YAPP-CD: Yet another protein-peptide complex database

Joon-Sang Park

Computer Engineering Dept., Hongik University
jsp@hongik.ac.kr

Abstract: Protein-peptide interactions are of great interest to the research community not only because they serve as mediators in many protein-protein interactions but also because of the increasing demand for peptide-based pharmaceutical products. Protein-peptide docking is a major tool for studying protein-peptide interactions, and several docking methods are currently available. Among various protein-peptide docking algorithms, template-based approaches, which utilize known protein-peptide complexes or templates to predict a new one, have been shown to yield more reliable results than template-free methods in recent comparative research. To obtain reliable results with a template-based docking method, the template database must be comprehensive enough; that is, there must be similar templates of protein-peptide complexes to the protein and peptide being investigated. Thus, the template database must be updated to leverage recent advances in structural biology. However, the template database distributed with GalaxyPepDock, one of the most widely used peptide docking programs, is outdated, limiting the prediction quality of the method. Here, we present an up-to-date protein-peptide complex database called YAPP-CD, which can be directly plugged into the GalaxyPepDock binary package to improve GalaxyPepDock's prediction quality by drawing on recent discoveries in structural biology. Experimental results show that YAPP-CD significantly improves GalaxyPepDock's prediction quality, e.g., the average Ligand/Interface RMSD of a benchmark set is reduced from 7.60 Å/3.62 Å to 3.47 Å/1.71 Å.

Keywords: Peptide docking; Protein-peptide Interaction; Binding Database

1. Introduction

Recently, peptide docking has drawn more attention as protein-peptide interactions play key roles in many biological processes, and there is an increasing demand for peptide-based pharmaceutical products [1,2]. Several computational tools for peptide docking have become available, such as GalaxyPepDock [3], pepATTRACT [4], MDockPeP [5], CABS-dock [6] ClusPro [7], PIPER-FlexPepDock [8], HPEPDOCK [9], InterPep [10], GOLD [11] Surflex-Dock [12], DynaDock [13], HADDOCK [14], PEP-FOLD 3 [15], DINC [16], and AutoDock CrankPep [17]. They can be categorized into template-based and template-free docking methods. Template-based docking methods use known protein-peptide complexes as templates to generate predictions of unknown protein-peptide complexes, whereas template-free docking methods perform searches for good fits between the given proteins and peptides without using templates. As expected, the quality of the predictions made by template-based methods depends on the existence of complex structures similar to those of the protein and peptide being investigated; if the coverage of the template database is wide enough, template-based methods can yield reliable results [18]. For this type of methods, the template database must be updated regularly to leverage recent discoveries in structural biology.

GalaxyPepDock is one of the most widely used template-based docking tools. It is free, easy to use, and accessible via a dedicated web server or downloadable precompiled binaries from <http://galaxy.seoklab.org/>. Unfortunately, the template database included in the GalaxyPepDock binary

package at present is based on an obsolete database called PepBind [19], which comprises structure coordinate files released before 2015. Approximately 10,000 new entries are appended to the Protein Data Bank (PDB) [20] annually, suggesting clear opportunities for the template database to expand significantly. By increasing the coverage of the template database, the quality of predictions made by GalaxyPepDock can be improved.

In this study, we present a protein-peptide complex database called Yet another protein-peptide complex database or YAPP-CD. It is made to be directly plugged into the GalaxyPepDock binary distribution and replace the native template database to improve GalaxyPepDock's prediction quality. It consists of 19,533 protein-peptide complexes with standard residues collected from PDB. The experimental results show that using our new database significantly reduced the average Ligand/Interface root-mean-square deviation (RMSD) from 7.560 Å / 3.620 Å to 3.469 Å / 1.712 Å, respectively. The database is freely available to download at <http://wnl.cs.hongik.ac.kr/yapp>.

2. Materials and Methods

Several protein-peptide complex databases have been reported in the literature. LEADS-PEP [21], PepPro [22], and PeptiDB [23] are small sets (with 100 or fewer entries) mostly utilized in the benchmarking of docking tools. Larger databases with 1,300–3,200 complexes, such as PepX [24], PepBind [14], and PixelDB [25], have been proposed, but they are no longer maintained and are thus obsolete. The template database distributed with GalaxyPepDock's binary package was based on PepBind. PepBDB [26] and Propedia [27] are recently released databases with over 13,000 complexes maintained to date. Some protein-peptide databases provide web interfaces for versatile functionalities, such as visualization and clustering, to help users understand protein-peptide interactions. In fact, protein-peptide complex datasets can also be derived from more comprehensive protein-ligand affinity databases such as BioLiP [28], because peptides can be considered a special type of ligand. BioLiP is comprised of 529,047 entries in total and 25,960 entries for peptide-type ligands (as of May 2021). Our proposed complex database YAPP-CD was built based on BioLiP. From the BioLiP annotation file, a list of complexes with peptide ligands was extracted, excluding peptides with non-standard residues, which resulted in 19,533 entries to be included in the database. Each complex entry consists of two coordinate files, one for the protein receptor and another for the peptide ligand, which are collected from BioLiP or PDB and manipulated to meet GalaxyPepDock's requirements; that is, residues are renumbered such that the residue number always starts with one. Peptides with non-standard residues were excluded from our database because some of them caused GalaxyPepDock execution failures. We noticed that there were peptide ligands with non-standard residues included in GalaxyPepDock's native template database; however, since GalaxyPepDock is distributed as a pre-compiled binary package, we were unable to debug and fix the problems with newly added non-standard residues. Therefore, we decided to exclude non-standard residue ligands from our database. In addition, it is unclear how non-standard residues contribute to the *interaction similarity* score-based template matching used in GalaxyPepDock. GalaxyPepDock generates complex structure predictions based on templates with the highest interaction similarity scores. Note that the *interaction similarity* score calculation makes use of the BLOSUM62 matrix, but it is not elucidated how residues that do not appear in the matrix—non-standard residues—are handled in the calculation [3].

GalaxyPepDock utilizes two kinds of molecular interactions found in each protein-peptide complex template to calculate the *interaction similarity* score: ionic and hydrophobic interactions. Information regarding the two types of molecular interactions for each protein-peptide complex is provided in our database. The interactions were identified using in-house Python scripts based on the following criteria:

hydrophobic interactions are presumed if hydrophobic residues (namely, ALA, VAL, LEU, ILE, MET, PHE, TRP, PRO, and TYR) fall within the 5 Å range, and ionic residues, that is, ARG, LYS, HIS, ASP, and GLU, are considered to contribute to ionic interactions if they fall within the 6 Å range.

3. Results

In this section, we demonstrate the advantage of YAPP-CD by comparing the docking qualities of GalaxyPepDock with YAPP-CD and its native template database in terms of Ligand-RMSD, Interface-RMSD, and the fraction of native contacts (Fnat). LRMSD/IRMSD/Fnat calculation and prediction quality classification (e.g., *acceptable*, *incorret*) were performed based on the Critical Assessment of Predicted Interactions (CAPRI) criteria (<http://www.ebi.ac.uk/msd-srv/capri/round28/round28.html>).

Table 1. Docking results of 56 PepPro complexes. (For Bound structures, each PDB target is indicated as PDB ID followed by the peptide chain ID and its receptor chain ID. For Unbound structures, each PDB ID is followed by the receptor chain ID only.)

Target PDB ID		YAPP-CD			Native Templates		
Bound	Unbound	LRMSD	IRMSD	Fnat	LRMSD	IRMSD	Fnat
5gtu_B:A	1w24_A	4.008	1.779	0.48	18.073	9.224	0.04
5fzt_B:A	2x0c_A	18.743	7.329	0.029	43.665	19.56	0
5epp_B:A	3hu3_A	1.859	1.292	0.686	2.511	1.627	0.514
5crw_B:A	2djk_A	4.052	2.089	0.591	25.824	11.136	0
4yl6_B:A	4ykd_A	2.238	1.52	0.545	15.754	9.704	0.091
4uwx_D:B	2bnx_A	1.843	0.944	0.821	2.581	0.852	0.607
4qqi_X:A	3so8_A	1.913	1.149	0.706	4.865	2.462	0.529
4qja_P:AB	3hvp_A	2.155	1.751	0.68	2.6	1.634	0.56
4q5u_C:A	4bw7_A	18.797	12.804	0.164	16.222	11.129	0.291
4oni_C:A	3tx7_B	3.757	1.495	0.52	3.565	1.335	0.56
4m5s_B:A	2y1y_A	2.099	0.941	0.852	10.202	5.262	0.148
4k0u_B:A	3utk_A	1.137	0.73	0.903	14.667	6.812	0.032
4j1v_G:A	4jiz_A	2.305	1.315	0.565	23.251	11.1	0.043
4htp_C:A	3w5o_A	1.657	1.087	0.727	17.772	7.187	0
4hh6_Z:A	4hh5_A	1.77	1.145	0.762	3.778	2.069	0.381
4eto_P:AB	2q91_AB	1.659	0.966	0.611	2.45	0.94	0.722
4dj9_B:A	1rke_A	5.515	3.599	0.238	5.244	3.653	0.381
3wg5_C:AB	3bpp_A	6.192	1.149	0.692	12.199	3.963	0.115
3ukx_C:B	4u5u_A	2.718	1.398	0.718	2.358	1.438	0.615
3ryb_B:A	3fto_A	5.547	2.532	0.34	5.954	2.884	0.34
3r7g_B:A	2ylf_A	1.154	2.218	0.8	0.759	2.358	0.743
3plv_C:A	1m94_A	2.873	1.355	0.65	31.633	11.413	0
3o37_E:A	4yat_A	1.938	0.759	0.9	2.633	0.766	0.867
3mhp_C:AB	1qfz_BA	6.332	2.344	0.5	19.608	8.59	0.068
3lu9_C:B	3u69_H	3.069	1.71	0.731	10.352	5.705	0.167
3kut_C:A	1g9l_A	3.599	1.87	0.469	8.292	1.952	0.406
3kj0_B:A	2mhs_A	1.806	0.871	0.762	2.681	0.951	0.714
3hbv_Z:P	3hb2_P	1.55	0.833	0.875	1.319	0.86	0.875

3h8a_E:AB	1e9i_AB	4.731	1.418	0.636	7.669	2.841	0.5
3dab_B:A	3lbj_E	1.191	1.056	0.88	1.427	1.185	0.84
3d9u_B:A	1qbh_A	4.266	2.348	0.438	4.505	2.486	0.438
3awr_C:A	5az9_A	3.26	1.388	0.8	3.571	1.52	0.733
3at0_B:A	3au0_A	1.64	0.839	0.84	1.641	0.835	0.86
2ybf_B:A	2yb6_A	1.484	1.119	0.684	9.849	4.977	0.184
2xum_S:A	3od4_A	2.136	1.442	0.698	2.297	1.944	0.581
2xs1_B:A	2oev_A	3.444	1.085	0.68	5.88	2.47	0.28
2qos_A:C	2ova_A	1.781	1.212	0.522	6.313	3.458	0.174
2qn6_C:A	4rjl_A	2.806	1.315	0.658	12.937	7.179	0.132
2pv1_B:A	1m5y_A	1.998	1.128	0.593	2.306	1.311	0.63
2puy_E:B	2yql_A	1.766	0.899	0.815	2.397	1.138	0.704
2peh_C:A	2pe8_A	2.937	0.691	0.625	5.094	2.343	0.5
2okr_C:A	3gc7_A	3.026	1.984	0.556	2.694	1.729	0.533
2nm1_B:A	1s0e_A	2.612	1.039	0.806	3.261	1.091	0.71
2hpl_B:A	2hpi_A	1.592	0.95	0.947	2.179	1.229	0.842
2g30_P:A	2iv9_A	2.604	0.967	0.704	3.949	1.329	0.704
2c5k_P:T	2c5j_A	9.999	1.732	0.424	6.709	1.979	0.485
2b1j_C:A	3rvq_A	4.37	3.118	0.435	4.311	3.063	0.435
1wkw_B:A	5gw6_A	3.912	2.073	0.643	5.365	2.456	0.643
1uhb_P:AB	2a31_A	6.021	2.498	0.308	6.129	2.606	0.231
1t0j_C:B	1t0h_B	0.744	0.481	0.867	0.828	0.608	0.933
1sqk_B:A	2q0u_A	2.24	1.075	0.706	2.605	1.062	0.676
1oai_B:A	1go5_A	3.861	0.889	0.682	3.795	1.184	0.682
1lb6_B:A	1lb4_A	1.976	1.165	0.733	1.953	0.995	0.733
1j2x_B:A	1i27_A	1.563	1.08	0.762	1.719	0.971	0.714
1gux_E:B	2r7g_C	1.734	0.859	0.654	1.815	1.032	0.615
1dkd_E:A	1kid_A	2.306	1.066	0.6	1.506	0.996	0.55
Average		3.469	1.712	0.643	7.599	3.618	0.462

As shown in Table 1, docking results for 56 PepPro [22] complexes demonstrated that YAPP-CD improved the prediction quality of GalaxyPepDock. By replacing GalaxyPepDock's native template database with YAPP-CD, the average Ligand-RMSD/Interface-RMSD was reduced from 7.599 Å / 3.618 Å to 3.469 Å / 1.712 Å, respectively, and Fn timer increased from 0.462 to 0.643. In addition, 47 (out of 56) complex structure predictions made by GalaxyPepDock with YAPP-CD could be classified as better than *acceptable* models according to the CAPRI classification, whereas only 30 predictions could be classified as better than *acceptable* models when docked with the native database. In other words, the number of successfully predicted complex structures using YAPP-CD was 17 more than the case without YAPP-CD. For instance, the best prediction made by GalaxyPepDock with YAPP-CD for the target PDB ID 5GTU showed an IRMSD of less than 2 Å and Fn timer greater than 0.2, which is considered an *acceptable* model; however, the best predicted structure made without YAPP-CD exhibited a very high IRMSD of 18 Å and a very low Fn timer of 0.04, which is considered an *incorrect* model. The other 16 targets successfully predicted only with YAPP-CD were 4YL6, 4QQI, 4M5S, 4K0U, 4J1V, 4HTP, 3WQ5, 3PLV, 3LU9, 3H8A, 2YBF, 2XS1, 2QOS, 2QN6, 2PEH, and 1WKW. In the docking experiments, 10 prediction models were generated by docking a specific peptide to the unbound structure and compared to the corresponding bound structure to calculate and find the best Ligand/Interface-RMSD for each target PDB ID listed in Table 1. (Details of peptide sequences are found in Appendix, Table A1.) In the PepPro benchmark database, 58 complexes were provided with their free form (or unbound) structures, but GalaxyPepDock failed to generate predictions with two targets, 5CQX and 3BEF, in our experiments. We used 2X0C instead of 5IC0 as the unbound structure for 5FZT because 2X0C is referred to as the free form of 5FZT in [29]. Note that 5IC0 is a triple-domain talin, whereas 5FZT and 2X0C are two-domain talin structures.

4. Discussion

Protein-peptide docking is a major tool for studying protein-peptide interactions, and several docking methods are currently available. Among the protein-peptide docking methods, template-based approaches such as GalaxyPepDock have been shown to yield more reliable results than template-free methods in recent studies. To obtain reliable results with a template-based docking method, the template database must be comprehensive enough and up-to-date to leverage recent advances in structural biology. In this paper, we present a protein-peptide complex database YAPP-CD, which is freely available at <http://wnl.cs.hongik.ac.kr/yapp>. YAPP-CD can be directly plugged into GalaxyPepDock to improve its prediction quality. Experimental results on a benchmark set of 56 targets showed that YAPP-CD improved GalaxyPepDock's prediction quality, e.g., average Ligand/Interface-RMSD was reduced from 7.60 Å /3.62 Å to 3.47 Å /1.71 Å. At present, complexes with non-standard residue-based ligand peptides are not included in YAPP-CD; however, there are numerous complexes with non-standard residues in PDB, and some of them are believed to serve as meaningful templates for predictions. Our immediate next step toward expanding YAPP-CD is to develop strategies for including peptides with non-standard residues in the database.

Appendix A

Table A1. Peptide sequences of 56 PepPro benchmark structures used in docking experiments.

PDB ID	Peptide Sequence	PDB ID	Peptide Sequence
5gtu_B:A	GQQDLMINNPLSQDEGSLWNKFFQDKE	3h8a_E:AB	QSPMPLTVAAASPELASGKVWIRYPIVR
5ftz_B:A	PELDDILYHVKGMRIVNQWSEK	3dab_B:A	SQETFSDLWKLLPEN
5epp_B:A	SPEEMRRQRLHRFDS	3d9u_B:A	AVPIAQ
5crw_B:A	GKTEGVLVYG	3awr_C:A	GPRLSRLSSAGC
4yl6_B:A	MDEQEALNSIMNDLVALQMNRR	3at0_B:A	GSWNSGSSGTGTGNQ
4uwx_D:B	TPRSARLERMAQALAQAGSP	2ybf_B:A	SKYRKKHKSEFQLLVQDARKGYKKIAG
4qqi_X:A	KAFVHMPTLPNLDHFHT	2xum_S:A	HLEVVKLLLEHGADVDAQDK
4qja_P:AB	RPGNFLQSRL	2xs1_B:A	SREKPYKEVTEDLLHLNSLF
4q5u_C:A	ARKEVIRNKIRAIKGMARVFSVLR	2qos_A:C	LYRDSTAERLY
4oni_C:A	QGAASRPAILYALLSSSLK	2qn6_C:A	SSEKEYVEMLDRLYSKLP
4m5s_B:A	GERTIPITRE	2pv1_B:A	WEYIPNV
4k0u_B:A	RTFRQVQSSISDFYD	2puy_E:B	ARTKQTARKS
4j1v_G:A	ALPAWARPDYNPPLVESWRR	2peh_C:A	KRKSrwDETP
4htp_C:A	DVPQWEVFFKR	2okr_C:A	IKIKKIEDASNPILLKRRKKARAL
4hh6_Z:A	KKWDSVYASLFKINLKK	2nm1_B:A	EDMFAKLKDKFFNEINK
4eto_P:AB	DAMNREVSSLKNKLRR	2hpl_B:A	DDLYG
4dj9_B:A	ETQVVLINAVKDVAKALGDLISATKAAAG	2g30_P:A	DDGLDEAFSRLAQSR
3wg5_C:AB	NVIVLMLPME	2c5k_P:T	KSLRVSSLNKDRRLREFYNLEN
3ukx_C:B	GSRRRRRRKRKREWDPPPKRRRLD	2b1j_C:A	MGDSILSQAEIDALLN
3ryb_B:A	SLSQSLSQS	1wkw_B:A	PGGTRIIYDRKFLMECRNSP
3r7g_B:A	KSlyKIKPRHDSGIKAKISMKT	1uhb_P:AB	GKDSCQGDS
3plv_C:A	SLSIETNELRASLGLKLIPP	1t0j_C:B	GAQQLEEDLKGyLDWITQAE
3o37_E:A	ARTKQTARKS	1sqk_B:A	DLPKVAENLSQLEGFNQDKLKNAS
3mhp_C:AB	KTEQPLSPYTAYDDLKPPSSPSPTKP	1oai_B:A	DSGFSFGSK
3lu9_C:B	ATNATLDPRSFLRNPNdkYEPFWE	1lb6_B:A	KQEPQEIDF
3kut_C:A	SNLNPNAAEFVPGVKYG	1j2x_B:A	SEADEMAKALEAELNDLM
3kj0_B:A	GSGGRPEIWYAQELRRIGDEFNAYYAR	1gux_E:B	DLYCYEQLN
3hbw_Z:P	AKASQAA	1dkd_E:A	SWMTTPWGLFHP

References

1. Petsalaki, E.; Russell, R. B. Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.* **2008**, *19*, 344–350, DOI: 10.1016/j.copbio.2008.06.004.
2. Muttenthaler M, King GF, Adams DJ, Alewood PF. Trends in peptide drug discovery. *Nature Reviews Drug Discovery* **2021**, 04/01;20(4):309-325.
3. Lee, H.; Heo, L.; Lee, M. S.; Seok, C. GalaxyPepDock: a protein-peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res.* **2015**, *43*, W431–W435, DOI: 10.1093/nar/gkv495.
4. De Vries, S.J.; Rey, J.; Schindler, C.E.M.; Zacharias, M.; Tuffery, P. The pepATTRACT web server for blind, large-scale peptide-protein docking. *Nucleic Acids Res.* **2017**, *45*, W361–W364, DOI: 10.1093/nar/gkx335.
5. Xu, X.J.; Yan, C.F.; Zou, X.Q. MDockPeP: An ab-initio protein-peptide docking server. *J. Comput. Chem.* **2018**, *39*, 2409–2413, DOI: 10.1002/jcc.25555.
6. Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43*, W419–W424, DOI: 10.1093/nar/gkv456.
7. Porter, K.A.; Xia, B.; Beglov, D.; Bohnuud, T.; Alam, N.; Schueler-Furman, O.; Kozakov, D. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. *Bioinformatics* **2017**, *33*, 3299–3301, DOI: 10.1093/bioinformatics/btx216.
8. Alam, N.; Goldstein, O.; Xia, B.; Porter, K.A.; Kozakov, D.; Schueler-Furman, O. High-resolution global peptide-protein docking using fragments-based PIPER-FlexPepDock. *PLoS Comput. Biol.* **2017**, *13*, e1005905, DOI: 10.1371/journal.pcbi.1005905.
9. Zhou, P.; Jin, B. W.; Li, H.; Huang, S. Y. HPEPDOCK: a web server for blind peptide-protein docking based on a hierarchical algorithm. *Nucleic Acids Res.* **2018**, *46*, W443–W450, DOI: 10.1093/nar/gky357.
10. Johansson-Åkhe, I.; Mirabello, C.; Wallner, B. InterPep2: global peptide-protein docking using interaction surface templates. *Bioinformatics* **2020** 04/15; 5/21;36(8):2458-2465.
11. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748, DOI: 10.1006/jmbi.1996.0897.
12. Jain, A.N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306, DOI: 10.1007/s10822-007-9114-2.
13. Antes, I. DynaDock: A new molecular dynamics-based algorithm for protein-peptide docking including receptor flexibility. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1084–1104, DOI: 10.1002/prot.22629.
14. Trellet, M.; Melquiond, A. S. J.; Bonvin, A. M. J. J. A Unified Conformational Selection and Induced Fit Approach to Protein-Peptide Docking. *PLoS One* **2013**, *8*, e58769, DOI: 10.1371/journal.pone.0058769.
15. Lamiabile, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuffery, P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44*, W449–W454, DOI: 10.1093/nar/gkw329.
16. Antunes, D.A.; Moll, M.; Devaurs, D.; Jackson, K.R.; Lizée, G.; Kavrakı, L. E. DINC 2.0: A New Protein-Peptide Docking Webserver Using an Incremental Approach. *Cancer Res.* **2017**, *77*, e55–e57, DOI: 10.1158/0008-5472.CAN-17-0511.
17. Zhang, Y.; Sanner, M. AutoDock CrankPep: Combining folding and docking to predict protein-peptide complexes. *Bioinformatics* **2019**, *35*, 5121–5127, DOI: 10.1093/bioinformatics/btz459.
18. Weng, G.; Gao, J.; Wang, Z.; Wang, E.; Hu, X.; Yao, X.; et al. Comprehensive evaluation of fourteen docking programs on protein-peptide complexes. *J Chem Theory Comput.* **2020**;16(6):3959–69.
19. Das, A.A.; Sharma, O.P.; Kumar, M.S.; Krishna, R.; Mathur, P.P. Pepbind: a comprehensive database and computational tool for analysis of protein-peptide interactions. *Genom Proteom Bioinform.* **2013**;11(4):241–6.
20. Rose, P.W.; Beran, B.; Bi, C.; Bluhm, W.F.; Dimitropoulos D, Goodsell DS, Prlić A, Quesada M, Quinn GB, Westbrook JD, et al. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.* **2011**;39:D392–D401.
21. Hauser, A.S.; Windshugel, B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J. Chem. Inf. Model.* **2016**, *56*, 188–200, DOI: 10.1021/acs.jcim.5b00234.
22. Xu, X.; Zou, X. Peppro: a nonredundant structure data set for benchmarking peptide-protein computational docking. *J Comput Chem.* **2020**;41:362–9.
23. London, N.; Movshovitz-Attias, D.; Schueler-Furman, O. The structural basis of peptide-protein binding strategies. *Structure.* **2010**;18(2):188–99.

24. Vanhee, P.; Reumers, J.; Stricher, F.; Baeten, L.; Serrano, L.; Schymkowitz, J.; Rousseau, F. Pepx: a structural database of non-redundant protein-peptide complexes. *Nucleic Acids Res.* **2010**;38(suppl-1):545–51.
25. Frappier, V.; Duran, M.; Keating, AE. Pixeldb: protein-peptide complexes annotated with structural conservation of the peptide binding mode. *Protein Sci.* **2018**;27(1):276–85.
26. Wen, Z.; He, J.; Tao, H.; Huang, S-Y. Pepbdb: a comprehensive structural database of biological peptide-protein interactions. *Bioinformatics.* **2019**;35(1):175–7.
27. Martins, P.M.; Santos, L.H.; Mariano, D.; Queiroz, F.C.; Bastos, L.L.; Gomes, I.d.S.; et al. Propedia: a database for protein-peptide identification based on a hybrid clustering algorithm. *BMC Bioinformatics* **2021**, 01/02;22(1):1.
28. Yang, J.; Roy, A.; Zhang, Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res* **2013**, 01;41:D1096-D1103.
29. Zacharchenko, T.; Qian, X.; Gault, B.T.; Jethwa, D.; Almeida, T.B.; Ballestrem, C.; et al. LD Motif Recognition by Talin: Structure of the Talin-DLC1 Complex. *Structure* **2016**, 6 July 2016;24(7):1130-1141.