A Virtual Library Approach To Investigate Protein Folding and Internal Packing

Martin A. Case and George L. McLendon*

Department of Chemistry, Princeton University
Princeton, New Jersey 08544

Received February 29, 2000

It is now possible to design synthetic proteins from first principles with predictable structures,[1] and even in certain cases with predictable function.[2] A well-packed protein interior is clearly critical to the stability of proteins[3] and also to issues of enzyme catalysis such as substrate recognition and activation. However, unequivocal determination of amino acid sequences which confer *optimal* packing of protein interiors remains elusive, because an optimal sequence can only be so judged in comparison with other possible sequences. In this respect, computational methods are proving valuable, but ultimately must be vetted against experimental data. The only reasonable empirical approach requires massively combinatorial construction of all possible sequence variants from which optimally packed candidates may be selected and characterized.[4]

In this communication we outline a novel strategy for exploring protein packing based on a *virtual* combinatorial library. Lehn[5] has contrasted *synthesis-directed* libraries with virtual, or dynamic libraries.[6] Directed libraries consist of objects created by synthetic combinatorial chemistry, in which individual members differ by their covalent connectivities. By contrast, virtual libraries are created by self-assembly processes and the resulting molecules differ by *non*-covalent interactions and may interconvert via facile equilibria. For a synthesis-directed library to be useful, screening methods must be developed to isolate a desired sub-population, whereas a virtual library might self-screen by spontaneous equilibration of the constituent members.

The library of interest encodes proteins of a parallel three-helix bundle motif,[7] created by the metal-templated assembly of proto-helical subunits.[8] Previous work suggested that, in this context, protein stability and metal-ion binding are thermodynamically coupled.[9] Three peptide subunits were synthesized and allowed to assemble into a small virtual library of 11 exchange-labile three-helix bundles directed by complexation with iron(II).[10] Figure 1 is a representation of an ideal α-helix onto which the amino acid sequences have been mapped.

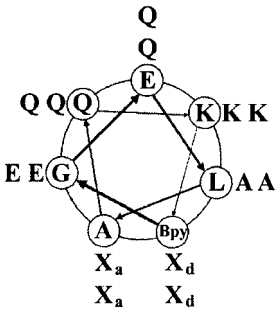

Figure 1. Helix wheel diagram of αpX peptides. Sequences used are the following: bpy-G-E-L-A-Q-K-**A**-E-Q-A-**A**-Q-K-**A**-E-Q-A-**A**-Q-K-NH_2 (αpA), bpy-G-E-L-A-Q-K-**A**-E-Q-A-**L**-Q-K-**A**-E-Q-A-**L**-Q-K-NH_2 (αpLA), and bpy-G-E-L-A-Q-K-**L**-E-Q-A-**L**-Q-K-**L**-E-Q-A-**L**-Q-K-NH_2 (αpL), where bpy is 2,2′-bipyridyl-5-carboxylic amide appended at the N-terminus for metal-ion binding. Variable residues are shown in bold. The helix is viewed looking along the helical axis from the N (metal-binding) to the C-terminus. αpA, X = ala; αpLA, X_a = leu, X_d = ala; αpL, X = leu.

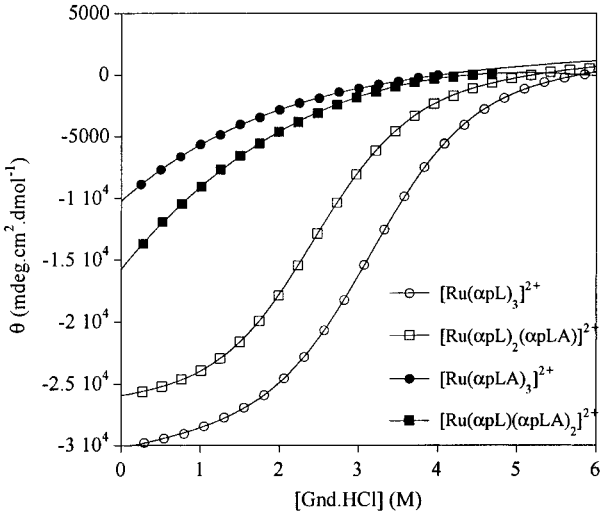

Figure 2. Guanidinium chloride unfolding of $[Ru(\alpha pX)_3]^{2+}$ three-helix bundles. Elipticities were recorded at 222 nm and the data fit using a two-state unfolding model to give $\Delta G_u°$ values as follows: $[Ru(\alpha pL)_3]^{2+}$ 3.0 kcal M^{-1}; $[Ru(\alpha pL)_2(\alpha pLA)]^{2+}$ 2.4 kcal M^{-1}; $[Ru(\alpha pLA)_3]^{2+}$ 0.2 kcal M^{-1}; $[Ru(\alpha pL)(\alpha pLA)_2]^{2+}$ 0.2 kcal M^{-1}. All values are ±0.1 kcal M^{-1}.

All eleven products were independently synthesized as ruthenium(II) complexes, which do not undergo facile equilibration.[11] Unfolding free energies ($\Delta G_u°$) of each of these metalloproteins were measured by chemical denaturation. The results are shown in Figure 2.

In the case of an ensemble of exchange-labile species, two limiting cases may be considered. Assembly could proceed as in a synthesis-directed library wherein all products are represented statistically. Alternatively, if subunit packing and metal-ion binding are thermodynamically coupled, equilibrium self-assembly will favor the most stable structures and the ensemble will function as a virtual library which self-screens for stability. To ascertain which case pertains, the *equilibrium* product distribution was monitored in two cases: 1/9 total peptide equivalent of labile

(1) (a) Degrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F.; Lombardi, A. *Annu. Rev. Biochemistry* **1999**, *68*, 779−819.

(2) (a) Mutz, M. W.; Case, M. A.; Wishart, J. F.; Ghadiri, M. R.; McLendon, G. L. *J. Am. Chem. Soc.* **1999**, *121*, 858−859. (b) Tommos, C.; Skalicky, J. J.; Pilloud, D. L.; Wand, A. J.; Dutton, P. L. *Biochemistry* **1999**, *38*, 9495−9507. (c) Shifman, J. M.; Moser, C. C.; Kalsbeck, W. A.; Bocian, D. F.; Dutton, P. L. *Biochemistry* **1998**, *37*, 16815−16827.

(3) (a) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401−1407. (b) Wagschal, K.; Tripet, B.; Lavigne, P.; Mant, C.; Hodges, R. S. *Protein Sci.* **1999**, *8*, 2312−2329.

(4) Pelletier, J. N.; Arndt, K. M.; Pluckthun, A.; Michnick, S. W. *Nature Biotech.* **1999**, *17*, 683−690.

(5) Huc, I.; Lehn, J.-M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2106−2110.

(6) Ganesan, A. *Angew. Chem., Int. Ed.* **1998**, *37*, 2828−2831.

(7) (a) Nautiyal, S.; Alber, T. *Protein Sci.* **1999**, *8*, 84−90. (b) Eckert, D. M.; Malashkevich, V. N.; Kim, P. S. *J. Mol. Biol.* **1998**, *284*, 859−865. (c) Ogihara, N. L.; Weiss, M. S.; Degrado, W. F.; Eisenberg, D. *Protein Sci.* **1997**, *6*, 80−88. (d) Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *371*, 80−83.

(8) (a) Ghadiri, M. R.; Soares, C.; Choi, C. *J. Am. Chem. Soc.* **1992**, *114*, 825−831. (b) Lieberman, M.; Tabet, M.; Sasaki, T. *J. Am. Chem. Soc.* **1994**, *116*, 5035−5044.

(9) Case, M. A.; Ghadiri, M. R.; Mutz, M. W.; McLendon, G. L. *Chirality* **1998**, *10*, 35−40.

(10) Considering only the *fac* Λ diastereoisomers, the eleventh species arises because of the topological nonequivalence of ABC and ACB systems.

(11) (a) Linton, B.; Hamilton, A. D. *Chem. Rev.* **1997**, *97*, 1669−1680. (b) Ghadiri, M. R.; Case, M. A. *Angew. Chem., Int. Ed.* **1993**, *32*, 1594−1597.

(12) (a) Duus, J. Ø.; Meldal, M.; Winkler, J. R. *J. Phys. Chem. B* **1998**, *102*, 6413−6418. (b) Förster, T. *Ann. Phys. (Leipzig)* **1948**, 2, 55−75.

10.1021/ja000722q CCC: $19.00
Published on Web 08/03/2000

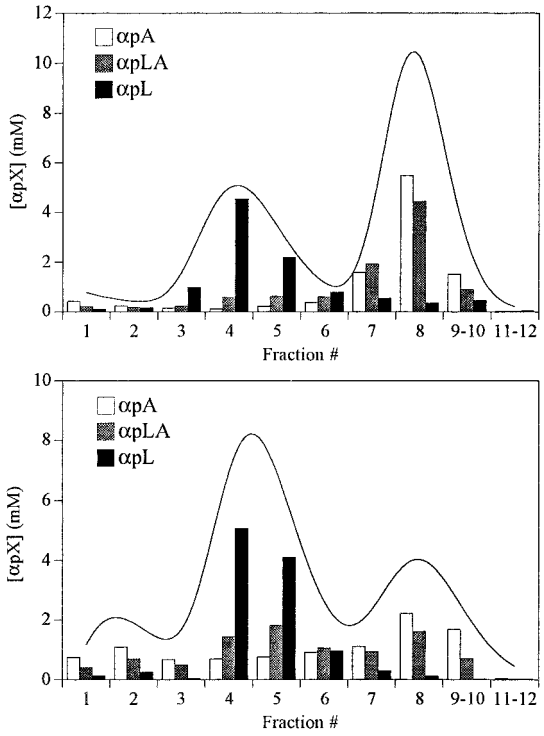

Figure 3. Size exclusion chromatograms and fraction composition for the equilibrium mixture $[Fe(\alpha pX)_3]^{2+}$. Top: 1/9 equivalent Fe^{2+}. Bottom; 2/9 equivalent Fe^{2+}.

metal ion (Fe^{2+}) and 2/9 total peptide equivalent. In both cases, three-helix bundles were separated from unreacted peptide by size-exclusion chromatography (SEC). The peptide composition of the bundles was then determined by HPLC analysis. The resulting peptide distributions are shown in Figure 3.

For a simple two-component system in which there is coupling between folding free energy and equilibrium metal binding, the predicted stability ratio of two products, P_1 and P_2, in an equimolar mixture follows the relationship:

$$-RT\ln(P_1/P_2) = c[\Delta G_u^\circ(P_1) - \Delta G_u^\circ(P_2)] \qquad (1)$$

where c is a measure of the coupling efficiency. If all the free energy of folding were translated into differential metal binding energy, then $c = 1$, if none, then $c = 0$.

The denaturation curves presented in Figure 2 clearly show that the exchange-inert metalloproteins exhibit a wide range of stabilities as a consequence of the relative hydrophobicities and/or efficiency of packing of their hydrophobic cores. Not unexpectedly, the more leucine-rich peptides combine to form bundles of higher stability, and these stabilities scale qualitatively with the amount of hydrophobic surface buried in the core. None of the ruthenium three-helix bundles containing αpA gives a measurable ΔG_u°. From the free energies of unfolding, the theoretical distribution of an equilibrium population of exchange-labile three-helix bundles containing peptides αpL and αpLA may be calculated. From this predicted equilibrium distribution, a theoretical *peptide* distribution among three-helix bundles at equilibrium was calculated. These data are presented in Table 1, along with the observed peptide distribution for the Fe^{2+} exchange-labile three-helix metalloprotein system.

The most striking feature of these experiments is the degree to which this simple protocol is capable of discriminating in favor of peptides with leucine-rich hydrophobic core patterns. This result is fully consistent with previous studies. Indeed, the key observation is that the virtual library preferentially returns the protein of optimal stability from the available possibilities. There is almost complete discrimination, in the case of limiting Fe^{2+}, between αpL and αpA peptides. The former are present almost exclusively in the three-helix structures which have earlier SEC retention times, and the latter are partitioned into a later band consisting of uncomplexed peptides. The chimeric αpLA peptide is strongly selected against in three-helix bundle formation but is nevertheless present in measurable amounts, as predicted from the stability data presented in Figure 2.

Table 1. Peptide Distributions in Fe(II) Three-Helix Bundles

	obsd distribution					
	fraction				predicted distribution	
peptide	3	4	5	total	$c = 1.00$	$c = 0.62$
αpL	0.985	4.543	2.201	7.729	0.91	0.84
αpLA	0.226	0.583	0.623	1.432	0.09	0.15
αpA	0.146	0.111	0.216	0.473		
αpL/αpLA				5.4	10.1	5.4

Equation 1 presents a model in which the degree of coupling between the free energy of metal binding and the free energy of protein folding may be explicitly determined from knowledge of the equilibrium distribution of species and their relative free energies of unfolding. The data in Table 1 yield a value of $c = 0.62$. While this value clearly indicates that metal binding energy is coupled to peptide folding, multiple uncertainties preclude a more detailed analysis.

To verify that the three-peptide system was indeed reaching equilibrium in the presence of Fe^{2+}, the kinetic course of the reaction between αpL, which forms the most stable three-helix bundle, and a fluorescent analogue, αpL-dsl, was followed in a pseudo-self-exchange experiment. The C-terminal alanine residue of αpL was replaced by an ornithine residue and a dansyl group appended to the side chain N^δ moiety. The dansylated peptide has a fluorescence emission maximum of 550 nm. The Fe^{2+} tris-bipyridyl moiety of the three-helix bundle has an absorbance maximum at 544 nm and quenches the fluorescence of the dansyl group by intramolecular energy transfer.[11] Introduction of αpL displaces αpL-dsl peptides from a preassembled $[Fe(\alpha pL\text{-}dsl)_3]^{2+}$ complex. The reaction proceeds until equilibrium, when the fluorescence signal becomes asymptotic. If sufficient time is allowed for this self-exchange process to equilibrate, all equilibria present in the 11-member virtual library are necessarily accommodated. At a concentration of 10 mM the half-life for this exchange is of the order of 5 min. Thirty minutes incubation was therefore deemed sufficient.

In summary, this work presents a proof of concept for the construction of exchange-labile, self-screening, virtual combinatorial libraries from which may be selected self-assembled proteins with optimal stability. Optimal packing in the parallel three-helix motif can now be explored using very large (2×10^8 member) self-screening libraries. Such studies are in progress.

Acknowledgment. This work was supported by grant CHE-9729124 from the NSF Division of Chemistry.

Supporting Information Available: Experimental details, syntheses, and characterization of the ruthenium complexes and fluorescence kinetics (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA000722Q