



Published in final edited form as:

Science. 2017 December 15; 358(6369): 1461–1466. doi:10.1126/science.aap7577.

Comprehensive computational design of ordered peptide macrocycles

Parisa Hosseinzadeh^{1,*}, Gaurav Bhardwaj^{1,*}, Vikram Khipple Mulligan^{1,*}, Matthew D. Shorridge², Timothy W. Craven¹, Fátima Pardo-Avila³, Stephen A. Rettie¹, David E. Kim¹, Daniel-Adriano Silva¹, Yehia M. Ibrahim⁴, Ian K. Webb⁴, John R. Cort⁴, Joshua N. Adkins⁴, Gabriele Varani², and David Baker^{1,5,†}

¹Department of Biochemistry and Institute for Protein Design, University of Washington, Seattle, WA 98195, USA

²Department of Chemistry, University of Washington, Seattle, WA 98195, USA

³Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

⁴Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

⁵Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

Abstract

Mixed-chirality peptide macrocycles such as cyclosporine are among the most potent therapeutics identified to date, but there is currently no way to systematically search the structural space spanned by such compounds. Natural proteins do not provide a useful guide: Peptide macrocycles lack regular secondary structures and hydrophobic cores, and can contain local structures not accessible with L-amino acids. Here, we enumerate the stable structures that can be adopted by macrocyclic peptides composed of L- and D-amino acids by near-exhaustive backbone sampling followed by sequence design and energy landscape calculations. We identify more than 200 designs predicted to fold into single stable structures, many times more than the number of currently available unbound peptide macrocycle structures. Nuclear magnetic resonance structures of 9 of 12 designed 7- to 10-residue macrocycles, and three 11- to 14-residue bicyclic designs, are close to the computational models. Our results provide a nearly complete coverage of the rich space of structures possible for short peptide macrocycles and vastly increase the available starting scaffolds for both rational drug design and library selection methods.

[†]Corresponding author. dabaker@u.washington.edu.

^{*}These authors contributed equally to this work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/358/6369/1461/suppl/DC1

Methods

Figs. S1 to S29

Tables S1 to S7

Databases S1 to S3

References (25–50)

The high stability, diverse functionality, and favorable pharmacokinetic properties of macrocyclic peptides make them promising starting points for targeted therapeutics (1–4). However, there are few well-characterized natural macrocycles, and they are difficult to repurpose for new functions. Thus, most current approaches focus on random library selection methods, which, although powerful (5–7), only cover a small fraction of the vast sequence space that can be accessed by even short sequences of L- and D-amino acids (8,9) and often yield peptides that are not structured in the absence of target. Methods are needed for designing ordered macrocycles with shapes precisely crafted to bind their targets and with functionalities common in medicinal chemistry, but absent in the natural 20 amino acids, positioned at critical interaction sites. Despite the progress in computational design of proteins (10–14) and constrained peptides as small as 18 residues (15), designing shorter peptide macrocycles had remained an unsolved challenge. The driving force for the folding of larger peptides and proteins is the sequestration of hydrophobic residues in a nonpolar core, enabled by regular secondary structures in which buried backbone polar groups can make hydrogen bonds. This principle has been the basis of almost all previous peptide and protein design work. However, the balance of forces is considerably different for 7- to 10-residue peptides: They are too small to have either a solvent-excluded hydrophobic core or α -helical and β -sheet (other than β -hairpin) secondary structures. Beyond these differences in the physics of folding, protein design methods often use the PDB (Protein Data Bank) as a source of local structural information, but native structures provide a poor guide for local structures that include noncanonical D-amino acids. On the other hand, short cyclic peptides are an attractive target for computational design because unlike larger systems, there is the possibility of obtaining a completeness of conformational sampling rare in any molecular design endeavor.

The local structure space relevant for cyclic peptides is quite different from that of proteins, so they cannot be systematically generated by assembling protein fragments. Instead, we used generalized kinematic closure (genKIC) (15–17) drawing from achiral flat-bottom backbone torsional distributions (fig. S1, A and B) to generate closed backbone structures starting from a polyglycine chain. For each structure, we used Monte Carlo simulated annealing to search for the lowest-energy amino acid sequence, restricting positions with negative values of the backbone torsion angle ϕ to L-amino acids (and rotamers) and those with positive values to D-amino acids, and disallowing glycine to maximize local sequence encoding of the structure. In preliminary calculations, we found that energy gaps greater than $\sim 10 k_B T$ (~ 6 kcal/mol), where k_B is the Boltzmann constant and T is temperature, could only be obtained for N -residue macrocycles if they contained at least $N/3$ backbone hydrogen bonds; hence, in subsequent calculations, back-bones with fewer hydrogen bonds were discarded. We carried out large-scale backbone generation and sequence design calculations for 7- to 10-residue backbone cyclized peptides, obtaining 50, 596, 12,374, and 49,571 distinct backbones for lengths 7, 8, 9, and 10, respectively, after clustering based on backbone torsion angle bins (ABXY, where torsion bin A is the helical region of Ramachandran space, B is the extended strand region, X is mirror of A, and Y is mirror of B) (Fig. 1A) and backbone hydrogen bond patterns (fig. S1, C to E). Because the sampling method is stochastic, there is no guarantee of completeness, but the symmetry of the system enables a convergence test: For each distinct peptide backbone conformation identified, the

mirror image should also be observed (figs. S2 and S3). As the amount of sampling increases, the number of clusters identified for which the mirror image is observed initially increases, as does the number of clusters with no mirror. The former then plateaus, whereas the latter decreases to near zero (Fig. 1B). Such convergence suggests near-complete coverage of the combined D- and L-space compatible with peptide closure with backbone hydrogen bonds and no steric clashes. We also sampled and designed structures for 11- to 14-residue macrocycles, but did not seek completeness owing to combinatorial explosion in the number of states.

The Monte Carlo simulated annealing sequence design calculations seek a sequence that minimizes the energy of the target backbone conformation, but there is no guarantee that the sequence found maximizes the energy gap between the target backbone conformation and alternative conformations. To assess the energy landscape for low-energy designs (from 21 designs for length 7 to 673 designs for length 10), 10^4 to 10^5 conformations were generated for each sequence, and the energy was minimized with respect to the backbone and side-chain torsion angles. The energy gap and Boltzmann-weighted probability of finding the peptide in or close to the designed main-chain conformation (P_{Near}) were estimated from the resulting energy landscapes (Fig. 1C). A total of 12, 22, 45, and 145 designs with distinct backbone structures had energy landscapes strongly funneled into the design target structure for 7-, 8-, 9, and 10-residue macrocycles, respectively (Fig. 1D and figs. S4 and S5). For comparison, in the PDB and CSD (Cambridge Structure Database), we only found four 7- to 10-residue macrocycle structures not bound to a target and composed only of the 20 canonical amino acids and their mirror images without cross-links or backbone modifications such as N-methylation (Fig. 1D and fig. S6).

Because of the constraints imposed by the cyclic backbone, the small size, and the presence of D-amino acids, the designs span a local structural space inaccessible or underexplored in native proteins. Recurrent features include hydrogen-bonded turnlike structures and proline-stabilized kinks, some of which are observed rarely or not at all in native proteins (Fig. 2 and fig. S7A), that can be viewed as macrocycle-generating building blocks (fig. S7B and table S1). Stepwise residue insertion preserves some of the building blocks and alters others, resulting in a complex propagation of features from the shorter macrocycles to the longer ones (Fig. 2).

It was not feasible to characterize each of the >200 macrocycle designs (database S1) experimentally. Instead, we chemically synthesized (fig. S8) a subset of 12 peptides [four 7-residue peptides (7mers), two 8mers, three 9mers, and three 10mers; tables S2 to S4] and experimentally characterized their structures by nuclear magnetic resonance (NMR) spectroscopy (table S5 and figs. S9 to S11). Ten of the 12 peptides had well-dispersed one-dimensional NMR spectra, with the number of backbone HN peaks expected for a single conformation (tables S2 to S5). We collected extensive nuclear Overhauser effect (NOE) data (fig. S11) for these peptides and solved their structures using XPLOR-NIH (18,19) followed by NOE-restrained molecular dynamics (MD) simulations [similar structures were obtained with an independent large-scale enumeration approach (fig. S12)]. As shown in Figs. 3 and 4 and described below, the experimental NMR structures closely matched the design models for nine of these peptides, and in unrestrained MD simulations, eight out of

the nine peptides are within 1 Å of the designed structure more than 75% of the time (figs. S13 to S15 and table S6). These data suggest that a large fraction of our >200 macrocycles are structured as designed.

Unlike proteins, macrocycles cannot be stabilized primarily by the hydrophobic effect as they are too small to form a core that can exclude solvent (20). How, then, do the sequences of the designs specify their structures? To address this question, we computed the effect on folding of every single substitution to a different amino acid with the same chirality, and to an alanine with opposite chirality, at each position, for all the designs with NMR-confirmed structures. For each of the 20xNres variants, full energy landscape calculations were carried out through large-scale backbone enumeration (Figs. 3 and 4; details in fig. S16). These computationally intensive calculations were carried out by using cellular phones and tablets of volunteers participating in the Rosetta@Home distributed computing project (<http://boinc.bakerlab.org/rosetta>). To evaluate the computed sequence-energy landscape experimentally, we used SLIM (structures for lossless ion manipulations), an ion-mobility mass spectrometry technique that can distinguish different conformations in small molecular structures (21). This technique requires only a small amount of unpurified sample and enables parallel evaluation of the effects of amino acid substitutions on folding. SLIM results from a set of variants with point mutations of design 7.1 at either the dPro⁴ or Thr⁵ position (Fig. 3 and fig. S17) were consistent with the sequence-energy landscape calculations: The structure was perturbed more by mutations at the dPro⁴ position than at the Thr⁵ position, consistent with the computed P_{Near} values.

Several general principles emerge from the comprehensive landscape calculations and from folding calculations on permuted sequences (fig. S18). First, L- and D-proline residues play a key role in structure specification: 52% of the positions in which substitutions disrupt the structure are proline residues in the design, and in almost all of the cases, the most destabilizing mutant of a nonproline residue is a substitution to proline (fig. S19). Proline is the most torsionally constrained amino acid, and placement of L- and D-proline residues favors specific turn and kink structures. Second, side-chain-to-backbone hydrogen bonds that either stabilize a local structural motif, such as Asp² in design 8.1, or connect two sides of the structure, such as Glu² in design 10.1 or Asp³ in design 10.2, are important for structural specification as removal of these interactions substantially reduces the energy gaps. Third, chirality in many cases plays a greater role in structure specification than side-chain identity: Replacing an amino acid residue with its mirror-image isomer is usually more disruptive than changing to a different amino acid with the same chirality. Fourth, for each design, usually fewer than three residues (often proline) are critical to defining the fold, leaving the remainder largely free for future functionalization (figs. S16 and S19). Even after mutation of the remaining residues to Ala (retaining chirality), a number of the sequences still encode the designed structure (fig. S20). Overall, this global analysis of the effect of substitutions on energy landscape topography defines the sequence determinants of the folding-energy landscape in exceptional detail.

It is instructive to consider these data in the context of the structures and sequences of the individual designs. The seven-residue macrocycles exhibit several recurrent backbone hydrogen-bonding patterns, often featuring a proline-nucleated $i, i + 3/i, i + 4$ motif (Fig. 2,

second row, and fig. S21). Of the four seven-residue designs experimentally tested, three had structures nearly identical to the design models (Fig. 3 and table S7), and MD and Rosetta calculations on the fourth (design 7.4) suggest that it also is close to the design model despite overlap of backbone NH-group NOEs (figs. S12 and S15). The energy landscape calculations show that the proline nucleating the $i, i + 3/i, i + 4$ turn is essential (Fig. 3, A and E). The remainder of the structure is largely specified by the designed amino acid chirality with the exception of dPro⁵ in design 7.2, which packs on the turn-nucleating dPro⁷. The eight-residue macrocycles (fig. S22) are dominated by two major classes, one featuring two $i, i + 3/i, i + 4$ motifs (Fig. 2, second row) and the other, two criss-cross $i, i + 3/i + 1, i + 4$ motifs (Fig. 2, third row). The two eight-residue macrocycles that were experimentally characterized both had NMR structures within 1 Å of the design model. Design 8.1 has multiple slow-exchanging (fig. S23) side-chain–side–chain and side-chain–backbone hydrogen bonds, with a structurally critical (Fig. 3, panels IV. A and E) hydrogen bond from Asp² to the backbone of Thr⁴, which along with Pro³ stabilizes a sharp kink in the chain. Adjacent to the kink is a BXX ($i, i + 3/i + 1, i + 4$) motif rare in proteins, anchored by the structurally critical dPro⁵. Design 8.2 has near-perfect sequence inversion symmetry; the sequence-symmetric version of this design with sequence Pro-Gln-dArg-Glu-dPro-dGln-Arg-dGlu and torsion string AAYBXXBY has half the number of NMR resonances (three backbone HN instead of six), consistent with structural S_2 symmetry (fig. S24). In contrast to the other seven- to eight-residue designs characterized, all residues in design 8.2 are important for structure specification (fig. S16 and Fig. 3, panel V. E), with residues involved in multiple side-chain–side–chain hydrogen bonds more essential than the two prolines.

As the macrocycle length increases (9 and 10 residues, Fig. 4), so does the entropic cost of folding, and more hydrogen bonds in increasingly diverse patterns (fig. S25) are required to stabilize the peptide in the folded state. Three of six experimentally characterized designs had structures close to computational models, one was disordered, and two had well-dispersed spectra, but the NOE data did not uniquely define the structures (table S5). Design 9.1 contains a YAA $i, i + 3/i, i + 4$ building block similar to those in the seven-residue macrocycles in which dPro⁵ plays a critical role (as in the L-Pro/D-Pro in design 7.3, the second proline plays a less critical role). The structure is expanded by insertion of a kink stabilized by Pro⁹; the remainder of the structure is completed by a tight AAA $i, i + 3/i, i + 4$ turn. Design 10.1 contains a five-residue distorted helix terminated by the critical dPro⁷. On one face, the structurally critical Glu² in the middle of the helix makes a long-range side-chain–backbone hydrogen bond to Arg⁸, and on the other, Ala³, dVal⁶, and dLeu⁹ form a nonpolar cluster. Design 10.2 contains BX, YA, and the rare YAX building blocks, each beginning with a proline residue; of these, Pro⁴ in the BX motif is the most critical. As with 10.1, the building blocks are held together by nonpolar interactions (between dVal² and dAla⁸) on one face, and a long-range side-chain–backbone hydrogen bond (from Asp³ to Asn⁸) on the other; both dVal² and Asp³ are critical for specifying the structure (Fig. 4, panel IV.E).

The entropic cost of folding continues to increase with increasing number of residues, and for 11- to 14-residue macrocycles, additional crosslinks to form bicyclic structures were required to obtain single states amenable to NMR structure determination (table S5). We solved the structures of three such designs (Fig. 4, row IV) that feature long-range

backbone-backbone hydrogen bonds. Design 11_SS has a $i, i + 3/i + 1, i + 4$ building block (Fig. 2, third row) with a critical proline (fig. S16) in the first position preceded by a cysteine that forms a critical disulfide to a cysteine preceding a YA turn. Design 12_SS has a rare BXAX $i, i + 4/i, i + 5$ turn, which exhibits higher flexibility in the NMR structure, and a disulfide between non-hydrogen-bonding residues. The more compact and complex 14_SS design has a network of interleaved local and nonlocal backbone hydrogen bonds (22), and a D-Cys to L-Cys disulfide bond.

The wide variety of shapes spanned by our macrocycle designs, together with their high stability (fig. S26) and high predicted tolerance for sequence mutations (figs. S16 and S20), makes them attractive starting points for developing new therapeutics. One approach to inhibitor design is scaffolding loops at binding interfaces in the PDB; such scaffolding can increase binding affinity by preorganizing the loops in the binding-competent conformation, enable additional interactions with the target, and improve cell permeability and oral bioavailability (23). We found that 907 of the 1017 “hot loops” identified at protein-protein interfaces by Kritzer and co-workers (24) (database S2) could be scaffolded by one or more of our designs (see fig. S27 for some examples).

The finding that 70% of the experimentally tested 7- to 10-residue macrocycle designs adopt single unique structures close to the computationally designed models suggests that most of the 200+ new macrocycle designs with high computed Boltzmann weights fold as designed, increasing the known repertoire of possible macrocycle structures by more than two orders of magnitude. Our results demonstrate that the principles and energy functions developed in recent years to design proteins have quite broad applicability, transferring over to much smaller systems even though (i) the factors dominating the folding of proteins (for example, the hydrophobic effect) differ considerably from those that stabilize conformations of small peptide macrocycles (local hydrogen bonding patterns and intrinsic conformational preferences of amino acid building blocks), and (ii) all designed proteins to date contain regular α -helix or β -sheet structures, whereas small peptide macrocycles lack these and instead contain a wide range of local structures, some of which are rarely or never observed in proteins.

There are two clear paths forward for engineering new macrocyclic therapeutics by exploiting the rigidity and stability of the designs together with the freedom to choose the identities of the nonstructure-specifying positions. The first is experimental: Libraries can be constructed in which, at each position, all residues compatible with the structure are allowed (identified as described above by using large-scale energy landscape calculations) and screened for target binding with current library selection methodologies. The second is computational: Each macrocycle can be docked against the target (using, for example, rigid body docking or “hot loop” superposition), and the interface residues designed to maximize binding affinity. Unnatural amino acids can be incorporated in either approach, but the second has the advantage that new functionalities—such as known active site binding groups—can be strategically placed to maximize binding affinity. Beyond binding, the control over geometry and chemistry provided by our approach should contribute to understanding the structural correlates of membrane permeability and other desirable pharmacological properties.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH Ruth Kirschstein F32 award no. F32GM120791-02 (P.H.), Washington Research Foundation (P.H. and T.W.C.), NIH award GM1157490 (F.P.-A.), American Cancer Society postdoctoral fellowship PF-13-056-01-RMC (M.D.S.), Pew Latin-American fellowship in biomedical sciences (D.-A.S.), CONACYT postdoctoral fellowship (D.-A.S.), NIH-National Institute of General Medical Sciences 5R01GM103834 (G.V.), and Howard Hughes Medical Institute (D.B.). The SLIM experiment was supported by Laboratory-Directed Research and Development (LDRD) and was performed in the Environmental Molecular Sciences Laboratory (EMSL), a U.S. Department of Energy (DOE)–Biological and Environmental Research national scientific user facility at Pacific Northwest National Laboratory. Computing resources were provided by the many volunteers who have donated the spare CPU (central processing unit) cycles of their cellular telephones and computers to the Rosetta@Home project; the Hyak supercomputer at the University of Washington; the XStream computational resource, supported by the National Science Foundation Major Research Instrumentation program (ACI-1429830); and Amazon Web Services (AWS) Credits for Research Grant (G.B). In addition, an award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. We thank A. Roy, H. Park, and many other members of the Baker lab for discussions; R. Flatland, S. Hari, W. J. Lloyd, L. Clithero, and A. Bucher for AWS resources and support; and B. Weitzner, A. Nayyeri, S. Ovchinnikov, T.J. Brunette, D. Alonso, L. Goldschmidt, and P. Vecchiato for help with coding and computing resources. Conformational sampling was done with the Rosetta simple_cycpep_prediction application and peptide design was carried out with the rosetta_scripts application, both of which are included in the Rosetta software suite. The Rosetta software suite is available free of charge to academic users and can be downloaded from <http://www.rosettacommons.org>. Instructions and inputs for running these applications, and all other data and coding necessary to support the results and conclusions, are in the supplementary materials. All the structures presented here are deposited in PDB and Biological Magnetic Resonance Band (BMRB) (design 7.1: PDB ID 6BE9, BMRB ID 30356; design 7.2: PDB ID 6BEW, BMRB ID 30364; design 7.3a: PDB ID 6BF3, BMRB ID 30365; design 7.3b: PDB ID 6BF5, BMRB ID 30366; design 8.1: PDB ID 6BE7, BMRB ID 30355; design 8.2: PDB ID 6BEN, BMRB ID 30357; design 9.1: PDB ID 6BEO, BMRB ID 30358; design 10.1: PDB ID 6BEQ, BMRB ID 30359; design 10.2: PDB ID 6BER, BMRB ID 30360; design 11_SS: PDB ID 6BES, BMRB ID 30361; design 12_SS: PDB ID 6BET, BMRB ID 30362; design 14_SS: PDB ID 6BEU, BMRB ID 30363) and the original designs are provided in database S3. D.B., G.B., P.H., and V.M. are inventors on a U.S. provisional patent application submitted by the University of Washington that covers computational design of small structured macrocycles. P.H., G.B, V.K.M., and D.B. developed research idea, the experimental approach, and generated the designed macrocycles. D.B., P.H., G.B, V.K.M., M.D.S., F.P.-A., T.W.C., S.A.R, D.-A.S., I.K.W., and Y.M.I. performed experiments. All authors analyzed data. P.H. and V.K.M. generated new computational tools with help from D.E.K. P.H., G.B., V.K.M, and D.B. wrote the manuscript, and all authors contributed suggestions. G.V. and D.B. supervised research.

REFERENCES AND NOTES

1. DrIggers EM, Hale SP, Lee J, Terrett NK. *Nat Rev Drug Discov*. 2008; 7:608–624. [PubMed: 18591981]
2. Thapa P, Espiritu MJ, Cabaltega C, Bingham JP. *Int J Pept Res Ther*. 2014; 20:545–551.
3. Fosgerau K, Hoffmann T. *Drug Discov Today*. 2015; 20:122–128. [PubMed: 25450771]
4. Craik DJ, Fairlie DP, Liras S, Price D. *Chem Biol Drug Des*. 2013; 81:136–147. [PubMed: 23253135]
5. Gray BP, Brown KC. *Chem Rev*. 2014; 114:1020–1081. [PubMed: 24299061]
6. Marasco D, Perretta G, Sabatella M, Ruvo M. *Curr Protein Pept Sci*. 2008; 9:447–467. [PubMed: 18855697]
7. Liu R, Li X, Xiao W, Lam KS. *Adv Drug Deliv Rev*. 2017; 110–111:13–37.
8. Obexer R, Walport LJ, Suga H. *Curr Opin Chem Biol*. 2017; 38:52–61. [PubMed: 28319812]
9. Passioura T, Suga H. *Chem Commun (Camb)*. 2017; 53:1931–1940. [PubMed: 28091672]
10. Huang PS, et al. *Science*. 2014; 346:481–485. [PubMed: 25342806]
11. Huang PS, et al. *Nat Chem Biol*. 2016; 12:29–34. [PubMed: 26595462]
12. Boyken SE, et al. *Science*. 2016; 352:680–687. [PubMed: 27151862]

13. Marcos E, et al. *Science*. 2017; 355:201–206. [PubMed: 28082595]
14. Huang PS, Boyken SE, Baker D. *Nature*. 2016; 537:320–327. [PubMed: 27629638]
15. Bhardwaj G, et al. *Nature*. 2016; 538:329–335. [PubMed: 27626386]
16. Coutsias EA, Seok C, Jacobson MP, Dill KA. *J Comput Chem*. 2004; 25:510–528. [PubMed: 14735570]
17. Mandell DJ, Coutsias EA, Kortemme T. *Nat Methods*. 2009; 6:551–552. [PubMed: 19644455]
18. Schwieters CD, Kuszewski JJ, Clore G Marius. *Prog Nucl Magn Reson Spectrosc*. 2006; 48:47–62.
19. Schwieters CD, Kuszewski JJ, Tjandra N, Clore G Marius. *J Magn Reson*. 2003; 160:65–73. [PubMed: 12565051]
20. Rizo J, Gierasch LM. *Annu Rev Biochem*. 1992; 61:387–418. [PubMed: 1497316]
21. Ibrahim YM, et al. *Analyst*. 2017; 142:1010–1021. [PubMed: 28262893]
22. Hutchinson EG, Thornton JM. *Protein Sci*. 1994; 3:2207–2216. [PubMed: 7756980]
23. Nielsen DS, et al. *Chem Rev*. 2017; 117:8094–8128. [PubMed: 28541045]
24. Gavenonis J, Sheneman BA, Siegert TR, Eshelman MR, Kritzer JA. *Nat Chem Biol*. 2014; 10:716–722. [PubMed: 25038791]

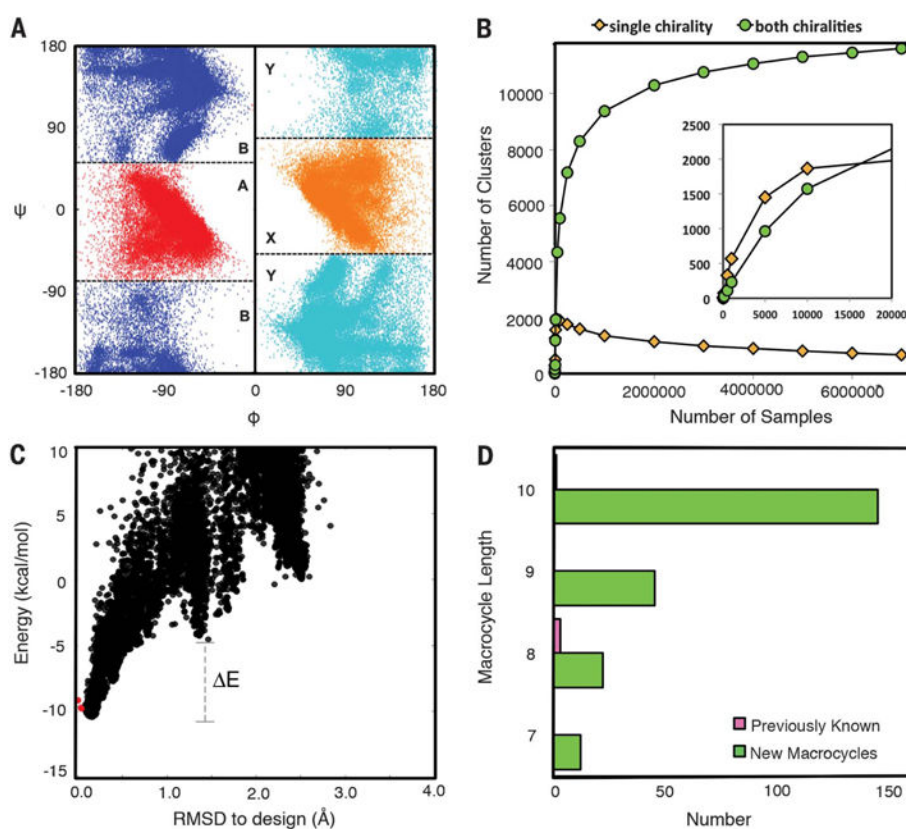


Fig. 1. Expanding the repertoire of ordered macrocycles by comprehensive sampling

(A) Torsion bin definitions (bin B = blue; A = red; Y = mirror of bin B, cyan; X = mirror of bin A, orange) overlaid on the generated nine-residue macrocycle backbone torsion angle distributions. The inversion symmetry is evident. (B) Convergence test. The number of clusters (torsion bin strings) observed for which the mirror image is not observed (yellow) or for which the mirror image is observed (cyan) was determined both for the complete set of 7,500,000 samples for nine-residue macrocycles and for randomly selected subsets of different sizes (x axis). For comparison, the less converged results obtained starting from a 20,000-sample subset and then down-sampling are shown in the inset: The fraction of single-chirality clusters is higher, and the number of both-chirality clusters does not plateau. (C) Energy landscape analysis using large-scale genKIC backbone generation followed by all-atom energy minimization. Each point is the result of an independent calculation; the energy is shown on the y axis and the root mean square deviation (RMSD) from the design model on the x axis. The extent of funneling (or convergence) of the energy landscape is quantified through the energy gap between conformations generated by stochastic sampling around the design model (red points) and high-RMSD structures, and the Boltzmann weight of the near-design model population. (D) Number of designed macrocycle clusters with distinct torsion bin strings and backbone hydrogen bond patterns with energy gaps less than -0.1 and Boltzmann weights greater than 0.85 ; the number of unbound peptide macrocycles in the PDB and CSD, consisting of only 20 canonical amino acids and their D-enantiomers, is indicated in pink.

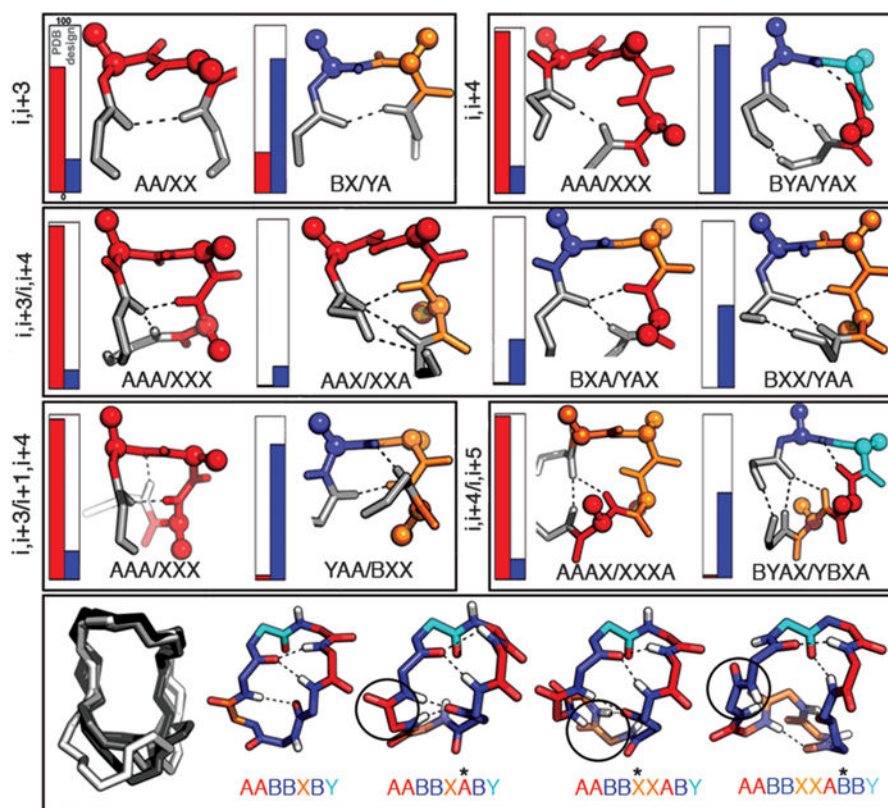


Fig. 2. Recurrent local structural motifs in designed macrocycles

Top three rows: frequencies in native proteins in the PDB (red) and in the designed macrocycles (blue) indicated by vertical bars (including both enantiomers); structures on the right are colored by torsion bin as in Fig. 1A. Bottom row: example of structure propagation with residue insertion for a 7- to 10-residue macrocycle series.

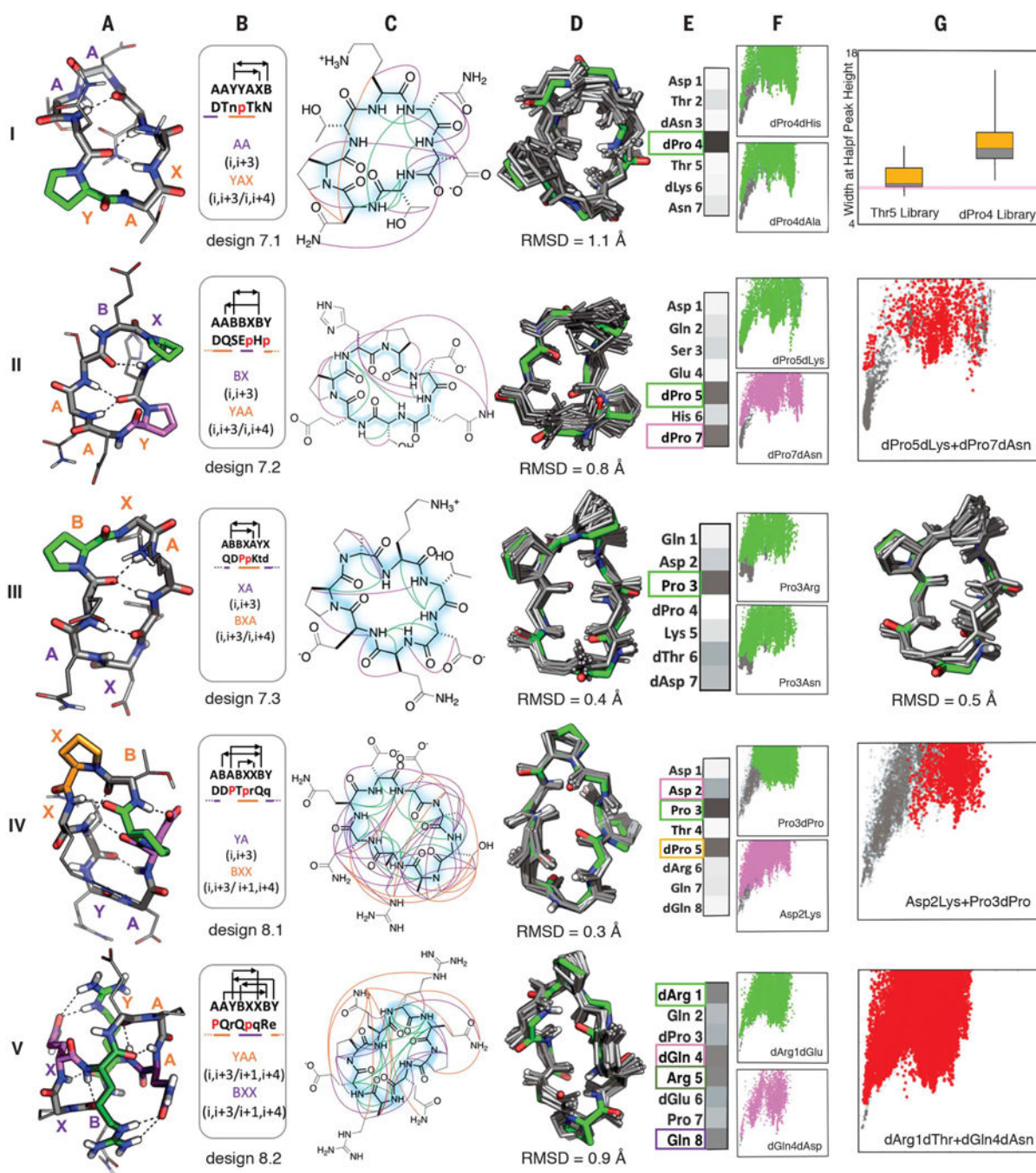


Fig. 3. Seven- to eight-residue macrocycle NMR structures are very close to design models
 Columns A: Design model. B: Amino acid sequence, torsion bin string, hydrogen bond pattern, and building block composition. C: Observed backbone-backbone (green), backbone-side-chain (purple), and side-chain-side-chain (orange) NOEs. D: Overlay of design model (green) on MD-refined NMR ensemble (gray; the average backbone RMSD to the NMR ensemble is indicated). E: Average decrease in the propensity to favor the designed state (P_{Near} , see methods) over all mutations at each position. Darker gray indicates larger decreases (P_{Near} values for each substitution at each position are in fig. S16); positions

particularly sensitive to mutation are boxed and indicated by color in the design model in column A. F: Representative energy funnels for mutations at key positions (colored points) as compared to the design sequence (gray points). Row I, column G: Experimental SLIM data. Distribution of peak width at half height for peptide libraries with all amino substitutions at positions 4 and 5; the position 4 library has a broader distribution consistent with the computed energy landscape in column F. Rows II, IV, V, column G: Representative energy landscapes for double substitutions (red) of critical residues overlaid on the original design landscape (gray). Row III, column G: Overlay of design model on alternative structure NMR ensemble (turn flip at bottom right).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

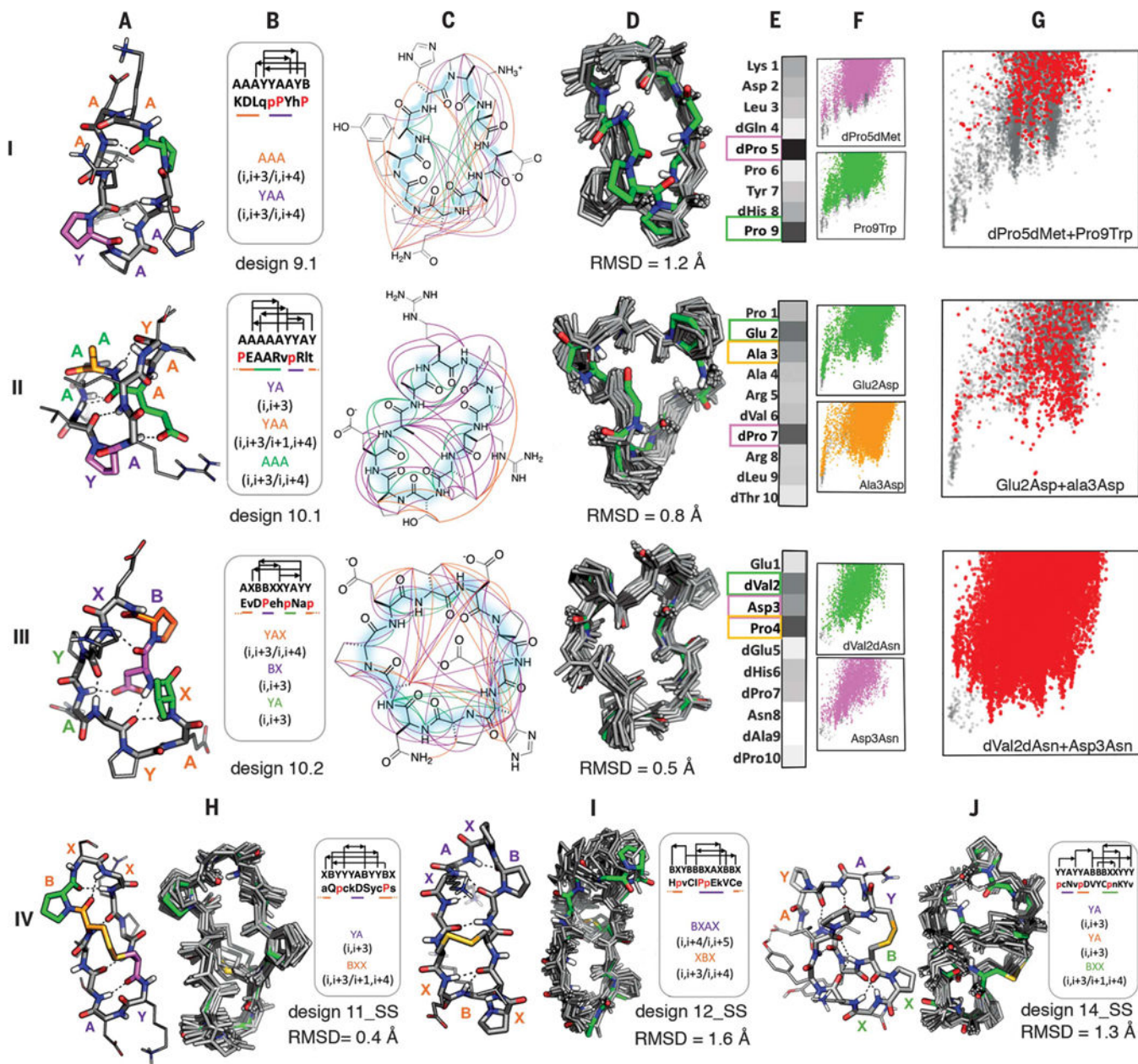


Fig. 4. Nine- to 14-residue macrocycle NMR structures are very close to design models
 Rows I to III: 9- and 10-residue designs. Columns A to G are as in Fig. 3 rows II, IV, and V.
 Row IV: Comparison of bicyclic design models and NMR structures.