entropy of unfolding relative to their mesophilic analogs.[1] In an effort to design an artificial HYP from scratch, we have developed a computer program (CORE) that optimizes the hydrophobic-packing efficiency and side chain rigidity of core residues.[2] Despite the potential importance of other stabilizing forces, efficient hydrophobic packing, and increased rigidity alone lead to enhanced thermal stability in natural thermophilic proteins.

We have chosen the backbone structure of the ubiquitous antiparallel four-helix bundle as the basic motif for our synthetic protein.[3] As a result of the pioneering work of several groups, construction of four-α-helical peptides has become almost routine.[4] Despite this, synthesis of proteins with "native-like" stability does not meet with consistent success and is still largely an unrealized goal. Our design incorporates two peptide chains, A and B, linked by a disulfide bond between a C-terminal cysteine on chain A and an N-terminal cysteine on chain B. This antiparallel helix-link-helix unit is designed to dimerize forming an antiparallel four-helix protein. Exterior charged residues (glutamate and lysine) were positioned to assure optimal electrostatic interactions following well-established rules.[4a,5] CORE was then employed to optimize the sequence of the remaining 24 (2 sets of 12) hydrophobic core residues. The resulting sequence[6a] is depicted as two-dimensional helical projections shown in Figure 1.

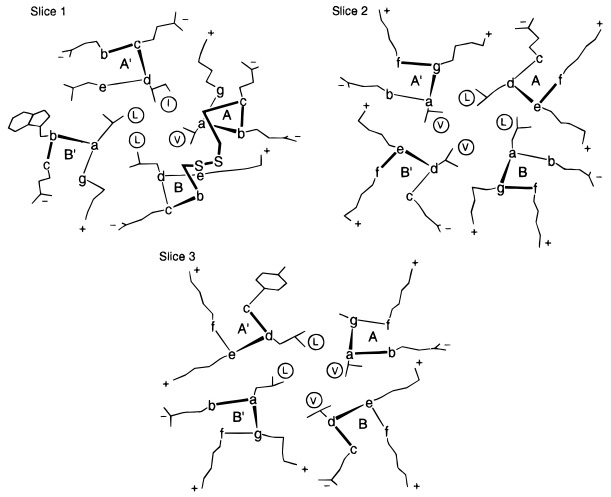

Figure 1. Helical projections of the three unique slices of the *de novo* designed protein. Charged residues were positioned to maximize interhelical electrostatic interactions: lysine (+) at positions "f" and "g" and glutamate (−) at positions "b" and "c". The identity of hydrophobic core residues, at positions "a" and "d" were determined using the computer program CORE (see text).

The three unique slices highlight hydrophobic interactions (determined by CORE) and potential charged interactions as well as the disulfide bond (S−S in slice 1) that links chains A and B. None of the slices contain only leucine residues, the traditional core residues exploited in *de novo* designed four-α-helical proteins. Each slice contains two leucines "balanced" by either two valines or one valine and one isoleucine; this arrangement (two leucines balanced by smaller hydrophobic residues) appears to optimize hydrophobic packing and minimize side chain mobility of the core residues.[6b] Although a similar conclusion was reached in systematic studies involving both the redesigning of a natural four-helix protein[7] and the *de novo* synthesis of a four-helix peptide,[4b] positioning of Leu and a smaller hydrophobic residue (Val or Ala) in the heptad "a" and

(1) (a) Cavagnero, S.; Zhou, Z. H.; Adams, M. W. W.; Chan, S. I. *Biochemistry* **1995**, *34*, 9865−73. (b) Adams, M. W. W. *Annu. Rev. Microbiol*. **1993**, *47*, 627−58. (c) Bradley, E. A.; Stewart, D. E.; Adams, M. W. W.; Wampler, J. E. *Protein Sci*. **1993**, *2*, 650−65. (d) Chen, C.-H.; Roth, L. G.; MacColl, R.; Berns, D. S. *Biophys. Chem*. **1994**, *50*, 313−21. (e) Vieille, C.; Zeikus, J. G. *Trends Biotechnol*. **1996**, *14*, 183−90.

(2) CORE utilizes a previously published program (Shenkin, P. S.; Farid, H.; Fetrow, J. S. *Proteins: Struct., Funct., Genet*. **1996**, *26*, 323−52) that accurately determines side chain conformations. Building on this program, CORE is designed to output protein sequences based on an inputted backbone structure via Metropolis-driven simulated-annealing and Monte Carlo runs. A typical simulated-annealing run produces a single sequence near or at the global minimum with respect to maximal hydrophobic packing and minimal side chain motion. A subsequent low-temperature Monte Carlo search, initiated with the simulated-annealing sequence, generates a collection of proteins (typically between 2 and 15) all within the global minimum well. The lowest energy sequence is then predicted to exhibit enhanced thermal stability. CORE successfully regenerates the native core residue sequence of several native hyperthermophilic proteins with no prior knowledge of the native core sequence (Pistor, E.; Farid, H.; Farid, R. S. Manuscript in preparation).

(3) The prototypical backbone structure of the naturally occurring DNA-binding protein Rop was used as a template. The Rop structure is an antiparallel dimer of a helix-turn-helix unit consisting of 63 amino acids.

(4) (a) Bryson, J. W.; Betz, S. F.; Lu, H. S.; Suich, D. J.; Zhou, H. X.; O'Neil, K. T.; DeGrado, W. F. *Science* **1995**, *270*, 935−41. (b) Betz, S. F.; DeGrado, W. F. *Biochemistry* **1996**, *35*, 6955−62. (c) Hill, C. P.; Andersong, D. H.; Wesson, L.; DeGrado, W. G.; Eisenberg, D. *Science* **1990**, *249*, 543−6. (d) DeGrado, W. F.; Wasserman, Z. R.; Lear, J. D. *Science* **1989**, *243*, 622−8. (e) Betz, S. F.; Raleigh, D. P.; DeGrado, W. F. *Curr. Opin. Struct. Biol*. **1993**, *3*, 601−10. (f) Robertson, D. E.; Farid, R. S.; Moser, C. C.; Urbauer, J. L.; Mulholland, S. E.; Pidikiti, R.; Lear, J. D.; Wand, A. J.; DeGrado, W. F.; Dutton, P. L. *Nature* **1994**, *368*, 425−32. (g) Ghadiri, M. R.; Soares, C.; Choi, C. *J. Am. Chem. Soc*. **1992**, *114*, 4000−2. (h) Desjarlais, J. R.; Handel, T. M. *Curr. Opin. Biotechnol*. **1995**, *6*, 460−6.

(5) (a) Monera, O. D.; Kay, C. M.; Hodges, R. S. *Protein Sci*. **1994**, *3*, 1984−91. (b) Fairman, R.; Chao, H.-G.; Lavoie, T. B.; Villafranca, J. J.; Matsueda, G. R.; Novotny, J. *Biochemistry* **1996**, *35*, 2824−9.

(6) (a) Peptide A is NH_2-KEEEILKKVEYLKKAVEELKKKVEEGC-C(O)NH_2 and peptide B is NH_2-CGELKKKLEEVKKKLEEVKKKL-WEKK-C(O)NH_2. (b) The two-dimensional representation of the protein in Figure 1 does not highlight interslice hydrophobic interactions. For example, in the model, Leu at position "a" in slice 2 (helix B) forms favorable hydrophobic interactions with Val at position "a" in slice 3 (helix A).

(7) (a) Munson, M.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci*. **1994**, *3*, 2015−22. (b) Munson, M.; Balasubramanian, S.; Fleming, K. G.; Nagi, A. D.; O'Brien, R.; Sturtevant, J. M.; Regan, L. *Protein Sci*. **1996**, *5*, 1584−93.

(8) Peptides were synthesized by FMOC chemistry and purified by reversed-phase HPLC, yielding pure material as assessed by analytical HPLC and ion spray mass spectrometry.

(9) See Supporting Information.

S0002-7863(96)03319-7 CCC: $14.00

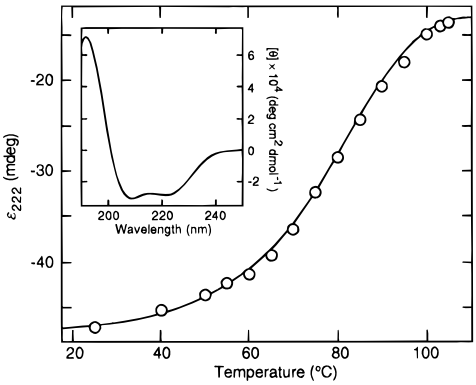

Figure 2. Temperature dependence of the far-UV CD signal at 222 nm of HYP-1 at 10 μM in 20 mM phosphate buffer, pH 2, 1 M NaCl. Inset: CD spectrum of the protein under the same conditions. The CD spectra at pH 2 and 7 are superimposable.

"d" positions was done without explicit consideration of interslice interactions; Leu was placed exclusively in position "d" and Val or Ala in position "a". In the present CORE-designed protein, Leu is found in both "a" and "d" positions.

Peptides A and B were synthesized[8] and mixed in equimolar amounts at pH 8.8 for 12 h yielding three disulfide-linked peptides, AA, AB, and BB, in a 1:2:1 ratio.[9] The designed peptide, AB, was isolated from the homodimers via HPLC, and its identity reconfirmed by MS. Size exclusion chromatography confirmed that AB dimerizes to form a four-helix structure, HYP-1. The 2D TOCSY spectrum of HYP-1 in D_2O (not shown) shows a single set of crosspeaks for the two aromatic residues (Tyr and Trp), confirming a symmetrical dimeric structure. The chemical shifts of the side chain protons for these residues are consistent with solvent exposed residues, as expected by the design (Figure 1).

The CD spectrum of HYP-1 at pH 7 (not shown) is typical of four-helix bundles and indicates that the protein is 90 ± 10% α-helical (92% α-helix is predicted from the sequence). Remarkably, at pH 2 and high ionic strength (>0.5 M NaCl), the CD spectrum (Figure 2) is nearly identical to the pH 7 spectrum, indicating little change in the secondary structure upon neutralization of the glutamate residues. This may indicate that the designed strong hydrophobic interactions are sufficient to define the structure and stability of HYP-1.

Urea and GdmCl denaturation of the protein are highly cooperative with m values of 2.1 ± 0.1 and 3.3 ± 0.3 kcal mol^{-1} M^{-1}, respectively.[9] Fitting the data to a two-state model reveals an average free energy of unfolding, ΔG_u, of +13 kcal (mol of dimer)$^{-1}$ (1 M standard state) at 298 K.[9] Since it has been shown that ΔG_u values are similar for thermophilic and mesophilic proteins,[1d] this rather large value of ΔG_u may not be significant; however, the m values, a measure of the cooperativity of melting, indicate a high degree of order in the protein.[5a,10]

The temperature dependence of the far-UV CD signal was measured at several pH values. At pH 7, the protein is approximately 10% denatured at 98 °C.[9] At pH 2 and 4, a cooperative thermal-unfolding transition is observed with a T_m of 78 °C (Figure 2) and 95 °C, respectively. A van't Hoff

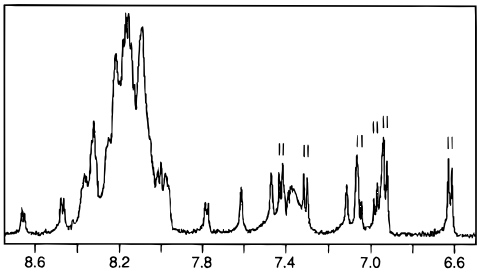

Figure 3. Downfield region of the ^{1}H NMR spectrum of HYP-1 at 22 °C in 10% D_2O, 10 mM sodium phosphate, pH 7. The marked peaks correspond to aromatic side chain protons of Tyr and Trp. The remaining peaks ranging from 7.1 to 8.6 ppm are assigned to the backbone amide protons.

analysis of the data gives a ΔH_m of 41 ± 3 kcal mol^{-1} at pH 2 and 65 ± 9 kcal mol^{-1} at pH 4. The fits are somewhat insensitive to the value of ΔC_p (heat capacity change upon unfolding); however, the overall ΔC_p at pH 2 can be fit to 8 ± 3 cal mol^{-1} K^{-1} res^{-1}, a value typical of native proteins with $T_m \leq 80$ °C.[4d] The ΔC_p at pH 4 could not be determined accurately due to lack of data near the end of the thermal melt. In order to determine the T_m and ΔH_m of the protein at pH 7, differential scanning calorimetry was employed.[11] These experiments revealed a T_m of 116 °C, a ΔH_{cal} of 95 kcal mol^{-1}, and a ΔC_p near 0[12] at a protein concentration of 190 μM.

The amide region of the ^{1}H NMR spectrum of HYP-1 (Figure 3) shows good dispersion of chemical shifts for an α-helical structure with substantial "sequence symmetry". The peaks are sharp and well-resolved, suggesting that the protein is not fluctional on the millisecond time scale.

The fluorescent probe ANS binds to hydrophobic sites that are somewhat solvent exposed and has been used to characterize the structure of synthetic[4b] and native proteins.[13] Rubredoxin isolated from the hyperthermophile *Pyrococcus furiosus* does not bind ANS at pH 7; while at pH 2 and 1 M NaCl, ANS binds weakly (K_d = 37 μM).[1a] Since it has been demonstrated that rubredoxin retains its neutral pH structure even under this extreme pH condition, it is clear that the hydrophobic core of this protein is only slightly more exposed to solvent. This phenomena appears to be duplicated in the present protein. HYP-1 does not bind ANS at pH 7 under typical conditions[9] and weakly binds ANS at pH 2, 1 M NaCl with a K_d of 40 ± 10 μM,[9] a value remarkably similar to that of rubredoxin.

This first generation of synthetic proteins, designed by the computer program CORE to yield a rigid protein with optimal hydrophobic packing efficiency, exhibits characteristic features of natural hyperthermophilic proteins. High thermal stability in HYP-1 appears to originate mainly from hydrophobic stabilization since at low pH the protein is still folded with a T_m of nearly 80 °C. Variants of HYP-1 are currently being designed that will test this hypothesis. In addition to probing the origin of hyperthermophilicity, we are implementing CORE to construct highly stable enzymes since the program can be utilized to design proteins that bind substrates in addition to transition state structures.

Acknowledgment. We thank Hany Farid, Dan Robertson, and Michael Therien for invaluable discussion. We are also grateful to Norma Greenfield and Sarah Hitchcock-DeGregory for access to their CD instrument. X.J. is grateful to the Department of Chemistry for fellowship support.

Supporting Information Available: HPLC chromatogram, data and fits to GdmCl, urea and temperature CD melts, and ANS binding (6 pages). See any current masthead page for ordering and Internet access instructions.

JA963319L

(10) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138−48.

(11) Measurements were performed using a Perkin-Elmer DSC-7 differential scanning calorimeter at heating rates of 5 or 10 °C/min. Samples (60 μL) were heated in hermetically sealed stainless steel containers.

(12) ΔC_p is temperature dependent; its value is maximal at approximately 50 °C and decreases as the temperature increases reaching a value near 0 at around 100 °C (see ref 9b; Makhatadze, G. I.; Kim, K.-S.; Woodward, C.; Privalov, P. L. *Protein Sci.* **1993**, *2*, 2028).

(13) (a) Kuwajima, K. *Proteins: Struct., Funct., Genet.* **1989**, *6*, 87−103. (b) Semisotnov, G. V.; Rodionova, N. A.; Razjulyaev, O. I.; Uversky, V. N. *Biopolymers* **1991**, *31*, 119−28.