Supplementary Information

Cyclic peptide structure prediction and design using AlphaFold2

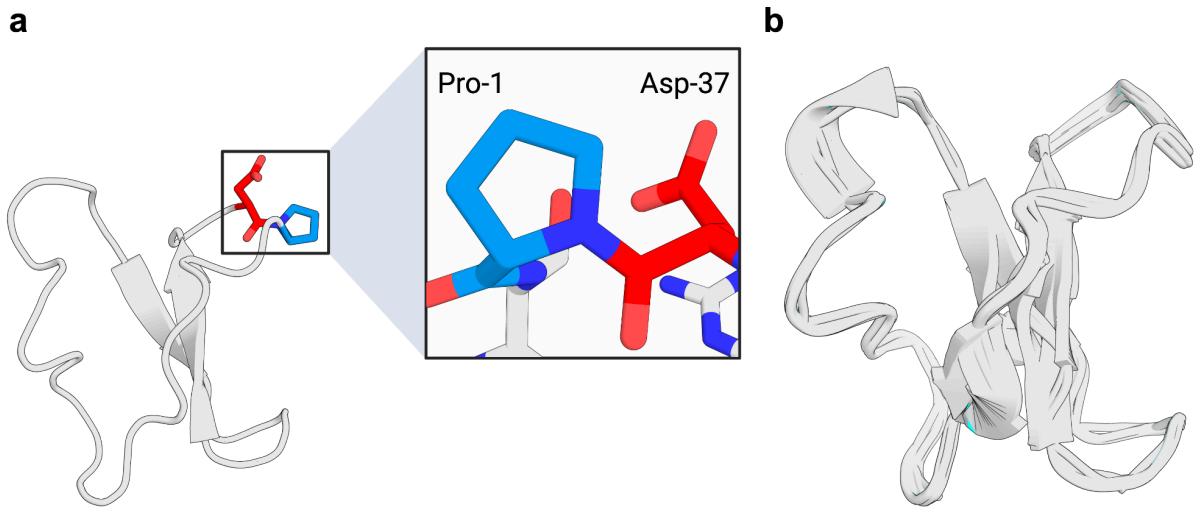
Stephen A. Rettie[1,2], Katelyn V. Campbell[2,3], Asim K. Bera[2], Alex Kang[2], Simon Kozlov[4], Yensi Flores Bueso[2,3,5,6], Joshmyn De La Cruz[2], Maggie Ahlrichs[2], Suna Cheng[2], Stacey R. Gerben[2], Mila Lamb[2], Analisa Murray[2], Victor Adebomi[2,4], Guangfeng Zhou[2,3], Frank DiMaio[2,3], Sergey Ovchinnikov[4,*], Gaurav Bhardwaj[1,2,5,*]

AFFILIATIONS

[1] Molecular and Cell Biology program, University of Washington, Seattle, WA, USA
[2] Institute for Protein Design, University of Washington, Seattle, WA, USA
[3] Department of Biochemistry, University of Washington, Seattle, WA, USA
[4] Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA
[5] Department of Medicinal Chemistry, University of Washington, Seattle, WA, USA
[6] Cancer Research @UCC, University College Cork, Cork, Ireland

*Corresponding authors: Sergey Ovchinnikov (so3@mit.edu), Gaurav Bhardwaj (gauravb@uw.edu)

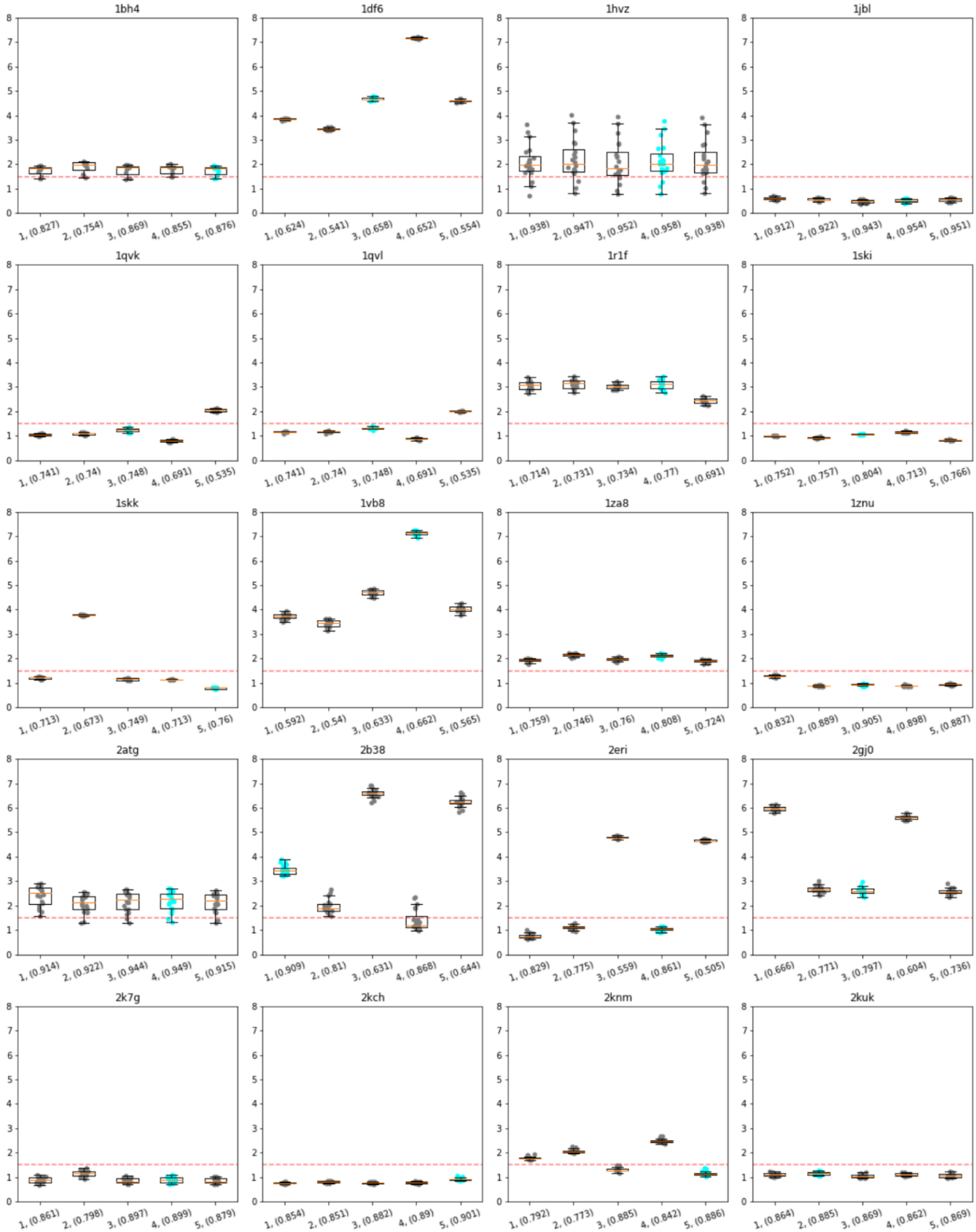
Supplementary Figure 1

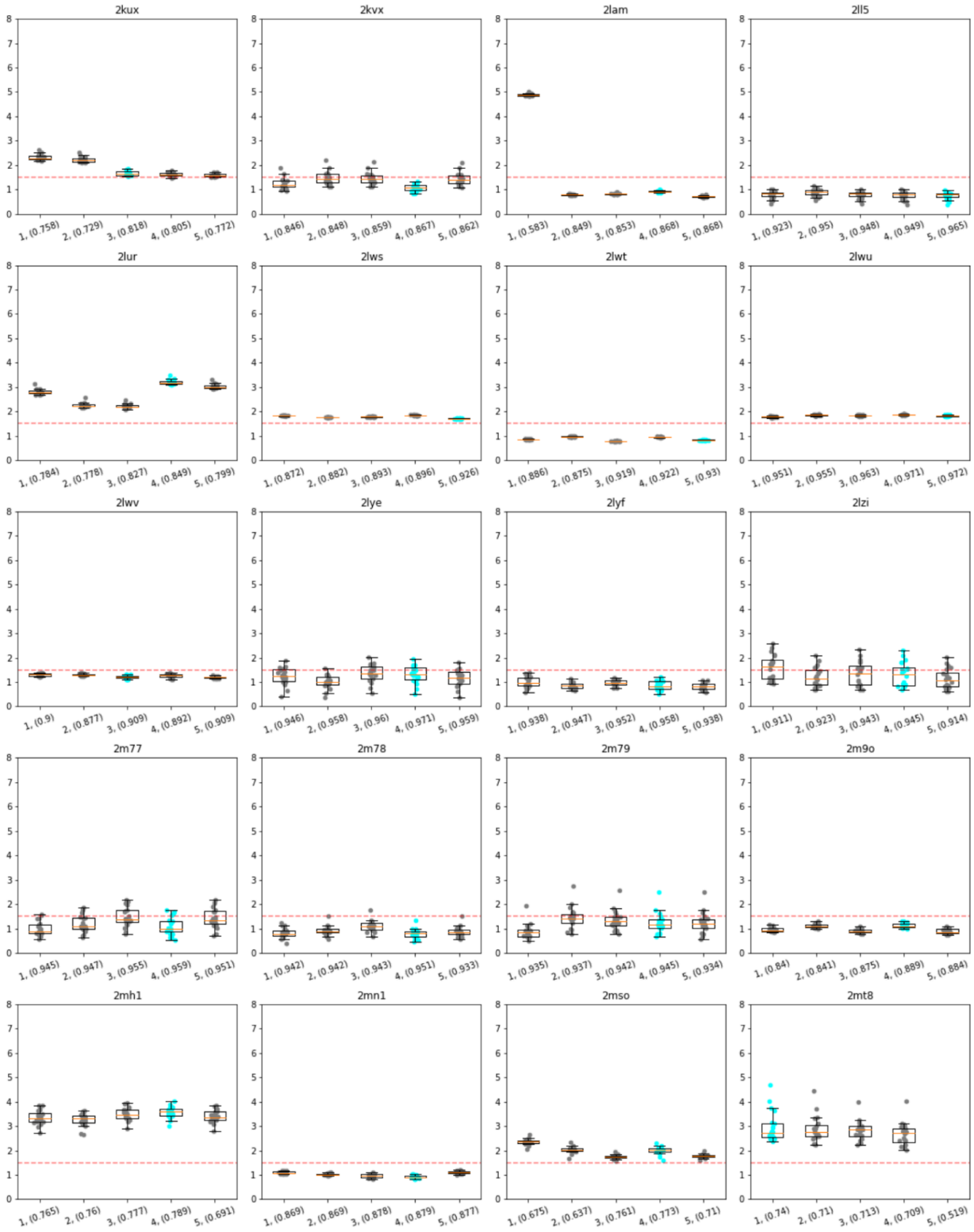


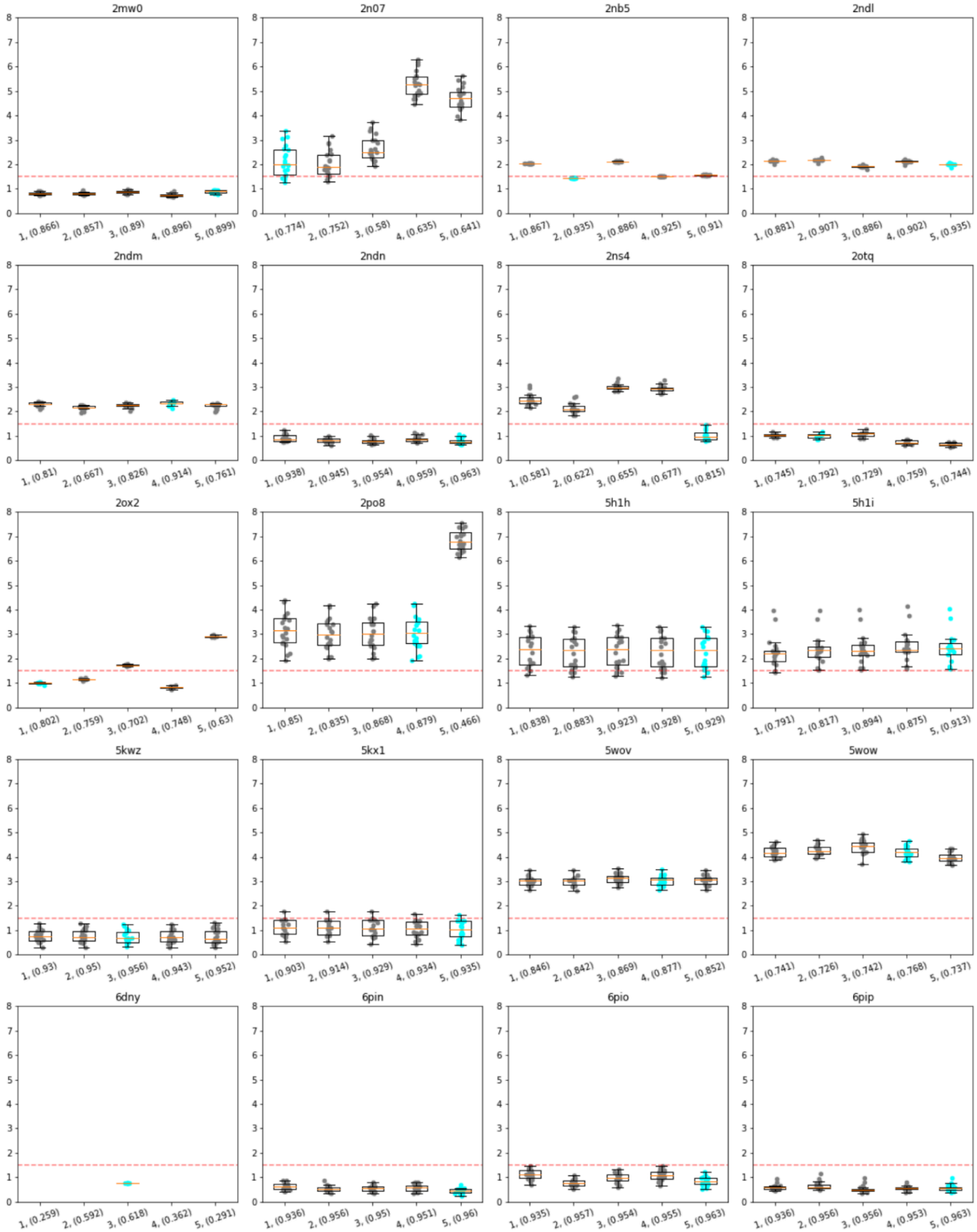
Supplementary Figure 1: Cyclic offset to relative positional encoding in AfCycDesign enforces N-to-C terminal cyclization and is invariant to circular permutations of the sequence

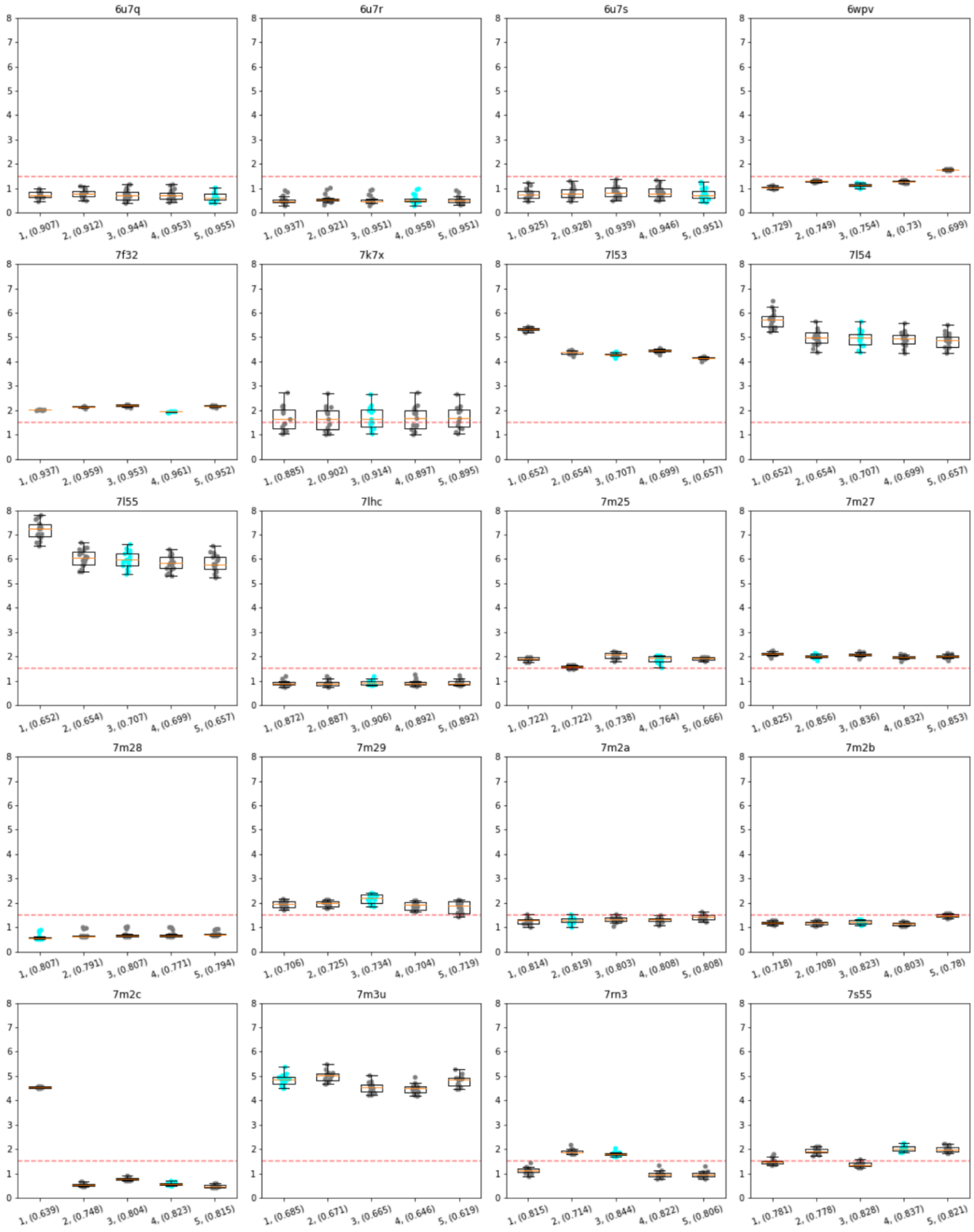
(A) Peptide structures predicted with AfCycDesign show correct bond connectivity and geometry at the termini. (B) Cartoon overlay of 37 predicted models for each circularly permuted sequence of a native peptide with a cyclic offset applied in AfCycDesign. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 2




2kux
1, (0.758)
2, (0.729)
3, (0.818)
4, (0.805)
5, (0.772)
2kvx
1, (0.846)
2, (0.848)
3, (0.859)
4, (0.867)
5, (0.862)
2lam
1, (0.583)
2, (0.849)
3, (0.853)
4, (0.868)
5, (0.868)
2ll5
1, (0.923)
2, (0.95)
3, (0.948)
4, (0.949)
5, (0.965)
2lur
1, (0.784)
2, (0.778)
3, (0.827)
4, (0.849)
5, (0.799)
2lws
1, (0.872)
2, (0.882)
3, (0.893)
4, (0.896)
5, (0.926)
2lwt
1, (0.886)
2, (0.875)
3, (0.919)
4, (0.922)
5, (0.93)
2lwu
1, (0.951)
2, (0.955)
3, (0.963)
4, (0.971)
5, (0.972)
2lwv
1, (0.9)
2, (0.877)
3, (0.909)
4, (0.892)
5, (0.909)
2lye
1, (0.946)
2, (0.958)
3, (0.96)
4, (0.971)
5, (0.959)
2lyf
1, (0.938)
2, (0.947)
3, (0.952)
4, (0.958)
5, (0.938)
2lzi
1, (0.911)
2, (0.923)
3, (0.943)
4, (0.945)
5, (0.914)
2m77
1, (0.945)
2, (0.947)
3, (0.955)
4, (0.959)
5, (0.951)
2m78
1, (0.942)
2, (0.942)
3, (0.943)
4, (0.951)
5, (0.933)
2m79
1, (0.935)
2, (0.937)
3, (0.942)
4, (0.945)
5, (0.934)
2m9o
1, (0.84)
2, (0.841)
3, (0.875)
4, (0.889)
5, (0.884)
2mh1
1, (0.765)
2, (0.76)
3, (0.777)
4, (0.789)
5, (0.691)
2mn1
1, (0.869)
2, (0.869)
3, (0.878)
4, (0.879)
5, (0.877)
2mso
1, (0.675)
2, (0.637)
3, (0.761)
4, (0.773)
5, (0.71)
2mt8
1, (0.74)
2, (0.71)
3, (0.713)
4, (0.709)
5, (0.519)

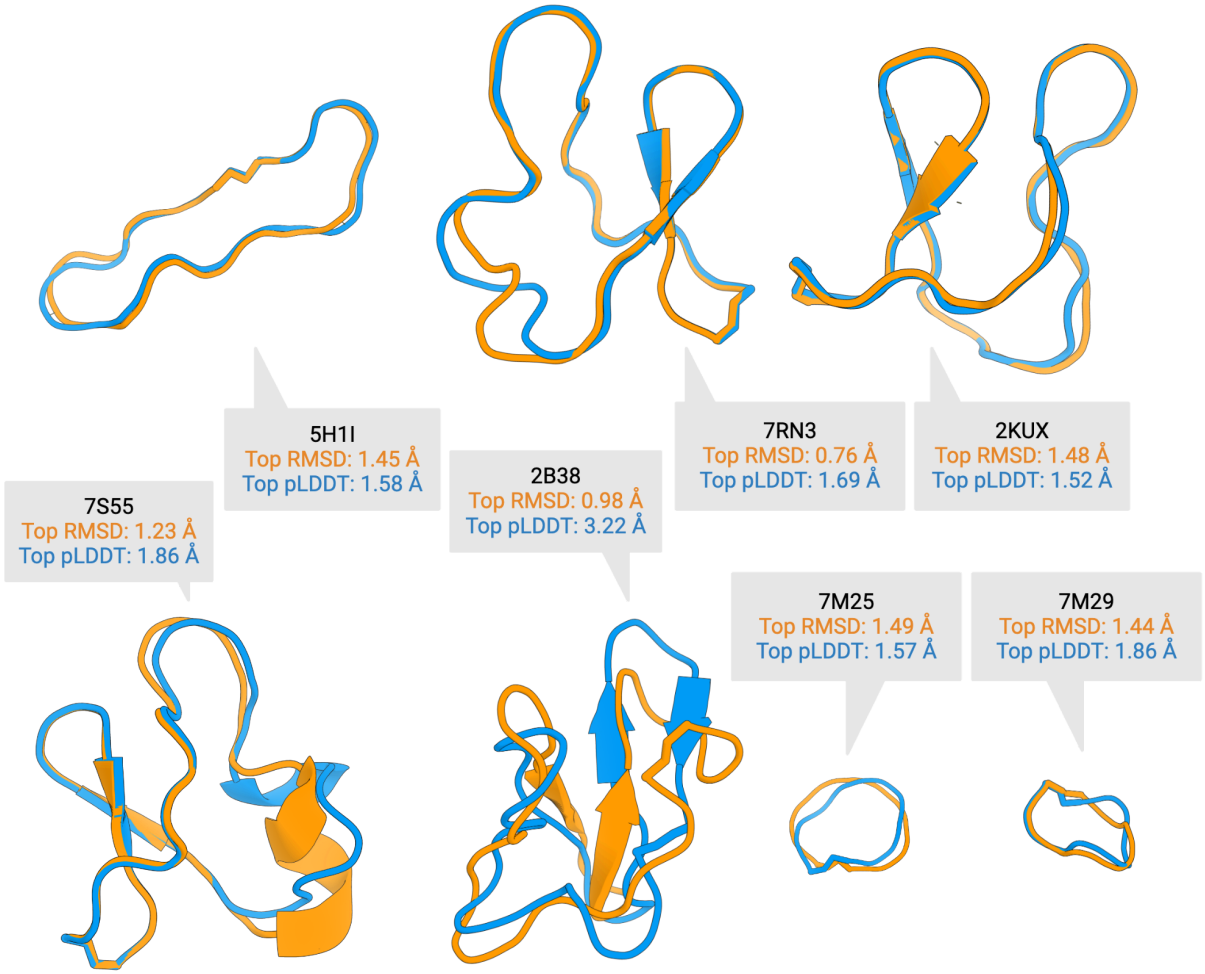

2mw0
1, (0.866)
2, (0.857)
3, (0.89)
4, (0.896)
5, (0.899)
2n07
1, (0.774)
2, (0.752)
3, (0.58)
4, (0.635)
5, (0.641)
2nb5
1, (0.867)
2, (0.935)
3, (0.886)
4, (0.925)
5, (0.91)
2ndl
1, (0.881)
2, (0.907)
3, (0.886)
4, (0.902)
5, (0.935)
2ndm
1, (0.81)
2, (0.667)
3, (0.826)
4, (0.914)
5, (0.761)
2ndn
1, (0.938)
2, (0.945)
3, (0.954)
4, (0.959)
5, (0.963)
2ns4
1, (0.581)
2, (0.622)
3, (0.655)
4, (0.677)
5, (0.815)
2otq
1, (0.745)
2, (0.792)
3, (0.729)
4, (0.759)
5, (0.744)
2ox2
1, (0.802)
2, (0.759)
3, (0.702)
4, (0.748)
5, (0.63)
2po8
1, (0.85)
2, (0.835)
3, (0.868)
4, (0.879)
5, (0.466)
5h1h
1, (0.838)
2, (0.883)
3, (0.923)
4, (0.928)
5, (0.929)
5h1i
1, (0.791)
2, (0.817)
3, (0.894)
4, (0.875)
5, (0.913)
5kwz
1, (0.93)
2, (0.95)
3, (0.956)
4, (0.943)
5, (0.952)
5kx1
1, (0.903)
2, (0.914)
3, (0.929)
4, (0.934)
5, (0.935)
5wov
1, (0.846)
2, (0.842)
3, (0.869)
4, (0.877)
5, (0.852)
5wow
1, (0.741)
2, (0.726)
3, (0.742)
4, (0.768)
5, (0.737)
6dny
1, (0.259)
2, (0.592)
3, (0.618)
4, (0.362)
5, (0.291)
6pin
1, (0.936)
2, (0.956)
3, (0.95)
4, (0.951)
5, (0.96)
6pio
1, (0.935)
2, (0.957)
3, (0.954)
4, (0.955)
5, (0.963)
6pip
1, (0.936)
2, (0.956)
3, (0.956)
4, (0.953)
5, (0.963)



Supplementary Figure 2: All vs. all comparison of NMR ensemble members and predicted models generated by AfCycDesign

Boxplots for backbone heavy atom RMSD between 5 models predicted by AfCycDesign and all members of deposited NMR ensembles (individual states as dots). Y-axis is backbone heavy atom RMSD (Å), and X-axis denotes the model number and pLDDT. Cyan coloring denotes the highest pLDDT model and the red dashed line shows the 1.5 Å cutoff. For structure 6DNY, four models are distorted and RMSD is much greater than 8 Å. The box region contains Q1 to Q3 of the interquartile range with the bisecting line showing the median. The whiskers are to the farthest point within 1.5 times the edge of the box. All data points are shown.

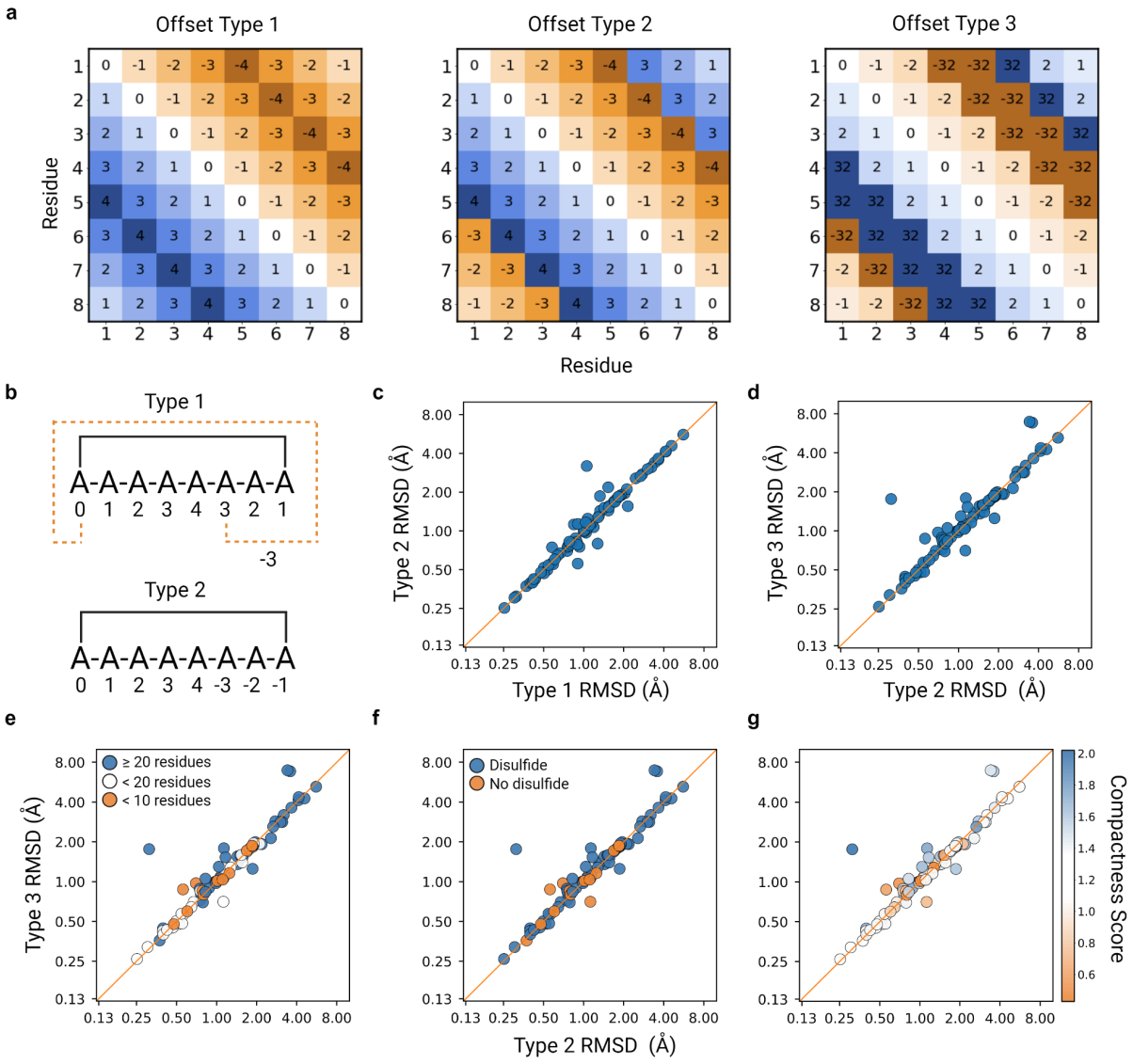
Supplementary Figure 3:



Supplementary Figure 3: Benchmark cases where an accurate prediction could have been obtained using a AfCycDesign model other than the highest pLDDT

Aligned structures of the benchmark cases where a model different from the highest pLDDT leads to a prediction with RMSD < 1.5 Å. Orange models are those closest to the deposited structure, blue is the model with the highest pLDDT. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 4:

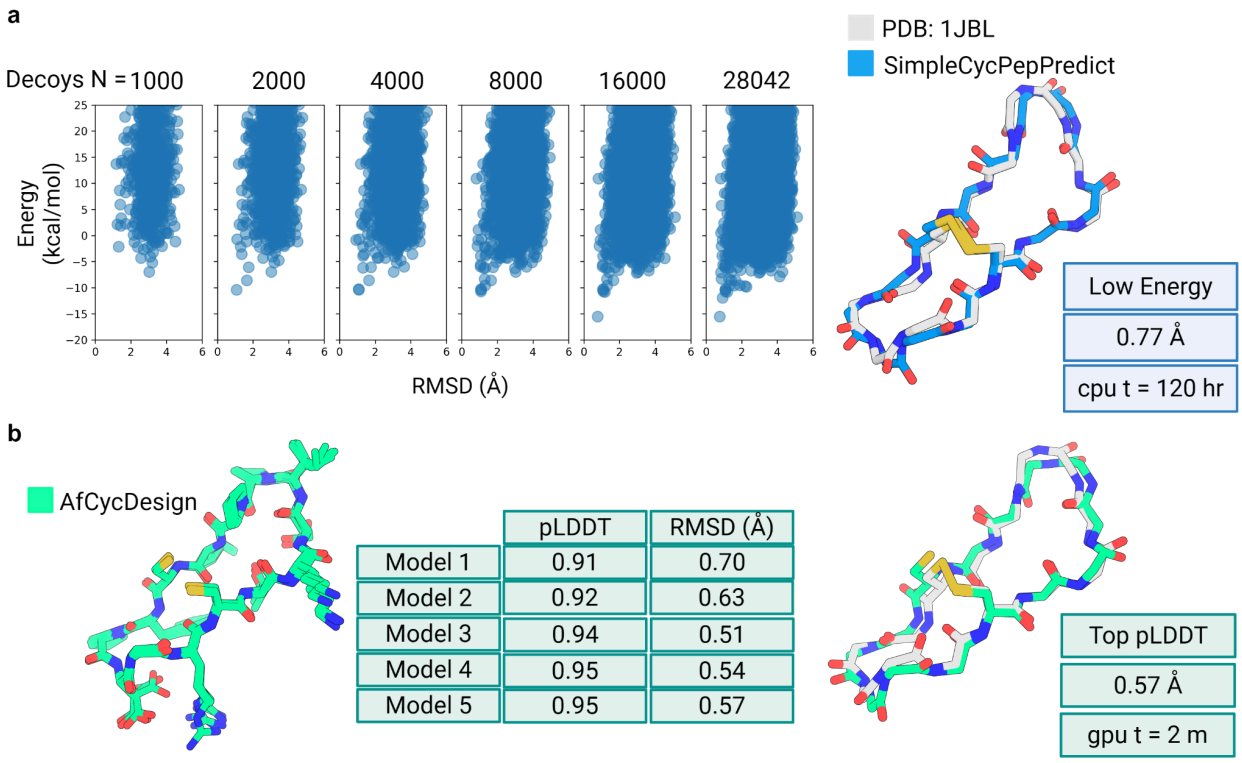


Supplementary Figure 4:: Long-range information in relative positional encoding does not significantly impact prediction accuracy

(A) Offset Type 1 used in the design sections of this work, for a hypothetical 8 residue peptide Visual representation of offset Type 1 relative positional encoding for an 8 residue poly-alanine peptide. Orange dashed line highlights a position that is incorrectly labeled with this offset. Offset Type 2 is the accurate labeling for this peptide. (C) Scatter plot shows offset Type 1 prediction accuracy of 80 PDB structures compared to offset Type 2. (D) Scatter plot shows offset Type 2 prediction accuracy of 80 PDB structures compared to offset Type 3. (E) The same scatter plot as in (D) but colored by binned peptide lengths. (F) The same scatter plot as in (D) but colored by whether or not the peptide has a disulfide. (G) The same scatter plot as in (D) but colored by compactness score. Compactness score was defined as the number of residue pairs within 6 Å Cα–Cα, not including direct neighbours, and normalized by number of residues. The median compactness score for all members of each deposited NMR ensemble is reported. The

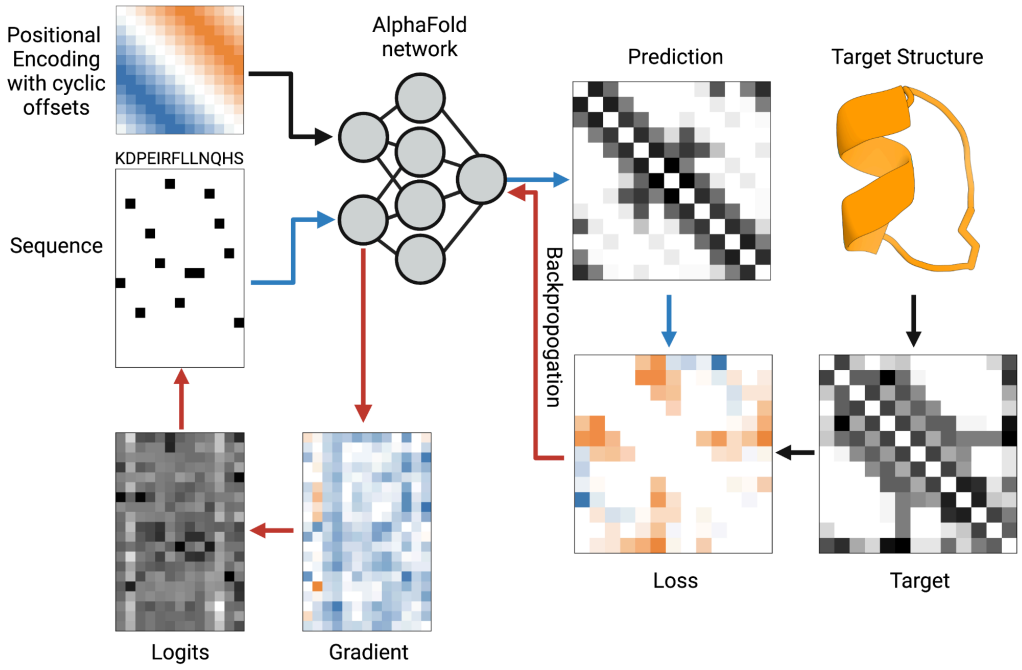
outlier with high compactness for which the Type 3 offset negatively impacts RMSD is PDB ID: 5KWZ. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 5:



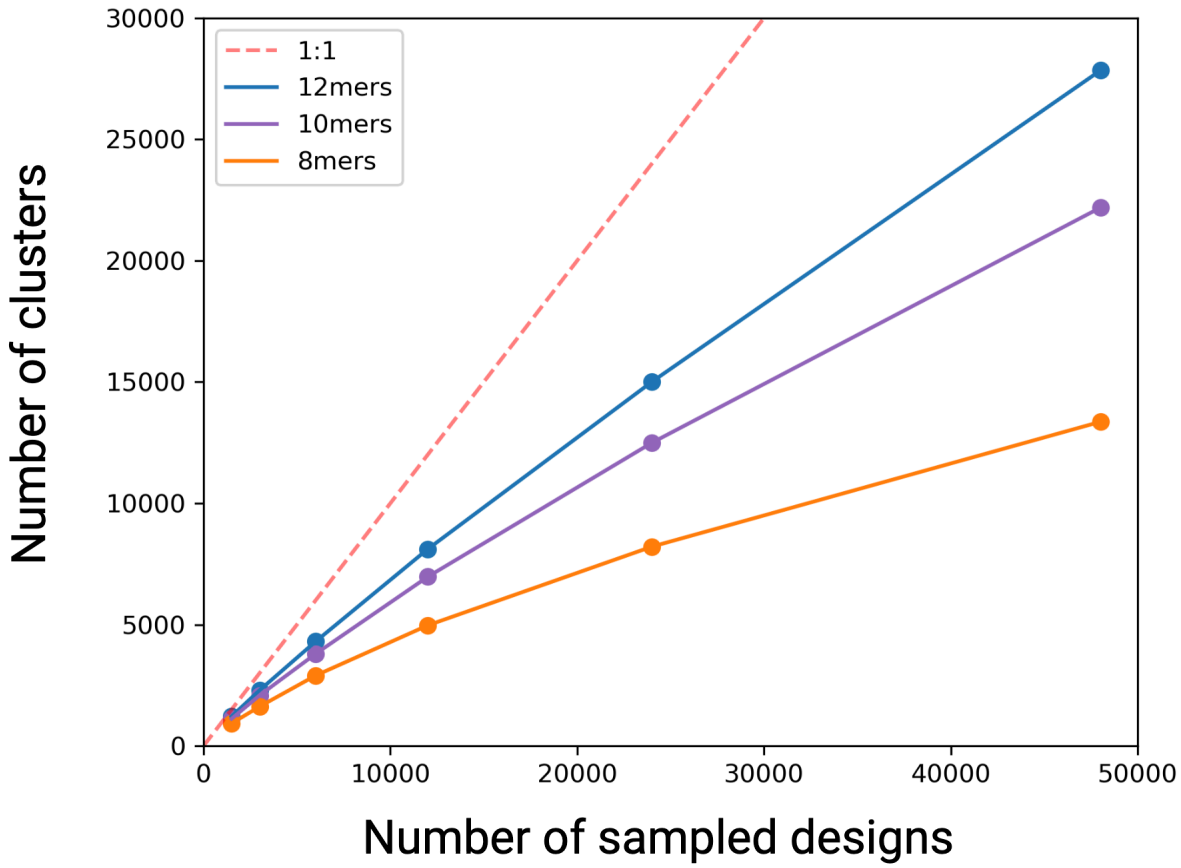
Supplementary Figure 5: Comparison of AfCycDesign and simple_cycpep_predict for a structure that both successfully predict.
(A) Energy landscape plots over the course of the SimpleCycPepPredesict run for PDB ID: 1JBL using 10 CPUs for 12 hours each, totalling 28,042 structures. The lowest energy structure from the final plot is shown aligned to the deposited PDB with side chains hidden for visual clarity. (B) Predicted model for the same peptide using AfCycDesign. pLDDT and RMSD for all five models. The top pLDDT model is shown aligned to the deposited PDB. Models were generated on a single GPU in 2 minutes. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 6:



Supplementary Figure 6: Overview of the sequence design approach in AfCycDesign
Given the input sequence and relative positional encoding with cyclic offset, AlphaFold2 is used to predict the distogram (distribution of distances between every pair of positions). The loss is defined as the categorical-cross entropy (CCE) between the output distogram and the desired distogram extracted from the desired structure. At each iteration, the gradient is computed to minimize the CCE. The gradient is then applied to a proxy variable we term "logits". The one-hot encoding of the argmax of the logits is then used as the input sequence for the next iteration. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

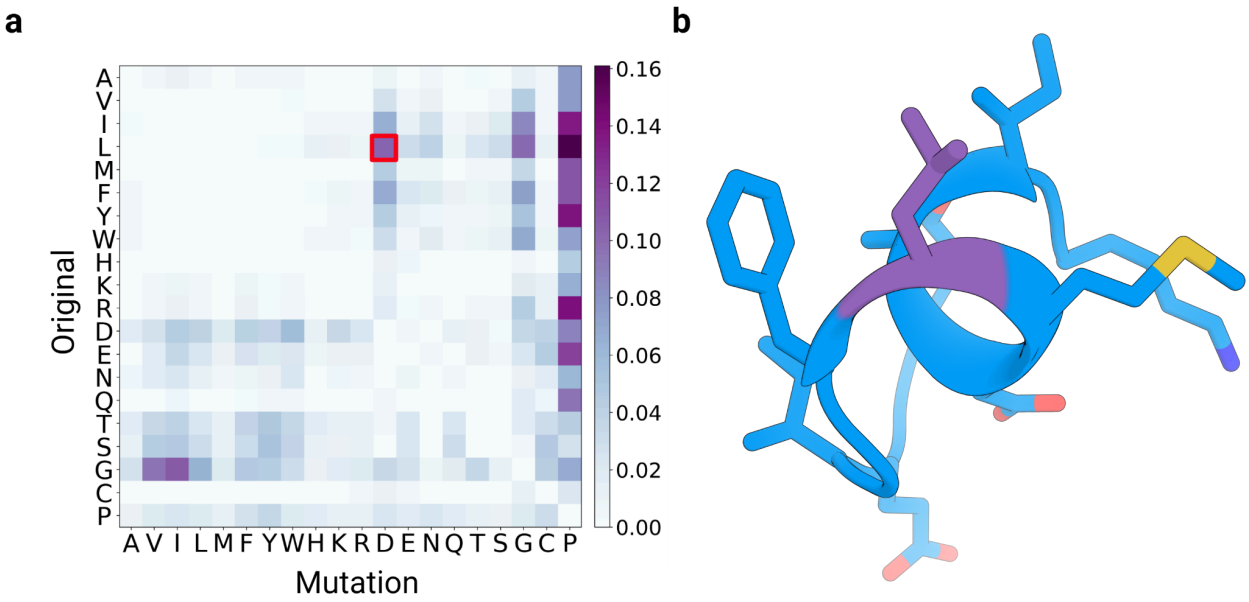
Supplementary Figure 7:



Supplementary Figure 7: The available diversity of unique structural conformations for macrocyclic peptides that can be sampled by hallucination is greater than what is observed in 48,000 designs for all peptide sizes

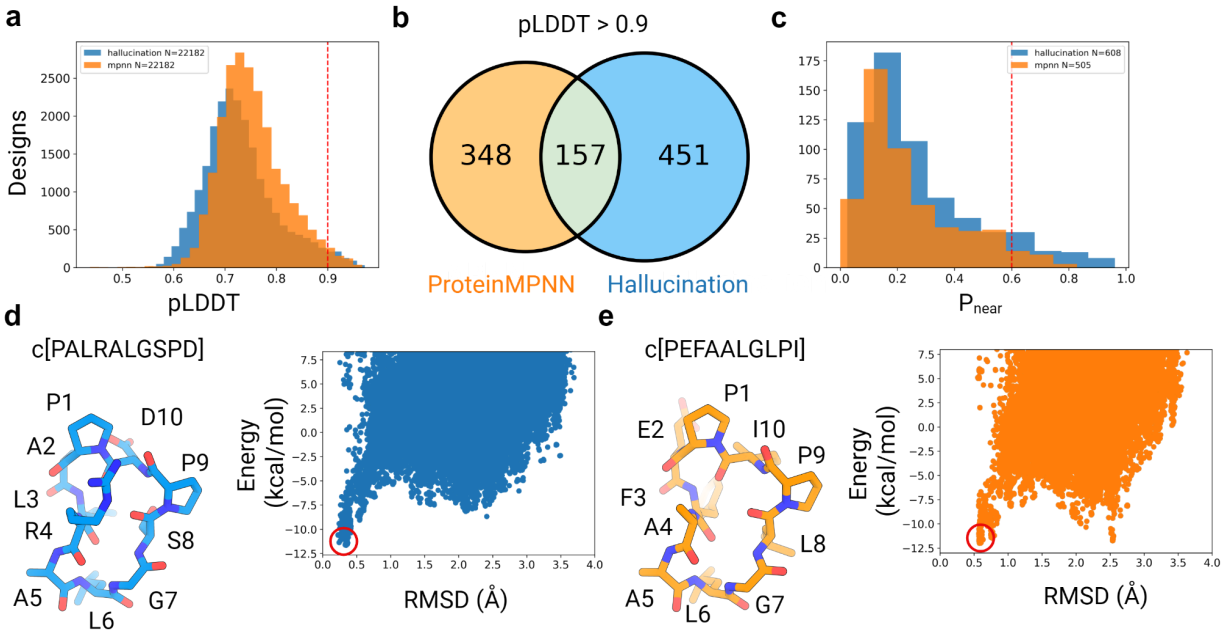
The number of sampled designs (X-axis) vs. the number of unique structural clusters shows that the number of unique structural clusters has not reached a plateau and there is still more structural diversity to be explored beyond 48,000 generated designs for all sizes (8-mers, 10-mers, and 12-mers) tested. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 8:



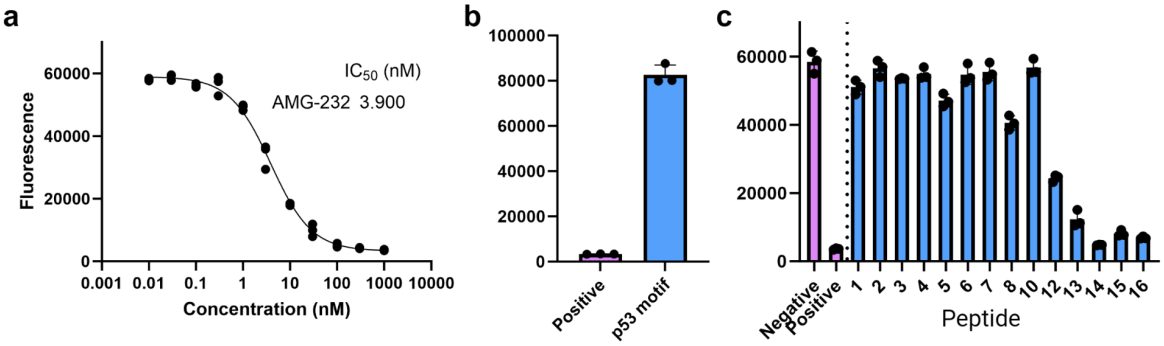
Supplementary Figure 8: Hallucinated scaffolds show high mutational tolerance *in silico*
The hallucinated set of scaffolds spanning sizes 7 to 13 residues in length and pLDDT > 0.9 were subjected to computational site saturation mutagenesis. (A) Each cell shows the fraction of such mutations that decrease the pLDDT by 0.2. Most single-point mutations are well tolerated, except the changes to prolines, and mutation of glycines to beta-branched hydrophobic residues. (B) Example of a design for which a leucine (purple) to aspartate mutation leads to a pLDDT loss of greater than 0.2. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 9



Supplementary Figure 9: *In silico* comparison of sequence design by ProteinMPNN
(A) Distribution of pLDDT for the same set of unique peptide backbones designed by hallucination and by using ProteinMPNN. (B) Venn diagram showing the union of backbones that have pLDDT > 0.9 when hallucinated or when designed using ProteinMPNN. (C) Distribution of P_{near} for backbones with pLDDT > 0.9. (D) Structure of selected backbone sequence designed by hallucination. R4 makes long-range side chain-to-main chain hydrogen bonds. Energy landscape with P_{near} of 0.91. (E) Structure of the same backbone from D but designed by ProteinMPNN, there is no long-range side chain-to-main chain hydrogen bond. The energy landscape with P_{near} of 0.75 exhibits a secondary low-energy state. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

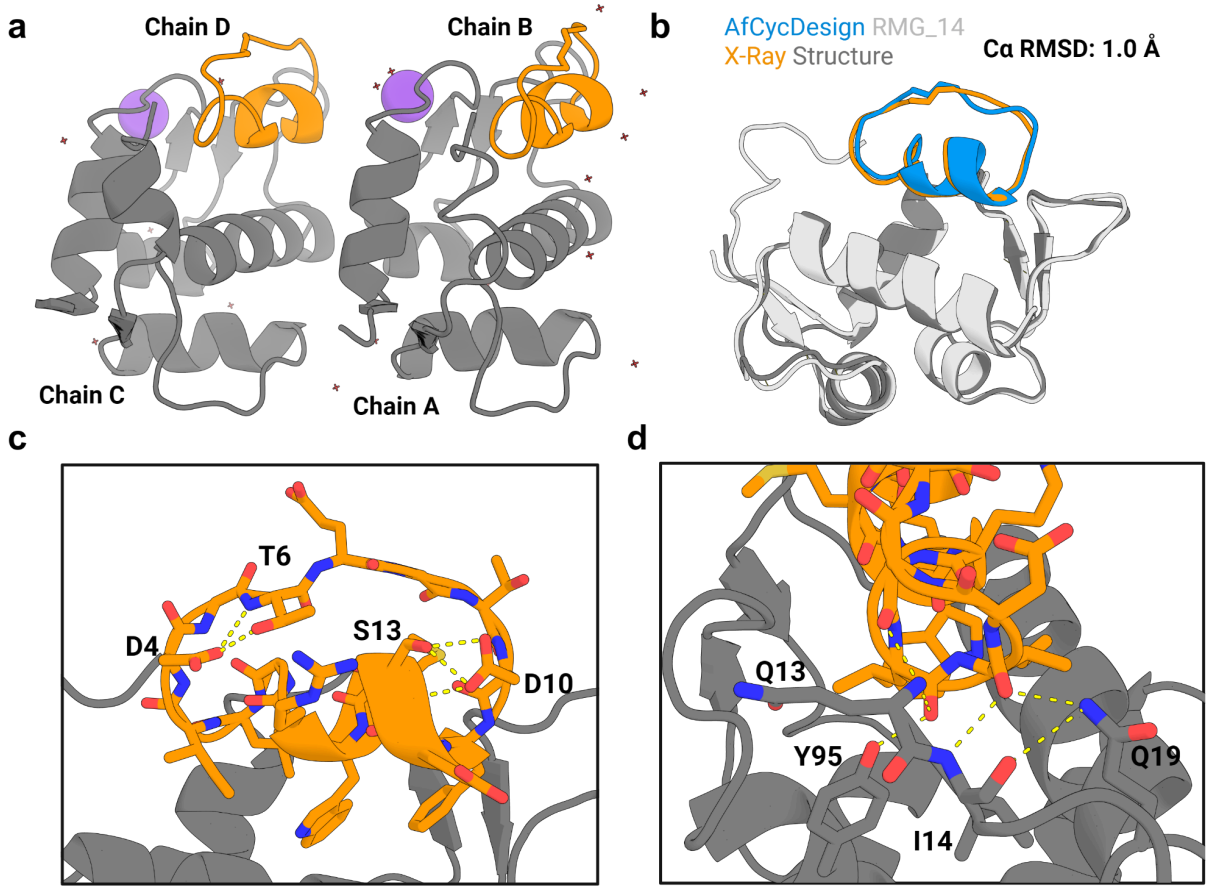
Supplementary Figure 10



Supplementary Figure 10: Grafted p53 motif alone does not bind MDM2

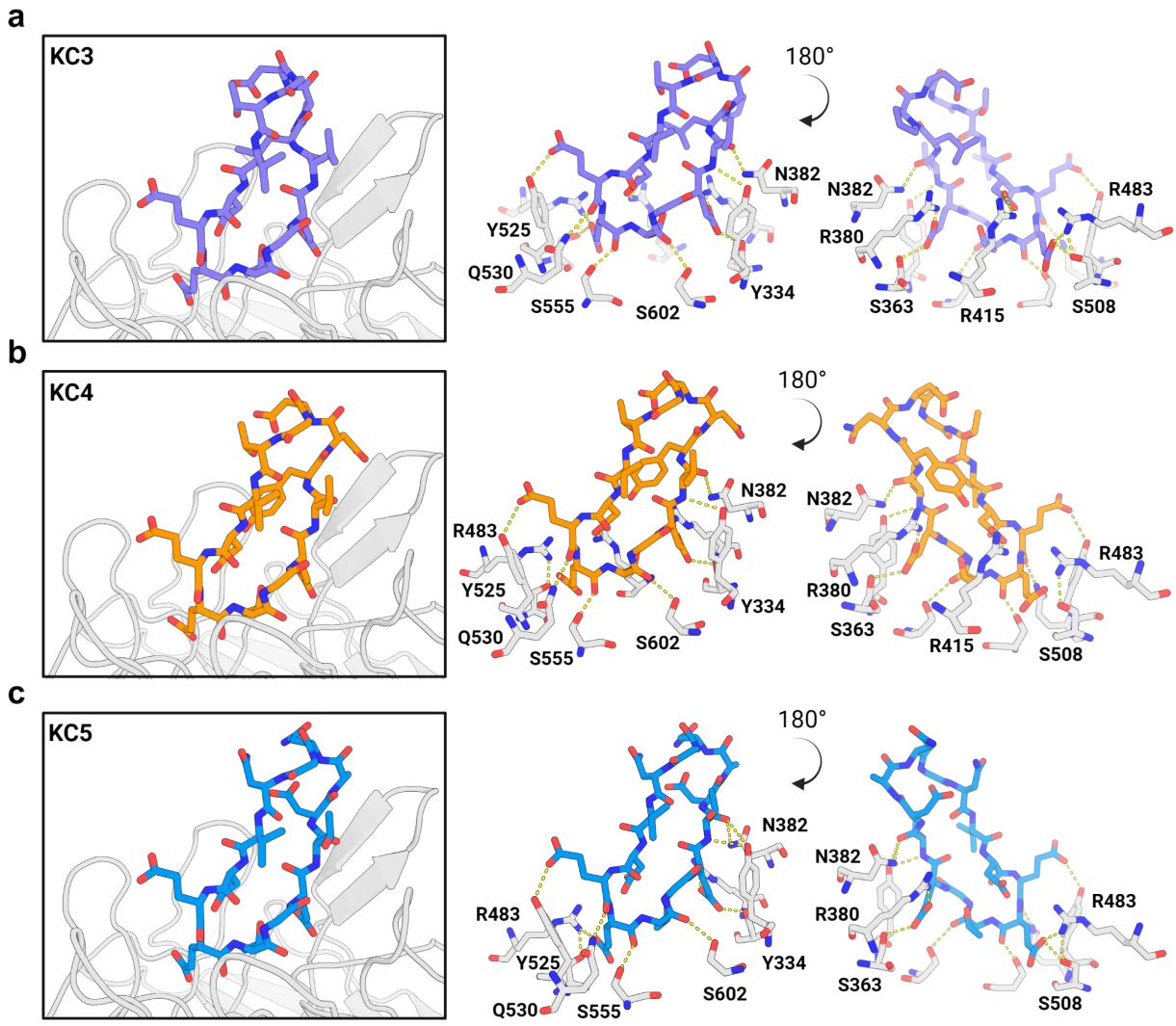
(A) AMG-232 control dose–response curve from MDM2 AlphaLISA kit. Expected IC_{50} from the kit manufacturer is 3.3 nM. (B) The five residue p53 motif (F-S-D-L-W) that was grafted onto cyclic peptide designs against MDM2 shows no activity in the AlphaLISA assay. Positive control is AMG-232 at 1 μM, p53 motif was tested at 50 μM. (C) Designs screened at 50 μM in the same assay, 12-16 show > 50% inhibition. Negative control is buffer alone, positive control is AMG-232 at 1 μM. For all plots column height: mean fluorescence shown across three replicates; Error bars: standard deviation across the three replicates. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

Supplementary Figure 11



Supplementary Figure 11: RMG_14 interaction to MDM2 in the X-ray crystal structure. (A) View of both copies of the MDM2-RMG_14 complex in the asymmetric unit. (B) Alignment of AfCycDesign model for RMG_14 to the X-ray crystal structure. (C) Sidechain mediated internal hydrogen bonding for RMG_14. Thr6 on the loop donates hydrogen bonds from the sidechain and backbone NH to the sidechain of Asp4 near the C-terminus of the helix, Asp10 caps the N-terminus of the helix. (D) A dense network of polar interactions at the C-terminus of the helix contributes to the binding of RMG_14 to MDM2. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

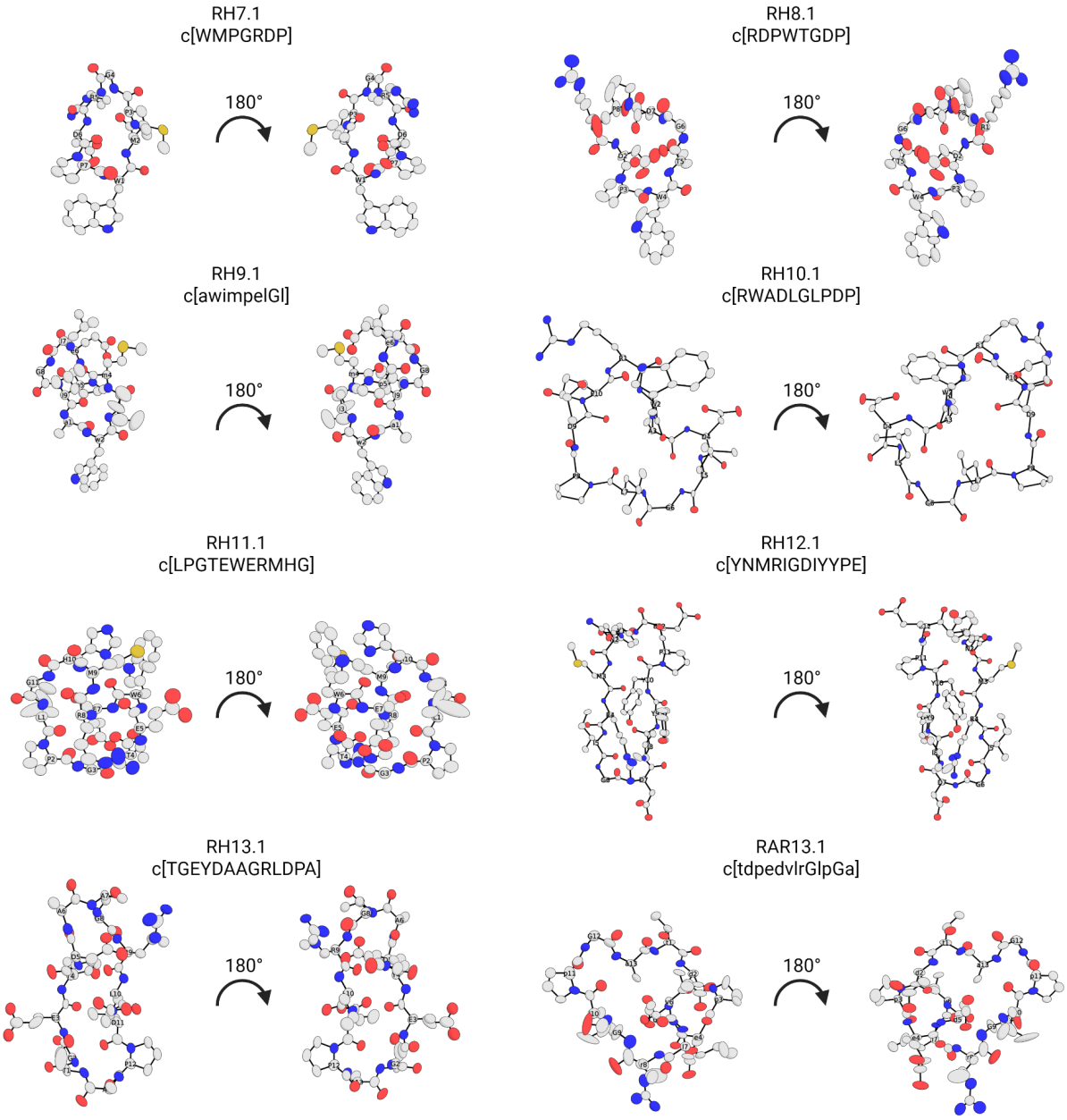
Supplementary Figure 12



Supplementary Figure 12: Interactions of designed Keap1 binders to Keap1 in design models

(A) Interactions of KC3 to Keap1. Left panel: KC3 in sticks, Keap1 as cartoon. Right panel: Rotated views of KC3 and Keap1 interacting residues shown as sticks. (B) Interactions of KC4 to Keap1. Left panel: KC4 in sticks, Keap1 as cartoon. Right panel: Rotated views of KC4 and Keap1 interacting residues shown as sticks. (C) Interactions of KC5 to Keap1. Left panel: KC5 in sticks, Keap1 as cartoon. Right panel: Rotated views of KC5 and Keap1 interacting residues shown as sticks. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

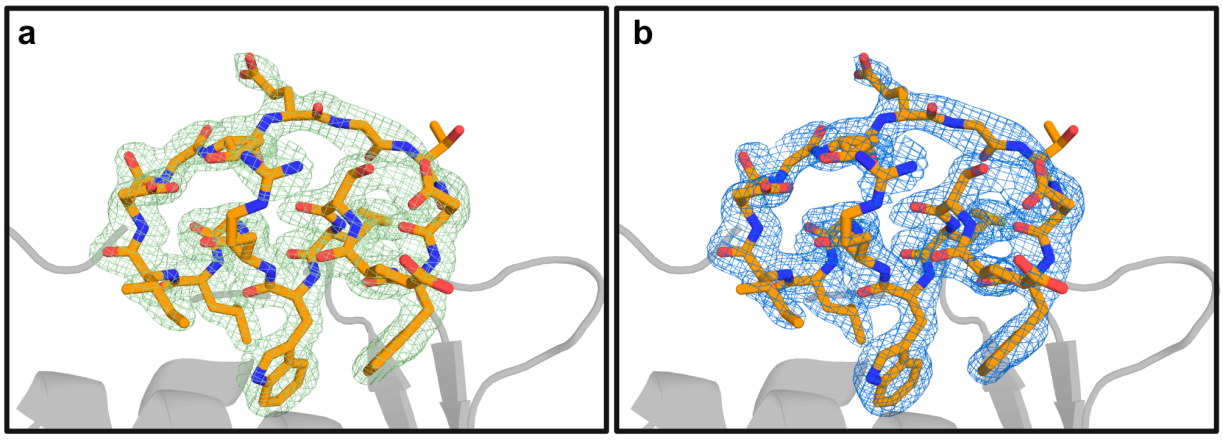
Supplementary Figure 13



Supplementary Figure 13: Probability ellipsoids for X-ray crystal structures of peptide RH7.1-13.1 and RAR13.1.

Thermal ellipsoids for crystal structures of hallucinated peptides. Each structure is shown with two views, rotated 180°. Amino acid identity and residue number labeled on Cα atoms. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

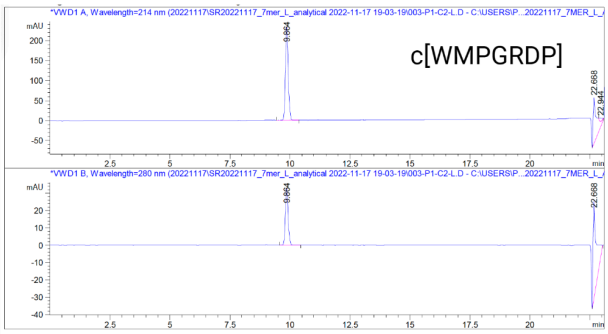
Supplementary Figure 14

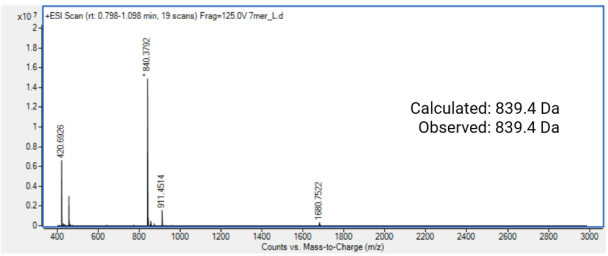


Supplementary Figure 14:
(a) 2Fo–Fc map (green mesh) at 1 σ contour level of RMG_14 bound to MDM2. (b) Omit map (blue mesh) at 1 σ contour level of RMG_14 bound to MDM2. Created in BioRender. Rettie, S. (2025) https://BioRender.com/dgyi674

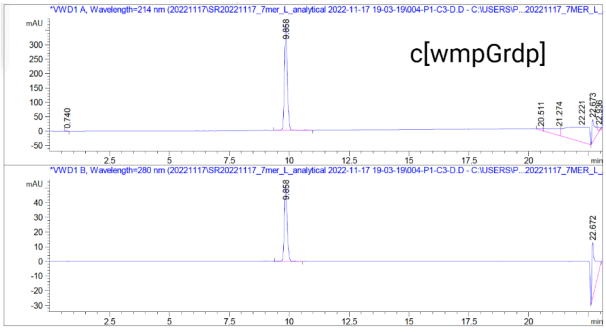
Peptide Purity and Identity

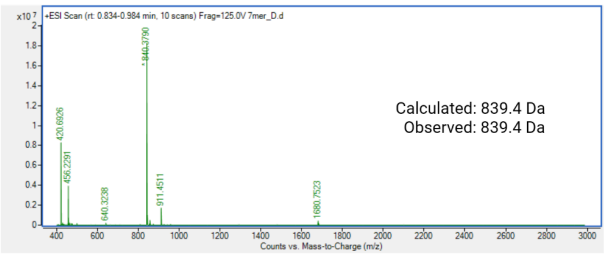
RH7.1 L



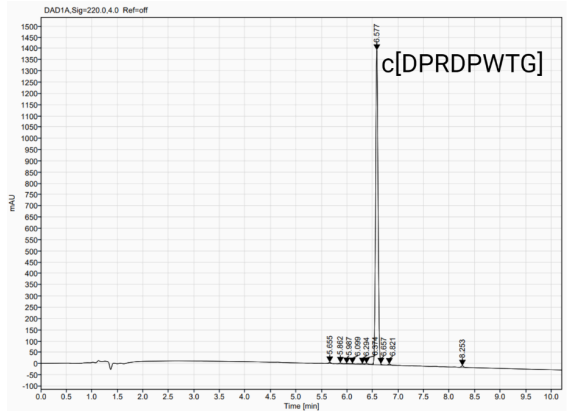


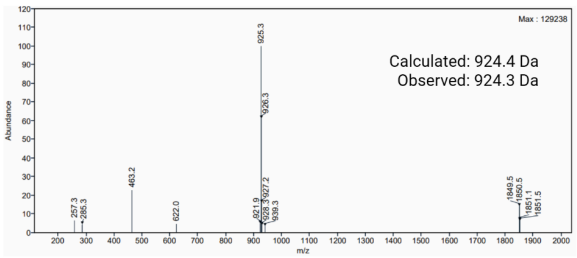
RH7.1 D



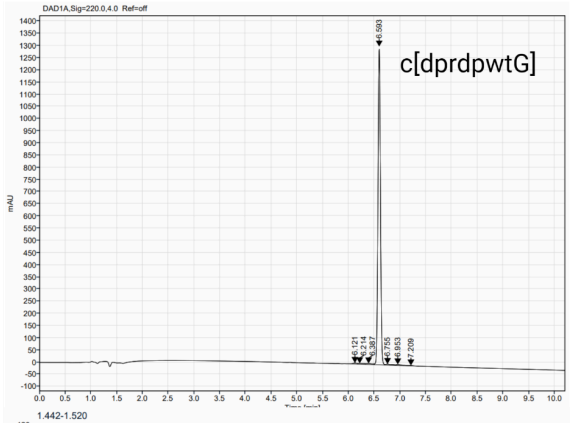


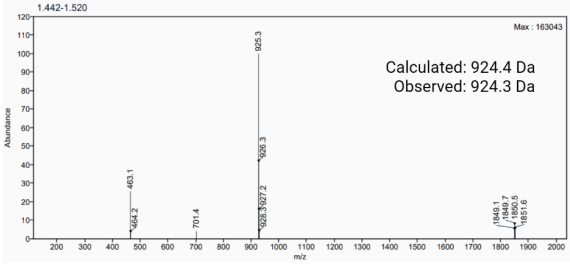
RH8.1 L



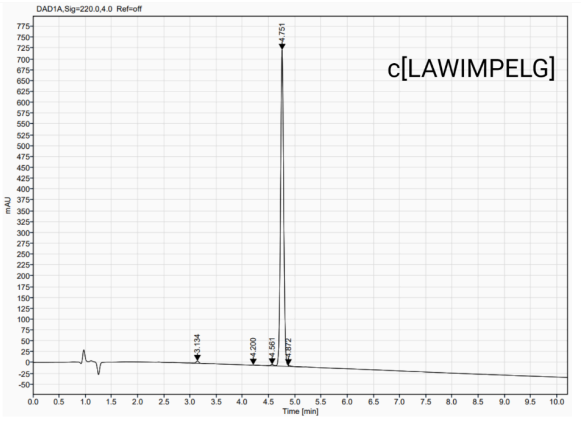


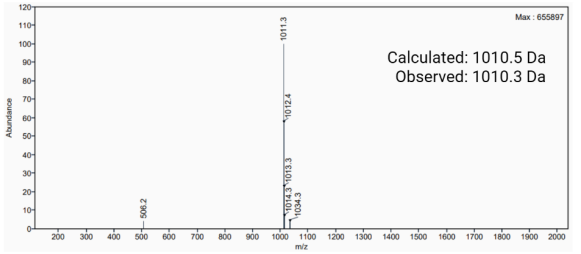
RH8.1 D



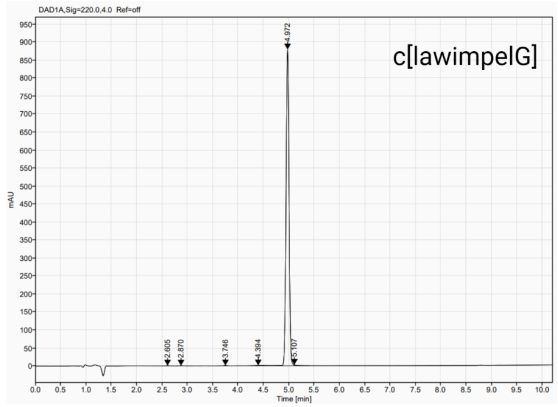


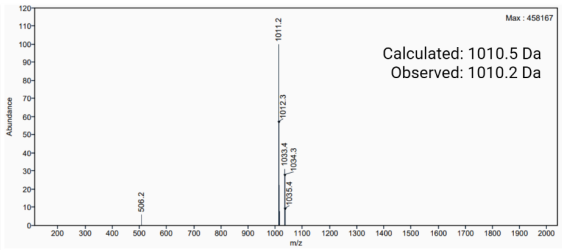
RH9.1 L



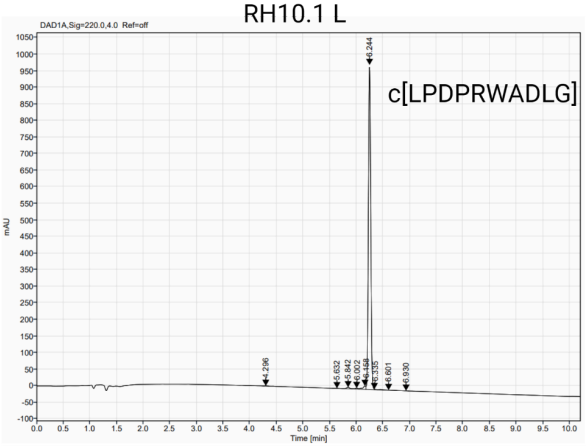


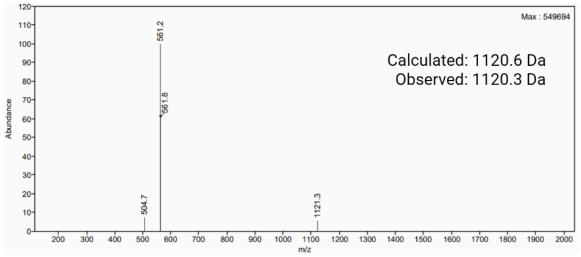
RH9.1 D



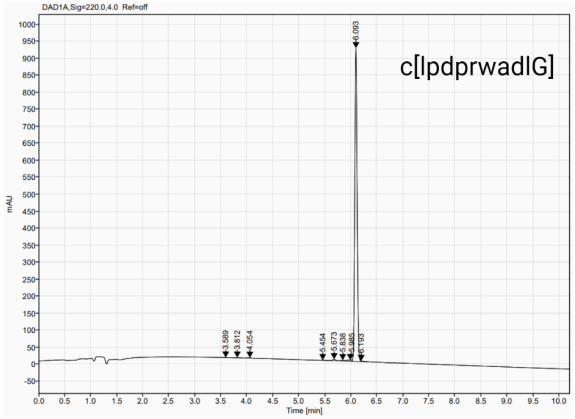


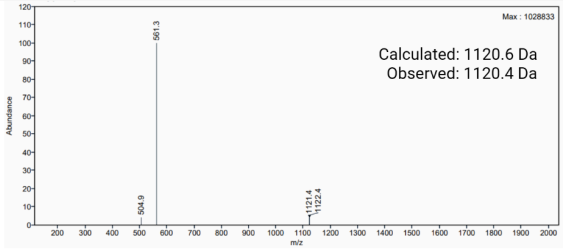
RH10.1 L



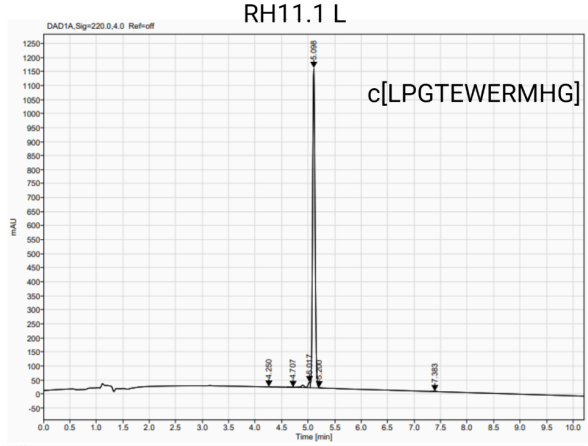


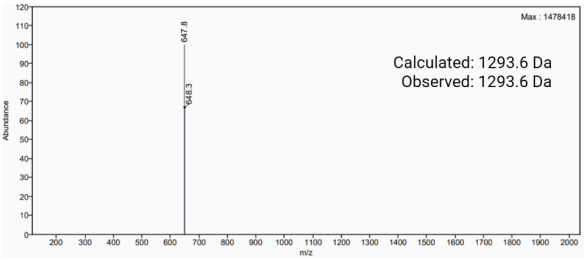
RH10.1 D



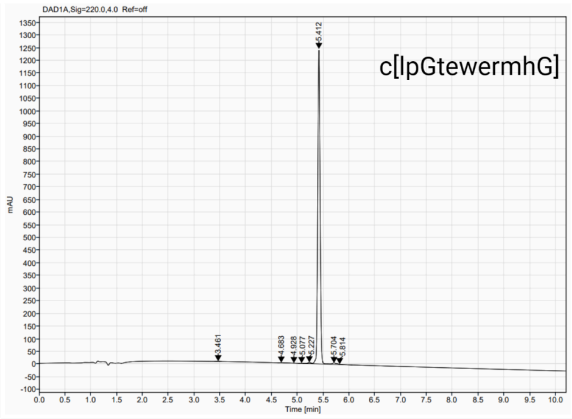


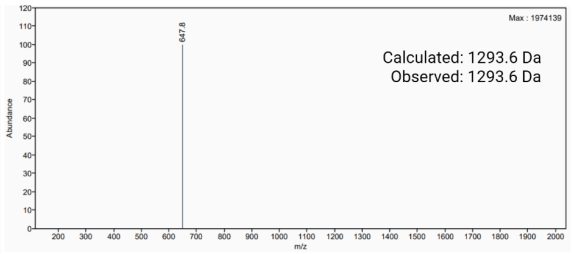
RH11.1 L



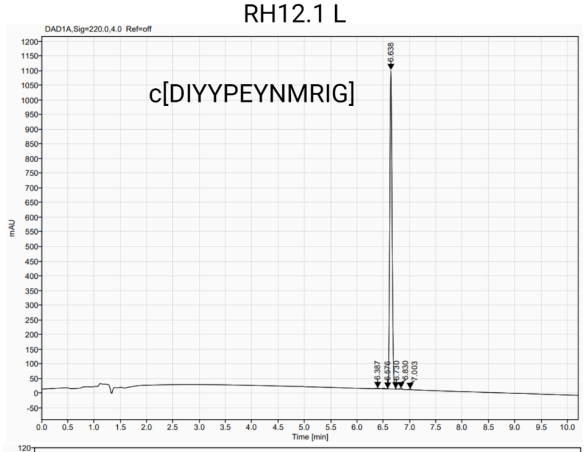


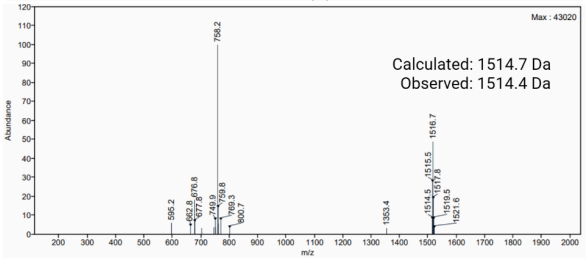
RH11.1 D



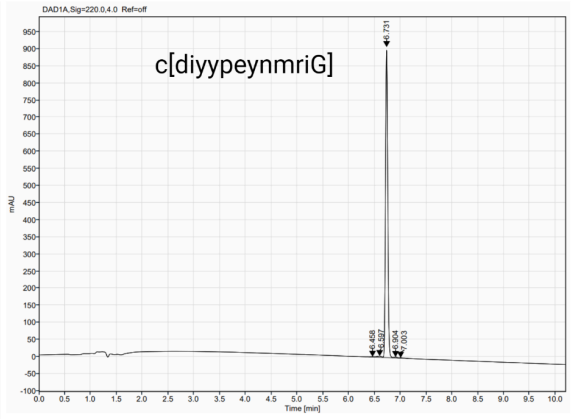


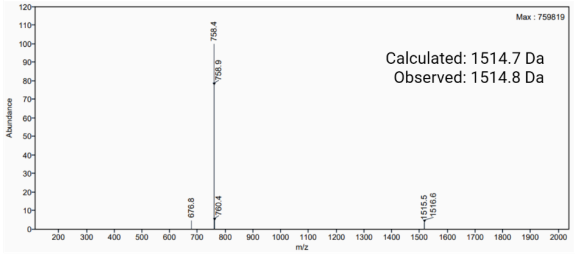
RH12.1 L



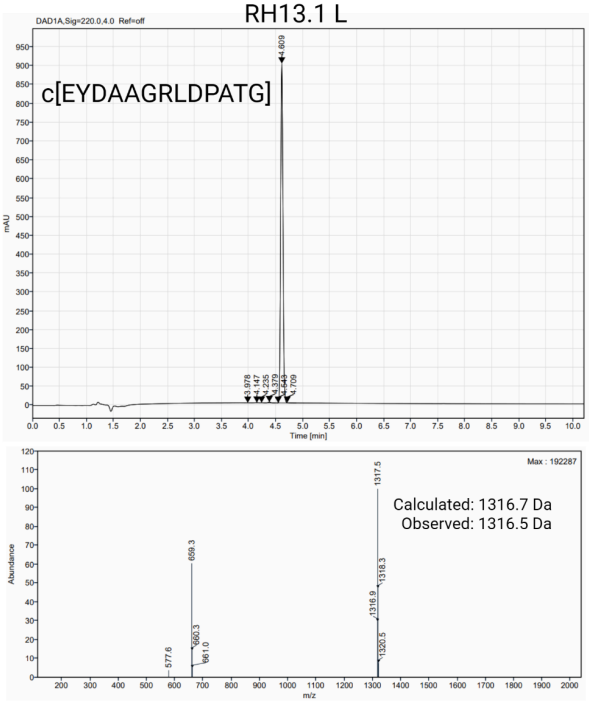


RH12.1 D

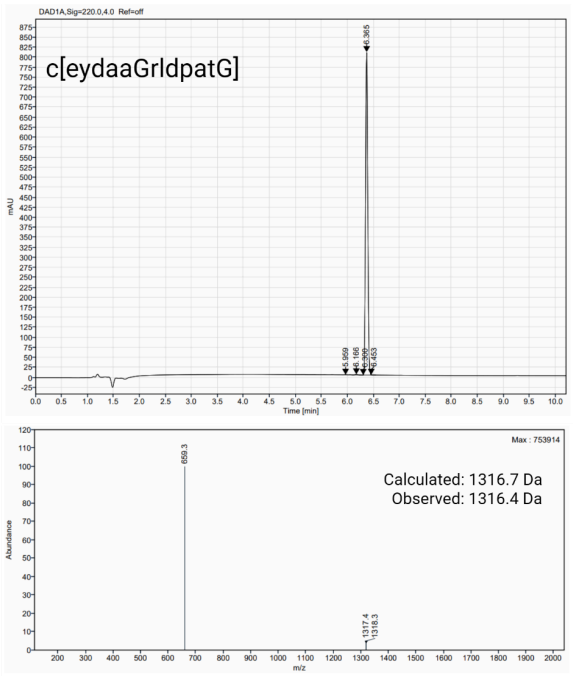




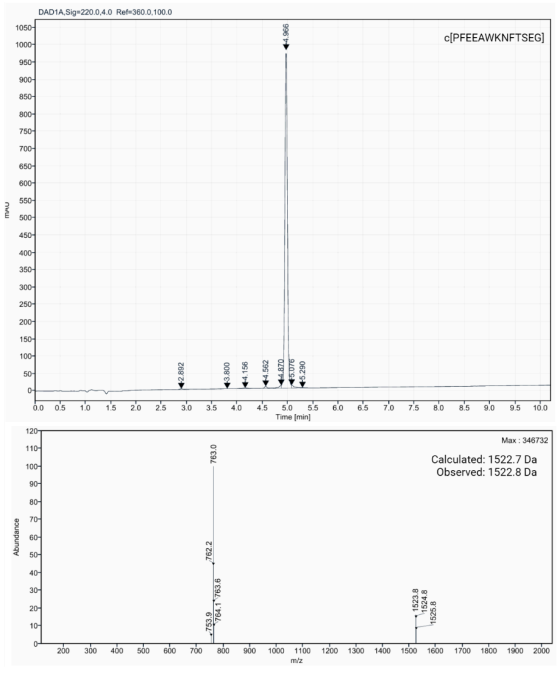
RH13.1 L



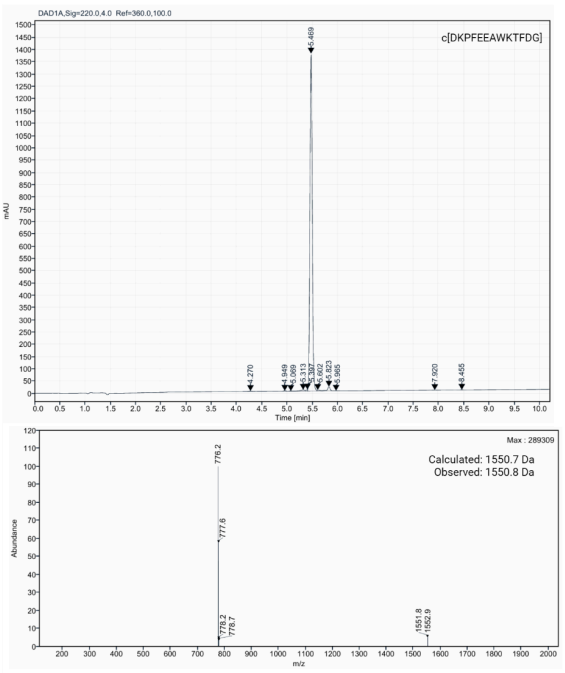
RH13.1 D



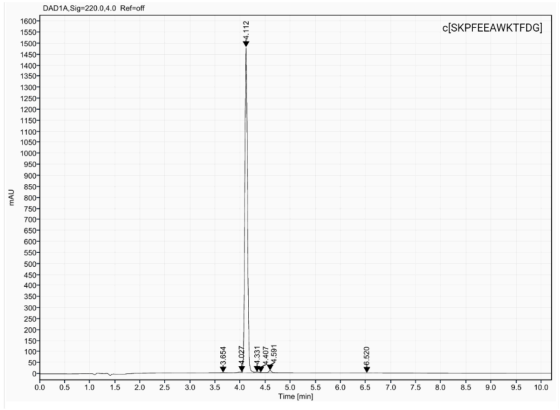
RMG_1

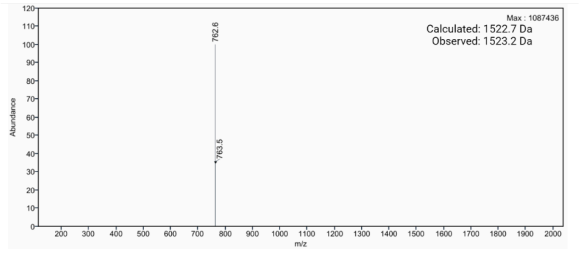


RMG_2

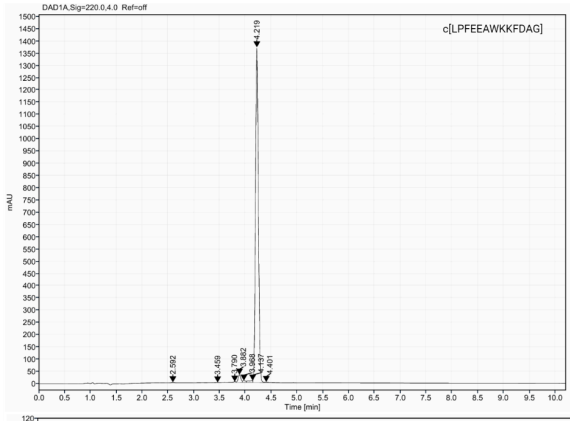


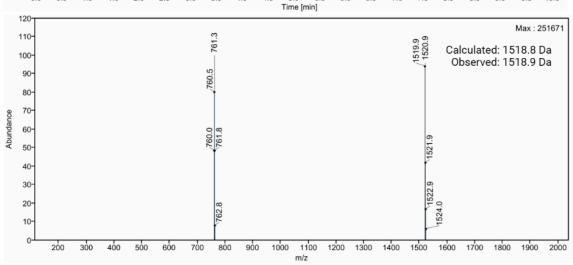
RMG_3



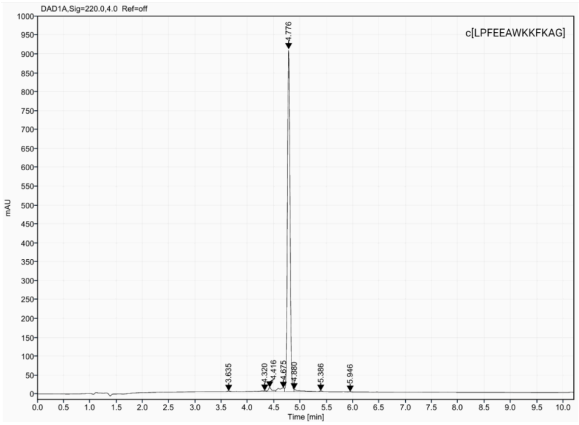


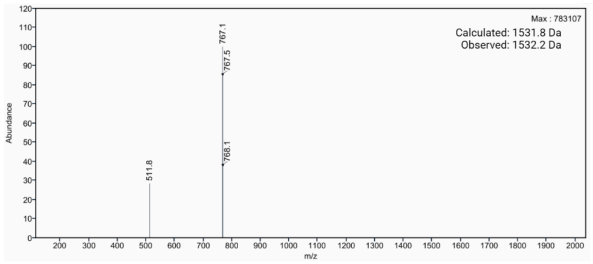
RMG_4



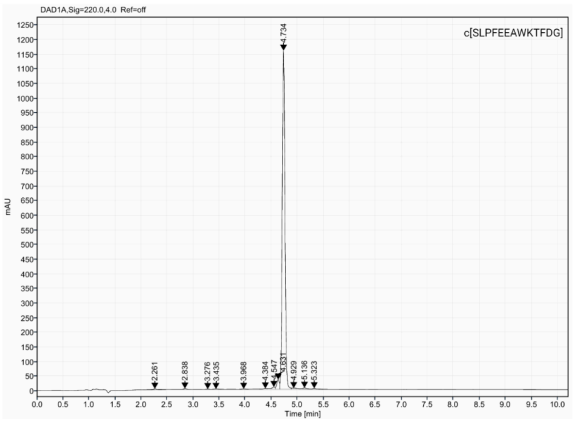


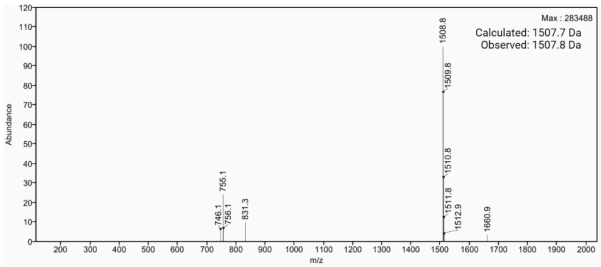
RMG_5



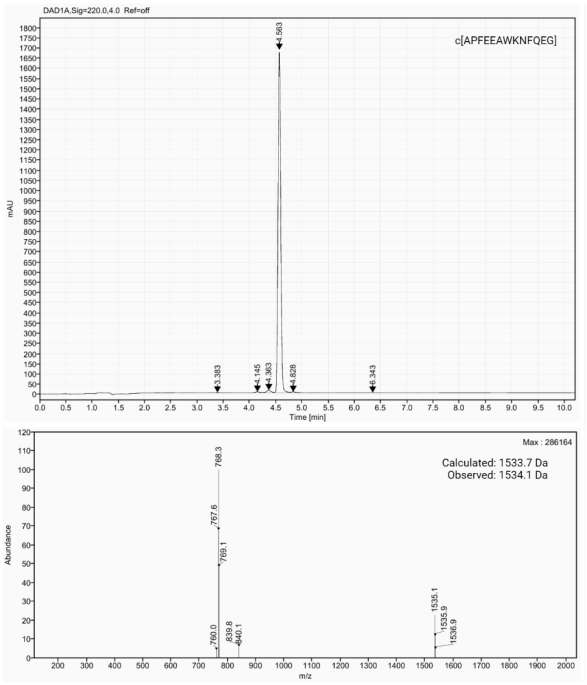


RMG_6

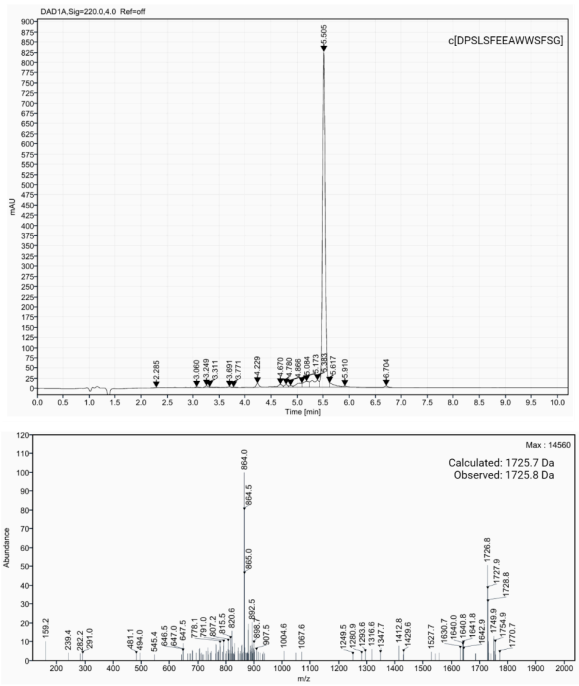




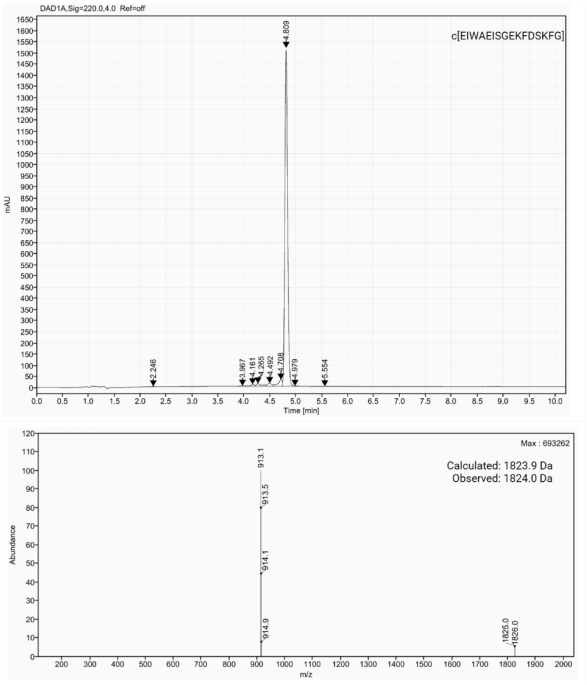
RMG_7



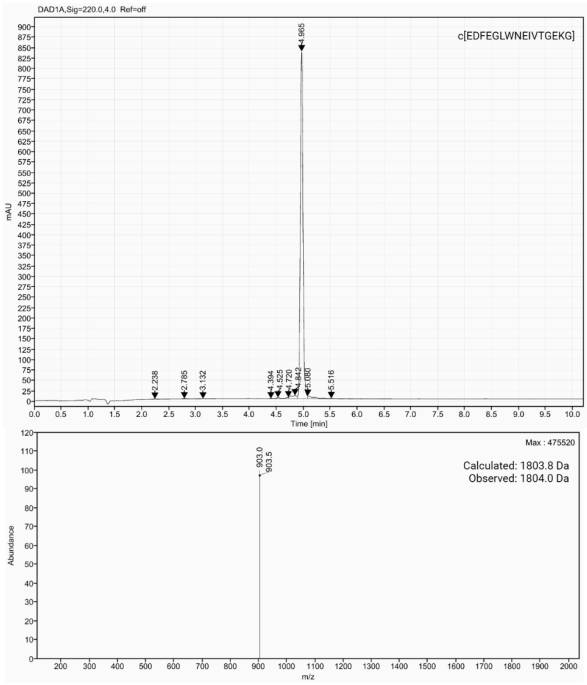
RMG_8



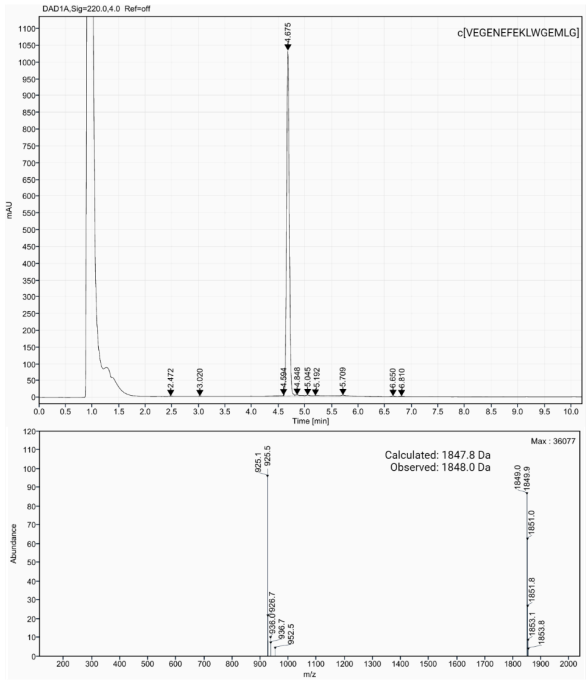
RMG_10



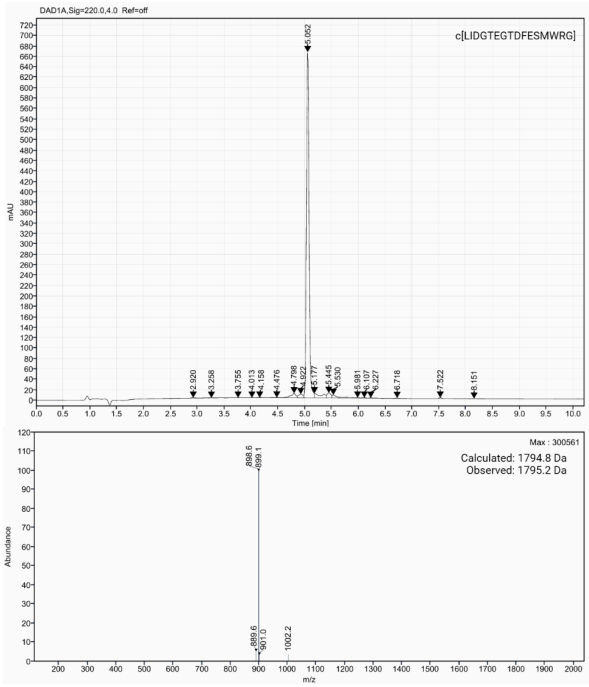
RMG_12



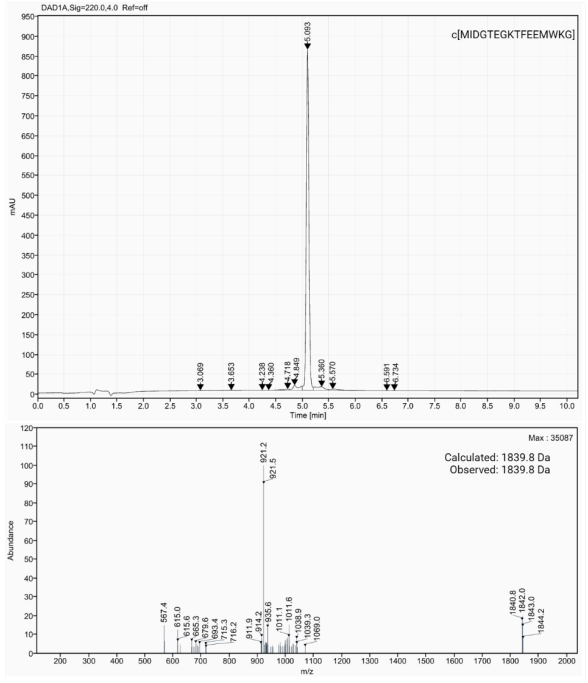
RMG_13



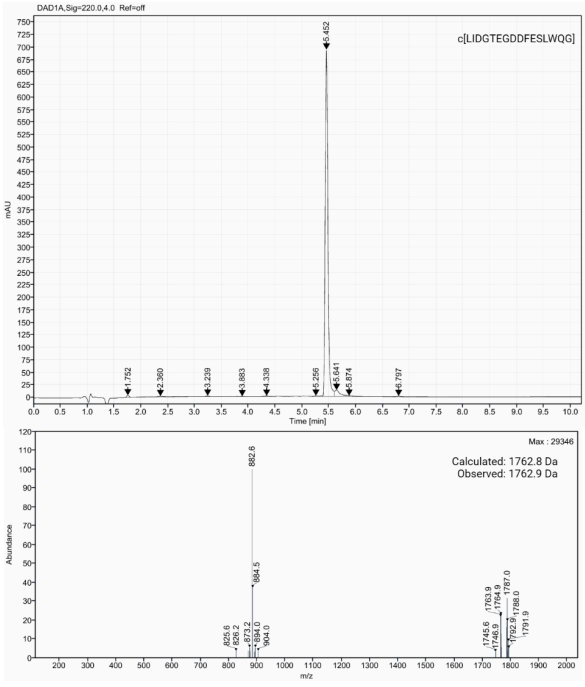
RMG_14



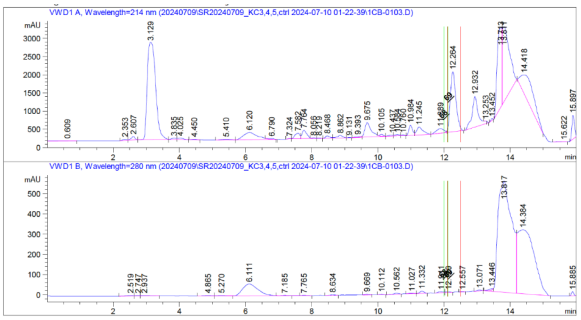
RMG_15

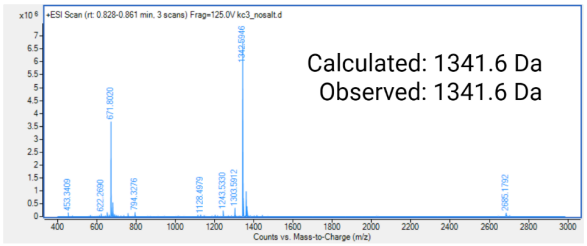


RMG_16

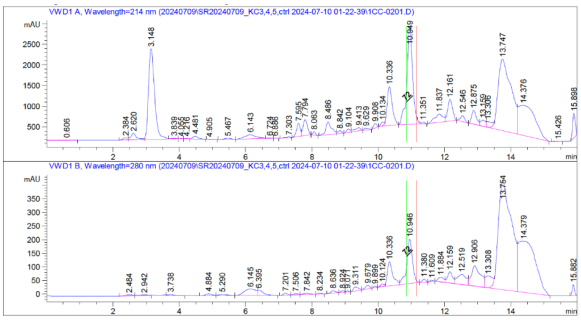


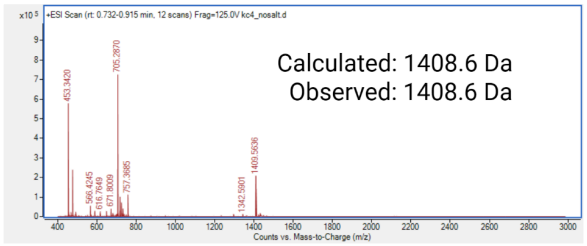
KC3 c[ETGEVLPDGTVDE]



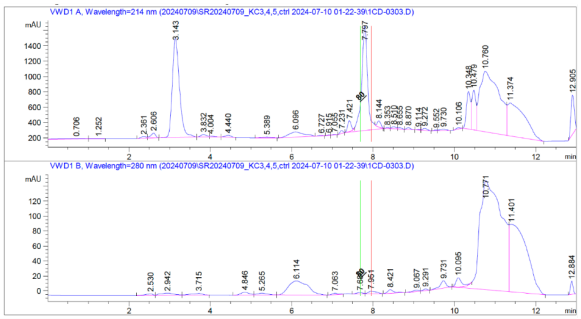


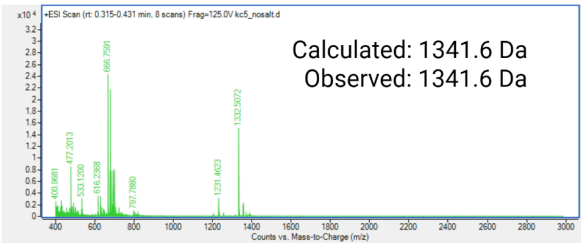
KC4 c[ETGEVYNDGTVDE]



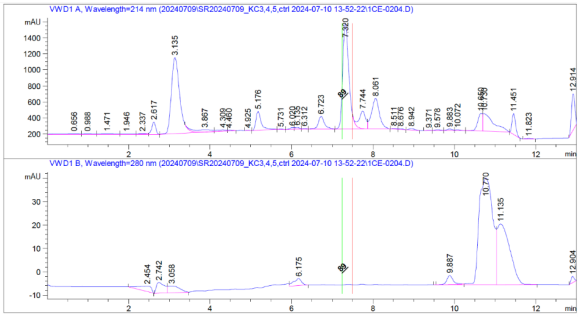


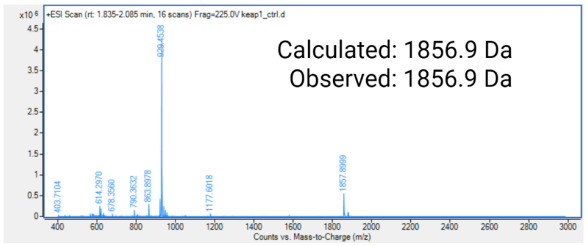
KC5 c[ETGETDANGNVDE]





Nrf2 16aa c[AFFAQLQLDEETGEFL]





Supplementary Table 1
Data collection and refinement statistics.

	MDM2-Peptide (PDB ID: 9CDZ)
Resolution range	24.47 - 1.7 (1.81 - 1.72)
Space group	P 2_1
Unit cell	36.78, 48.93, 49.27; 90, 111.64, 90
Unique reflections	15978 (2539)
Multiplicity	6.8 (6.7)
Completeness (%)	91.40 (100.00)
Mean I/sigma(I)	11.1 (1.1)
Wilson B-factor	34.83
R-merge	0.068 (1.659)
R-pim	0.030 (0.754)
$CC_{1/2}$	0.998 (0.529)
Reflections used in refinement	15929 (1609)
R-work	0.2216 (0.4051)
R-free	0.2704 (0.4517)
Number of non-hydrogen atoms	1816

macromolecules	1798
solvent	16
Protein residues	220
RMS(bonds)	0.005
RMS(angles)	0.700
Ramachandran favored (%)	97.64
Ramachandran allowed (%)	2.36
Ramachandran outliers (%)	0.00
Average B-factor	48
macromolecules	48
solvent	53

The highest-resolution shell are shown in parentheses.