J|A|C|S COMMUNICATIONS

Published on Web 12/30/2008

Designed α-Helical Tectons for Constructing Multicomponent Synthetic Biological Systems

Elizabeth H. C. Bromley,*,† Richard B. Sessions,‡ Andrew R. Thomson,† and Derek N. Woolfson†,‡

School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, U.K, and, Department of Biochemistry, School of Medical Sciences, University of Bristol, University Walk, Bristol, BS8 1TD, U.K.

Received June 11, 2008; E-mail: Beth.Bromley@bristol.ac.uk

One route to develop new synthetic-biological systems is to assemble discrete nanoscale objects from programmed peptide-based building blocks, or tectons.[1] This would require tectons that exhibit high fidelity interactions with specified partners, and the potential to be functionalized. Such a set of peptidic building blocks could be used to create more-complex assemblies analogous to the way that DNA has been used design a wide range of 2D and 3D nanostructures.[2−6] While DNA-based assemblies have reached the level of sophistication to allow control of structure and even motion,[7−9] the use of peptidic building blocks is far less advanced. This is because the rules relating peptide sequence to structure are less well established than the exquisite specificity of base pairing in DNA. Although the design of peptidic tectons is more difficult, the potential benefits include ease of production, functionalization, and versatility. Here we present an algorithm to search coiled-coil specificity space, along with a few modest multicomponent structures to demonstrate our ability to build peptide-based nanostructures from the bottom up.

We take the α-helical coiled coil as a starting point for design because the rules governing its structure, oligomerization state, and partner specificity are arguably the most developed for any protein-folding motif.[10−12] Indeed, scoring matrices that allow the prediction of melting temperatures, or rank ordering of partner sequences based on interactions within and straddling the hydrophobic core are becoming available (Figure 1).[13,14]

To produce building blocks that interact in specified combinations, we must design coiled-coil sequences that preferentially bind to their partner sequence over any other sequence present. In other words, we need to find the set of coiled-coil sequences that maximize the energy gap between the least favorable desirable and the most favorable undesirable interactions. In the first instance we wish to make multicomponent sequences via peptide synthesis and are therefore limited to considering coiled-coil segments that are a maximum of three heptads—the seven-residue, ***a***−***g***, signature repeat of coiled-coil sequences—in length. These are too long to be considered in an exhaustive search of sequence space.[15] We consider only a limited number of heptad positions that impact directly on the peptide−peptide interface, and a reduced amino-acid alphabet.

As a starting point for assigning theoretical interaction scores, we created a matrix based on the bCIPA algorithm (see Supporting Information).[13] The matrix gives a score for the interaction between two heptads from either the same or different sequences. The interactions considered and their scores are shown in Figure 1B. We allowed only Asn or Ile at position ***a***, Leu at position ***d***, and only Gln, Lys, or Glu at positions ***g*** and ***e***. There are 18 possible heptads that can be made using these residues and sequence positions. Combining three such heptads with the constraint that Asn was not used in terminal heptads due to its capping propensity,[16] generated 1458 candidate sequences for the algorithm to rank.

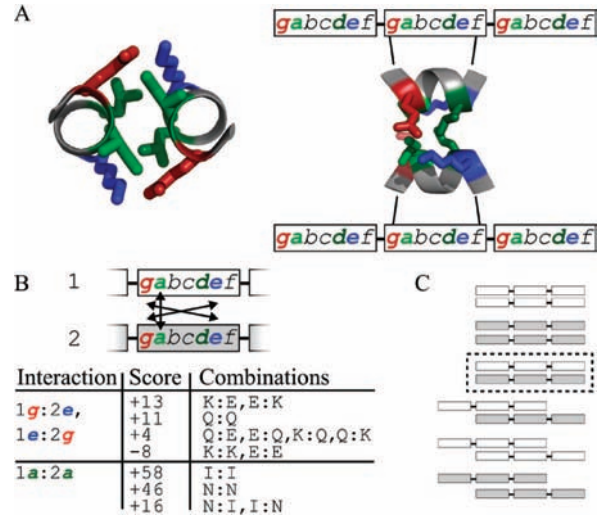

Figure 1. Principles of coiled-coil design. (A) View of a dimeric coiled-coil heptad from the N- to C-terminus. The specifying residues at the ***g***, ***a***, ***d***, and ***e*** positions are highlighted indicating their heptad repeat positions. (B) Interactions considered in generating a heptad-based score. (C) Some possible configurations scored in the algorithm. The target blunt-ended heterodimer is highlighted (dotted box).

The maximum score for one strand interacting with another was calculated by scoring all of the possible homo-, hetero-, blunt-ended, sticky-ended, parallel, and antiparallel configurations (Figure 1C). For the purposes of this work a successful pair of sequences is one for which the blunt-ended, parallel, heterodimer is the highest scoring. These criteria can be extended to a set of two or more dimeric coiled coils (four or more sequences). For this system, we found that a maximum of three heterodimers could be generated.

We chose a specific group of peptide sequences from a range of the highest scoring hits based on secondary information. First, we considered the preferred ordering for Glu and Lys pairings in human BZip protein sequences (Glu at ***g*** Lys at ***e***).[17] Second, we gave a bias toward positively charged peptides to assist purification and solubility. Third, we considered the interactions that would be involved once these peptides were linked together to allow further assembly (Supporting Information). Once a group had been chosen, the ***b***, ***c***, and ***f*** positions, which were not specified by the algorithm, were filled with a mixture of high α-helix propensity Ala and Gln, and Tyr to give each peptide a chromophore and a unique mass (Figure 2). The scores for each sequence as a homodimer and the 15 heterodimer combinations were calculated (Figure 2).

† School of Chemistry.
‡ School of Medical Sciences.

10.1021/ja804231a CCC: $40.75

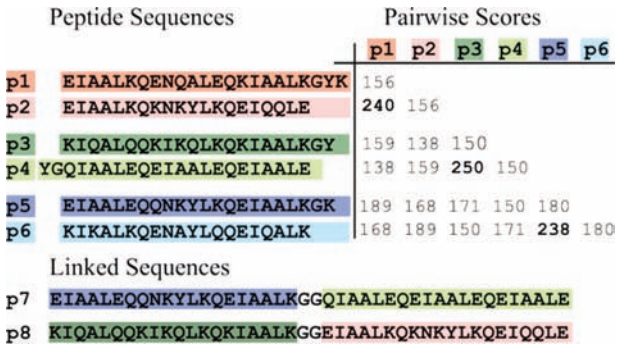

Figure 2. Results of the design algorithm. The score for each of the 21 pairwise combinations along with the sequences of these peptides and the linked pairs of peptides.

To test the success of the algorithm, the six sequences were synthesized and characterized for folding and assembly. The folding of the 21 peptide combinations was probed by circular dichroism (CD) spectroscopy (Figure 3A). In both shape and magnitude these spectra indicated that the individual peptides and the 12 nonspecified combinations were less than half-folded as α-helices. In contrast, all of the target interactions showed a higher level of α-helical folding, and, at 100 μM, were the only mixtures for which T_M values could be determined (p1/p2, 21 °C; p3/p4, 50 °C; and p5/p6, 15 °C). The weaker pairings also exhibited concentration dependence as expected for oligomeric assemblies (Figure S2, Supporting Information).

To test the fidelity of the pairings, *N*-terminally Cys-labeled versions of the six peptides were made.[18] These peptides were mixed under redox buffered conditions designed to allow each peptide to be either reduced or oxidized. The species produced were monitored using reverse-phase HLPC and identified by mass spectrometry (Figure 3B). Under denaturing conditions, only a trace amount of the most stable disulfide-bonded dimer was found, with the majority of the peptide present in either the reduced form or conjugated to glutathione. However, under folding conditions only 3 out of the possible 21 disulfide-bonded dimers were found in significant quantities, and their masses corresponded to the three designed pairs (p1/p2, p3/p4, and p5/p6). Furthermore, these disulfide-bound pairs were fully folded at 5 °C and exhibited concentration-independent folding, demonstrating that they were parallel dimeric species as expected (Figure S2−S7, Supporting Information). The mixture of all three pairs showed no change in stability over the average of the isolated mixtures indicating that even in the presence of all of the possible competing interactions the system successfully generates only the three specified interactions (Figure S3, Supporting Information).

To demonstrate the utility of the designed 3-heptad coiled coils as tectons for synthetic biology, we synthesized two longer peptides: p7 comprising p5-Gly-Gly-p4, and p8 comprising p3-Gly-Gly-p2 (Figure 2). This enabled us to build a set of discrete nanostructures with more than one coiled-coil domain: mixing p3 and p4 should render a 3 nm rod; p7, p3, and p6 gives a 6 nm rod; and p7, p8, p1, and p6 gives a 9 nm rod (Figure 4). A Gly−Gly linker was chosen to allow flexibility and space for the shorter peptides to dock adjacent to one another.

To probe the assemblies in more detail, we measured the dominant diffusion coefficient using dynamic light scattering (DLS) (Figure 5B). For comparison, we calculated a theoretical diffusion coefficient for each of the assemblies using coordinate files based on appropriate coiled-coil models, which were then processed using the program *Hydropro*.[19] Although the theoretical values slightly

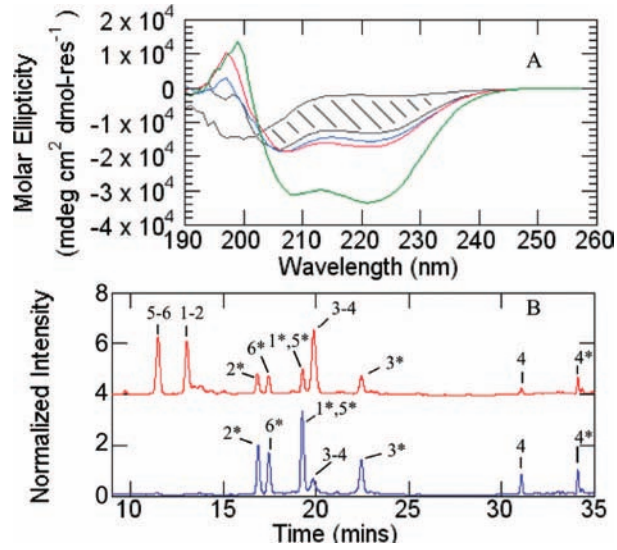

Figure 3. Partner specificity. (A) CD spectra obtained at 100 μM of each peptide, 5 °C and 10 mM MOPS. Grey hatched area shows the region occupied by the spectra from the 6 individual peptides and the 12 nonspecified interactions (Figure S1, Supporting Information): red, p1/p2; green, p3/p4; and blue, p5/p6. (B) Normalized HPLC traces from the mixture of all six Cys-linked peptides (100 μM of each peptide, 5 °C, 10 mM MOPS, 1 mM reduced glutathione and 4 mM oxidized glutathione), blue line, with 6 M guanidine HCl, red line, without (normalized intensity + 4). Numbers refer to peptides: (−) disulfide bond; (∗) glutathione adduct.

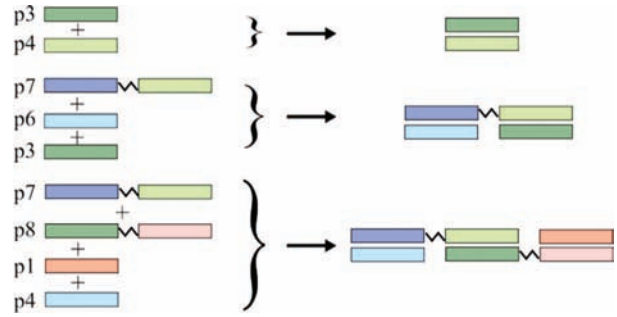

Figure 4. Scheme for nanostructure assembly; color as in Figure 2.

underestimated the peak of experimentally determined size distributions (possibly due to an underestimate of the solvent layer, or an overestimate of the density) the trend and correlations are clear (Figure 5B).

Analytical ultracentrifugation (AUC) was also performed on the three mixtures (Figure 6). All three samples modeled well as single species and the observed masses (4980, 9500, and 14650 Da) compared well with those calculated from sequence (4996, 9572, and 14463 Da). Together with the demonstrated fidelity of the tecton pairs, these CD, DLS, and AUC data indicate that we have successfully produced discrete nanoscale objects by combining programmed peptidic building blocks.

In conclusion, we present a method that applies rules relating sequence and stability in dimeric coiled coils to produce peptidic tectons for use in self-assembly and synthetic biology. The method uses a reduced alphabet and considers only the ***g***, ***a***, ***d***, and ***e*** positions of the heptad repeat. In this way we designed six peptides to interact preferentially with their designed partners over any of the other possible combinations and so demonstrated control over specificity.

Our system also presents a framework for extending and testing scoring matrices for sequence-to-stability relationships in coiled-

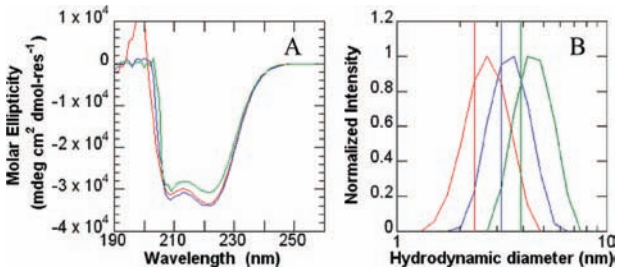

Figure 5. Characterization of multicomponent assemblies I. (A) CD spectra (red, p3 + p4; blue, p7 + p3 + p6; green, p7 + p8 + p1 + p6). (B) DLS data indicating the size distribution for each mixture. Vertical bars indicate the predicted diameters.

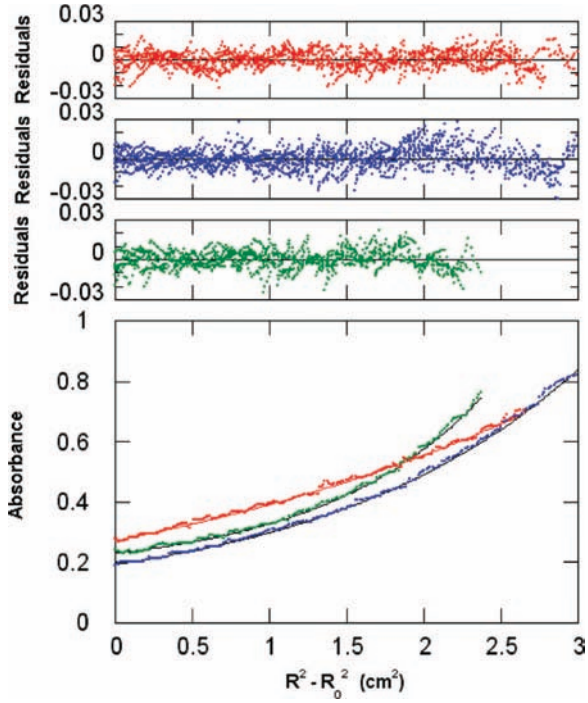

Figure 6. Characterization of multicomponent assemblies II. Lower panel: AUC data of the three more-complex combinations (red, p3 + p4; blue, p7 + p3 + p6; green, p7 + p8 + p1 + p6). Data from runs at 33000 rpm are shown along with the single ideal species fit.[20] Upper panels: the residuals for each fit.

coil systems. For example, we note that the difference in folding between the most-folded nonspecified interaction and the least-folded specified interaction, and the T_M values are quite modest with respect to natural coiled-coil proteins. We also find that bCIPA, on which our algorithm is based, fails to accurately predict T_M values for our 3-heptad coiled coils. These discrepancies are likely due to the fact that bCIPA is trained on natural coiled coils.[13] The majority of these proteins derive their stability and specificity from sequences that are several heptads longer than those we had to use herein. We anticipate that extending this algorithm to design sequences with four heptads instead of three—while necessitating a move to either ligation strategies or recombinant methods to produce the linked peptides of the type described—should increase both the stability and the number of specified pairs that can be generated, and give better correspondence between the predicted and experimental stabilities. For such longer constructs it should also be possible to incorporate a wider range of specifying motifs including recently demonstrated de novo examples.[21−23]

By linking the designed coiled-coil segments into larger sequences, we have begun to build self-assembling nanostructures. This concept is readily expandable both to longer coiled coils and to larger numbers of domains. These extensions could provide both the extra versatility and stability necessary to achieve programmed multicomponent self-assembling systems. In future, we aim to extend the number and functionality of the available building blocks, thus developing a toolkit and a basis for an exciting bottom-up approach to synthetic biology.

Acknowledgment. We thank J. Spencer, H. Hoerber, R. Ehlich, and the Woolfson Group for input to the project and the EPSRC (GR/T09224/01) and HFSP (RPG31/2007) for funding.

Supporting Information Available: Experimental methods, supporting CD, AUC, and MS data and algorithm details. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N. A.C.S. *Chem. Biol.* **2008**, *3*, 38–50.
(2) Rothemund, P. W. K. *Nature* **2006**, *440*, 297–302.
(3) Goodman, R. P.; Schaap, A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J. *Science* **2005**, *310*, 1661–1665.
(4) Erben, C. M.; Goodman, R. P.; Turberfield, A. J. *J. Am. Chem. Soc.* **2007**, *129*, 6992.
(5) Jaeger, L.; Chworos, A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 531–543.
(6) Ke, Y.; Lindsay, S.; Chang, Y.; Liu, Y.; Yan, H. *Science* **2008**, *319*, 180–183.
(7) Bath, J.; Turberfield, A. J. *Nat. Nanotechnol.* **2007**, *2*, 275–284.
(8) Seeman, N. C. *Trends Biochem. Sci.* **2005**, *30*, 119–125.
(9) Venkataraman, S.; Dirks, R. M.; Rothemund, P. W.; Winfree, E.; Pierce, N. A. *Nat. Nanotechnol.* **2007**, *2*, 490–494.
(10) Harbury, P. B.; Kim, P. S.; Alber, T. *Nature* **1994**, *371*, 80–83.
(11) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401–1407.
(12) Woolfson, D. N. *Adv. Protein Chem.* **2005**, *70*, 79–112.
(13) Mason, J. M.; Schmitz, M. A.; Muller, K. M.; Arndt, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8989–8994.
(14) Newman, J. R. S.; Keating, A. E. *Science* **2003**, *300*, 2097–2101.
(15) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34*, 11645–11651.
(16) Chakrabartty, A.; Doig, A. J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 11332–11336.
(17) Fong, J. H.; Keating, A. E.; Singh, M. *Genome Biol.* **2004**, *5*, R11.
(18) O'Shea, E. K.; Rutkowski, R.; Stafford III, W. F.; Kim, P. S. *Science* **1989**, *245*, 646–648.
(19) de la Torre, J. G.; Huertas, M. L.; Carrasco, B. *Biophys. J.* **2000**, *76*, 719–730.
(20) Demeler, B. In *Modern Analytical Ultracentrifugation:Techniques and Methods*; Scott, D. J., Harding, S. E., Rowe, A. J., Eds.; Royal Society of Chemistry: 2005; p 210−229.
(21) Diss, M. L.; Kennan, A. J. *J. Am. Chem. Soc.* **2008**, *130*, 1321–1327.
(22) Havranek, J. J.; Harbury, P. B. *Nat. Struct. Biol.* **2003**, *10*, 45–52.
(23) Diss, M. L.; Kennan, A. J. *Org. Lett.* **2008**, *10*, 3797–3800.

JA804231A