J|A|C|S COMMUNICATIONS

Published on Web 06/10/2010

Protein Core Packing by Dynamic Combinatorial Chemistry

Liton Roy and Martin A. Case*

Department of Chemistry, The University of Vermont, Burlington, Vermont 05405

Received April 8, 2010; E-mail: mcase@uvm.edu

Control of protein stability is of critical importance for applications in biotechnology and medicine. Successful approaches to optimizing folding stability have included optimization of native packing interactions by rational design[1] and selection-based strategies.[2] Selection-based strategies employing random gene libraries can provide formidable numbers of sequence variants, but they offer no means of critically assessing whether the best of all possible sequences has been returned or of describing the distribution of stabilities across the library of sequence variants. Here we describe an ex vivo method for unambiguously identifying the optimal amino acid sequences of the most stable members of a library of 8436 proteins using a selection protocol based only on thermodynamic folding stability.

The key concept is the use of a dynamic combinatorial library (DCL) of proteins assembled from combinations of members of a more modest library of peptide subunits. The composition of the DCL is determined by the thermodynamic stability of each of the library members, which makes DCLs ideal for discovering optimally stable systems.[3] We constructed a small combinatorial library of 36 peptides (six hydrophobic amino acid substitutions at two positions) and allowed it to assemble under thermodynamic control to generate a much larger DCL of 8436 trimeric proteins. The DCL sampled all of the possible combinations of the constituent peptide library, and the challenge was to capture the optimally stable members. Capture was accomplished by coupling DCL stability to a second, cooperative stabilizing effect that was introduced substoichiometrically, augmenting the stability of the most stable fraction of the DCL and enabling separation. We have previously shown that this can be accomplished by appending metal-binding ligands to the peptide subunits and using substoichiometric metal ion to increase the stability of the optimal members of the DCL by formation of an exchange-labile complex.[4] The stable trimeric DCL members can then be separated from the rest of the library by size-exclusion or ion-exchange chromatography. Multiple iterations of the selection process were effected by recovery of the selected peptides followed by further substoichiometric metal-ion-assisted selection. In contrast to combinatorial genetic methods, the dynamic combinatorial method necessarily explores the relative stabilities of all of the library members and returns all of the optimal sequences.

We optimized six core residues of a parallel three-helix polypeptide bundle in which the oligomerization state and the interhelical register are controlled by covalently appended 2,2-bipyridyl ligands. Addition of hexacoordinate Fe^{2+} ion results in obligate formation of a trimeric assembly that undergoes hydrophobic collapse with concomitant induction of α-helicity.[5] The peptide sequences are shown in Figure 1. The peptides were constructed using three repeats of an *abcdefg* heptad in which positions *a* and *d* are hydrophobic amino acids. Upon induction of the α-helical secondary structure, this patterning leads to sequestration of the *a* and *d* positions on the interior. The identities and optimal placement of solvent-exposed residues have been described elsewhere.[6] The

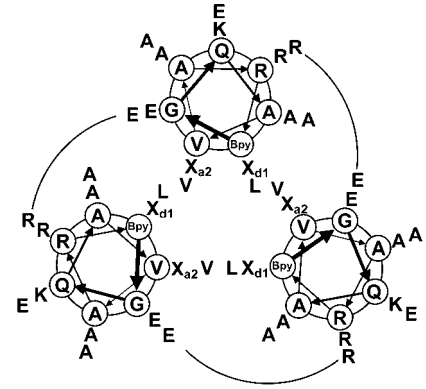

Figure 1. Peptide sequences used in this study. X indicates randomized hydrophobic amino acids. k(NBD) represents nitrobenzofurazan at N_ε of D-lysine. Helices are viewed from the N-termini, and arcs indicate potential salt-bridging interactions.

hydrophobic residues at d_1 and a_2 were varied over the six amino acids glycine, alanine, valine, leucine, isoleucine, and phenylalanine, resulting in a library of 36 peptides. The peptides are named according to the amino acids at the randomized positions. For example, the iron(II) homotrimer of the sequence with leucine at d_1 and valine at a_2 is designated Fe:$\mathbf{LV}_3$.

We first established the exchange kinetics for the DCL. The metallopeptide exchange rate was measured for a fluorescently labeled Fe:$\mathbf{LV\text{-}NBD}_3$ homotrimer[7] in the presence of the **LV** peptide. Formation of the iron(II) complex quenches the fluorescence of the **LV-NBD** peptide, and the fluorescence recovers upon displacement of **LV-NBD** by unlabeled **LV** peptides. Exchange was followed by fluorescence as shown in Figure 2 and gave a $t_{1/2}$ of 22 h at a total peptide concentration of 80 μM. In a second experiment, exchange of an isotopically labeled Fe:$\mathbf{LI}(16^2H)_3$ homotrimer with the **LI** peptide was followed by electrospray ionization mass spectrometry (ESI-MS) and gave a $t_{1/2}$ of 28 min at a total peptide concentration of 330 μM (see the Supporting Information for details). The DCL was allowed to equilibrate for 24 h at a total peptide concentration of 5 mM.

The 36-peptide library was constructed on polystyrene resin beads using standard pool-and-split techniques.[8] The mass degeneracy between leucine and isoleucine was eliminated by using *N*-Fmoc($^{13}C_1$)leucine. The HPLC chromatogram of the crude peptide library is shown in Figure 3, together with peak assignments derived from tandem mass spectrometry. The DCL was created by adding 3.3 mol % iron(II) to select the optimal 10 mol % of the DCL as iron(II) trimers. After 24 h of equilibration, the metallotrimer pool was separated from the uncomplexed peptides by size-exclusion chromatography. Metal ions were removed from the selected peptide complexes by addition of excess EDTA, and the

10.1021/ja1029717

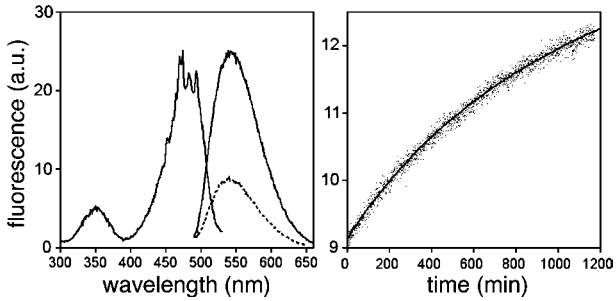

Figure 2. (left) Excitation and emission spectra of **LV-NBD** (solid lines, 40 μM) and Fe:**LV-NBD$_3$** (dashed line, 13.3 μM). (right) Time course for displacement of **LV-NBD** from Fe:**LV-NBD$_3$** after addition of 40 μM **LV** (λ_{ex} = 470 nm, λ_{em} = 550 nm).

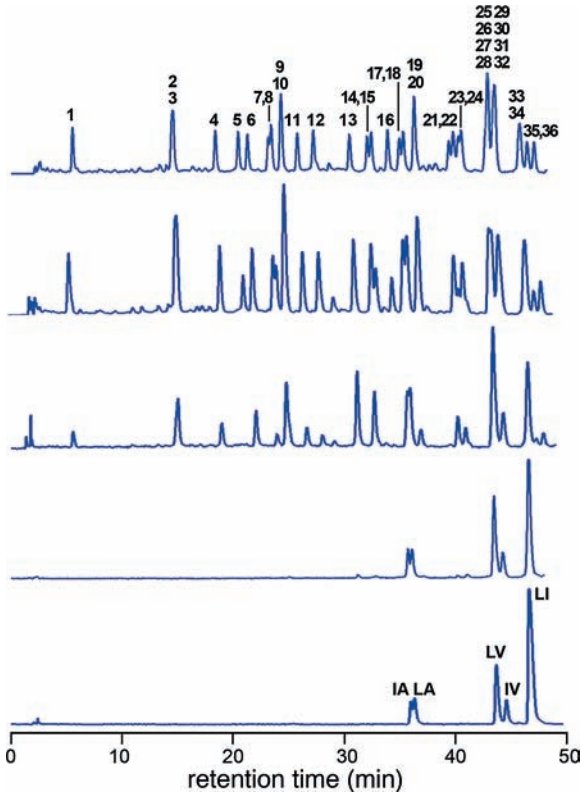

Figure 3. Evolution of the peptide distribution in the DCL from the initial peptide distribution (top) through four rounds of selection (bottom). Key: 1 **GG**; 2,3 **AG**,**GA**; 4 **VG**; 5 **FG**; 6 **GV**; 7 **IG**; 8 **AA**; 9,10 **GF**,**LG**; 11 **GI**; 12 **GL**; 13 **VA**; 14 **AV**; 15 **FA**; 16 **AF**; 17 **IA**; 18 **LA**; 19,20 **AL**,**AI**; 21 **VV**; 22 **VF**; 23 **FV**; 24 **FF**; 25,26,27,28 **LV**,**VL**,**LF**,**VI**; 29,30,31,32 **IV**,**FL**,**FI**,**IF**; 33,34 **LI**,**LL**; 35 **IL**; 36 **II**.

composition of the selected apopeptides was assessed using LC-MS^2. Addition of 3.3 mol % iron(II) to the selected apopeptides gave a second round of enrichment of the DCL. Four rounds of selection were necessary to approach convergence. HPLC traces showing the peptide distribution over four rounds of selection are presented in Figure 3. To ensure that the selection was not biased by apopeptide aggregates, the selection process was repeated in 2 M urea solution with identical results. The first round of selection did not appear to have a significant effect on the peptide distribution beyond the suppression of sequences containing phenylalanine. After the second round, phenylalanine-containing sequences had been lost and glycine-containing sequences suppressed. The third round of selection reduced the library to five dominant peptide sequences: **LV**, **LA**, **LI**, **IA**, and **IV**. These peptides can combine to form up to 35 different iron(II) trimers. We prepared a set of

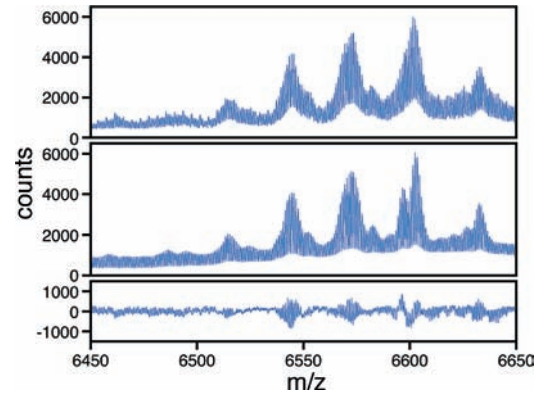

Figure 4. (top) ESI mass spectrum of the isotopically labeled resynthesized DCL after four rounds of selection, (middle) the fitted trimer distribution, and (bottom) the residuals from the fit.

Table 1. Observed and Calculated Average Ellipticities and Folding Free Energies for Iron(II) Trimers Returned after Four Rounds of Selection

	$[\Theta]_{222}$ (deg cm^2 dmol^{-1})		
trimer	observed	calcd avg	ΔG^2_{fold} (esd) (kcal mol^{-1})
Fe:**LI**$_3$	−15800		−1.73 (7)
Fe:**LV**$_3$	−16300		−1.50 (5)
Fe:**IV**$_3$	−10000		−1.4 (1)
Fe:**IA**$_3$	−2200		−0.66 (5)[a]
Fe:**LA**$_3$	−4300		+0.23 (7)[a]
Fe:**LI**$_2$:**IA**	−15900	−11300	−1.81 (8)
Fe:**LV**:**LI**$_2$	−16300	−16000	−1.68 (8)
Fe:**LV**$_2$:**LI**	−15900	−16100	−1.63 (8)
Fe:**LV**$_2$:**IV**	−15000	−14200	−1.67 (7)
Fe:**LI**$_2$:**LA**	−14700	−12300	−1.36 (8)
Fe:**LV**$_2$:**IA**:	−16000	−11600	−1.32 (7)
Fe:**LI**$_2$:**IV**	−15900	−13900	−1.24 (8)
Fe:**LV**:**LI**:**IA**	−16000	−11400	−1.8 (1)
Fe:**LV**:**LI**:**IV**	−15900	−14000	−1.7 (1)
Fe:**LI**:**IV**:**LA**	−12500	−10200	−1.7 (1)
Fe:**LV**:**LI**:**LA**	−14000	−12100	−1.6 (1)
Fe:**LV**:**IV**:**IA**	−13700	−9500	−1.5 (1)

[a] Not observed by ESI-MS.

isotopically modified **LV**, **LA**, **LI**, **IA**, and **IV** peptides that would give a unique mass for each of the 35 possible trimers (see the Supporting Information for details). The five analogues were mixed in the relative proportions returned from the final selection, and their Fe^{2+} complexes were allowed to equilibrate for 24 h. ESI-MS (Figure 4) revealed that only 15 of the 35 possible combinations of metallotrimers were formed.[9] The observed trimer sequences are listed in Table 1, and several features of the selected set are worth noting. The d_1 position is predominantly occupied by leucine (33 of 45 d_1 residues were leucine, 12 were isoleucine). The a_2 position is more tolerant of variation: the observed alanine/isoleucine/valine ratio was 7:16:22. The statistical consensus homotrimer was thus Fe:**LV**$_3$, a repeating V_aL_d pattern that has been previously reported for parallel homotrimeric α-helices.[10] The stable set can be constructed by allowing dimers of **LV** and **LI** to partner with any of **LV**, **LI**, **LA**, **IV**, or **IA** as the third helix. This set can be described as {(**LV** ∨ **LI**) & (**LV** ∨ **LI**) & (**LV** ∨ **LI** ∨ **IA** ∨ **LA** ∨ **IV**)}. One member of this set, Fe:**LV**$_2$:**LA**, was absent. Two additional heterotrimers, Fe:**LV**:**IV**:**IA** and Fe:**LI**:**IV**:**IA**, were also present. All of the selected trimers are highly α-helical according to their circular dichroism (CD) spectra (Figure 5), and the positive feature at 315 nm implies diastereoselection of the Λ geometry at the metal center, presumably induced by a left-handed supercoil.[11] The stabilities of selected trimers were measured by

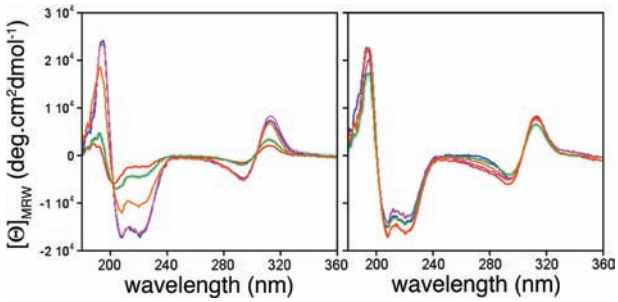

Figure 5. CD spectra of selected metallotrimers: (left) homotrimers Fe:**IA**$_3$ (red), Fe:**IV**$_3$ (orange), Fe:**LA**$_3$ (green), Fe:**LI**$_3$ (blue), Fe:**LV**$_3$ (magenta); (right) 1:1:1 heterotrimers Fe:**LV**:**LI**:**IA** (red), Fe:**LV**:**LI**:**IV** (orange), Fe:**LV**:**LI**:**LA** (green), Fe:**LV**:**IV**:**IA** (blue), Fe:**LI**:**IV**:**IA** (magenta). The CD spectra of the 2:1 heterotrimers are essentially superimposable and have been omitted for clarity.

urea denaturation monitored by CD (details of all fitting algorithms are provided in the Supporting Information). Fe:**LI**$_3$ is the most stable homotrimer ($\Delta G^{\circ}_{\text{fold}} = -1.73$ kcal mol^{-1}). The statistical consensus homotrimer Fe:**LV**$_3$ has modest folding stability, appearing roughly midway in the stability rankings. The homotrimers of **IA** and **LA** have negligible folding stabilities. The unfolding curves for the heterotrimers are weighted sums of the unfolding curves of an equilibrium population of homo- and heterotrimers. For 2:1 heterotrimers, the stabilities of the individual species were extracted from the ensemble unfolding curves by fitting the two known homotrimeric components and two unknown heterotrimeric species to the observed unfolding curve. Fe:**LI**$_2$:**IA** was revealed to be the most stable 2:1 heterotrimer, with a folding free energy of -1.8 kcal mol^{-1}. The folding free energies of the 2:1 heterotrimers were used in a bootstrap calculation of the folding free energies of the 1:1:1 heterotrimers. Fe:**LV**:**LI**:**IA** was found to be the most stable 1:1:1 heterotrimer, with a folding free energy of -1.8 kcal mol^{-1}. The calculated folding free energies for the homotrimers and heterotrimers are presented in Table 1.

In summary, we have used a dynamic combinatorial library to exhaustively probe the folding stabilities of over 8000 homologous protein sequences. Within this library there is no single sequence that confers optimal stability. The set of optimally stable sequences contains 15 of the 8436 library sequences (0.18%). Alanine is present in the hydrophobic cores of the two most stable heterotrimers, suggesting that the small side chain is accommodated by "jigsaw" packing with bulkier residues and that stability is not conferred simply according to buried hydrophobic surface area.

Acknowledgment. The authors are grateful to Bruce O'Rourke at UVM and Haiteng Deng at Rockefeller University for acquisition of mass spectra. The authors thank the National Science Foundation (CHE 0106342) and VT EPSCoR (EPS 0236976) for financial support and the National Science Foundation for funding the acquisition of an ABI 4000 QTrap mass spectrometer (CHE 0821501).

Supporting Information Available: Full experimental details, CD spectra, urea melts, and data fitting procedures. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Park, S.; Yang, X.; Saven, J. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 487–494. (b) Dantas, G.; Kuhlman, B.; Callender, D.; Wong, M.; Baker, D. *J. Mol. Biol.* **2003**, *332*, 449–460. (c) Mendes, J.; Guerois, R.; Serrano, L. *Curr. Opin. Struct. Biol.* **2002**, *12*, 441–446. (d) Street, A. G.; Mayo, S. L. *Struct. Folding Des.* **1999**, *7*, R105–R109. (e) DeGrado, W. F.; Summa, C. M.; Pavone, V.; Nastri, F. *Annu. Rev. Biochem.* **1999**, *68*, 779–819.

(2) (a) Chockalingam, K.; Lu, H. D.; Banta, S. *Anal. Biochem.* **2009**, *388*, 122–127. (b) Jumawid, M. T.; Takahashi, T.; Yamazaki, T.; Ashigai, H.; Mihara, H. *Protein Sci.* **2009**, *18*, 384–398. (c) Graziano, J. J.; Liu, W. S.; Perera, R.; Geierstanger, B. H.; Lesley, S. A.; Schultz, P. G. *J. Am. Chem. Soc.* **2008**, *130*, 176–185. (d) Wunderlich, M.; Schmid, F. X. *J. Mol. Biol.* **2006**, *363*, 545–557. (e) Campbell-Valois, F. X.; Tarassov, K.; Michnick, S. W. *J. Mol. Biol.* **2006**, *362*, 151–171. (f) Benitez-Cardoza, C. G.; Stott, K.; Hirshberg, M.; Went, H. M.; Woolfson, D. N.; Jackson, S. E. *Biochemistry* **2004**, *43*, 5195–5203. (g) Feng, H. Q.; Bai, Y. W. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 426–429. (h) Ventura, S.; Serrano, L. *Proteins: Struct., Funct., Bioinf.* **2004**, *56*, 1–10. (i) Magliery, T. J.; Regan, L. *Eur. J. Biochem.* **2004**, *271*, 1595–1608. (j) Matsuura, T.; Ernst, A.; Zechel, D. L.; Pluckthun, A. *ChemBioChem* **2004**, *5*, 177–182. (k) Roy, S.; Hecht, M. H. *Biochemistry* **2000**, *39*, 4603–4607.

(3) (a) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652–3711. (b) Hadley, E. B.; Testa, O. D.; Woolfson, D. N.; Gellman, S. H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 530–535.

(4) Case, M. A.; McLendon, G. L. *J. Am. Chem. Soc.* **2000**, *122*, 8089–8090.

(5) (a) Doerr, A. J.; Case, M. A.; Pelczer, I.; McLendon, G. L. *J. Am. Chem. Soc.* **2004**, *126*, 4192–4198. (b) Lieberman, M.; Tabet, M.; Sasaki, T. *J. Am. Chem. Soc.* **1994**, *116*, 5035–5044. (c) Ghadiri, M. R.; Soares, C.; Choi, C. *J. Am. Chem. Soc.* **1992**, *114*, 825–831.

(6) Roy, L.; Case, M. A. *Chem. Commun.* **2009**, 192–194.

(7) Ghosh, P. B.; Whitehouse, M. W. *Biochem. J.* **1968**, *108*, 155–156.

(8) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487–493.

(9) Cooper, H. J.; Case, M. A.; McLendon, G. L.; Marshall, A. G. *J. Am. Chem. Soc.* **2003**, *125*, 5331–5339.

(10) (a) Dublin, S. N.; Conticello, V. P. *J. Am. Chem. Soc.* **2008**, *130*, 49–81. (b) Kammerer, R. A.; Kostrewa, D.; Progias, P.; Honnappa, S.; Avila, D.; Lustig, A.; Winkler, F. K.; Pieters, J.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13891–13896. (c) Zimenkov, Y.; Conticello, V. P.; Guo, L.; Thiyagarajan, P. *Tetrahedron* **2004**, *60*, 7237–7246. (d) Potekhin, S. A.; Melnik, T. N.; Popov, V.; Lanina, N. F.; Vazina, A. A.; Rigler, P.; Verdini, A. S.; Corradin, G.; Kajava, A. V. *Chem. Biol.* **2001**, *8*, 1025–1032. (e) Pandya, M. J.; Spooner, G. M.; Sunde, M.; Thorpe, J. R.; Rodger, A.; Woolfson, D. N. *Biochemistry* **2000**, *39*, 8728–8734. (f) Ogihara, N. L.; Weiss, M. S.; DeGrado, W. F.; Eisenberg, D. *Protein Sci.* **1997**, *6*, 80–88. (g) Krylov, D.; Mikhailenko, I.; Vinson, C. *EMBO J.* **1994**, *13*, 2849–2861.

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

JA1029717