Supporting information for:

Computational prediction of cyclic peptide structural ensembles and application to the design of Keap1 binders

Francini Fonseca Lopez,[1] Jiayuan Miao,[1] Jovan Damjanovic,[1] Luca Bischof,[2] Michael B. Braun,[2] Yingjie Ling,[1] Marcus D. Hartmann,[2] Yu-Shan Lin[*,1] and Joshua A. Kritzer[*,1]

[1]Department of Chemistry, Tufts University, Medford, MA, USA 02155

[2]Department of Protein Evolution, Max Planck Institute for Biology, 72076 Tübingen, Germany; Interfaculty Institute of Biochemistry, Tübingen University, 72076 Tübingen, Germany

*Corresponding authors: yu-shan.lin@tufts.edu and joshua.kritzer@tufts.edu

Table of Contents

Materials and Methods

Charge-generation protocol.

Generalized AMBER force field (GAFF) was used for the linkers pfl, oxy, mxy, and pxy; **Figure S1**.[12] To generate partial atomic charges for each linker, we followed the protocol on this website http://ambermd.org/antechamber/pro4.html. Thirty different conformations of each linker moiety were built in the following format: each of the two free termini of the linker to be parametrized was connected to the sulfur atom of a cysteine which was capped by an acetyl group on the N-terminus and to a N-methyl group on the C-terminus. One example conformation for the linker oxy is shown in **Figure S1F**. The obtained partial charges, together with non-bonded and bonded parameters found in the Generalized AMBER force field (GAFF) were incorporated into the RSFF2 force field of the Gromacs software.[10,12] Specifically, we modified the aminoacids.rtp, aminoacids.hdb, residuetypes.dat, specbond.dat, atomtypes.atp, ffnonbonded.itp, and the ffbonded.itp files to include the new parameters. After the modification, the linker residues can be recognized by Gromacs as other amino acid residues. Now we describe the exact protocol we used to develop the partial charges on the example of oxy below. The same protocol was used to develop the partial charges for mxy, pxy, and pfl. The 30 structures we built for oxy were saved as oxy_2c_i.pdb, where i = 1, 2, 3, … 30.

Step 1: In the Antechamber program in AmberTools, we generated mol2 files for each conformation:

  antechamber -i oxy_2c_i.pdb -fi pdb -o oxy_2c_i.mol2 -fo mol2 -c bcc -s 2

Step 2: We used the mol2 file to generate the Gaussian input file:

  antechamber -fi mol2 -fo gzmat -i oxy_2c_i.mol2 -o oxy_2c_i.gau

Step 3: We performed geometry optimization in the Gaussian09 software[25] using the HF/6-31G* level of theory:

  g09 < oxy_2c_i.gau > oxy_2c_i.out

Step 4: We extracted the electrostatic potential (ESP) information from the Gaussian output file, and output an esp file to be read by the resp program:

  espgen -i oxy_2c_i.out -o oxy_2c_i.esp

Step 5: We concatenated the 30 esp files into one file:

  cat oxy_2c_1.esp oxy_2c_2.esp … oxy_2c_30.esp > oxy_2c.esp

Step 6: We generated an ac file from the Gaussian output file:

antechamber -fi gout -fo ac -i oxy_2c_1.out -o oxy_2c_1.ac -c resp

Step 7: We prepared input files for two-stage restrained electrostatic potential (RESP) fitting:

respgen -i oxy_2c_1.ac -o oxy_2c-step1.respin -f resp1 -n 30

respgen -i oxy_2c_1.ac -o oxy_2c-step2.respin -f resp2 -n 30

Step 8: We ran resp program to get the RESP charges:

resp -O -i oxy_2c-step1.respin -o oxy_2c-step1.respout -e oxy_2c.esp -t qout_stage1

resp -O -i oxy_2c-step2.respin -o oxy_2c-step2.respout -e oxy_2c.esp -q qout_stage1 -t qout_stage2

Step 9: We generated an ac file from the Gaussian output file, reading in RESP charges:

antechamber -fi gout -fo ac -i oxy_2c_1.out -o oxy_2c.ac -c rc -cf qout_stage2

Step 10: We prepared a maichain.oxy file, following the instructions on the website: http://ambermd.org/antechamber/pro4.html.

Step 11: We generated the prep file:

prepgen -i oxy_2c.ac -o oxy.prepi -f prepi -m mainchain.oxy -rn OXY -rf oxy.res

The partial charges can be found in the oxy.prepi file, and they are shown in **Figure S2**.

Peptide synthesis.

Peptides were synthesized by Fmoc solid-phase peptide synthesis on either an automated Tribute peptide synthesizer (Gyros Protein Technologies) or an automated Prelude peptide synthesizer (Gyros Protein Technologies). Peptides were synthesized on Rink amide resin with Fmoc deprotection in 20 % piperidine in N,N-dimethylformamide (DMF). 5 equivalents of amino acid were used for coupling along with 5 equivalents of coupling reagent (2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 5 equivalents of the racemization suppression reagent hydroxybenzotriazole (HOBt), and 10 equivalents of N,N-diisopropylethylamine (DIPEA). DMF and dichloromethane washes were performed between each step. To couple biotin to peptides, the N-terminus was deprotected in 20% piperidine in DMF for 7 min followed by 2 h incubation in 5 equivalents of biotin-N-hydroxysuccinimideester (Millipore Sigma) and 10 equivalents of DIPEA in DMF. Repeated couplings were performed with the same conditions if MALDI-TOF mass spectrometry analysis showed incomplete conversion after the first coupling. All peptides were globally deprotected and cleaved using a trifluoroacetic acid (TFA) cleavage cocktail (95 : 2 : 2:1 TFA:water:ethane 1,2-dithiol:triisopropylsilane) for 3 h. Cleaved peptides were precipitated in cold diethyl ether, pelleted, and washed with additional cold diethyl ether. Peptides were then lyophilized and resuspended in water/acetonitrile (50 : 50) for reverse-phase HPLC purification on a preparative-scale C8 column using a gradient of 5–100 %

acetonitrile with 0.1% trifluoroacetic acid over 20 min. Linear peptides were purified to at least 80% purity as determined by analytical HPLC. MALDI-TOF mass spectrometry was used to verify peptide mass in a matrix of 15 mg/mL α-cyano-4-hydroxycinnamate in 50:50 water:acetonitrile with 0.1 % TFA. After purification, each linear peptide was lyophilized and stored at –20 °C until use. Concentrations of peptide solutions were quantified based on absorbance at 280 nm using a Nanodrop 1000 (ThermoFisher).

For stapling of perfluoroaryl-linked (pfl) peptides, linear peptide was dissolved in a solution of hexafluorobenzene (25 equivalents, 100 mM) in acetonitrile plus an equal volume of Tris base (12.5 mM) in 50:50 DMF/acetonitrile. The vessel was shaken for 60 s and left at room temperature for 18 h. For stapling of *ortho*-, *meta*-, and *para*-dibromomethylbenzene linked peptides, a 1 mM solution of peptide was prepared in 50:50 acetonitrile:water buffered with 20 mM ammonium bicarbonate, pH 8.8. 1.5 equivalents of the linker was dissolved in 1-2 mL of acetonitrile and added to the peptide, and the vessel was agitated for 2 h. CPs (**Table S1**) were purified to at least 95% purity as determined by analytical HPLC using the purification methods described above. MALDI-TOF mass spectrometry was used to confirm the mass of each CP. Following purification, CPs were again lyophilized and resuspended in DMSO to make working stocks, which were quantified based on their A280 when diluted at least 1:100 in aqueous buffer. Final purification traces can be seen in **Figure S5**.

Protein expression and purification for biolayer interferometry.

Recombinant, His-tagged Kelch binding domain of Keap1 was expressed in BL21 (DE3) *E. coli* transformed with pET15b expression plasmids encoding the protein.[19] The plasmid was obtained from Mark Hannink, and the same plasmid was used for the crystal structure obtained by Lo and others.[17] Transformed cells were plated on ampicillin agar plates and incubated at 37 °C overnight. Individual colonies were picked and grown overnight shaking at 37 °C in 5 mL of LB culture medium with 100 μg/mL ampicillin. Each 5 mL culture was then added to 1 L of LB culture medium with 100 μg/mL ampicillin and incubated with shaking, at 37 °C until the OD_{600} measured roughly 0.6. At this point, protein expression was induced by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Following induction, cells were incubated shaking for 3 h at 37°C. Cells were then pelleted and stored at -80°C. To purify the protein, cells were resuspended in 40 mL of lysis buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM imidazole, 0.1 mM ethylenediaminetetraacetic acid in deionized water) with 80 mg lysozyme, 2 protease inhibitor cocktail pellets (Roche), and 4 μL of universal nuclease (Pierce). Resuspended cells were sonicated in cycles of 10 s on and 10 s off for 20 min, and lysed cells were centrifuged to separate

the lysate and cellular debris. Clarified lysate was purified using batch affinity purification with HisPur Ni-NTA resin (ThermoFisher). Resin was incubated with the lysate at 4°C for 1 h before washing with 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.1 mM ethylenediaminetetraacetic acid in deionized water. Protein was then eluted from the resin in elution buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 300 mM imidazole, 0.1 mM ethylenediaminetetraacetic acid in deionized water). Protein purity and mass were evaluated by SDS-PAGE. Protein was buffer exchanged with a desalting column into a storage buffer (10 mM 1,4-dithiothreitol, 10 mM sodium phosphate monobasic at pH 8.8). A typical yield after batch purification and desalting was roughly 24 mg protein per L culture. Protein was then aliquoted, flash frozen in liquid nitrogen and stored at –80°C. Before each use, protein was thawed on ice and concentration was quantified by absorbance at 280 nm using a Thermo Scientific Nanodrop 1000.

Biolayer interferometry (BLI).

BLI assays were performed on an Octet K2 System (Forté Bio). Biotinylated peptides were diluted in assay buffer (PBS pH 7.4, 0.05% Tween-20, 1 mM EDTA, 1 mg/mL BSA, 100 mM 1,4-dithiothreitol) to a final volume of 200 µL in a flat, black 96-well polypropylene plate (Greiner Bio-One). Biotinylated peptides were loaded onto streptavidin-coated BLI tips, tips were washed, and then association to Keap1 was measured. Serial dilutions of protein were prepared in assay buffer to a final volume of 200 µL. Loading tests were performed for each peptide to inform the optimal concentration to load onto the sensor tip. Different ranges of protein concentrations were also tested to optimize the concentration of Keap1. Assays were run at 30 °C with shaking at 1000 rpm. A reference well measuring buffer association to the immobilized peptide was subtracted from the data. Responses were aligned to the average of the baseline step, and inter-step correction was applied to correct for jumps in responses between steps. Octet Data Analysis HT 10.0 software (FortéBio) was used to generate curve fits using a 1:1 global fitting model for combined association and disassociation curves for each peptide:protein pair with R_{max} unlinked. Multiple protein concentrations were tested in each independent replicate, and only experiments where the residual of the curve fit was less than 10% of the maximal response were included in the global fit. At least 3 concentrations were used for a global fit that calculated k_{on}, k_{off}, and K_d for the peptide-protein interaction. All peptides were tested in at least three independent replicates and average K_d, k_{on}, and k_{off} values and standard errors were calculated.

Protein expression and purification for crystallography.

For crystallization, a Kelch construct with the mutations E540A, E542A, C319S, C613S, C622S and C624S was used as previously described.[20] *E. coli* BL21 (DE3) cells were transfected with a pET30a(+) vector containing the construct. Cells were grown in LB-media supplemented with 50 μg/L Kanamycin at 37°C to an OD of 0.6. After induction with 1 mM IPTG the temperature was reduced to 20°C and the cells were harvested the next day. After lysis by sonification in a 50 mM Tris-HCl 8.0, 500 mM NaCl, 20 mM imidazole and 2.5 mM 1,4-dithiothreitol (DTT) containing buffer and centrifugation the protein was purified using a Ni-NTA column. The protein was eluted with lysis buffer supplemented with 250 mM imidazole. Afterwards, the protein was dialyzed and the His-tag was cleaved using TEV-protease. Uncleaved protein as well as the TEV-protease was removed in a second Ni-NTA chromatography. SEC was done using a SD200 16/600 column and a buffer containing 25 mM Tris-HCl 8.0, 150 mM NaCl and 5 mM DTT. Finally, the buffer was exchanged to 25 mM Tris-HCl 8.0, 5 mM DTT with a PD MiniTrap G25 column.

X-ray crystallography.

Crystals were grown in a vapor diffusion hanging drop setup by mixing 5.1-6.2 mg/mL protein solution in 25 mM Tris-HCl 8.0, 5 mM DTT with 0.2 % PEG 550, 0.1 M BisTris-HCl 6.5 and 1.5 M NH_4SO_4 at 20°C. The crystals were incubated for 24 h in reservoir solution supplemented with 15 % glycerol as cryoprotectant and either 1 mM or 3 mM cyclic peptides.

Diffraction data was collected at beamline X06SA at the Swiss light source, Villigen, Switzerland at 100K using an EIGER 16M X detector (Dectris). The data was processed using XDS.[21] Molecular replacement was done with Phaser using pdb:5WFI as model. Coot[22] and phenix.refine[23] were used for model building and refinement, while additional restraints were generated using AceDRG.[24] Data collection and refinement statistics are summarized in **Table S4** and structures deposited in the RCSB Protein Data Bank under the accession codes 8PKU (CP3), 8PKV (CP4), 8PKW (CP5), 8PKX (CP11).
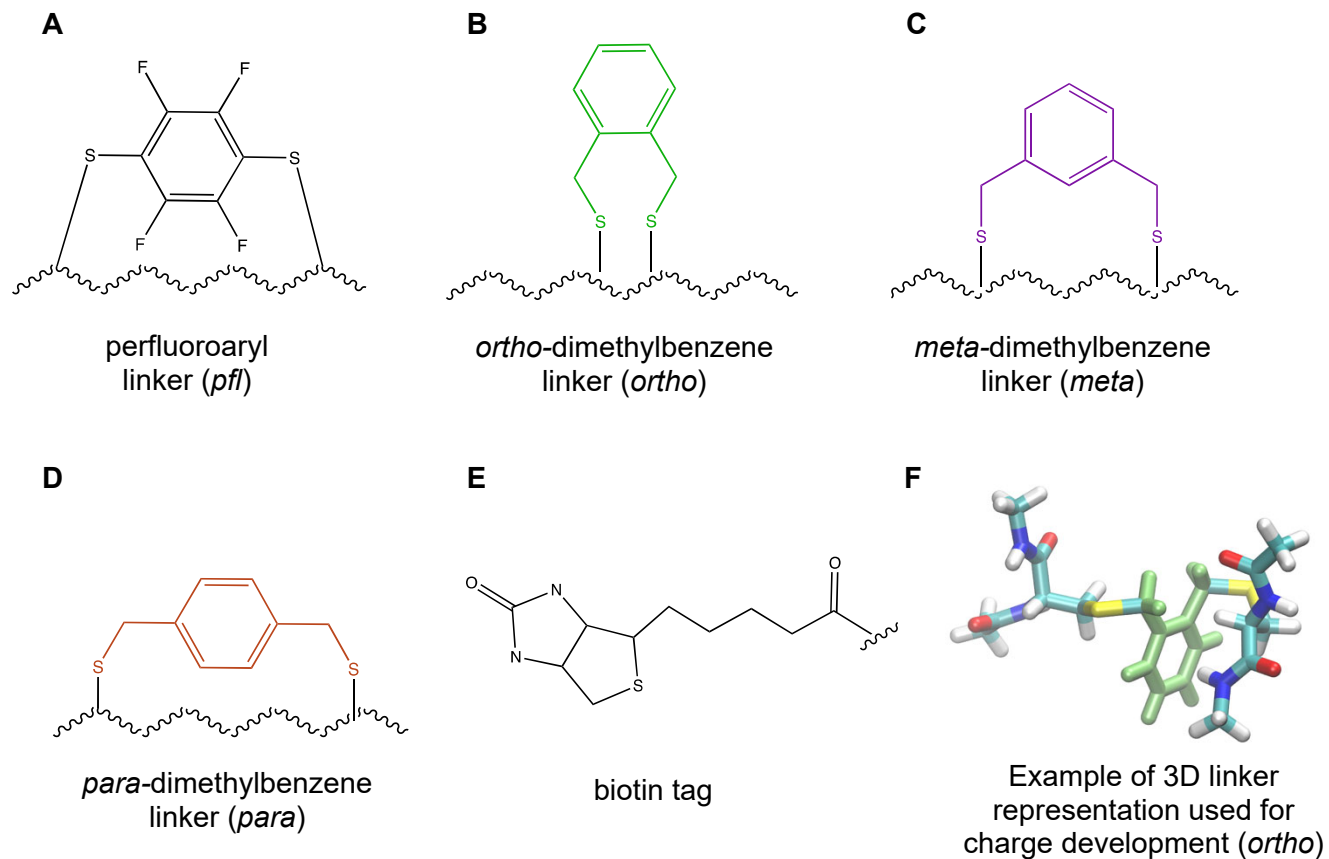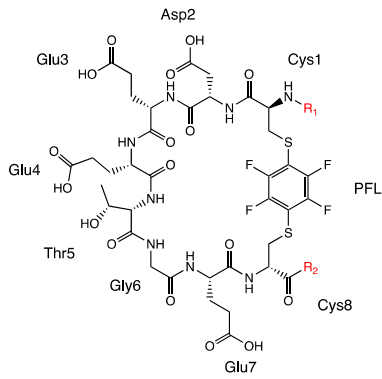
Figure S1. Chemical structures of linkers and tags. A. perfluoroaryl linker **B.** *ortho*-dimethylbenzene linker **C.** *meta*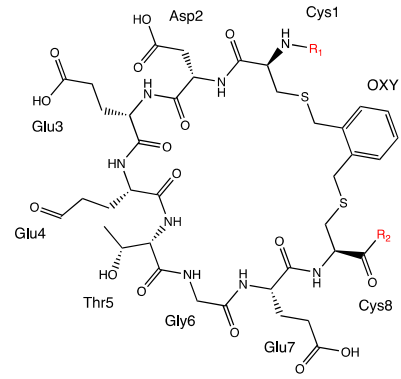-dim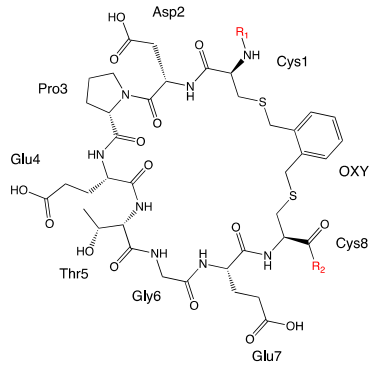ethylbenzene linker **D.** *para*-dimethylbenzene linker **E.** biotin tag coupled on the N-terminus of all peptides **F.** Example of a compound that was used to develop the partial atomic charges. In this example, the *ortho*-dimethylbenzene linker was attached to the sulfur atoms of cysteine residues that were capped by an acetyl group and an N-methyl group on the N- and C-terminus, respectively. Analogous compounds were used to develop partial atomic charges for *meta*-dimethylbenzene, *para*-dimethylbenzene, and perfluoroaryl linkers.

A perfluoroaryl
linker (*pfl*)

B *ortho*-dimethylbenzene
linker (*ortho*)

C *meta*-dimethylbenzene
linker (*meta*)

D *para*-dimethylbenzene
linker (*para*)
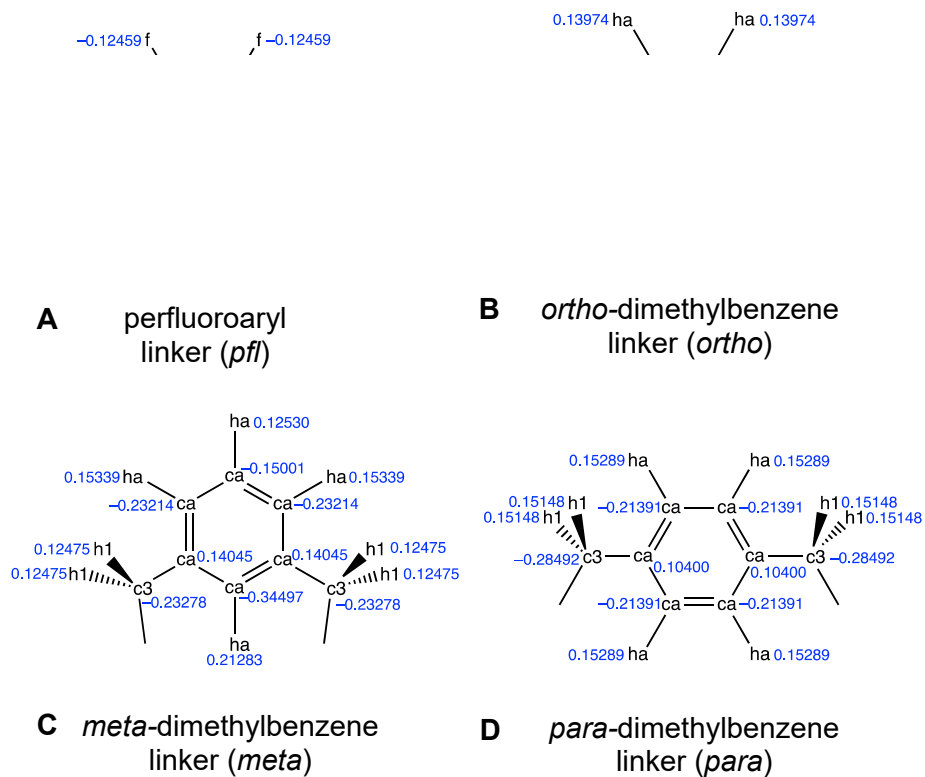
Figure S2. Atom types and partial charges for linkers. A. perfluoroaryl linker **B.** *ortho*-dimethylbenzene linker **C.** *meta*-dimethylbenzene linker **D.** *para*-dimethylbenzene linker
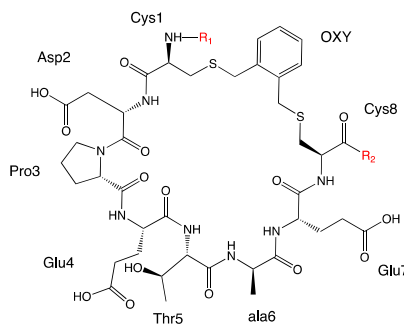
CP1 = pfl-CDEETGEC
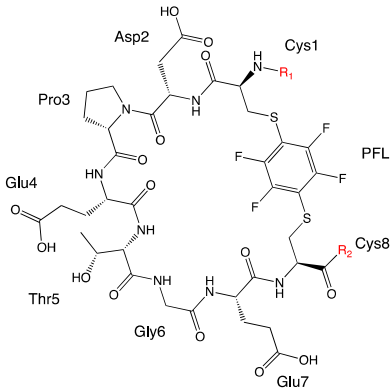
Glu3
Asp2
Cys1
R1
Glu4
Thr5
PFL
Gly6
Cys8
R2
Glu7

CP2 = pfl-CDPETGEC

Asp2
Cys1
Pro3
R1
Glu4
PFL
Thr5
Cys8
R2
Gly6
Glu7

CP3 = oxy-CDEETGEC

Asp2
Cys1
R1
Glu3
OXY
Glu4
Thr5
Cys8
R2
Gly6
Glu7

CP4 = oxy-CDPETGEC

Asp2
R1
Cys1
Pro3
Glu4
OXY
Thr5
Cys8
R2
Gly6
Glu7

CP5 = oxy-CDPETaEC

Cys1
R1
OXY
Asp2
Cys8
Pro3
R2
Glu4
Glu7
Thr5
ala6

CP6 = mxy-CDPETaEC

Cys1
NH—R1
Asp2
MXY
Pro3
Cys8
R2
Glu4
Glu7
Thr5
ala6

R1 =

acetylated cap   OR

biotin-Trp-Arg cap

R2 = —NH2   OR   —NH
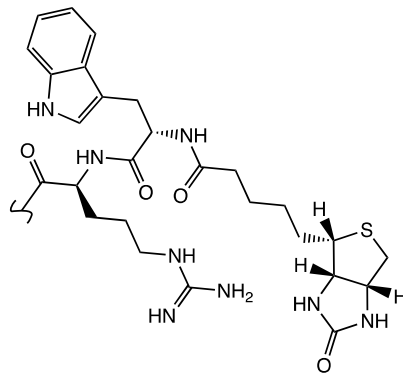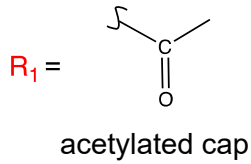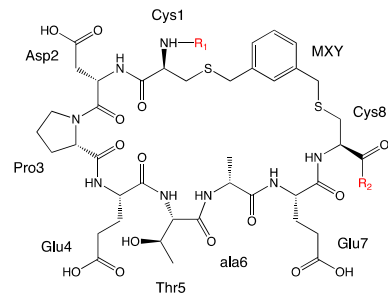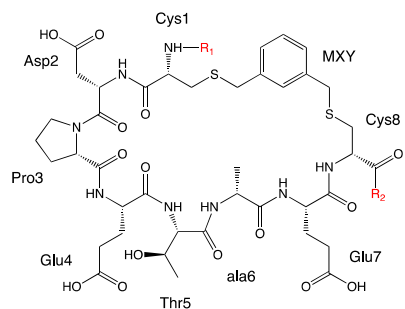
amino cap   N-methylated cap
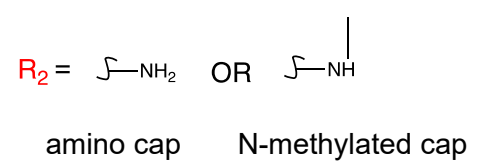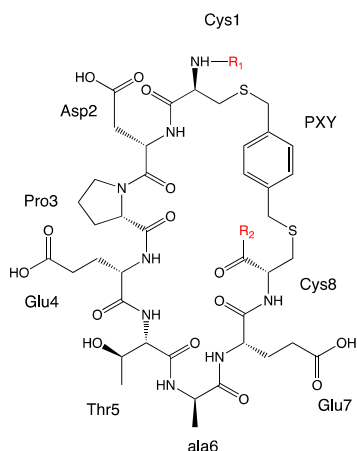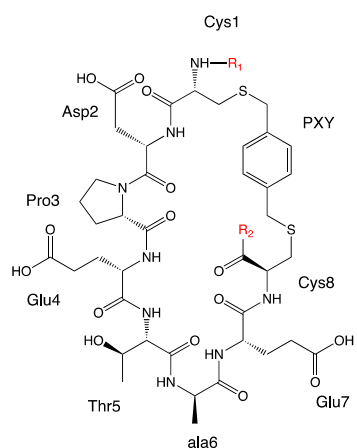
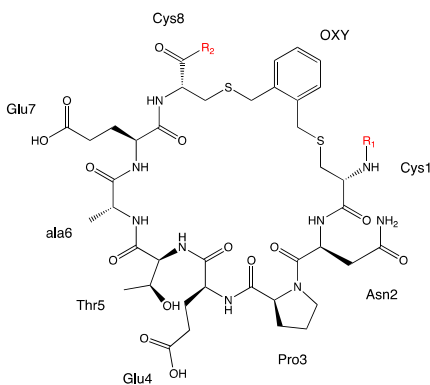CP7 = pxy-CDPETaEC   CP8 = oxy-cDPETaEc   CP9 = mxy-cDPETaEc
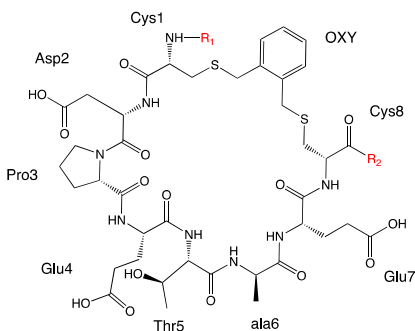

CP7 = pxy-cDPETaEc   CP8 = oxy-CNPETaEC

Figure S3. Chemical structures of cyclic peptides. Peptides were synthesized with R_1 = NHS-biotin-Trp-Arg cap at the N-terminus and R_2 = amino cap at the C-terminus. Peptides were simulated with R_1 = acetylated cap and R_2 = N-methylated cap. Synthetic peptides were prepared with the biotin-Trp-Arg cap at the N-terminus to allow for binding assay measurements (biotin), concentration measurements (Trp), and improved characterization via mass spectrometry (Arg). Ionizable groups are shown at neutral charge.
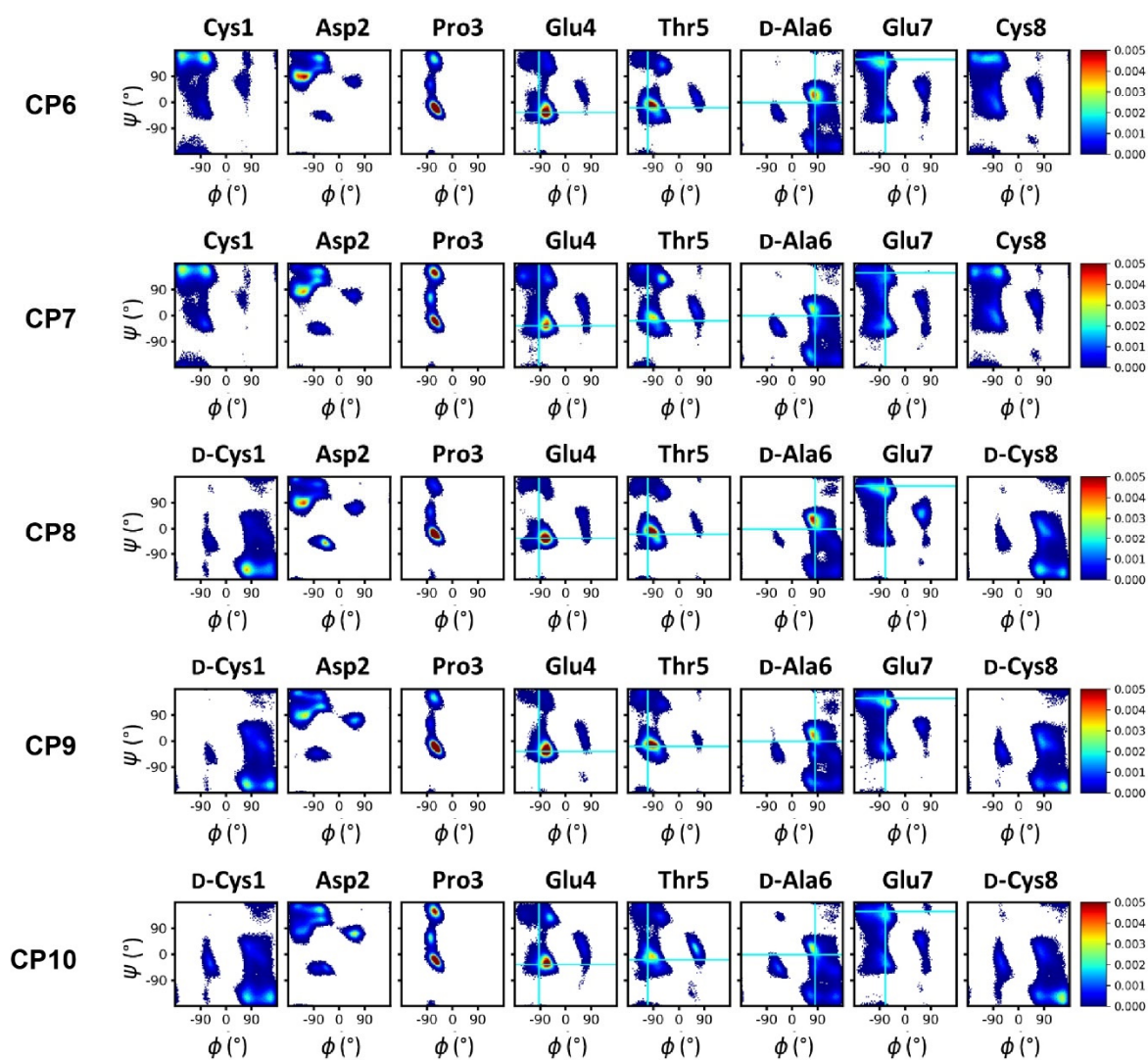
Figure S4. Ramachandran plots showing entire simulated ensembles for CP6–CP10. Cyan lines indicate the (ϕ, ψ) values in the desired conformation for the residues in the core binding sequence ETGE.

Table S1. Expected and observed masses of synthesized cyclic peptides.

CP name	Sequence	Expected Mass [M+H]	Observed Mass [M+H]
CP1	biotin-pfl-WRCDEETGEC	1599.4	1598.7
CP2	biotin-pfl-WRCDPETGEC	1567.4	1568.2
CP3	biotin-ortho-WRCDEETGEC	1555.4	1556.3
CP4	biotin-ortho-WRCDPETGEC	1523.4	1524.0
CP5	biotin-ortho-WRCDPETaEC	1537.4	1537.6
CP6	biotin-meta-WRCDPETaEC	1537.4	1537.7
CP7	biotin-para-WRCDPETaEC	1537.4	1538.2
CP8	biotin-ortho-WRcDPETaEc	1537.4	1537.4
CP9	biotin-meta-WRcDPETaEc	1537.4	1537.3
CP10	biotin-para-WRcDPETaEc	1537.4	1538.4
CP11	biotin-ortho-WRCNPETaEC	1536.4	1537.4

CP1, method 3
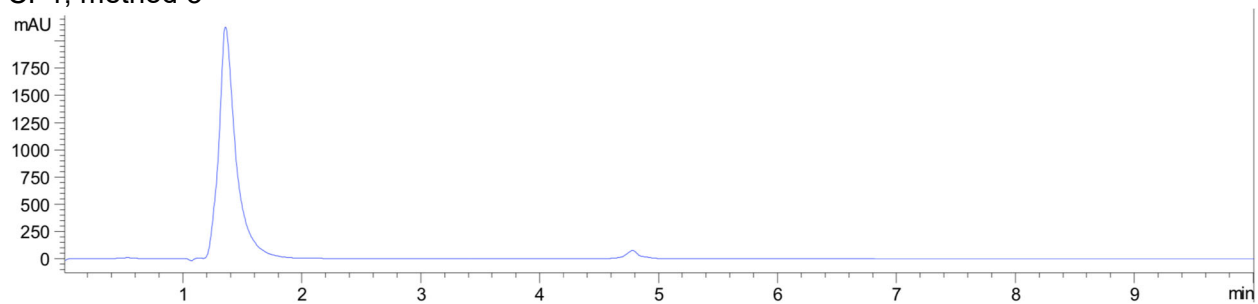
CP2, method 3
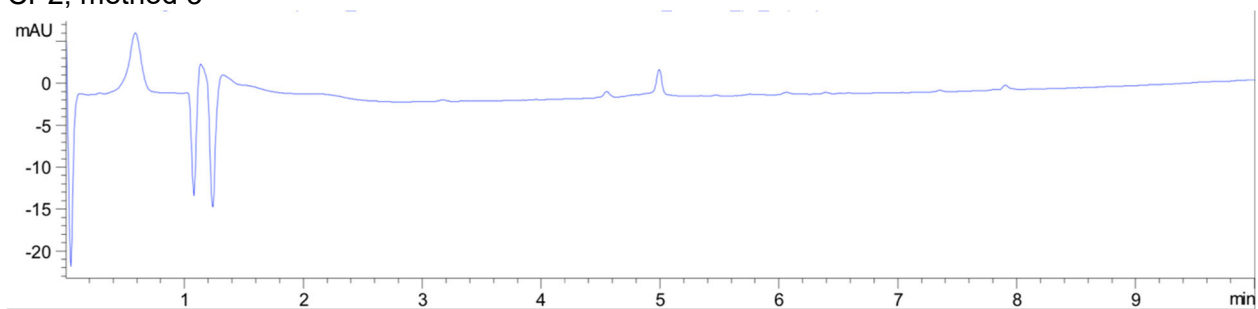
CP3, method 3
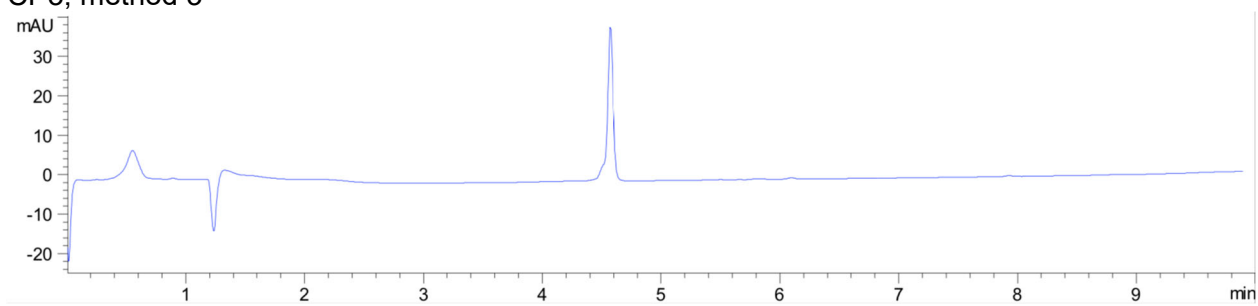
CP4, method 3
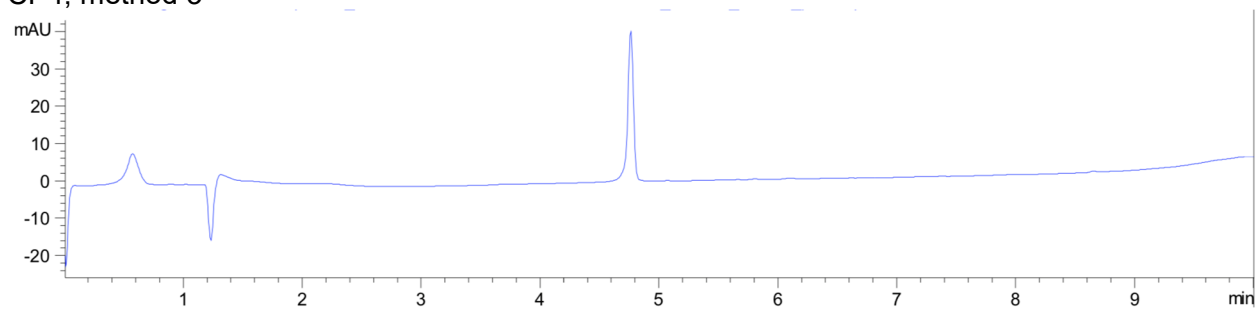
CP5, method 3
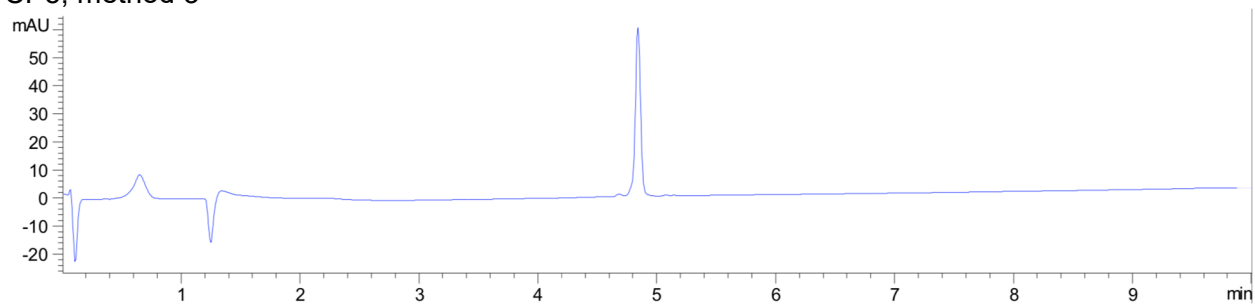
CP6, method 3
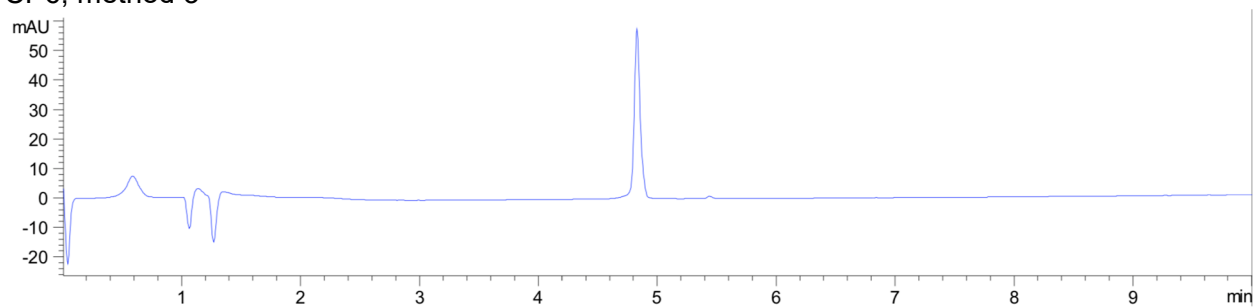
CP7, method 3

CP8, method 3

14
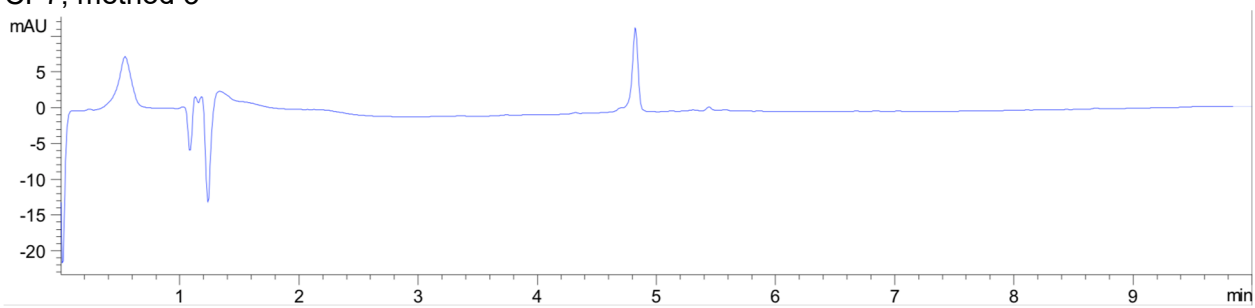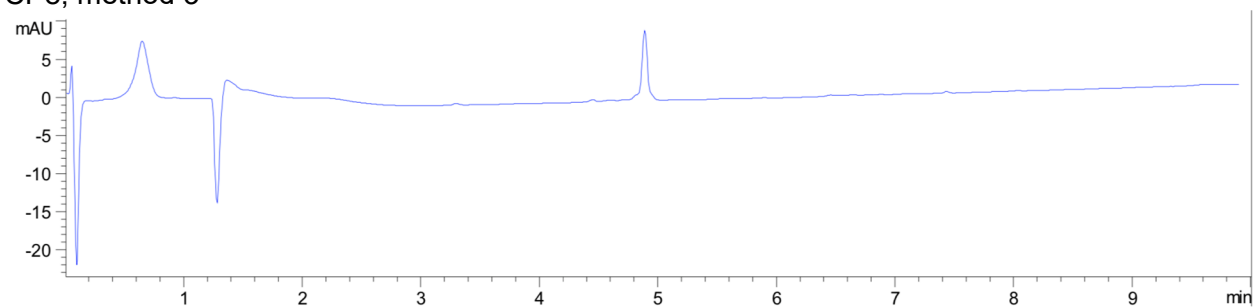
CP9, method 3
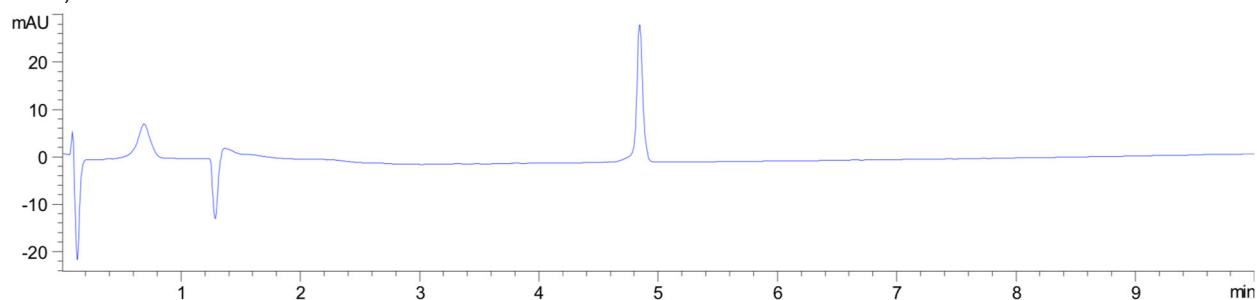
CP10, method 3
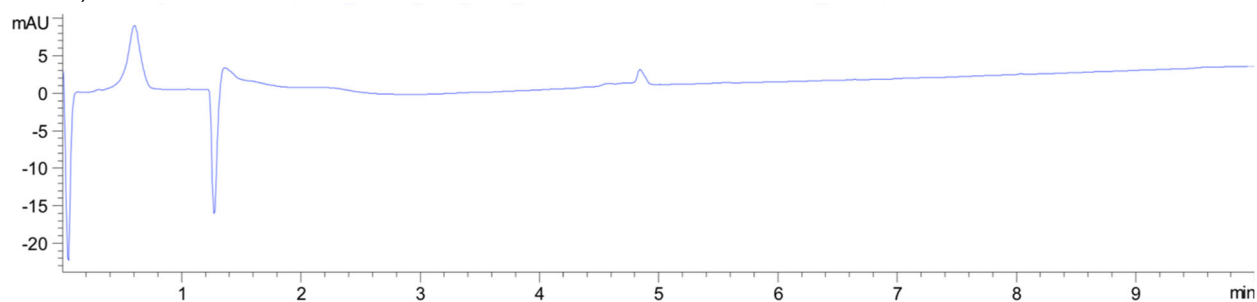
CP11, method 3
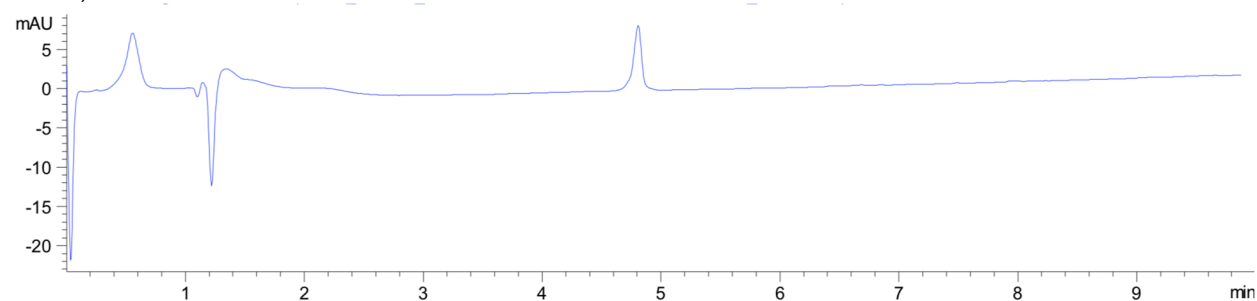
Figure S5. HPLC Chromatograms of biotinylated cyclic peptides. HPLC methods were as follows: **method 1**, 5µm, 4.6 x 250 mm Zorbax C18 column at 1 mL/min, 5–100% acetonitrile gradient over 20 minutes; **method 2**, 5µm, 4.6 x 250 mm Zorbax C18 column at 1 mL/min, 30–60% acetonitrile gradient over 20 minutes; **method 3**, 3.5 µm, 4.6 x 100 mm Eclipse Plus C18 column at 1 mL/min, 5-100% acetonitrile gradient over 10 minutes.

Table S2. Molecular dynamics simulation results for selected CPs and controls. MD results were reported with the standard error of the mean. **Column 3 (Pre-organized population (%))** shows the preoganized population measured as the percentage of time each CP occupies the desired ETGE conformation with RMSD below 0.5 Å. **Column 4 (Average RMSD for the pre-organized population (Å))** shows the average RMSD for for the pre-organized population. **Column 5 (Average number of clashes in pre-organized population)** indicates the average number of atoms in Keap1 that clashes when the pre-organized population of the CP is aligned with the ETGE conformation from Nrf2 in the binding pocket of Keap1.

CP name	Sequence	Pre-organized population (%)	Average RMSD for the pre-organized population (Å)	Average number of clashes in pre-organized population
CP3	ortho_CDEETGEC	43 ± 4	0.258 ± 0.003	5 ± 1.0
WR-3	ortho_WRCDEETGEC	21 ± 1	0.289 ± 0.006	26 ± 1.0
CP4	ortho_CDPETGEC	53 ± 5	0.242 ± 0.002	3 ± 0.3
WR-4	ortho_WRCDPETGEC	55 ± 4	0.229 ± 0.001	15 ± 4.0
CP11	ortho_CNPETaEC	87 ± 3	0.206 ± 0.001	2 ± 0.0
WR-11	ortho_WRCNPETaEC	79 ± 2	0.216 ± 0.001	7 ± 0.4

CP1

CP3

CP4

CP5

17
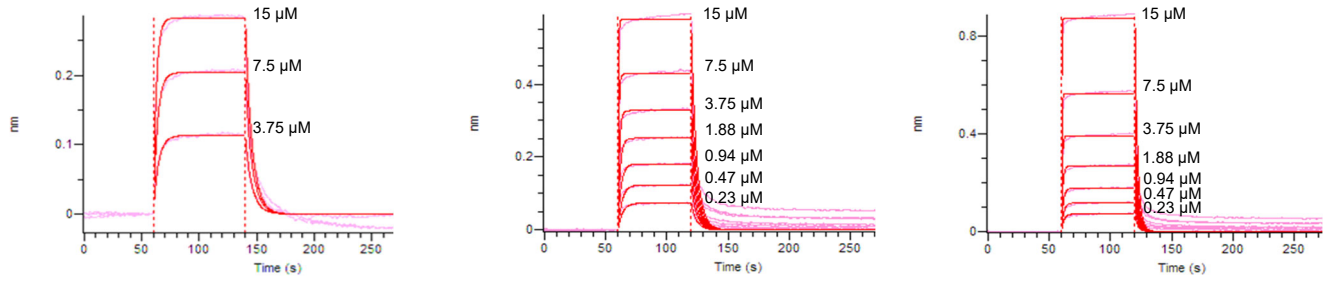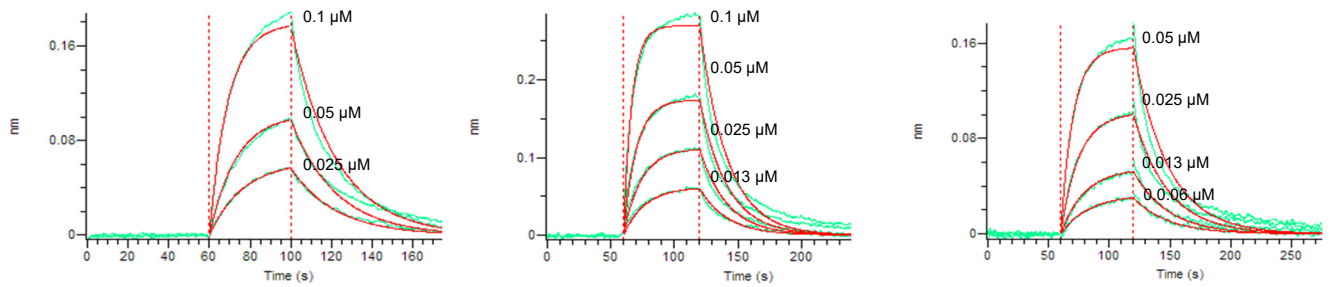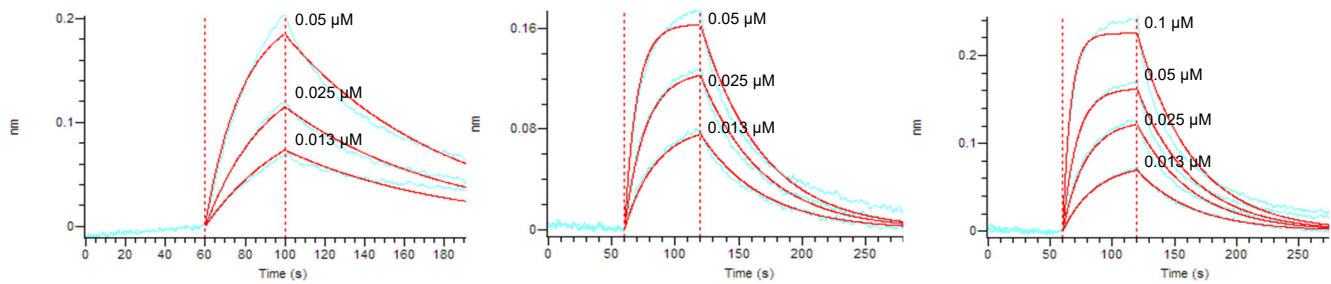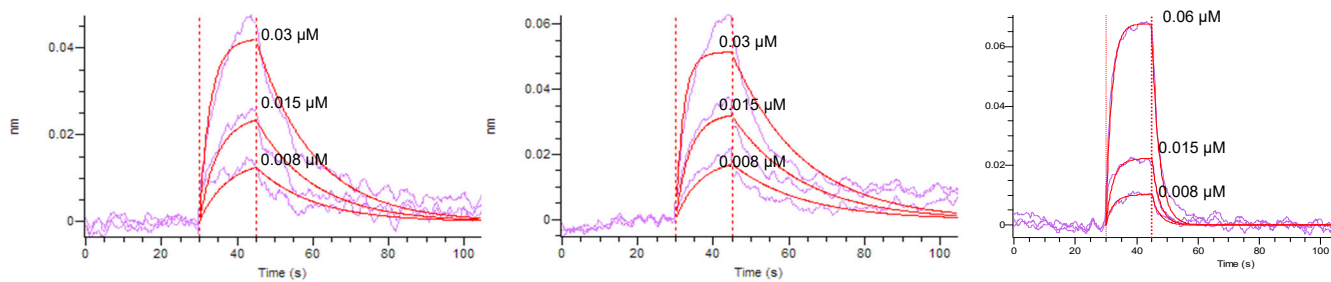
CP6
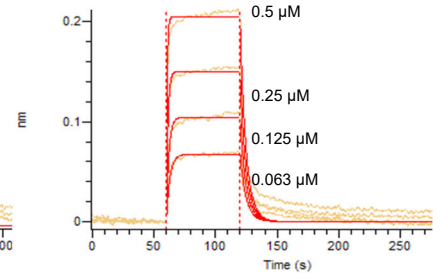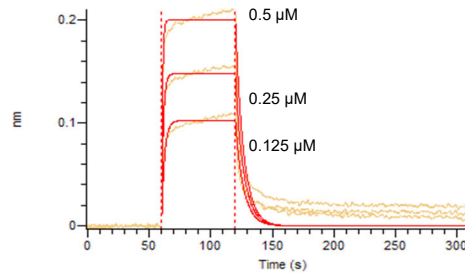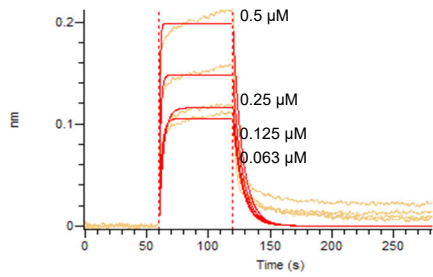
CP7
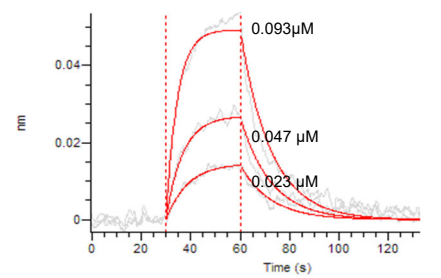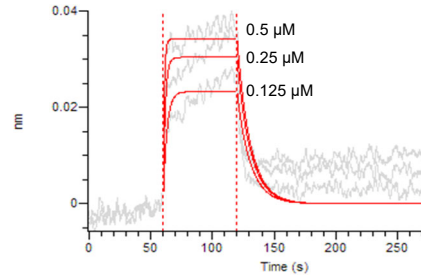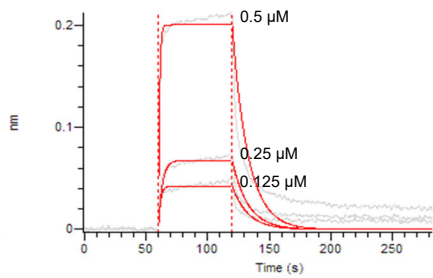
CP11
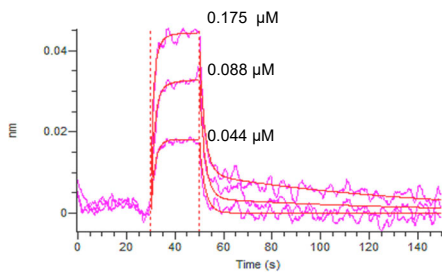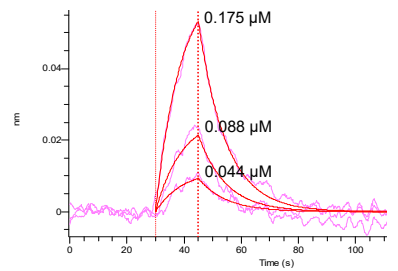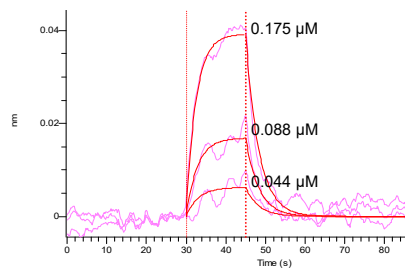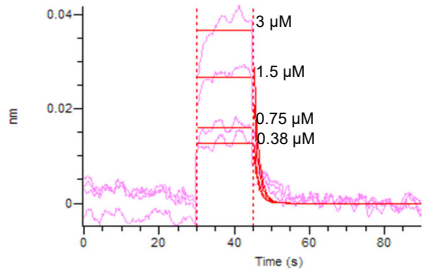
Figure S6. Complete biolayer interferometry data. Raw data (various colors) and curve fits (red) are shown for cyclic peptides binding to recombinantly expressed Keap1 Kelch domain. Three or four independent trials are shown, and each trial tests at least three different protein concentrations.

Table S3. Kinetics measurements. k_{off}, k_{on}, and K_d measurements determined via BLI experiments.

CP name	Sequence	k_{off} (s^{-1})	k_{on} (10^5 M^{-1} s^{-1})	K_d (nM)
CP1	PFL_WRCDEETGEC	0.277 ± 0.065	1.94 ± 0.37	1400 ± 82.2
CP2	PFL_WRCDPETGEC	Not meas.	Not meas.	Not meas.
CP3	ortho_WRCDEETGEC	0.0424 ± 0.003	9.71 ± 1.44	48.1 ± 10.1
CP4	ortho_WRCDPETGEC	0.0183 ± 0.003	12.4 ± 1.98	15.1 ± 1.09
CP5	ortho_WRCDPETaEC	0.0928 ± 0.027	105 ± 22.5	8.46 ± 0.61
CP6	meta_WRCDPETaEC	0.163 ± 0.018	22.6 ± 0.9	72.5 ± 9.17
CP7	para_WRCDPETaEC	0.0841 ± 0.004	17.6 ± 0.11	48.2 ± 2.99
CP8	ortho_WRcDPETaEc	Not obs.	Not obs.	Not obs.
CP9	meta_WRcDPETaEc	Not obs.	Not obs.	Not obs.
CP10	para_WRcDPETaEc	Not obs.	Not obs.	Not obs.
CP11	ortho_WRCNPETaEC	0.278 ± 0.206	3970 ± 319	4.7 ± 1.69

Table S4. Crystallography data collection and processing. Values in parenthesis refer to the outer resolution shell.

	CP3	CP4	CP5	CP11
	ortho-WRCDEETGEC	ortho-WRCDPETGEC	ortho-WRCDPETaEC	ortho-WRCNPETaEC
PDB code	8PKU	8PKV	8PKW	8PKX
Wavelength (Å)	1.000	1.000	1.000	1.000
Space group	C2	C2	C2	C2
a, b, c (Å)	161.96, 68.59, 77.72	162.95, 68.86, 77.95	162.68, 68.97, 77.97	161.58, 68.57, 77.84
α, β, γ (°)	90, 117.5, 90	90, 117.7, 90	90, 117.9, 90	90, 117.7, 90
Resolution range (Å)	50–1.73 (1.83–1.73)	50–1.55 (1.64–1.55)	50–1.54	50–1.78
Total No. of reflections	530063 (78745)	756510 (120235)	763245 (121058)	501113 (81084)
No. of unique reflections	78891 (12656)	107030 (16859)	112741 (18156)	72288 (11545)
Completeness (%)	99.7 (99.7)	96.4 (94.6)	99.8 (99.9)	99.6 (99.0)
CC1/2	99.9 (59.2)	100.0 (49.5)	99.9 (56.7)	99.9 (58.0)
⟨I/σ(I)⟩	16.08 (1.09)	18.77 (1.06)	16.86 (1.09)	16.53 (1.27)
Mosaicity (°)	0.15	0.07	0.08	0.17
R_{meas} (%)	6.0 (147.2)	5.0 (173.4)	5.3 (136.7)	6.9 (157.0)
Wilson B-factor (A^2)	41.3	36.0	35.412	39.92
Refinement statistics				
R_{work}/ R_{free} (%)	18.1/20.6	15.7/19.2	15.8/19.9	15.8/19.9
No. atoms	4933	5120	4957	4913
Protein A/B	2363/2137	2404/2170	2334/2124	2319/2164
Ion	26	36	19	11
Peptide	86	81	91	82
Water	259	383	369	289
Other	62	46	20	48
Average B factors (A^2)	57.8	44.3	47.9	52.7
Protein A/B (A^2)	42.0/74.5	34.3/52.9	36.08/59.9	40.1/65.5
Ion (A^2)	95.9	93.5	70.0	51.2
Peptide (A^2)	67.6	50.9	52.4	61.2
Water (A^2)	53.6	49.5	50.7	52.3
Other (A^2)	70.5	68.1	53.6	69.5
Ramachandran favored/allowed (%)	96.0/4.0	96.9/3.1	97.2/2.8	97.2/2.8
R.m.s.d. Bond lengths (Å)	0.006	0.005	0.005	0.005
R.m.s.d Angles (°)	0.83	0.78	0.78	0.78
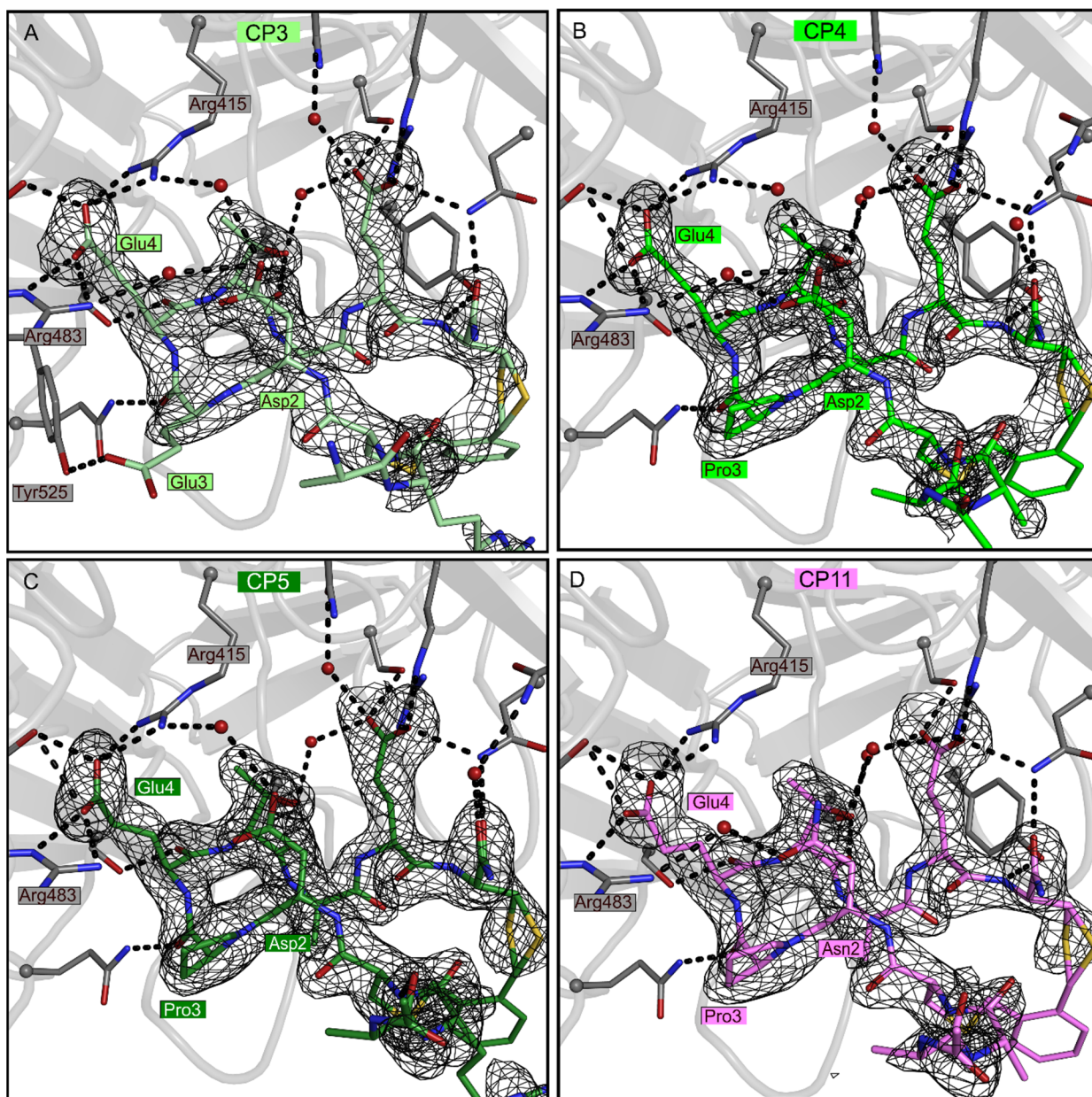
Figure S7. Crystal structures of cyclic peptides bound to the Keap1 Kelch domain, highlighting intermolecular hydrogen bond interactions. Simulated annealing omit maps are shown in grey, contoured at 3 σ. Asp2/Asn2 interacts with Arg483 and Arg415 via different bridging waters. The cysteines linked via the *ortho*-dimethylbenzene linker are well-defined in the electron density, while the linker itself is largely unresolved.
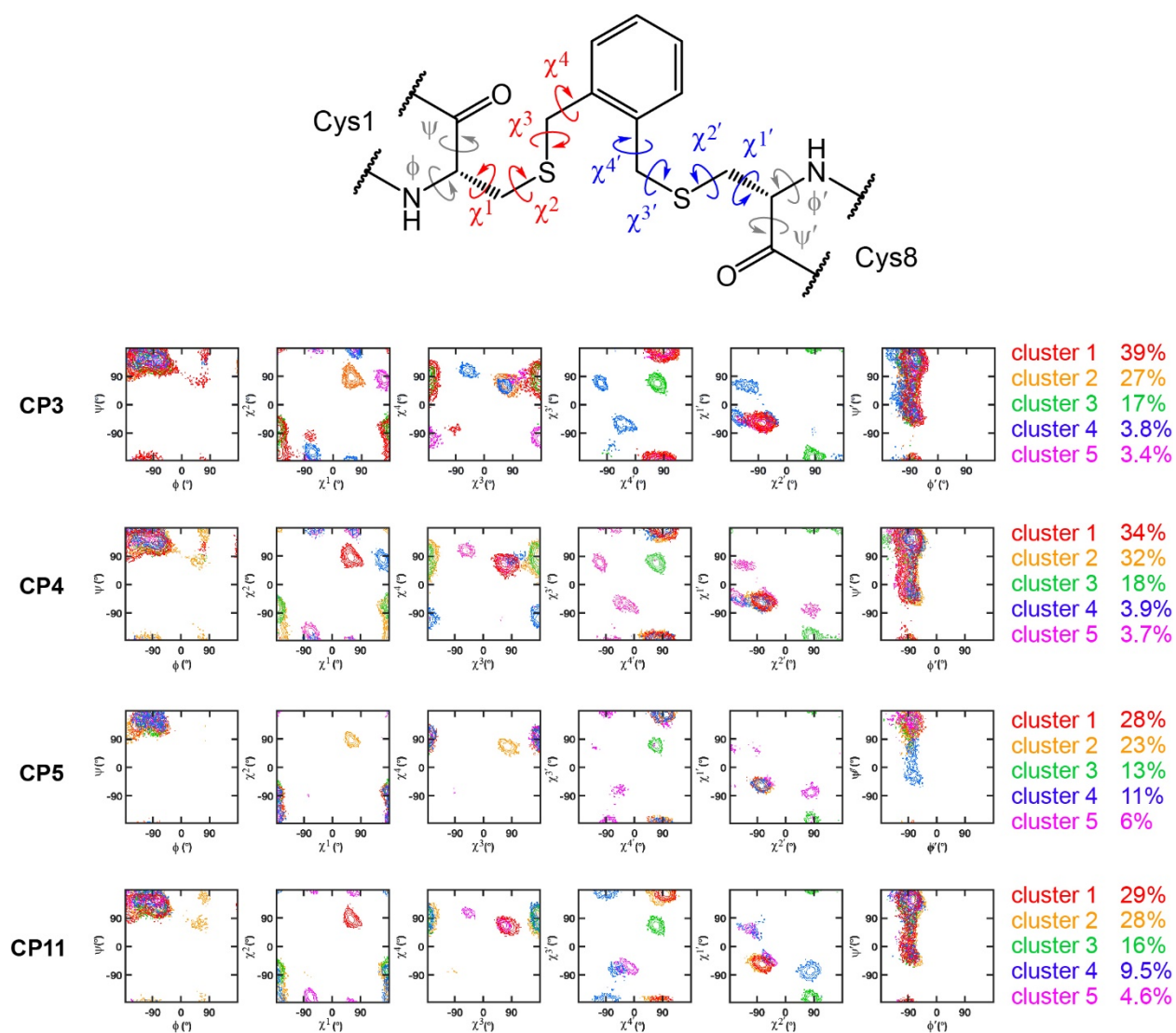
Figure S8. Linker geometries in CP simulations. The pre-organized populations of CP3, CP4, CP5, and CP11 were each clustered using the torsional angles in the linker region. This analysis revealed that all four had over 50% of the pre-organized population using one of two very similar linker geometries, but that alternate geometries for the linker were also tolerated. These predictions are consistent with the observation that the linker could assume multiple geometries when bound to Keap1.

Table S5. Comparing the Keap1-bound structures of CP3, CP4, CP5 and CP11 to the ensembles predicted by molecular dynamics. Pre-organized population is defined as the subset of the ensemble for which the ETGE backbone atoms have an RMSD less than 0.5 Å compared to the Nrf2-bound ETGE motif.

	CP3	CP4	CP5	CP11
pre-organized population (% of total ensemble)	43 ± 4	53 ± 5	62 ± 2	87 ± 3
median RMSD of ETGE motif between pre-organized population and Keap1-bound structure	0.246 ± 0.004	0.243 ± 0.002	0.231 ± 0.001	0.198 ± 0.001
median RMSD of entire peptide backbone between pre-organized population and Keap1-bound structure	0.376 ± 0.003	0.370 ± 0.001	0.665 ± 0.178	0.387 ± 0.003
percentage of frames from the pre-organized population with H-bond between Asp2/Asn2 amide and Glu7 carbonyl	30 ± 4	33 ± 5	41 ± 1	71 ± 2
percentage of frames from the pre-organized population with H-bond between the C-term carboxamide and the N-term acetyl group	30 ± 4	33 ± 5	41 ± 1	71 ± 2

References

(1) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111* (17), 4553–4559. https://doi.org/10.1021/jp067873l.

(2) Todorova, N.; Yarovsky, I. Molecular Modelling of Peptide Folding, Misfolding and Aggregation Phenomena. *Procedia Computer Science* **2010**, *1* (1), 1185–1193. https://doi.org/10.1016/j.procs.2010.04.132.

(3) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. Protein Folding and Ligand-Enzyme Binding from Bias-Exchange Metadynamics Simulations. *CPC* **2012**, *2* (1), 79–91. https://doi.org/10.2174/1877946811202010079.

(4) Ichiye, T.; Karplus, M. Collective Motions in Proteins: A Covariance Analysis of Atomic Fluctuations in Molecular Dynamics and Normal Mode Simulations. *Proteins* **1991**, *11* (3), 205–217. https://doi.org/10.1002/prot.340110305.

(5) Yu, H.; Lin, Y.-S. Toward Structure Prediction of Cyclic Peptides. *Phys. Chem. Chem. Phys.* **2015**, *17* (6), 4210–4219. https://doi.org/10.1039/C4CP04580G.

(6) Cossio, P.; Marinelli, F.; Laio, A.; Pietrucci, F. Optimizing the Performance of Bias-Exchange Metadynamics: Folding a 48-Residue LysM Domain Using a Coarse-Grained Model. *J. Phys. Chem. B* **2010**, *114* (9), 3259–3265. https://doi.org/10.1021/jp907464b.

(7) Damas, J. M.; Filipe, L. C. S.; Campos, S. R. R.; Lousa, D.; Victor, B. L.; Baptista, A. M.; Soares, C. M. Predicting the Thermodynamics and Kinetics of Helix Formation in a Cyclic Peptide Model. *Journal of Chemical Theory and Computation* **2013**, *9* (11), 5148–5157. https://doi.org/10.1021/ct400529k.

(8) Miao, J.; Descoteaux, M. L.; Lin, Y.-S. Structure Prediction of Cyclic Peptides by Molecular Dynamics + Machine Learning. *Chem. Sci.* **2021**, *12* (44), 14927–14936. https://doi.org/10.1039/D1SC05562C.

(9) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4* (3), 435–447. https://doi.org/10.1021/ct700301q.

(10) Zhou, C.-Y.; Jiang, F.; Wu, Y.-D. Residue-Specific Force Field Based on Protein Coil Library. RSFF2: Modification of AMBER Ff99SB. *J. Phys. Chem. B* **2015**, *119* (3), 1035–1047. https://doi.org/10.1021/jp5064676.

(11) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(12) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(13) D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. Amber2022, 2022.

(14) Sittel, F.; Jain, A.; Stock, G. Principal Component Analysis of Molecular Dynamics: On the Use of Cartesian vs. Internal Coordinates. *The Journal of Chemical Physics* **2014**, *141* (1), 014111. https://doi.org/10.1063/1.4885338.

(15) Mu, Y.; Nguyen, P. H.; Stock, G. Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins* **2004**, *58* (1), 45–52. https://doi.org/10.1002/prot.20310.

(16) Rodriguez, A.; Laio, A. Clustering by Fast Search and Find of Density Peaks. *Science* **2014**, *344* (6191), 1492–1496. https://doi.org/10.1126/science.1242072.

(17) Lo, S.-C.; Li, X.; Henzl, M. T.; Beamer, L. J.; Hannink, M. Structure of the Keap1:Nrf2 Interface Provides Mechanistic Insight into Nrf2 Signaling. *EMBO J* **2006**, *25* (15), 3605–3617. https://doi.org/10.1038/sj.emboj.7601243.

(18) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(19) Steel, R. J.; O'Connell, M. A.; Searcey, M. Perfluoroarene-Based Peptide Macrocycles That Inhibit the Nrf2/Keap1 Interaction. *Bioorganic & Medicinal Chemistry Letters* **2018**, *28* (16), 2728–2731. https://doi.org/10.1016/j.bmcl.2018.03.003.

(20) Zhong, M.; Lynch, A.; Muellers, S. N.; Jehle, S.; Luo, L.; Hall, D. R.; Iwase, R.; Carolan, J. P.; Egbert, M.; Wakefield, A.; Streu, K.; Harvey, C. M.; Ortet, P. C.; Kozakov, D.; Vajda, S.; Allen, K. N.; Whitty, A. Interaction Energetics and Druggability of the Protein–Protein Interaction between Kelch-like ECH-Associated Protein 1 (KEAP1) and Nuclear Factor Erythroid 2 Like 2 (Nrf2). *Biochemistry* **2020**, *59* (4), 563–581. https://doi.org/10.1021/acs.biochem.9b00943.

(21) Kabsch, W. XDS. *Acta Cryst D* **2010**, *66* (2), 125–132. https://doi.org/10.1107/S0907444909047337.

(22) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010**, *66* (Pt 4), 486–501. https://doi.org/10.1107/S0907444910007493.

(23) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr D Struct Biol* **2019**, *75* (Pt 10), 861–877. https://doi.org/10.1107/S2059798319011471.

(24) Long, F.; Nicholls, R. A.; Emsley, P.; Gražulis, S.; Merkys, A.; Vaitkus, A.; Murshudov, G. N. AceDRG: A Stereochemical Description Generator for Ligands. *Acta Crystallogr D Struct Biol* **2017**, *73* (Pt 2), 112–122. https://doi.org/10.1107/S2059798317000067.

(25) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox,. Gaussian 09, Revision A.02, 2016.