



Strategies to Overcome the Induced Fit Effects in Molecular Docking

I. Kufareva, R. Abagyan

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 1-6, 2008.

© 2008 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume40>

Strategies to Overcome the Induced Fit Effects in Molecular Docking

Irina Kufareva and Ruben Abagyan

The Scripps Research Institute, La Jolla, CA 93037, USA

E-mail: {kufareva, abagyan}@scripps.edu

Protein flexibility and induced fit effects present a major obstacle to the development of better molecular docking algorithms and scoring functions. Here we present several modeling and sampling strategies to generate high quality ligand-induced receptor conformations. We demonstrate that the repulsive-density approach, the alanine-scanning (SCARE) algorithm, and the ligand-guided receptor modeling significantly improve the ligand pose prediction, de novo inhibitor finding via virtual screening, and ligand profiling results.

1 Introduction

Recent years have been characterized by great advances of structural proteomics and exponential growth of the number of newly solved protein structures. However, these structures still represent only a small fraction of therapeutically relevant proteome, with other proteins only accessible in the form of approximate homology models. Moreover, as static snapshots of protein dynamic flexibility, they are often not fully compatible with ligands of interest. Ligand docking, screening and profiling efforts invariably fail in such cases.

General structure prediction methods of molecular dynamics simulations partially address the problem, however, they are still incapable of identifying a reasonably small set of dockable and screenable ligand binding pocket conformations. Alternatively, concurrent pocket/ligand global optimization can provide an exhaustive solution to the problem, however, it is impractical due to time and resource requirements. A reasonable alternative is the so-called multiple receptor conformation (MRC) docking¹. In this approach, the protein flexibility is represented by a series of rigid snapshots. For the best experimentally explored cases, experimental snapshots (e.g. multiple crystallographic structures or NMR ensembles) can be used as input for the MRC docking.

In the absence of multiple experimental structures, *ab initio* or ligand-guided simulations are needed to computationally generate an ensemble of receptor conformations. Available approaches to this task include molecular dynamics (e.g. Eyrisch and Helms²), normal mode analysis (e.g. Cavasotto, Kovacs and Abagyan³), and internal coordinate sampling (e.g. Abagyan and Totrov⁴). All these methods, however, tend to produce far too many models, most of them inappropriate for ligand docking due to insufficient volume and shape of the binding pocket. Furthermore, the large number of pocket conformers tends to decrease, rather than increase, the number of false positives in both pose prediction and compound scoring and screening. In addition the large conformational ensembles quickly become overwhelming for docking algorithms. Therefore, it is important to either compress the generated conformational ensemble or reliably select its most informative representatives.

We here present several modeling and sampling strategies to generate high quality ligand-induced receptor conformations. The first of them allows generation of more drug-

gable pocket models by introducing a pocket controlling device in the form of repulsive density. This method and its successful application to protein kinase CK2 are described in Section 2. In some cases, while a major fraction of a crystallographic or generated pocket very closely resembles its bound conformation, the structural rearrangements in the remaining fraction creates steric hindrances with the ligand and makes it impossible to reproduce the correct ligand binding geometry. In such cases, computational removal of the incorrectly placed elements provides a partially correct/partially empty pocket with sufficient number of native contacts for the ligand to dock in the near-native pose. The common wisdom underlying this approach is “better no atom than a wrong atom”. This approach was successfully implemented in the *SCan Alanines and REfine* (SCARE) flexible receptor docking algorithm presented in Section 3. Computational excision, rather than modeling, of structural elements of the receptor is also beneficial in cases where ligands bind to pockets previously occupied by these elements. The so-called type II inhibitors of protein kinases form a well-studied family of such interactions. In Section 3, we also present DOLPHIN (*Deletion Of Loop from PHe-IN*) kinase models as powerful devices for ligand development, screening, and activity profiling.

Known strong small molecule binders to the protein of interest can provide a valuable information for both generation and selection of receptor pocket conformations in MRC docking, and help reduce the complexity and dimensionality of the problem. One way to employ this information is a so-called ligand-guided receptor selection, where among the multiple conformations, only those are chosen for screening that are selective towards known ligands. An overview of several successful applications of this strategy is described in Section 4.

Overall, the described approaches to the MRC generation and selection, along with proper methods of ligand docking and scoring, represent a great advance in the field of induced fit docking.

2 *Ab Initio* Pocket Ensemble Generation

A common drawback of simulation methods is the large number of generated models, with only a small fraction of them being compatible with ligand binding. Clearly, there is a need in computational techniques that can successfully guide the simulation procedures towards more druggable pockets. Experimental information in the form of known ligands targeting the pocket and their activities may be of great help. However, for *de novo* pockets such information is not available. We developed a computational protocol that addresses this problem by introducing a repulsive density in the pocket as an independent energy/penalty term. Such density represents a generic ligand and prevents the simulation procedure from generating conformations in which elements of the structure collapse inside the pocket. The algorithm of density generation includes the following steps:

1. Identification of residues whose sidechains form the pocket.
2. Simultaneous conversion of these residues to Ala.
3. Construction of atom density grid map for the obtained protein.
4. Repeated spatial averaging of the map in order to obtain a smoothed density cloud filling the cavities of the original protein.

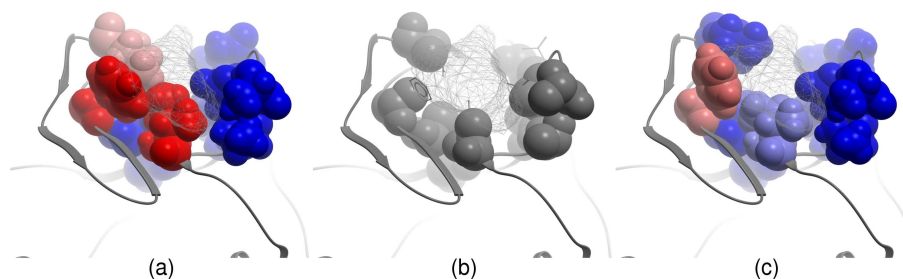


Figure 1. Repulsive density blob represents a generic ligand and guides side-chain sampling procedure towards more open pocket conformations. (a) - Initial closed pocket conformation, (b) - pocket sidechain alanine mutation and density generation, (c) - resulting open pocket conformation.

5. Taking a difference of the smoothed and original maps.

The procedure results in blobs of density filling in the cavities of the protein. Due to side-chain alanine conversion, the generated density represents the maximal volume of the pocket achievable without backbone rearrangements. The density is further used as an additional energy/penalty term in montecarlo side-chain simulation (Fig. 1).

This protocol was applied for *de novo* finding of compounds that bind to the N-terminal lobe of protein kinase CK2 and prevent its interaction with the regulatory subunit CK2 β . Screening a large virtual chemical database against the generated ensemble of four most druggable conformations yielded a series of compounds that were experimentally validated and confirmed to inhibit the subunit interaction in a dose-dependent manner.

3 Better Deleted Than Displaced

The generated receptor conformations will rarely cover a 100% of the protein pocket conformational space. It is safe to assume that any generated conformation is only partially correct, with some elements still interfering with ligand binding. Therefore, a customized or systematic removal of those parts may lead to “dockable” pocket models in which the ligand binding geometry is successfully reproduced, provided that the removed parts is not the main determinant of the ligand-receptor interaction. This part can later be brought in as a part of the refinement, if needed.

3.1 SCARE

The majority of the induced fit changes involve a few protein side-chains and only a minor adjustment of the backbone. Therefore removal the pocket parts at the side-chain level is frequently sufficient. We developed an algorithm that systematically scans pairs of neighboring side chains in the binding pocket, replaces them by alanines, and docks the ligand to each “gapped” version of the pocket. All docked positions are scored, refined with original side chains and flexible backbone and re-scored. The optimal SCARE (SCan Alanines and

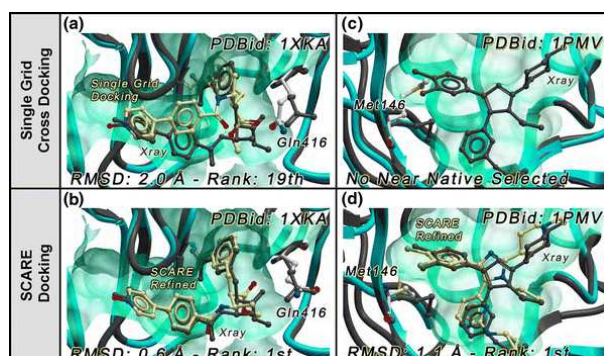


Figure 2. The SCARE algorithm successfully reproduces ligand binding geometry in difficult cases where traditional rigid-receptor docking fails.

REfine)⁵ protocol identifies a near native conformation (under 2 Å RMSD) as the lowest rank for as many as 90% of the cross-docking complexes (compare to 50% success rate with optimal single receptor cross-docking) (Fig. 2). The procedure predicts not only the binding pose of a ligand, but also conformational changes induced by its binding, therefore producing a new highly relevant protein conformation that can be used in VLS along with the original one.

3.2 DOLPHIN

In the recent years, there is a great interest to a specific type of protein kinase inhibitors, the so-called *type II inhibitors* that induce a transition of the kinase activation loop from its active, DFG-in, position, to the DFG-out state. This transition is too large to be modeled by any existing computational method. However, we found that the above “better deleted than displaced” strategy is very helpful in induced fit docking of type II ligands to DFG-in structures. We showed that *Deletion-Of-Loop from PHe-IN* (DOLPHIN, Fig. 3) kinase modification leads to models which

1. Reproduce the correct binding geometry of the existing type II ligands.
2. Selectively score active type II ligands higher than inactive and decoys.
3. Provide means for *in silico* ligand activity profiling.

Fig. 4 illustrates a great potential of the DOLPHIN kinase models as screening and profiling devices for type II kinase inhibitors.

4 Ligand Guided Model Selection and Generation

Multiple receptor conformations for high throughput ligand docking can be generated with one or several ligands actually present in the binding site. Fully receptor-flexible docking of a few known ligands can be performed to force the receptor into alternative conformations. The irrelevant generated conformations can then be filtered out by evaluating

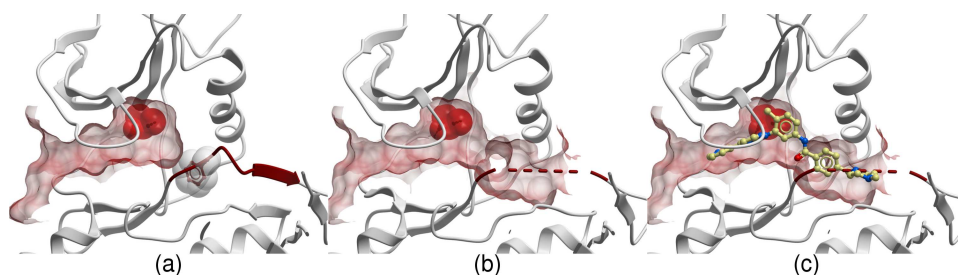


Figure 3. The DOLPHIN kinase models are powerful screening devices for type II kinase inhibitors. (a) - Initial DFG-in structure, (b) - modified structure, (c) - docking of a type-II ligand.

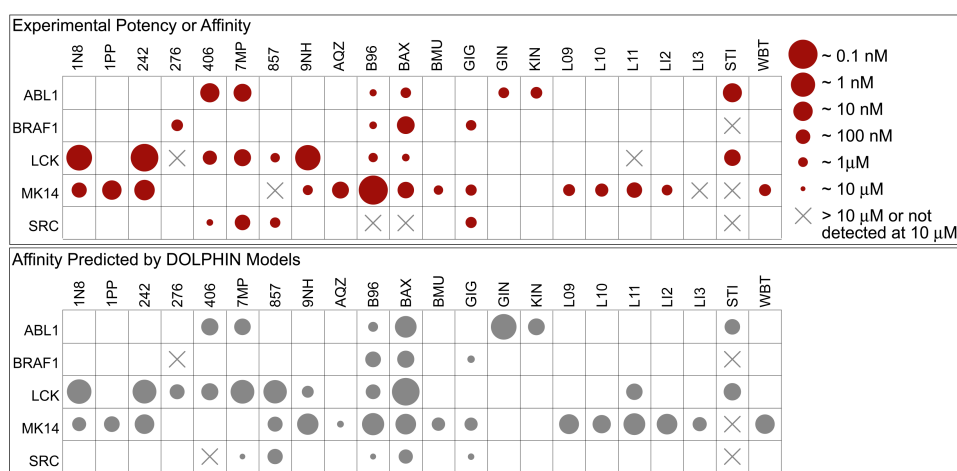


Figure 4. Comparison of experimental activities of known type II kinase inhibitors with their binding affinities predicted by the DOLPHIN kinase models.

the enrichment factor on a test set. Bisson et al.⁶ generated multiple conformations of androgen receptor with two different antagonists by Monte-Carlo sampling in ICM. Each conformation was tested for its ability to discriminate between AR binders and non-binders in a panel of 88 nuclear receptor ligands. The two AR conformations with the best enrichment characteristics were then used for virtual ligand screening of the marketed drugs for potential androgen receptor antagonists. Three identified antipsychotic drugs exhibited anti-androgenic activity and were then rationally re-purposed to nonsteroidal molecules with improved AR antagonism and marked reduction in affinity for dopaminergic and serotonergic receptors. Similar procedure led to productive models of two G-protein coupled receptors, M2 muscarinic receptor and melanin concentrating hormone receptor⁷, and resulted in successful *de novo* identification of new modulator chemotypes.

5 Conclusion

We described a series of computational approaches that allow to model receptor flexibility in molecular docking. These methods present a practical alternative to concurrent protein and ligand simulation, and lead to productive receptor models suitable for ligand docking, screening, and profiling. The new approaches were successfully applied to find novel inhibitors in several difficult cases.

Acknowledgments

This work was funded by NIH/NIGMS grants 5-R01-GM071872-02 and 1-R01-GM074832-01A1.

References

1. M. Totrov and R. Abagyan, *Flexible ligand docking to multiple receptor conformations: a practical alternative*, Curr. Opin. Struct. Biol., **18**, 178–184, Apr 2008.
2. S. Eyrisch and V. Helms, *Transient pockets on protein surfaces involved in protein-protein interaction*, J. Med. Chem., **50**, 3457–3464, Jul 2007.
3. C.N. Cavasotto, J.A. Kovacs, and R.A. Abagyan, *Representing receptor flexibility in ligand docking through relevant normal modes*, J. Am. Chem. Soc., **127**, 9632–9640, Jul 2005.
4. R. Abagyan and M. Totrov, *Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins*, J Mol Biol, **235**, no. 3, 983–1002, Jan 1994.
5. G. Bottegoni, I. Kufareva, M. Totrov, and R. Abagyan, *A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE)*, J. Comput. Aided Mol. Des., **22**, 311–325, May 2008.
6. W.H. Bisson, A.V. Cheltsov, N. Bruey-Sedano, B. Lin, J. Chen, N. Goldberger, L.T. May, A. Christopoulos, J.T. Dalton, P.M. Sexton, X.K. Zhang, and R. Abagyan, *Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs*, Proc. Natl. Acad. Sci. U.S.A., **104**, 11927–11932, Jul 2007.
7. C.N. Cavasotto, A.J. Orry, N.J. Murgolo, M.F. Czarniecki, S.A. Kocsi, B.E. Hawes, K.A. O'Neill, H. Hine, M.S. Burton, J.H. Voigt, R.A. Abagyan, M.L. Bayne, and F.J. Monsma, *Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening*, J. Med. Chem., **51**, 581–588, Feb 2008.