Palacký University
Olomouc

Prof. Rüdiger Ettrich, Ph.D.

Center for Nanobiology and Structural Biology
Institute of Microbiology of the Czech Academy of Sciences
Nove Hrady, Czech Republic

Palacký University Olomouc

Qs 1: Several figures in the introduction lack any citation in their legend (for example Figure 1.1., 1.3., 3.2., 3.4.). This either means that these figures were prepared by the candidate using however information that should be cited (source of the number and distribution of the approved covalent drugs, for example) or the figure was taken over/ modified from a textbook or article that should be cited. Please clarify?

Answer:

I agree that citations are not mentioned in their legend. However, citations are mentioned in the text. Figure 1.1 and 1.3 have been made by myself. For Figure 3.2 & 3.4 are also prepared by myself using the data available in the Ref. 108 and Ref. 170-172.

Qs 2: In the introduction you write "To be usable for SBDD, the crystal structure needs to show the details of protein-ligand binding and have a reasonable resolution (< 2.5 Å)." this is of course true, but I miss a hint of what is meant with "details of protein-ligand binding". I think at this place it would be appropriate to explain how actually binding works, briefly describing the theory behind conformational selection (or in rare cases induced fit). Then it would come clearer that "details of protein-ligand binding" actually mean that the crystal structure should represent a binding-competent conformation, which is one of the main drawback in in-silico drug screening, as crystal structures do not necessarily represent binding-competent conformations, and sometimes not even the most-populated conformation in solution. In most cases we find conformational selection as the mechanism of how ligands bind, which makes it especially hard to predict the binding-competent state in cases where no complex structure is known.

Answer :

In this thesis our main focus was the protein-ligand non-covalent interactions. We need the positions of the atoms to be sufficiently reliable (i.e. electron densities are well defined, especially for the ligand) so that physics-based modeling makes sense. Of course, the reviewer is right in that protein conformational landscape should be taken into account. We are currently working on it using Induced Fit Docking and Restrained Molecular Dynamics-based approaches.

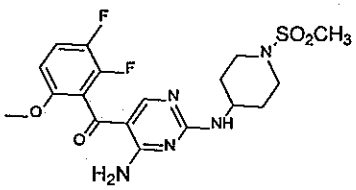
Qs 3: Any idea for what ligands the classical SF fail resp. report false positives? Is there a specific ligand type, for example large, flexible and charged ligands that causes the failing?

Answer:

In the case of neutral ligands, the classical SFs performed worse in identifying the native binding poses.

Causes of failing:

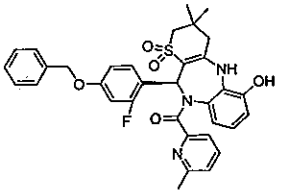
- Rotatable bond.



2FVD (CDK2)

Rotatable Bond : 6

PM6-D3H4X/COSMO : 0
DFTB3-D3H4X/COSMO :0
AutoDock4 : 4
UCSF Dock : 0
AutoDock Vina : 3
Glide XP : 40

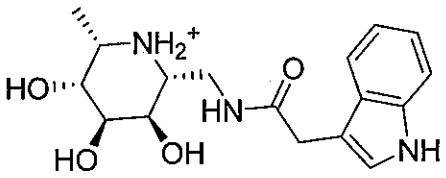


3GNW(hepatitis C virus NS5B RNA-dependent RNA polymerase)

Rotatable Bond : 5

PM6-D3H4X/COSMO : 1
DFTB3-D3H4X/COSMO :1
AutoDock4 : 71
UCSF Dock : 28
AutoDock Vina : 14
Glide XP : 1

Palacký University
Olomouc



In the case of Charged ligands, the classical SFs had larger trouble in identifying the native binding poses.

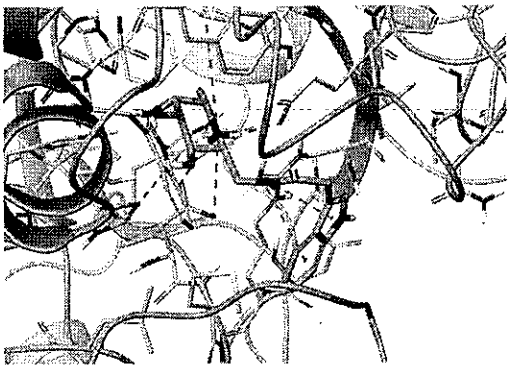
Causes of failing:

- +Ve Charged System
- High number of hydroxyl group

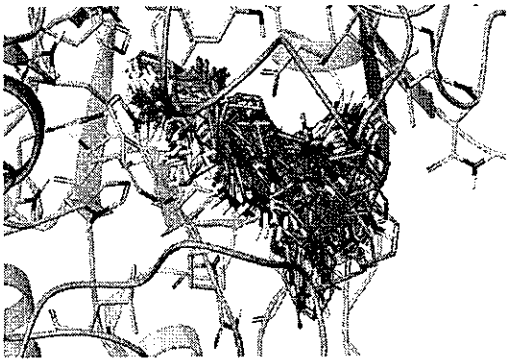
2ZX6 (α-L-fucosidase)

Rotatable Bond : 4

PM6-D3H4X/COSMO : 5
DFTB3-D3H4X/COSMO :0
AutoDock4 : 75
UCSF Dock : 26
AutoDock Vina : 36
Glide XP : 140



Std. Binding Pose



Docked Binding Poses

Qs 4: Do you have a guess why UCSF DOCK by far outnumbers the other SF in terms of false positives? Would you consider that a common feature for all anchor-and-grow based docking programs?.

Answer:

For UCSF Dock, four system, 2P4Y, 4GID, 2VOT and 3NOX, yielded in total 403 FPs, which is 70% of FPs for charged ligands.

Reasons : Morpholino group has high flexibility and fewer non-covalent interaction with the protein due to this reason poses with higher RMSD could not score well. Large distribution of docked poses in binding site due to anchor-and-grow algorithm.

Protein	DFTB3	PM6	UCSF
2P4Y	5	20	98
4GID	3	2	98
2VOT	1	9	103
3NOX	8	10	104

3NOX: Std pose and Docked poses from UCSF Dock

Přírodovědecká
fakulta

Review of PhD thesis

Computer-Aided Drug Design: Quantum Mechanical Investigation of Protein-Ligand and Protein-Ligand-Water Complexes

Submitted by M.S. Haresh Ajani

The PhD thesis of Haresh Ajani describes application and development of the structure-based drug design methods, in particular molecular docking and rescoring by more accurate semi-empirical QM methods. The goal of the thesis is threefold: The first section summarize results of the project, in which molecular docking was used in order to find the proper binding mode of the ligand in several selected protein enzymes. In the second section, the author tested efficiency of a posteriori rescoring the binding poses by two semi-empirical methods, and finally in the third section an attention is paid to the role of active site water molecules to the binding of the ligand and application of the docking methods taking these waters into account.

The PhD thesis is supported by six articles published in impact journals, from which in two of them the candidate is the first author. To this day, the candidate published in total 10 articles in impact journals (two of them as the first author), which were cited according Web of Science to this dat 79 times leading to H-index of the author equal to 6. These scientometric data are considered as very good for finishing PhD student.

The thesis it-self is written as a brief introduction to the structure-based drug design methods complemented by short comments to the articles supporting the PhD thesis. This composition in some sections hampered the text flow, so that the reader has to read the corresponding articles first to fully understand the text. In addition, some section would benefit from improving the English. Nonetheless, all aims of the thesis are clearly stated and all conclusions are fully justified by data shown either in thesis or in attached articles, so the thesis also fulfill all criteria for the purpose of the candidate to be awarded a PhD degree.

I have few points I would like to ask the candidate, while leaving the other minor points for discussion during the defense:

1. Could you compare the computational demands of evaluation of the standard scoring functions with respect to the semi-empirical QM ones? What is the rate between time demands required conformational search (docking) by standard scoring function and the subsequent rescoring by semi-empirical QM method? What is the sensitivity of the rescoring by semi-empirical methods to the quality of the binding poses obtained by classical docking, i.e., are these poses accurate enough?
2. What ligand/decoy testing set was used for calculation of the enrichment factor in section 4.2.2 (Figure 4.4)? What RMSD cutoff is according to your opinion most relevant for comparison of the enrichments factors obtained by different methods?
3. Is it possible to combine tools for taking the active site water molecules into account such as WaterMap with algorithms for flexible docking, in which at least part of the protein

Department of Physical Chemistry| Faculty of Science
17. listopadu1192/12 | CZ 771 46 Olomouc | Czech Republic | T: +420 585 634 756
www.fch.upol.cz

active site is allowed to move during the conformational search?

Finally I would like to state that applicant fulfills all criteria for being awarded a PhD degree and thus I fully

recommend

the thesis for further defense.

In Olomouc 12. 6. 2018

doc. Mgr. Pavel Banáš, Ph.D.

Department of Physical Chemistry| Faculty of Science
17. listopadu1192/12 | CZ 771 46 Olomouc | Czech Republic | T: +420 585 634 756
www.fch.upol.cz

Review of the dissertation thesis " **Computer-Aided Drug Design: Quantum Mechanical Investigation of Protein-Ligand and Protein-Ligand-Water Complexes"** proposed by Haresh Ajani.

The thesis is a dedicated collection of works on a rather large subject – Computer Aided Drug Design by means of computational methods. In all described and discussed cases the basic condition fulfilled is the knowledge of the 3D structure of the molecular target. This information is a cornerstone of the docking and scoring methodologies used for finding of an optimal ligand and its optimal pose in the target protein molecule. However, the central concept of this thesis is "applicability" rather than "development" of more or less advanced methodologies to predict correct binding mode and quantify binding affinity.

The format of the thesis is a brief summary or better say comment of the attached papers already published in per review journals. The introductory part provides an explanation of the whole CADD concept and its history, techniques used for docking and prediction of the binding modes, philosophy and construction of a scoring functions and role of water molecules in the process of binding and interactions. Brief list of the non-covalent interactions and its explanation closes this part of the thesis and from my point of view is rather redundant. Scopes of the thesis are divided into two parts - the first part of the thesis is based on structure-based drug design in the discovery, structural activity relationship (SAR), accurate prediction of binding mode of small drug-like molecules whereas the second part is testing of developed (not by the author) structure-based drug design methods followed by the application of SQM based scoring functions on diverse protein-ligand complexes with the implicit COSMO solvation model.

The more detailed part of the summary has title *"STRUCTURE-BASED DRUG DESIGN: A DIRECT APPROACH"* and only describes the already mentioned categories and terms defined in the introduction part. What I do appreciate is quite comprehensive description of existing scoring functions and their classification. From a certain point of view the role of water is nothing but another case of ligand docking, in the proposed thesis the role of explicit waters is discussed separately and it importance is particularly stressed.

First part of the results is dedicated to papers describing Binding Mode Analysis and Design and Discovery of Kinase and Hydrolase Inhibitors , the second part is fully devoted to the role of explicit water in the docking procedures. The exception is a paper briefly described

the SQM based scoring functions, a deterministic scoring algorithm, employing a complete non-covalent interaction and iterative pose prediction via different available docking programs.

The proposed intro or summary is a kind of minimalistic which is not meant completely as an objection. It only means that it would be probably better to make for example the scoring function the central element of the methodology and demonstrate it in various examples. The thesis is from graphics point of view at standard level, so the text. I have almost no comments to the "papers part" of proposed thesis but I do a few questions which seems to me important to be answered.

1) There is no discussion in the proposed thesis about a role of natural cations/anions in the process of protein-ligand binding or their influence on the explicit water molecules placement. I would like to get an explanation how the cations/anions are treated and if it is already known how much they can influence the binding and corresponding methodology treatment.
2) There is only one paper dedicated to the development or testing of the new scoring function. The presented SQM based scoring functions is based on a deterministic scoring algorithm, employing a complete non-covalent interaction and iterative pose prediction via different available docking programs. My question is how much is the process sensitive to the used datasets and how these datasets differ between each other. Last but not least, what was the role of the author on different parts of this methodology development and how the whole effort was synchronized with other computational tasks necessary to achieve the goal.

Prague 13.6.2018

.....

RNDr Jiri Vondrasek CSc