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

Supplementary Materials for

Comprehensive computational design of ordered peptide macrocycles

Parisa Hosseinzadeh,* Gaurav Bhardwaj,* Vikram Khipple Mulligan,* Matthew D. Shortridge, Timothy W. Craven, Fátima Pardo-Avila, Stephen A. Rettie, David E. Kim, Daniel-Adriano Silva, Yehia M. Ibrahim, Ian K. Webb, John R. Cort, Joshua N. Adkins, Gabriele Varani, David Baker†

*These authors contributed equally to this work.
†Corresponding author. Email: dabaker@u.washington.edu

Published 15 December 2017, *Science* **358**, 1461 (2017)
DOI: 10.1126/science.aap7577

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:
(available at www.sciencemag.org/content/358/6369/1461/suppl/DC1)

Databases S1 to S3 as zipped archives: Comp_Folded_Macrocycles, Hot_Loop_Matches, Designed_Macrocycles_Structures

Methods:

Backbone conformational sampling

Conformations of 7- to 14-residue polyglycine backbones were sampled using the previously-described Rosetta `simple_cycpep_prediction` application(*15*), with key modifications. Unlike the Rosetta *ab initio* method used for protein structure prediction(*25*), `simple_cycpep_predict` does not make use of fragments of proteins of known structure, since such fragments poorly cover the conformational space accessible to chains of mixtures of L- and D-amino acids. Instead, it uses an efficient kinematic closure-based algorithm(*17*, *26*) that samples only closed conformations to limit the search space. Briefly, the sampling process consisted of the following steps: first, a linear chain of glycine residues was constructed, one residue of which was selected randomly to be the "anchor" residue for subsequent loop closure steps. The N- and C-terminal residues were excluded from being the anchor residue. This residue's mainchain φ and Ψ dihedral values were drawn randomly from a flat, symmetric Ramachandran distribution based on the glycine Ramachandran map (fig. S1A-B). Second, a bond was declared between the nitrogen of the N-terminal residue and the carbonyl carbon of the C-terminal residue, and the Rosetta generalized kinematic closure (GenKIC) module was invoked to close the loop consisting of all residues but the anchor residue. During this process, the φ and Ψ dihedral values of all but three residues in the loop were randomized, biased by the same flat, symmetric distribution used to randomize the anchor residue, and the φ and Ψ dihedral values for the remaining three residues were determined algebraically to ensure loop closure with ideal peptide bond geometry at the cutpoint (the bond between the first and last residues). In preliminary design calculations, we found that unique low-energy structures with energy gaps greater than ~10 k_BT (~6 kcal/mol) could only be obtained for macrocycles containing at least N/3 backbone hydrogen bonds; therefore, in subsequent sampling calculations, of the many closure solutions found, those with mainchain hydrogen bond counts below the threshold value were discarded. Third, the cyclic backbone was relaxed with the Rosetta FastRelax protocol(*27*) using the all-atom Rosetta energy function "ref2015"(*28,29*), with the `rama_prepro` and `p_aa_pp` mainchain potentials made symmetric, as described previously(*15*). Up to 10^8 samples were attempted, not all of which yielded closed solutions with the desired minimum number of hydrogen bonds.

Sampling was carried out on the "Mira" Blue Gene/Q supercomputer (Argonne labs) or Amazon Web Service (AWS) . For efficiency, a new multi-level hierarchical job distribution and data reduction scheme was implemented for use on massively parallel architecture. In performance benchmarks, this yielded linear performance scaling up to at least 250,000 CPUs. A sample command line for backbone enumeration is shown below. Note that this and all other Rosetta command lines provided here were tested with Rosetta Git SHA f44f45d89d204553bd7074af9b580e3c9c214c12 (Friday, 23 October 2017).

Listing S1: Sample command line for sampling polyglycine backbones with the `simple_cycpep_predict` Rosetta application

```
mpirun -np <processes count> <path to binary>/simple_cycpep_predict.mpi.linuxgccrelease /
  -database <path to database> @rosetta.flags
```

In the above, `<processes count>` must be replaced with the number of parallel instances of simple_cycpep_predict to run, and `<path to binary>` and `<path to database>` must be replaced with the paths to the `Rosetta/main/source/bin/` and `Rosetta/main/database/` directories, respectively. Rosetta must first be compiled for Message Passing Interface (MPI) execution. The `mpirun` command shown may need to replaced with a parallel execution command specific for the hardware of the cluster on which the run is carried out. Options for the run are contained in the file `rosetta.flags`. The example file shown below is for a 1 hour run (with sampling halted after 45 minutes, with the remaining time available for data collection and output), using a flattened symmetric proline Ramachandran table for sampling, on 131,072 CPUs on the Blue Gene/Q architecture, with the lowest-energy 1% of samples written to disk.

Listing S2: File rosetta.flags containing command-line options for sampling conformations of a 7-residue polyglycine cyclic backbone

```
-nstruct 10000000
-cyclic_peptide:MPI_stop_after_time 2700
-out:file:silent out.silent
-MPI_batchsize_by_level 100 1
-MPI_processes_by_level 1 361 130710
-cyclic_peptide:MPI_output_fraction 0.01
-cyclic_peptide:MPI_choose_highest false
-cyclic_peptide:MPI_sort_by energy
-cyclic_peptide:sequence_file inputs/seq.txt
-score:symmetric_gly_tables true
-cyclic_peptide:genkic_closure_attempts 250
-cyclic_peptide:genkic_min_solution_count 1
-cyclic_peptide:use_rama_filter true
-cyclic_peptide:rama_cutoff 3.0
-cyclic_peptide:default_rama_sampling_table flat_symm_pro_ramatable
-cyclic_peptide:min_genkic_hbonds 2
-cyclic_peptide:min_final_hbonds 2
-mute all
-unmute protocols.cyclic_peptide_predict.SimpleCycpepPredictApplication_MPI_summary
```

A polyglycine input sequence was also provided. The number of glycine residues determined the size of the macrocycle. A sample input file for a 7-residue macrocycle is shown below:

Listing S3: Sample sequence file `seq.txt` specifying a 7-residue polyglycine macrocycle

```
GLY GLY GLY GLY GLY GLY GLY
```

Energy-based clustering and data reduction

The sampling described above yields up to millions of backbones, making the problem of identifying repeatedly-sampled conformations a difficult problem in data reduction. While many algorithms for clustering large datasets have been developed(*30*–*32*), this particular problem has an interesting feature: Rosetta's energy calculations can be used to establish a rank order for the degree to which elements in the dataset are "interesting", providing a useful means of selecting cluster centers without performing a prohibitively expensive all-to-all RMSD calculation.

We developed a simple energy-based clustering algorithm for this problem: first, the energy of each input structure is scored using the Rosetta all-atom energy function (ref2015), and minimal backbone information for every structure is stored in an unclustered pool. Second, the lowest-energy structure in the unclustered pool is selected as the center of the first cluster. This structure is moved from the unclustered pool into the first cluster, and the backbone RMSD between this structure and every circular permutation of every structure remaining in the unclustered pool is calculated. Those structures for which at least one circular permutation lies within a threshold RMSD from the current cluster center are also removed from the unclustered pool and added to the new cluster. For our purposes, we typically used an RMSD threshold of 1.25 Å. Third, the lowest-energy structure remaining in the unclustered pool is selected as the center of the next cluster, and the second step is repeated. This process continues for subsequent clusters until no structures remain in the unclustered pool. Note that, unlike Voronoi clustering schemes, this "cookie-cutter" approach deliberately gives precedence to lower-energy clusters. Although simple, we found that this approach worked well for our large datasets, yielding lower-energy clusters that were particularly easy to stabilize with suitable amino acid sequences.

The algorithm described above is implemented in the Rosetta software suite as the `energy_based_clustering` application. A sample command line for clustering a large dataset stored in the Rosetta binary structure file `dataset.silent` is shown below:

Listing S4: Sample command line for clustering a large dataset of macrocycle backbones

```
<path to binary>/energy_based_clustering.default.linuxgccrelease @clustering.flags
```

In the above, `<path to binary>` must be replaced with a suitable path to the `Rosetta/main/source/bin` directory, and `linuxgccrelease` must be replaced with the particular operating system, compiler, and compilation mode for the version of Rosetta used. (For example, on a Macintosh system on which Rosetta was compiled using the clang compiler, the string would be `macosclangrelease`.) The file `clustering.flags`, shown below, contains settings that guide the clustering process:

Listing S5: File `clustering.flags`, containing command-line flags used to control the clustering process

```
-in:file:silent dataset.silent
-in:file:fullatom
-symmetric_gly_tables true
-cluster:energy_based_clustering:cyclic true
-cluster:energy_based_clustering:cluster_cyclic_permutations true
-cluster:energy_based_clustering:cluster_radius 1.25
-cluster:energy_based_clustering:limit_structures_per_cluster 10
-cluster:energy_based_clustering:use_CB false
```

Torsion bin-based clustering

We developed a custom PyRosetta Python script (provided in the file `torsion_bin_clustering.py`) for re-clustering the cluster centers from the previous, RMSD-based clustering step. Briefly, this script assigns a torsion bin string to each input structure, sorting all circular permutations in both chiralities of the bin string alphabetically and selecting the first in order to allow structures with different circular permutations to be compared easily. A string representing a hydrogen bonding pattern is also assigned to each input structure, circularly permuted to match the circular permutation of the torsion bin string. The structure is then assigned to a cluster with the same torsion bin string and hydrogen bonding pattern, or, if no such cluster has yet been encountered, a new cluster is created and the structure is assigned to that new cluster. The process is repeated until all input structures have been assigned to clusters.

A sample command line for running this script is shown below:

Listing S6: Command line for running the torsion bin clustering script

```
python torsion_bin_clustering dataset.silent
```

Computational sequence design

The Rosetta `FastDesign` module was used for sequence design. FastDesign performs alternating rounds of side-chain identity and rotamer optimization (using the Rosetta `Packer` module) and torsion-space energy minimization (using the Rosetta `Minimizer` module), with the repulsive term of the Rosetta energy function, `fa_rep`, ramped from 2% of its normal value to 100% of its normal value from round to round.

FastDesign seeks to minimize the energy of a designed structure. However, there were additional requirements that we wished to impose during the design process. Some such requirements were intended to limit the conformational flexibility of the designs produced, and to maximize the chances of the designed structure representing a unique low-energy conformation. To this end, we wished to require a minimum L- or D-proline content, for example. Other requirements were practical needs for synthesis (*e.g.* the need for at least one L-aspartate or L-glutamate in the sequence to allow resin tethering during cyclization), or for

characterization (*e.g.* the need for at least one positively-charged residue to facilitate mass spectrometry).

To this end, we implemented a non-pairwise-decomposable term, called `aa_composition`, which allowed users to define a nonlinearly-ramping penalty for deviation from a desired amino acid composition to guide the `Packer` to find sequences with desired compositions. This allowed us to require a minimum proline count, and at least one L-aspartate or L-glutamate and one positively charged residue per design. The `Packer` required significant modifications to be compatible with non-pairwise decomposable score terms. These will be described elsewhere.

We also implemented two new residue selectors, called the `PhiSelector` and `BinSelector`, to provide additional control over the `Packer`. We used these to require that the `Packer` consider only L-amino acid residues at positions with a mainchain φ value less than zero, and only D-amino acid residues at positions with a mainchain φ value greater than zero. These tools, and their RosettaScripts(*33*) interface, are described fully in the Rosetta online documentation wiki (https://www.rosettacommons.org/docs/latest/Home).

During early design runs, we found that Rosetta's normally pairwise-decomposable scoring function would erroneously favor structures in which more than two hydrogen bond donors made bonds to a single acceptor. Since it is difficult to change the hydrogen bonding architecture to give favourable scores to a maximum of two donors binding to an oxygen acceptor (since such scoring would necessarily be non-pairwise-decomposable), we instead implemented a filter, called the `OversaturatedHbondAcceptorFilter`, to discard designs with this pathology.

A sample RosettaScripts script used for design is shown below.

Listing S7: Sample RosettaScripts script used for flexible-backbone sequence design.

```
<ROSETTASCRIPTS>
        # The SCOREFXNS section defines scoring functions that will be used later in the script:
        <SCOREFXNS>
                # The current Rosetta default scorefunction:
                <ScoreFunction name="ref" weights="ref2015" />
                # The default scorefunction with increased hydrogen bond weights, and
                # with the aa_composition, aspartimide_penalty, and constraint score
                # terms activated.
                <ScoreFunction name="ref_highhbond" weights="ref2015" >
                        <Reweight scoretype="hbond_lr_bb" weight="5.0" />
                        <Reweight scoretype="hbond_sr_bb" weight="5.0" />
                        <Reweight scoretype="atom_pair_constraint" weight="1.0" />
                        <Reweight scoretype="dihedral_constraint" weight="1.0" />
                        <Reweight scoretype="angle_constraint" weight="1.0" />
                        <Reweight scoretype="aa_composition" weight="1.0" />
                        <Reweight scoretype="aspartimide_penalty" weight="1.0" />
                </ScoreFunction>
        </SCOREFXNS>
        # The RESIDUE_SELECTORS section allows users to configure tools to select
        # residues,which are used when setting up other Rosetta modules.
        <RESIDUE_SELECTORS>
```

```
        # Select residues with mainchain phi torsion values greater than zero.
        # These positions will be restricted to becoming D-amino acids during
        # design:
        <Phi name="posPhi" select_positive_phi="true" />
        # Select residues with mainchain phi torsion values less than zero.
        # These positions will be restricted to becoming L-amino acids during
        # design:
        <Phi name="negPhi" select_positive_phi="false" />
</RESIDUE_SELECTORS>
# The FILTERS section allows users to configure filters.  These measure
# properties of a structure and make decisions, based on the measured
# properties, about whether to discard the current structure.
<FILTERS>
        # Filter to avoid score function artifact of having more than two
        # hydrogen bonds to carbonyls:
        <OversaturatedHbondAcceptorFilter name="oversat" scorefxn="ref"
          max_allowed_oversaturated="0" consider_mainchain_only="false"/>
</FILTERS>
# The TASKOPERATIONS section allows users to configure task operations, which
# are Rosetta modules that control the behavior of Rosetta's "Packer" module.
# The Packer,in turn, is used for side-chain identity and rotamer optimization.
# (As such, it is the primary tool used for sequence design.):
<TASKOPERATIONS>
        # Task operation to read a resfile defining the D-amino acids, which will
        # be used for design at positions with mainchain phi torsion values
        # greater than zero:
        <ReadResfile name="d_res" filename="inputs/d_res.txt" selector="posPhi"/>
        # Task operation to read a resfile defining the L-amino acids, which will
        # be used for design at positions with mainchain phi torsion values less
        # than zero:
        <ReadResfile name="l_res" filename="inputs/l_res.txt" selector="negPhi"/>
</TASKOPERATIONS>
# The MOVERS section allows users to define movers, which are Rosetta modules
# that modify a structure in some way:
<MOVERS>
        # A mover to declare a bond connecting the termini (i.e. to cyclize the
        # peptide). In the context of Rosetta, declaring a bond tells Rosetta
        # that two atoms should
        # not have van der Waals interactions computed, but does not constrain
        #the bond geometry in any way.  Note that the variable %%Nres%%,
        # specified on the command line, is used to specify the index of the C-
        # terminal residue:
        <DeclareBond name="peptide_bond1" res1="1" atom1="N" atom2="C" res2="%%Nres%%"
add_termini="true" />
        # The following three movers are used to set up torsion, angle, and #
        # length constraints for the terminal peptide bond, ensuring that good
        # bond geometry is preserved during relaxation.  Again, the command-line
        # variable %%Nres%% is used to specify the index of the C-terminal
        # residue:
        <CreateTorsionConstraint name="peptide_torsion_constraint">
                <Add res1="%%Nres%%" res2="%%Nres%%" res3="1" res4="1" atom1="CA"
                  atom2="C" atom3="N" atom4="CA"
                  cst_func="CIRCULARHARMONIC 3.141592654 0.005" />
        </CreateTorsionConstraint>
        <CreateAngleConstraint name="peptide_angle_constraints">
                <Add res1="%%Nres%%" atom1="CA" res_center="%%Nres%%" atom_center="C"
                  res2="1" atom2="N" cst_func="CIRCULARHARMONIC 2.02807247 0.005" />
                <Add res1="%%Nres%%" atom1="C" res_center="1" atom_center="N" res2="1"
                  atom2="CA" cst_func="CIRCULARHARMONIC 2.12406565 0.005" />
```

```
        </CreateAngleConstraint>
        <CreateDistanceConstraint name="N_To_C_dist_cst">
               <Add res1="%%Nres%%" res2="1" atom1="C" atom2="N"
                 cst_func="HARMONIC 1.32865 0.01" />
        </CreateDistanceConstraint>
        # Composition constraints are used with the aa_composition score term in
        # order to add a nonlinearly-ramping penalty for deviation from a desired
        # amino acid composition.  In this case, we use them to require at least
        # two proline residues (L- or D-), at least one L-aspartate or L-
        # glutamate, and at least one positively-charged residue.
        <AddCompositionConstraintMover name="addcompcsts"
    filename="inputs/desired_makeup.comp" />
        # The FastDesign mover performs alternating rounds of sequence design and
        # torsion-space energy minimization, while ramping the repulsive term in
        # the scorefunction (fa_rep).  We use it here with a modified
        # scorefunction with constraints and aa_composition energy terms
        # activated:
        <FastDesign name="fdes" scorefxn="ref_highhbond" repeats="3"
          task_operations="d_res,l_res" ramp_down_constraints="false" >
               # A MoveMap is used to specify which degrees of freedom can move
               # and which are fixed during energy minimization.  Here, we
               # indicate that all mainchain torsions (bb) and all sidechain
               # torsions (chi) can move:
               <MoveMap name="fdes_mm" >
                      <Chain number="1" chi="true" bb="true" />
               </MoveMap>
        </FastDesign>

    </MOVERS>
    # The PROTOCOLS section is the section in which the user invokes the modules
    # defined above in linear sequence to define a protocol:
    <PROTOCOLS>
    <Add mover="peptide_bond1" />
    <Add mover="peptide_torsion_constraint" />
    <Add mover="peptide_angle_constraints" />
    <Add mover="N_To_C_dist_cst" />
    <Add mover="addcompcsts"/>
    <Add mover="fdes" />
        # A side-effect of the DeclareBond mover is the correction of positions
        # of H and O atoms that depend on the peptide bond.  We re-invoke it here
        # for that purpose:
        <Add mover="peptide_bond1" />
        <Add mover="peptide_bond1" />
        <Add filter="oversat" />
    </PROTOCOLS>
    # The OUTPUT section allows the user to define output settings.  Here, we
    # specify the scoring function that will be used to score the output structure
    # for the score written in the output PDB file.
    <OUTPUT scorefxn="ref"/>
</ROSETTASCRIPTS>
```

The additional files used in the script (which must be put into an `inputs/` directory) are as follows:

Listing S8: File `L_res.txt`, used to specify permitted L-amino acid residues for design

---

```
RESET NOTAA CG
start
```

Listing S9: File `D_res.txt`, used to specify permitted D-amino acid residues for design

```
EMPTY NC DAL NC DAS NC DGU NC DPH NC DHI NC DIL NC DLY NC DLE NC DME NC DAN /
NC DGN NC DAR NC DSE NC DTH NC DVA NC DTR NC DTY NC DPR
start
```

Listing S10: File `desired_makeup.comp`, used to place constraints on the desired amino acid composition for design

```
# At least two proline residues.
# These can be L- or D- (or mixed).
PENALTY_DEFINITION
TYPE PRO DPR
DELTA_START -2
DELTA_END 1
PENALTIES 500 10 0 0
ABSOLUTE 2
BEFORE_FUNCTION QUADRATIC
AFTER_FUNCTION CONSTANT
END_PENALTY_DEFINITION

# At least one L-asp or L-glu.
PENALTY_DEFINITION
TYPE ASP GLU
DELTA_START -1
DELTA_END 1
PENALTIES 200 0 0
ABSOLUTE 1
BEFORE_FUNCTION QUADRATIC
AFTER_FUNCTION CONSTANT
END_PENALTY_DEFINITION

# At least one positively-charged residue
PENALTY_DEFINITION
TYPE LYS ARG DLY DAR
DELTA_START -1
DELTA_END 1
PENALTIES 200 0 0
ABSOLUTE 1
BEFORE_FUNCTION QUADRATIC
AFTER_FUNCTION CONSTANT
END_PENALTY_DEFINITION
```

Assuming that cyclic peptide backbones from earlier clustering steps are contained in the Rosetta binary file `inputs/backbones.silent`, and the script itself is file `inputs/design_script.xml`, then the command line to run the script above for 7-residue cyclic peptides is shown below:

Listing S11: Sample command line for running the design RosettaScripts script shown above

```
<path to binary>/rosetta_scripts.default.linuxgccrelease -parser:protocol design.xml /
  -in:file:silent inputs/backbones.silent -in:file:fullatom -parser:script_vars /
  Nres=7 -parser:protocol inputs/design_script.xml
```

Computational validation: Energy landscape analysis of designed macrocycles and their mutants

For each torsion bin string and hydrogen bonding pattern, the lowest-energy sequence designed was picked as a representative of that cluster. A subset of such low-energy structures (from 44% of all designs for length 7 to 3% of all designs for length 10) was subjected to a final round of computational validation using the `simple_cycpep_predict` application, as described previously(*15*). As for the sampling of polyglycine conformations, large numbers of backbone conformations were sampled for each sequence tested, but this time, the sampling was biased based on the Ramachandran map of each amino acid residue in the sequence. Each sample was subjected to full side-chain rotamer optimization and energy minimization using the Rosetta `FastRelax` protocol(*27*). The "foldability" of each macrocycle was evaluated based on the estimated fractional occupancy of the native state (a value that we call P_{Near}), and on the energy gap between the native structures and other low energy models, as reported previously(*15*). A P_{Near} value of > 0.9 and energy gap of < -0.1 was selected as the basic threshold for acceptance. Additionally, the plot of energy *vs.* RMSD was then visually inspected.

For a subset of macrocycles, large-scale landscape analysis was performed. Each residue in the initial sequence was systematically mutated to the other 18 amino acid residues of the same chirality, and to alanine with mirror chirality, in the input sequences provided to the `simple_cycpep_predict` application. These large scale computational analyses of the energy landscape was performed using the Berkeley Open Infrastructure for Network Computing (BOINC) as part of the Rosetta@Home project, mostly using volunteer cellular telephones as the computing hardware (though some earlier predictions were carried out using volunteer desktop computers, or using the Argonne "Mira" Blue Gene/Q system used for poly-Gly conformational sampling).

Scrambled sequences were generated by randomly assigning residues to different positions in the structure (see fig. S13 for details of assignment).

After generation of results, site-saturation mutagenesis plots were generated based on P_{Near} values (see equation below for P_{Near}) for each structure, with λ set to 1 Å and a value of k_BT of 0.62 kcal/mol (equivalent to 37 °C). For two of the macrocycles, different combinations of λ and k_BT (0.5, 0.75, 1, 1.5 for l and 0.5, 0.75, 1, 2 for k_BT) were tested and the value with more dynamic range (*i.e.* values that showed the difference between a high-quality *vs.* low-quality energy funnel best) were selected. Double mutants were generated and analyzed using similar methods described above.

Equation 1: Definition of P_{Near}, a measure of the quality of an energy function. P_{near} approximates the Boltzmann-weighted probability of finding the structure in a conformation near the native conformation.

$$P_{near} = \frac{\sum_{i=1}^{N} exp\left(-\frac{rmsd_i^2}{\lambda^2}\right) exp\left(-\frac{E_i}{k_B T}\right)}{\sum_{j=1}^{N} exp\left(-\frac{E_j}{k_B T}\right)}$$

Turn type analysis and measurement of RMSD to hot loops

We defined a turn as a semi-independent part of a macrocycle structure that is connected internally through backbone-to-backbone hydrogen bonds, but which lacks hydrogen bonds to other parts of the structure. For each structure, different turn types were defined by their torsion bin strings and hydrogen bond patterns. Similar analysis was performed on a subset of structures from the PDB, and the frequencies were then calculated and compared. The redundancy of the PDB subset was reduced to 30% -- that is, no two PDB chains in the set had more than 30% sequence identity.

From all the hot loops generated by Kritzer and coworkers(*24*), those that contained continuous stretches of amino acids were selected. Each loop, and small truncations of it (one residue shorter from each side) were then compared to a library of macrocycles that passed computational consistency check. For every motif and scaffold, a matrix of pairwise distances between C-alpha atoms and a vector of dihedral angles for every four consecutive C-alpha atoms was computed. For every possible alignment of linear motif to cyclic scaffold, Root-Mean-Square of the differences of both the distance matrices (distance RMS) and the dihedral vectors (dihedral RMS) is reported. Macrocycles that at least in one position had a distance RMSD of less than 1 Å and a dihedral RMSD of less than 10 degrees (*i.e.* contained a portion matching the motif backbone) were considered to be plausible stabilizing scaffolds for the given motif. A complete list of these hot loops and the results are available as a supplementary file.

Synthesis, purification, and mass spectrometry of macrocycles

All peptides were synthesized using standard Fmoc solid phase peptide synthesis (SPPS) on preloaded and sidechain-linked Fmoc-Asp(Wang resin LL)-ODmab or Fmoc-Glu(Wang resin LL)-ODmab resin. Linear, protected peptides were built on a CEM Liberty Blue Peptide Synthesizer with microwave heating at coupling and deprotection steps. After the final Fmoc deprotection, the resin was treated with 2% (v/v) hydrazine monohydrate in dimethylformamide (DMF) to remove the C-terminal Dmab protecting group; the N- and C-termini were then joined on-resin by a coupling reaction. A cleavage cocktail of TFA:Water:TIPS:DODT (92.5 : 2.5 : 2.5 : 2.5) used for global deprotection of side-chains and to cleave the peptide from the resin. After the removal of residual TFA by evaporation, peptides were ether precipitated and further purified using RP-HPLC.

Crude peptides were purified using an Agilent Infinity Preparative HPLC with an Agilent Zorbax SB-C18 column (9.4 mm X 250 mm). A linear gradient of 1%/min for Solvent B (ACN with 0.1% TFA) and flow rate of 5 ml/min was used for purification to collect fractions with pure peptides. Mass and purity of peptides were were confirmed using electrospray ionization mass spectrometry (ESI-MS) on a Thermo Scientific TSQ Quantum Access mass spectrometer.

For disulfide-stapled peptides, cyclic reduced peptides were air-oxidized in 0.1 M ammonium bicarbonate buffer (pH 8.3) for 48 hours, and purified again using RP-HPLC. Some of the disulfide-containing peptides were synthesized with Fmoc-Cys(Acm)-OH at the cysteine positions. Following synthesis and cyclization, the resin was treated with 8 eqs. of iodine in 4:1 DMF:methanol overnight to remove the Acm protecting groups and facilitate disulfide bond formation. After iodine treatment, the resin was washed with 2% w/v ascorbic acid in DMF, rinsed with dichloromethane (DCM) and cleaved and purified as normal.

Nuclear magnetic resonance (NMR) spectroscopy studies of the designed macrocycles

Each peptide macrocycle was dissolved at concentrations of ~5 mg/mL at a pH between 3.0 and 5.5 in 10% D_2O, with up to 5% glycerol-d_8 added. All NMR data were collected on a DRX 500 MHz, an Avance III 600 MHz, or an Avance III 800 MHz spectrometer, equipped with TCI cryoprobe and triple-axis gradient (Bruker). Unless otherwise noted, all NMR data were collected at 5 ^oC and 25 ^oC using pulse sequences with excitation sculpting water suppression. Data were processed with TOPSPIN v. 3.5 (Bruker) or NMRPIPE(*34*) and visualized with Sparky. Initial screening of designed cyclic peptides for discrete structure involved recording 1D ^1H spectra at 25 ^oC and selecting peptides with sharp, and well dispersed backbone amide resonances. The small size of the peptides (<=14 residues) selected for structural analysis allowed for complete proton backbone and side chain resonance assignment using 2D [^1H,^1H] TOCSY; including many stereospecific assignments. To facilitate quantitative evaluation of internuclear distances, sample temperatures were dropped to 5 ^oC and both 2D [^1H,^1H] -ROESY with a 200 ms mixing time and 2D [^1H,^1H] NOESY spectra were collected using mixing times of 100 ms and 500 ms. For designs 8.1 and 14_SS a full NOESY buildup curve (50-75 ms mixing time) was collected to ensure linear behavior of the glycerol containing samples of small peptides (fig. S28). Because it is currently not economical to prepare uniformly ^{13}C and ^{15}N-labelled peptides using solid phase methods, and because natural abundance experiments are resource-intensive, only a set of ^{15}N assignments were measured using natural abundance 2D [^{15}N,^1H] SOFAST HMQC for designs 7.1, 7.2, 8.1, 8.2, 12_SS, 14_SS. For longer peptide designs or designs with clear overlap in the 2D [^1H,^1H] TOCSY we also collected natural abundance 2D [^{13}C,^1H] HMQC.

Nuclear Overhauser Effect (NOE) constraint consistency check

To evaluate whether NOE constraints alone can predict the designed structure, we first used Rosetta to relax 5 macrocycles from the Protein Data Bank (PDB) and Cambridge Structural Database (CSD) that shared the same criteria as our peptides (4ME6 from the PDB and CUQYUI, DUYTIA, JIANGO, and UZUKUW from the CSD); this was repeated 20 times. Based

on the observed distribution of energies after relaxation, we set the following filters for the score terms below and selected structures that passed these filters from our previous landscape analysis:

```
omega=1, fa_rep=10, fa_intra_rep=0.5, pro_close=5, rama_prepro=3
```

Each structure was then rescored, using Rosetta, based solely on how well it satisfied the NOE constraints and the scores *vs.* RMSD to design were plotted, as shown in fig. S12.

NMR structure determination of designed macrocycles

A set of 200 structures were calculated for well-behaved designs with the Xplor-NIH software package using torsion angle dynamics and simulated annealing. Initial folding was conducted from a single starting template of randomized torsional angles for the cyclic peptide after patching L- or D- stereoisomers. Distance restraints were derived from NOE intensities at 100 and 500 ms mixing times in 2D [^{1}H,^{1}H] NOESY spectra recorded at 500 or 800 MHz and were sorted into Strong (2.5 0.7 0.7), Medium (3.5 1.5 1.5) and Weak (4.5 2.0 2.0) bins based on relative peak intensities to aromatic resonance signals. A soft square potential was used for NOE restraints for initial folding and convergence was established when there were no NOE violations greater than 0.5 Å of the calculated structures.

After initial folding, hydrogen-bonding restraints were inferred from proximal atoms, identified by cross-strand or nearest neighbor amide NOE cross peaks in the 2D [^{1}H-^{1}H] NOESY or monitoring slow exchanging protons with 1D ^{1}H CLEANEX-PM pulse sequences (mixing time 0-500 ms). After backbone hydrogen bonding was established, structures were re-calculated as described incorporating hydrogen bonds as NOE restraints using a biharmonic potential. Throughout folding and refinement, only NOE and van der Waals terms were active during structure calculation. Due to lack of uniform labeling and peak overlap we were not able to make clear predictions of backbone dihedral angle restraints or coupling constants. The torsional database constraints were also left unrestrained due to lack of sufficient information for handling D-amino acids. To refine the structures based on NMR experiments, we launched MD simulations with NOE constraints. In particular, for each structure, a simulated annealing from 350K to 310K followed by a 10-ns production run was performed(*35*). For each atom pair measured by NOE, a distance restraint (k=1000 kJ mol^{-1} nm^{-1}) was applied throughout the simulation. The 20 conformations with the lowest total energy were selected for further analysis.

MD simulation of designed macrocycles

Molecular Dynamics simulations were performed using GROMACS 2016.1(*36*, *37*) with the Amber 99SB-ILDN forcefield(*38*). Each peptide was solvated in a dodecahedron box of explicit TIP3P waters(*39*) and neutralized with either sodium or chloride ions. The solvated systems were energy-minimized using the steepest descent minimization method. Next, the system was equilibrated for 1 ns under the NPT ensemble with position restraints (1000 kJ mol^{-1} nm^{-1}) applied on all the heavy atoms of the peptide. During this equilibration, pressure coupling to 1

atm was performed with the Berendsen barostat(*40*), and temperature coupling to 310 K using the velocity-rescaling thermostat(*41*). From each equilibrated system, 10 simulations of 100 ns were performed in the NVT ensemble. The systems were simulated using periodic boundary conditions. A cutoff at 10 Å was used for van der Waals and short-range electrostatic interactions. The Particle-Mesh Ewald (PME) summation method was used for the long-range electrostatic interactions(*42*). The Verlet cut-off scheme was used(*43*). All chemical bonds were constrained using the LINCS algorithm(*44*). The integration time-step was 2 fs, and simulations were analyzed using GROMACS tools. We calculated the root-mean-square deviation (RMSD) of the position of the C_α atoms of the peptides, compared to the initial conformation, using *gmx rms*. The peptides were aligned to the C_α of the initial conformation. The Ramachandran plots were calculated using *gmx chi,* and plotted using the Matplotlib histogram2d function (https://matplotlib.org/citing.html).

For two of the structures, design 8.1 and 10.1, we also performed our analysis for the mirrored structure of the designs to make sure that our calculations are not energetically biased against L- or D-amino acids. As shown in fig. S29, the results are comparable; thus, we only performed simulations of the designed structure (and not its mirror image) for the rest of the macrocycles.

For designs 7.3 and 7.4, we performed long (>1 μs) molecular dynamics simulations to analyze the dynamics of folding and different conformations explored by the macrocycles. The Markov state model that captures movement of the macrocycle was generated by MSM builder, and the dynamics of movement were described using a time-structure independent component analysis (tlCA) model(*44*–*47*).

Ion mobility spectrometry analysis

The single-site mutant libraries of design 7.1 were synthesized with a process similar to that described above with an additional step. For the residues for which the mutation was made (dPro4 or Thr5), the resin was removed from the synthesizer and split into 6 pools. Each pool had its respective amino acid coupled individually using the synthesizer (D-Pro,D-Ser, D-Asn, D-Asp, D-Met, D-Arg for position 4 and Thr, Ser, Leu, Gln, Glu, Trp for position 5). After all pools of resin were loaded with the desired amino acid they were recombined and the remaining amino acids in the sequence were coupled as normal. Cleavage of the resin was performed using the same cleavage cocktail described above. All expected species were confirmed by mass spectrometry.

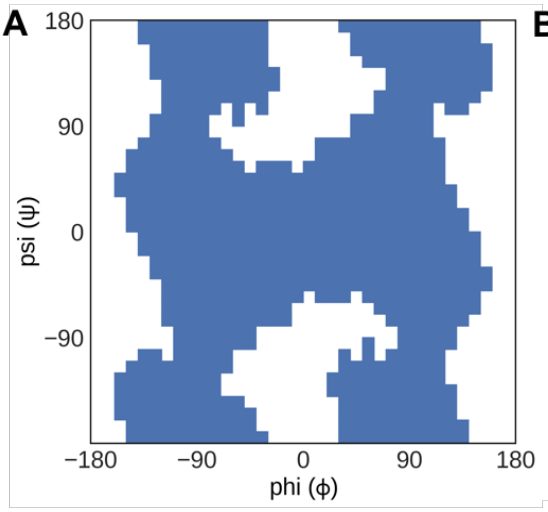
All samples were prepared in 50% aqueous methanol acidified with 0.1% formic acid. The solutions were infused at an infusion rate of 300 nL/min and electrosprayed in the positive mode using an etched emitter (20 mm i.d.). The formed ions were transmitted through a heated inlet capillary (130^{o}C) into a high-resolution Structures for Lossless Ion Manipulations Ion Mobility Mass Spectrometer (SLIM IM-MS) platform for high resolution ion mobility spectrometry(*21*). Ions were accumulated in an ion funnel trap(*48*) for 2 ms and then released to SLIM IM-MS. The SLIM module was similar to that of the SLIM serpentine design previously reported(*49*, *50*), but has a path length of 15.9 m that allows for multiple passes through the serpentine path for

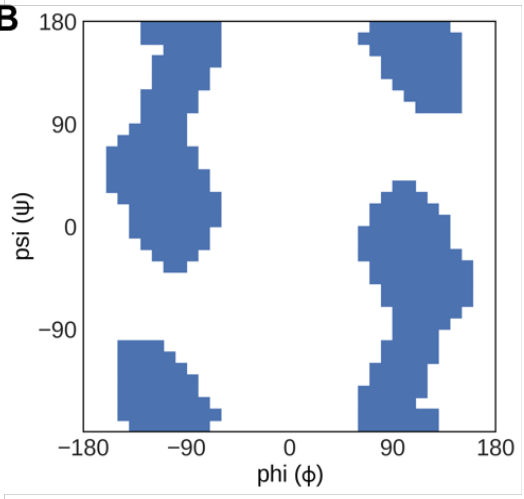
higher ion mobility spectrometry resolution. The SLIM module was integrated with an Agilent 6224 TOF MS equipped with a 1.5 m extended flight tube via a rear ion funnel and RF-only quadrupole. All SLIM separations were performed at ~2.5 Torr N_2 with the following parameters: wave speed of 160 m/s, wave amplitudes of 40 V, guard electrode voltage of 6 V, and RF frequency of 1.0 MHz and amplitude of 380 V_{p-p}. Data were acquired on an 8-bit ADC (analog-to-digital converter) using a control software developed in-house.

Protease assay:

Protease assay was performed using PRONASE® Protease derived from *Streptomyces griseus* from EMD Millipore (product# 53702). 0.2 µmole of each peptide tested was added to 200 µl of 50 mM ammonium acetate buffer, pH 8, supplemented with 0.01 M calcium acetate. 5 µl of this starting material was mixed with µl TFA and kept as the time 0 sample. To this mixture we added 2 µl of 2 mg/ml protease mix stock (prepared by dissolving in water) and incubated at 37 ˚C. At different time points, 5 µl of the reaction mixture was taken out and quenched by addition of 5 µl TFA. To track protease cleavage, each sample was analyzed by LC/MS (Thermo Scientific Accela HPLC system connected to Thermo Scientific TSQ Quantum Access mass spectrometer) using an Agilent ZORBAX StableBond 300 C18, 4.6 x 150 mm, 5 µm as the chromatography column.

Supplementary figures

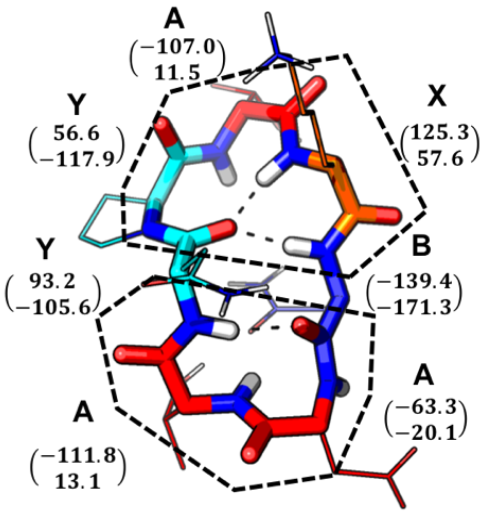




C

AAAYBYB	AA AAAa		AAYBXXB	BXXb	AA
AAAYYAB	YA	AAAa	AAYBXYY	BX BX	YA
AABBXBY	YAAa		AAYYAXB	YAXa	AA
AABBYBY	YA YAAa		ABABXYY	ABX ABX	YA
AABYBYB	AA		ABBBXBY	BX	YA
AAXABYY	YA YAAa		ABBXAYY	BXAa	YA
AAXBYAB	YA YA	AA AAXa	ABBYAYY	BYA YA YA	YA YA
AAXYBXY	BX	AAXa	ABXYABY	YA YA	YA
AAXYYAB	YA	AAX		YA	YA
AAYAAYY	YAAb YAAb YA YA AA AA AA YAAa	YA YAAb YA YAAb YA YAAa YAAab	ABXYABY	YA	
		AA	ABXYBYB	BX BX	ABX
			ABXYBYB	ABX	
			ABYABYY	YA	YA
			ABYBXAY	BX	YA
			ABYBYBY	YA	
AAYABYY	YAAa		AXBYAYY	YA YAXa	
AAYBXAY	BXAa BX	AA YAAb			

D



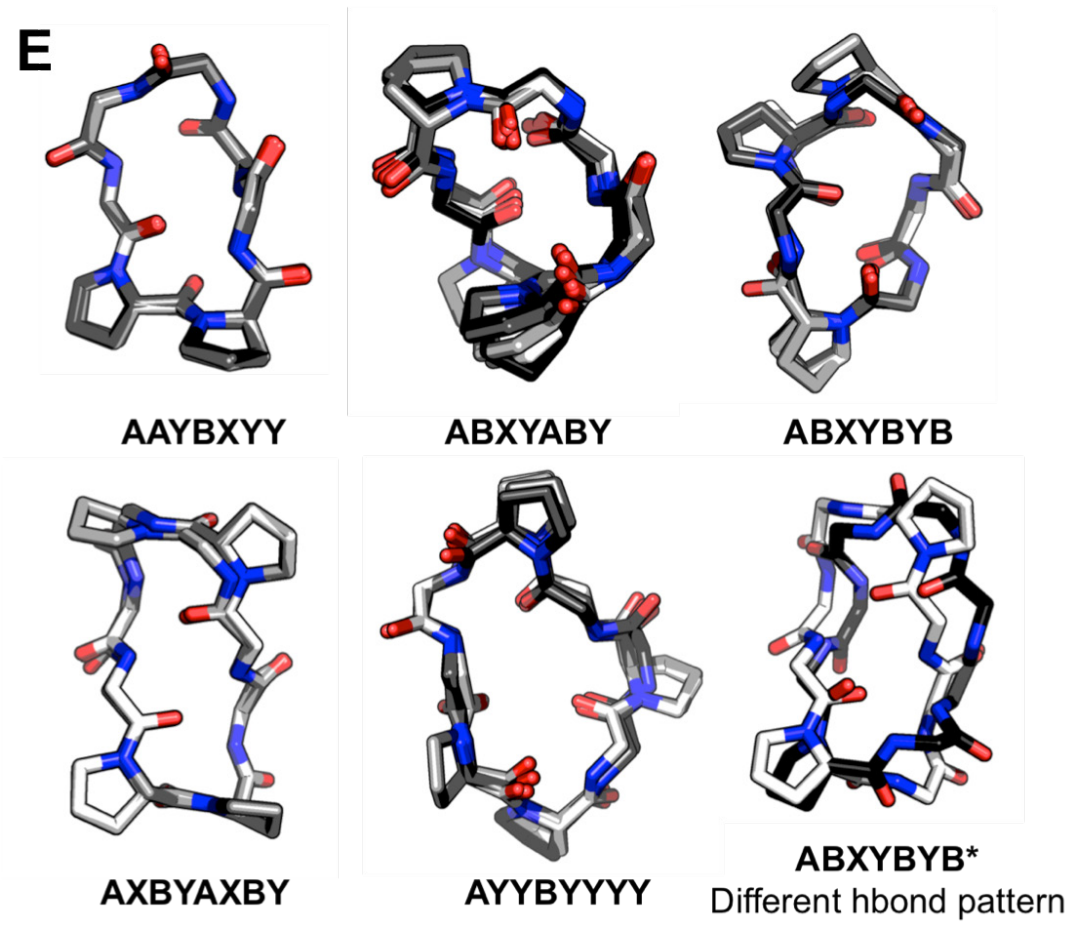


fig. S1. (A), (B) the flat-bottomed Gly (A) and Pro (B) symmetric Ramachandran table used for sampling backbone conformations. (C) shows all the clusters obtained for macrocycles with 7 residues. Designs were clustered based on the backbone torsion bin strings and then sub-clustered based on turn types and hydrogen bond patterns. The red box in the table corresponds to the structure in (D), for which the torsion angle bin strings, the torsion angles, as well as the two turn types are depicted. (E) members of the clusters with same hydrogen bond pattern overlay with each other. Both the torsion bin string and the hydrogen bond pattern are important for defining a cluster.

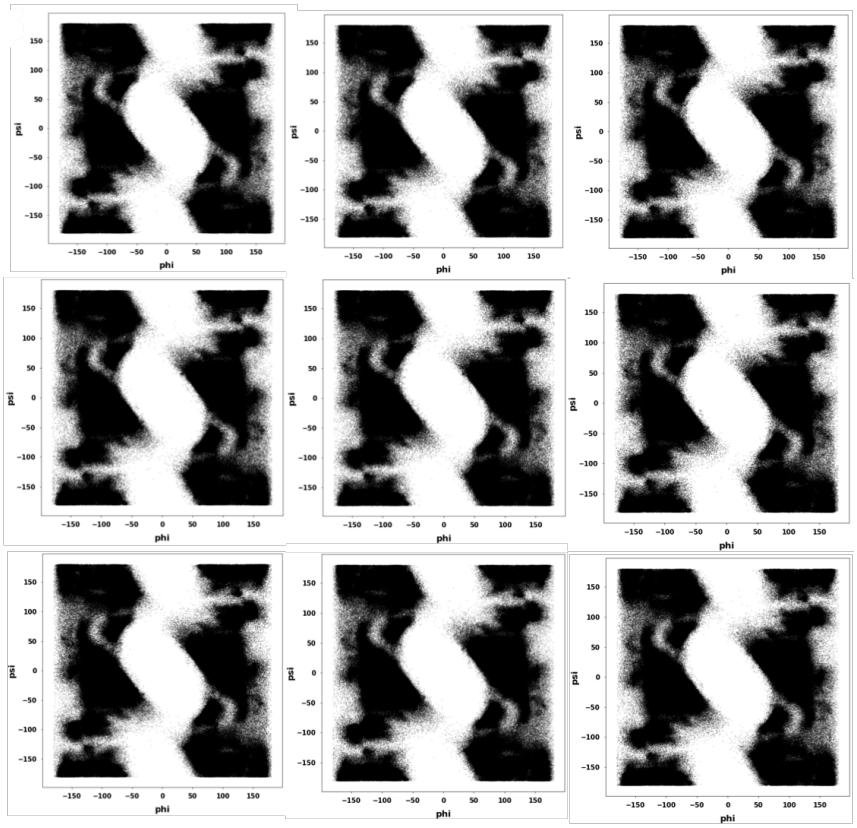


fig. S2. Phi-psi distributions at individual residue positions for 9 residue macrocycle samples. Sampling is clearly complete at the individual residue level.

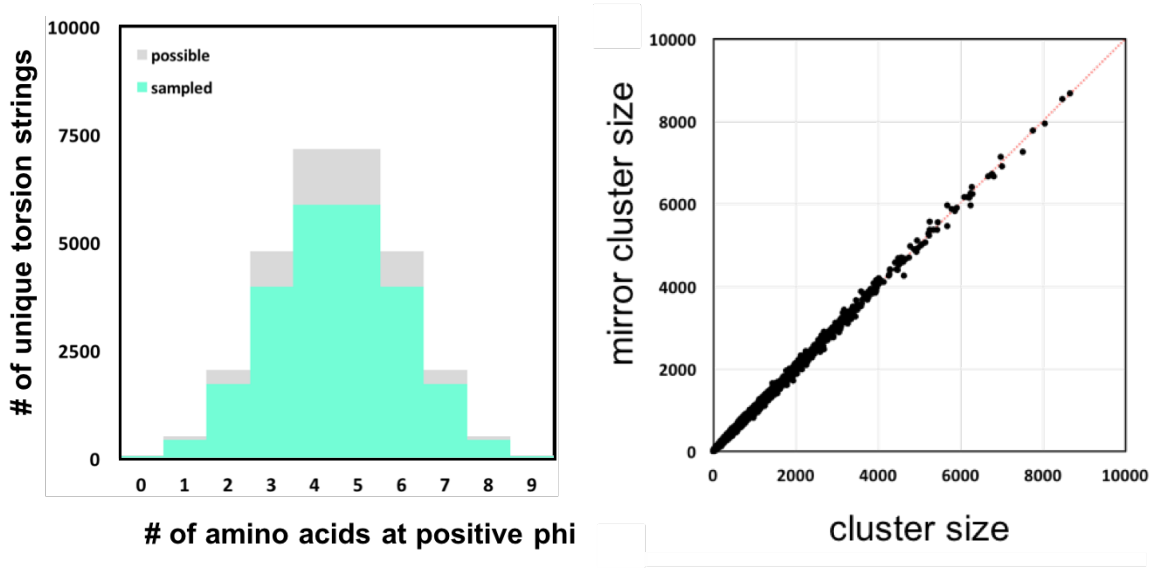


fig. S3. Cluster distribution. Left panel: Number of unique clusters for 9 residue macrocycles as a function of number of D-amino acids for all possible torsion bin strings and those sampled with more than 10 members. The unsampled strings correspond to backbone conformations not compatible with chain closure (for example the helical bin string AAAAAAAAA). Right panel: Designs were clustered based on the torsion bin strings, and the number of designs in each cluster plotted against the number of designs with the mirror image bin string. The similarity in the frequency of sampling of both chiralities is another indication of the convergence of sampling. Tests on individual clusters sampled at very different frequencies show that the differences in sampling frequency arise from differences in the frequency with which the corresponding conformations can be closed by the kinematic closure algorithm.

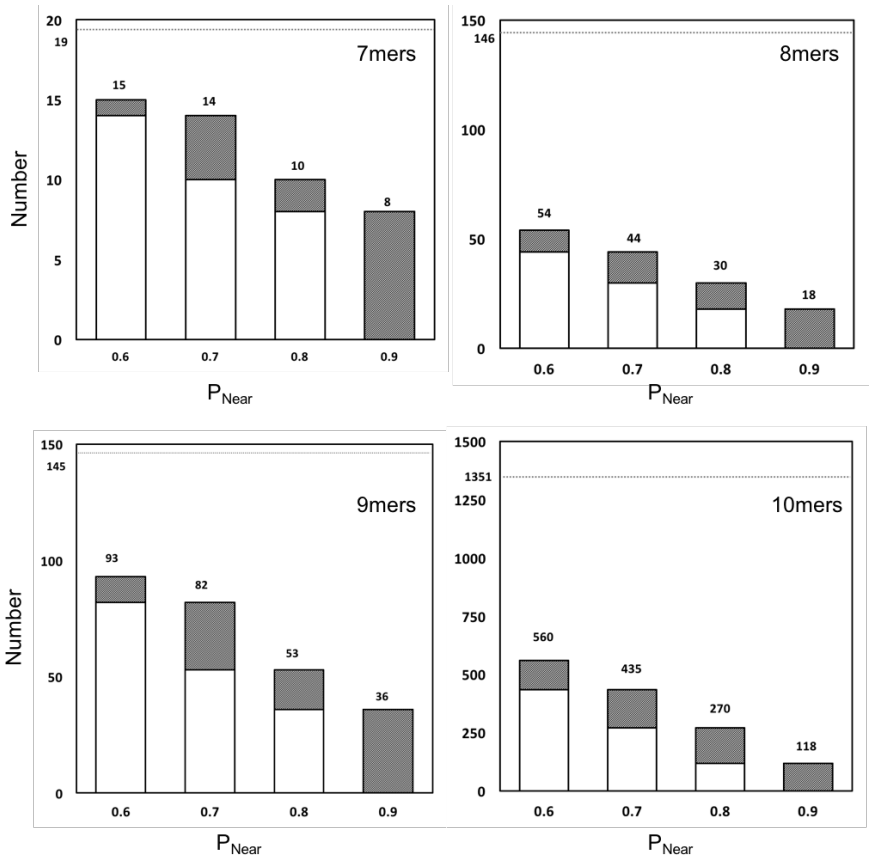


fig. S4. Computed energy gaps for designed macrocycles of different lengths. The total number of designs for which energy landscape calculations were carried out for each length is indicated by the dashed line near the top of each panel. The number of designs in different P_{Near} ranges is shown in grey bars, and the combined grey plus white bars indicate the cumulative distribution. As expected given the much smaller conformational space accessible to 7 residue version 10 residue macrocycles, for the former, roughly half had P_{Near} > 0.9 and nearly all had P_{Near}> 0.6, whereas for the latter, less than a tenth had P_{Near} > 0.9.

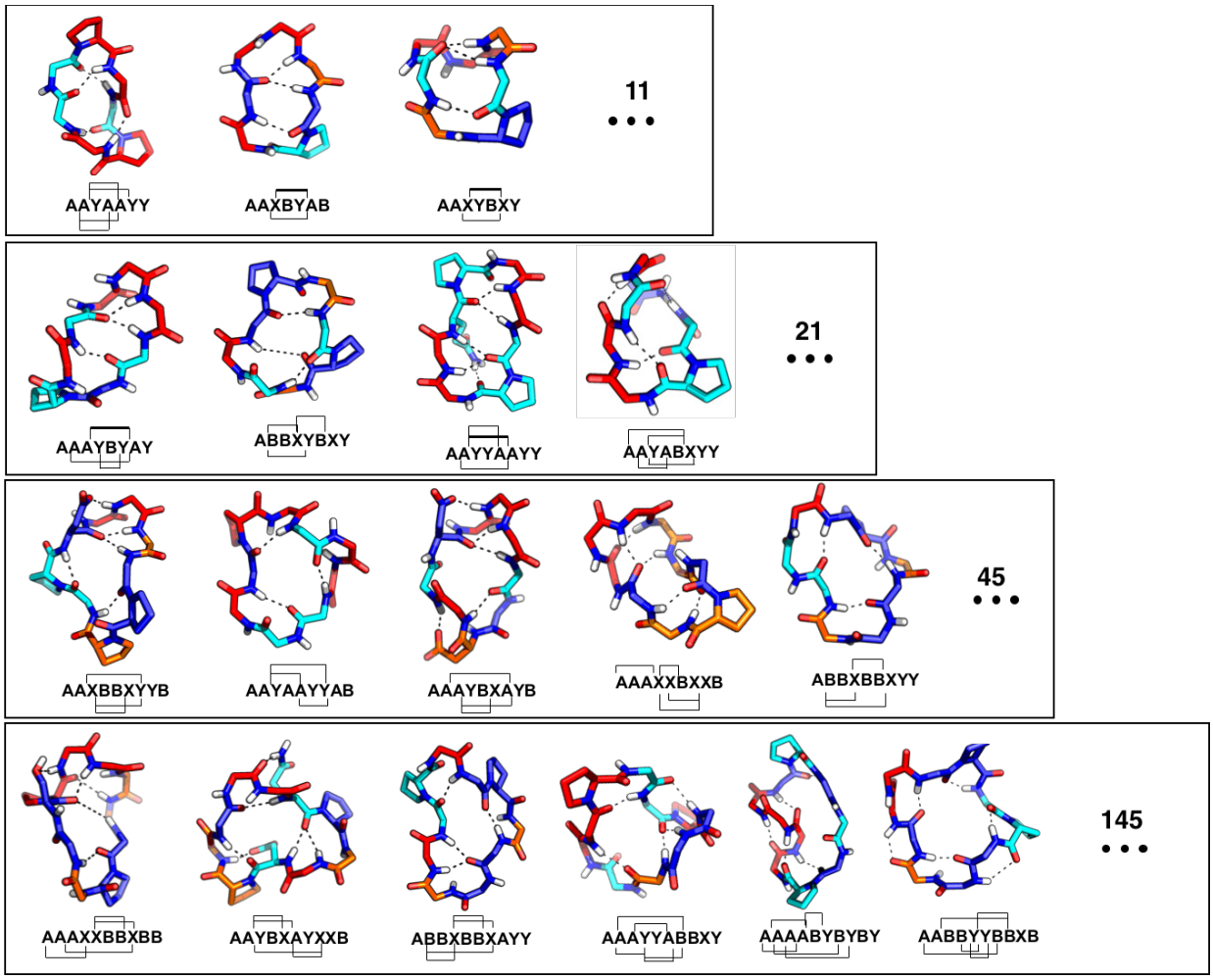


fig. S5. Examples of designed macrocycles with large computed energy gaps. 7-10 residue macrocycles are shown in rows 1-4, respectively. The torsion bin strings and the backbone-backbone hydrogen bonds are shown under each structure. Each amino acid is colored based on its backbone torsion angle (fig. 1A, red=A, blue=B, orange=X, cyan=Y).

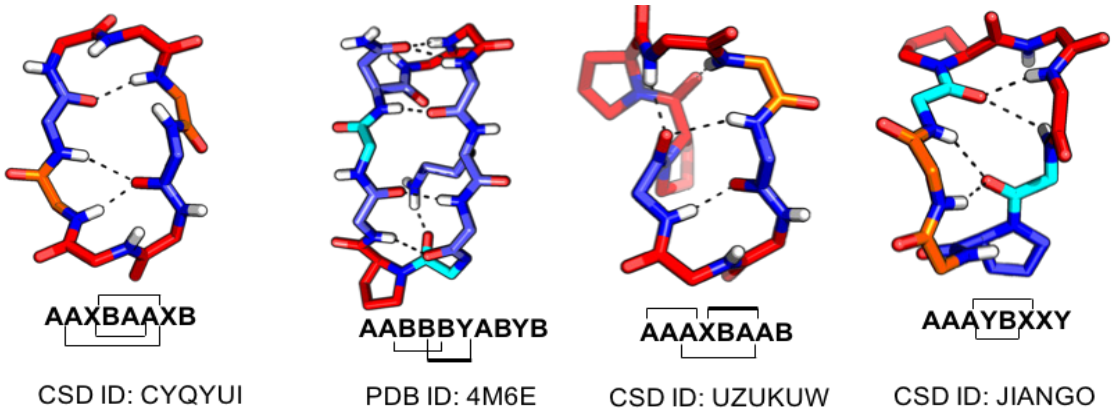


fig. S6. Structures of previously described macrocycles in size range of 7-10 with no crosslinkers (including ester linkages (depsipeptides) and thioether staples), no backbone N-methylation, and not in the context of a binding partner. The structure ID, torsion bin strings, and backbone-backbone hydrogen bonds are shown under each structure. Each amino acid is colored based on its backbone torsion angle (red=A, blue=B, orange=X, cyan=Y).

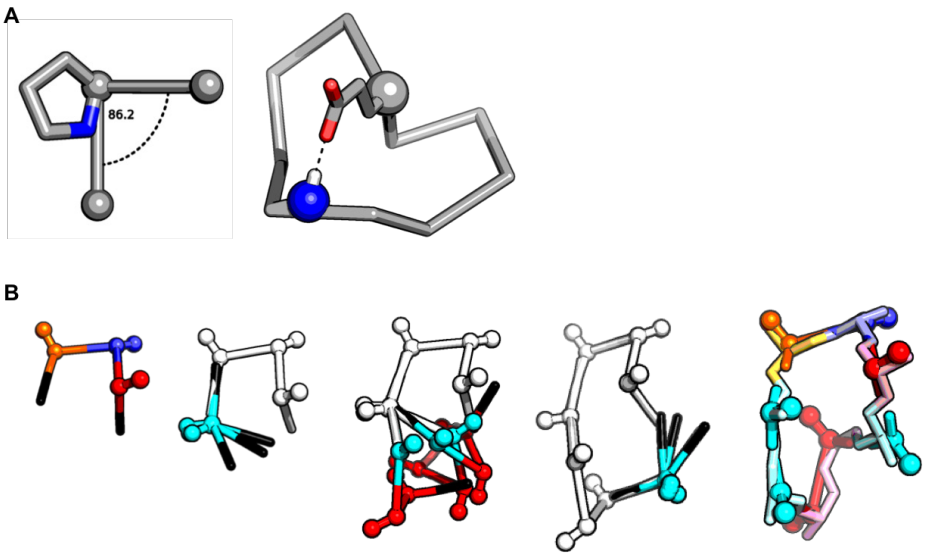
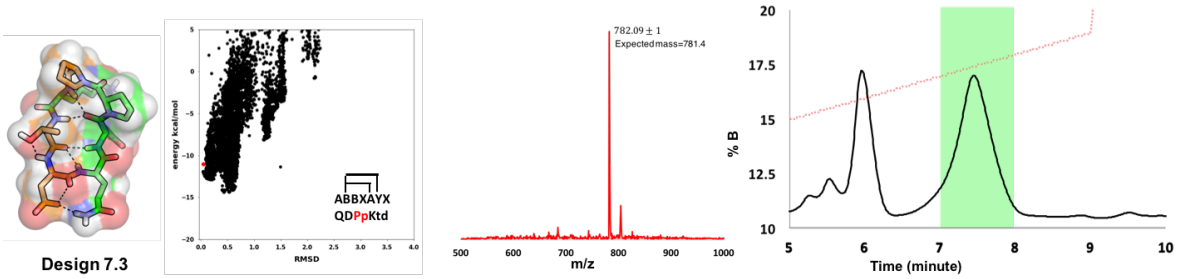
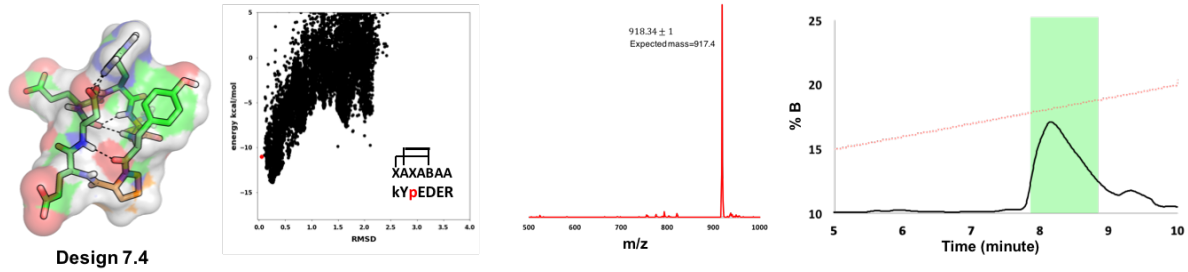
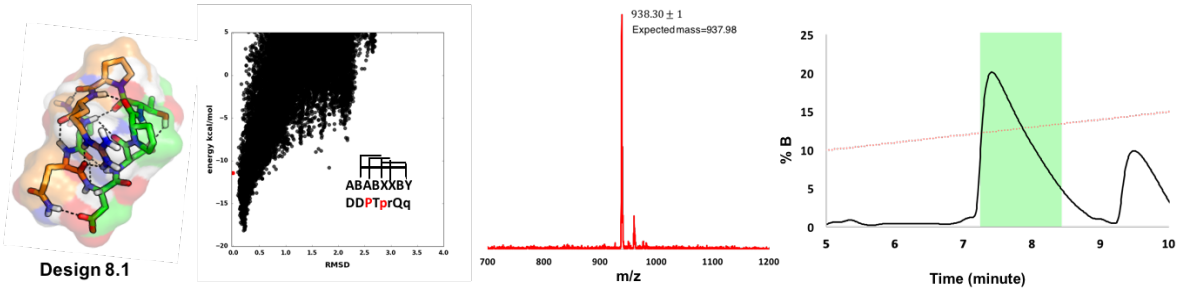
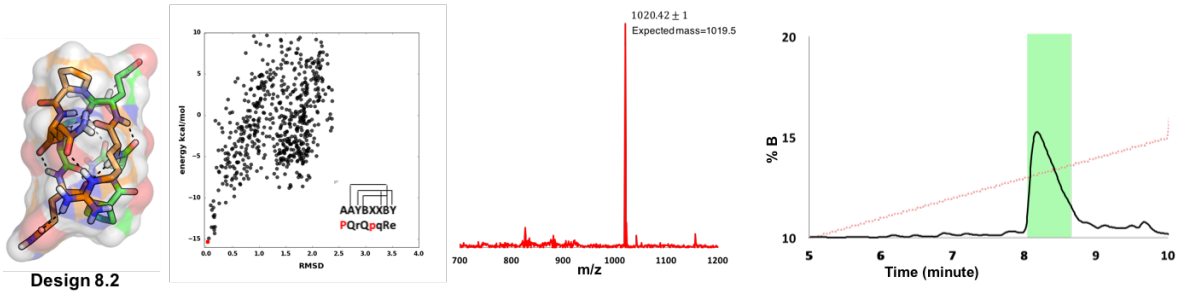


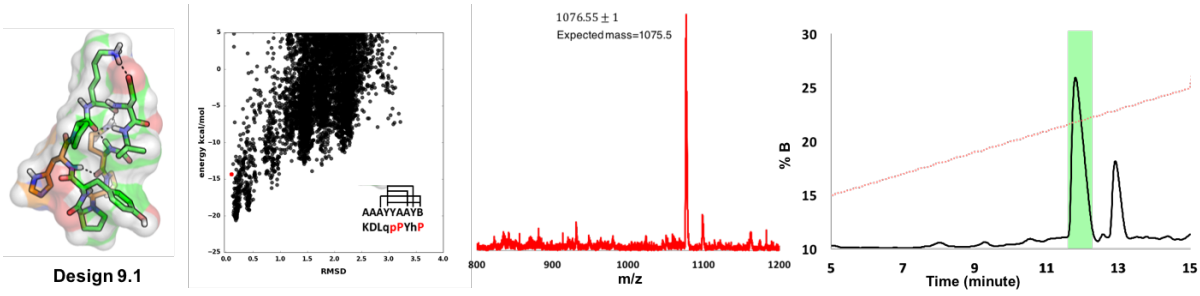
fig. S7. Macrocycle building blocks. Top: Representative structure for two non-turn building blocks: proline-stabilized kinks on the left and long range sidechain-backbone hydrogen bonds on the right. Bottom: Macrocycle structures can be viewed as sequences of rigid building blocks. Starting from an ABX turn (the leftmost panel), addition of building blocks generate a structure for a potential design with bin string ABXYYAAY. For each linkage point, different torsions of the first amino acid are explored and the final solution (shown as discrete ball and stick models in the rightmost figure) is the one that can generate a closed cycle. As shown in the final panel on the right, the assembled structure matches the design. The assembled building blocks are from different backbones.


Design 7.3
energy kcal/mol
RMSD
ABBXAYX
QDPpKtd
782.09 ± 1
Expected mass=781.4
m/z
% B
Time (minute)


Design 7.4
energy kcal/mol
RMSD
XAXABAA
kYpEDER
918.34 ± 1
Expected mass=917.4
m/z
% B
Time (minute)


Design 8.1
energy kcal/mol
RMSD
ABABXXBY
DDPTprQq
938.30 ± 1
Expected mass=937.98
m/z
% B
Time (minute)


Design 8.2
energy kcal/mol
RMSD
AAYBXXBY
PQrQpqRe
1020.42 ± 1
Expected mass=1019.5
m/z
% B
Time (minute)


Design 9.1
energy kcal/mol
RMSD
AAAYYAAYB
KDLqpPYhP
1076.55 ± 1
Expected mass=1075.5
m/z
% B
Time (minute)

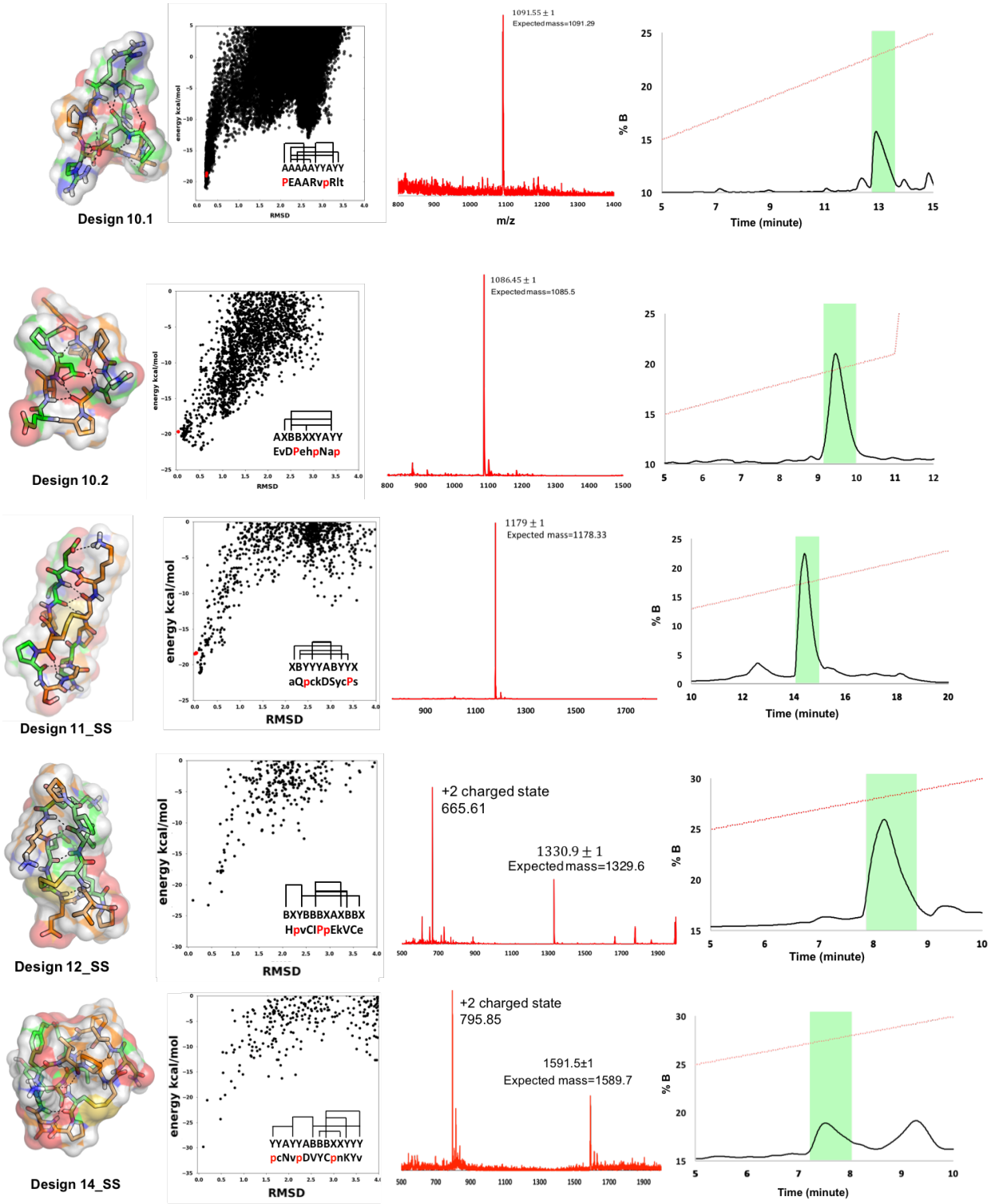


fig. S8. For each peptide with experimental structure within 1.5 Å of the designed structure, the initial design, computational folding landscape, ESI spectra and the HPLC traces are shown from left to right, respectively. The bin string and sequence of each peptide as well as the main

chain hydrogen bonds are shown as well. The dotted red line in the HPLC trace shows the acetonitrile gradient and the green box highlights the peak collected for further NMR analysis.

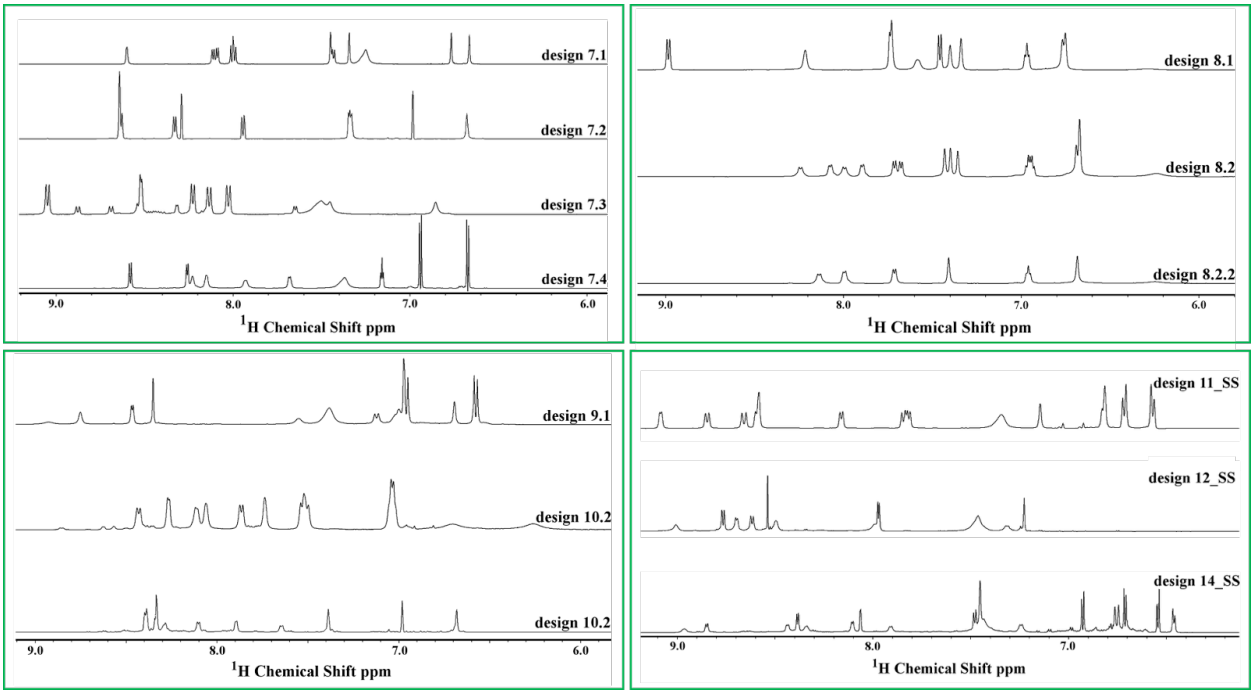
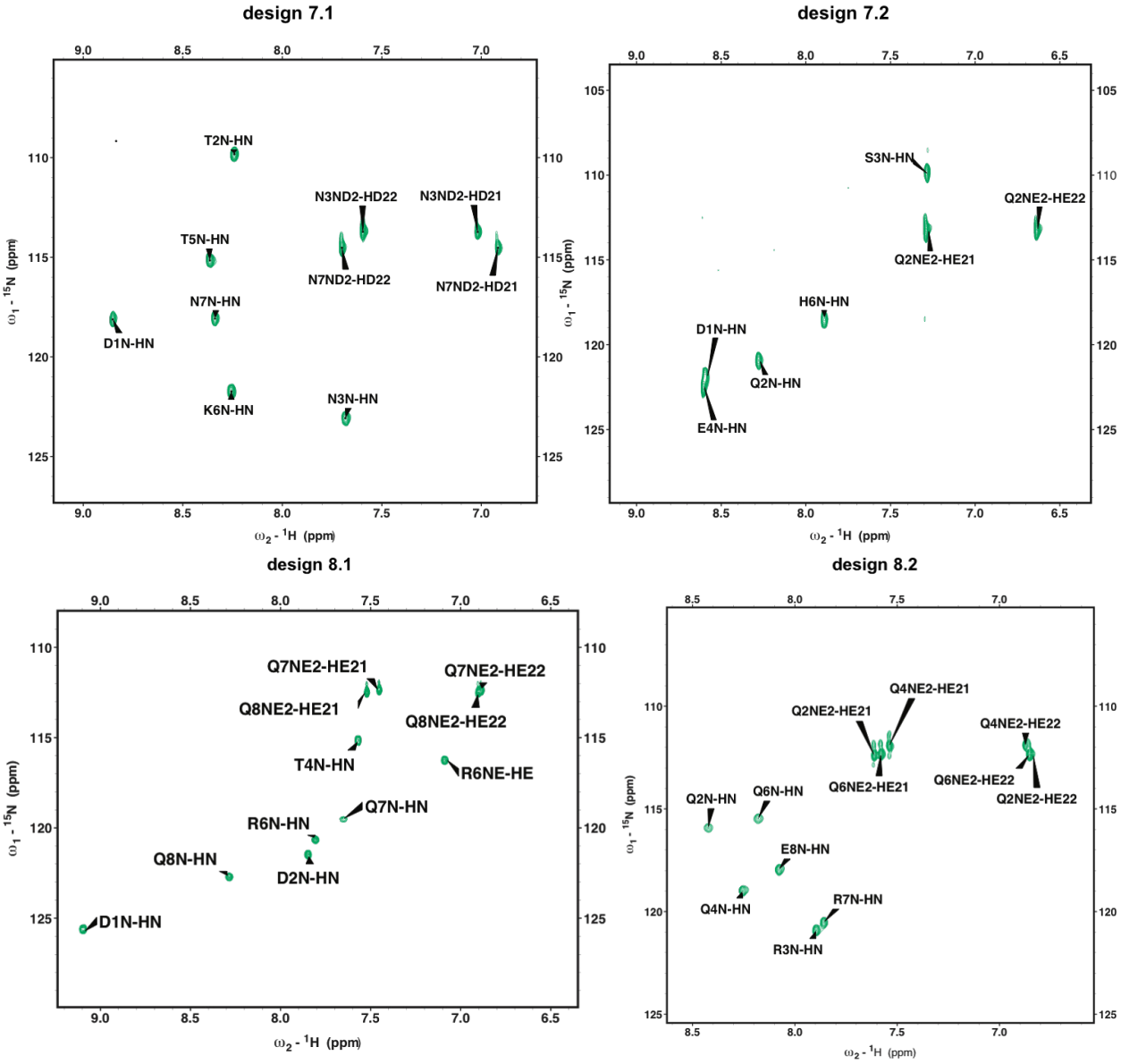


fig. S9. 1D NMR spectra of peptides described in this study with NMR structures within 1.5 Å of design.


design 7.1
T2N-HN
N3ND2-HD22
N3ND2-HD21
T5N-HN
N7ND2-HD22
N7ND2-HD21
N7N-HN
D1N-HN
K6N-HN
N3N-HN
design 7.2
S3N-HN
Q2NE2-HE22
Q2NE2-HE21
H6N-HN
D1N-HN
Q2N-HN
E4N-HN
design 8.1
Q7NE2-HE21
Q7NE2-HE22
Q8NE2-HE21
Q8NE2-HE22
T4N-HN
R6NE-HE
Q7N-HN
R6N-HN
Q8N-HN
D2N-HN
D1N-HN
design 8.2
Q4NE2-HE21
Q2NE2-HE21
Q4NE2-HE22
Q6NE2-HE21
Q6NE2-HE22
Q2NE2-HE22
Q2N-HN
Q6N-HN
E8N-HN
Q4N-HN
R7N-HN
R3N-HN
ω1 - 15N (ppm)
ω2 - 1H (ppm)

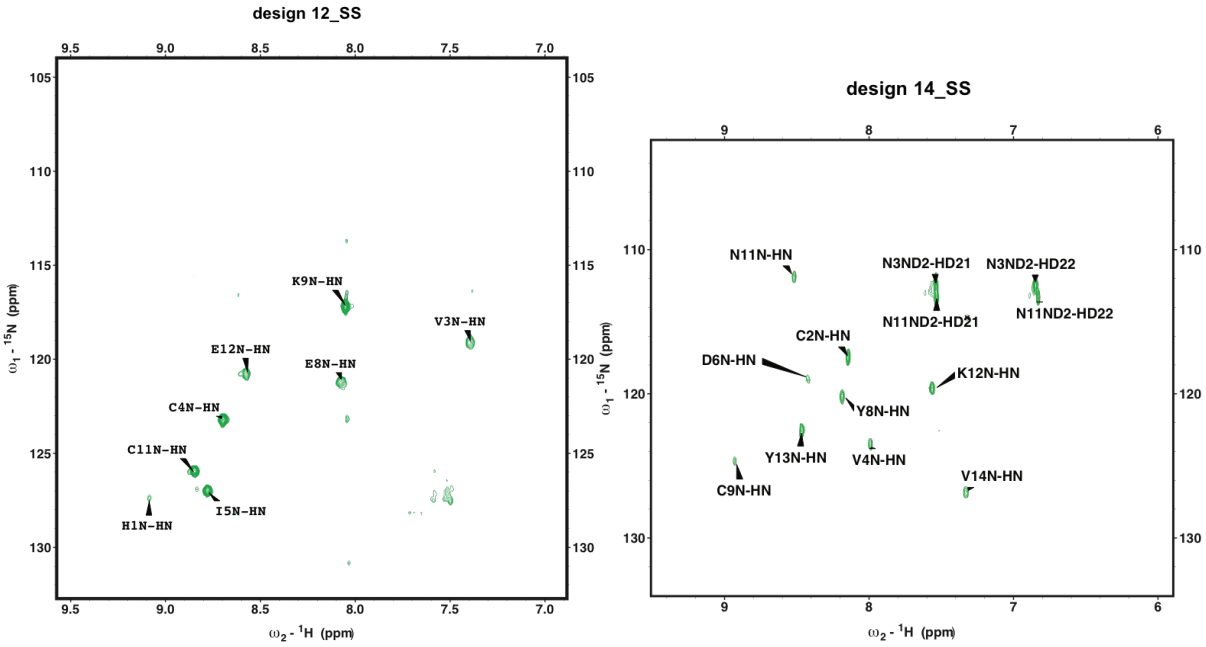
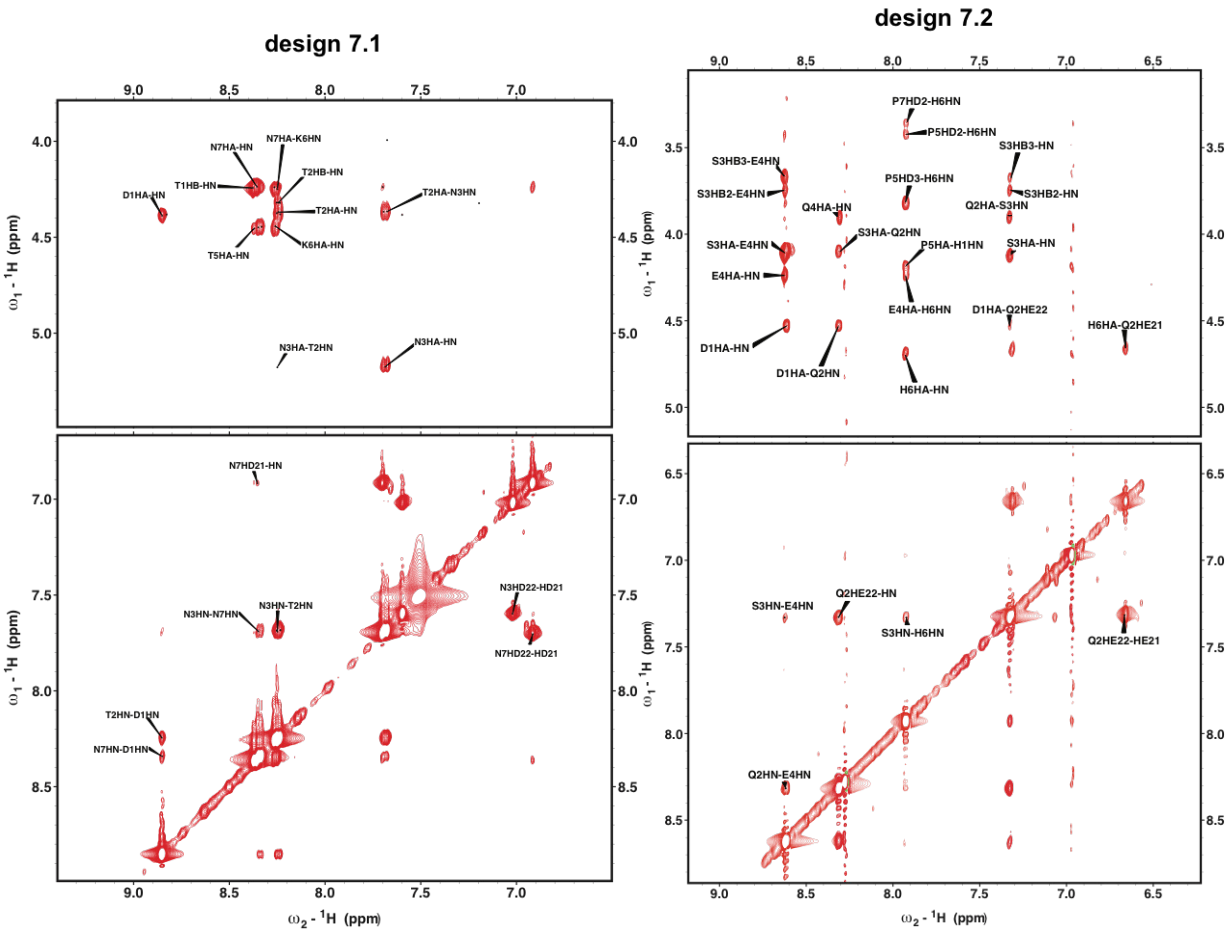
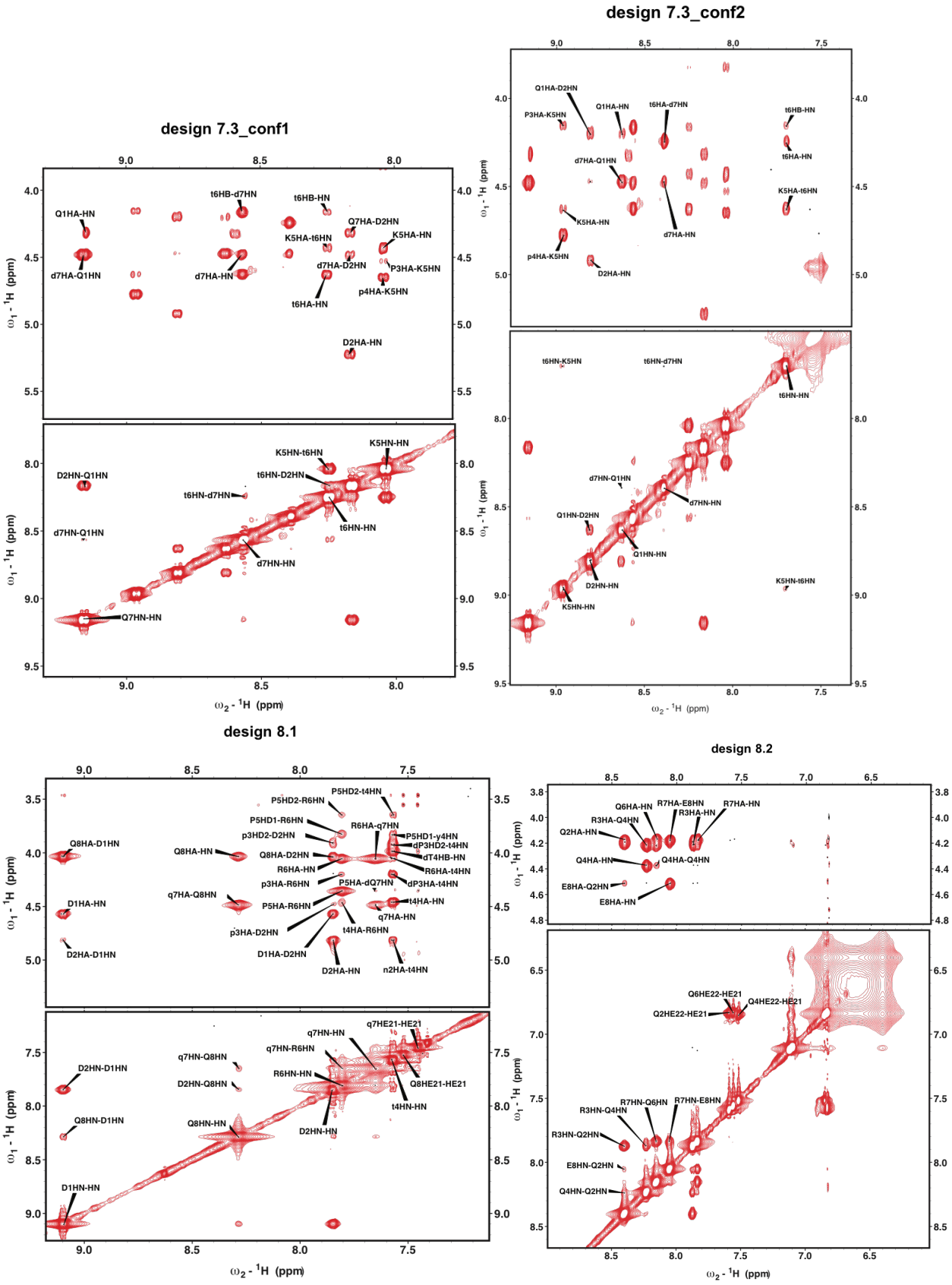
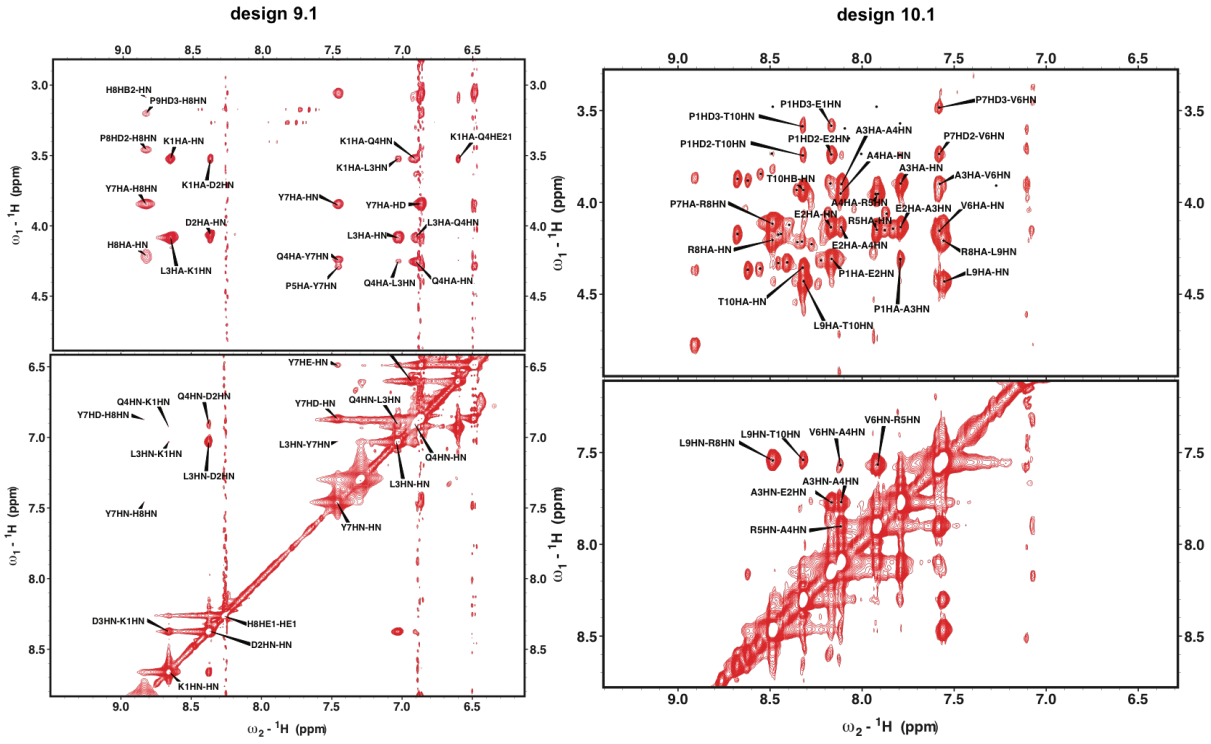
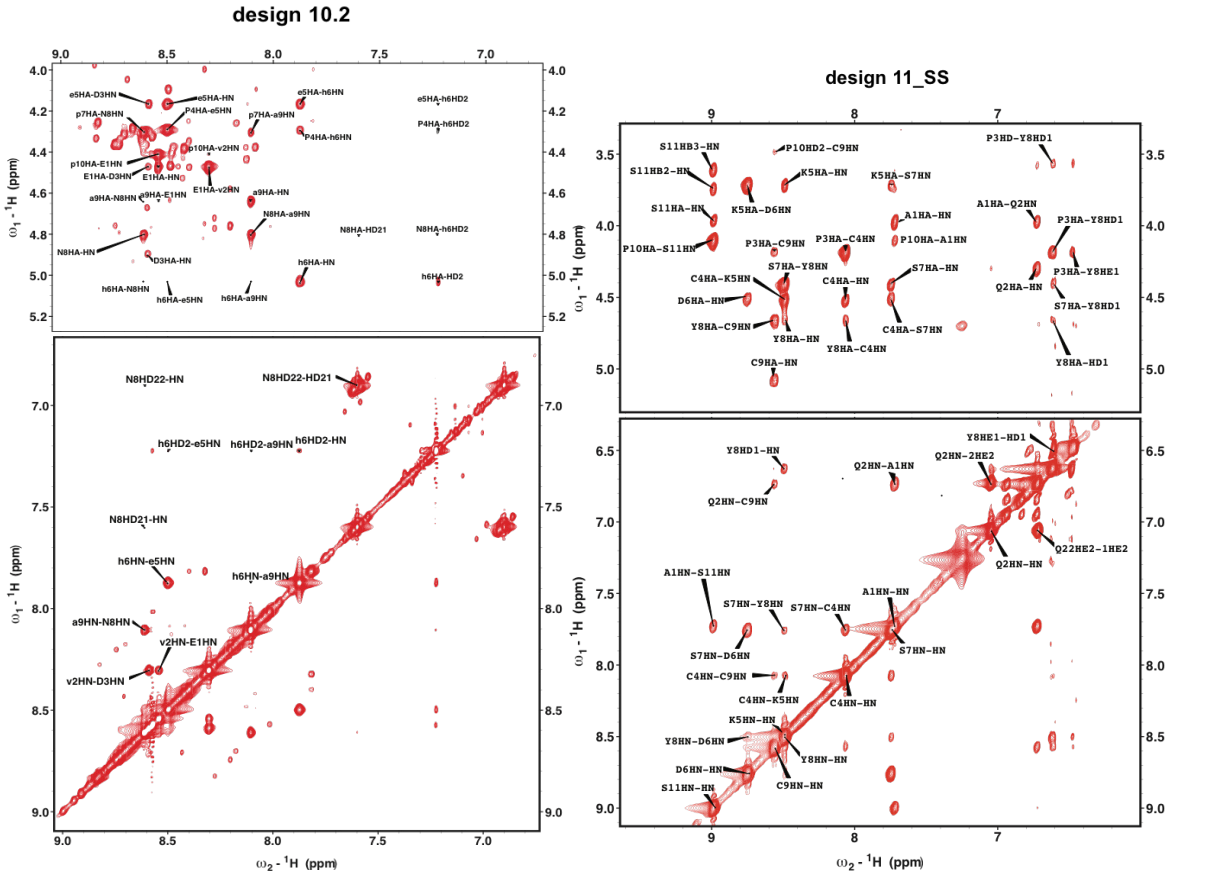


fig. S10. Representative HMQC spectra for several peptides in this study.


design 7.1
design 7.2
N7HA-HN
N7HA-K6HN
T1HB-HN
T2HB-HN
D1HA-HN
T2HA-N3HN
T2HA-HN
K6HA-HN
T5HA-HN
N3HA-T2HN
N3HA-HN
P7HD2-H6HN
P5HD2-H6HN
S3HB3-HN
S3HB3-E4HN
P5HD3-H6HN
S3HB2-E4HN
S3HB2-HN
Q4HA-HN
Q2HA-S3HN
S3HA-Q2HN
S3HA-E4HN
P5HA-H1HN
S3HA-HN
E4HA-HN
E4HA-H6HN
D1HA-Q2HE22
H6HA-Q2HE21
D1HA-HN
D1HA-Q2HN
H6HA-HN
N7HD21-HN
N3HD22-HD21
N3HN-N7HN
N3HN-T2HN
N7HD22-HD21
T2HN-D1HN
N7HN-D1HN
Q2HE22-HN
S3HN-E4HN
S3HN-H6HN
Q2HE22-HE21
Q2HN-E4HN
ω1 - 1H (ppm)
ω2 - 1H (ppm)


design 7.3_conf1
design 7.3_conf2
design 8.1
design 8.2
ω2 - 1H (ppm)
ω1 - 1H (ppm)


design 9.1
design 10.1
ω_1 - ^{1}H (ppm)
ω_2 - ^{1}H (ppm)


design 10.2
design 11_SS
ω_1 - ^{1}H (ppm)
ω_2 - ^{1}H (ppm)

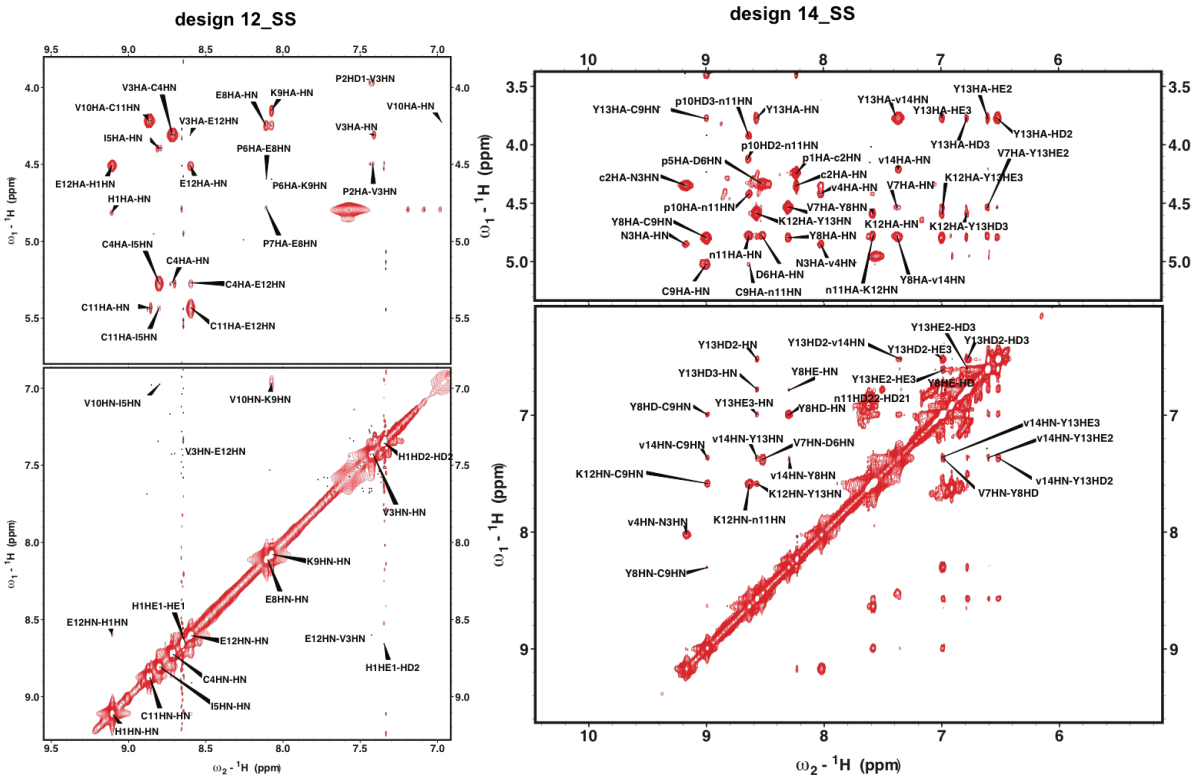


fig. S11. NOESY NMR spectra of peptides mentioned in this study with NMR structures within 1.5 Å of design.

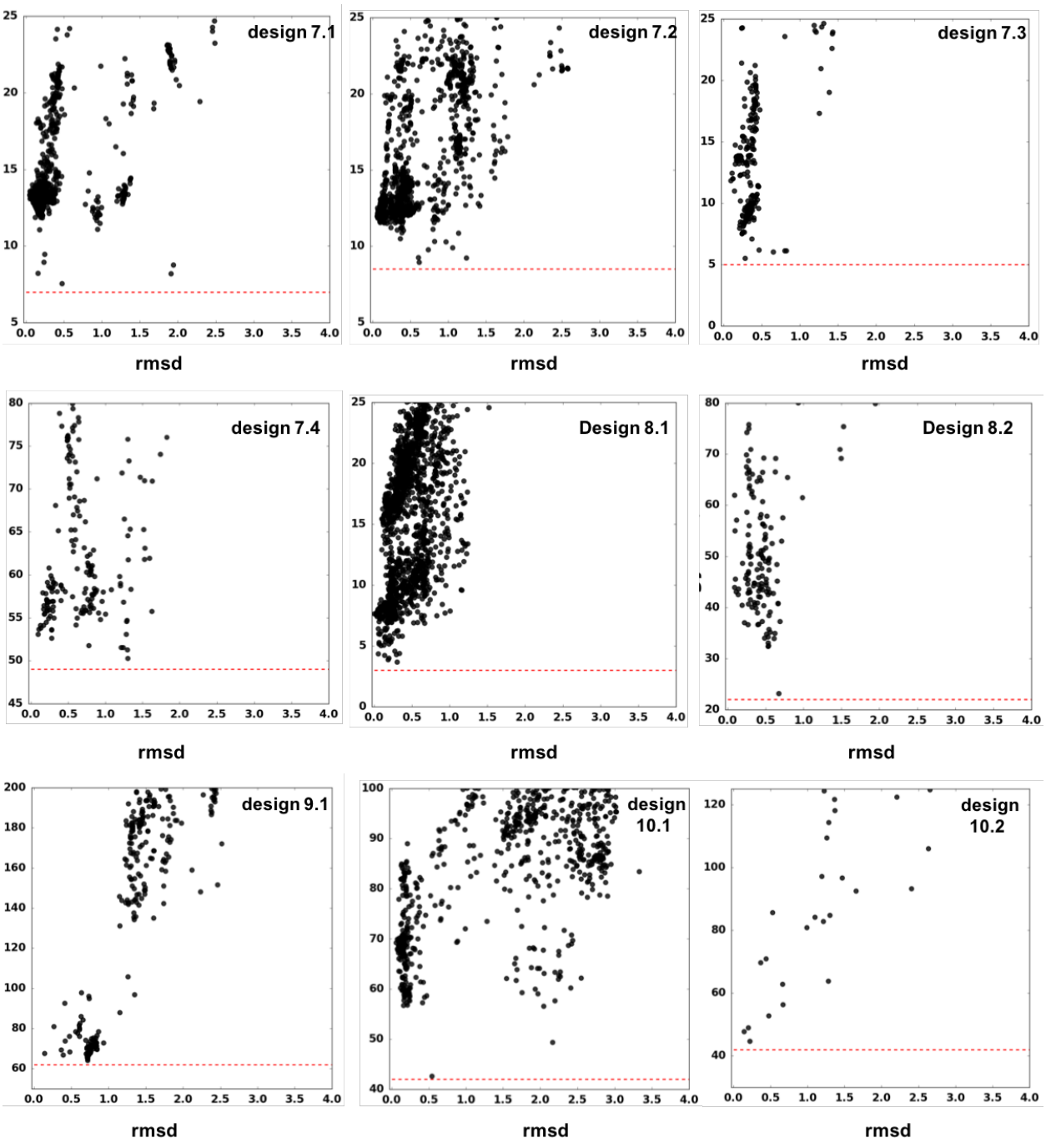
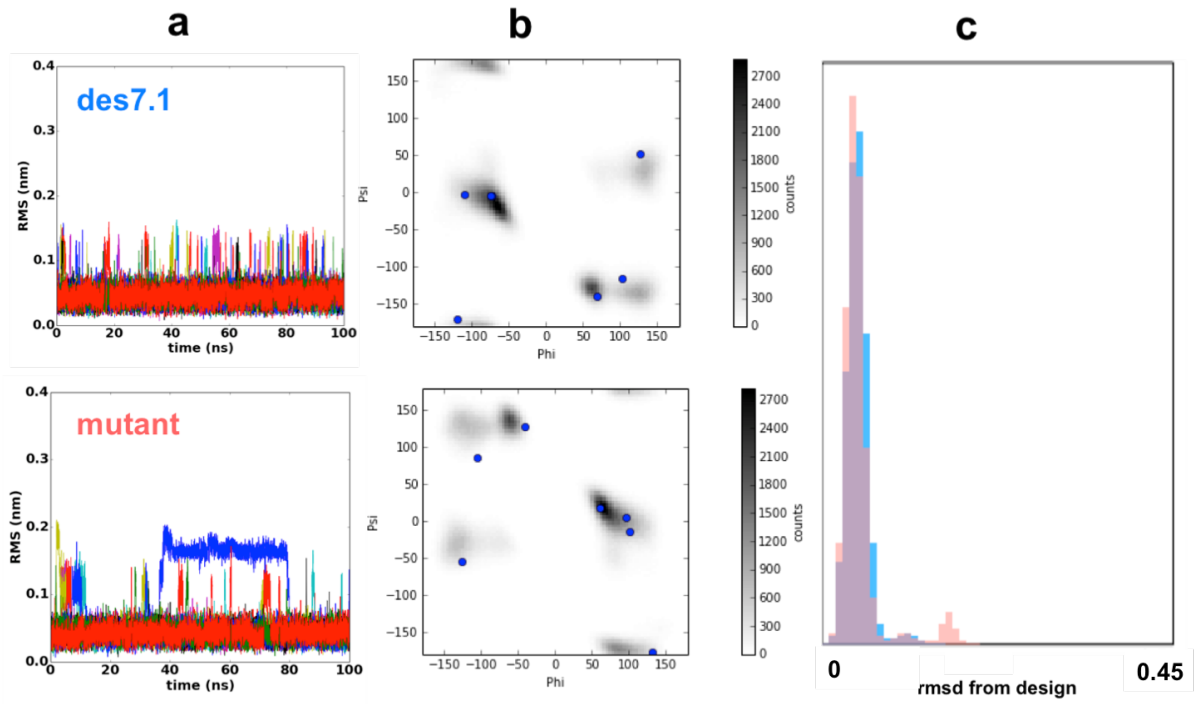
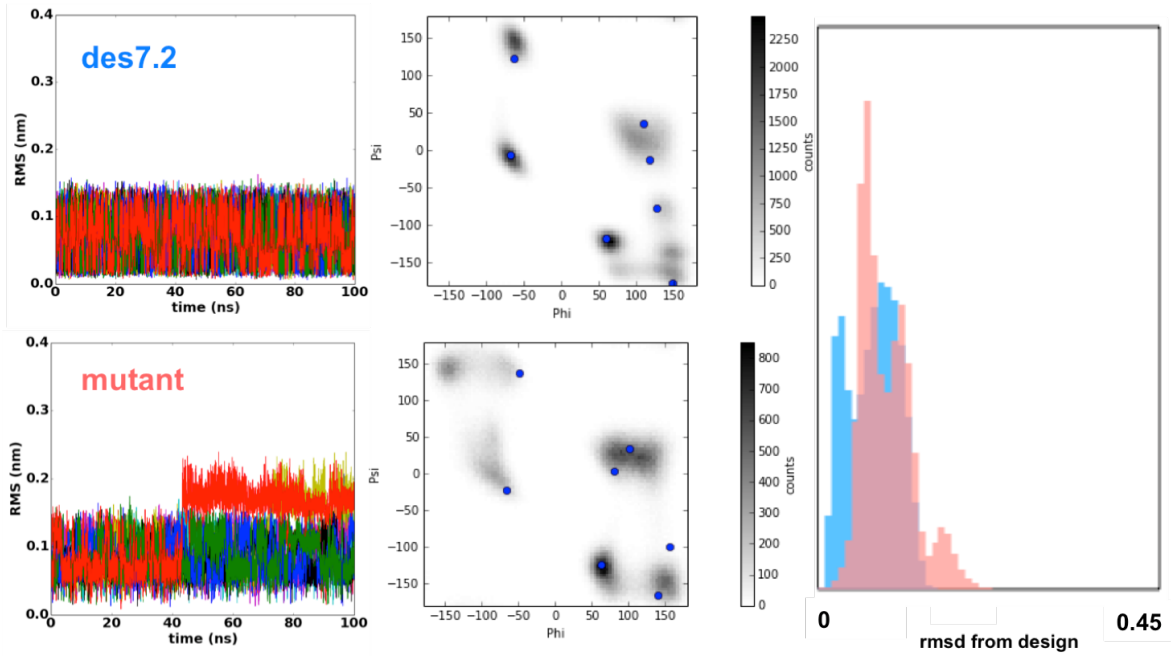
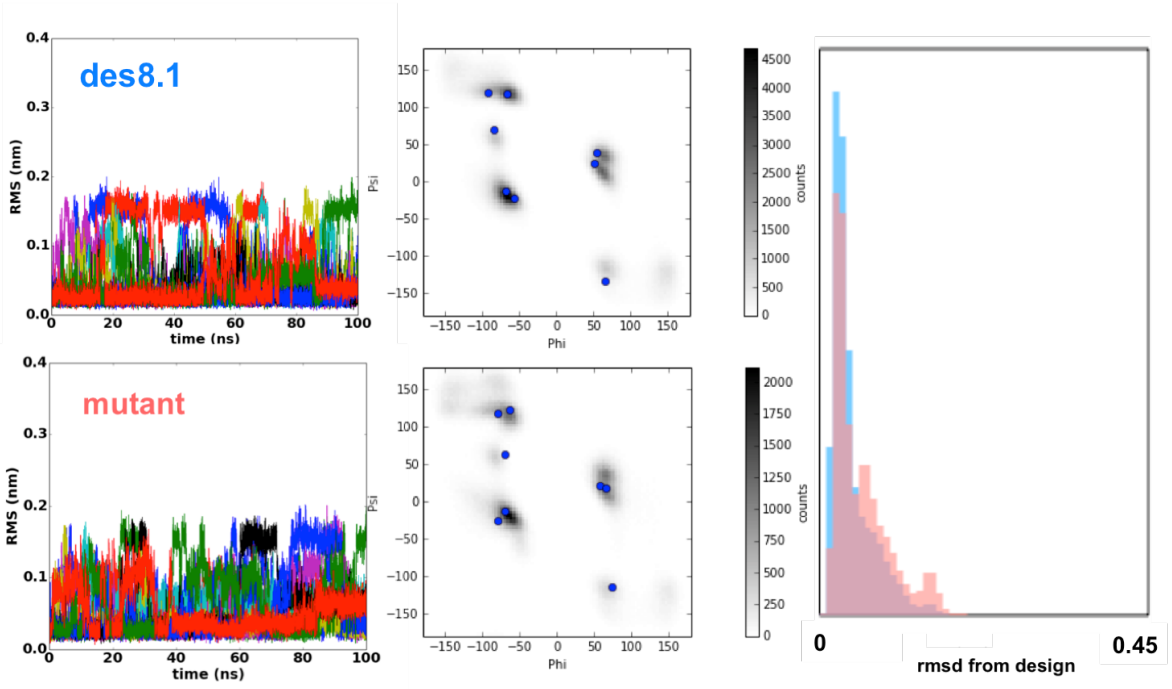
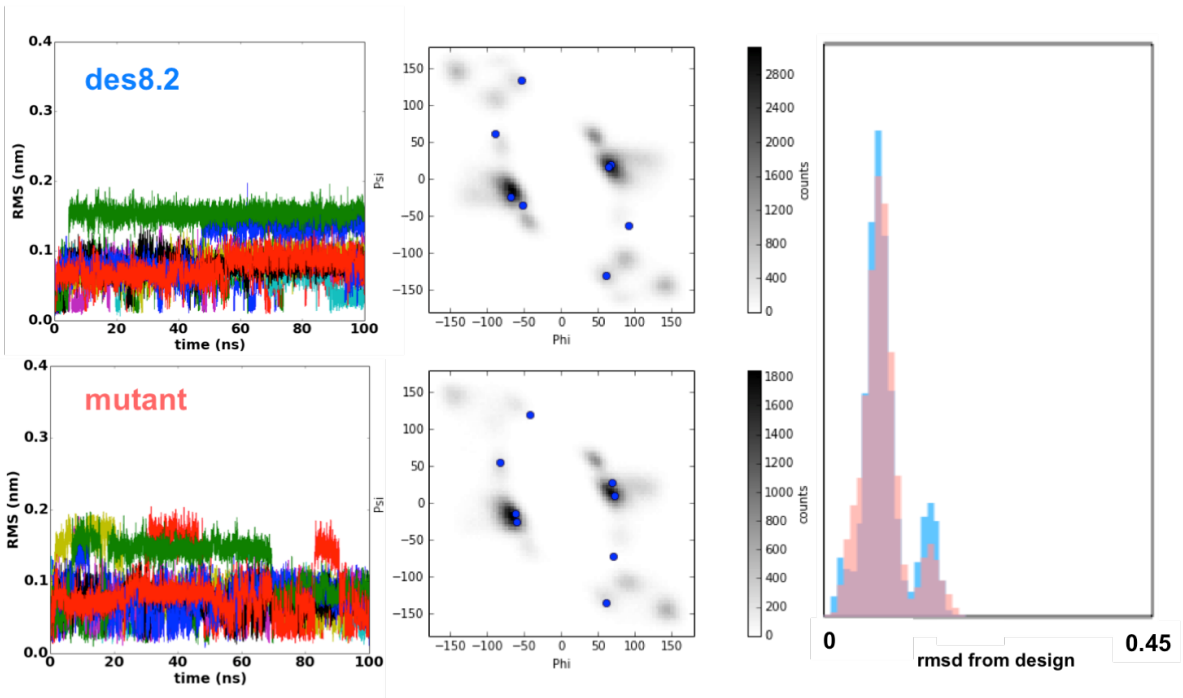


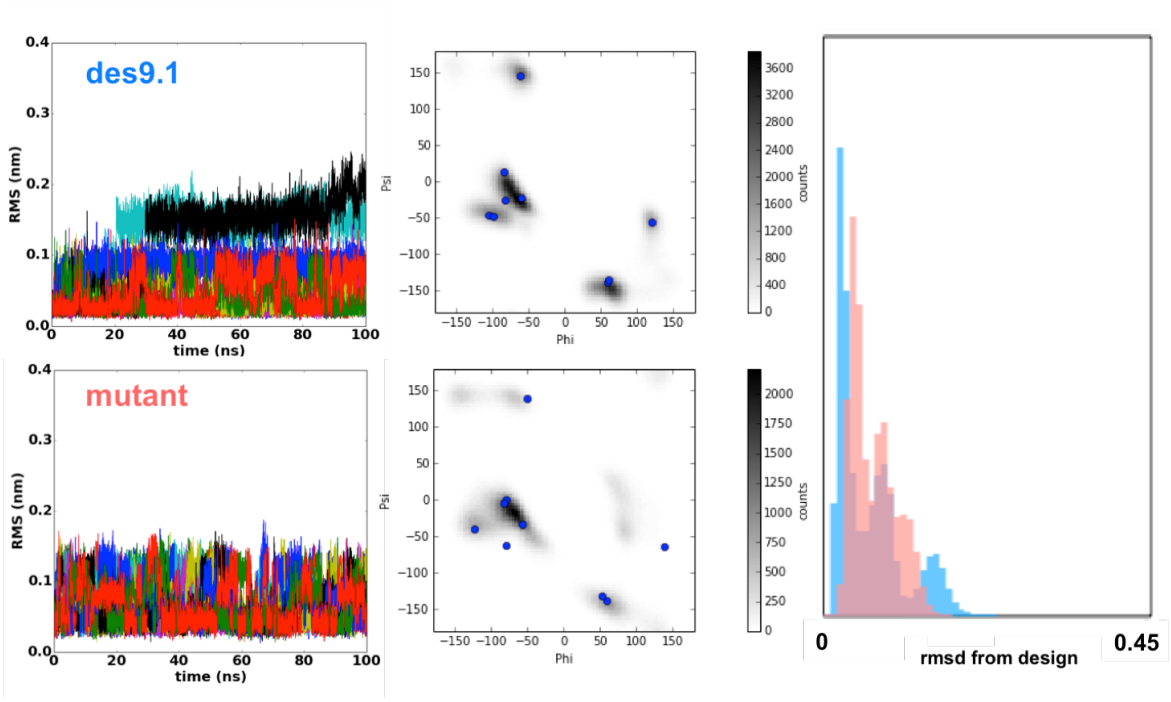
fig. S12. The NOE constraints identify structures close to the design models in pre-generated ensembles. For each design, an ensemble of possible cyclic structures pre-generated for energy landscape analysis was rescored based on how well they fit the NOE constraints (lower scores are better fits) with a penalty for interatomic clashes (`omega=1`, `fa_rep=10`, `fa_intra_rep=0.5`, `pro_close=5`, `rama_prepro=3`). For most of the designs, this simple filtering of the pre-generated ensembles yielded structures within 1Å of the designed models, providing an independent validation of the NMR structure calculations.

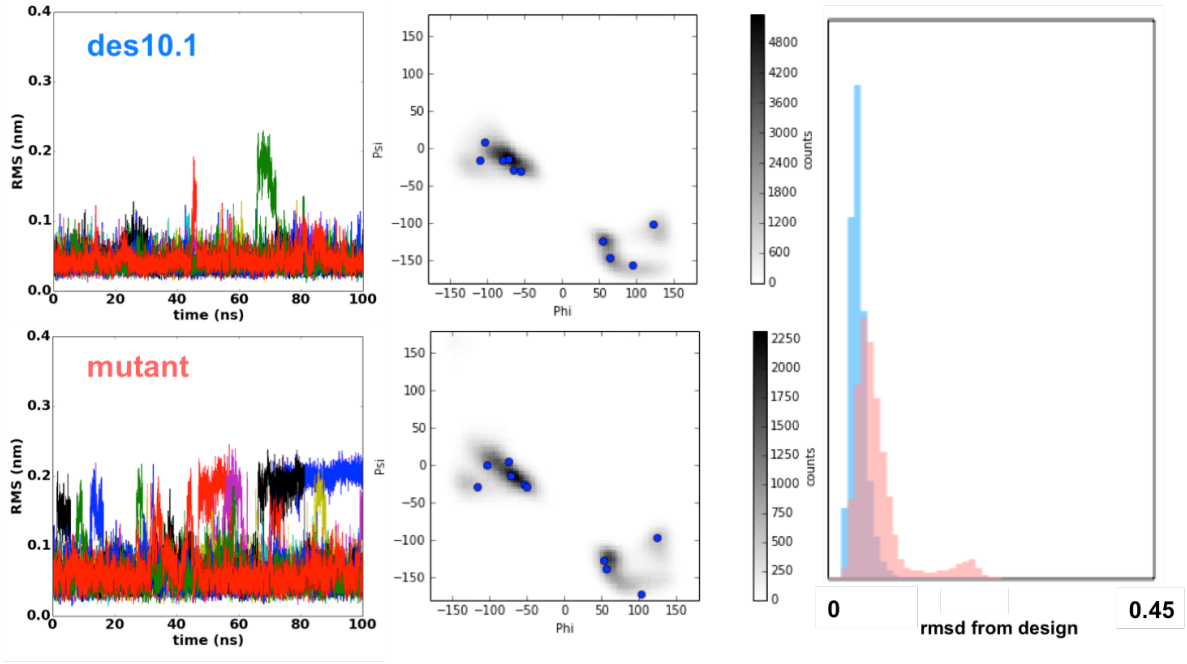

a
b
c
des7.1
mutant
RMS (nm)
time (ns)
Psi
Phi
counts
0
0.45
rmsd from design


des7.2
mutant
RMS (nm)
time (ns)
Psi
Phi
counts
0
0.45
rmsd from design


des8.1
RMS (nm)
time (ns)
Psi
Phi
counts
mutant
RMS (nm)
time (ns)
Psi
Phi
counts
0
0.45
rmsd from design


des8.2
RMS (nm)
time (ns)
Psi
Phi
counts
mutant
RMS (nm)
time (ns)
Psi
Phi
counts
0
0.45
rmsd from design


des9.1
mutant
RMS (nm)
time (ns)
Psi
Phi
counts
0
0.45
rmsd from design


des10.1
mutant
RMS (nm)
time (ns)
Psi
Phi
counts
0
0.45
rmsd from design

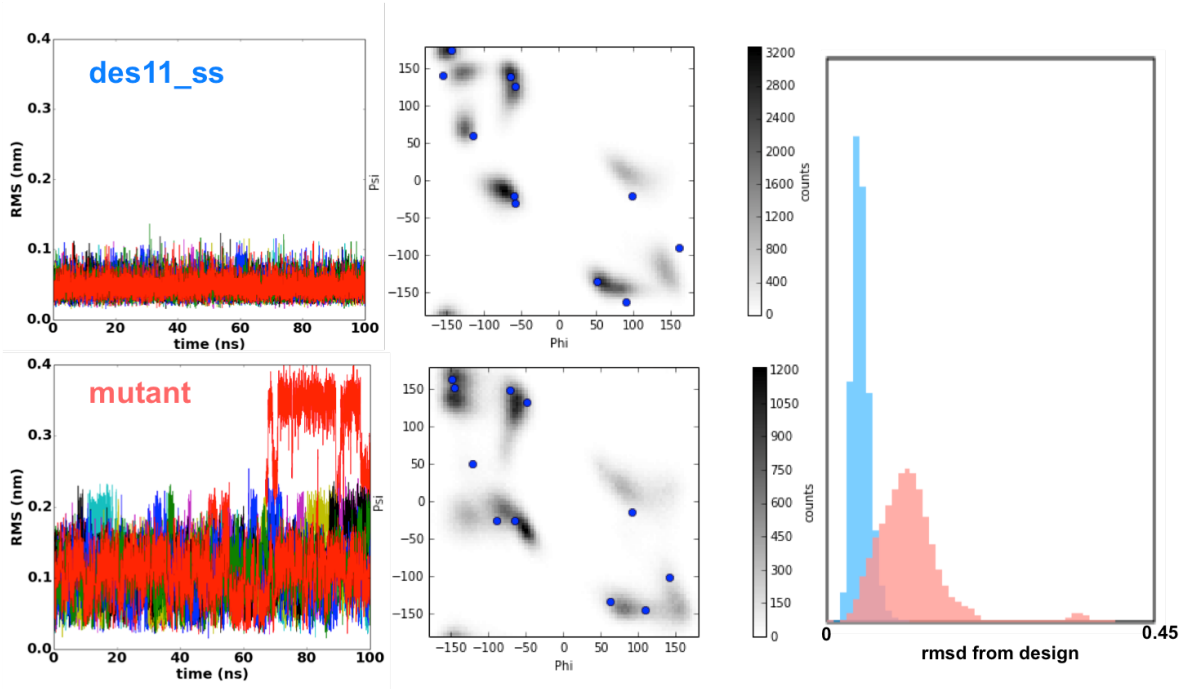


fig. S13. (a) RMSD to the starting conformation of MD trajectories (10x100 ns) for designed macrocycles with NMR structures within 1.5 Å of design as well as their mutants. (b) The torsional space that the macrocycles and their mutants explore during simulation. (c) Histograms compare the frequency of observing structures from the entire trajectory versus their RMSD to the initial structure for initial design (cyan) and mutants (pink). The designs spend the majority of their time in the designed conformation. Analysis of MD trajectories for mutants of these macrocycles shows more flexibility in most cases, some with prominent changes (design 11_SS). Design 10.2 showed a completely open structure in unrestrained MD that violated experimental NOEs, thus not shown.

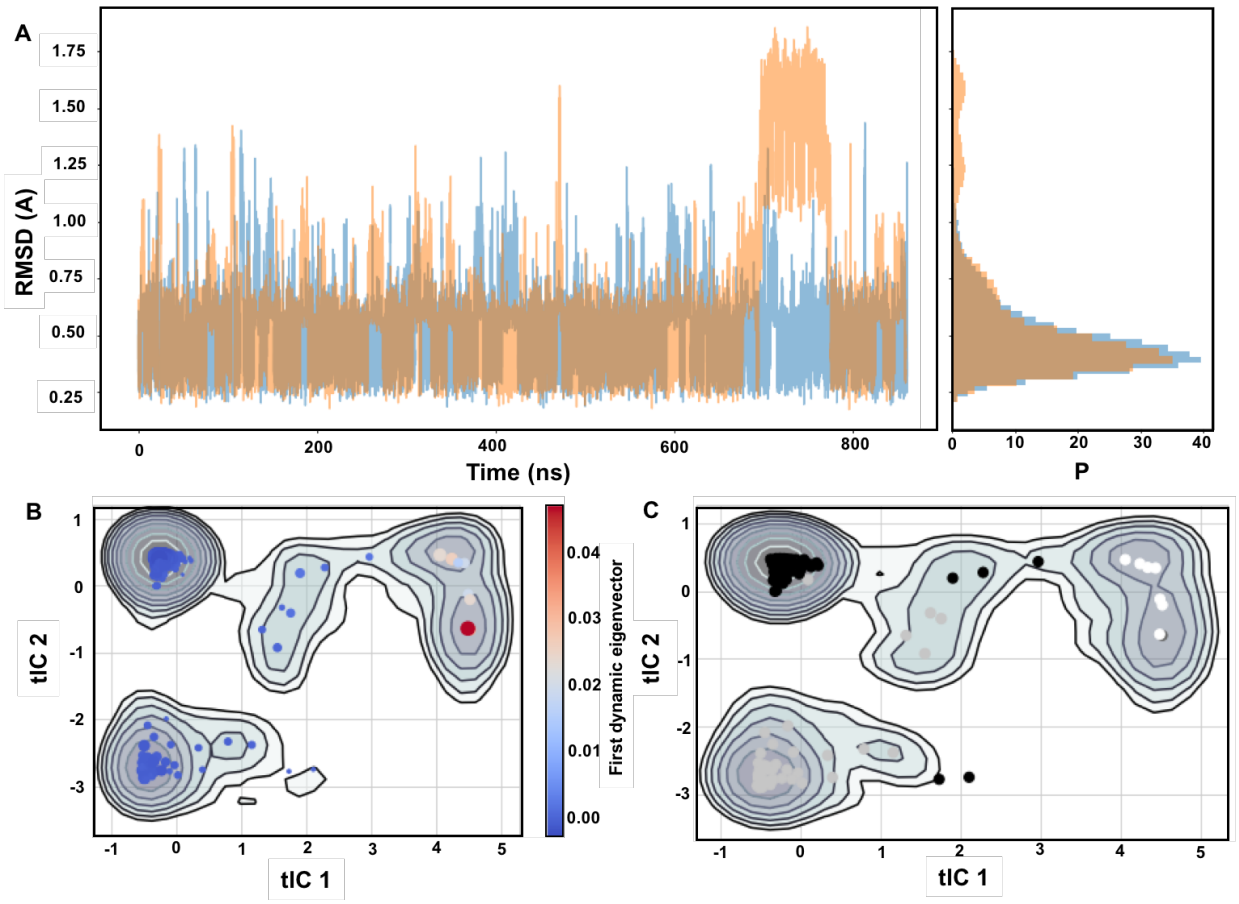


fig. S14. Long time simulation of design 7.3. Plot A shows the RMSD of the backbone over two ~1 microsecond trajectories. MSM was generated by projecting the backbone conformations (phi,psi) using the tICA methods. figure B shows the projection of the 2 main tICs, and the centers of the clustered conformations (200 microstates), the size of the dots is proportional to its populations, and a red color means that transitions to it are slow if coming from a blue-colored dot. The 200-microstates are clustered in 3 macrostates shown in part C. These three models are later used to extract the retention time of the macrocycle at the designed structure.

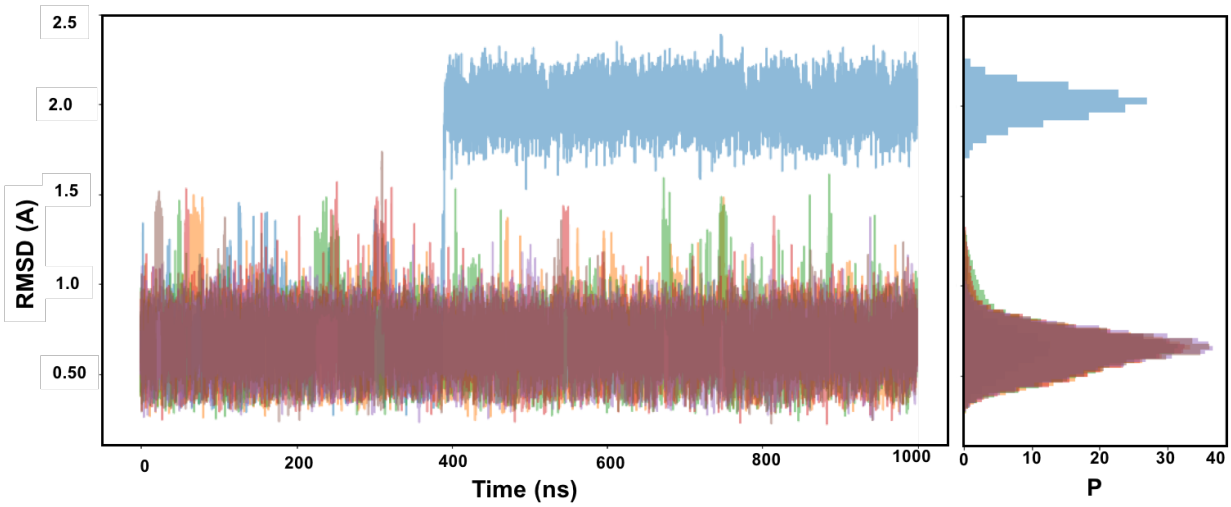
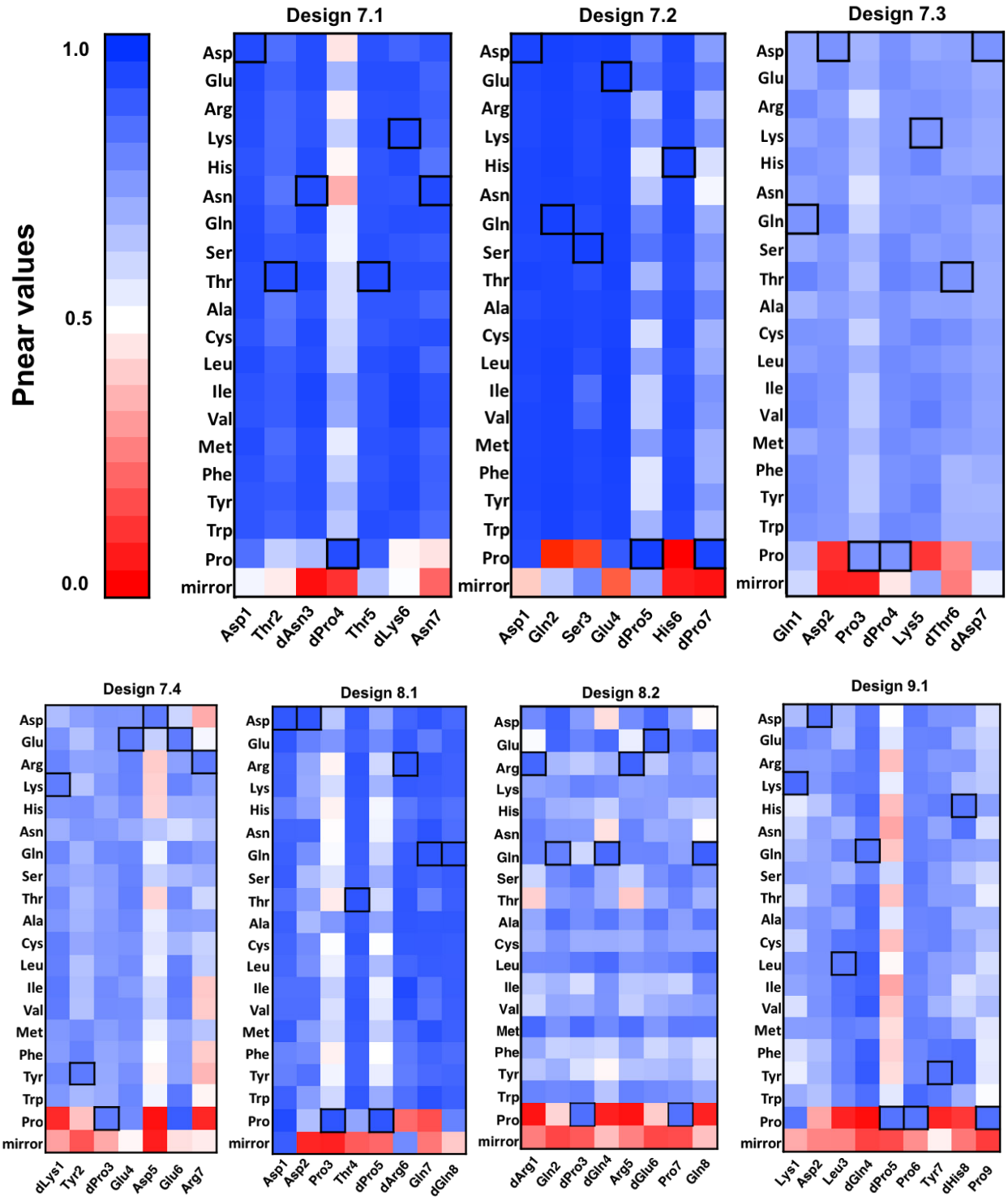


fig. S15. Trajectory of 6x1 microsecond MD simulation runs on design 7.4 shows that the structure is spending most of its time within 1 Å of the designed structure but there is another kinetically stable conformation that shows up after 400 ns in one of the trajectories and persists throughout the simulation.


Pnear values
1.0
0.5
0.0
Design 7.1
Design 7.2
Design 7.3
Design 7.4
Design 8.1
Design 8.2
Design 9.1
Asp
Glu
Arg
Lys
His
Asn
Gln
Ser
Thr
Ala
Cys
Leu
Ile
Val
Met
Phe
Tyr
Trp
Pro
mirror
Asp1 Thr2 dAsn3 dPro4 Thr5 dLys6 Asn7
Asp1 Gln2 Ser3 Glu4 dPro5 His6 dPro7
Gln1 Asp2 Pro3 dPro4 Lys5 dThr6 dAsp7
dLys1 Tyr2 dPro3 Glu4 Asp5 Glu6 Arg7
Asp1 Asp2 Pro3 Thr4 dPro5 dArg6 Gln7 dGln8
dArg1 Gln2 dPro3 dGln4 Arg5 dGlu6 Pro7 Gln8
Lys1 Asp2 Leu3 dGln4 dPro5 Pro6 Tyr7 dHis8 Pro9

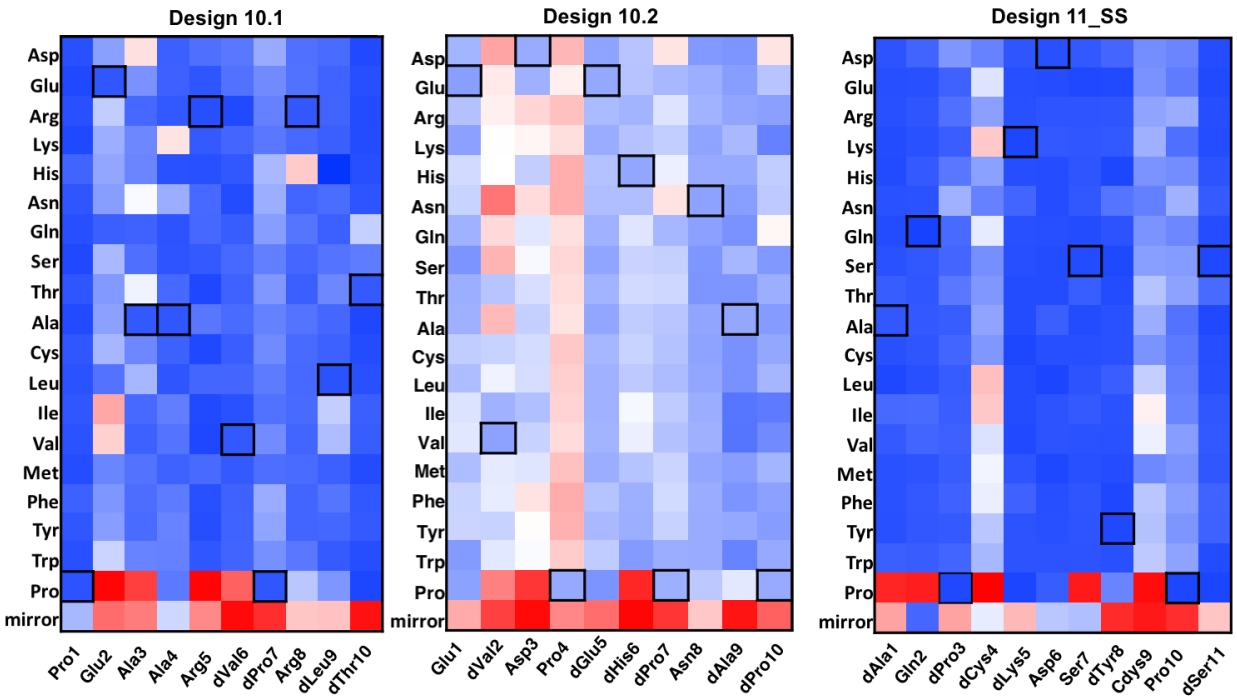


fig. S16. Large-scale landscape analysis of macrocycles with validated NMR structure. Each square box corresponds to a different full energy landscape analysis using the Rosetta@Home cell phone network for the indicated amino acid substitution at the indicated position. Colors indicate the P_{Near} values obtained from these energy landscapes; disruptive substitutions with P_{Near} < 0.4 (sequence predicted not to populate designed structure) are shown in red, and neutral substitutions with P_{Near} > 0.8 are shown in blue. These tables provided a guide to library generation without structure generation: blue substitutions may be freely included, whereas red substitutions should be avoided.

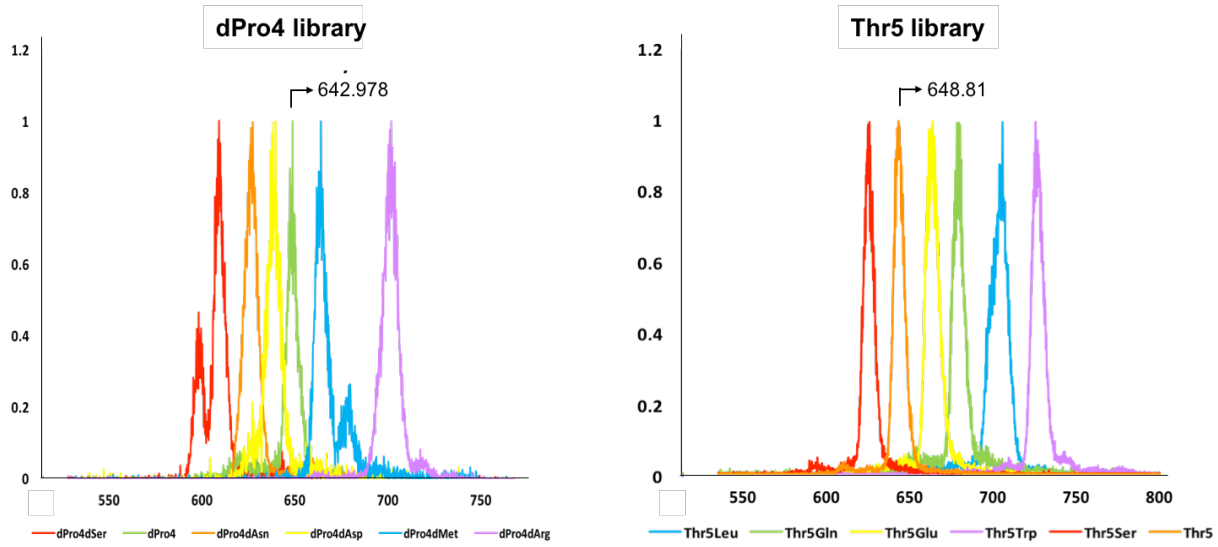


fig. S17. SLIM ion mobility results. Intensity of observed peaks as a function of arrival time (ms) for two crude libraries of design 7.1. dPro4 and Thr5 are the original designs, for which the arrival time is shown. As depicted in the figure, dPro4 library (dPro4 to dSer, dPro, dAsn, dAsp, dMet, dArg) clearly has members with more than one conformation.

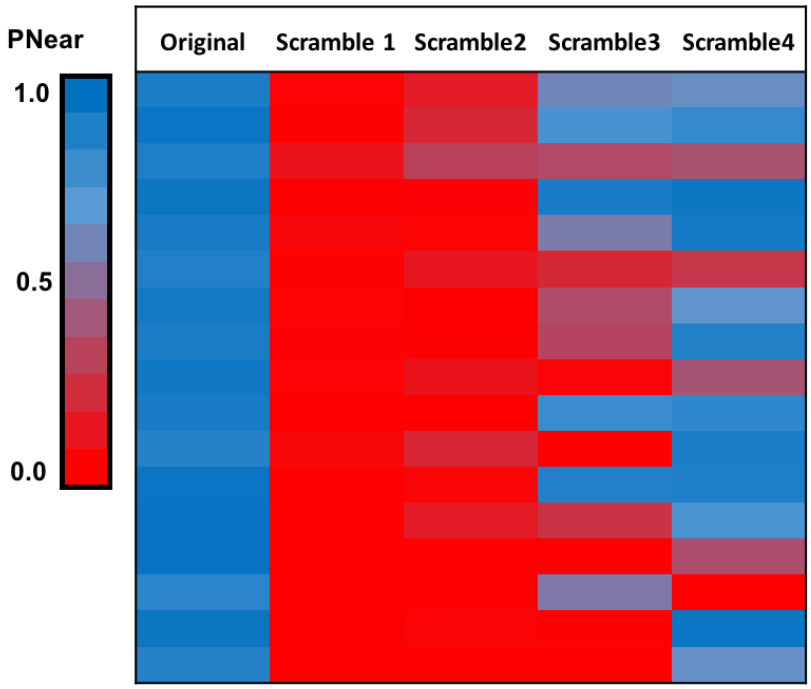


fig. S18. Computed energy gaps (P_{Near}) for scrambled versions of a subset of the designed macrocycles. Scramble 1 indicates random permutations of the original sequence. In scramble 2, the identity of amino acids at each given position are permuted but the chirality at each position retained. Scramble 3 is similar to 2, but proline residues are retained. In scramble 4, prolines are retained along with other residues found to decrease P_{Near} to less than 0.4 in the individual substitution scans (fig. S16). For most but not all of the designed macrocycles, specifying only the individual residues identified in the fig. S16 tables is sufficient to determine the structure.

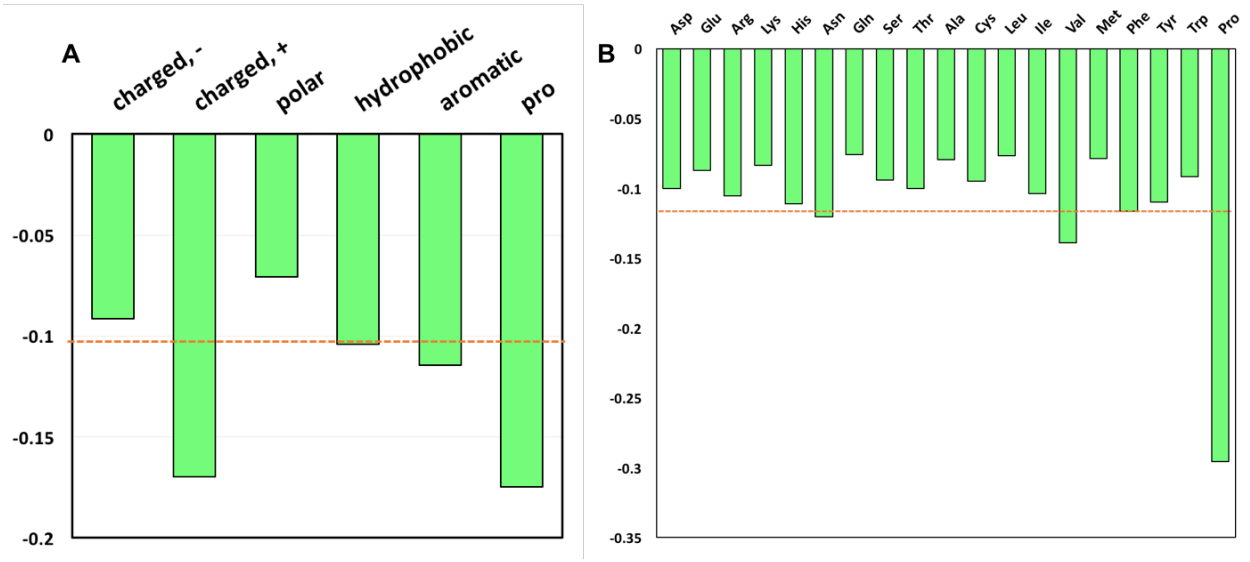


fig. S19. Contribution of amino acid types to energy gaps of designed macrocycles. The Y axis is the average change in P_{Near} upon mutation. (A) shows the average effect of mutating the original amino acids (top, grouped based on class) to all other amino acids. (B) shows the effect of mutating amino acids in different positions to those indicated at top. Mutations to proline produce the largest effects. Orange lines represent the average P_{Near} deviation for all the mutations.

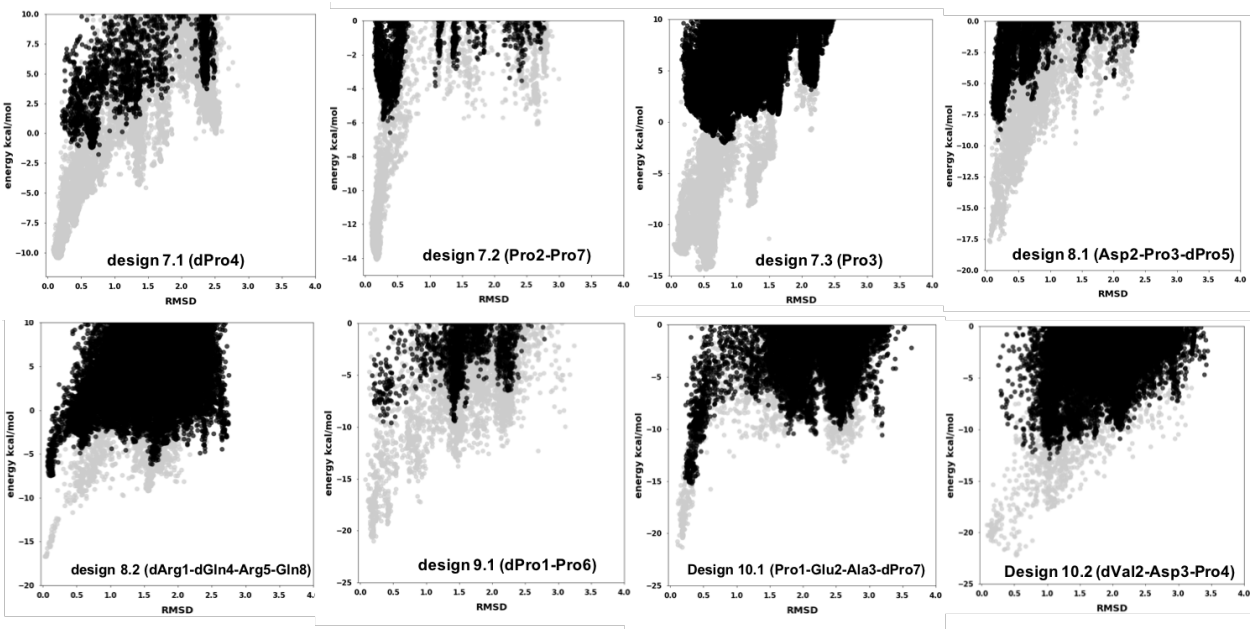


fig. S20. Energy landscape of selected designs with all residues mutated to alanine with the same chirality except those (indicated under the plot for each design), found to be important for folding in the comprehensive individual substitution landscape analyses (Fig. 3-4, and Fig. S16). As shown in the figure, specifying the identity of only few amino acids is in some case sufficient to determine the structure. The grey plot shows the original landscape before mutation.

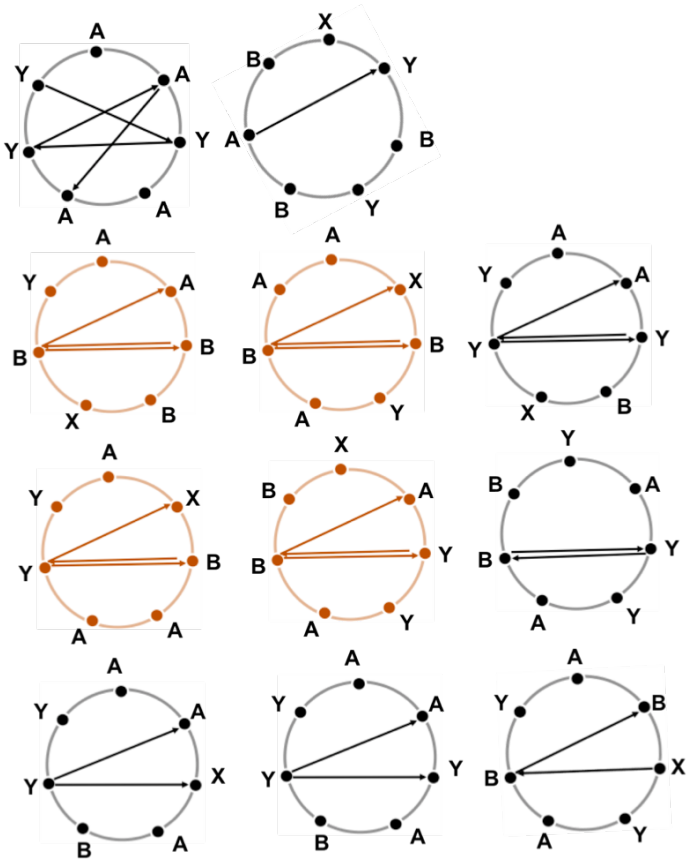


fig. S21. Backbone hydrogen bond patterns of macrocycles with 7 residues which passed our computational folding analysis. The arrow direction is from C=O to NH. Red circles indicate those for which NMR structures were determined. The torsion strings are unique but the overall hydrogen bond patterns quite similar.

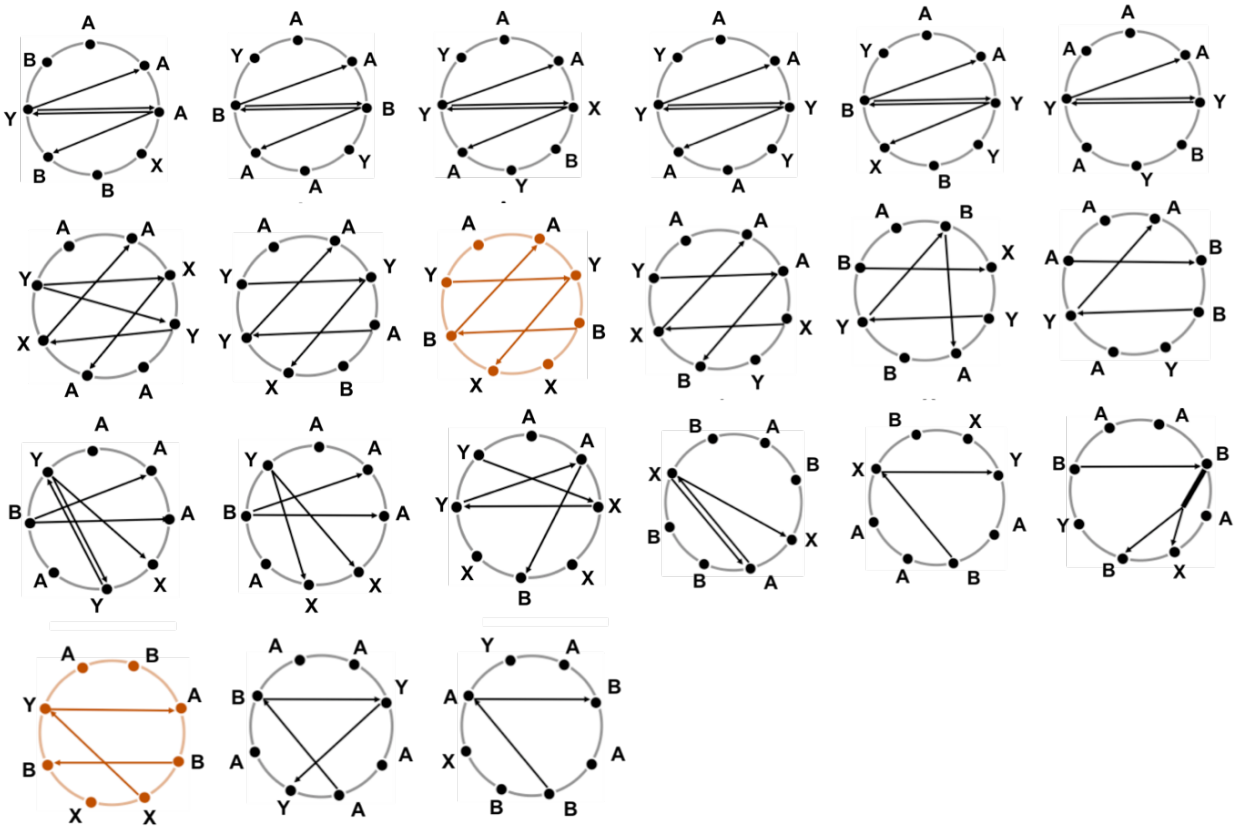


fig. S22. Backbone hydrogen bond patterns of 8 residue designed macrocycles with P_{Near} > 0.8. Representation is as in Fig S21. The thick line indicates a sidechain. There are several distinct hydrogen bond patterns that recur with different torsion bin strings.

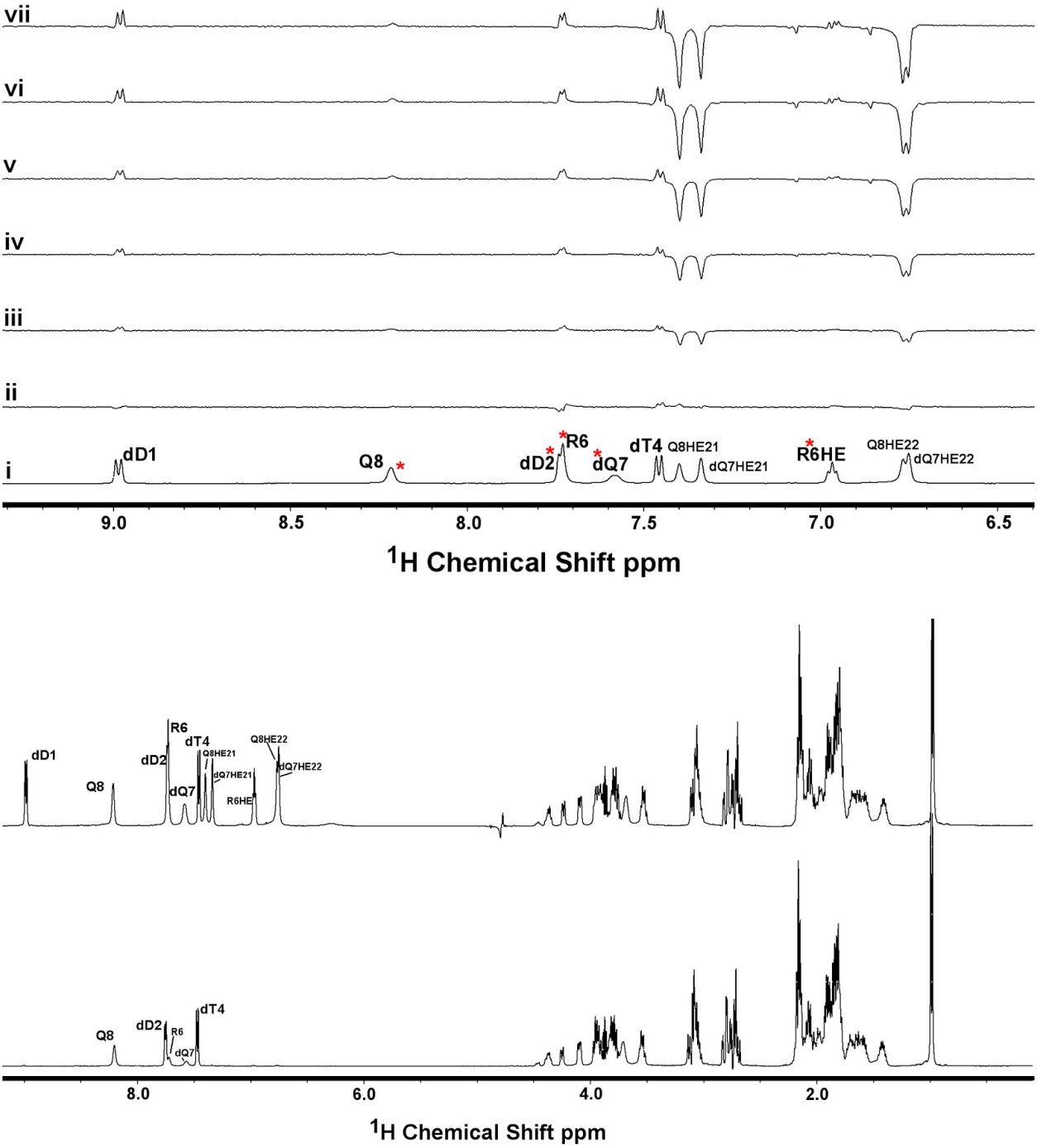


fig. S23. A) 1D 1H CLEANEX-PM experiment on design 8.1 showing slow exchanging protons with reduced signal (red star). Data were collected with mixing times 0-500ms (i-vii) at 25 C. B-top, design 1D 1H spectrum of design 8.1 in 90% H_2O/ 10% D_2O at 25 C, B-bottom, the sample was lyophilized to dryness and exchanged with 99.99% D_2O. Three of the slowly exchanging backbone amide peaks remain visible after fully exchanging with D_2O. These are those involved in backbone-backbone hydrogen bonds.

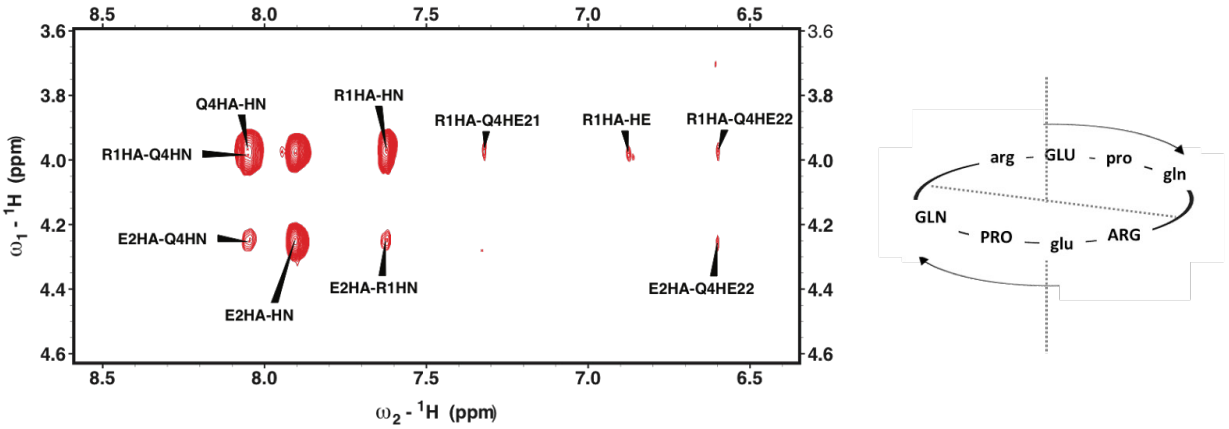


fig. S24. NOESY spectrum of design 8.2.1 showing only 3 (out of 6) backbone NH protons, suggesting symmetry in the peptide (sequence shown in the panel on the right).

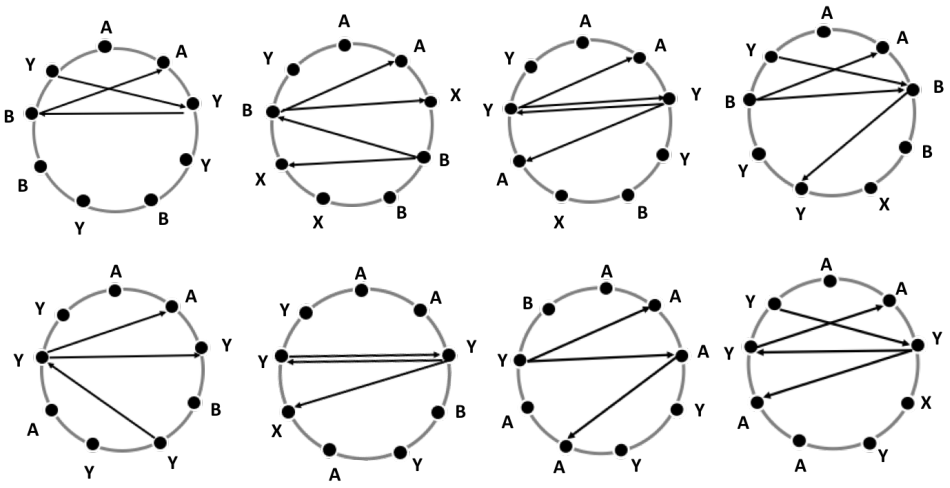


fig. S25. Backbone hydrogen bond patterns of a subset of 9 residue macrocycles that passed our energy landscape analysis shows increased diversity in the longer lengths. The arrow direction is from C=O to NH.

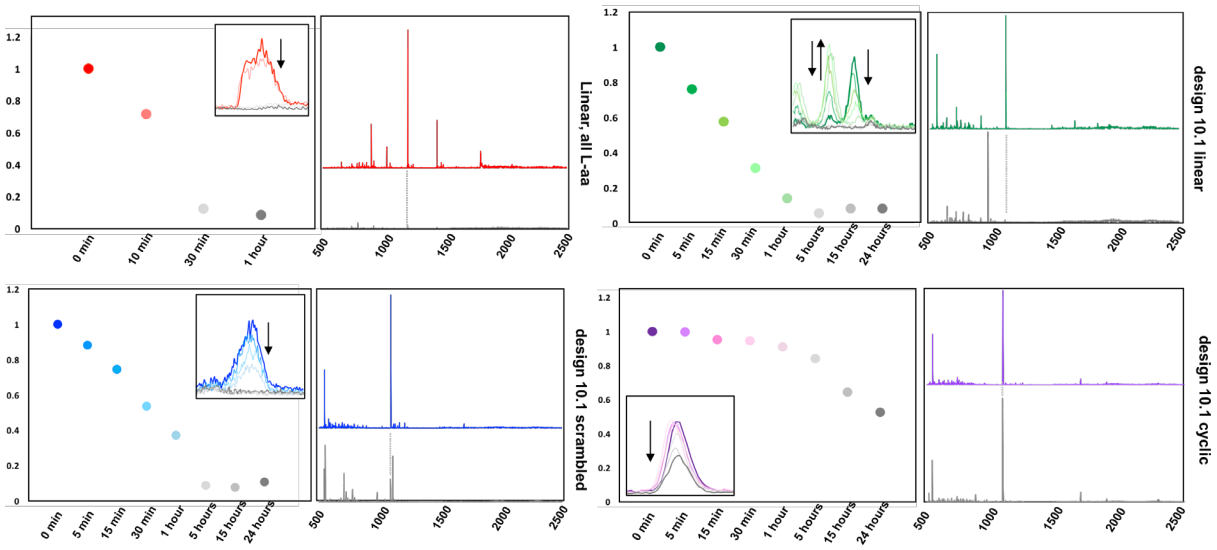
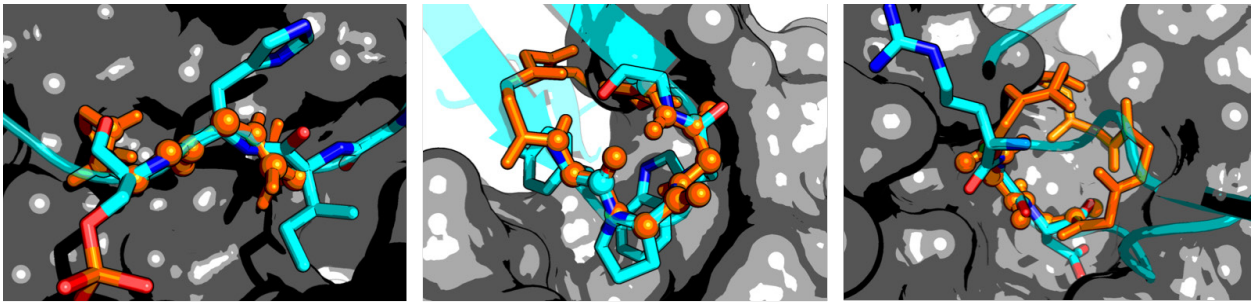


fig. S26. The designed macrocycles are highly protease resistant. To test whether cyclization, mixed chirality, and structured nature of the designed macrocycle adds to its stability, we performed protease cleavage test on one of our designs along with some controls. We used Pronase, a cocktail of different proteases, and the controls were a linear small protein with all L-amino acids, the linear version of the design, and a cyclic version with scrambled sequence (expected to have no rigid structure), and tested the cleavage at different time points using LC/MS. After 1 day, small fractions of the design was cleaved but the majority remained intact whereas the linear control with all L-amino acids was completely degraded after 1 hour. The MS results for initial and last time points are shown as well.



1PP2, ß-catenin **2QNR, septin 2** **4MMX, RGD loop**

fig. S27. The designed macrocycles can mimic and stabilize known protein binding loops. The native loops are shown as color cartoons at the grey interfaces and the peptide is overlayed on them, shown in orange. The spheres show the peptide backbone atoms at the hot spot residues.

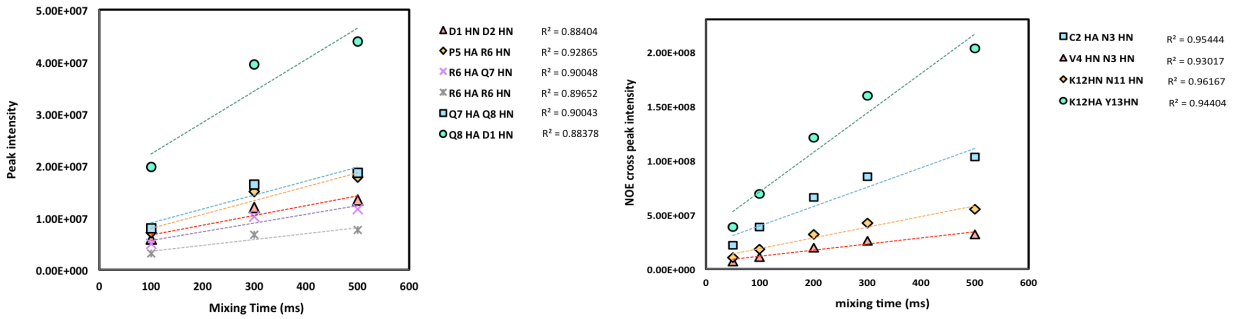


fig. S28. The addition of 5% glycerol to the aqueous peptide NMR samples reduced the overall molecular tumbling rate resulting in improved uniform linear response in peak intensity for NOE build up curves of backbone NOEs in both design 8.1 (left) and 14_SS (right). This suggests the range in NOE mixing times used were appropriate for NOE assignments and structure determination.

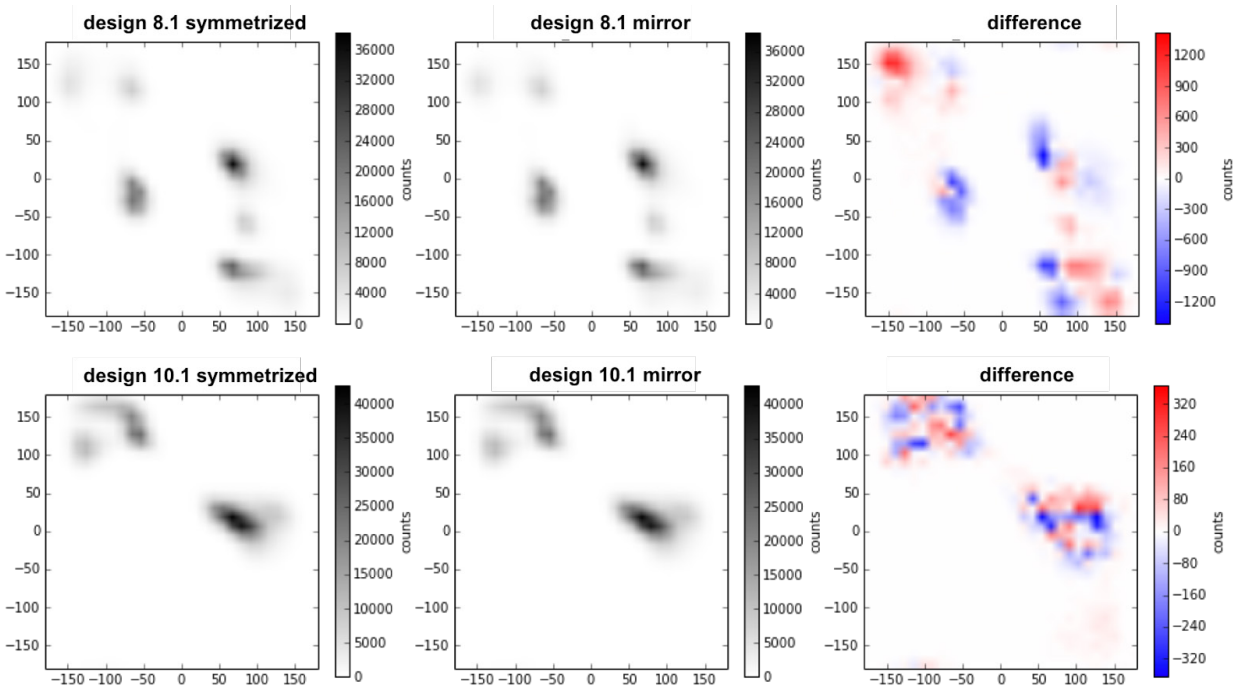


fig. S29. MD simulations of design 8.1 and design 10.1 and their mirrored structures show that the two structures sample similar regions in the Ramachandran plot (when the rama plot of one is symmetrized), suggesting that our MD simulation method is agnostic with respect to amino acid chirality.

Table S1. Frequency of different turn types within PDB and designs and the distribution of sub-turns within each turn.

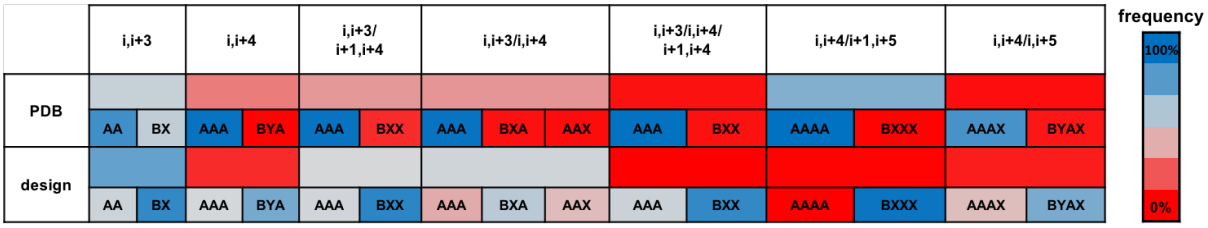


Table S2. NMR statistics for reported structures for size 7 and 8 residues.

	design 7.1	design 7.2	design 7.3_a	design 7.3_b	design 7.4	design 8.1	design 8.1
NMR Constraints							
Total NOEs	72	64	66	47	112	170	88
Intra-residual	42	38	38	27	49	84	46
Sequential (i – j = 1)	20	14	16	13	41	48	23
Short (1 < i – j < 5)	8	6	12	4	17	30	14
Long (i – j ≥ 5)	2	6	0	3	5	8	5
hydrogen bonds[a]	0	1	1	1	1	3	0
Structure Statistics							
Violations							
Distance constraints (Å)	0.024 ± 0.0080	0.11 ± 0.0084	0.13± 0.0083	0.18 ± 0.0099	0.13 ± 0.0047	0.099 ± 0.0046	0.071 ± 0.0049
Max. Distance constraint violation (Å)	0.038	0.12	0.15	0.19	0.14	0.1	0.084
Deviations from idealized geometry							
Bond lengths (Å)	0.0017± 0.00045	0.012 ± 0.00061	0.011± 0.00078	0.010 ± 0.00043	0.010 ± 0.00025	0.012 ± 0.00027	0.0030 ± 0.00029
Bond angles (°)	0.47 ± 0.023	1.26 ± 0.074	1.17 ± 0.026	1.16 ± 0.054	1.04 ± 0.027	1.25 ± 0.043	0.58± 0.031
Impropers (°)	0.31 ± 0.011	0.79 ± 0.24	0.67 ± 0.043	0.54 ± 0.028	0.47 ± 0.033	0.897 ± 0.025	0.35 ± 0.026
Average pairwise r.m.s.d. (Å)							
NMR ensemble NOE only (50 structures)							
Backbone	0.72	0.81	0.47	0.97	0.53	0.075	0.89
Heavy atom	1.46	1.8	1.18	2.53	1.34	0.76	2.01
Average pairwise r.m.s.d. (Å)							
MD Refined ensemble (20 structures)							
Backbone	0.62	0.39	0.29	0.34	-	0.27	0.29
Heavy atom	1.59	0.93	1.16	1.62	-	0.79	1.15
Averager.m.s.d. (Å)							
from design to MD ensemble	1.1	0.8	0.4	0.5	-	0.3	0.9

a) Hydrogen bonds were only enforced if cross strand HN-HN noes were also observed

Table S3. NMR statistics for reported structures for size 9 and 10 residues.

	design 9.1	design 10.1	design 10.2
NMR Constraints			
Total NOEs	91	100	186
Intra-residual	44	37	59
Sequential (i – j = 1)	24	40	86
Short (1 < i – j < 5)	18	10	34
Long (i – j ≥ 5)	5	13	7
hydrogen bonds[a]	1	0	5
Structure Statistics			
Violations			
Distance constraints (Å)	0.069 ± 0.0037	0.048 ± 0.0079	0.062 ± 0.0067
Max. Distance constraint violation (Å)	0.076	0.059	0.078
Deviations from idealized geometry			
Bond lengths (Å)	0.0036 ± 0.00018	0.0026 ± 0.00028	0.0083 ± 0.00031
Bond angles (°)	0.63 ± 0.018	0.55 ± 0.019	1.02 ± 0.021
Impropers (°)	0.38 ± 0.015	0.35 ± 0.35	0.44 ± 0.030
Average pairwise r.m.s.d. (Å)			
NMR ensemble NOE only (50 structures)			
Backbone	0.24	1.06	0.25
Heavy atom	1.27	2.03	0.79
Average pairwise r.m.s.d. (Å)			
MD Refined ensemble (20 structures)			
Backbone	0.44	0.43	0.5
Heavy atom	1.43	1.23	1.14
Averager.m.s.d. (Å)			
from design to MD ensemble	1.2	0.8	0.5

a) Hydrogen bonds were only enforced if cross strand HN-HN noes were also observed

Table S4. NMR Statistics for reported structures for macrocycles containing disulfide crosslinks. Design14_SS displayed high average backbone RMSD in the preliminary structural ensemble as intial NMR structure calculations had difficulty converging. However, restrained MD simulations starting with lowest-energy NMR structure converge very well on the designed topology (Fig. 4).

	11_SS	12_SS	14_SS
NMR Constraints			
Total NOEs	85	103	255
Intra-residual	38	48	121
Sequential (i – j = 1)	25	30	85
Short (1 < i – j < 5)	15	11	27
Long (i – j ≥ 5)	7	14	22
hydrogen bonds[a]	2	0	0
Structure Statistics			
Violations			
Distance constraints (Å)	0.086 ± 0.0033	0.134 ± 0.0088	0.099 ± 0.0085
Max. Distance constraint violation (Å)	0.094	0.16	0.11
Deviations from idealized geometry			
Bond lengths (Å)	0.012 ±0.0005	0.011 ± 0.0005	0.011 ± 0.00063
Bond angles (°)	1.05 ± 0.012	1.14 ± 0.054	1.23± 0.054
Impropers (°)	0.65 ±0.030	0.51 ± 0.051	0.63 ± 0.081
Average pairwise r.m.s.d. (Å)			
NMR ensemble NOE only (50 structures)			
Backbone	0.54	1.64	1.35
Heavy atom	1.61	2.84	2.29
Average pairwise r.m.s.d. (Å)			
MD Refined ensemble (20 structures)			
Backbone	0.45	0.69	0.59
Heavy atom	1.14	1.5	1.41
Averager.m.s.d. (Å)			
from design to MD ensemble	0.4	1.6	1.3

a) Hydrogen bonds were only enforced if cross strand HN-HN noes were also observed

Table S5. Macrocycles evaluated in this study. Designs 8.1 and 10.1 are selected using an older protocol before clustering of large dataset and picking the lowest energy variants.

size	name	synthesis/MS	HPLC	1D NMR	2D NMR
macrocycles					
7mers	design 7.1				
	design 7.2				
	design 7.3		two peaks		two conformations, one matches the design
	design 7.4				did not converge
8mers	design 8.1				
	design 8.2				
9mers	design 9.1				
	design 9.2		split peak on hplc	additional peaks	
	design 9.3.2		not so clean	well dispersed but higher assembly	
10mers	design 10.1				
	design 10.2				additional minor conformation. Major matches the design
	design 10.3				one conformation but did not converge
11mers	design 11.1			minor additional peaks	one conformation but did not converge
	design 11.2			many other peaks	
13mers	design 13.1			additional Trp conformation. Hard to annotate.	
	design 13.2			lots of extra peak. Smeared	
bicycles					
11mers	11_SS				
12mers	12_SS				
	12_SS_D2				unable to solve the structure
	12_SS_D3			additional peaks	
13mers	13_SS_D1				not enough data to solve structure
14mers	14_SS				
	14_SS_D2				structure different from design

Table S6. % time that each structure spends in designed state during MD simulations

name	% time below 1 Å
design 7.1	98
design 7.2	75
design 8.1	79
design 8.2	85
design 9.1	82
design 10.1	99
design 10.2	2
design 11_SS	99

Table S7. Different structural features observed for experimentally verified designs.

name	sequence	number of turns	sc-mediated hbond	critical residues	additional important residues
design 7.1	Asp-Thr-dAsn-dPro-Thr-dLys-Asn A A Y Y A X B	AA (i,i+3) YAX (i,i+3/i,i+4)	Asn7---Thr2(NH) Asn7---Thr5(C=O)	dPro4	
design 7.2	Asp-Gln-Ser-Glu-dPro-His-dPro A A B B X B Y	BX (i,i+3) YAA (i,i+3/i,i+4)		dPro5, dPro7	
design 7.3	Gln-Asp-Pro-dPro-Lys-dThr-dAsp A B B X A Y Y	XA (i,i+3) BXA (i,i+3/i,i+4)		Pro3	dThr6: polar interaction with dPro4
design 7.4	dLys-Tyr-dPro-Glu-Asp-Glu-Arg X A X A B A A	XA (i,i+3) AAX (i,i+3/i,i+4)	Asp5---Arg(NH) Asp5---Arg	Asp5	Arg7: hbond to Asp5
design 8.1	Asp-Asp-Pro-Thr-dPro-dArg-Gln-dGln A B A B X X B Y	YA (i,i+3) BXX (i,i+3/i+1,i+4)	dArg6---dGln8(C=O) Gln7---dPro5(C=O) Asp2---Thr4(NH)	Asp2, Pro3, dPro5	
design 8.2	Pro-Gln-dArg-Gln-dPro-dGln-Arg-dGlu A A Y B X X B Y	YAA(i,i+3/i+1,i+4) BXX (i,i+3/i+1,i+4)	Gln2---Arg5---Gln4 dArg3---dGln6---dGlu8	most	S2 symmetric backbone
design 9.1	Lys-Asp-Leu-dGln-dPro-Pro-Tyr-dHis-Pro A A A Y Y A A Y B	AAA (i,i+3/i,i+4) YAA (i,i+3/i,i+4)	dGln4---Lys1(C=O)	dPro5, Pro9	Pro9 stabilizes a bulge Tyr7 packs against Pro6
design 10.1	Pro-Glu-Ala-Ala-Arg-dVal-dPro-Arg-dLeu-dThr A A A A A Y Y A Y Y	YA (i,i+3) YAA (i,i+3/i+1,i+4) AAA (i,i+3/i,i+4)	Glu2---Arg6(NH) Glu2---Arg6		long range bb to bb hydrogen bond Ala2, dVal6, dLeu9: hydrophobic packing
design 10.2	Glu-dVal-Asp-Pro-dGlu-dHis-dPro-Asn-dAla-dPro A X B B X Y Y A Y Y	YA (i,i+3) BX(i,i+3) YAX (i,i+3/i,i+4)	Asp3---Asn8(NH)		dVal2,dAla8: hydrophobic interaction

Database S1. Comp_Folded_Macrocycles

The provided zip file contains the ~200 macrocycle that passed our computational folding criteria.

Database S2. Hot_Loop_Matches

The provided excel file shows the hot loops from Kritzer and coworker(*24*) that we tested in this study to evaluate the potential of our macrocycles in stabilizing them. All the columns are adapted from (*24*). Last column shows the results of our analysis. Blue color indicates that we found at least one macrocycle whose backbone matches that of the hot loop. Red color indicates failure in finding such macrocycle.

Database S3. Designed_Macrocycles_Structures

The provided zip file contains the original designed version of all the macrocycles in the main text for which NMR structure is provided.

References and Notes

1. E. M. Driggers, S. P. Hale, J. Lee, N. K. Terrett, The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat. Rev. Drug Discov.* **7**, 608–624 (2008). doi:10.1038/nrd2590 Medline

2. P. Thapa, M. J. Espiritu, C. Cabalteja, J.-P. Bingham, The emergence of cyclic peptides: The potential of bioengineered peptide drugs. *Int. J. Pept. Res. Ther.* **20**, 545–551 (2014). doi:10.1007/s10989-014-9421-0

3. K. Fosgerau, T. Hoffmann, Peptide therapeutics: Current status and future directions. *Drug Discov. Today* **20**, 122–128 (2015). doi:10.1016/j.drudis.2014.10.003 Medline

4. D. J. Craik, D. P. Fairlie, S. Liras, D. Price, The future of peptide-based drugs. *Chem. Biol. Drug Des.* **81**, 136–147 (2013). doi:10.1111/cbdd.12055 Medline

5. B. P. Gray, K. C. Brown, Combinatorial peptide libraries: Mining for cell-binding peptides. *Chem. Rev.* **114**, 1020–1081 (2014). doi:10.1021/cr400166n Medline

6. D. Marasco, G. Perretta, M. Sabatella, M. Ruvo, Past and future perspectives of synthetic peptide libraries. *Curr. Protein Pept. Sci.* **9**, 447–467 (2008). doi:10.2174/138920308785915209 Medline

7. R. Liu, X. Li, W. Xiao, K. S. Lam, Tumor-targeting peptides from combinatorial libraries. *Adv. Drug Deliv. Rev.* **110-111**, 13–37 (2017). doi:10.1016/j.addr.2016.05.009 Medline

8. R. Obexer, L. J. Walport, H. Suga, Exploring sequence space: Harnessing chemical and biological diversity towards new peptide leads. *Curr. Opin. Chem. Biol.* **38**, 52–61 (2017). doi:10.1016/j.cbpa.2017.02.020 Medline

9. T. Passioura, H. Suga, A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chem. Commun. (Camb.)* **53**, 1931–1940 (2017). doi:10.1039/C6CC06951G Medline

10. P.-S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014). doi:10.1126/science.1257481 Medline

11. P.-S. Huang, K. Feldmeier, F. Parmeggiani, D. A. F. Velasco, B. Höcker, D. Baker, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016). doi:10.1038/nchembio.1966 Medline

12. S. E. Boyken, Z. Chen, B. Groves, R. A. Langan, G. Oberdorfer, A. Ford, J. M. Gilmore, C. Xu, F. DiMaio, J. H. Pereira, B. Sankaran, G. Seelig, P. H. Zwart, D. Baker, De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **352**, 680–687 (2016). doi:10.1126/science.aad8865 Medline

13. E. Marcos, B. Basanta, T. M. Chidyausiku, Y. Tang, G. Oberdorfer, G. Liu, G. V. T. Swapna, R. Guan, D.-A. Silva, J. Dou, J. H. Pereira, R. Xiao, B. Sankaran, P. H. Zwart, G. T. Montelione, D. Baker, Principles for designing proteins with cavities formed by curved β sheets. *Science* **355**, 201–206 (2017). doi:10.1126/science.aah7389 Medline

14. P.-S. Huang, S. E. Boyken, D. Baker, The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016). doi:10.1038/nature19946 Medline

15. G. Bhardwaj, V. K. Mulligan, C. D. Bahl, J. M. Gilmore, P. J. Harvey, O. Cheneval, G. W. Buchko, S. V. S. R. K. Pulavarti, Q. Kaas, A. Eletsky, P.-S. Huang, W. A. Johnsen, P. J. Greisen, G. J. Rocklin, Y. Song, T. W. Linsky, A. Watkins, S. A. Rettie, X. Xu, L. P. Carter, R. Bonneau, J. M. Olson, E. Coutsias, C. E. Correnti, T. Szyperski, D. J. Craik, D. Baker, Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335 (2016). doi:10.1038/nature19791 Medline

16. E. A. Coutsias, C. Seok, M. P. Jacobson, K. A. Dill, A kinematic view of loop closure. *J. Comput. Chem.* **25**, 510–528 (2004). doi:10.1002/jcc.10416 Medline

17. D. J. Mandell, E. A. Coutsias, T. Kortemme, Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6**, 551–552 (2009). doi:10.1038/nmeth0809-551 Medline

18. C. D. Schwieters, J. J. Kuszewski, G. Marius Clore, Using Xplor–NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006). doi:10.1016/j.pnmrs.2005.10.001

19. C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. Marius Clore, The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65–73 (2003). doi:10.1016/S1090-7807(02)00014-9 Medline

20. J. Rizo, L. M. Gierasch, Constrained peptides: Models of bioactive peptides and protein substructures. *Annu. Rev. Biochem.* **61**, 387–418 (1992). doi:10.1146/annurev.bi.61.070192.002131 Medline

21. Y. M. Ibrahim, A. M. Hamid, L. Deng, S. V. B. Garimella, I. K. Webb, E. S. Baker, R. D. Smith, New frontiers for mass spectrometry based upon structures for lossless ion manipulations. *Analyst* **142**, 1010–1021 (2017). doi:10.1039/C7AN00031F Medline

22. E. G. Hutchinson, J. M. Thornton, A revised set of potentials for β-turn formation in proteins. *Protein Sci.* **3**, 2207–2216 (1994). doi:10.1002/pro.5560031206 Medline

23. D. S. Nielsen, N. E. Shepherd, W. Xu, A. J. Lucke, M. J. Stoermer, D. P. Fairlie, Orally absorbed cyclic peptides. *Chem. Rev.* **117**, 8094–8128 (2017). doi:10.1021/acs.chemrev.6b00838 Medline

24. J. Gavenonis, B. A. Sheneman, T. R. Siegert, M. R. Eshelman, J. A. Kritzer, Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nat. Chem. Biol.* **10**, 716–722 (2014). doi:10.1038/nchembio.1580 Medline

25. P. Bradley, K. M. S. Misura, D. Baker, Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868–1871 (2005). doi:10.1126/science.1113801 Medline

26. E. A. Coutsias, C. Seok, K. A. Dill, Using quaternions to calculate RMSD. *J. Comput. Chem.* **25**, 1849–1857 (2004). doi:10.1002/jcc.20110 Medline

27. F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker, F. Players, Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18949–18953 (2011). doi:10.1073/pnas.1115898108 Medline

28. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R.

Bonneau, P. Bradley, R. L. Dunbrack Jr., R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017). doi:10.1021/acs.jctc.7b00125 Medline

29. H. Park, P. Bradley, P. Greisen Jr., Y. Liu, V. K. Mulligan, D. E. Kim, D. Baker, F. DiMaio, Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016). doi:10.1021/acs.jctc.6b00819 Medline

30. S. C. Li, Y. K. Ng, Calibur: A tool for clustering large numbers of protein decoys. *BMC Bioinformatics* **11**, 25 (2010). doi:10.1186/1471-2105-11-25 Medline

31. B. North, A. Lehmann, R. L. Dunbrack Jr., A new clustering of antibody CDR loop conformations. *J. Mol. Biol.* **406**, 228–256 (2011). doi:10.1016/j.jmb.2010.10.030 Medline

32. J. J. Gray, S. Moughon, C. Wang, O. Schueler-Furman, B. Kuhlman, C. A. Rohl, D. Baker, Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299 (2003). doi:10.1016/S0022-2836(03)00670-3 Medline

33. S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E.-M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler, D. Baker, RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLOS ONE* **6**, e20161 (2011). doi:10.1371/journal.pone.0020161 Medline

34. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995). doi:10.1007/BF00197809 Medline

35. O. F. Lange, N.-A. Lakomek, C. Farès, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B. L. de Groot, Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008). doi:10.1126/science.1157092 Medline

36. M. J. Abraham *et al.*, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25 (2015/9).

37. S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl, in *Solving Software Challenges for Exascale* (Springer, Cham, Switzerland, 2014), pp. 3–27.

38. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010). Medline

39. J. Leszczynski, M. K. Shukla, *Practical Aspects of Computational Chemistry: Methods, Concepts and Applications* (Springer Science+Business Media, 2009).

40. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984). doi:10.1063/1.448118

41. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007). doi:10.1063/1.2408420 Medline

42. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995). doi:10.1063/1.470117

43. S. Páll, B. Hess, A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun.* **184**, 2641–2650 (2013). doi:10.1016/j.cpc.2013.06.003

44. B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997). doi:10.1002/(SICI)1096-987X(199709)18:12<1463:AID-JCC4>3.0.CO;2-H

45. C. R. Schwantes, V. S. Pande, Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000–2009 (2013). doi:10.1021/ct300878a Medline

46. B. Cronkite-Ratcliff, V. Pande, MSMExplorer: Visualizing Markov state models for biomolecule folding simulations. *Bioinformatics* **29**, 950–952 (2013). doi:10.1093/bioinformatics/btt051 Medline

47. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013). doi:10.1063/1.4811489 Medline

48. Y. Naritomi, S. Fuchigami, Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: The case of domain motions. *J. Chem. Phys.* **134**, 065101 (2011). doi:10.1063/1.3554380 Medline

49. Y. Ibrahim, M. E. Belov, A. V. Tolmachev, D. C. Prior, R. D. Smith, Ion funnel trap interface for orthogonal time-of-flight mass spectrometry. *Anal. Chem.* **79**, 7845–7852 (2007). doi:10.1021/ac071091m Medline

50. L. Deng, Y. M. Ibrahim, A. M. Hamid, S. V. B. Garimella, I. K. Webb, X. Zheng, S. A. Prost, J. A. Sandoval, R. V. Norheim, G. A. Anderson, A. V. Tolmachev, E. S. Baker, R. D. Smith, Ultra-high resolution ion mobility separations utilizing traveling waves in a 13 m serpentine path length structures for lossless ion manipulations module. *Anal. Chem.* **88**, 8957–8964 (2016). doi:10.1021/acs.analchem.6b01915 Medline