Cooperative Thermal Denaturation of Proteins Designed by Binary Patterning of Polar and Nonpolar Amino Acids

Sushmita Roy[‡] and Michael H. Hecht*

Department of Chemistry, Princeton University, Princeton, New Jersey 08544-1009

Received October 6, 1999; Revised Manuscript Received December 22, 1999

ABSTRACT: We previously reported a combinatorial strategy for designing α-helical proteins by assigning only the binary patterning of polar or nonpolar residues [Kamtekar, S., Schiffer, J. M., Xiong, H. Y., Babik, J. M., and Hecht, M. H. (1993) *Science 262*, 1680−1685]. Here we describe the finding that approximately half of the proteins in the original collection display some level of cooperativity in their thermal denaturation profiles. Many are monomeric in solution, demonstrating that the observed cooperativity is not merely a consequence of oligomerization. These findings demonstrate that although the combinatorial nature of the design strategy precludes explicit design of side-chain packing, binary patterning incorporates sufficient sequence information to generate de novo proteins with cooperatively folded structures. As binary partitioning of polar and nonpolar amino acids is an intrinsic part of the genetic code, these findings may bear on the early evolution of native proteins.

What are the minimal features of an amino acid sequence that are required to specify a uniquely folded three-dimensional structure? Attempts to answer this question have relied on four approaches: (i) Comparison of evolutionarily related amino acid sequences (*1*); (ii) mutagenesis of natural proteins (*2*−*7*); (iii) construction of de novo proteins composed of simplified sequences (*8, 9*); and (iv) theoretical studies using simplified models (*10, 11*). All four approaches have led to the realization that proteins are robust: Much of the information in a typical amino acid sequence is not essential, and many different sequences can fold into the same structure.

To exploit the tolerance of protein structures to a wide variety of different sequences, we developed a combinatorial strategy for protein design that aims to produce libraries of de novo sequences compatible with a chosen target structure (*12*). Our strategy was based on the premise that appropriate binary patterning of polar and nonpolar residues can drive a polypeptide chain to fold into amphiphilic elements of secondary structure, which anneal together to form the desired globular structure. Our initial application of the binary code strategy aimed to construct a library of four-helix bundle proteins (*12*). In that earlier work, we designed a sequence pattern of polar and nonpolar residues compatible with the four-helix bundle motif and then constructed a library of synthetic genes to encode this binary pattern. Residues designed to be on the surface of the four-helix bundle were encoded by the degenerate DNA codon NAN, which encodes the polar amino acids Lys, His, Glu, Gln, Asp, and Asn. Residues designed to be buried in the interior of the bundle were encoded by the degenerate codon NTN, which encodes the nonpolar amino acids Met, Leu, Ile, Val, and Phe. (N represents a specified mixture of the DNA nucleotides A, G, T, and C.) Initial characterization of three proteins from the collection demonstrated that they were soluble, α-helical, and moderately stable (*12*).

While it seems reasonable that binary patterning can direct the formation of amphiphilic α-helices that pack together into a bundle, our initial experiments did not enable us to assess "which of our designed proteins possess flexible interiors, and which are nativelike" (*12*). Indeed, since native protein structures typically maintain precisely packed interiors, one might expect that combinatorial strategies, which cannot design specific packing interactions, would not yield cooperatively folded structures.

Our initial attempts to probe this issue led to the characterization of a single protein (M60) from the original binary code collection. Through a series of experiments, we demonstrated that protein M60 possessed nativelike features including (i) chemical shift dispersion in the NMR spectrum; (ii) protection of amide protons from exchange with solvent; (iii) minimal fluorescence in the presence of the hydrophobic dye 1-anilinonaphthalene-8-sulfonate (ANS); (iv) formation of crystals; and (v) cooperative thermal denaturation (*13*). However, M60 was only one protein from the original collection, and its properties may or may not be representative of the binary code collection as a whole. The question remained: Can binary patterning of polar and nonpolar residues incorporate sufficient sequence information to encode cooperatively folded protein structures?

In the current report we demonstrate that M60 is not unique. Over half the proteins in the original collection of Kamtekar et al. (*12*) form cooperatively folded structures.

EXPERIMENTAL PROCEDURES

Cell Growth and Freeze−*Thaw Protocol.* Cultures (1 L) of transformed cells (X90/DE3) were grown at 37 °C in 2× YT medium and 100 μg/mL ampicillin to an OD_{600} between

* Correspondence should be addressed to this author: email hecht@princeton.edu; Fax (609) 258-6746.

‡ Present address: Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037.

10.1021/bi992328e CCC: $19.00
Published on Web 03/29/2000

Table 1: Thermodynamic and Solution Properties of the Binary Code Proteins[a]

protein	T_m (°C)	ΔH (kcal/mol)	oligomeric state
		Cooperative	
I	61	22	M
8	59	41	M
12	59	53	H
15	52	18	M
16	NA	NA	D
17	56	56	M/D/H
51	58	47	M/I
52	62	26	M/I
M60	55	25	M/D/T
63	NA	NA	M
76	52	12	M
83	59	31	M/I
85	59	37	M/I
86	42	13	M
		Noncooperative	
B			M
F			M
G			M
N			M
U			M/D
Y			M/H
Z			D
10			M
13			M/D
30			undetermined
49			M/D
90			M

[a] Kamtekar et al. *(12)* reported 29 sequences. We present data on 26 of these. Three proteins (K, 11, and 24) gave poor yields upon purification; hence their properties are not reported. Sequence M60 is a modification of the original 60, in which a tyrosine has been inserted following the initiator methionine *(30)*. [b] M = monomer; D = dimer; T = tetramer; H = higher-order oligomer; I = impurity or higher-order oligomer. [c] The denaturation profiles for proteins 16 and 63 did not provide good fits to the Gibbs−Helmholtz equation. Hence, the T_m and ΔH for these proteins were not calculated.

0.7 and 1. Cultures were induced with 100 μg/mL IPTG for 3 h and cells were harvested by centrifugation at 6000*g* for 10 min at 4 °C. Protein was released from the harvested cells by a modified version of the freeze−thaw protocol described by Johnson and Hecht (*14*). Harvested cells were subjected to three cycles of freezing (dry ice/ethanol bath) and thawing (10 °C water bath) for 10 min each. Following the third cycle, 10 mL of elution buffer (100 mM $MgCl_2$) was added. Samples were kept on ice for 1 h with gentle inversion to resuspend cells. Samples were then centrifuged at 6000*g* for 10 min at 4 °C and the supernatant containing the desired protein was carefully decanted. These freeze−thaw supernatants were then subjected to either anion-exchange or cation-exchange chromatography.

Anion-Exchange Chromatography. Proteins Y, 8, 12, 15, 16, 17, 51, 52, 63, 83 and 85 were purified by anion exchange chromatography. The freeze−thaw supernatants were exchanged into 50 mM Tris-HCl buffer, pH 8, and concentrated to 4 mL in Centricon 3 concentrators (Amicon). Samples were purified by anion-exchange chromatography on an HQ-POROS 20 column (PerSeptive Biosystems) in 50 mM Tris-HCl, pH 8, with 1.5 M NaCl in the elution buffer.

Acid Precipitation and Cation-Exchange Chromatography. Proteins B, F, G, I, N, U, Z, 10, 13, 30, 49, M60, 76, 86, and 90 were purified by cation-exchange chromatography. A solution of 1 M sodium acetate/acetic acid buffer at pH 4.0 was added to the freeze−thaw supernatants to a final concentration of 50 mM to precipitate the bulk of contaminating *Escherichia coli* proteins. Samples were kept on ice for 15 min and centrifuged at 6000*g* for 15 min at 4 °C. Supernatants were concentrated in Centricon 3 concentrators. After the 20 mL volume was reduced to 4 mL, samples were subjected to cation-exchange chromatography on a HS−POROS column in 50 mM sodium acetate/acetic acid buffer at pH 4.0 with 1.5 M NaCl in the elution buffer. Fractions containing a single band by Coomassie-stained SDS−PAGE were pooled and exchanged into a buffer containing 50 mM sodium phosphate and 200 mM NaCl at pH 7.0.

Circular Dichroism and Thermal Denaturation. CD spectra were measured in a 1 mm cuvette with an Aviv 62 DS spectropolarimeter. α-Helicity was characterized by double minima at 208 and 222 nm. For thermal denaturation experiments, ellipticity was monitored at 222 nm and data points were collected every 0.5 or 1 °C from 0 to 100 °C after 1.5 min equilibration at each temperature.

Size-Exclusion Chromatography. A Superdex 75 gel-filtration column (Pharmacia) was equilibrated with 50 mM sodium phosphate buffer (pH 7.0) and 200 mM NaCl. Purified samples were injected at concentrations identical to those in the thermal denaturation studies. Columns were run at room temperature. Proteins that eluted between aprotonin (MW 6500) and cytochrome b_{562} (MW 12 500) were classified as monomeric. Peaks eluting before cytochrome b_{562} were classified as dimers or higher-order oligomers, depending on the elution time. Absorbance was monitored at 235 nm. In some cases small amounts of impurities with higher molecular weights could not be distinguished from higher-order oligomers. These are indicated in Table 1.

RESULTS

Twenty-six proteins from our original binary code library (*12*) were expressed in *E. coli* and purified. The amino acid sequences are shown in Figure 1. All 26 proteins were well-behaved and water-soluble; none formed inclusion bodies. This observation is in marked contrast to proteins isolated from randomly generated sequences, which frequently form insoluble aggregates (*9*, *15*). Secondary structure was probed by circular dichroism spectroscopy (*16*). Spectra for all 26 proteins displayed the double minima at 208 and 222 nm, characteristic of α-helical secondary structure (data not shown).

Thermal denaturations of the proteins were monitored at physiological conditions (50 mM sodium phosphate and 200 mM NaCl, pH 7.0). Under these conditions, all proteins were readily soluble and the vast majority displayed reversible denaturation. The loss of secondary structure, as indicated by ellipticity at 222 nm, is plotted as a function of temperature in Figure 2. Fourteen proteins display sigmoidal thermal denaturation profiles (Figure 2A,B), while twelve proteins denature noncooperatively (Figure 2C). Thus we observe some level of cooperativity in approximately half of the binary code proteins.

According to the van't Hoff equation $[d(\ln K)/d(T) = \Delta H/RT^2]$, the more cooperative a thermal transition, the larger the ΔH of unfolding. Since $\Delta H_{unfolding}$ measures the disruption of enthalpically favorable interactions in the folded state, a

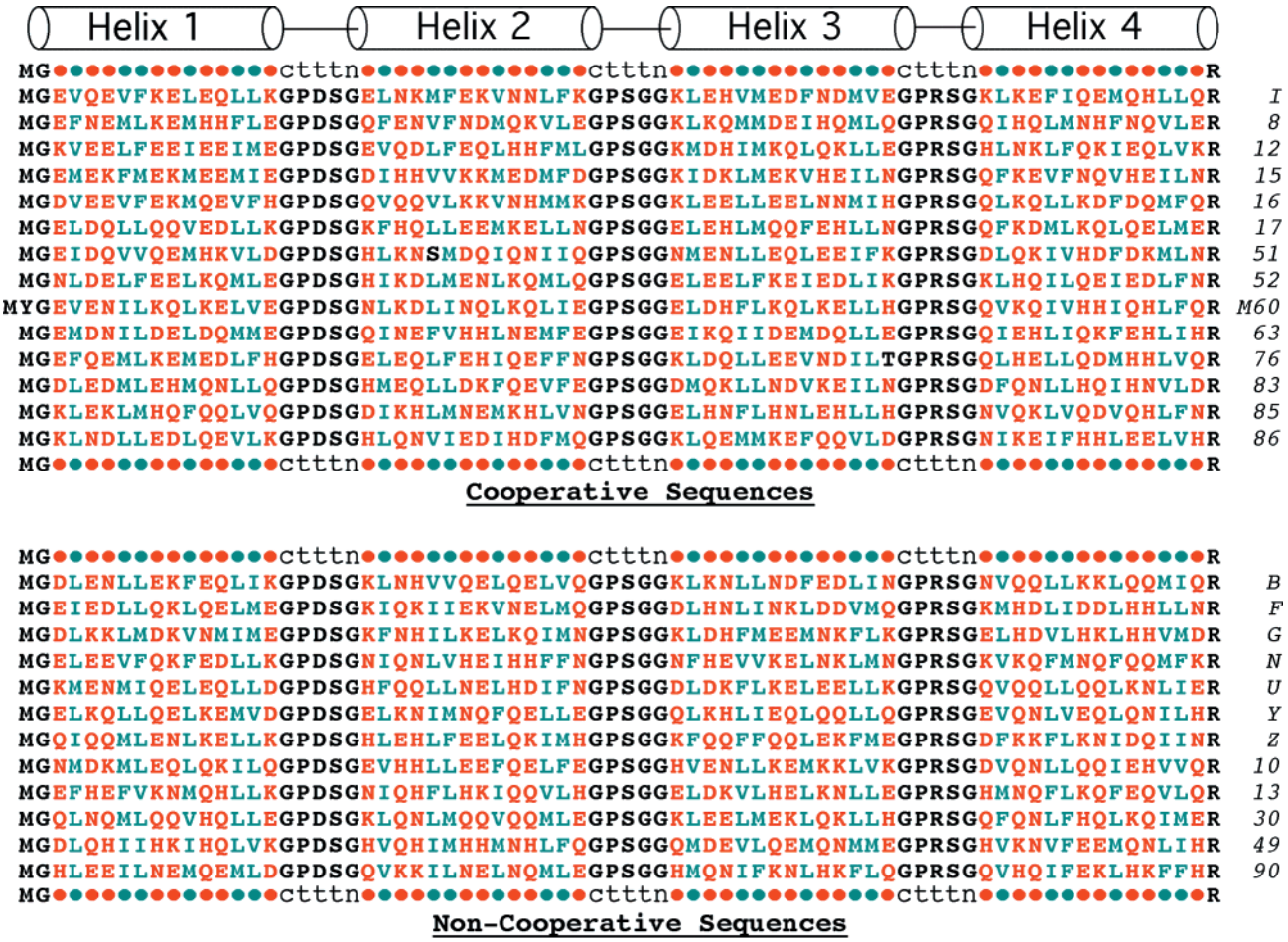

FIGURE 1: Amino acid sequences (*12*) of cooperative and noncooperative proteins. The binary pattern of polar and nonpolar residues (in single-letter code) is indicated with polar residues in red and nonpolar residues in teal. In the first and last lines t, n, and c indicate turn, N-cap, and C-cap residues, respectively.

larger $\Delta H_{unfolding}$ suggests a folded structure stabilized by energetically favorable interactions in the hydrophobic core. Therefore, cooperativity in thermal denaturation profiles is often regarded as an indication of nativelike packing. Indeed the cooperativity of the thermal denaturation is sometimes used as a diagnostic (*17*, *18*) for distinguishing between native structures, which are well packed, and molten globules, which sample an ensemble of related structures (*19*, *20*) and hence are not stabilized by enthalpically favorable interactions in the hydrophobic core. By this criterion, the binary code proteins shown in Figure 2C appear molten globule-like, whereas those shown in Figure 2A,B appear more nativelike.

For the cooperatively folded proteins (Figure 2A,B), $\Delta H_{unfolding}$ was estimated by fitting the thermal denaturation data to a modification of the integrated Gibbs−Helmholtz equation (*21*). The calculated $\Delta H_{unfolding}$ and the midpoint of the transition (T_m) are shown in Table 1. The lowest T_m (42 °C) was observed for protein 86, and the highest (62 °C) for protein 52. A range of enthalpies was observed, with protein 76 having the lowest value of 12 kcal/mol and protein 17 having the highest value of 56 kcal/mol. Proteins 8, 12, 17, and 51 all have values of ΔH above 40 kcal/mol. These values of ΔH are comparable to those reported for small natural proteins such as the B1 and B2 domains of streptococcal protein G, which denature at pH 5.4 with $\Delta H_{unfolding}$ of 62 and 57 kcal/mol, respectively (*22*). Overall, these binary code proteins display enthalpies of unfolding that are comparable and in some cases higher than those reported for several other de novo proteins (*23*−*25*).

While cooperative thermal denaturation curves can be regarded as an indication of nativelike structures (*17*, *18*), similar curves might also be expected for thermally induced disruption of oligomeric structures associated into large cooperative units. Therefore it is important to ascertain whether the binary code proteins form oligomers.

The oligomeric states of the 26 proteins were evaluated either by sedimentation equilibrium ultracentrifugation or by size-exclusion chromatography. Using sedimentation equilibrium, protein F was characterized to be monomeric, protein B was characterized to be in a monomer−dimer equilibrium (*26*), and protein M60 was found to exist in an equilibrium between monomer, dimer, and tetramer (*13*). The remaining 23 proteins were analyzed by gel-filtration chromatography under conditions identical to the thermal denaturation experiments. The concentration of each protein was kept constant for the thermal denaturation and gel-filtration studies, although absolute concentrations varied from protein to protein. Table 1 shows the oligomeric state of each protein. Twelve proteins are fully monomeric (B, F, G, I, N, 8, 10, 15, 63, 76, 86, 90). Another 10 proteins contain mixtures of monomers and dimers (or higher-order oligomers) with the monomeric form accounting for >50% of the sample (U, Y, 13, 17, 49, 51, 52, M60, 83, 85). Only two proteins are

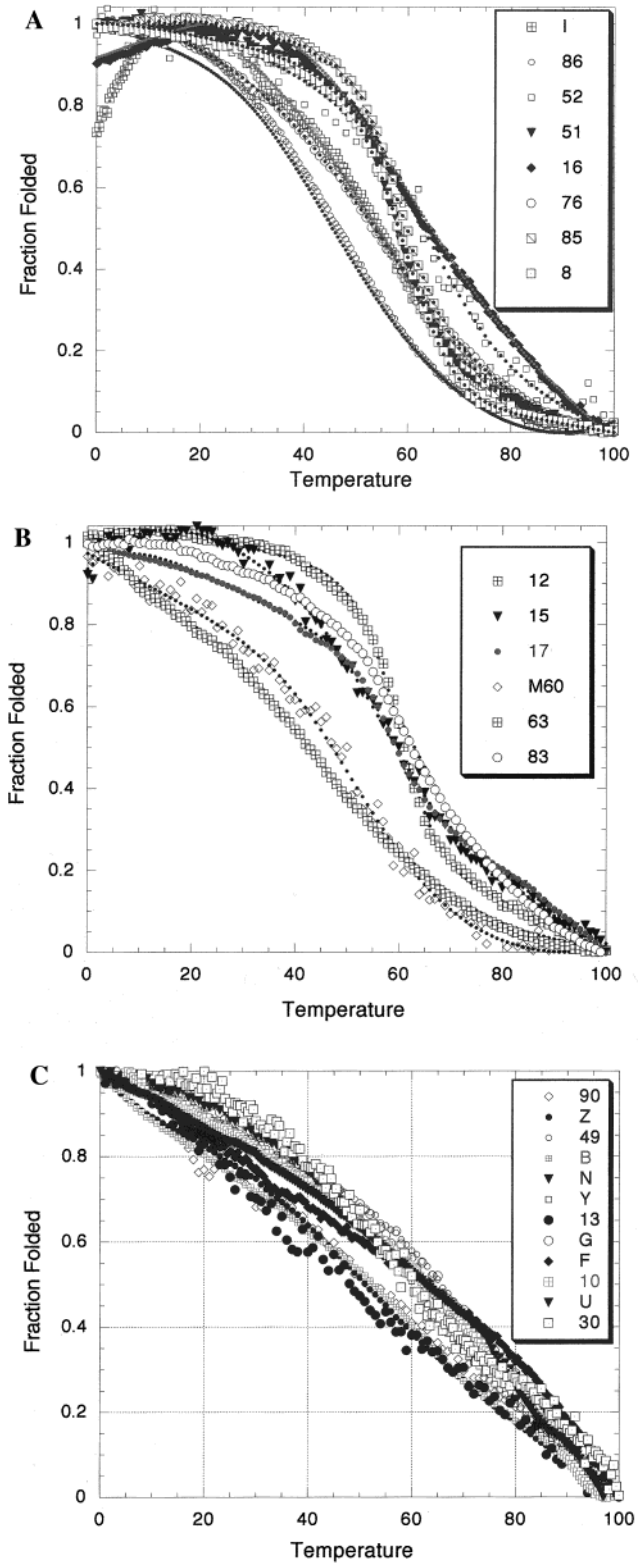

FIGURE 2: (A, B) Fourteen proteins displaying cooperative thermal denaturation profiles. (C) Twelve proteins displaying noncooperative thermal denaturation profiles. Ellipticity at 222 nm was monitored as a function of temperature for proteins dissolved in native buffer (50 mM sodium phosphate and 200 mM NaCl, pH 7.0). Experimental data are shown as symbols. For the cooperative profiles (A and B), the fit to a modified Gibbs−Helmholtz equation (*21*) is shown as points (·).

primarily dimeric (Z and 16), and only one (12) formed mostly higher-order oligomers.

As expected for a combinatorial library of diverse sequences, the data in Table 1 reveal a range of behaviors. Some of the cooperatively folded proteins exist as monomers in equilibrium with dimers or higher-order oligomers. For these proteins, fitting the data to a model for the unfolding of a monomeric protein is an oversimplification, and the observed cooperativity may include contributions from the disruption of a multimeric cooperative unit. However, several of the cooperative proteins (I, 8, 15, 63, 76, 86) are fully monomeric at the concentrations used for thermal denaturation experiments. For these sequences, the observed cooperativity does not require oligomerization and presumably results from favorable packing interactions in these monomeric structures.

DISCUSSION

The binary patterning designed into our initial library of 74-residue α-helical proteins allowed five nonpolar residues at 24 buried positions and six polar residues at 32 exposed positions (the interhelical turns were not combinatorially varied, but were held constant). This patterning is consistent with a hypothetical library containing 5×10^{41} (i.e., $5^{24} \times 6^{32}$) different sequences. In contrast, if an alternative library had been constructed without any elements of rational design, and all 74 residues had been chosen at random, then 2×10^{96} (i.e., 20^{74}) sequences would have been possible. By constraining the library to sequences satisfying the binary code, we reduced the theoretical sequence space by >54 orders of magnitude (*27*). In return for this reduction of diversity, what has been gained?

To address this question it is interesting to compare the properties observed for the 26 proteins from our designed library to the properties expected for sequences chosen entirely at random. De novo sequences isolated from random combinatorial libraries are typically insoluble in native buffers and must be characterized in the presence of urea or guanidine hydrochloride (*9*, *15*, *28*). In contrast, our binary code sequences are expressed as soluble proteins and fold into α-helical structures in native buffers. Moreover, as shown in Figure 2, thermal denaturations performed in native buffer frequently yield cooperative profiles. Thus, a substantial amount of folding information is encoded by polar/nonpolar patterning.

The diverse sequences displayed in Figure 1 display a similarly diverse range of properties. Some are cooperative monomers, while others are cooperative oligomers. For the monomers, it seems reasonable to correlate cooperativity with some degree of nativelike packing. However, for those sequences that exist as mixtures of monomers and dimers, the situation is more complex. For some of the multimeric proteins, cooperativity may be due solely to the disruption of a large cooperative unit into unfolded monomers. In such cases, the thermodynamic analysis used to determine ΔH would be an oversimplification. In other cases, however, the observed cooperativity of a sample containing a mixture of monomers and dimers may indeed reflect the inherent cooperativity of the monomeric unit. For example, detailed characterization of protein M60 (*13*) showed that cooperativity was not diminished by reducing protein concentration to a range that favored monomers. Thus, even for some of the proteins that contain a mixture of monomers and dimers

(Table 1), the observed cooperativity may reflect features of nativelike packing.

Which features in the sequences of these proteins cause some of them, but not others, to denature cooperatively? We have found no obvious differences in sequence composition between the different classes of proteins. For the monomeric proteins in our collection (Table 1) we presume that the cooperativity of the thermal denaturation correlates with the ability of a sequence to form nativelike tertiary interactions. Although this correlation cannot be confirmed in the absence of high-resolution structures, the data presented above suggest that binary code libraries can yield proteins with some level of nativelike interactions.

Can the binary code strategy be extended to applications beyond the folding of α-helical proteins? Might it be possible to generate functional proteins? We have demonstrated previously that among the collection of α-helical proteins, many bind the heme cofactor (*29*), and recently we have shown that several of these de novo heme proteins display peroxidase activity (Moffet et al., manuscript in preparation).

It should be stressed that the cooperatively folded structures described in this communication, the heme binding proteins described by Rojas et al. (*29*), and the peroxidases described by Moffet et al. (manuscript in preparation) were isolated without the use of a genetic selection. Sequences were chosen for characterization merely by their ability to be expressed in *E. coli* at reasonable levels. As most of the binary code sequences (∼60%) express at reasonable levels (*12*), this requirement introduces only a modest bias into the choice of sequences. Therefore, we suggest that the 26 sequences characterized here are not atypical. They probably represent a fairly unbiased sampling of the proteins that can be encoded by the pattern of polar and nonpolar amino acids designed into the initial library. Therefore, we expect that among this vast library of possible sequences most will be α-helical and many will denature cooperatively. Moreover, a significant fraction of this library can be expected to bind heme and catalyze reactions (*29*; Moffet et al., manuscript in preparation). These results indicate that sequences capable of recapitulating key properties of natural proteins are not unusual among binary code libraries. Moreover, since the binary distribution of polar and nonpolar amino acids is inherent in the organization of the genetic code, these results suggest that binary patterning may have played a significant role in the early evolution of natural proteins.

ACKNOWLEDGMENT

We thank Satwik Kamtekar for initiating this project.

REFERENCES

1. Chothia, C., and Lesk, A. M. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 399−405.
2. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A., and Sauer, R. T. (1990) *Science 247*, 1306−1310.
3. Matthews, B. W. (1993) *Annu. Rev. Biochem. 62*, 139−160.
4. Rennell, D., Bouvier, S. E., Hardy, L. W., and Poteete, A. R. (1991) *J. Mol. Biol. 222*, 67−87.
5. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5590−5594.
6. Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., and Regan, L.(1996) *Protein Sci. 5*, 1584−1593.
7. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., and Baker, D. (1997) *Nat. Struct. Biol. 4*, 805−809.
8. Regan, L., and DeGrado, W. F. (1988) *Science 241*, 976−978.
9. Davidson, A. R., and Sauer, R. T. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 2146−2150.
10. Dill, K. A., Bromberg, S., Yue, K., Fieberg, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995) *Protein Sci. 4*, 561−602.
11. Wolynes, P. G. (1997) *Nat. Struct. Biol. 4*, 871−874.
12. Kamtekar, S., Schiffer, J. M., Xiong, H., Babik, J. M., and Hecht, M. H. (1993) *Science 262*, 1680−1685.
13. Roy, S., Ratnaswamy, G., Boice, J. A., Fairman, R., McLendon, G., and Hecht, M. H. (1997) *J. Am. Chem. Soc. 119*, 5302−5306.
14. Johnson, B. H., and Hecht, M. H. (1994) *Bio/Technology 12*, 1357−1360.
15. Mandecki, W. (1990) *Protein Eng 3*, 221−226.
16. Greenfield, N., and Fasman, G. D. (1969) *Biochemistry 8*, 4108−4116.
17. Betz, S. F., Raleigh, D. P., and DeGrado, W. F. (1993) *Curr. Opin. Struct. Biol. 3*, 601−610.
18. Betz, S. F., Bryson, J. W., and DeGrado, W. F. (1995) *Curr. Opin. Struct. Biol. 5*, 457−463.
19. Ohgushi, M., and Wada, A. (1983) *FEBS Lett. 164*, 21−24.
20. Brazhnikov, E. V., Chirgadze, Y. N., Dolgikh, D. A., and Ptitsyn, O. B. (1985) *Biopolymers 24*, 1899−1907.
21. Cohen, D. S., and Pielak, G. J. (1994) *Protein Sci. 3*, 1253−1260.
22. Alexander. P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry 31*, 3597−3603.
23. Dalal, S., Balasubramanian, S., and Regan, L. (1997) *Nat. Struct. Biol. 4*, 548−552.
24. Raleigh, D. P., Betz, S. F., and DeGrado, W. F. (1995) *J. Am. Chem. Soc. 117*, 7558−7559.
25. Dahiyat, B. I., and Mayo, S. L. (1997) *Science 278*, 82−87.
26. Kamtekar, S. (1995) Ph.D. Thesis, Department of Chemistry, Princeton University, Princeton, NJ.
27. Beasley, J. R., and Hecht, M. H. (1997) *J. Biol. Chem. 272*, 2031−2034.
28. Davidson, A. R., Lumb, K. J., and Sauer, R. T. (1995) *Nat. Struct. Biol. 2*, 856−864.
29. Rojas, N. R., Kamtekar, S., Simons, C. T., McLean, J. E., Vogel, K. M., Spiro, T. G., Farid, R. S., and Hecht, M. H. (1997) *Protein Sci. 6*, 2512−2524.
30. Xiong, H. (1995) Ph.D. Dissertation, Department of Chemistry, Princeton University, Princeton, NJ.

BI992328E