Supporting Information

Mou et al. 10.1073/pnas.1505072112

SI Materials and Methods

Protein Docking and Computational Design. The ENH crystal structure (PDB ID code 1ENH) was used as the scaffold for homodimerization, with side-chain atoms beyond C_β deleted and atomic radii of the remaining atoms adjusted as follows: N, 1.4 Å; O, 1.3 Å; C′, 1.75 Å; C_α, 2.35 Å; and C_β, 2.15 Å. A symmetric docking program based on an FFT algorithm was applied (1). Arrays used for docking calculations were 64 × 128 × 128 for each of the *X*, *Y*, and *Z* dimensions, with each element corresponding to 1 Å^3. Each round of searching consisted of extensive translational dockings followed by 1° increments about the *Y* and *Z* axes. Using shape complementarity as the criterion, the top 20 conformations for each rotational position were identified, combined into a set containing all of the top 20 conformations, and then ranked. The top 200 of these conformations were clustered into 11 groups based on structural similarity (rmsd values), the clusters were visually inspected, and one high-scoring model was selected for computational designs. ORBIT CPD software was used for stability designs for both ENH and NC3-NCap. Initial interface designs were also done using ORBIT, and subsequent designs and analyses were done using our improved CPD programs PHOENIX and TRIAD. Sequence optimization was performed using an improved version of FASTER (2) and a rotamer library based on the backbone-dependent library of Dunbrack and Karplus (3).

Construct Preparation, Protein Expression, and Purification. Oligonucleotides (Integrated DNA Technologies) containing ∼20-bp overlapping segments were assembled via a modified version of the method developed by Stemmer et al. (4) using KOD Hot Start Polymerase (Novagen) to generate genes for ENH, ENH-c2a, and ENH-c2b. For SDS/PAGE analysis, a His_6 or StrepII tag was added to generate His_6-ENH, ENH-c2a-Strep, and ENH-c2b-Strep constructs. For all other biophysical characterizations, intact ENH-c2b was used. The construct for solution NMR (ENH-c2b-Strep) was ENH-c2b with an extra N-terminal sequence (MEKRPR) and an extra C-terminal Gly followed by a Strep-tag II. All proteins were expressed using BL21 DE3 cells transformed by pET plasmids with 1 mM isopropyl β-D-1-thiogalactopyanoside in standard LB at 16 °C (His_6-ENH and ENH-c2a-Strep) or 37 °C (all other proteins). The $^{13}C/^{15}N$-labeled ENH-c2b-Strep for NMR experiments was prepared by growing BL21 DE3 cells in 1 L of LB until the OD_{600} reached ∼0.6 and then transferring the cells to 250 mL of M9 medium with ^{13}C glucose and ^{15}N ammonium chloride. Purification of ENH-c2b was accomplished by fusing it to His_6-ubiquitin, running the construct on an Ni^{2+}-NTA column (Qiagen), and then cleaving His_6-ubiquitin off using ubiquitin carboxyl-terminal hydrolase isozyme L3 protease (37 °C overnight). Strep-Tactin Sepharose (IBA) and Superdex 75 (Amersham Pharmacia) columns were used for Strep-tag affinity chromatography and size exclusion chromatography, respectively.

CD Spectroscopy. CD studies were performed on an Aviv 62A DS spectropolarimeter equipped with a thermoelectric temperature controller. Samples were prepared in 100 mM sodium chloride and 20 mM sodium phosphate buffer at pH 7.5. Wavelength scans and temperature denaturations were carried out in cuvettes with a 0.1-cm pathlength at a protein concentration of ∼10 μM. Three wavelength scans were performed at 25 °C for each sample and averaged. The thermal denaturation curve was collected at 222 nm from 0 to 99 °C, sampling every 1 °C separated by 2-min equilibration times (signal averaging time was 1 s). The refolding curve was collected after the thermal denaturation experiment using the same sample.

Analytical Ultracentrifugation. ENH-c2b was analyzed on an XL-1 analytical ultracentrifuge equipped with an AnTi60 rotor (Beckman Coulter). Two-channel Epon-filled centerpieces were used for the sedimentation velocity experiment. Cells were torqued to 130 pounds per inch and run at 60,000 rpm. Data were acquired at 230 nm and 20 °C in continuous mode. Data were first fit to the c(s) model (continuous distribution of sedimentation coefficient) and then converted to the c(M) model (continuous distribution of molecular mass). Time-invariant noises and baseline offsets were corrected before fitting. A maximum entropy regularization confidence level of 0.95 was used in all of the size distribution analyses.

Fluorescence Polarization Assay. Fluorescence polarization was measured at room temperature with a Fluorolog-3 spectrofluorometer (HORIBA). ENH-c2b was serially diluted in buffer containing 100 mM NaCl and 20 mM Tris·HCl at pH 8.0. Fluorescence anisotropy was measured for each sample, and the G-factor was determined individually. Data were analyzed according to a simple monomer-dimer equilibrium model and fit with KaleidaGraph software. Polarization values for the completely monomeric and dimeric states were fit to be 251 mA and 12 mA, respectively.

X-Ray Crystallography. ENH-c2b variant containing the 21-residue N-terminal sequence MGSSHHHHHHSSGLVPRGSHM was crystalized at room temperature in 1% tryptone, 20% (wt/vol) polyethylene glycerol 3350, and 0.05 M Hepes sodium at pH 7.0 using hanging-drop diffusion. Needle-like crystals appeared within 1 wk. The crystals were soaked in glycerol cryoprotectant and flash-frozen by cold nitrogen stream. Diffraction data were collected at beamline BL13C1 at the National Synchrotron Radiation Research Center in Taiwan. The best diffraction data had a resolution of ∼2.2 Å. However, due to X-ray overexposure, the overall data quality was not ideal, so the data were truncated to 3.5 Å for better refinement. Phases were obtained through molecular replacement using ENH (PDB ID code 1ENH) as the searching model. Further refinement was done with PHENIX (5). The crystal structure and statistics are shown in Fig. S3 and Table S2, respectively. Final coordinates were deposited in the PDB with ID code 4NDL.

Solution NMR Experiments. All spectra were acquired at 310 K on a Bruker Avance III 800 spectrometer equipped with a 5-mm z-gradient TCI (^1H, ^{13}C, and ^{15}N) CryoProbe (Bruker). ENH-c2b-Strep (1.9 mM protein in 300 μL) was dissolved in 100 mM NaCl, 5 mM $CaCl_2$, 10 mM DTT, 0.02% NaN_3, 5% D_2O, and 20 mM NH_4OAc at pH 4.5 in a Shigemi NMR tube. Assignment of main-chain and side-chain chemical shifts was based on ^1H-^{15}N HSQC, ^1H-^{13}C HSQC, CBCA(CO)NH, HNCACB, HNCO, HNCACO, HCCH-COSY, HCCH-TOCSY, HBHANH, HNHA(CO)NH, (H)CC(CO)NH, H(C)CCONH, HNHA, CACO, CON, and ^{15}N-TOCSY-HSQC experiments. NOE distance restraints were obtained from ^{15}N-edited NOESY, ^{13}C-edited NOESY (aliphatic), and ^{13}C-edited NOESY (aromatic) for intrachain or interchain contacts. An asymmetrically labeled dimer was prepared by mixing 1:1 uniformly $^{13}C/^{15}N$-labeled and unlabeled ENH-c2b. This sample was used for a 3D $^{13}C/^{15}N$-filtered NOESY-^1H-^{13}C-HSQC experiment (Fig. S4) to extract the interchain NOE restraints. The

^{1}H chemical shifts were referenced to 4,4-dimethyl-4-silapentane-1-sulfonate as the external standard. The ^{15}N and ^{13}C chemical shifts were referenced using the consensus ratios of the zero-point frequencies at 310 K. Data were processed with Topspin (Bruker) for Fourier transformations and analyzed with CCPN for chemical shift assignments (6).

Solution Structure Determination. TOLOS+ (7) was used for φ/ψ restraints predicted by backbone chemical shifts. Backbone hydrogen-bond restraints were created between consecutive i/i + 4 helical residues. The φ/ψ restraints, hydrogen-bond restraints, and a set of partially manual NOE assignments were used as the initial input for ARIA2.3 (8). Automated NOE cross-peak assignments and structure calculations were then applied iteratively by ARIA2.3. For regular NOESY experiments, all NOE cross-peaks were treated ambiguously as inter- or intramolecular restraints. For the ^{13}C/^{15}N-filtered NOESY-^{1}H-^{13}C-HSQC, the NOE cross-peaks were treated as intermolecular restraints only. A soft square potential was used in the simulated annealing protocol with automated determination of weights for NOE-derived restraints. Two identical chains of ENH-c2b-Strep were included in calculations using the CNS protocol in ARIA2.3. The homodimer symmetry was enforced by a C2-symmetry restraint energy. In addition, a packing restraint was applied between the centers-of-mass of the two chains to facilitate the association of dimer. The seven highest scoring structures of a total of 32 generated structures were chosen for every cycle to obtain assignment statistics. A total of eight cycles were run, and the 10 lowest energy models were refined in explicit water at the end to obtain the final NMR ensemble. The NMR statistics are shown in Table S3.

1. Huang P-S, Love JJ, Mayo SL (2005) Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J Comput Chem* 26(12):1222–1232.
2. Allen BD, Mayo SL (2006) Dramatic performance enhancements for the FASTER optimization algorithm. *J Comput Chem* 27(10):1071–1075.
3. Dunbrack RL, Jr, Karplus M (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol* 230(2):543–574.
4. Stemmer WP, Crameri A, Ha KD, Brennan TM, Heyneker HL (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164(1):49–53.
5. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
6. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59(4):687–696.
7. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 44(4): 213–223.
8. Rieping W, et al. (2007) ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23(3):381–382.

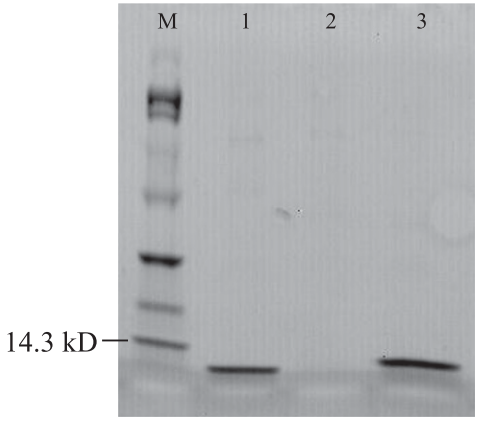

Fig. S1. SDS/PAGE of purified proteins from soluble fractions. The labels are as follows: M, marker; 1, ENH; 2, ENH-c2a; and 3, ENH-c2b. ENH-c2a is not expressed in the soluble fraction. Molecular masses for ENH and ENH-c2b are as expected (~6.5 kDa).

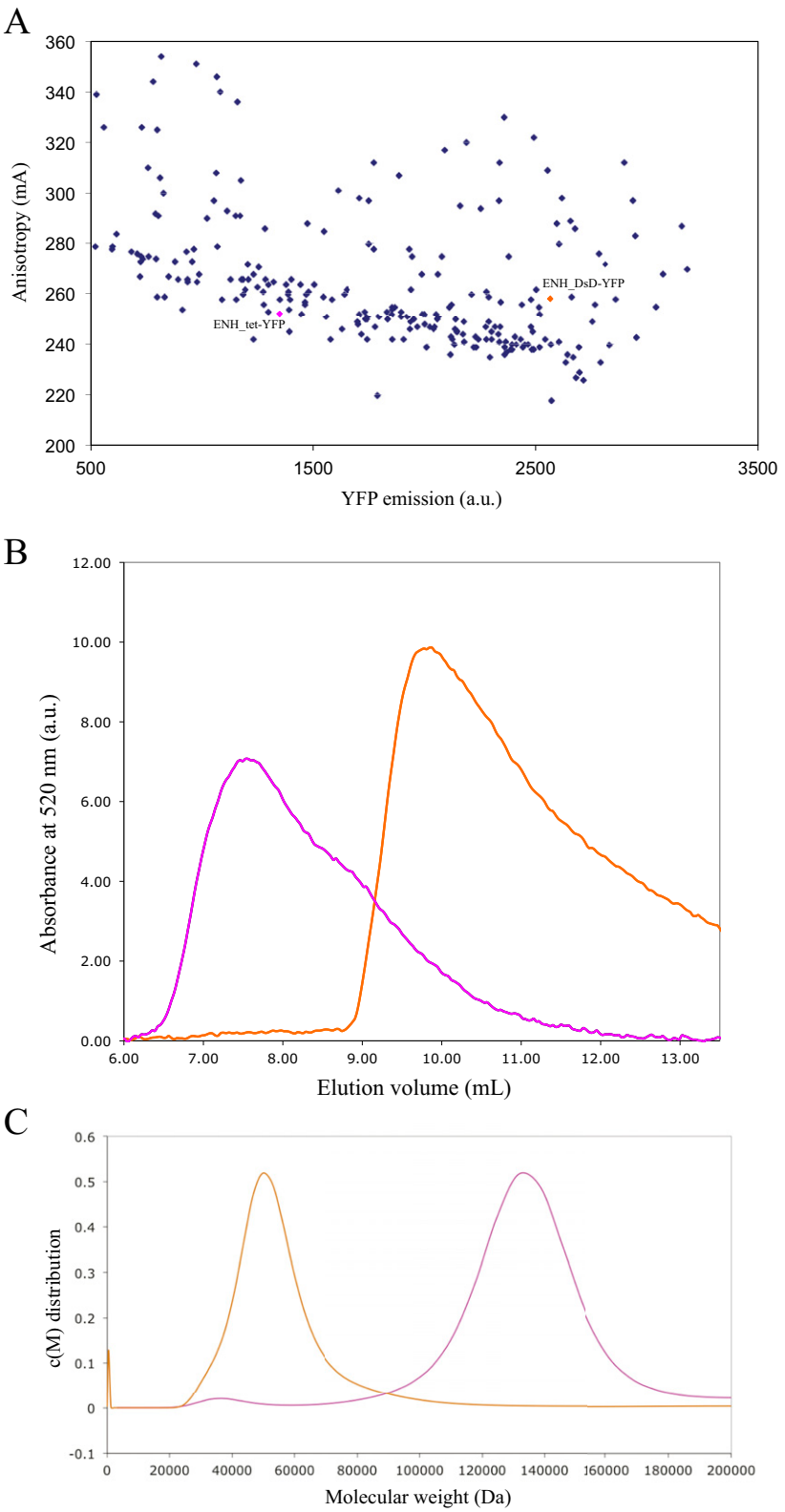

Fig. S2. Experimental characterizations of the computational library design. (*A*) Homo-FRET assay of the library members. Each dot represents one member in the library. Because of the homo-FRET effect, members with lower anisotropic values are likely to have stronger dimer affinities or higher oligomeric states. Two representative members, shown in orange and magenta (named ENH_DsD-YFP and ENH_tet-YFP, respectively), were chosen for further biophysical characterization. (*B*) Size exclusion chromatography shows that ENH_DsD-YFP (orange curve) is likely a dimer and ENH_tet-YFP is a higher order oligomer. (*C*) Sedimentation velocity experiments show that ENH_DsD-YFP is in equilibrium between monomeric and dimeric states (orange curve) and that ENH_tet-YFP is a tetramer (magenta curve). The molecular masses of the ENH_DsD-YFP and ENH_tet-YFP monomers are nearly identical (~34.4 kDa).

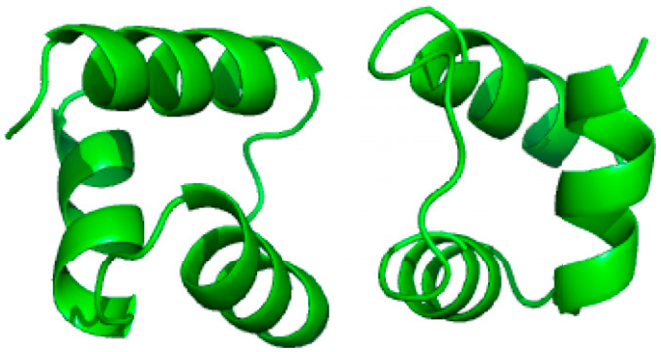

Fig. S3. Crystal structure of an ENH-c2b variant containing the 21-residue N-terminal sequence MGSSHHHHHHSSGLVPRGSHM (PDB ID code 4NDL). In the asymmetric unit, two chains form crystal contacts between helices 1 and 2. Evaluation by protein interfaces, surfaces, and assemblies (PISA) indicates that these contacts represent crystal packing rather than a biological interface, mainly because of the small area (699 Å^2) (the average interface area for a homodimer is 2,740 ± 1,240 Å^2).

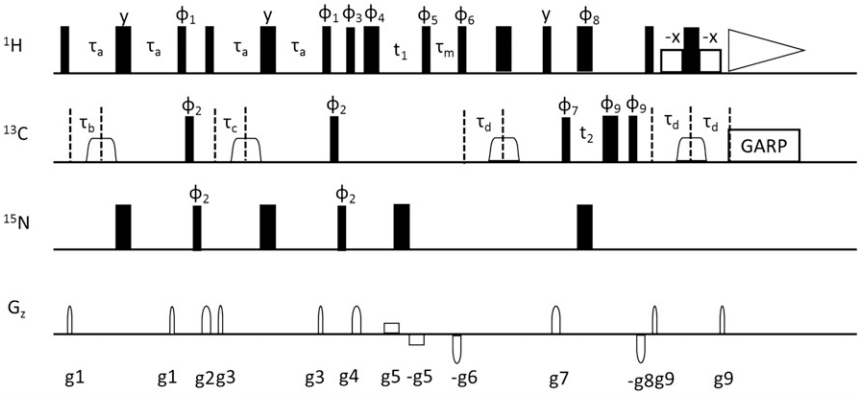

Fig. S4. Pulse sequence of the 3D ^{13}C/^{15}N-filtered NOESY-^{1}H-^{13}C-HSQC experiment. The frequencies of proton, ^{13}C, and ^{15}N pulses are set at 4.7, 39, and 118.5 ppm, respectively, except for the third and fourth ^{13}C smoothed Chirp pulses, whose frequencies are set at 69 ppm. The duration of the ^{13}C smoothed Chirp pulses with a 60-kHz frequency sweep is 500 μs, and the duration of ^{1}H soft 90° pulses is 1 ms. The phase cycle is as follows: $\varphi_1 = -x$; $\varphi_2 = (x, y, -x, -y)$; $\varphi_3 = 4(x)$, $4(-x)$; $\varphi_4 = 4(y)$, $4(-y)$; $\varphi_5 = 16(x)$, $16(-x)$; $\varphi_6 = 8(x)$, $8(-x)$; $\varphi_7 = (x, -x)$; $\varphi_8 = (x, x, -x, -x)$; $\varphi_9 = (x, x, -x, -x)$; $\varphi_{rec} = (x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, -x, x, x, -x, -x, x, -x, x, x, -x)$. The duration and strength of the smoothed square gradient pulses are g1 = (500 μs, 1.59 G/cm), g2 = (1 ms, 5.83 G/cm), g3 = (500 μs, 1.06 G/cm), g4 = (1 ms, 3.71 G/cm), g6 = (1 ms, 10.6 G/cm), g7 = (1 ms, 21.2 G/cm), g8 = (1 ms, 26.5 G/cm), and g9 = (500 μs, 31.8 G/cm). The strength of g5 is 1.06 G/cm. The mixing time (τ_m) is set as 300 ms. The periods of τ_a, τ_b, τ_c, and τ_d are set as 2.78, 2.00, 1.66, and 1.79 ms, respectively. GARP, globally optimized alternating-phase rectangular pulse.

Table S1. Library of 128 NC3-NCap variants

9	10	13	14	16	17	18	25	28	32
K	AT	L	A	FY	FV	DFVY	AW	FY	R

Computational library design for homodimerization. Interfacial residues shown above were chosen for sequence optimization. The library size was set to 128 members.

Table S2. Data collection and refinement statistics for the crystal structure ENH-c2b with an extra 21-residue tag at the N terminus (PDB ID code 4NDL)

Statistics	Value
Data collection	
Space group	$C222_1$
Cell dimensions	
a, *b*, *c*; Å	87.6, 167.8, 29.7
α, β, γ; °	90.0, 90.0, 90.0
Resolution, Å	25.9–2.2
R_{merge}	0.147
I/σ	8.0
Completeness, %	94.1
Multiplicity	5.3
Refinement	
Resolution, Å	24.6–3.5
No. of reflections	5,697
R_{work}/R_{free}, %	31/32
No. of molecules in asymmetric unit	3
Number of atoms	2,024
Protein	2,023
Water	1
B factors, $Å^2$	27.3
Protein	27.4
Water	17.1
rmsd	
Bond lengths, Å	0.009
Bond angles, °	1.77
Ramachandron map analysis	
Most favored regions, %	96.9
Additional allowed regions, %	3.1
Disallowed regions, %	0.0

I/σ, average signal-to-noise ratio; R, Σ | |F(obs)| − |F(calc)| |/Σ|F(obs)|; R_{free}, R with 5% of reflections sequestered before refinement; R_{merge}, ΣhklΣi| I(hkl)i − | /ΣhklΣi over i observations of a reflection hkl; R_{work}, Σ | |F(obs)| − |F(calc)| l/ΣlF(obs)l.

Table S3. NMR statistics for the structure ENH-c2b-Strep (PDB ID code 2MG4)

Summary of restraints	
Total NOE distance restraints	1,347
Intramolecular unambiguous	1,134
Intramolecular ambiguous	186
Intermolecular unambiguous	26
Intermolecular ambiguous	1
Hydrogen bonds	42
Dihedral angle restraints, ϕ/ψ	82/82
rmsd from restraints	
NOE restraints, Å	0.061 ± 0.021
H-bond restraints, Å	0.026 ± 0.010
Dihedral restraints, °	1.16 ± 0.26
rmsd from idealized geometry	
Bonds, Å	0.0051 ± 0.0002
Angles, °	0.72 ± 0.03
Improper, °	1.75 ± 0.17
Coordinate precision rmsd, Å	
Backbone, secondary structure	0.64 ± 0.23
Heavy atoms, secondary structure	1.19 ± 0.19
Backbone, all	1.40 ± 0.40
Heavy atoms, all	2.20 ± 0.34
Ensemble Ramachandran statistics, %	
Residues in most-favored region	87.9
Additionally allowed region	7.6
Generally allowed region	4.5
Disallowed region	0.0

Table S4. Completeness of NOEs at different distance shells

Shell, Å	Observed NOE	Expected NOE	NOE completeness, %	Cumulative NOE completeness, %
0.00–2.50	72	121	59.5	59.5
2.50–3.00	172	308	55.8	56.9
3.00–3.50	188	483	38.9	47.4
3.50–4.00	211	846	24.9	36.6
4.00–4.50	250	1,346	18.6	28.8
4.50–5.00	206	1,926	10.7	21.8

Table S5. Intermolecular NOE restrains that have no or one violation against all 10 NMR models

Proton in chain A	Proton in chain B	Average distance*	No. of models violated
12Glu-HB	12Glu-HG	4.4	0
12Glu-HB	19Leu-HD	5	1
12Glu-HG	12Glu-HB	4.4	0
15Lys-H	16Ala-HB	4.1	0
15Lys-HB	16Ala-HB	3.1	1
15Lys-HD	16Ala-HB	5.4	1
15Lys-HD	40Gly-HA	4	0
15Lys-HE	38Arg-HB	3	1
15Lys-HE	39Leu-HA	3.1	1
15Lys-HE	39Leu-HB	3.5	0
15Lys-HE	39Leu-HD	2.7	1
15Lys-HE	39Leu-HG	4.8	1
15Lys-HG	38Arg-HB	3.8	1
16Ala-H	16Ala-HB	3.9	1
16Ala-H	39Leu-HD	4	1
16Ala-HA	16Ala-HB	2.1	0
16Ala-HA	16Ala-HB	2.8	0
16Ala-HA	19Leu-HB	3	0
16Ala-HA	19Leu-HD	3.3	0
16Ala-HB	15Lys-HD	5.4	1
16Ala-HB	16Ala-H	3.9	1
16Ala-HB	16Ala-HA	2.1	0
16Ala-HB	16Ala-HA	2.8	0
16Ala-HB	19Leu-HD	2.5	1
16Ala-HB	39Leu-HD	4.7	1
17Leu-HD	19Leu-HD	3.2	0
18Asp-H	19Leu-HD	2.2	0
19Leu-H	19Leu-HD	2.2	0
19Leu-H	26Arg-HB	4.9	1
19Leu-HA	19Leu-HD	5.2	1
19Leu-HB	19Leu-HD	3.5	0
19Leu-HD	12Glu-HB	5	1
19Leu-HD	15Lys-HG	3.8	0
19Leu-HD	16Ala-HA	2.7	0
19Leu-HD	16Ala-HA	3.3	0
19Leu-HD	16Ala-HB	2.5	1
19Leu-HD	17Leu-HD	3.2	0
19Leu-HD	18Asp-H	2.2	0
19Leu-HD	18Asp-HA	4.2	1
19Leu-HD	19Leu-HA	5.2	1
19Leu-HD	19Leu-HB	3.5	0
19Leu-HD	19Leu-HD	3.8	0
19Leu-HD	19Leu-HG	5.5	1
19Leu-HD	20Ala-H	3.4	1
19Leu-HD	20Ala-HA	3.8	1
19Leu-HD	20Ala-HB	2.2	0
19Leu-HD	25Arg-HA	5.4	1
19Leu-HD	35Leu-HB	3.7	1
19Leu-HG	19Leu-HD	5.5	1
20Ala-H	16Ala-HB	5	0
20Ala-HB	19Leu-HD	2.2	0
26Arg-HB	19Leu-H	4.9	1
35Leu-HB	19Leu-HD	3.7	1
38Arg-HB	15Lys-HE	3	1
39Leu-HB	15Lys-HE	3.5	0
39Leu-HD	15Lys-HE	2.7	1
39Leu-HD	16Ala-HB	4.7	1
39Leu-HD	39Leu-HD	5.9	1
39Leu-HG	15Lys-HE	4.8	1

H, amide hydrogen; HA, alpha hydrogen; HB, beta hydrogen; HD, delta hydrogen; HE, epsilon hydrogen; HG, gamma hydrogen.

*The proton-proton distance is the average value of the 10 models in the NMR ensemble.

Table S6. List of expected NOEs with a 4-Å cutoff distance

Proton in chain A	Proton in chain B	Average distance*	NOE observed	Reason for missing NOE
12Glu-HG	12Glu-HG	3.7	No	Unclear
15Lys-HB	16Ala-HB	3.1	Yes	
15Lys-HD	38Arg-HB	3.5	No	Unclear
15Lys-HD	39Leu-HA	3.1	No	Buried by H_2O
15Lys-HE	38Arg-HB	3	Yes	
15Lys-HE	39Leu-HA	3.1	Yes	
15Lys-HE	39Leu-HB	3.5	Yes	
15Lys-HE	39Leu-HD	2.7	Yes	
15Lys-HG	38Arg-HB	3.8	Yes	
16Ala-H	16Ala-HB	3.9	Yes	
16Ala-HA	16Ala-HA	2.8	No	Buried by H_2O
16Ala-HA	16Ala-HB	2.1	Yes	
16Ala-HA	19Leu-HB	3	Yes	
16Ala-HA	19Leu-HD	3.3	Yes	
16Ala-HB	16Ala-H	3.9	Yes	
16Ala-HB	16Ala-HA	2.1	Yes	
16Ala-HB	19Leu-HD	2.5	Yes	
17Leu-HA	19Leu-HD	3.8	No	Buried by H_2O
17Leu-HD	19Leu-HD	3.2	Yes	
19Leu-HB	16Ala-HA	3	No	Unclear
19Leu-HB	19Leu-HB	3.3	No	Unclear
19Leu-HB	19Leu-HD	3.5	Yes	
19Leu-HB	20Ala-H	3.9	No	Unclear
19Leu-HB	20Ala-HA	3.4	No	Unclear
19Leu-HD	16Ala-HA	3.3	Yes	
19Leu-HD	16Ala-HB	2.5	Yes	
19Leu-HD	17Leu-HA	3.8	No	Unclear
19Leu-HD	17Leu-HD	3.2	Yes	
19Leu-HD	19Leu-HB	3.5	Yes	
19Leu-HD	20Ala-H	3.4	Yes	
19Leu-HD	20Ala-HA	3.8	Yes	
19Leu-HD	20Ala-HB	2.2	Yes	
19Leu-HD	34TYR-HE	3.6	No	δ(34Tyr-HE) not assigned
19Leu-HD	35Leu-HD	3.5	No	Unclear
19Leu-HG	20Ala-HB	4	No	δ(19Leu-HG) not assigned
20Ala-H	19Leu-HB	3.9	No	δ(20Ala-N) not assigned
20Ala-H	19Leu-HD	3.4	No	δ(20Ala-N) not assigned
20Ala-HA	19Leu-HB	3.4	No	Buried by H_2O
20Ala-HA	19Leu-HD	3.8	No	Buried by H_2O
20Ala-HB	19Leu-HD	2.2	Yes	Unclear
20Ala-HB	19Leu-HG	4	No	δ(19Leu-HG) not assigned
34TYR-HE	19Leu-HD	3.6	No	δ(34Tyr-HE) not assigned
35Leu-HD	19Leu-HD	3.5	No	Unclear
38Arg-HB	15Lys-HD	3.5	No	Unclear
38Arg-HB	15Lys-HE	3	Yes	
38Arg-HB	15Lys-HG	3.8	No	Unclear
39Leu-HA	15Lys-HD	3.1	No	Buried by H_2O
39Leu-HA	15Lys-HE	3.1	No	Unclear
39Leu-HB	15Lys-HE	3.5	Yes	
39Leu-HD	15Lys-HE	2.7	Yes	

*The proton-proton distance is the average value of 10 models in the NMR ensemble.