Supporting information

De Novo Design of Covalently Constrained Meso-size Protein Scaffolds with Unique Tertiary Structures

Bobo Dang[1,3], Haifan Wu[1,3], Vikram K. Mulligan[2,3], Marco Mravic[1], Yibing Wu[1], Thomas Lemmin[1], Alexander Ford[2], Daniel-Adriano Silva[2], David Baker[2], William F. DeGrado[1,4]

1. Department of Pharmaceutical Chemistry, UCSF, San Francisco, CA 94158

2. Department of Biochemistry, University of Washington, Seattle, WA 98195

3. These authors contributed equally to this work

4. Corresponding author   Dr. William F. DeGrado

    Email: william.degrado@ucsf.edu

    Fax: +001 415 476 8173

Contents

1. Structure Design

1.1 Bicyclic CovCore protein

***2H-2 design*.** The helix-loop-helix motif from the crystal structure of the Rop protein(1) (PDB ID 4DO2) was used as the starting template for the design of **2H-2** (Fig. 2). First, the anti-parallel helical dimer (Ala12-Ser51 fragment) was extracted and duplicated (shown in Fig. 2). Then, to build the symmetric and cyclic structure of **2H-2**, we superposed the backbone atoms of the residues 13-20 (blue helix in Fig. 2) from the first fragment onto residues 39-46 (red helix in Fig. 2) of the second fragment(2). The RMSD between the two superposed helical fragments was smaller than 0.5 Å, allowing us to easily merge both fragments easily (purple helix).

1.2 Tricyclic CovCore protein 3H1

***Development of New Rosetta Modules*.** We implemented two new Rosetta modules that were first used for this project. Although these will be described in greater detail in other publications, we provide a brief overview here.

***Parametric Design Tools*.** Parametric design approaches generate molecular geometry using generating expressions, the parameters of which can be varied to sample variant conformations. For protein design, expressions developed by Francis Crick, describing helices following a superhelical path through space, have been used previously to generate helical bundle (coiled-coil) topologies (3-5). Initial forays into parametric design used Python scripts to generate crude C_α traces from the Crick equations. These traces were subsequently imported into Rosetta and converted to full-atom structures by aligning fragments of protein geometry to the traces and using the Rosetta Cartesian-space energy minimization algorithm to pull the fragments into place. We have added parametric geometry generation capabilities to the Rosetta software suite, permitting much more rapid backbone generation without any computationally-expensive minimization steps. A module called the `BundleGridSampler` has been written to automate the enumeration of conformations corresponding to regular N-dimensional grids of parameter values. This module interfaces with the `rosetta_scripts` application, permitting automatic parallelization of sampling, which enables very rapid exploration of large numbers of conformations on large computer clusters such as the University of Washington Hyak supercomputer. The speedup, parallelization, and omission of slow steps involving writing C_α traces to disk permit us to sample five to six orders of magnitude more backbone conformations than we could using the Python-based workflow. In addition, accessibility to the `rosetta_scripts` application permits parametric design to be used in a single continuous protocol with other Rosetta modules for new applications that may or may not involve helical bundles. The work presented here is one such example of a parametric workflow that is not intended to produce coiled-coil topologies.

***Fragment Lookup Filter*.** To assess viability of non-parametrically generated backbone segments rapidly, we used a nonparametric model of coordinate root-mean-square deviation from polypeptide backbone conformations within the PDB. This model evaluates all overlapping fixed-length backbone fragments within the model, considering any fragments with observed PDB conformations within a fixed RMSD threshold "valid" and those without observed neighbors "invalid". The query fragment length and RMSD neighbor threshold radius are application-specific tuned parameters. Fragment comparison is implemented *via* an optimized superposition algorithm.(6)

Given the size of the observed fragment space, $\sim 5 \times 10^6$ fragments within the PDB, this nonparametric model is significantly more expensive to evaluate than Rosetta's existing statistical score terms, precluding a naïve implementation. To reduce the working fragment set size, an iterative K-centroids clustering algorithm was used to generate representative fragment sets covering the observed fragment space at fixed RSMD tolerances. This dramatically reduced both score-time memory

requirements and scoring CPU cost, as the majority of observed fragments clustered around highly favorable secondary structure conformations.

For evaluation of candidate KIC solutions, a library of four-residue fragments clustered to a 0.33 Å RSMD radius was used, in which any candidate fragment within 0.33 Å of a cluster centroid is considered "valid". KIC solutions containing any invalid fragments were discarded. The filtering method is implemented in the Rosetta software suite as the `FragmentLookupFilter`, which is accessible to the `rosetta_scripts` application. The clustered fragment set used, which is a needed input for this filter, is included compressed in the file `backbone_profiler_database_06032014.tgz`.

C3-Symmetric Polypeptide Design Workflow. Triangular polypeptides with C3 symmetry were designed using the Rosetta software suite and a single continuous protocol developed using the RosettaScripts scripting language.(7) As shown in **Fig. 4** in the main text, the design protocol started with a single parametrically-generated α-helix created using the Crick parameterization described previously.(3-5) Although the Crick parameterization describes coiled-coil geometry, for short helices and small values of the twist parameter, ω_0, the parameterization provides a convenient way to sample conformations of a helix that is offset from, and slightly inclined relative to, the Z-axis.

We made use of Rosetta's parametric grid-sampling tool, the `BundleGridSampler`, to sample a regular 4-dimensional grid of Crick parameter values corresponding to a range of conformations of a 20-amino acid α-helix. The parameter ranges sampled are shown in **Table S1**.

For each helical conformation sampled, we mutated a single amino acid residue in the middle of the helix to a sidechain-conjugated cysteine variant (called "CYX" in Rosetta), and appended one-third of the TBMB cross-linker to the side-chain of this CYX residue. We next set up C3 symmetry by triplicating the helix, with its appended TBMB fragment, with rotation about the Y-axis. Rosetta's symmetry machinery is able to enforce symmetry throughout a protocol, propagating changes to one asymmetric subunit to all other subunits during subsequent design or energy-minimization steps, so from this point forward, the structure remained C3-symmetric.(8) At this point, we filtered based on geometric clashes, discarding those structures which, when symmetrized, produced sterically impossible geometry. For those structures passing this filter, we then added constraints linking the TBMB fragments together and enforcing planarity of the TBMB phenyl ring, then used the symmetric Rosetta gradient-descent energy minimization algorithm (called the minimizer) to pull the TBMB fragments together and to orient the CYX side-chain while keeping the backbone rigid and immobile. Structures in which the TBMB failed to close (*i.e.* for which the backbone geometry was incompatible with proper presentation of the TBMB linker) were discarded at this point. Since these initial steps took on the order of a CPU-second (most of which was needed for the minimization step), the majority of the 311,600 samples could be considered and discarded with low computational cost, prior to attempting any computationally expensive sequence design.

At this point, we used Rosetta's generalized kinematic closure (`GeneralizedKIC`) module(9) to attempt to link the C-terminus of one symmetry repeat with the N-terminus of the next. Given a chain of atoms with fixed start and end points, `GeneralizedKIC` allows the user to perturb all but six degrees of freedom of the chain in any of many ways, then solves analytically for the values of the six remaining degrees of freedom in order to ensure that the chain remains closed. In this case, four types of perturbation were used. First, the bond distance between the terminal carbonyl carbon of one chain and the terminal amide nitrogen of the next was set to the ideal value for a peptide bond (1.32 Å). Second, the terminal C_α-C-N and C-N-C_α bond angles were set to ideal values for a peptide bond (114º and 123º, respectively). Third, the ω dihedral angle between the two lobes was set to 180º. Finally, backbone φ and ψ dihedral angles were drawn randomly, biased both by the Ramachandran conformational

preferences of L-amino acids and by the relative probabilities, based on Protein Data Bank statistics, of observing an amino acid residue at position i in a particular region of Ramachandran space (a Ramachandran "bin") and a residue at position $i+1$ in another Ramachandran bin. We divided Ramachandran space into four bins, corresponding to α-helices and β-strand conformations (A and B bins, respectively, with ψ < 0º), and the equivalent bins with φ and ψ values inverted (A' and B' bins, respectively, with ψ ≥ 0º). A single `GeneralizedKIC` closure attempt can return anywhere from 0 to 16 closed solutions, and 100 attempts were made per design trajectory.

For each closure solution, a low-cost protocol was applied. First, closure solutions that clashed internally or with the rest of the structure were discarded. Second, closure solutions that did not resemble loop conformations observed in the Protein Data Bank, based on comparison to a library of 4-residue fragments, were also discarded using the `FragmentLookupFilter` described above. For those loops remaining, we carried out a rapid round of sequence design, using a scoring function variant with the repulsive part of the Lennard-Jones term down-weighted. For this and all subsequent design steps, designable positions were divided into "core", "boundary", and "surface" groups based on side-chain solvent-accessible surface area. We used Rosetta's `LayerDesign` method(7) to restrict the amino acids allowed in each group: hydrophobic and small aliphatic types for "core" positions (Phe, Ala, Met, Ile, Leu, Tyr, Val, and Trp); small aliphatic types (excluding alanine), tyrosine, proline, and uncharged polar types for "boundary" positions (Ile, Leu, Tyr, Val, Thr, Ser, Asn, Gln, and Pro), and all polar types plus proline for "surface" positions (Asp, Glu, Arg, Lys, Thr, Ser, Asn, Gln, and Pro). We then energy-minimized side-chains using the unmodified talaris2014 scoring function. The top solution from 100 attempts was selected by `GeneralizedKIC` for subsequent refinement.

At this point, trajectories that returned no loop closure solutions were aborted. Those remaining were subjected to several rounds of more computationally-expensive sequence and structure refinement. We made use of the `FastDesign` protocol, which performs many alternating rounds of sequence design and energy minimization, gradually ramping the repulsive part of the Lennard-Jones term in the scoring function from round to round. We carried out 1 repeat (4 design-and-minimization rounds) of `FastDesign` keeping the backbone rigid and immobile, and 3 additional repeats (12 design-and-minimization rounds) of `FastDesign` with full backbone minimization. Amino acid identities were restricted by layer as described above.

A total of 311,600 parametric arrangements of helices were sampled, resulting in 3,881 conformations able to present TBMB, and 3,881 designed sequences. We chose the most promising candidates by Rosetta energy and by manual inspection, and used Rosetta's *ab initio* structure prediction algorithm for final computational validation. Only those designs with funnel-shaped energy landscapes, even in the absence of the head-to-tail cyclization or the TBMB cross-linker (which the *ab initio* algorithm cannot model) were considered as final candidates for synthesis. Since the *ab initio* algorithm did not take into account these stabilizing features, this served as a particularly stringent test, likely resulting in many false negatives.

The RosettaScripts XML that defines the protocol is shown in **Listing S1**. After finding initial "designable" regions of parameter space producing plausible-looking designs, we re-ran the script using tighter parameter ranges in these regions to increase the diversity of designs produced.

Note that we have subsequently added a new Rosetta module, called the `CrosslinkerMover`, which provides a simpler method of carrying out the steps performed by the `add_linker`, `linkerbond1`, `linkerbond2`, `linkerbond3`, `constrain_linker`, `frelax_linker`, and `bad_linker_filter` movers and filters in **Listing S1**. Please refer to the Rosetta help documentation

at http://www.rosettacommons.org for more information, or run the `rosetta_scripts` executable with the flag `-info CrosslinkerMover` to get a description of this mover's usage.

The scoring function used for design was a slight modification of talaris2014, the Rosetta scoring function that was default at the time. (Note that a new energy function, called ref2015, has subsequently become the default in Rosetta.) The reference energies for methionine, alanine, and tryptophan were raised by 2.5, 2.5, and 1.0 energy units, respectively, in order to discourage excessive numbers of amino acids of these types from being added. Similarly, the reference energy for proline was lowered by 1.5 energy units in order to encourage placement of proline, particularly in loops. **Listing S2** shows the weights file for the modified talaris2014 scoring function.

Symmetry is handled internally in Rosetta by the `SymmetricConformation` object, which automatically propagates changes to one asymmetric subunit in a structure to the entire symmetric structure.(8) Scoring is simplified by calculating internal interactions within an asymmetric subunit and inter-subunit interactions across a single interface, and then multiplying the calculated energy by the appropriate factor to derive that for the whole structure. C3 symmetry requires a symmetry definition file, shown in **Listing S3**.

When the TBMB linker is placed, the default talaris2014 scoring function has no knowledge of the torsional preferences of the new bonds; nor is it able to enforce reasonable bond distances and bond angles between the TBMB and the CYX side-chain. Similarly, constraints are needed to enforce reasonable peptide bond geometry between the asymmetric subunits of the polypeptide during backbone energy-minimization steps. Constraints files, listed in **Listing S4** and **S5**, are used for this purpose. Note that the CYX side-chain possess a "virtual atom" attached to the SG atom, and TBMB similarly possesses a number of "virtual atoms". By constraining these to lie atop real atoms (*i.e.* to have atom-pair distances of 0 between the virtual and a real atom), it is possible to enforce atom-pair distance constraints and bond angle constraints simultaneously, reducing manual setup of constraints.

The Foldtree is Rosetta's representation of the kinematic relationships between parts of a molecule. During minimization, rotation of dihedral angles moves atoms that are downstream in the Foldtree, and keeps fixed atoms that are upstream. Connections can either be through bonds or by rigid-body transformations through space (which we call "jumps"). For our purposes, we set up a Foldtree rooted on the center CYX residue of a helix, with through-bond connections to both ends of the helix and jumps from the CYX to the TBMB fragment. The Foldtree definition is shown in **Listing S6**. Note that the Foldtree is set up before symmetrization, and it persists for the asymmetric unit of the structure after symmetrization. The symmetry machinery adds virtual atoms, to which the asymmetric units are parented *via* jumps, and which in turn have a single virutal atom as the root of the symmetric Foldtree.

Sequence design is carried out using Rosetta's "packer" module, which optimizes side-chain rotamers and side-chain identities given a list of possible rotamers and identities at each position and a fixed backbone. Three packer definition files ("resfiles") are necessary for all design steps in the protocol used: one defining amino acids that are allowed at buried positions, one defining amino acids allowed at semi-buried positions, and one defining amino acids allowed at solvent-exposed positions. These are shown in **Listings S7**, **S8**, and **S9**, respectively.

In order to execute the script shown in **Listing S1**, the files `design.xml` (**Listing S1**), `talaris2014_modified.wts` (**Listing S2**), `foldtree1.txt` (**Listing S6**), and the three resfiles (**Listings S7**, **S8**, and **S9**) must be placed in an `xml/` sub-directory of the working directory. The symmetry definition file, `c3_y.symm` (**Listing S3**) must be placed in a `symmetry/` sub-directory of the working directory, and the constraints files (**Listings S4** and **S5**) must be placed in a

`constraints/` sub-directory of the working directory. The contents of `backbone_profiler_database_06032014.tgz` must be decompressed in the working directory (where it will create a sub-directory with the same name). A Rosetta flags file providing commandline flags, shown in **Listing S10**, must also be added to the `xml/` directory. Note that these flags include an "nstruct" option. This flag, which specifies the number of independent replicates of the RosettaScripts protocol that will be run, must be set to be equal to the total number of samples performed by the `BundleGridSampler` (311,600, in our case). The `BundleGridSampler` will divide its samples over independent RosettaScripts trajectories automatically.

A quirk of RosettaScripts is the requirement for an input file (in PDB format, FASTA format, or Rosetta's proprietary "silent file" format). The `BundleGridSampler` mover disregards the input geometry and generates new geometry, but despite this, the requirement for input geometry remains. For this reason, a "dummy" FASTA file must be placed in the `xml/` sub-directory and provided to the `rosetta_scripts` application, and this is listed in **Listing S11**.

Once all of these files are in place, the command to run the `rosetta_scripts` application is shown in **Listing S12**. Note that, depending on your operating system and compiler, the "linuxclangrelease" part of this command may need to be replaced with an appropriate alternative (*e.g.* "rosetta_scripts.default.macosclangrelease", "rosetta_scripts.default.linuxgccrelease", *etc.*).

The `rosetta_scripts` application may also be compiled and run in MPI mode, which uses the Message Passing Interface (MPI) for efficient parallel execution of independent RosettaScripts trajectories. Since the `BundleGridSampler` automatically divides its samples over different RosettaScripts trajectories, and these trajectories will be divided automatically over available CPUs in MPI mode, this is an easy way to parallelize the sampling of Crick parameter space on large HPC clusters. **Listing S13** shows the commandline execution of the MPI-mode `rosetta_scripts` application. Please refer to the Rosetta web documentation for information on compiling the `rosetta_scripts` application with MPI support.

1.3 Tricyclic CovCore protein 3H2

The **3H2** protein was redesigned based on coordinates of the lowest energy conformer of the calculated **3H1** NMR structures. We first searched the PDB for close structural matches (< 1 Å) for the first and last four residues of each helix (e.g. [16]DDSS[19] and [23]PEAE[26]). To perform the search, we used backbone coordinates of these helices as a query in the search algorithm MASTER(10) to a non-redundant set of the structures from the Protein Data Bank created as described previously(11). Briefly, structures were chosen to have less than 30% sequence identity, R-factors less than 0.25, and resolution better than 3.2 Å from among 18,306 isolated protein chains from x-ray crystal structures downloaded 01/29/2016. This search identified that 3 residues is the most common length of loop bridging terminal helical amino acids in the geometry observed in the 3H1 NMR structure. These loop conformations were clustered using a semi-greedy mini-batch k-medoids algorithm, identifying the most commonly observed loop conformation. The loop coordinates were grafted onto positions 20-22 (ignoring the symmetric portions of the structure) after alignment to the corresponding helical termini. Amino acids frequency patterns were calculated from the unique sequences that fold into this particular loop conformation, shown as a logo plot in **Fig. S1**. The most common loop sequence, excluding those with proline, was NGD followed by GGD.

To redesign the sequence, we used Rosetta's fixed backbone side chain design application, `fixbb`, (12) with a limited amino acid alphabet at each position, shown in **Table S2**. (The `fixbb` application is documented here: https://www.rosettacommons.org/docs/latest/application_documentation/design/fixbb). The grafted loop coordinates were minimized with the C-alpha atom position restrained (limiting, but not completely

prohibiting, backbone motion) prior to rotamer trials. At solvent-exposed positions, we selected amino acids to redesign to reduce the protein's charge and surface-exposed side-chain entropy. We redesigned Lys, Arg, and Glu positions, but continued to allow Lys, Asp, and Glu where potential inter-helical salt-bridges could form. At loop positions, amino acids were selected based on their presence in the frequency distributions of amino acids observed in the PDB for the native loop structure. The hydrophobic core was allowed to sample apolar amino acids. We perfomed 50,000 side chain design trajectories, with 31,093 models passing a filter that penalizes poor side chain packing and protein void space; these models had a "packstat" score greater than 0.6.(13) These 31,093 trajectories converged to only 9 unique sequences. The most frequently observed sequence was (NGDKAEEIARCKKLLEALKS)[3] (25,683 trajectories, with the greatest packstat score at 0.821). The loop sequence NGD was output unanimously in the ROSETTA design trajectories, consistent with our bioinformatics search. **Table S2** shows the lowest scoring sequence, along with the amino acids also observed at their respective positions in the other 8 sequences. Various solvent exposing residues were further changed to less hydrophilic amino acids to reduce the solubility of the redesigned protein molecule.

Due to the expected difficulties of -Asn-Gly- aspartimide formation during synthesis, Asn was altered to Gly in the final sequence for the redesign of **3H1**. This change in sequence was consistent with the second most common amino acid combination for the loop conformation determined bioinformatically above, *i.e.* –GGD- versus –NGD-. As well, several surface exposed positions were manually mutated from charged amino acids to apolar or polar uncharged amino acids. Those changes and the final **3H2** sequence are shown in **Table S2**. The final sequence identity between **3H1** and **3H2** was 45%.

2. Experimental Section

2.1 General information

Reagents. All reagents were used without further treatment. Fmoc-protected amino acids were purchased from GL Biochem. N-(9-Fluorenylmethoxycarbonyloxy)succinimide (Fmoc-OSu), Boc-Cys(Trt)-OH, Boc-Thz-OH, 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU), trifluoroacetic acid (TFA), and hydroxybenzotriazole hydrate (HOBt) were purchased from Chem-Impex International. O-Methylhydroxylamine Hydrochloride and 1,3,5-tris(bromomethyl)benzene (TBMB) were purchased from TCI america. 4-methylpiperidine were purchased from Acros Organics. Rink Amide-ChemMatrix resin (0.5 mmol/g loading) was purchased from Biotage. All other reagents including 1,3-Bis(bromomethyl)benzene(BBMB), 4-Mercaptophenylacetic acid (MPAA), 2-mercaptoethanesulfonate (MESNa), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), DL-Dithiothreitol (DTT), hydrazine, tetrakis(triphenylphosphine)palladium(0), borane dimethylamine complex, and p-nitrophenyl chloroformate (p-NPC) were purchased from Sigma-Aldrich, SYPRO™ Orange(S6650) was purchased from Thermo Fisher.

Reverse-phase HPLC and Mass Spectrometry. Analytical reverse-phase HPLC analyses were performed on an Agilent 1100 series HPLC system using a Phenomenex Aeris WIDEPORE 3.6 μm C4, 250 x 4.6 mm column. Chromatographic separations were obtained using a linear gradient of 5-45% acetonitrile (with 0.08% TFA) in water (with 0.1% TFA) over 40 min, with column at room temperature. Flow rates were controlled at 1.0 mL/min. Peptide detection was based on UV absorption at 220 nm, and mass spectrometry data were obtained using Shimadzu AXIMA Performance MALDI-TOF spectrometer in a linear mode with sinapinic acid as matrix or reflectron mode with α-Cyano-4-hydroxycinnamic acid as matrix. LC-MS data were collected on Applied Biosystems 3200 Q TRAP coupled with Shimadzu UFLC system.

Preparative HPLC. Products from solid-phase peptide synthesis and polypeptide products from ligation reactions were purified using a Phenomenex Jupiter 5.0 μm C4, 250 x 10.0 mm column. A shallow gradient of acetonitrile (with 0.08% TFA) versus water (with 0.1% TFA) was designed for each peptide based on its elution characteristics. Flow rates were controlled at 5 mL/min. Fractions containing the desired pure peptide were identified by analytical HPLC and mass spectrometry, then combined and lyophilized.

2.2 Synthesis of bicyclic CovCore protein (2H).

Fmoc-Dbz-OH Synthesis. To a solution of 3,4-diaminobenzoic acid (1 g) and NaHCO$_3$ (3 eq., 1.65 g) in 50 mL water was added Fmoc-OSu (1 eq., 2.1 g in 50 mL CH$_3$CN). The dropwise addition of Fmoc-OSu was completed within 1 h, then the reaction mixture was allowed to stir at room temperature overnight. Upon completion, CH$_3$CN was removed by rotary evaporation. 6 M HCl was subsequently added to adjust the solution pH to 2. White solid was collected by filtration, followed by washing with H$_2$O and drying. The material was directly used in the next step of the reaction.

Fmoc-Dbz(Alloc)-OH Synthesis. Fmoc-Dbz-OH (1g) and NaHCO$_3$ (3 eq., 1.65 g) was suspended in 20 mL H$_2$O/THF (3:1, v/v). Allyl chloroformate (1.2 eq., 0.34 mL) in 5 mL THF was added dropwise. The reaction mixture was allowed to stir at room temperature for 5 h. 6 M HCl was then added to adjust the pH to 2. The mixture was extracted with ethyl acetate (10 mL, 3 times). The organic layer then was combined, washed with brine, and dried over anhydrous Na$_2$SO$_4$. The crude product was obtained after rotary evaporation. The product was purified by flash column (0% to 10% methanol in dichloromethane [DCM]) to give a white solid (0.85 g, 75% yield). **¹H NMR** (DMSO-d$_6$, 900 MHz): δ 12.880 (s, 1H), 9.224 (s, 1H), 9.129 (s, 1H), 8.098 (s, 1H), 7.909 (d, J = 7.2 Hz, 2H), 7.785-7.742 (m, 4H), 7.427 (t, J = 7.2 Hz, 2H), 7.346 (t, J = 7.2 Hz, 2H), 6.015-5.972 (m, 1H), 5.395 (d, J = 18 Hz, 1H), 5.257 (d, J = 9.9

Hz, 1H), 4.650 (d, J = 5.4 Hz, 2H), 4.471 (s, 2H), 4.332 (t, J = 6.3 Hz, 1H). ^{13}C NMR (DMSO-d$_6$, 225 MHz): δ 166.68, 154.22, 154.14, 153.44, 153.35, 143.73, 140.83, 140.81, 133.17, 133.03, 133.02, 127.75, 127.14, 126.05, 125.24, 125.21, 120.22, 117.91, 117.90, 65.29, 65.17, 46.65.

Solid-phase peptide synthesis (SPPS). All peptides were synthesized on a 0.1 mmol scale using a Biotage Initiator+ Alstra peptide synthesizer. Rink Amide-ChemMatrix resin (Biotage, 0.5 mmol/g loading) was used for the synthesis. Fmoc-Dbz(Alloc)-OH or Fmoc-Dbz-OH was first attached on resin as the C-terminal residue. A typical SPPS reaction cycle includes Fmoc deprotection, washing, coupling, and post-coupling washing steps. The deprotection was carried out for 5 min at 70 °C with 4.5 mL 20% 4-methylpiperidine in dimethylformamide (DMF). A standard coupling step (for all amino acids except histidine and cysteine) was done for 5 min at 75 °C with 5 equivalents Fmoc-protected amino acids, 4.98 equivalents HCTU, and 10 equivalents DIEA (relative to the amino groups on resin) in DMF at a final concentration of 0.125 M amino acids. Cysteine coupling was conducted at 50 °C to minimize racemization.

Alloc deprotection. In the case of Fmoc-Dbz(Alloc)-OH containing syntheses, after SPPS, the resin was washed with DCM three times. The Alloc deprotection reaction was carried out by treating resin with Pd(PPh$_3$)$_4$ (0.2 eq.) and (CH$_3$)$_2$NH·BH$_3$ (6 eq.) in 4 mL dry DCM. The reaction was conducted at room temperature for 15 min under a N$_2$ atmosphere. The deprotection step was repeated to ensure completion.

Dbz activation. The activation of Dbz followed a published procedure.(14) The resin was first treated with p-nitrophenyl chloroformate (p-NPC, 5 eq.) in 4 mL DCM for 1 h at room temperature. After washing with DCM and DMF, 0.5 M DIEA in DMF (5 mL) was added and the resin was mixed for 30 min.

Peptide cleavage. Resin was washed with DCM and dried before cleavage. Peptide cleavage was carried out in 10 mL TFA/H20/TIS/EDT (95:2:2/1, v/v/v/v) for 2 h at room temperature. The crude peptide was obtained after precipitation in cold diethyl ether.

Head-to-tail cyclization. The linear peptide was dissolved in a pH 7.0 buffer (6 M Gu•HCl, 200 mM Na$_2$HPO$_4$, 100 mM MPAA, and 20 mM TCEP) at a concentration of 0.4 mg/mL. The mixture was reacted at room temperature overnight. DTT was then added to a final concentration of 100 mM. After 1 h, TFA was added to adjust the pH to 2.

Peptide crosslinking. After purification and lyophilization, the corresponding cyclic peptide was dissolved in a 1:1 (v/v) mixture of CH$_3$CN and 100 mM NH$_4$HCO$_3$ solution (pH 8) at a concentration of 0.5 mg/mL and then treated with TCEP (1.1 eq.) for 1h at room temperature. BBMB (1.5 eq. in 100 μL DMF) was added. The reaction progress was monitored by analytical HPLC every 2 h. Upon completion, TFA was added to adjust the pH to 2-3.

2.3 Synthesis of tricyclic CovCore protein (3H1, 3H2).
Solid phase peptide synthesis was performed as described above. Fmoc-Dbz-OH was used in this case.

Peptide Hydrazide preparation. Crude peptide Thz11-Arg30-Nbz (9.0 mg) was dissolved in 3 mL premade buffer (6 M Gu•HCl, 0.2 M Na$_2$HPO$_4$, 1.5 M Hydrazine, pH 8.0), after 20 minutes stirring at room temperature, MeONH$_2$•HCl (78 mg) was added and pH was adjusted to 4.0. The reaction was left for 4 h to deprotect N-terminal thiazolidine to afford Cys11-Arg30-NHNH$_2$.

Thz11-Arg30-COSCH₂CH₂SO₃ thioester preparation. Crude peptide Thz11-Arg30-Nbz (9.0 mg) was dissolved in 3 mL premade buffer (6 M Gu•HCl, 0.2 M Na_2HPO_4, 0.2 M MESNa, pH 7.1), the pH of the solution was readjusted to ~7.0 and it was stirred at room temperature for 3 h before purification.

Linear polypeptide synthesis. The synthesis of the **3H1** linear polypeptide was carried out through one-pot reactions (**Fig. S3**). Native chemical ligation was performed by dissolving Cys11-Arg30-*NHNH₂* (5.0 mg, 2.2 μmol) and Thz11-Arg30-Nbz (6.7 mg, 2.7 μmol) in ligation buffer (1.0 mL) containing 6 M Gu•HCl, 200 mM Na_2HPO_4, 100 mM MPAA and 30 mM TCEP hydrochloride at pH 7.0; reaction for 7 h at room temperature gave ligation product. The Thz-peptide was converted to the Cys-peptide by addition of 50 mM MeONH₂•HCl at pH 4.0 and reaction for 24 h at room temperature. Then peptide Thz11-Arg30-*COSCH₂CH₂SO₃* (7.9 mg, 3.3 μmol) was added to the reaction mixture, and the solution was adjusted to pH 7.0 by careful addition of 5 N NaOH. After 4 h reaction at room temperature, 200 mM methoxylamine hydrochloride was added with pH adjustment to 4.0 to convert N-terminal Thz to Cys, solid phase extraction was performed in the end to afford crude **3H1**-*NHNH₂* (~10 mg).

Macrocycliztion. Crude linear peptide **3H1**-*NHNH₂* (10.0 mg, 1.5 μmol) was dissolved in 4 mL buffer (6 M Gu•HCl, 200 mM Na_2HPO_4, pH 3.0), the solution was cooled to -15 °C for 15 minutes using ice salt bath. Then 60 μL of 0.4 M $NaNO_2$ was added to oxidize hydrazide for 15 minutes at -15 °C. 4 mL Ligation buffer (6 M Gu•HCl, 200 mM Na_2HPO_4, 0.2 M MPAA, pH 7.0) was added, the solution pH was later readjusted to 7.0 for macrocyclization. The cyclization reaction was stirred at room temperature for 6 h before TCEP•HCl was added to reduce any disulfides back to free cysteines.(15) Solid phase extraction was then performed to give crude cyclized peptide (~3.5 mg).

Crosslinking. In a typical cyclization reaction, 3.5 mg of crude cyclized polypeptide was dissolved in water (7 mL) and acetonitrile (2.3 mL). 280 μL Ammonium bicarbonate buffer (1 M, pH 8.1) and TBMB (98 μL, 2.4 mg/mL) were then added to initiate the reaction. (16) The reaction was complete within 1 h. Semi-prep purification was then performed to give ~3 mg cyclized and crosslinked product.

3H2 was prepared using the same method as **3H1.**

For the preparation of crosslinked but not cyclized **3H1** and **3H2**, linear polypeptide was reacted with TBMB to give the product. No backbone cyclization reaction was performed here.

2.4 Biophysical characterizations

Circular dichroism (CD) spectroscopy. CD spectra were recorded on a Jasco J-810 CD spectrometer using 1 mm cuvettes with the following parameters: 1 nm data pitch, 20 nm/min scanning speed, 4s response, 2 nm bandwidth, and 3 accumulations. Samples were prepared in phosphate buffer (10 mM, pH 7.4) or acetate buffer (10 mM, pH 4.0). Peptide concentrations were determined by weight considering the TFA salt and 10% water contents. CD melting experiments were performed with temperature range 5 °C to 95 °C with 5 °C increments for **2H** and **2H-5** and 25 °C to 95 °C with 5 °C increments for **3H1** and **3H2** monitored at 222 nm**.**

Analytical ultracentrifugation. Analytical ultracentrifugation of **3H2** was carried out by equilibrium sedimentation performed at 25 °C using a Beckman XL-I analytical ultracentrifuge. **3H2** solutions were prepared at 1.0 mM in phosphate buffer (25 mM, pH 7.4). Centrifugation was conducted at speeds of 25K, 30K, 35K, 40K and 45 K r.p.m. with an An-60 Ti rotor and the radial gradient profiles were acquired by absorbance scans at 254 nm. Data were globally fitted to equilibrium sedimentation models of a single-species and/or monomer-dimer mixtures by a nonlinear least-squares method using Igor Pro (WaveMetrics).

NMR spectroscopy. NMR spectra were recorded on a Bruker Avance II 900 MHz spectrometer equipped with a cryogenic probe. Samples were prepared at a concentration of 0.7 mM in a 50 mM acetate buffer (pH 4.0, 5% D_2O). 1D and 2D spectra were recorded with standard pulse sequences. Backbone and side-chain chemical shifts were determined from 2D ^1H-^1H TOCSY and ^1H-^1H NOESY spectra. A 2D ^1H-^1H NOESY experiment was carried out with mixing time of 100 ms, $t_{1,max} = 23.4$ μs, and $t_{2,max} = 147$ μs (16 scans). A 2D ^1H-^1H TOCSY experiment was carried out with mixing time of 80 ms, $t_{1,max} = 23.4$ μs, and $t_{2,max} = 147$ μs (16 scans). The ^1H carrier frequency was set at the water frequency, and chemical shifts were referenced to the water frequency. All spectra were processed using NMRPipe.(17) Prior to Fourier transform, time-domain data were multiplied by sine square bell window functions shifted by 90 degrees and zero-filled twice. Time domain data of indirect dimensions were extended by linear prediction only using the first third of the data points. Spectra were analyzed with XEASY(18) and visualized by Sparky.(19)

Distance restrains derived from cross-peak intensities in NOESY spectra were used for the structural calculation of **2H** and **2H-4**. The statistics was summarized in **Table S4**. NOE calibration was done using CYANA.(20) Structures were calculated using XPLOR-NIH.(21, 22) The 20 lowest-energy structures were obtained with the restrained simulated annealing protocol and then used for further explicit water refinement (**Fig. S7**). **2H** showed a symmetric structure, since it had one set of resonances (*i.e.* equivalent residues in the two C2-symmetric halves were in identical chemical environments). Thus, protons of only 19 (out of 38) residues needed to be assigned, and **2H** was treated as a homo-dimer in the structural calculation. Non-crystallographic symmetry (NCS) was used for the refinement. The packing between two helices within **2H** was defined by unambiguous inter-monomer NOEs (**Fig. S7D**). For **3H1**, ^{13}C assignments were obtained by natural ^{13}C HSQC. In addition to distance constraints, backbone and χ_1 dihedral angle constraints were derived from chemical shifts using the program TALOS(23) for residues located in well-defined secondary structure elements. NMR structure quality was assessed with the Protein Structure Validation Software Suite (PSVS)(24) (**Table S4**).

The backbone order parameters (S^2) plot. The ^{13}C chemical shifts were obtained from natural abundance 2D [^{13}C,^1H]-HSQC. The 2D HSQC spectrum comprising aliphatic signals was recorded along $t_1(^{13}C)$ and $t_2(^1H)$ with, respectively, 72 and 4096 complex points, respectively, and $t_{1,max}(^{13}C) = 8$ ms and $t_{2,max}(^1H) = 189$ ms with 128 scans. The ^{13}C carrier frequency was set to 45 ppm. The order parameters S^2 were predicted by Talos-N[23]. Usually, the residues with S^2 greater than 0.8 are in a rigid structure without backbone motions of significant amplitude on a picosecond timescale (**Fig. S9**).

X-ray crystallography. Initial crystal screens were set up using TTPLabtech Mosquito Nanoliter Dropsetters. Crystallization of **3H2** was attempted using a screen of 288 conditions from the Hampton Index (HR2-144), AmSO4 screen and Wizard I &II screen. Purified **3H2** was dissolved in water to a concentration of 12 mg/mL. Each condition used 0.1 mL of well solution with hanging drops consisting of 0.1 μL well solution and 0.1 μL protein solution. The plates were kept in an incubator at 20° C. Two conditions (0.1 M MES, pH 6.5, 2.0 M (($NH_4)_2SO_4$, 5 %w/v PEG 400) and (0.1 M HEPES, pH 7.5,1.6 M ($NH_4)_2SO_4$, 2 %w/v PEG 1K) gave crystals within one week. We then grew crystals in a 24 well plate using 1 mL well solution and 2 uL hanging drop solution to obtain bigger crystals. Diffraction data were collected at the Advanced Light Source beam 8.3.1 at 100K with a 15 keV beam and a detector-to-sample distance of 125 mm. Diffraction was observed to a resolution of 1.20 Å. Diffraction intensity statistics revealed that **3H2** crystallized in space group P1, with six molecules in the asymmetric unit. The beam size was 100 μm for all data sets; the detector used was a Dectris Pilatus3 S 6M. Because each **3H2** was expected to contain three helices, we used Arcimboldo(25) to search for 18 copies of a 14-residue helix. An overnight run on a 3.7 GHz intel core with i7-4820K CPU gave a direct-method solution. The models were then manually built to completion. The statistics was summarized in **Table S5**.

References

1. Amprazi M, *et al.* (2014) Structural plasticity of 4-alpha-helical bundles exemplified by the puzzle-like molecular assembly of the Rop protein. *Proc Natl Acad Sci U S A* 111(30):11049-11054.
2. Engel DE & DeGrado WF (2005) Alpha-alpha linking motifs and interhelical orientations. *Proteins* 61(2):325-337.
3. Crick FH (1953) The Fourier Transform of a Coiled Coil. *Acta Cryst* 6.
4. Grigoryan G & Degrado WF (2011) Probing designability via a generalized model of helical bundle geometry. *J Mol Biol* 405(4):1079-1100.
5. Huang PS, *et al.* (2014) High thermodynamic stability of parametrically designed helical bundles. *Science* 346(6208):481-485.
6. Theobald DL (2005) Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallogr A* 61(Pt 4):478-480.
7. Fleishman SJ, *et al.* (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* 6(6):e20161.
8. DiMaio F, Leaver-Fay A, Bradley P, Baker D, & Andre I (2011) Modeling symmetric macromolecular structures in Rosetta3. *PLoS One* 6(6):e20450.
9. Bhardwaj G, *et al.* (2016) Accurate de novo design of hyperstable constrained peptides. *Nature* 538(7625):329-335.
10. Zhou J & Grigoryan G (2015) Rapid search for tertiary fragments reveals protein sequence-structure relationships. *Protein Sci* 24(4):508-524.
11. Zheng F, Zhang J, & Grigoryan G (2015) Tertiary structural propensities reveal fundamental sequence/structure relationships. *Structure* 23(5):961-971.
12. Kuhlman B, *et al.* (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302(5649):1364-1368.
13. Sheffler W & Baker D (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* 18(1):229-239.
14. Blanco-Canosa JB & Dawson PE (2008) An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in native chemical ligation. *Angew Chem Int Ed Engl* 47(36):6851-6855.
15. Fang GM, *et al.* (2011) Protein chemical synthesis by ligation of peptide hydrazides. *Angew Chem Int Ed Engl* 50(33):7645-7649.
16. Heinis C, Rutherford T, Freund S, & Winter G (2009) Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat Chem Biol* 5(7):502-507.
17. Delaglio F, *et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293.
18. Bartels C, Xia TH, Billeter M, Guntert P, & Wuthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 6(1):1-10.
19. Lee W, Tonelli M, & Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31(8):1325-1327.
20. Guntert P, Braun W, & Wuthrich K (1991) Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. *J Mol Biol* 217(3):517-530.
21. Schwieters CD, Kuszewski JJ, & Clore GM (2006) Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Mag Res Sp* 48(1):47-62.
22. Schwieters CD, Kuszewski JJ, Tjandra N, & Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65-73.

23.   Shen Y & Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227-241.
24.   Bagaria A, Jaravine V, Huang YJ, Montelione GT, & Guntert P (2012) Protein structure validation by generalized linear model root-mean-square deviation prediction. *Protein Sci* 21(2):229-238.
25.   Millan C, Sammito M, & Uson I (2015) Macromolecular ab initio phasing enforcing secondary and tertiary structure. *IUCrJ* 2(Pt 1):95-105.
26.   Crooks GE, Hon G, Chandonia JM, & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188-1190.
27.   Jackson S*, et al.* (1994) Template-Constrained Cyclic-Peptides - Design of High-Affinity Ligands for Gpiib/Iiia. *Journal of the American Chemical Society* 116(8):3220-3230.

Supporting schemes and figures
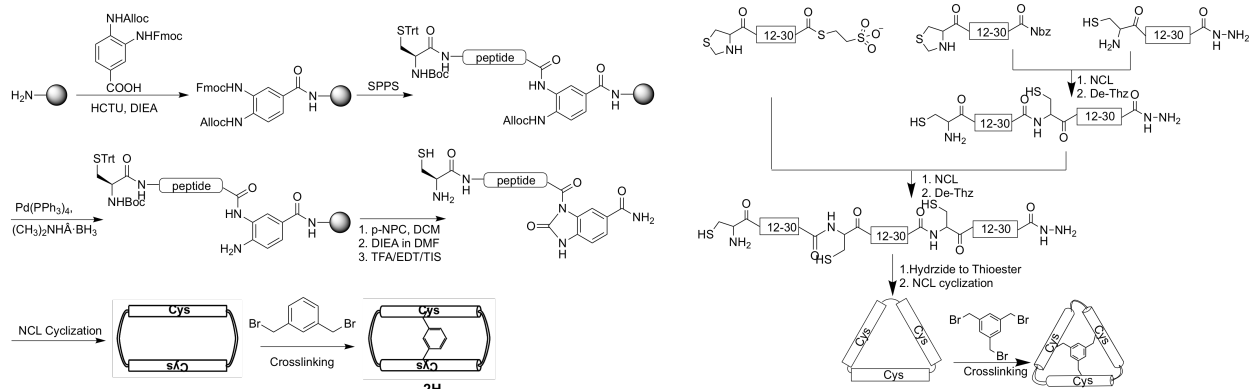
Scheme S1. Synthetic schemes. Synthetic scheme of anti-parallel helical dimers (left); synthetic scheme of tricyclic protein molecules (right). Three segments one-pot native chemical ligation was carried out to prepare linear polypeptide during the synthesis of **3H1** and **3H2**.
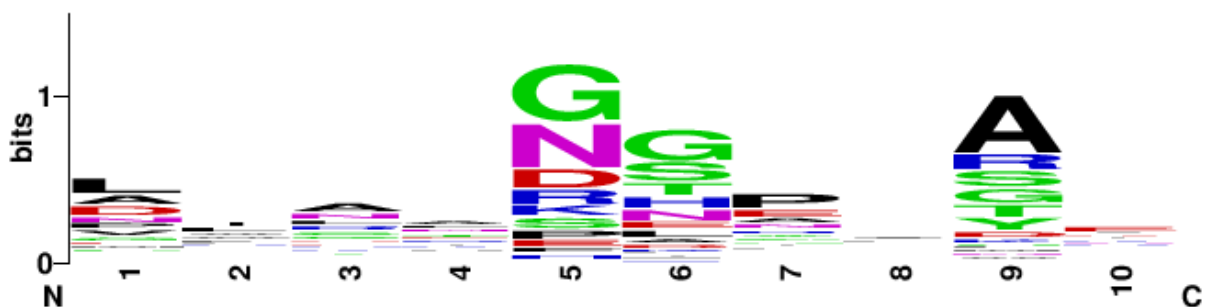
Fig. S1. Sequence profile of protein fragments closely matching the **3H1** NMR structure containing a new loop conformation grafted at positions -1,1, and 2 (or 19-22) of the first of three symmetric helix-loop-helix junctions. The loop residues are denoted in the weblogo(26) by the red underline, and other 7 positions represent those helical residues found in the NMR structure. These positions correspond to the ALNSGGDLAN fragment of the **3H2** sequence and the DSSSKNPEAE fragment of the **3H1** sequence.
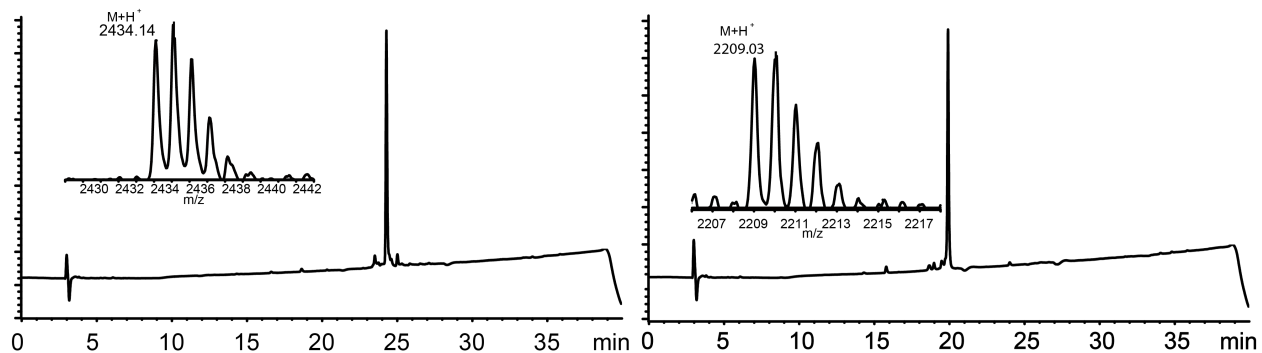
Fig. S2. Analytical HPLC traces and mass spectrometry analysis of **3H1** and **3H2** peptide fragments used for ligation reactions. Left: **3H1** peptide Thz11-Arg30-Nbz, Obsd. 2433.14 ± 0.1 Da, Calc. 2433.16 Da. Right: **3H2** peptide Thz11-Arg30-MESNa, Obsd. 2208.03 ± 0.1 Da, Calc. 2208.03 Da. All calculated masses are monoisotope composition.
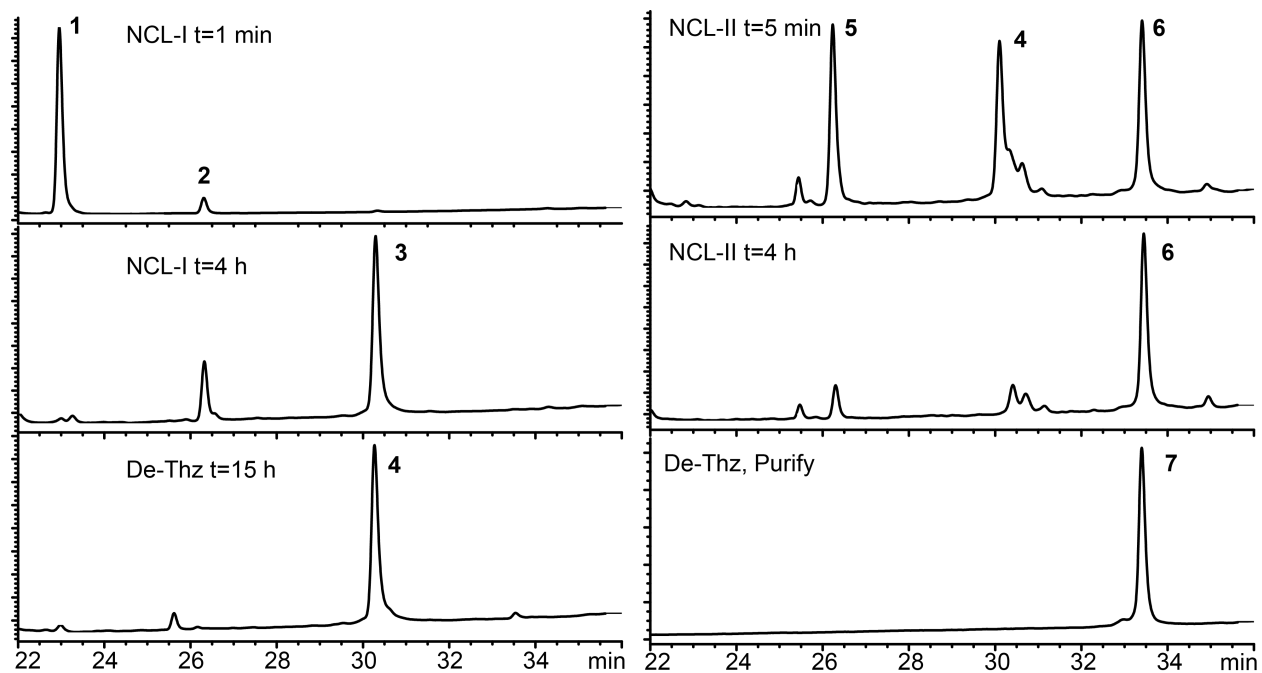
Fig. S3. Analytical HPLC data for the chemical synthesis of linear **3H1**.*CONHNH₂*. **1** Cys11-Arg30. *CONHNH₂*; **2** Thz11-Arg30-Nbz; **3** Thz11-Arg30-Cys11-Arg30.*CONHNH₂*; **4** Cys11-Arg30-Cys11-Arg30.*CONHNH₂*; **5** Thz11-Arg30-MESNa; **6** Thz11-Arg30-Cys11-Arg30-Cys11-Arg30.*CONHNH₂*; **7** Cys11-Arg30-Cys11-Arg30-Cys11-Arg30.*CONHNH₂*
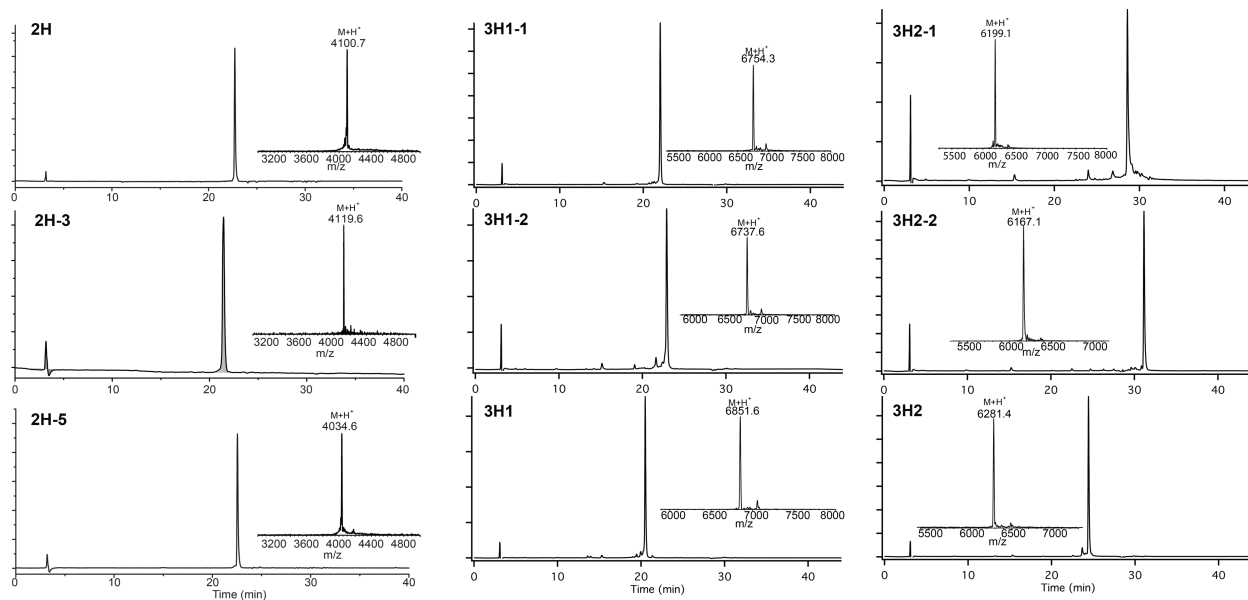
Fig. S4. Analytical HPLC traces and mass spectrometry data of helical dimers synthesis (left), **3H1** (middle) synthesis and **3H2** (right) synthesis. **2H** Mass: Obsd. 4100.7 ± 0.5 Da, Calc. 4100.9 Da. **2H-3** Mass: Obsd. 4118.6 ± 0.5 Da, Calc. 4118.9 Da. **2H-5** Mass: Obsd. 4034.6 ± 0.5 Da, Calc. 4035.0 Da. **3H1-1**.*CONH₂* Mass: Obsd. 6753.5 ± 0.5 Da, Calc. 6753.3 Da. **3H1-2** Mass: Obsd. 6736.5 ± 0.5 Da, Calc. 6736.6 Da. **3H1** Mass: Obsd. 6850.8 ± 0.5 Da, Calc. 6850.6 Da. **3H2-1**.*NHNH₂* Mass: Obsd. 6198.1 ± 0.5

16

Da, Calc. 6198.0 Da. **3H2-2** Mass: Obsd. 6166.1± 0.5 Da, Calc. 6166.0 Da. **3H2** Mass: Obsd. 6280.4± 0.5 Da, Calc. 6280.2 Da. All calculated masses are average isotope composition.
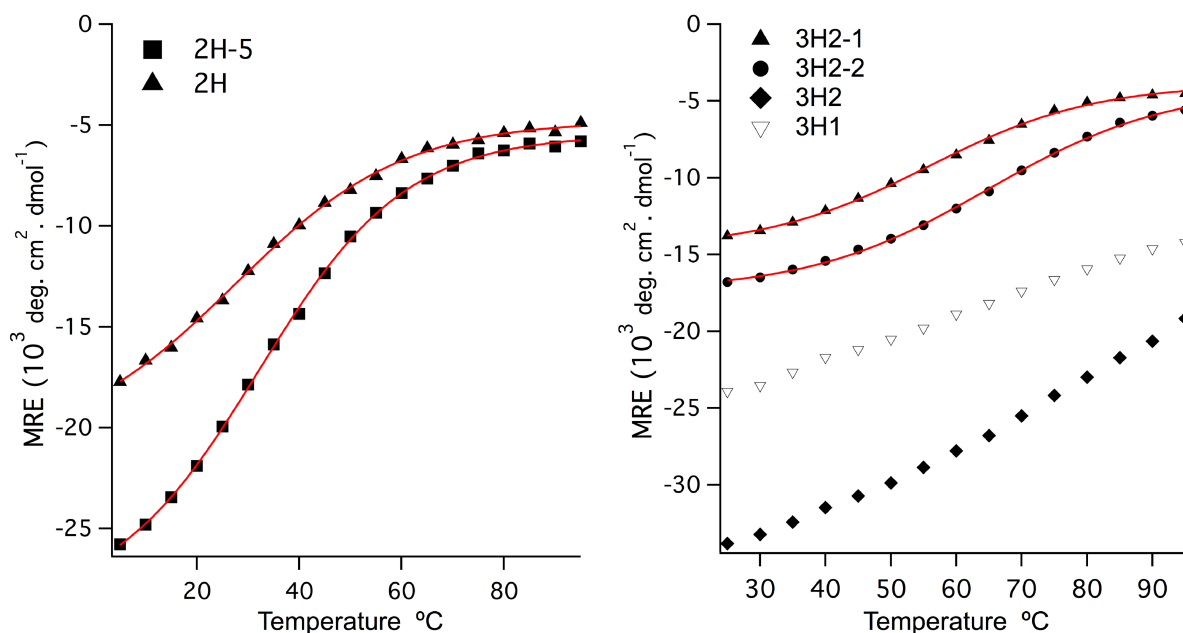
Fig. S5. Thermal melting of helical hairpins (left) and tricyclic protein **3H1** and **3H2**(right). Thermal melting temperatures were obtained by using sigmoidal fitting for **2H-5** (32.1°C), **2H** (27.5°C), 3H2-1 (55.8°C) and **3H2-2** (64.9°C).
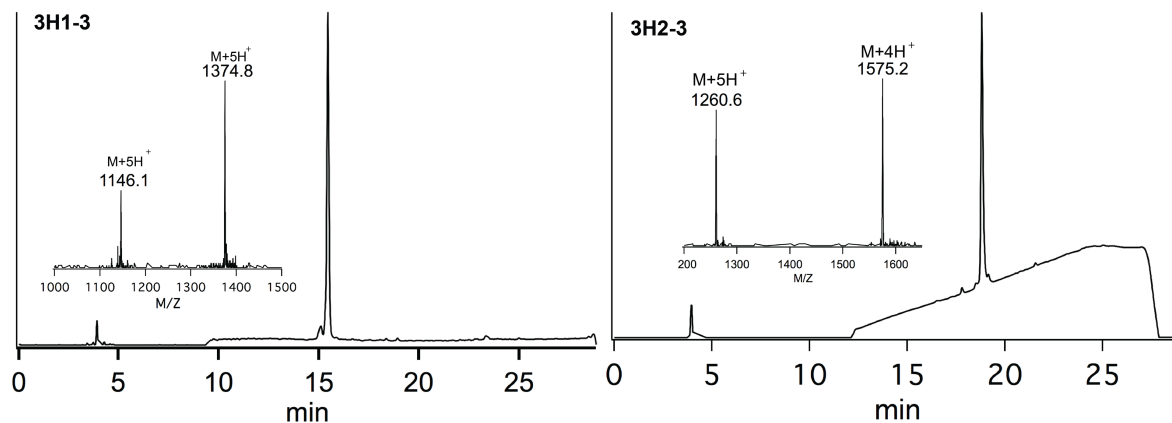
Fig. S6. Analytical HPLC traces, mass spectrometry data of crosslinked but not cyclized **3H1-3** (left) and **3H2-3** (right). Crosslinked **3H1-3** Mass: Obsd. 6869.8 ± 2.0 Da, Calc. 6867.6 Da. Crosslinked **3H2-3** Mass: Obsd. 6297.4± 2.0 Da, Calc. 6297.2 Da. All calculated masses are average isotope composition.
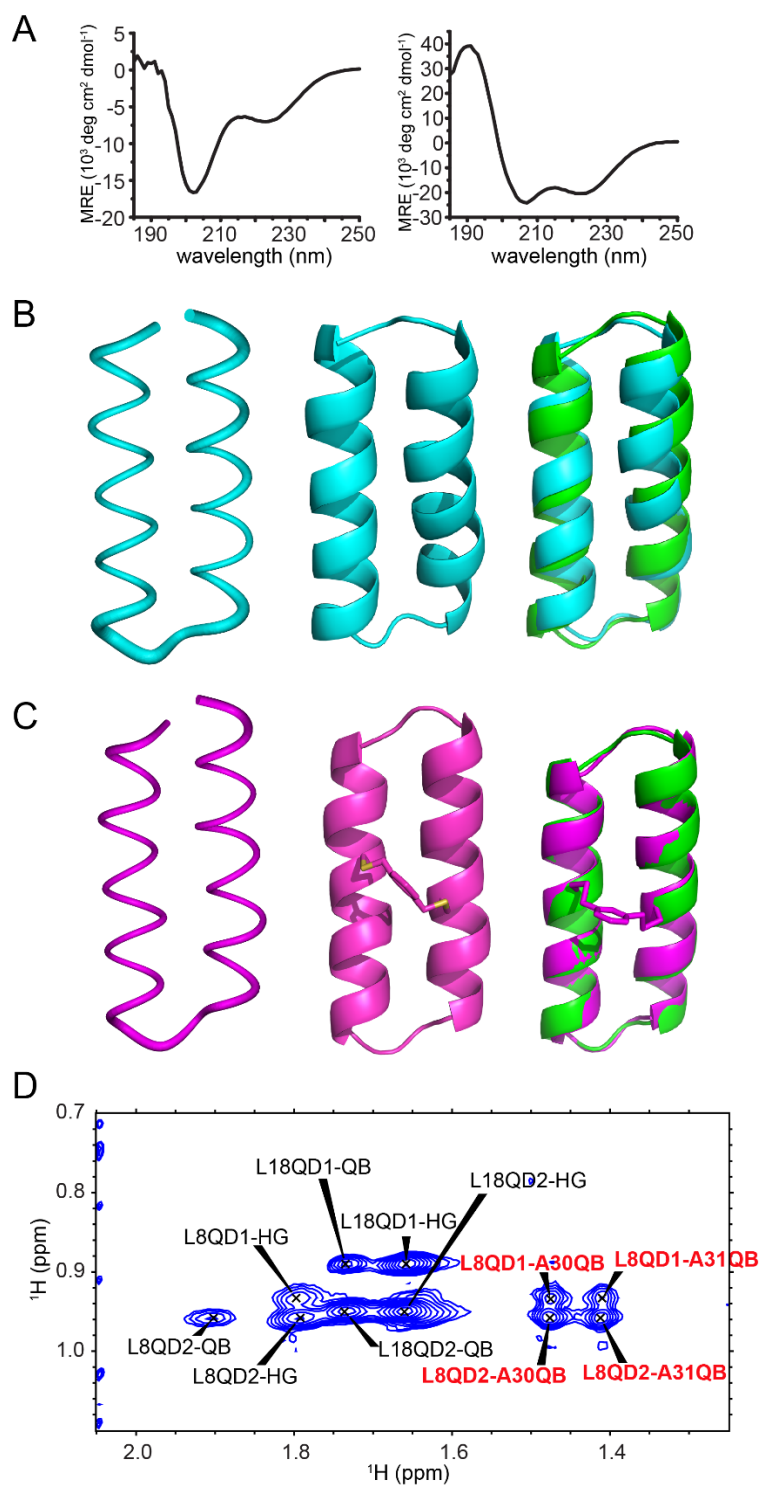
Fig. S7. (A) CD spectra of **2H-4** (left) and **2H-5** (right). Structures of **2H-5** (B) and **2H** (C). Left: ensembles of 20 lowest energy structures (only the backbone is shown); mid: the lowest energy structure; right: comparison with the designed model (green). (D) The NOESY spectrum of **2H**. Red: long range NOEs define the contacts between residues L8 and A30/A31.
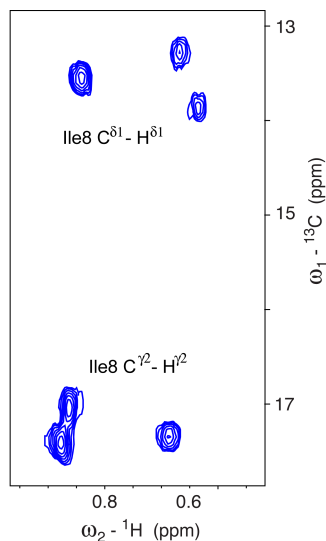
Fig. S8. ^{13}C-HSQC of **3H2** isoleucine residues. The natural abundance ^{13}C-HSQC was recorded at 285 K with 256 scans and a t_1 relaxation time of 8 ms in the ^{13}C dimension on a Bruker 900, in the same buffer that we used to solve the **3H1** structure. There is only one Ile per repeat. Nevertheless, the ^{13}C-HSQC of **3H2** shows three peaks for each Ile $C_{\delta1}$-$H_{\delta1}$ and $C_{\gamma2}$-$H_{\gamma2}$, indicating that the peptide's C3 symmetry is broken in solution. The peaks have different chemical shifts and intensities, indicating that the three repeats have different structures and dynamics.
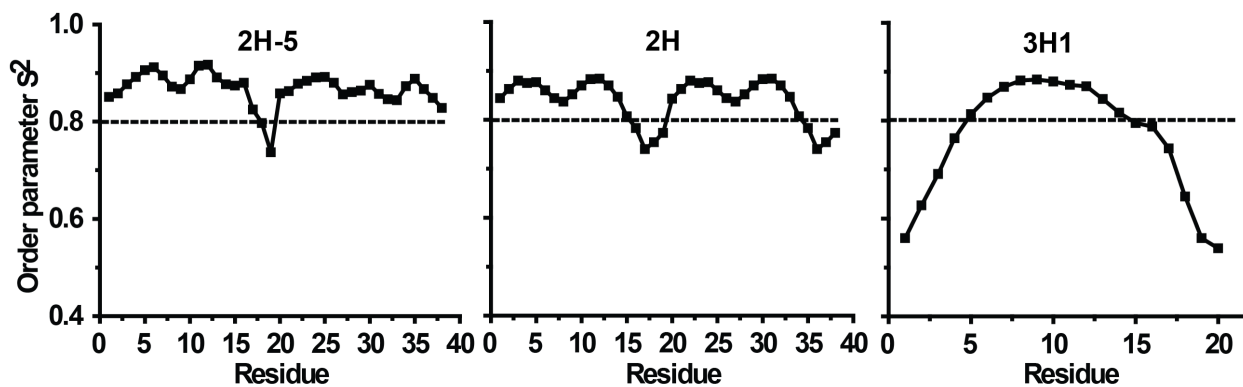
Fig. S9. The backbone order parameters S^2 are plotted as a function of residues. The S^2 prediction was based on ^{13}C chemical shifts by Talos-N.
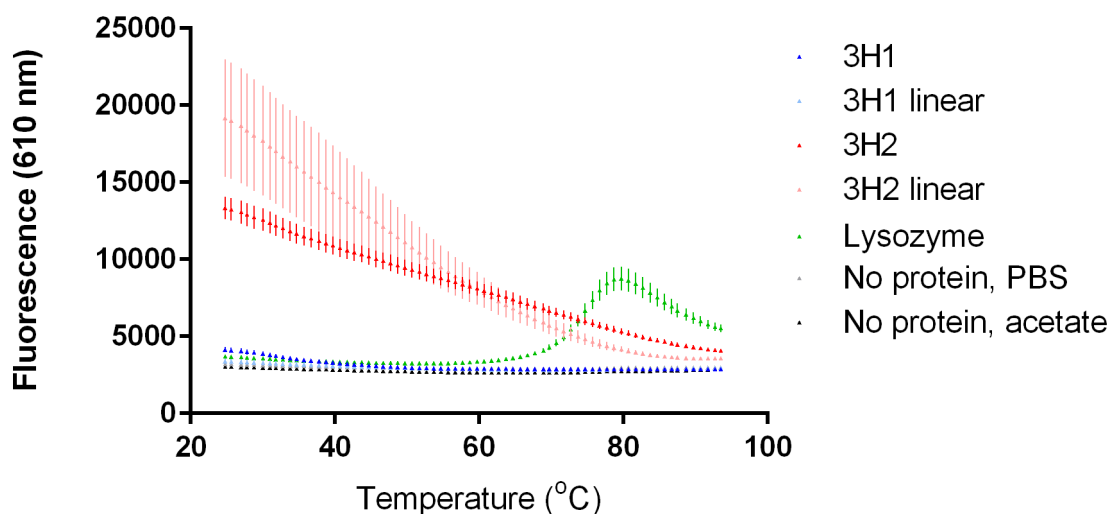
Fig. S10. Differential scanning fluorimetry of samples containing 15μM protein and 75μM SYPRO[TM] Orange (Thermo Fisher S6650). **3H1** (royal), **3H1** linear (light blue), **3H2** (red), and **3H2** linear (pink) samples prepared in PBS, pH 7.40; lysozyme (green)—a model globular DSF protein—prepared in 100 mM NaOAc, 150 mM NaCl, pH 5.0. "No protein" reactions contained only 75μM SYPRO[TM] Orange in buffer. All reactions performed at 25 μl volume in a white PCR plate (USA Scientific #1402-9590) on a Stratagene Mx3005P RT-PCR. SYPRO[TM] Orange fluorescence (ex: 492 nm, em: 610 nm) was measured as samples were heated from 25°C to 95 °C (ramp rate 2 °C per minute). Each condition performed in triplicate; error bars show standard deviation.
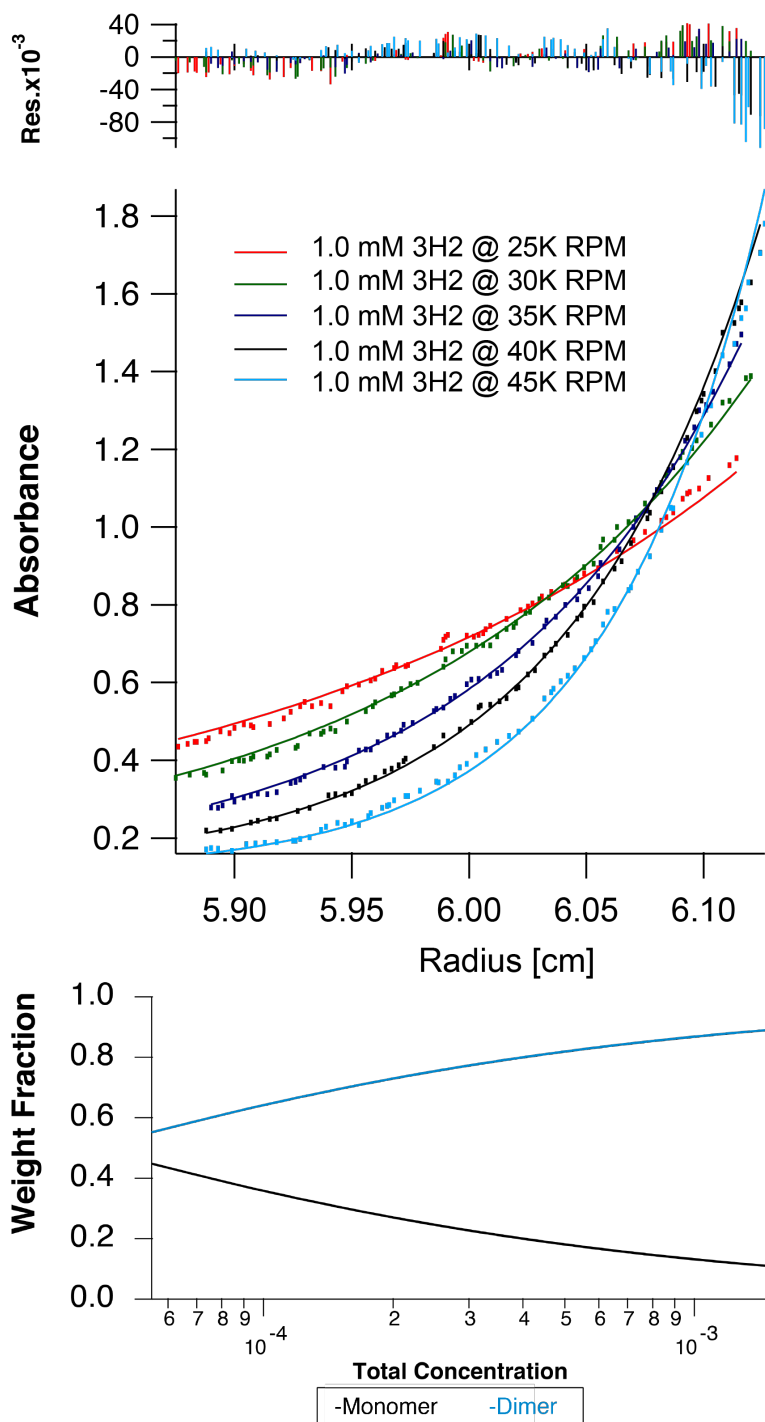
Fig. S11. Analytical ultracentrifugation of **3H2**. **3H2** (molecular weight as 6280.4 Da) at 1.0 mM (upper) in phosphate buffer (25 mM, pH 7.4), are centrifuged at five different speeds. Single-species fitting shows that the average molecular weight of **3H2** is 10500 ± 40 Da at 1.0 mM. Global fits as a monomer-dimer equilibrium yield Kd ~50 μM (lower) (27).

Supporting tables

Table S1: Crick parameter ranges sampled using the `BundleGridSampler`.

Parameter name	Description	Minimum value	Maximum value	Samples
r_0	Superhelix radius, or helix offset from the Z-axis.	4.5 Å	9.0 Å	19
ω_0	Superhelical twist about the Z-axis. For short helices and small values of ω_0, this is similar to a small inclination about the X-axis.	-10.0º	10.0º	41
$\Delta\omega_1$	Rotation of the helix about its own axis.	0.0º	337.5º	16
$z_{1,offset}$	Displacement of the helix along its own axis. (Note that this affects the placement of the symmetric copies of the helix after rotation about the Y-axis when C3 symmetry is set up.)	-3.0 Å	3.0 Å	25
			Total samples:	**311,600**

Table S2. Description of amino acids allowed and sampled during the fixed-backbone design of **3H2** based on the **3H1** NMR structure with a grafted loop. The final column shows the sequence with the lowest Rosetta energy, with amino acids in parentheses denoting the other residue identities output from this design protocol. Only 8 unique sequences were observed after 50,000 design trajectories. The asymmetric segment of the **3H1** and **3H2** sequences that were experimentally characterized are shown for reference.

	3H1 identity	Amino acids allowed	Structural description	Final **3H2** sequence	Top designed sequence
1	K	AFGIKLPRSTVHYN	loop	G	N
2	N	GSTHN	loop	G	G
3	P	DAKLIVE	loop/N-cap	D	D (KL)
4	E	ADEFGIKLNPQRSTVWYH	interface boundary	L	L
5	A	A	buried	A	A
6	E	E	surface	N	E
7	E	E	surface	E	E
8	I	AILVM	buried	I	I
9	T	ANTS	surface	A	A (N)
10	R	R	surface	R	R
11	C	C	covalent site	C	C
12	K	YWSTHALFKD	surface	T	K
13	K	K	surface	K	K
14	L	AILVM	buried	L	L
15	L	AILVM	buried	L	L
16	D	NDKE	surface	N	N (E)
17	D	ADEIKLNQRSTVGY	interface boundary	A	A
18	S	ADEIKLNQRSTVY	interface boundary	L	L
19	S	AGHNRKSTHDEVL	surface/c-cap -1	N	K
20	S	ST	c-cap	S	S

Table S3. Backbone dihedral angles of loop residues in **2H**, **2H-4,** and the designed model.

		Ser1/Cys1	Gly19	Ser20	Gly38
2H	**phi**	-136.4 ± 4.0 [a]	120.0 ± 6.4	-142.4 ± 7.0	120.3 ± 6.6
	psi	73.6 ± 3.3	54.1 ± 7.9	73.4 ± 3.2	48.4 ± 5.3
2H-4	**phi**	-98.8 ± 16.3	78.4 ± 9.9	-86.1 ± 51.5	72.7 ± 6.1
	psi	101.7 ± 7.0	12.3 ± 46.6	93.3 ± 9.8	21.1 ± 16.9
Design	**phi**	-109	121	-109.6	103.9
	psi	103.3	26.9	101.3	29

[a] average value ± standard deviation of 20 NMR structures.

Table S4. Statistics of NMR Structures.

	2H	2H-4	3H1
Conformationally-restricting distance constraints			
Intraresidue [$i = j$]	324	259	210
Sequential [$|i - j| = 1$]	86	67	51
Medium Range [$1 < |i - j| < 5$]	102	66	75
Long Range [$|i - j| > 5$]	0	29	3
Protein-Crosslinker	8	0	36
Inter-helices	8	0	15
Dihedral angle constraints ($\varphi/\psi/\chi_1$)	(26/24/0)	(32/32/0)	(60/60/27)
Average number of distance constraints violations per Xplor-NIH conformer			
0.2 – 0.5 Å	0.0	0.0	0.0
> 0.5 Å	0.0	0.0	0.0
Average number of dihedral-angle constraint violations per Xplor-NIH conformer			
> 5°	0.0	0.0	0.0
Average RMSD to the mean coordinates [Å]			
Regular secondary structure elements, backbone heavy atoms [a]	0.30 ± 0.11	0.38 ± 0.11	0.44 ± 0.14
Regular secondary structure elements, all heavy atoms [a]	0.87 ± 0.10	1.00 ± 0.09	0.98 ± 0.15
All backbone heavy atoms	0.31 ± 0.11	0.39 ± 0.12	0.47 ± 0.16
All heavy atoms	0.85 ± 0.09	0.98 ± 0.09	0.95 ± 0.15
Ramachandran plot summary [%]			
most favored regions	97.0	100	100.0
Additionally allowed regions	3.0	0.0	0.0
generously allowed regions	0.0	0.0	0.0
disallowed regions	0.0	0.0	0.0
RMS deviation for bond angles:	1.5°	0.5°	1.4°
RMS deviation for bond lengths	0.012 Å	0.007 Å	0.011 Å

[a] **2H/2H-4**: residue 2-18, 21-37; **3H1**: residues 4-17, 24-37, 44-57.

Table S5. X-ray data collection and refinement statistics.

Wavelength (Å)	0.8265
Resolution range (Å)	54.81– 1.20 (1.24 – 1.20)
Space group	P 1
Cell dimensions	
a, b, c (Å)	36.77, 55.34, 58.04
α, β, γ (°)	89.79, 108.15, 109.52
Solvent content (%)	52.64
Mol/asymmetric	6
Mol/unit cell	6
Unique reflections	127440
Average Redundancy	3.0 (1.4)
Completeness (%)	92.4 (92.2)
Mean I/sigma(I)	7.8 (1.3)
Rmeas	6.7 (82.7)
$CC_{1/2}$	99.9(52.9)
Refinement statistics	
Resolution range (Å)	54.81 – 1.20 (1.24 – 1.20)
R-factor	0.144
R-free	0.178
Number of reflections	114886 (8435)
Number of protein atoms	2684
Number of water	357
Protein residues	360
RMS(bonds)	0.03
RMS(angles)	2.90
Ramachandran favored (%)	98.9
Ramachandran allowed (%)	1.1
Ramachandran outliers (%)	0
Average B-factor	20.07

Rosetta scripts

Listing S1: RosettaScripts XML file `design.xml` defining the design protocol that was used to create the 3H1 polypeptide.

```
<ROSETTASCRIPTS>
        # 3H1 design script, written 3 June 2015 by
        # Vikram K. Mulligan, Ph.D. (vmullig@uw.edu),
        # Baker laboratory, Institute for Protein Design,
        # University of Washington.
        # Script updated on 28 February 2017 to ensure
        # compatibility with Rosetta 3.8.
        <SCOREFXNS>
                # This function defines the various scoring functions that are used in different
                # parts of this protocol.  All are variants on the default Rosetta talaris2014
                # scoring function.

                # Rosetta talaris2014 scorefunction with slightly modified reference weights:
                <ScoreFunction
                        name="tala"
                        weights="xml/talaris2014_modified.wts"
                >
                </ScoreFunction>

               # Symmetric version of the above scorefunction (for use with symmetric structures):
                <ScoreFunction
                        name="tala_symm"
                        weights="xml/talaris2014_modified.wts"
                        symmetric="true"
                >
                </ScoreFunction>

                # Modified version of the above scorefunction with the repulsive weight
                # reduced.  Used for the initial quick round of design.
                <ScoreFunction
                        name="tala_symm_soft"
                        weights="xml/talaris2014_modified.wts"
                        symmetric="true"
                >
                        <Reweight scoretype="fa_rep" weight="0.1" />
                </ScoreFunction>

                # Modified version of the tala_symm scorefunction with atom-pair distance, bond
                # angle, and dihedral angle constraint score terms turned on.
                <ScoreFunction
                        name="tala_symm_cst"
                        weights="xml/talaris2014_modified.wts"
                        symmetric="true"
                >
                        <Reweight scoretype="coordinate_constraint" weight="0.333" />
                        <Reweight scoretype="atom_pair_constraint" weight="0.333" />
                        <Reweight scoretype="angle_constraint" weight="0.333" />
                        <Reweight scoretype="dihedral_constraint" weight="0.333" />
                </ScoreFunction>

                # The constraints terms alone.  Used for filtering based on linker geometry.
                <ScoreFunction
                        name="symm_cst_only"
                        weights="empty.wts"
                        symmetric="true"
                >
                        <Reweight scoretype="coordinate_constraint" weight="0.333" />
                        <Reweight scoretype="atom_pair_constraint" weight="0.333" />
                        <Reweight scoretype="angle_constraint" weight="0.333" />
                        <Reweight scoretype="dihedral_constraint" weight="0.333" />
                </ScoreFunction>

        </SCOREFXNS>
        <RESIDUE_SELECTORS>
                # Residue selectors provide rules for defining sets of amino acid positions
```

```
# in a structure.  They can be used to control many Rosetta modules.
# A residue selector to select buried positions, based on solvent-accessible
# surface area:
<Layer name="select_core"
        select_core="true" select_boundary="false" select_surface="false"
        use_sidechain_neighbors="false" ball_radius="2.0"
        core_cutoff="30" surface_cutoff="50"
/>

# A residue selector to select semi-buried positions, based on solvent-accessible
# surface area:
<Layer name="select_boundary"
        select_core="false" select_boundary="true" select_surface="false"
        use_sidechain_neighbors="false" ball_radius="2.0"
        core_cutoff="30" surface_cutoff="50"
/>

# A residue selector to select solvent-exposed positions, based on solvent
# accessible surface area:
<Layer name="select_surface"
        select_core="false" select_boundary="false" select_surface="true"
        use_sidechain_neighbors="false" ball_radius="2.0"
        core_cutoff="30" surface_cutoff="50"
/>

</RESIDUE_SELECTORS>
<TASKOPERATIONS>
        # Task operations control the Rosetta packer module, which optimizes side-chain
        # rotamers given a fixed backbone.  Task operations provide restrictions on the
      # amino acid identities and rotamer types that will be considered at each position.

        # Only consider rotamers of the amino acid type that is already present at each
        # position:
        <RestrictToRepacking name="nodesign" />

        # Do not allow the amino acid identity of the cysteine residues that connect to
        # the linker to change:
        <RestrictResiduesToRepacking name="only_repack_linker" residues="21,42,63" />

        # At buried positions selected with the "select_core" residue selector, only allow
        # hydrophobic amino acids or alanine (FAMILYVW):
        <ReadResfile name="core_resfile" selector="select_core"
                filename="xml/core.resfile"
        />

      # At boundary positions, only allow proline, tyrosine, small aliphatic amino acids,
        # or polar uncharged amino acids (ILYVTSNQP):
        <ReadResfile name="boundary_resfile" selector="select_boundary"
                filename="xml/boundary.resfile"
        />

        # At solvent-exposed positions, only allow polar (charged or uncharged) amino
acids
        # or proline (DERKTSNQP):
        <ReadResfile name="surface_resfile" selector="select_surface"
                filename="xml/surface.resfile"
        />

</TASKOPERATIONS>
<FILTERS>
        # Filters allow a trajectory to be aborted if a particular condition is not met.

        # Terminate trajectories in which the initial backbone geometry clashes with
itself:
        <ScoreType name="initial_clash_check" scorefxn="tala_symm"
                score_type="fa_rep" threshold="100"
        />

        # Terminate trajectories in which energy-minimization cannot pull the TBMB
        # cross-linker into a "closed" state (i.e. in which the initial backbone
        # geometry is incompatible with presenting the TBMB cross-linker):
```

27

```
<ScoreType name="bad_linker_filter" scorefxn="symm_cst_only"
        score_type="total_score" threshold="40"
/>

# Discard sampled loop conformations that do not resemble 4-residue loops observed
# in the Protein Data Bank (used during GenKIC sampling):
<FragmentLookupFilter name="faulty_fragments_tolerant"
        lookup_name="source_fragments_4_mer_tolerant"
        store_path="backbone_profiler_database_06032014"
        lookup_mode="first" threshold="0" confidence="1"
/>


# Discard sampled loop conformations that do not resemble 4-residue loops observed
# in the Protein Data Bank (used as a final filter):
<FragmentLookupFilter name="faulty_fragments"
        lookup_name="source_fragments_4_mer"
        store_path="backbone_profiler_database_06032014"
        lookup_mode="first" threshold="0" confidence="1"
/>

</FILTERS>
<MOVERS>
        # Movers modify a structure (called a "pose" in Rosetta) in some way.  They can
add
        # or remove geometry, move atoms or rotate bonds, add constraints, or otherwise
        # alter the pose.

        # Generate the initial backbone geometry using the Crick parameterization.
        # Note that since the "nstruct_mode" option is used, each repeated iteration
through
        # the RosettaScripts protocol will sample a different grid-point in the set of
        # parameter value combinations to be sampled:
        <BundleGridSampler name="bgs1" helix_length="20" scorefxn="tala"
                set_bondlengths="true" set_bondangles="true" set_dihedrals="true"
                nstruct_mode="true"         nstruct_repeats="1"         use_degrees="true"
max_samples="311600"
                r0_min="4.5" r0_max="9" r0_samples="19"
                omega0_min="-10" omega0_max="10" omega0_samples="41"
                delta_omega0="0"
                delta_omega1_min="0" delta_omega1_max="337.5" delta_omega1_samples="16"
                delta_t="0"
                z1_offset_min="-3" z1_offset_max="3" z1_offset_samples="25"
        >
                <Helix />
        </BundleGridSampler>

        # Add a sidechain-conjugated cysteine type at position 11.  The TBMB linker
        # will be attached here:
        <MutateResidue name="makecys" target="11" new_res="CYX" />

        # Add the fragment of the TBMB linker (one-third of the overall geometry) to
        # the side-chain of the cysteine residue added by the previous mover.
        <PeptideStubMover name="add_linker" reset="false">
                <Append resname="TBMBSYMM" anchor_rsd="11"
                        anchor_atom="SG" connecting_atom="CM1"
                />
        </PeptideStubMover>

        # Temporarily add terminal types to the ends of the parametrically-generated
        # helix.  Rosetta can sometimes do bad things if a chain does not end in
        # an NH3+ or COO- group.  These termini will be stripped later when the
        # polypeptide is cyclized.
        <DeclareBond name="add_termini"
                res1="1" res2="2" atom1="C" atom2="N" add_termini="true"
        />

        # Link the fragments of the TBMB linker together:
        <DeclareBond name="linkerbond1"
                res1="21" res2="42" atom1="C2" atom2="C1" add_termini="true"
        />
```

```
                    # Link the fragments of the TBMB linker together:
                    <DeclareBond name="linkerbond2"
                            res1="42" res2="63" atom1="C2" atom2="C1" add_termini="true"
                    />

                    # Link the fragments of the TBMB linker together:
                    <DeclareBond name="linkerbond3"
                            res1="63" res2="21" atom1="C2" atom2="C1" add_termini="true"
                    />

                    # Define a new foldtree.  A foldtree defines the kinematic relationships
                    # between parts of a molecule:
                    <AtomTree name="foldtree1" fold_tree_file="xml/foldtree1.txt" />

                    # Set up the C3 symmetry about the Y-axis.  After this mover is applied, Rosetta
                    # continues to think of the structure as C3-symmetric, enforcing symmetry when
                    # design or energy-minimization operations are applied:
                    <SetupForSymmetry name="setup_symm" definition="symmetry/c3_y.symm" />

                    # Carry  out  initial  minimization  to  bring  the  fragments  of  the  TBMB  linker
together,
                    # orienting the cysteine side-chains in the process:
                    <FastRelax name="frelax_linker"
                            task_operations="nodesign"
                            min_type="lbfgs_armijo_nonmonotone"
                            scorefxn="tala_symm_cst" repeats="1"
                    >
                            <MoveMap name="frelax_linker_mm" >
                                    <Jump number="1" setting="1" />
                                    <Jump number="2" setting="1" />
                                    <Jump number="3" setting="1" />
                                    <Jump number="4" setting="0" />
                                    <Jump number="5" setting="0" />
                                    <Jump number="6" setting="0" />
                                    <Jump number="7" setting="0" />
                                    <Jump number="8" setting="0" />
                                    <Jump number="9" setting="0" />
                                    <Span begin="1" end="999" chi="1" bb="0" />
                            </MoveMap>
                    </FastRelax>

                    # Create a peptide bond between the end of one lobe and the start of the next:
                    <DeclareBond name="linkends1"
                            res1="20" res2="22" atom1="C" atom2="N" add_termini="false"
                    />

                    # Create a peptide bond between the end of one lobe and the start of the next:
                    <DeclareBond name="linkends2"
                            res1="41" res2="43" atom1="C" atom2="N" add_termini="false"
                    />

                    # Create a peptide bond between the end of one lobe and the start of the next:
                    <DeclareBond name="linkends3"
                            res1="62" res2="1" atom1="C" atom2="N" add_termini="false"
                    />

                    # Carry  out  a  quick  round  of  sequence  design,  using  a  version  of  the  scoring
function
                    # with a reduced repulsive term.  This is done for each loop conformation that
                    # GenKIC considers:
                    <SymPackRotamersMover name="quickpack"
                      scorefxn="tala_symm_soft"

task_operations="core_resfile,boundary_resfile,surface_resfile,only_repack_linker"
                    />

                    # Carry out a quick round of sidechain energy-minimization.  This is done for each
                    # loop conformation that GenKIC considers (holding the backbone fixed):
                    <SymMinMover name="quickmin" scorefxn="tala_symm_cst"
                            bb="false" chi="true" jump="0"
```

```
                        type="lbfgs_armijo_nonmonotone" tolerance="0.001"
        >
                <MoveMap name="quickmin_mm" >
                        <Span begin="1" end="999" bb="0" chi="1" />
                        <Jump number="1" setting="1" />
                        <Jump number="2" setting="1" />
                        <Jump number="3" setting="1" />
                        <Jump number="4" setting="0" />
                        <Jump number="5" setting="0" />
                        <Jump number="6" setting="0" />
                        <Jump number="7" setting="0" />
                        <Jump number="8" setting="0" />
                        <Jump number="9" setting="0" />
                </MoveMap>
        </SymMinMover>

        # The composite protocol applied to every loop conformation that GenKIC considers.
        # This uses a series of filters and movers that have been defined above.
        <ParsedProtocol name="quickpackmin" >
                <Add filter="initial_clash_check" />
                <Add filter="faulty_fragments_tolerant" />
                <Add mover="quickpack" />
                <Add mover="quickmin" />
        </ParsedProtocol>

        # The Generalized Kinematic Closure (GenKIC) mover, used to sample loop
conformations.
        # For every loop conformation considered, it applies the "quickpackmin" protocol,
        # defined above, to discard loops with bad geometry and to carry out a rapid round
        # of sequence design and side-chain energy minimization.
        <GeneralizedKIC name="genkic1"
                closure_attempts="100"
                selector="lowest_energy_selector"
                selector_scorefunction="tala_symm"
                pre_selection_mover="quickpackmin"
        >
                <AddResidue res_index="17" />
                <AddResidue res_index="18" />
                <AddResidue res_index="19" />
                <AddResidue res_index="20" />
                <AddResidue res_index="22" />
                <AddResidue res_index="23" />
                <SetPivots res1="17" atom1="CA" res2="19" atom2="CA" res3="23" atom3="CA"
/>
                <CloseBond res1="20" res2="22" atom1="C" atom2="N"
                        bondlength="1.32" angle1="114" angle2="123" torsion="180"
                />
                <AddPerturber effect="randomize_backbone_by_bins" bin_params_file="ABBA" >
                        <AddResidue index="17" />
                        <AddResidue index="18" />
                        <AddResidue index="19" />
                        <AddResidue index="20" />
                        <AddResidue index="22" />
                        <AddResidue index="23" />
                </AddPerturber>
                <AddFilter type="loop_bump_check" />
        </GeneralizedKIC>

        # The computationally-expensive "FastDesign" protocol applied to the best loop
        # conformation that GenKIC returns.  This first round keeps the backbone fixed
while
        # carrying out multiple rounds of side-chain design and side-chain energy
        # minimization, gradually ramping the repulsive term in the score function from
        # round to round.
        <FastDesign name="fdes1" task_operations="core_resfile,boundary_resfile,
          surface_resfile,only_repack_linker" disable_design="false"
          min_type="lbfgs_armijo_nonmonotone" scorefxn="tala_symm_cst" repeats="1"
        >
                <MoveMap name="fdes1_mm" >
                        <Jump number="1" setting="1" />
                        <Jump number="2" setting="1" />
```

```
                                    <Jump number="3" setting="1" />
                                    <Jump number="4" setting="0" />
                                    <Jump number="5" setting="0" />
                                    <Jump number="6" setting="0" />
                                    <Jump number="7" setting="0" />
                                    <Jump number="8" setting="0" />
                                    <Jump number="9" setting="0" />
                                    <Span begin="1" end="999" chi="1" bb="0" />
                        </MoveMap>
                </FastDesign>

                # Add atom-pair distance, bond angle, and dihedral constraints enforcing ideal
                # peptide bond geometry between the lobes of the polypeptide.
                <ConstraintSetMover name="constrain_pepbonds"
                        add_constraints="true" cst_file="constraints/pepbonds.cst"
                />

                # Add atom-pair distance and dihedral constraints for the TBMB cross-linker:
                <ConstraintSetMover name="constrain_linker"
                        add_constraints="true" cst_file="constraints/linker.cst"
                />

                # The computationally-expensive "FastDesign" protocol applied to the best loop
                # conformation that GenKIC returns.  This second round allows the backbone to
relax
                # while carrying out multiple rounds of side-chain design and full-structure
                # energy minimization, gradually ramping the repulsive term in the score function
                # from round to round.
                <FastDesign name="fdes2" task_operations="core_resfile,boundary_resfile,
                  surface_resfile,only_repack_linker" disable_design="false"
                  min_type="lbfgs_armijo_nonmonotone" scorefxn="tala_symm_cst"
                  ramp_down_constraints="false" repeats="3"
                >
                        <MoveMap name="fdes2_mm">
                                    <Jump number="1" setting="1" />
                                    <Jump number="2" setting="1" />
                                    <Jump number="3" setting="1" />
                                    <Jump number="4" setting="1" />
                                    <Jump number="5" setting="1" />
                                    <Jump number="6" setting="1" />
                                    <Jump number="7" setting="1" />
                                    <Jump number="8" setting="1" />
                                    <Jump number="9" setting="1" />
                                    <Span begin="1" end="999" chi="1" bb="1" />
                        </MoveMap>
                </FastDesign>

        </MOVERS>
        <PROTOCOLS>
                # This section specifies the order in which the movers and filters defined above
                # will be applied.

                <Add mover="bgs1" />
                <Add mover="makecys" />
                <Add mover="add_linker" />
                <Add mover="add_termini" />
                <Add mover="foldtree1" />
                <Add mover="setup_symm" />
                <Add mover="linkerbond1" />
                <Add mover="linkerbond2" />
                <Add mover="linkerbond3" />
                <Add mover="constrain_linker" />
                <Add mover="frelax_linker" />
                <Add filter="bad_linker_filter" />
                <Add mover="linkends1" />
                <Add mover="linkends2" />
                <Add mover="linkends3" />
                <Add mover="genkic1" />
                <Add mover="fdes1" />
                <Add mover="constrain_pepbonds" />
                <Add mover="fdes2" />
```

```
                <Add mover="linkends1" />
                <Add mover="linkends2" />
                <Add mover="linkends3" />
                <Add filter="faulty_fragments" />

        </PROTOCOLS>

        # For final output, the resulting structure should be scored with the tala_symm scoring
        # function, and the score written in the output file:
        <OUTPUT scorefxn="tala_symm" />
</ROSETTASCRIPTS>
```

Listing S2: Rosetta scoring function weights file `talaris2014_modified.wts`, defining the Rosetta talaris2014 scoring function with modified reference weights.

```
# Modified to upweight methionine (+2.5) alanine (+2.5) and tryptophan (+1) while
# down-weighting proline (-1.5).
METHOD_WEIGHTS ref 3.273742 0.443793 -1.63002 -1.96094 0.61937 0.173326 0.388298 1.0806 -0.358574
0.761128 3.249477 -1.19118 -1.750485 -1.51717 -0.32436 0.165383 0.20134 0.979644 2.23413 0.162496
fa_atr 1
fa_rep 0.55
fa_sol 0.9375
fa_intra_rep 0.005
fa_elec 0.875
pro_close 1.25
hbond_sr_bb 1.17
hbond_lr_bb 1.17
hbond_bb_sc 1.17
hbond_sc 1.1
dslf_fa13 1.25
rama 0.25
omega 0.625
fa_dun 0.7
p_aa_pp 0.4
yhh_planarity 0.625
ref 1
```

Listing S3: Symmetry definition file `c3_y.symm`, defining C3 symmetry about the Y-axis.

```
symmetry_name c3
subunits 3
number_of_interfaces  1
E = 3*VRT0001 + 3*(VRT0001:VRT0002)
anchor_residue COM
virtual_transforms_start
start -1,0,0 0,1,0 0,0,0
rot Ry_angle -120
virtual_transforms_stop
connect_virtual JUMP1 VRT0001 VRT0002
connect_virtual JUMP2 VRT0002 VRT0003
set_dof BASEJUMP x(50) angle_x(0:360) angle_y(0:360) angle_z(0:360)
```

Listing S4: Constraints file `linker.cst` listing atom-pair distance and dihedral constraints for the TBMB cross-linker.

```
AtomPair SG 11 V1 21 HARMONIC 0.0 0.01
AtomPair SG 32 V1 42 HARMONIC 0.0 0.01
AtomPair SG 53 V1 63 HARMONIC 0.0 0.01
AtomPair V1 11 CM1 21 HARMONIC 0.0 0.01
AtomPair V1 32 CM1 42 HARMONIC 0.0 0.01
AtomPair V1 53 CM1 63 HARMONIC 0.0 0.01


AtomPair V3 21 C1 42 HARMONIC 0.0 0.01
AtomPair V4 21 C2 42 HARMONIC 0.0 0.01
AtomPair V5 21 C1 63 HARMONIC 0.0 0.01
AtomPair V6 21 C2 63 HARMONIC 0.0 0.01


AtomPair V3 42 C1 63 HARMONIC 0.0 0.01
AtomPair V4 42 C2 63 HARMONIC 0.0 0.01
```

```
AtomPair V5 42 C1 21 HARMONIC 0.0 0.01
AtomPair V6 42 C2 21 HARMONIC 0.0 0.01

AtomPair V3 63 C1 21 HARMONIC 0.0 0.01
AtomPair V4 63 C2 21 HARMONIC 0.0 0.01
AtomPair V5 63 C1 42 HARMONIC 0.0 0.01
AtomPair V6 63 C2 42 HARMONIC 0.0 0.01

Dihedral CM1 21 SG 11 CB 11 CA 11 FADE 2.09439510239 4.18879020479 1.0 -0.5 0.5
Dihedral CM1 21 SG 11 CB 11 CA 11 FADE -4.18879020479 -2.09439510239 1.0 -0.5 0.5
Dihedral CM1 21 SG 11 CB 11 CA 11 FADE 0.0 2.09439510239  1.0 -0.2 0.2
Dihedral CM1 21 SG 11 CB 11 CA 11 FADE -2.09439510239 0.0 1.0 -0.2 0.2
Dihedral CM1 21 SG 11 CB 11 CA 11 FADE 4.18879020479 6.28318530718 1.0 -0.2 0.2
Dihedral C1 21 CM1 21 SG 11 CB 11 FADE 2.09439510239 4.18879020479 1.0 -0.5 0.5
Dihedral C1 21 CM1 21 SG 11 CB 11 FADE -4.18879020479 -2.09439510239 1.0 -0.5 0.5
Dihedral C1 21 CM1 21 SG 11 CB 11 FADE 0.0 2.09439510239  1.0 -0.2 0.2
Dihedral C1 21 CM1 21 SG 11 CB 11 FADE -2.09439510239 0.0 1.0 -0.2 0.2
Dihedral C1 21 CM1 21 SG 11 CB 11 FADE 4.18879020479 6.28318530718 1.0 -0.2 0.2
Dihedral V6 21 C1 21 CM1 21 SG 11 FADE -3.141592654 0.0 1.5 -0.5 0.5
Dihedral V6 21 C1 21 CM1 21 SG 11 FADE 0.0 3.141592654  1.5 -0.5 0.5

Dihedral CM1 42 SG 32 CB 32 CA 32 FADE 2.09439510239 4.18879020479 1.0 -0.5 0.5
Dihedral CM1 42 SG 32 CB 32 CA 32 FADE -4.18879020479 -2.09439510239 1.0 -0.5 0.5
Dihedral CM1 42 SG 32 CB 32 CA 32 FADE 0.0 2.09439510239  1.0 -0.2 0.2
Dihedral CM1 42 SG 32 CB 32 CA 32 FADE -2.09439510239 0.0 1.0 -0.2 0.2
Dihedral CM1 42 SG 32 CB 32 CA 32 FADE 4.18879020479 6.28318530718 1.0 -0.2 0.2
Dihedral C1 42 CM1 42 SG 32 CB 32 FADE 2.09439510239 4.18879020479 1.0 -0.5 0.5
Dihedral C1 42 CM1 42 SG 32 CB 32 FADE -4.18879020479 -2.09439510239 1.0 -0.5 0.5
Dihedral C1 42 CM1 42 SG 32 CB 32 FADE 0.0 2.09439510239  1.0 -0.2 0.2
Dihedral C1 42 CM1 42 SG 32 CB 32 FADE -2.09439510239 0.0 1.0 -0.2 0.2
Dihedral C1 42 CM1 42 SG 32 CB 32 FADE 4.18879020479 6.28318530718 1.0 -0.2 0.2
Dihedral V6 42 C1 42 CM1 42 SG 32 FADE -3.141592654 0.0 1.5 -0.5 0.5
Dihedral V6 42 C1 42 CM1 42 SG 32 FADE 0.0 3.141592654  1.5 -0.5 0.5

Dihedral CM1 63 SG 53 CB 53 CA 53 FADE 2.09439510239 4.18879020479 1.0 -0.5 0.5
Dihedral CM1 63 SG 53 CB 53 CA 53 FADE -4.18879020479 -2.09439510239 1.0 -0.5 0.5
Dihedral CM1 63 SG 53 CB 53 CA 53 FADE 0.0 2.09439510239  1.0 -0.2 0.2
Dihedral CM1 63 SG 53 CB 53 CA 53 FADE -2.09439510239 0.0 1.0 -0.2 0.2
Dihedral CM1 63 SG 53 CB 53 CA 53 FADE 4.18879020479 6.28318530718 1.0 -0.2 0.2
Dihedral C1 63 CM1 63 SG 53 CB 53 FADE 2.09439510239 4.18879020479 1.0 -0.5 0.5
Dihedral C1 63 CM1 63 SG 53 CB 53 FADE -4.18879020479 -2.09439510239 1.0 -0.5 0.5
Dihedral C1 63 CM1 63 SG 53 CB 53 FADE 0.0 2.09439510239  1.0 -0.2 0.2
Dihedral C1 63 CM1 63 SG 53 CB 53 FADE -2.09439510239 0.0 1.0 -0.2 0.2
Dihedral C1 63 CM1 63 SG 53 CB 53 FADE 4.18879020479 6.28318530718 1.0 -0.2 0.2
Dihedral V6 63 C1 63 CM1 63 SG 53 FADE -3.141592654 0.0 1.5 -0.5 0.5
Dihedral V6 63 C1 63 CM1 63 SG 53 FADE 0.0 3.141592654  1.5 -0.5 0.5
```

Listing S5: Constraints file `pepbonds.cst` defining atom-pair distance, bond angle, and dihedral constraints for the peptide bonds connecting the three lobes of the cyclic peptide.

```
Dihedral CA 20 C 20 N 22 CA 22 CIRCULARHARMONIC 3.141592654 0.02
AtomPair C 20 N 22 HARMONIC 1.32 0.02
Angle CA 20 C 20 N 22 CIRCULARHARMONIC 1.98967535 0.02
Angle C 20 N 22 CA 22 CIRCULARHARMONIC 2.14675498 0.02

Dihedral CA 41 C 41 N 43 CA 43 CIRCULARHARMONIC 3.141592654 0.02
AtomPair C 41 N 43 HARMONIC 1.32 0.02
Angle CA 41 C 41 N 43 CIRCULARHARMONIC 1.98967535 0.02
Angle C 41 N 43 CA 43 CIRCULARHARMONIC 2.14675498 0.02

Dihedral CA 62 C 62 N 1 CA 1 CIRCULARHARMONIC 3.141592654 0.02
AtomPair C 62 N 1 HARMONIC 1.32 0.02
Angle CA 62 C 62 N 1 CIRCULARHARMONIC 1.98967535 0.02
Angle C 62 N 1 CA 1 CIRCULARHARMONIC 2.14675498 0.02
```

Listing S6: Foldtree definition file `foldtree1.txt`, defining the kinematic relationships between parts of the molecule.

```
FOLD_TREE EDGE 10 1 -1 EDGE 10 20 -1 EDGE 11 21 1
```

Listing S7: Packer definition file `core.resfile`, defining the allowed amino acid types at buried positions.

```
PIKAA FAMILYVW
start
```

Listing S8: Packer definition file `boundary.resfile`, defining the allowed amino acid types at semi-buried positions.

```
PIKAA ILYVTSNQP
start
```

Listing S9: Packer definition file `surface.resfile`, defining the allowed amino acid types at solvent-exposed positions.

```
PIKAA DERKTSNQP
start
```

Listing S10: Commandline flags file `rosetta.flags` needed for RosettaScripts execution.

```
-in:file:fasta xml/dummy.fasta
-parser:protocol xml/design.xml
-write_all_connect_info
-connect_info_cutoff 0
-ex1
-ex2
-jd2:failed_job_exception false
-nstruct 311600
-extra_res_fa                                    sidechain_conjugation/CYX.params
crosslinker/1.3.5_trisbromomethylbenzene_symm.params
```

Listing S11: Initial FASTA file `dummy.fasta` needed for RosettaScripts execution.

```
AAAAA
```

Listing S12: Commandline execution of the 3H1 design protocol

```
<path_to_Rosetta_dir>/main/source/bin/rosetta_scripts.default.linuxclangrelease
@xml/rosetta.flags
```

Listing S13: Commandline execution of the 3H1 design protocol with parametric sampling distributed automatically via the Message Passing Interface (MPI)

```
mpirun -np <number_of_processes_to_launch>
<path_to_Rosetta_dir>/Rosetta/main/source/bin/rosetta_scripts.mpi.linuxgccrelease
@xml/rosetta.flags
```