

RESEARCH

Open Access



A protocol for CABS-dock protein–peptide docking driven by side-chain contact information

Mateusz Kurcinski^{1†}, Maciej Blaszczyk^{1†}, Maciej Pawel Ciemny^{1,2}, Andrzej Kolinski¹ and Sebastian Kmiecik^{1*}

From 4th International Work-Conference on Bioinformatics and Biomedical Engineering-IWBBIO 2016 Grenada, Spain. 20–22 April 2016

*Correspondence:

sekmi@chem.uw.edu.pl

[†]Mateusz Kurcinski and Maciej Blaszczyk contributed equally to this work

¹ Faculty of Chemistry, Biological and Chemical Research Center, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland
Full list of author information is available at the end of the article

Abstract

Background: The characterization of protein–peptide interactions is a challenge for computational molecular docking. Protein–peptide docking tools face at least two major difficulties: (1) efficient sampling of large-scale conformational changes induced by binding and (2) selection of the best models from a large set of predicted structures. In this paper, we merge an efficient sampling technique with external information about side-chain contacts to sample and select the best possible models.

Methods: In this paper we test a new protocol that uses information about side-chain contacts in CABS-dock protein–peptide docking. As shown in our recent studies, CABS-dock enables efficient modeling of large-scale conformational changes without knowledge about the binding site. However, the resulting set of binding sites and poses is in many cases highly diverse and difficult to score.

Results: As we demonstrate here, information about a single side-chain contact can significantly improve the prediction accuracy. Importantly, the imposed constraints for side-chain contacts are quite soft. Therefore, the developed protocol does not require precise contact information and ensures large-scale peptide flexibility in the broad contact area.

Conclusions: The demonstrated protocol provides the extension of the CABS-dock method that can be practically used in the structure prediction of protein–peptide complexes guided by the knowledge of the binding interface.

Keywords: Protein–peptide interactions, Molecular docking, Flexible docking, Protein–peptide complexes

Background

The prediction of protein–peptide complexes is a demanding modeling challenge, particularly when significant conformational changes occur in the binding process. The modeling of large-scale dynamics during binding cannot be effectively performed with standard simulation tools of all-atom resolution. A significant speed-up in flexible docking simulations can be achieved using coarse-grained protein models [1]. The

CABS-dock is a method based on a coarse-grained model that is one of the most effective approaches to the simulations of large conformational changes during protein binding [1–3]. The CABS-dock is available as a web server [4–6]. The method doesn't use any knowledge about peptide structure or a peptide binding site. Additional information on the protein–peptide interaction interface (obtained from experiments or theoretical predictions) may significantly improve the docking accuracy [7]. For example, the majority of state-of-the-art protein–peptide docking tools, like Rosetta FlexPepDock [8] or HADDOCK [9], follow the data-driven docking paradigm. The Rosetta FlexPepDock method enables selection of the “anchoring residue”, a residue that will be constrained during simulation on a given anchoring position. On the other hand, the HADDOCK approach uses so-called “ambiguous interaction restraints” that label receptor residues as “active” or “passive” in peptide binding.

In the CABS-dock method, the most intuitive way to introduce information about protein–peptide contact(s) is to apply distance constraint(s) on a chosen residue pair during the simulation. The side-chain contact information may be derived either directly from structural experiments or with bioinformatics tools. The possible approaches include binding site prediction [10], similarity based docking [11] or analysis of protein sequence co-evolution [12]. In this work, we present a strategy for incorporating the information on protein–peptide side-chain interactions into the CABS-dock procedure. The developed protocol for docking driven by side-chain contact information leads to a significant improvement in modeling accuracy as compared with CABS-dock docking in the default mode.

Methods

CABS model

The CABS-dock uses a CABS coarse-grained protein model for flexible docking simulations. The main features of the CABS model (described in detail elsewhere [13] and also in recent review [1]) are summarized below:

1. Coarse-grained representation of molecules: each amino acid residue is represented by three pseudo-atoms: Carbon Alpha ($C\alpha$), carbon Beta and the Side-chain. To mimic the peptide bond, the fourth center of interactions is defined in the geometrical center of the virtual $C\alpha$ – $C\alpha$ bond. Positions of the $C\alpha$ atoms are restricted to the cubic lattice, whereas other pseudoatoms are placed off the lattice.
2. Statistical force field: the energy of the complex models is related to the frequency of interactions observed in already solved structures available in the PDB [14];
3. Sampling of the configurational space is controlled by the Replica Exchange Monte Carlo scheme.

Such a design of the CABS model leads to significant simulation speed-up, by three to four orders of magnitude with regard to all-atom molecular dynamics. At the same time, reasonable resolution of modeled structures is preserved, as coarse-grained models may be easily rebuilt to realistic all-atom representation. The CABS model was successfully applied to a variety of modeling tasks including: protein structure prediction [15, 16],

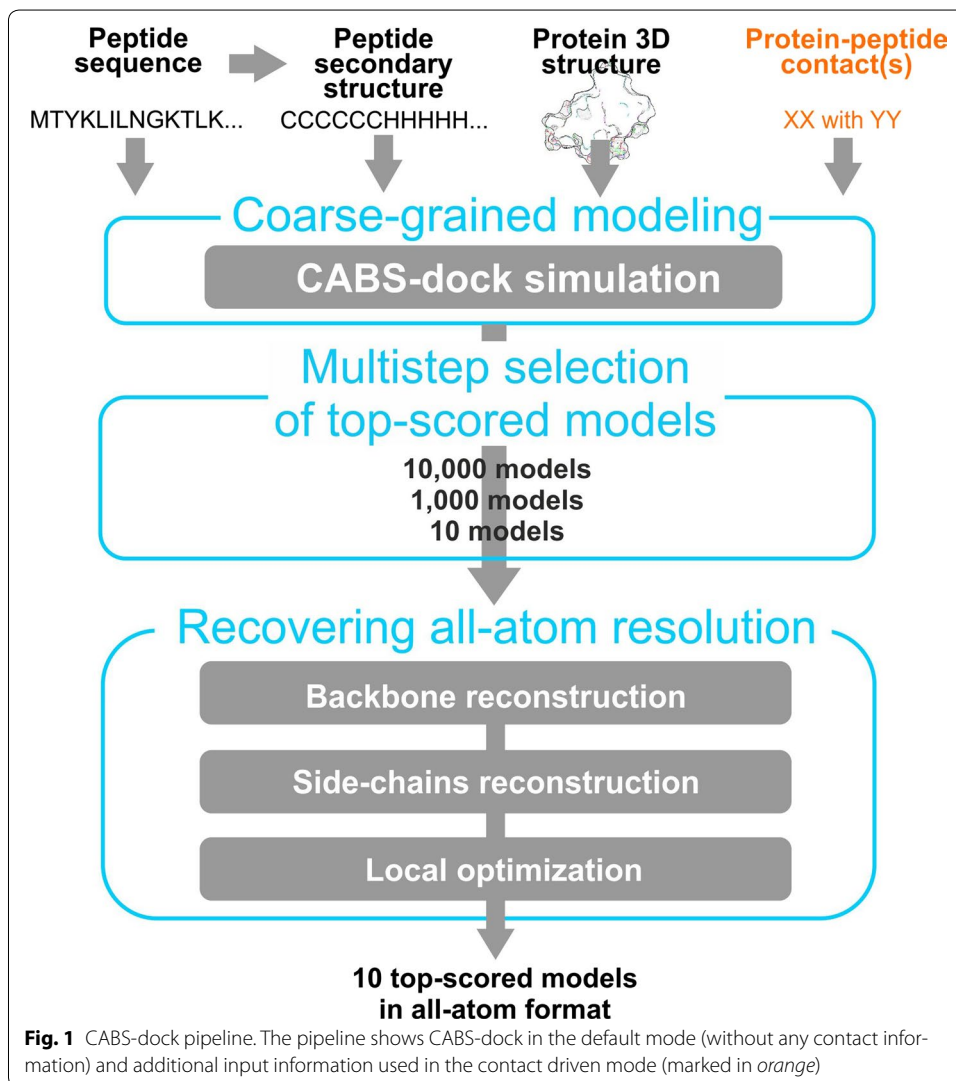
simulations of folding mechanisms [17–20], flexibility of globular proteins [21, 22] and modeling of protein–protein and protein–peptide complexes [23–28].

CABS-dock docking procedure

The pipeline of the CABS-dock method for protein–peptide docking [4–6] is presented in Fig. 1. The modeling procedure consists of four steps: (1) initial setup, (2) coarse-grained simulation, (3) model selection and (4) model refinement.

Initial setup

In the initial setup the receptor structure is translated into coarse-grained representation. Subsequently, ten copies of the peptide in random conformations are generated for the replica exchange method and also transformed into coarse-grained representation. As in the default docking mode, random peptide conformations are randomly scattered around the receptor at distances up to 20 Å from the receptor molecular surface. The



information about the side-chain contacts between the receptor and the peptide is transformed into soft distance restraints imposed on the modeled molecules.

Simulation

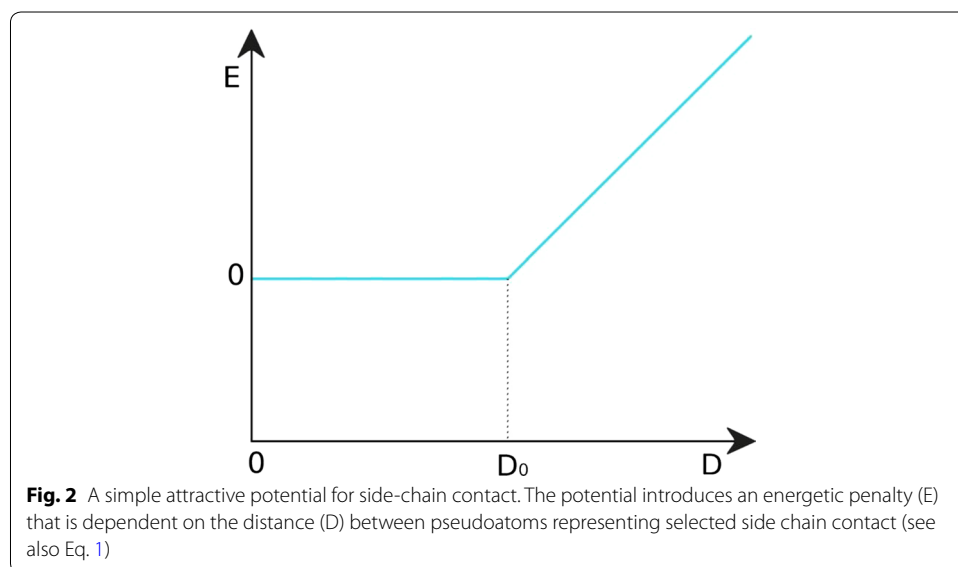
The coarse-grained simulation of the system is carried out in ten copies at different temperatures, with the exchange of coordinates between copies every given number of simulation cycles. The peptide molecule is fully flexible during the docking simulation. In the contact-driven mode of the CABS-dock method, we introduced a simple contact potential described by the following formula:

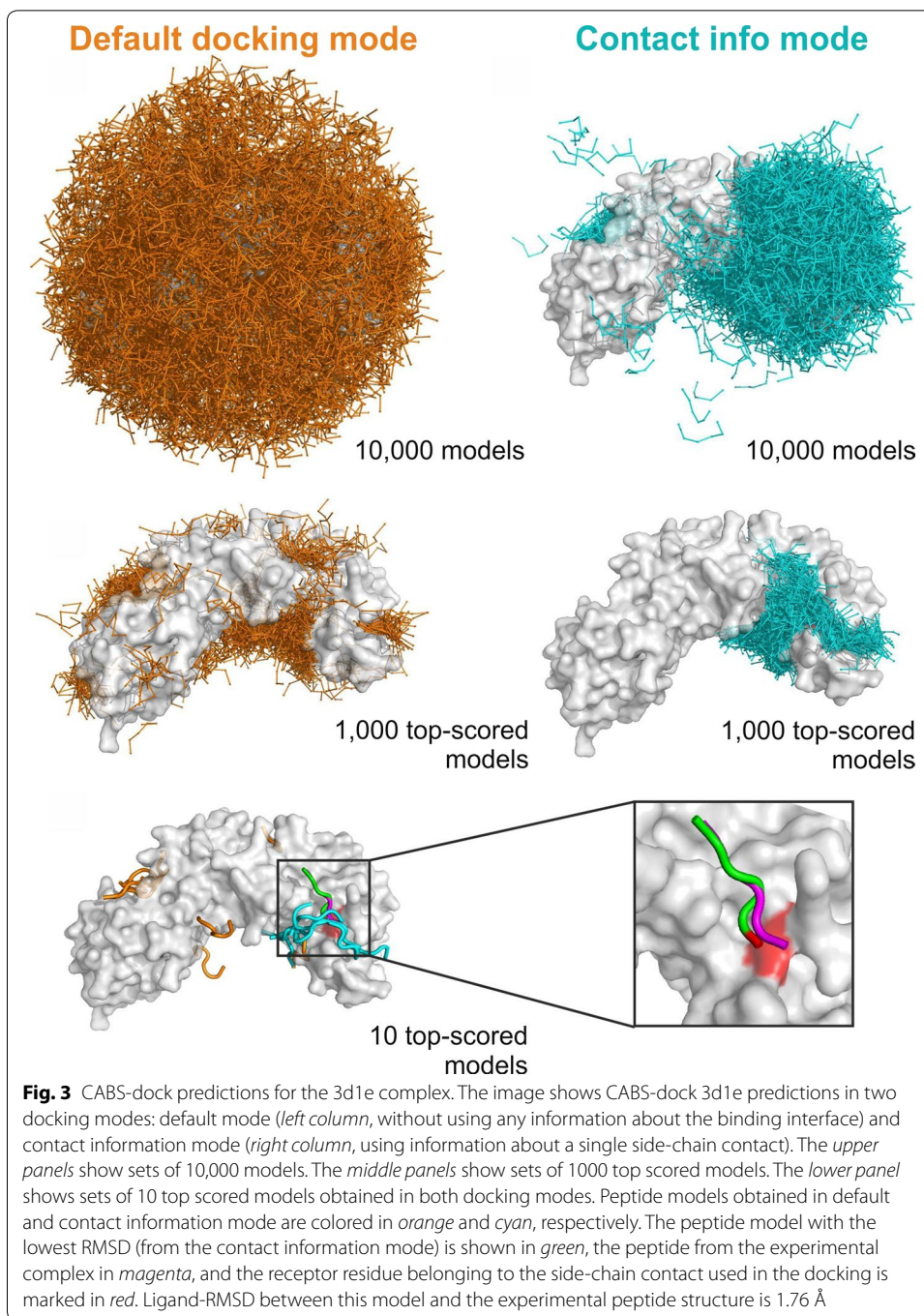
$$E = \begin{cases} 0 & \text{if } D \leq D_0 \\ s(D - D_0) & \text{if } D > D_0 \end{cases} \quad (1)$$

where D is the observed distance between pseudoatoms representing side chains, D_0 is the distance below which the potential vanishes and s is the slope of the potential line. This potential is also depicted in Fig. 2. Its role is to draw the ligand molecule to the binding site, but not to contribute to the final conformational energy of the complex. As in the default CABS-dock modeling mode [4, 5] the receptor molecule is also flexible, both on the side-chain and backbone level, but kept in near native conformation by distance restraints.

Model selection

CABS-dock simulation provides 10,000 alternative models of the complex. From this set the 1000 top scored complexes (with the lowest CABS interaction energy) are selected for the next step. Final selection is done by clustering the 1000 models using the k-medoid procedure with $k = 10$ and ligand RMSD (root mean square deviation of peptide coordinates after superposition of receptor molecules) as the measure of model similarity. The medoids from each cluster are selected for the next step as 10 top ranked





models. The ranking from 1st to 10th is based on cluster density values (number of cluster models divided by their average difference within a cluster). Figure 3 shows consecutive stages of model selection.

Refinement

Finally, 10 top ranked models are reconstructed to all-atom representation. For this task, CABS-dock method uses an automated Modeller procedure [29].

Results and discussion

We tested the developed protocol (for driving CABS-dock docking with side-chain(s) contact information) on several protein–peptide complexes from previous CABS-dock tests (without contact information, default docking settings) [4, 5]. The results, together with comparison (default docking vs. docking with contact information) are presented in Table 1. In each case, a single protein–peptide contact for driving the docking was chosen randomly (see Table 2). The parameters of the attractive potential for side-chain contacts (see Eq. 1) were set as: $D_0 = 5.0$ Angstroms, $s = 1.0$. Like in our previous CABS-dock tests [4, 5], preferred secondary structures of the peptides were taken from the native structures of the complexes.

For most of the docking cases, we noted significant improvement (see Table 1). One of the cases (PDB ID: 3d1e) is shown in Fig. 3. For this test case, in the default CABS-dock mode (without any contact information), the accuracy of predictions in the set of ten top-scored models was very low (RMSD¹⁰ was 18.82 Å, the peptides are shown in

Table 1 Comparison of CABS-dock docking performance without (default) vs. with information about a randomly selected contact

PDB	Docking without contact information (default CABS-dock settings)				Docking driven by random contact information			
	RMSD ^{10k}	RMSD ^{1k}	RMSD ¹⁰⁰	RMSD ¹⁰	RMSD ^{10k}	RMSD ^{1k}	RMSD ¹⁰⁰	RMSD ¹⁰
2v3s	2.42	2.42	3.48	8.89	1.30	1.37	1.65	1.77
2vj0	2.09	2.96	4.12	3.91	2.71	3.00	3.00	3.40
2zjd	2.35	2.69	2.79	4.60	1.77	2.03	2.03	3.06
3bfq	10.20	11.53	14.22	13.48	1.47	1.47	2.34	2.89
3bu3	6.87	7.06	7.71	8.86	3.62	4.45	5.33	5.47
3bwa	2.42	2.75	3.17	3.94	2.00	2.32	2.36	3.55
3cvp	4.52	4.67	8.47	10.14	2.37	2.98	3.91	4.29
3d1e	4.39	6.59	8.18	18.82	1.76	1.76	2.01	1.76
3d9t	3.56	4.34	7.25	10.06	1.96	2.69	3.42	3.72

The table shows RMSD values showing the lowest RMSD value from: 10,000 CABS-dock models (RMSD^{10k}), 1000 top-scored CABS-dock models (RMSD^{1k}), 100 top-scored CABS-dock models (RMSD¹⁰⁰), 10 top-scored CABS-dock models (RMSD¹⁰)

Table 2 Input data for CABS-dock protein–peptide docking using information about side-chain contacts

PDB	Receptor chain	Peptide sequence	ID of contact residues	
			Receptor	Peptide
2v3s	B	GRFQVT	449	4
2vj0	A	PKGWVTFE	782	3
2zjd	A	GGDDDWTHLS	35	9
3bfq	G	ADSTITIRGYVRDNR	117	5
3bu3	A	YNPYPEDYGDIEIG	1181	11
3bwa	A	FPTKDVAL	66	1
3cvp	A	NRASKL	557	4
3d1e	A	GQLGLF	364	1
3d9t	B	ATPFQE	307	4

Single contacts were randomly selected from native contacts (defined using 5 Å distance cut-off based on positions of heavy atoms)

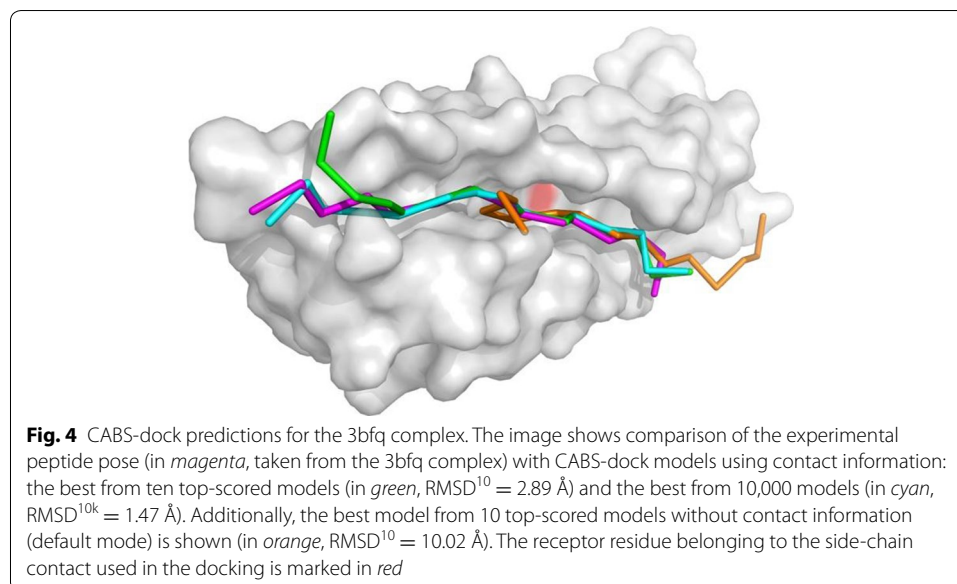
orange). Side-chain contact information enables restraining the conformational sampling of a peptide to the broad neighborhood of the contact. This resulted in the selection of 10 top-scored peptides that were much closer to the binding site than in the default docking mode.

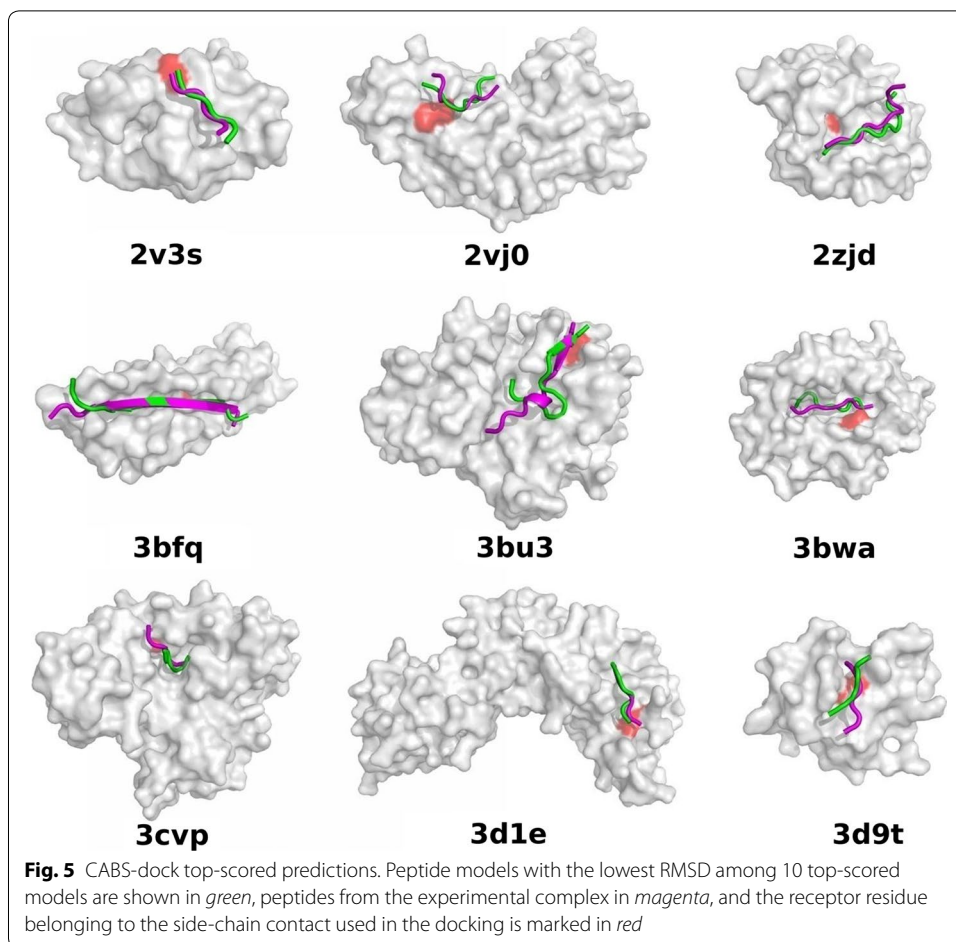
Another docking example of 3bfq complex is presented in Fig. 4. In this case, a much longer peptide (15 residues) was docked. As compared to docking in the default mode, the use of contact information enabled significant improvement of the docking accuracy, however, there is still room for improvement. Namely, the lowest RMSD model out of the 10,000 models is much more accurate than that out of the 10 top-scored models, which is also the case for other modeled complexes (see Table 1). CABS-dock top-scored predictions for all testing cases are presented in Fig. 5.

Conclusions

The accurate characterization of protein–peptide interfaces is important for understanding the molecular basis of life and rational design of peptide therapeutics [30]. Also, the lessons learnt from protein–peptide molecular docking can be extremely valuable in addressing important questions regarding the modeling of protein–protein interactions [31, 32].

In this work we demonstrated how very sparse and easily accessible data may improve structure prediction of protein–peptide complexes with the CABS-dock method. We introduced a simple protocol that transforms information about expected protein–peptide contacts into soft restraints, which enable extensive sampling of the peptide conformational space in a large area around the defined contact. Further development of the protocol will provide a promising tool for high-throughput studies, incorporated into a publicly available CABS-dock server. Our protocol can be easily combined with other bioinformatics tools, for contact prediction, or with experimental data [7]. The former is an especially promising approach as there already are numerous methods





that could be incorporated in such a pipeline (for example binding site prediction tools [7, 33, 34]). Additional improvements can be achieved using better scoring and selection procedures that would be able to fish out the best accuracy peptide models out of a large set of CABS-dock predictions. This can be done in various ways, for example, using external force-fields (e.g. all-atom molecular dynamics [5]) or machine learning approaches [35].

Declarations

Authors' contributions

MK designed and wrote the software used in the study, MB performed the modeling. All authors contributed to the analysis of the data, writing and editing the manuscript. All authors read and approved the final manuscript.

Author details

¹ Faculty of Chemistry, Biological and Chemical Research Center, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland. ² Faculty of Physics, University of Warsaw, Pasteura 5, Warsaw 02-093, Poland.

Acknowledgements

The authors acknowledge support from the National Science Center (NCN, Poland) Grant [MAESTRO2014/14/A/ST6/00088].

Competing interests

The authors declare that they have no competing interests.

Funding

Publication of this article was funded by the the National Science Center (NCN, Poland) Grant [MAESTRO2014/14/A/ST6/00088].

About this supplement

This article has been published as part of *BioMedical Engineering OnLine* Volume 16 Supplement 1, 2017: Selected articles from the 4th International Work-Conference on Bioinformatics and Biomedical Engineering-IWBBIO 2016. The full contents of the supplement are available online at <https://biomedical-engineering-online.biomedcentral.com/articles/supplements/volume-16-supplement-1>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Published: 18 August 2017

References

1. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained protein models and their applications. *Chem Rev*. 2016;116:7898–936.
2. Ciemny MP, Debinski A, Paczkowska M, Kolinski A, Kurcinski M, Kmiecik S. Protein-peptide molecular docking with large-scale conformational changes: the p53-MDM2 interaction. *Sci Rep*. 2016;6:37532.
3. Antunes DA, Didier D, Kavrazi LE. Understanding the challenges of protein flexibility in drug design. *Expert Opin Drug Discov*. 2015;10:1301–13.
4. 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:W419–24.
5. Blaszczyk M, Kurcinski M, Kouza M, Wieteska L, Debinski A, Kolinski A, et al. Modeling of protein-peptide interactions using the CABS-dock web server for binding site search and flexible docking. *Methods*. 2016;93:72–83.
6. Ciemny MP, Kurcinski M, Kozak JK, Kolinski A, Kmiecik K. Highly flexible protein-peptide docking using CABS-Dock. *Methods Mol Biol*. 2017;1561:69–94.
7. Trellet M, Melquiond ASJ, Bonvin AMJJ. Information-driven modeling of protein-peptide complexes. *Methods Mol Biol*. 2015;1268:221–39.
8. London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O. Rosetta FlexPepDock web server—high resolution modeling of peptide–protein interactions. *Nucleic Acids Res*. 2011;39:W249–53.
9. Trellet M, Melquiond ASJ, Bonvin AMJJ. A unified conformational selection and induced fit approach to protein-peptide docking. *PLoS ONE*. 2013;8:e58769.
10. Petsalaki E, Evangelia P, Alexander S, Eduardo G-U, Russell RB. Accurate prediction of peptide binding sites on protein surfaces. *PLoS Comput Biol*. 2009;5:e1000335.
11. Lee H, Hasup L, Lim H, Lee MS, Chaok S. GalaxyPepDock: a protein–peptide docking tool based on interaction similarity and energy optimization. *Nucleic Acids Res*. 2015;43:W431–5.
12. Hopf TA, Schärfe CPI, Rodrigues JPGLM, Green AG, Kohlbacher O, Sander C, et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *Elife*. 2014;3:e03430. doi:10.7554/eLife.03430.
13. Kolinski A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim Pol*. 2004;51:349–71.
14. Kmiecik S, Kolinski A. One-dimensional structural properties of proteins in the coarse-grained CABS model. *Methods Mol Biol*. 2017;1484:83–113.
15. Blaszczyk M, Jamroz M, Kmiecik S, Kolinski A. CABS-fold: server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Res*. 2013;41:W406–11.
16. Kmiecik S, Jamroz M, Kolinski M. Structure prediction of the second extracellular loop in G-protein-coupled receptors. *Biophys J*. 2014;106:2408–16.
17. Kmiecik S, Kolinski A. Characterization of protein-folding pathways by reduced-space modeling. *Proc Natl Acad Sci USA*. 2007;104:12330–5.
18. Kmiecik S, Kolinski A. Folding pathway of the b1 domain of protein G explored by multiscale modeling. *Biophys J*. 2008;94:726–36.
19. Kmiecik S, Jamroz M, Kolinski A. Multiscale approach to protein folding dynamics. In: Kolinski A, editor. *Multiscale Approaches to Protein Modeling*. New York: Springer; 2011. p. 281–293.
20. Jamroz M, Kolinski A, Kmiecik S. Protocols for efficient simulations of long-time protein dynamics using coarse-grained CABS model. *Methods Mol Biol*. 2014;1137:235–50.
21. Jamroz M, Kolinski A, Kmiecik S. CABS-flex: server for fast simulation of protein structure fluctuations. *Nucleic Acids Res*. 2013;41:W427–31.
22. Jamroz M, Kolinski A, Kmiecik S. CABS-flex predictions of protein flexibility compared with NMR ensembles. *Bioinformatics*. 2014;30:2150–4.
23. Kurcinski M, Kolinski A. Steps towards flexible docking: modeling of three-dimensional structures of the nuclear receptors bound with peptide ligands mimicking co-activators' sequences. *J Steroid Biochem Mol Biol*. 2007;103:357–60.
24. Kurcinski M, Kolinski A. Hierarchical modeling of protein interactions. *J Mol Model*. 2007;13:691–8.
25. Kurcinski M, Kolinski A. Theoretical study of molecular mechanism of binding TRAP220 coactivator to Retinoid X Receptor alpha, activated by 9-cis retinoic acid. *J Steroid Biochem Mol Biol*. 2010;121:124–9.
26. Horwacik I, Kurcinski M, Bzowska M, Kowalczyk AK, Czaplicki D, Kolinski A, et al. Analysis and optimization of interactions between peptides mimicking the GD2 ganglioside and the monoclonal antibody 14G2a. *Int J Mol Med*. 2011;28:47–57.

27. Steczkiewicz K, Zimmermann MT, Kurcinski M, Lewis BA, Dobbs D, Kloczkowski A, et al. Human telomerase model shows the role of the TEN domain in advancing the double helix for the next polymerization step. *Proc Natl Acad Sci USA*. 2011;108:9443–8.
28. Kurcinski M, Kolinski A, Kmiecik S. Mechanism of folding and binding of an intrinsically disordered protein as revealed by ab initio simulations. *J Chem Theory Comput*. 2014;10:2224–31.
29. Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinform*. 2016;54:5.6.1–6.37.
30. Fosgerau K, Hoffmann T. Peptide therapeutics: current status and future directions. *Drug Discov Today*. 2015;20:122–8.
31. Verschueren E, Vanhee P, Rousseau F, Schymkowitz J, Serrano L. Protein-peptide complex prediction through fragment interaction patterns. *Structure*. 2013;21:789–97.
32. London N, Raveh B, Movshovitz-Attias D, Schueler-Furman O. Can self-inhibitory peptides be derived from the interfaces of globular protein–protein interactions? *Proteins*. 2010;78:3140–9.
33. Lavi A, Ngan CH, Movshovitz-Attias D, Bohnuud T, Yueh C, Beglov D, et al. Detection of peptide-binding sites on protein surfaces: the first step toward the modeling and targeting of peptide-mediated interactions. *Proteins*. 2013;81:2096–105.
34. Saladin A, Rey J, Thévenet P, Zacharias M, Moroy G, Tufféry P. PEP-SiteFinder: a tool for the blind identification of peptide binding sites on protein surfaces. *Nucleic Acids Res*. 2014;42:W221–6.
35. Ain QU, Aleksandrova A, Roessler FD, Ballester PJ. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip Rev Comput Mol Sci*. 2015;5:405–24.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

