Supplementary data for

AutoDock CrankPep: Combining folding and docking to predict protein-peptide complexes

Yuqi Zhang[1] and Michel Sanner[1*]

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, United States of America

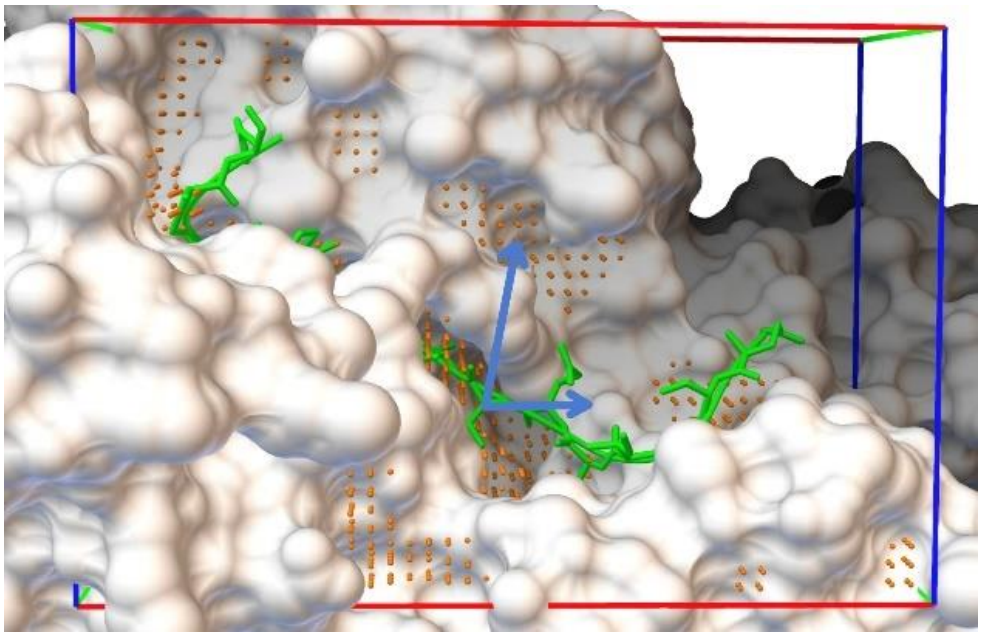

Figure S1. *AutoDock CrankPep* additional Monte Carlo moves. The affinity maps calculated by *AutoGridFR* define the 3D space the peptide can explore. The translational points are promising position for placing the peptide root atom.

Chart S1. Flow chart describing the algorithm for updating the pose cache used in AutoDock CrankPep to escape local minima.

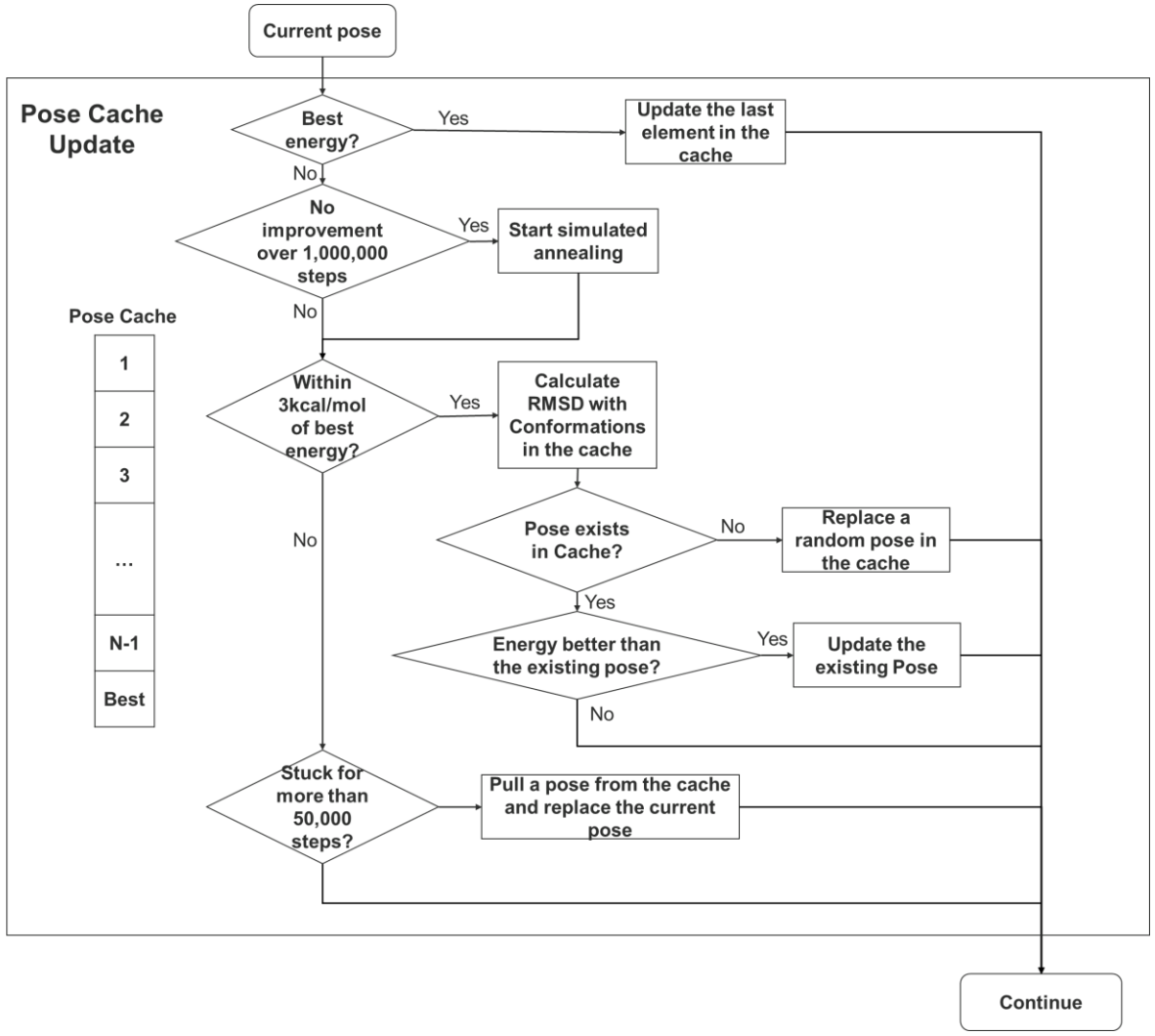

Table S1. Details information of the 12 systems containing longer peptides queried from PDB.

PDB ID	Peptide Chain	Contacting Receptor Segment(s):Chain(s)*	BioMolecule Number	Length	Apo PDB ID	Apo Receptor Chain
1CM1	B	A	0	18	1PRW	A
2F31	B	A	0	20		
2IVZ	F	B	1	16	1CRZ	A
2OBH	C	A-B-D	0	16		
2XAP	D	A-B-E	0	16		
4AK4	P	I-J-K-L-M-N-O	1	16		
4RS9	B	A	0	19		
4YZ6	B	A	0	18		
5N4B	D	A	0	17	5N4F	A
5UWI	D	A-B-C	0	16	4HB2	C
6CIT	D	A-B-C	0	17	4HB2	C

* This column lists the receptor chains contacting the peptide in the biomolecule. The segment information is created as chains are duplicated in the process of building the biomolecules.

Docking pose clustering

The docking poses from *AutoDock CrankPep* can be clustered based on either peptide backbone RMSD or contacts between predicted peptide and the receptor. The clustering algorithm is as following:

1. Identify the cluster seed as the pose with the best score in the solution pool;
2. Form a cluster containing all poses that are similar to the cluster seed. Depending on the user specified similarity measure, poses are deemed similar if (i) they have a backbone RMSD with the cluster seed pose below than a user-define value or, (ii) they reproduce more than a user-defined percentage of contacts identified in the cluster seed pose;
3. Remove the clustered poses from the solution pool;
4. Repeat until all poses are clustered.

Table S2. System specific comparison between *AutoDock CrankPep* and *HPepDock* on the complexes with peptides longer than 4 amino acids from the *Leads_Pep* dataset.

Size	PDB	*ADCP*			*HPepDock*			Median *ADCP* MC-replica time* (mins)
		Top 1	Top 10	Top 1000	Top1	Top10	Top 1000	
5	1NVR	3.2	1.2	0.7	9.9	1.5	0.8	3.88
5	2HPL	3	0.9	0.8	1.3	1.2	0.7	7.82
5	2V3S	3	3	1	1.8	1.4	0.9	44.02
5	3NFK	8.3	1	0.4	1.4	0.8	0.5	41.17
5	3T6R	8.2	5.9	0.9	1.9	1	1	62.40
5	4V3I	5.3	4.1	2.3	4.5	1.8	0.9	54.78
6	1SVZ	1.6	0.7	0.7	2.3	1.9	1.7	20.72
6	3D1E	1.1	1.1	0.6	2.2	1.1	1.1	17.55
6	3IDG	1.3	1	0.8	1.8	1.8	1.4	40.33
6	3LNY	1.7	1.1	0.7	0.8	0.8	0.7	21.13
6	4NNM	1.4	1.3	0.8	2.1	0.9	0.9	10.13
6	4Q6H	1.1	0.8	0.5	2.9	1.1	0.5	53.22
7	3MMG	0.5	0.5	0.5	1.6	1.4	1.4	74.03
7	3NJG	1.6	1	0.7	1.2	0.5	0.5	72.18
7	3Q47	0.7	0.5	0.5	6.9	1.9	1.1	15.35
7	3UPV	1.4	1.3	0.7	5.1	2.4	1.7	24.15
7	4QBR	3.1	1	0.6	1.3	1.3	0.9	91.78
8	1ELW	4.7	1.6	1.2	2.5	1.7	0.8	29.40
8	1N7F	1.5	0.7	0.6	4.3	1.4	0.8	69.53
8	1OU8	1.1	0.9	0.7	2.9	2.9	1.9	28.43
8	3CH8	0.5	0.5	0.5	1.1	1.1	1.1	29.98
8	4WLB	1.4	0.7	0.7	0.9	0.7	0.3	83.05
9	2QAB	4.1	1.8	1	1.9	1.2	0.5	117.55
9	2W0Z	15.3	6.7	0.7	1.6	1.1	1	96.03
9	3OBQ	3.8	1.5	0.8	2.8	2	1.4	12.78
9	4BTB	3.2	1.5	1.1	11.4	3.1	1.1	9.30
9	4N7H	1.4	1.4	0.9	14.1	2.7	2.2	29.13
10	1H6W	1	0.8	0.8	3.6	3.6	3.6	127.68
10	1NTV	6.1	0.8	0.8	15.7	7.8	1.8	133.63
10	2O02	3.7	2.4	1.9	7.6	4.9	2.5	28.03
10	3BRL	1	0.8	0.8	8.4	5.4	1.9	72.27
10	4DS1	1	1	0.9	1.1	0.7	1.6	50.58
11	1N12	0.6	0.6	0.6	23.3	4.3	4.3	111.15
11	2XFX	1.9	1.6	1.6	7	4.5	3.5	139.72
11	3BFW	0.7	0.7	0.7	2	1	0.9	159.83
11	3DS1	1.2	0.6	0.6	1.1	0.7	0.7	57.93
11	4EIK	4.5	3.6	2	5.5	3.1	1.9	152.18
12	2B9H	4	2.7	2.4	2.1	2.1	2.1	239.43
12	2W10	15.3	1.8	0.8	15.4	2.8	2.4	137.73
12	3JZO	8.9	2.5	1.4	13	4.2	1.6	81.73
12	4DGY	9.1	1.6	1.6	7.1	6.3	4.6	78.47
12	4J8S	5.3	2.3	1.9	5	5	3.8	245.48
bbRMSD ≤ 1.0		**8**	**20**	**32**	**2**	**9**	**18**	
bbRMSD ≤ 2.5		**22**	**36**	**42**	**21**	**28**	**37**	

* The reported runtimes are median values as the various MC-replicas making up an ADCP docking were performed on a heterogeneous cluster comprising Intel Xeon CPUs dating from 2007 to 2012. The performances of these chips vary significantly (up to a factor 8) between the older and newer CPUs. This runtime is for one MC replica on one CPU. The total CPU time depends on how many replicas the users chooses to perform.