Forschungszentrum Jülich GmbH
Institute for Advanced Simulation (IAS)
Jülich Supercomputing Centre (JSC)

From Computational Biophysics to Systems Biology (CBSB11) – Celebrating Harold Scheraga's 90th Birthday

Proceedings, 20 – 22 July 2011 | Jülich, Germany

edited by Paolo Carloni, Ulrich H.E. Hansmann, Thomas Lippert, Jan H. Meinke, Sandipan Mohanty, Walter Nadler, Olav Zimmermann

Schriften des Forschungszentrums Jülich
IAS Series Volume 8

ISSN 1868-8489 ISBN 978-3-89336-748-1

Predicting Protein-protein Interactions with DrugScore^{PPI}: Docking, Scoring, and *in silico* Alanine-scanning

Dennis M. Krüger[1], José Ignacio Garzón[2], Pablo Chacón[2], and Holger Gohlke[1]

[1] Department of Mathematics and Natural Sciences, Heinrich-Heine-University
40225 Düsseldorf, Germany
E-mail: gohlke@uni-duesseldorf.de

[2] Department of Biological Chemical Physics, Rocasolano Physical Chemistry Institute, CSIC
Serrano 119, Madrid 28006, Spain

Protein-protein complexes play key roles in many cellular processes. Therefore, knowledge of the 3D structure of protein-complexes is of fundamental importance. A key goal in protein-protein docking is to identify near-native protein-complex structures using an appropriate scoring function. In this work, we tested our recently developed knowledge-based approach DrugScore^{PPI} (I) for identifying hotspots in protein-protein interfaces, (II) as a scoring function for rescoring a large dataset of pre-generated protein-protein docking decoys, and (III) as an objective function in combination with a fast spherical harmonics-based protein-protein docking tool for predicting 3D structures of protein-protein complexes.

1 Introduction

Protein-protein interactions have important implications in most cellular signaling processes. Therefore, protein interfaces become more and more important as drug targets[1]. Several studies were published addressing the properties of binding sites in protein surface areas and small molecule inhibitors targeting these protein interaction sites[2,3]. From these, it is known that protein-protein complex formation frequently relies on a few interface residues (hotspots) that account for most of the binding free energy. Various protein-protein docking algorithms have been developed during the last years that allow predicting the binding mode of two protein complex partners[4,5]. There are two main aspects in protein-protein docking: (I) Searching for possible docking configurations and (II) assessing each predicted configuration with a scoring function. For the latter, a fast and accurate scoring function is required that identifies a (the) pose that comes close (the closest) to the native structure on the first (on one of the first) scoring rank(s). Despite many advances in the field of protein-protein docking in the past years, the results are still not satisfying: Whereas the generation of several thousands of docking solutions is just a question of computational resources available, the identification of near-native complex structures by a scoring function is still challenging.

In this work, we present a fast and accurate computational approach to predict protein-protein interactions. The approach is based on DrugScore^{PPI}, a knowledge-based scoring function for which pair potentials were derived from 851 complex structures and adapted against 309 experimental alanine scanning results. In our hands, the DrugScore approach has been proven successful already for scoring and predicting protein-ligand and protein-RNA complexes[6,7]. In part, this has been attributed to the implicit, well-balanced consideration of several different types of interactions. Obtaining such a delicate balance is

also considered crucial for successfully predicting protein-protein complexes[8]. To the best of our knowledge, an aspect that has never been addressed is whether these scoring functions are able to predict the change in binding free energy upon single alanine mutations (alanine scanning). Currently, there are only a few computational approaches that address the prediction of hotspots in protein interfaces. Three well-known approaches, FoldX[9], MM/PBSA[10] and the recently published CC/PBSA[11] method, were used for comparison to our QSAR-adapted scoring function DrugScore^{PPI} [12].

2 Results and Discussion

2.1 *In silico* Alanine-scanning

DrugScore^{PPI} was used for computational alanine-scanning on a dataset of 18 protein-protein complexes with a total of 309 mutations to predict changes in the binding free energy upon mutations in the interface[12]. Computed and experimental values showed a correlation of r^2 = 0.58. To improve the predictive power, a QSAR-model was built based on 24 residue-specific atom types. This improves the correlation to r^2 = 0.73, with a root mean square deviation of 1.23 kcal/mol. A Leave-One-Out analysis yields a correlation coefficient of q^2 = 0.64. For further validation, alanine-scanning with DrugscorePPI was performed on two pharmaceutically important systems, Ras and Raps, which signal to a number of distinct pathways by interacting with diverse downstream effectors. We note that these two systems were not included in the QSAR training set. The results (Tab. 1) demonstrate that DrugscorePPI outperforms other state-of-the-art methods not only with respect to predictive power but also in terms of computational times of 3 seconds per residue on a standard CPU[12].

Method	Ras/RalGDS	Raps/Raf
DrugScore^{PPI}	**0.62**	**0.45**
MM-PBSA	0.46	n.a.
FoldX	0.52	0.07
CC/PBSA	0.23	0.22

Table 1. Correlation coefficients of predicted vs. experimental relative binding free energies for the two external test datasets.

Based on these findings, we developed the DrugScore^{PPI} webserver[12], accessible at `http://cpclab.uni-duesseldorf.de/dsppi`, which allows identifying hotspot residues in protein-protein interfaces and performing computational alanine scanning of a protein-protein interface within a few minutes.

2.2 Rescoring of Protein-protein Docking Decoys

When used as a scoring function to evaluate decoys of a non-redundant dataset of 54 protein-protein complexes for which ”unbound perturbation” solutions have been generated[5], funnel-shaped DrugScore^{PPI} score vs. rmsd curves were obtained for the majority

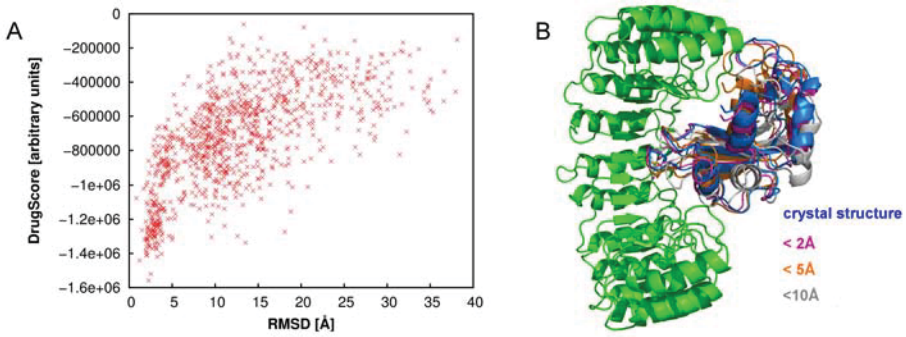

Figure 1. (A) Energy funnel for rescoring docking decoys generated for protein-protein complexes of porcine kallikrein A / bovine pancreatic trypsin inhibitor. (B) Comparison of decoy quality from docking a ribonuclease inhibitor to ribonuclease A. The RMSD is given with respect to the native structure (blue).

of cases (Fig. 1). Accordingly, DrugScore^{PPI} was able to rank a near-native solution, i.e. rmsd $<$ 10 Å, in the top ten (five) in 94.1% (64.8%) of the cases. When applied to a dataset of "unbound docking" solutions, DrugScore^{PPI} was able to rank a near-native solution in the top ten (five) in 100% (73.3%) of the cases. These results compare favorably with those obtained by Baker et al.[5].

2.3 Protein-protein Docking

Finally, DrugScore^{PPI} was applied as an objective function in FRODOCK[4] in order to predict 3D structures of protein-protein complexes. For this, precalculated knowledge-based potential grids by DrugScore^{PPI} were used to sample protein-protein configurations and to identify near-native complex configurations (Fig. 1). When applied to a subset of 97 bound-bound test cases of the ZDOCK benchmark 3, convincing results were obtained (docking success rate for complexes in the top ten with rmsd $<$ 10 Å: 69.1%). When comparing the docking results obtained for 76 cases of the ZDOCK benchmark 2 to ZDOCK version 2.3, version 3.0, and the original FRODOCK docking procedure[13,14], our approach outperforms the other methods (Tab. 2).

IRMSD	<=10.0 Å	<=2.5 Å	<= 2.5 Å	< 4 Å
	Top10	Top20	Top100	Top100
DrugScore^{PPI} [a]	**69.1**	**65.6**	**76.0**	**80.2**
DrugScore^{PPI} [b]	**61.8**	**61.8**	**69.7**	**72.4**
ZDOCK2.3	-	18.4	31.6	-
ZDOCK3.0	-	25.0	-	50.0
FRODOCK	-	17.1	30.3	50.0

Table 2. Protein-protein docking results. Success rates in %.

Results achieved for [a]97 cases of ZDOCK benchmark 3 and [b]76 cases of ZDOCK benchmark 2.

Acknowledgments

We thank the Zentrum für Informations- und Medientechnologie (ZIM) at Heinrich-Heine-University, Düsseldorf, for computational support.

References

1. Gonzalez Ruiz, D. & Gohlke, H. *Targeting protein-protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding*, Curr. Med. Chem. **13**, 2607-2625, 2006.
2. Yura, K. & Hayward, S. *The interwinding nature of protein-protein interfaces and its implication for protein complex formation*, Bioinformatics **25**, 3108-13, 2009.
3. Berg, T. *Small-molecule inhibitors of protein-protein interactions*, Curr Opin Drug Discov Devel **11**, 666-74, 2008.
4. Garzón, J. I., Lopez-Blanco, J. R., Pons, C., Kovacs, J., Abagyan, R., Fernandez-Recio, J. & Chacón, P. *FRODOCK: a new approach for fast rotational protein-protein docking*, Bioinformatics **25**, 2544-51, 2009.
5. Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A. & Baker, D. *Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations*, J Mol Biol **331**, 281-99, 2003.
6. Gohlke, H., Hendlich, M. & Klebe, G. *Knowledge-based scoring function to predict protein-ligand interactions*, J Mol Biol **295**, 337-56, 2000.
7. Pfeffer, P. & Gohlke, H. *DrugScore^{RNA} - knowledge-based scoring function to predict RNA-ligand interactions*, J Chem Inf Model **47**, 1868-76, 2007.
8. Kurcinski, M. & Kolinski, A. *Hierarchical modeling of protein interactions*, J Mol Model **13**, 691-8, 2007.
9. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. *The FoldX web server: an online force field*, Nucleic Acids Res **33**, W382-8, 2005.
10. Fogolari, F., Brigo, A. & Molinari, H. *Protocol for MM/PBSA molecular dynamics simulations of proteins*, Biophys J **85**, 159-66, 2003.
11. Benedix, A., Becker, C. M., de Groot, B. L., Caflisch, A. & Bockmann, R. A. *Predicting free energy changes using structural ensembles*, Nat Methods **6**, 3-4, 2009.
12. Krüger, D.M. & Gohlke, H. *DrugScore^{PPI} webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions*, Nucleic Acids Res **38**, W480, 2010.
13. Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R. & Weng, Z.*Integrating statistical pair potentials into protein complex prediction*, Proteins **69**, 511-520, 2007.
14. Pierce, B. & Weng, Z.*A combination of rescoring and refinement significantly improves protein docking performance*, Proteins **72**, 270-279, 2008.