Supplementary for

Predicting protein-peptide complex structures by accounting for peptide flexibility and physicochemical environment

Xianjin Xu, Xiaoqin Zou[*]

Dalton Cardiovascular Research Center, Department of Physics and Astronomy, Department of Biochemistry, Institute for Data Science and Informatics, University of Missouri, Columbia, MO 65211, USA

*Correspondence to: Xiaoqin Zou, Email: zoux@missouri.edu

Table S1 Percentages of the contact residues in the peptides from the PepPro dataset. An amino acid is defined as a contact residue if more than 1/3 of its surface area is buried.

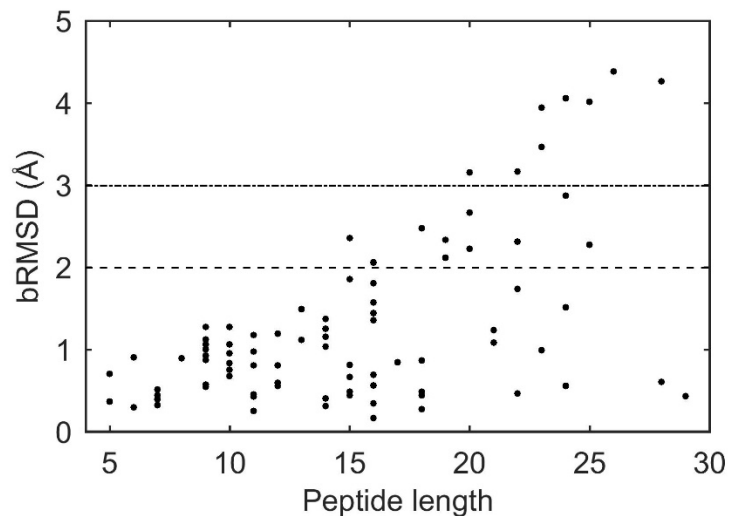| Amino Acid	Number of contact residues	Total number of residues	Percentage of contact residues (%)
ALA	48	85	56.5
ARG	59	101	58.4
ASN	35	69	50.7
ASP	34	86	39.5
CYS	4	5	80.0
GLU	33	93	35.5
GLN	37	70	52.9
GLY	27	55	49.1
HIS	7	14	50.0
ILE	52	63	82.5
LEU	113	140	80.7
LYS	54	110	49.1
MET	21	28	75.0
PHE	46	51	90.2
PRO	56	91	61.5
SER	51	98	52.0
THR	31	55	56.4
TRP	20	26	76.9
TYR	33	44	75.0
VAL	54	68	79.4
Total	815	1352	60.3

Figure S1 Distribution of the bRMSD values between the best fragment hits (i.e., the fragment with the lowest bRMSD) and the corresponding bound peptide structures. The broken line and the dished line represent 2.0 Å and 3.0 Å, respectively.
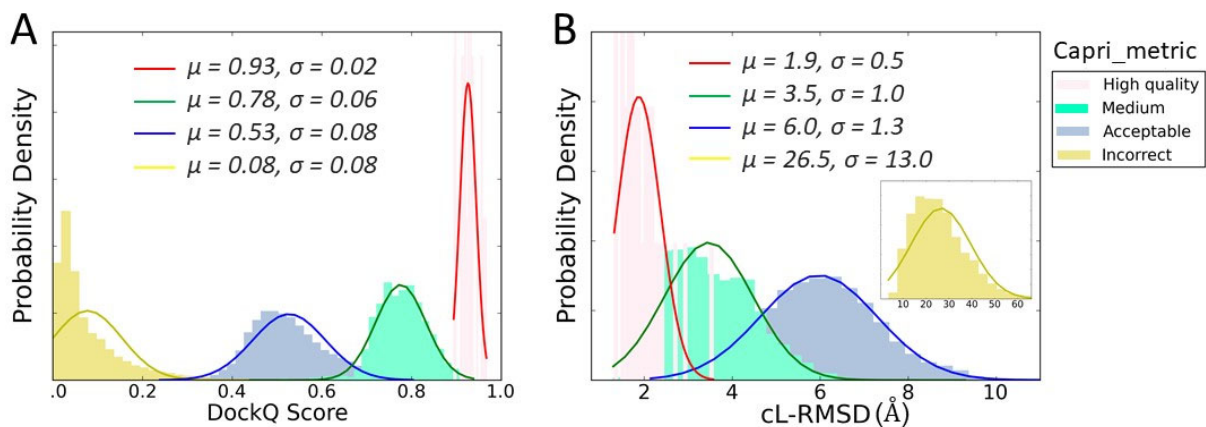

Figure S2 Distributions of DockQ Scores (A) and cL-RMSD values (B) for each group of models classified by capri_metric. The results were based on all sampled models (a total of ~8.4×10^6 models) of bound cases in the PepPro dataset. The total numbers of high quality, medium, and acceptable models were 34, 2088, and 47888, respectively. The probability densities were normalized by the total number of models in each group.
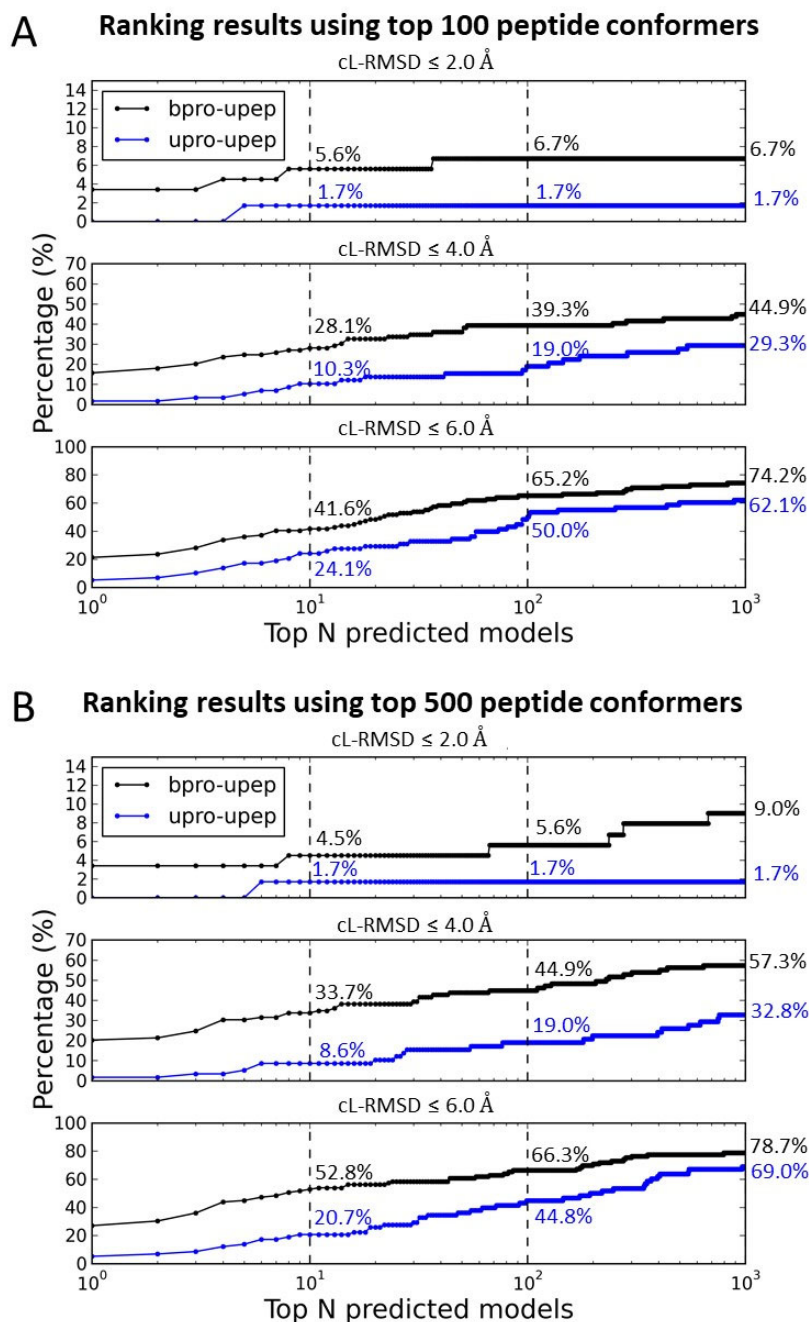
Figure S3 Ranking of the peptide binding modes generated using top 100 (A) and 500 (B) peptide conformers in the sampling stage for both bound docking (bpro-upep) and unbound docking (upro-upep) in the PepPro dataset. The success rates were calculated based on top 10, 100, or 1000 ranked models of each prediction. The subpanels show the ranking performances using different thresholds of cL-RMSD (2.0 Å, 4.0 Å, and 6.0 Å). The percentages (%) reported in the figures are the ranking success rates when top 10, 100, 200, or 1000 models were considered for each prediction.
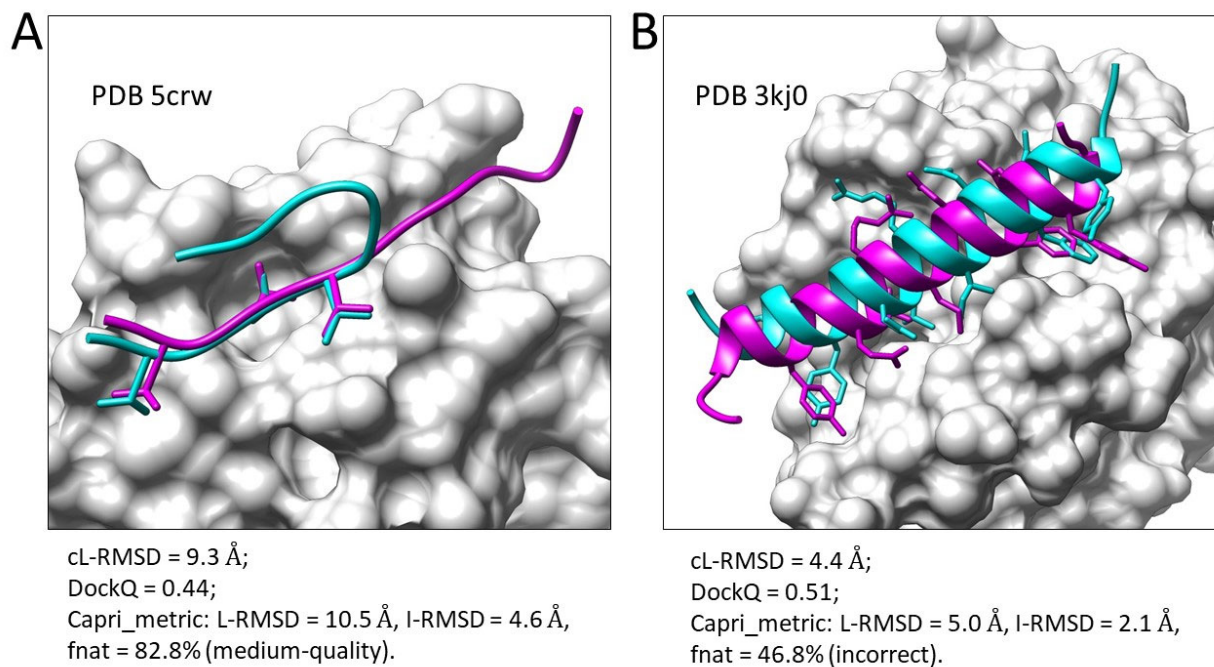
A PDB 5crw

cL-RMSD = 9.3 Å;
DockQ = 0.44;
Capri_metric: L-RMSD = 10.5 Å, I-RMSD = 4.6 Å,
fnat = 82.8% (medium-quality).

B PDB 3kj0

cL-RMSD = 4.4 Å;
DockQ = 0.51;
Capri_metric: L-RMSD = 5.0 Å, I-RMSD = 2.1 Å,
fnat = 46.8% (incorrect).

Figure S4 Two examples of peptide binding models sampled by MDockPeP2 bound docking. Proteins are displayed by the surface and colored light gray. Experimental bound peptide structures are colored cyan and predicted peptide binding modes are colored magenta. The side chains of the contact peptide residues are represented by the stick model.