Academy and University Center Nové Hrady

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

Prof. RNDr. Rüdiger H. Ettrich, PhD.

☎ +420 386 361 297 +420 386 361 279 ✉ ettrich@nh.cas.cz

Nové Hrady, 17.2.2016

Oponentský posudek doktorské disertační práce

Adam Pecina: Quantum Chemical Approach for *In Silico* Drug Design.

In his PhD thesis Adam Pecina contributes to the broad field of drug design from a computational perspective. The thesis deals with three topics, supported by eight publications, which are logically connected: The first part discusses the nature of nonclassical non-covalent interactions - so called σ-hole bonding. In this part high-level QM methods, the coupled cluster QM method (CCSD(T)/CBS) and the symmetry adapted perturbation theory (SAPT), are applied to study the strength and origin of halogen-, chalcogen- and pnicogen bonded model systems. The second part then uses a hybrid QM/MM approach to study three protein-ligand systems, HIV-1 protease, secreted aspartic protease and carbonic anhydrase. Corrected DFT and semi-empirical quantum mechanical (SQM) methods are used to explain fundamental differences in the binding modes of inhibitors of these enzymes. The last part then contributes to the virtual screening of compound libraries by adapting/simplyfying a SQM-based scoring function for this purpose. This is followed by benchmarking this new physics-based method against eight standardly used scoring functions. The results are convincing and promise that the SQM/COSMO filter might become an effective tool in the field.
The thesis consists of 88 pages and is written in solid english with only a minor number of misspellings, typos or errors. 6 pages make up the introduction that starts with an intro to ligand binding, binding affinities and free energy of binding, and then right-away switches to computational methods how to determine free energy of binding etc. Then 13 pages of methods are describing in detail all methods used in the thesis with giving the essential methodology background. Then 48 pages of results are followed by 4 pages of conclusions and final remarks. It is visible that the author invested a lot of time and care in writing up his research and did care even about small details that are often overseen. Adam Pecina has eleven publications so far, eight of those are included in the thesis. On five of these papers Adam Pecina is the first author or shares first authorship.
The broadness of the whole thesis demonstrates the broad theoretical and computational knowledge of the candidate and the reader is left with the feeling that the candidate knows what he is writing about.

Questions and remarks for the defense that should be addressed by the candidate:

1. In the introduction I miss a hint how actually binding works. I think in a thesis that discusses ligand-binding the theory behind conformational selection (or in rare cases induced fit) should be at least mentioned, as the main drawback in in-silico drug screening is still the fact, that a crystal structure does not necessarily represent a binding-competent conformation, especially not if it has not been crystallized with a ligand as a complex but just as the enzyme itself. Most biochemistry textbooks still discuss key and lock principles and induced fit as general models, even though we know thanks to MD and NMR work in the 90s that in most cases we find conformational selection as the mechanism, which makes it especially hard to predict the binding-competent state in cases where no complex structure is known. For this reason I consider it important that especially we in the computational field emphasize this fact repeatedly and thus contribute a bit to the "enlightenment" in the ligand-binding field. This is just a comment for discussion.

2. On page p.39 you write: "The enlarging of the size of the QM region did not influence the results. The sizes of the 6 and 8 Å surroundings of the ligand were energetically consistent with DFT and PM6- DH2 results on the

small region. However in the regions bigger than 10 Å, the unrelated structural changes occurred far from the active site that affected also the relative stabilities." It is not clear to me what you mean with unrelated structural changes? Unrelated from what? Could you describe the character of those changes and how they affect the relative stabilities, as this is not really clear to me what is meant here?

3. On page 44 you explain your "virtual glycine scanning" which was inspired by "computational alanine scanning" In this method you calculated the energy contributions of the active site side chains (ΔΔG'int) as the difference between the original ΔG'int with the wildtype amino acid and the new ΔG'int with the mutated glycine residue. I understand why to use glycine, as the contribution from the single hydrogen will be close to zero and it more represents a deletion of the residue side-chain as a whole. On the other hand, a substitution to an alanine might be more matching an experimental screening, as it doesn't show the flexibility of a glycin and has to keep the angles in the Ramachandran plot. Also the methyl-group fills the space, not allowing water to take the open space. Even if this a bit of a hypothetical question, do you have an estimate how big the difference from your values would be if using alanine instead of glycine and would this change the qualitative results you got?

4. In two cases, AR and HIV protease the widely used docking programs Autodock VINA and Glide from Schroedinger did not only produce the highest number of false positives (while all other scoring functions performed quite okay) but failed also with respect to the RMSDmax. Could you discuss what makes specifically these two programs fail in those two cases? What might be the interaction/binding contribution that is under-/overestimated? Could you conclude that for specific cases (which one?) these programs are not suited at all?

5. Question 4 actually raised also another follow-up question: For the HIV protease practically all widely used scoring-functions fail (apart from your method, the Amber/GB, ASP and CS) In this case we have a large, flexible and charged ligand, but also structural water in the binding site. What causes the failing in those cases, the size in combination with the flexibility, the charge or the water? Do you think the failing has the same origin in all cases or is for different reasons?

6. AutoDock VINA has a parameter that is called 'exhaustiveness' and sometimes it helps to increase the parameter. It represents the time spent on the search, and a linear increase should decrease the probability of not finding the minimum exponentially. Did you use the standard value or did you also try an increased value? Would that have improved the performance with respect to the RMSD max? However, I understand if you can't answer this question as your focus was not fine-tune the performance of the benchmarking programs but rather use them with their default values.

Finally, it is my pleasure to state that Adam Pecina until now conducted internationally recognized high quality science, clearly manifested in the 46 citations that his publications gained so far (Web of Science, 17.2. 2016). The well written thesis tells a nice story, the results are presented and discussed convincingly, and the publications that back up the thesis show, without leaving any doubts, that the applicant fulfills all criteria for being awarded a PhD degree. Therefore I can fully recommend Adam Pecina for being awarded the PhD degree.

(Český doplneťk: Adam Pecina jásně prokázal tvůrčí schopností, prácé bez sebemenších pochybů splňuje požadavky kládené na disertační přáce v oboru biochemie)

Prof. RNDr. Rüdiger H. Ettrich, Ph.D.

Center for Nanobiology and Structural Biology,
Institute of Microbiology,
Academy of Sciences of the Czech Republic
and
Faculty of Sciences, University of South Bohemia