Validation of Docking Poses via Interaction Motif Searching

Aim

Given a set of docking poses for a single ligand can we separate correct poses from high scoring but incorrect poses by comparing protein-ligand interaction motifs?

Introduction

Accurate modelling of the way molecules interact, either in crystals, or in protein-ligand complexes usually requires highly computationally intensive methodologies. Yet, in rational drug design, effective virtual screening requires a rapid analysis of many ligand-protein configurations and so highly intensive methods are not useful. Available docking methodologies are fast and, although reasonably accurate, they are very far from being 100% reliable. Techniques which are capable of validating high ranking docking poses to identify with near certainty the pose most likely to be correct, therefore have value.

Currently practitioners might do this by employing several different scoring functions, using a rescoring or a consensus approach[1]. Alternatively, they might validate with a sophisticated force-field or a quantum mechanical approach[2]. We introduce here the idea of using a knowledge-based validation approach.

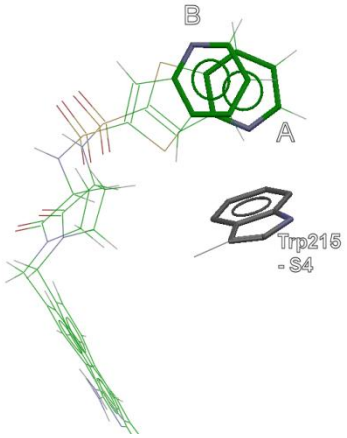
Knowledge-based scoring functions already exist, that are parameterised using atom-atom distance statistics derived from crystallographic databases[3]. However, the data reduction step of collapsing 3-dimensional information into a single dimensional score means much useful configurational data is lost. What we propose to do instead is to choose one or more important interaction types and, using all the available 3-dimensional data for these interaction motifs in the binding pose, test the validity of these interactions between protein and ligand, against the data available in the Cambridge Structural Database. Choice of interaction type is important. Geometries of intermolecular hydrogen bonds are usually generated with reasonable accuracy by docking programs, being easy to calculate and, because important to get right, parameterised well. So if we wish to choose an interaction type to validate, it should be of a more diffuse nature and harder to calculate, yet remain very common. An example of such an interaction is the pi-pi stacking interaction which, it is well known, can adopt various configurations (T-shaped, parallel stacked and so on) depending on the partners that are interacting[4]. The pi-pi stacking interaction is extremely common both in small molecule crystals and in protein-ligand complexes, yet, because it depends heavily on the quadrupole moments of the interacting species, it is difficult to calculate accurately without using time consuming quantum chemical methods[5]. Other motifs are also suitable for this purpose, cation-pi interactions or metal coordination geometries are examples.

We will look at two docking cases in which the docking package GOLD generates the incorrect pose as the top ranking pose, and the correct pose is lower ranked. In the first we will use a single pi-pi interaction to validate the poses, in the second we will look at a metal coordination motif.

Method

Pi-pi Interaction: Poses were taken from the GOLD docking of the factor Xa inhibitor in the PDB structure 1f0r, back into the cognate protein model. 1f0r is one of the structures within the Astex Diverse set and has been validated as a high quality structure. The docking run was carried out under default settings using the GoldScore scoring function. Ten Genetic Algorithm attempts were run and ten poses saved. The top two poses have respectively GoldScores of 89.9 and 84.6. In Figure 1a we see the two binding poses superimposed. Also displayed is the indole of the tryptophan 215 amino acid residue found in the S4 pocket of factor Xa. Although the poses are very similar we can observe that the terminal pyridothiophene is placed in two different and opposite orientations.

Mol2 files of each pose+protein were exported from Hermes. In addition a mol2 file representing the 1f0r structure was also prepared. Each complex was imported into Mercury and a Packing Feature search was set up. This comprised the six pyridyl heavy atoms on the ligand aryl group and the nine heavy atoms of the Trp indole. The pyridyl N was optionally allowed to be C or N and the number of bonded atoms was set as unspecified for this atom to allow a search for phenyl as well as pyridyl groups. All restrictions on the numbers of hydrogens on query atoms were removed. Required closeness of fit to the query was set at 'Low'. The search was carried out over the entire CSD v531. RMSDs of matches, for all three complexes, are displayed in histogram form in Figure 1b.



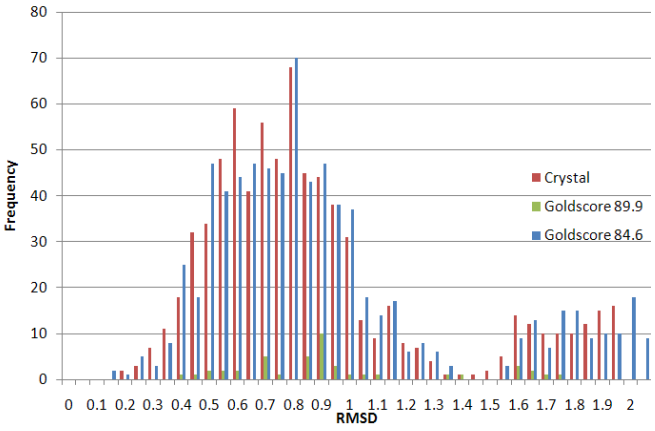
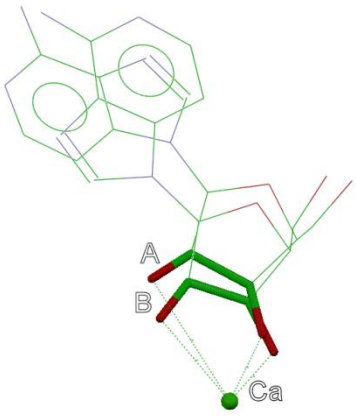


Figure 1a – Superposition of top two binding modes of factor Xa inhibitor in 1f0r,

Figure 1b - RMSD distribution of aryl-indole interaction for pose A (green) and pose B (blue), compared with similar interactions in the CSD. The distribution for the same interaction found in the 1f0r crystal structure is in maroon.

Metal coordination: Poses were taken from GOLD docking runs of the nucleoside hydrolase inhibitor inhibitor in the PDB structure 1hp0, back into the cognate protein model. 1hp0 is one of the structures within the Astex Diverse set and has been validated as a high quality structure. The docking run was carried out under default settings using the GoldScore scoring function. Ten Genetic Algorithm attempts were run and ten poses saved. The first and the third poses have respectively GoldScores of 71.8 and 69.6 (the second pose saved is almost identical in geometry to the first and is not shown). In Figure 2a we see the two binding poses superimposed. Also displayed is the calcium ion that the sugar moiety of the ligand makes a bidentate interaction to. The terminal purine base is in opposite orientation in the two poses and the metal coordination is slightly different.

Mol2 files of each pose+protein were exported from Hermes. In addition a mol2 file representing the 1hp0 structure was also prepared. Each complex was imported into Mercury and a Packing Feature Search was set up. This comprised the diol fragment of the inhibitor and the calcium ion. All restrictions on the numbers of hydrogens on query atoms were removed. Required closeness of fit to the query was set at 'Low'. The search was carried out over the entire CSD v531. RMSDs of matches, for all three complexes, are displayed in histogram form in Figure 2b.



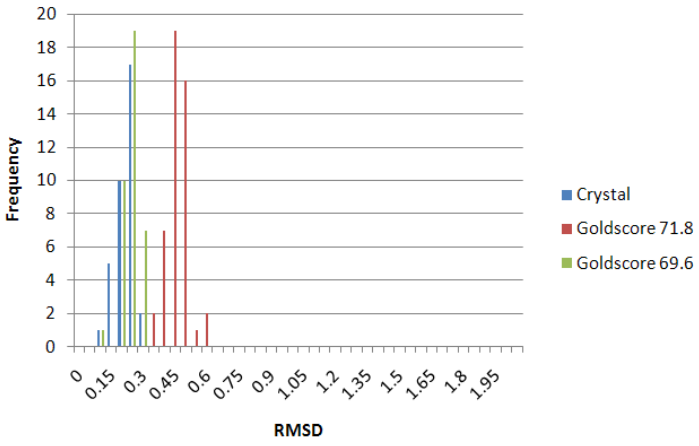


Figure 1a – Superposition of top two binding modes of nucleoside hydrolase inhibitor in 1hp0,

Figure 1b - RMSD distribution of Ca-diol interaction for pose A (maroon) and pose B (green), compared with similar interactions in the CSD. The distribution for the same interaction found in the 1hp0 crystal structure is in blue.

Results

Pi-pi interaction: The search on the highest scoring docking poses only finds a very small proportion of similar stacking interactions within 2 Å RMSD (green bars). The search on the second ranked pose (blue bars) finds a great many hits below RMSD 0.5. The crystal structure (red bars) has a slightly greater number of stacking interactions at low RMSD. This may be not significant as in fact these two structures are almost superimposed in reality. Nevertheless it might be an indication that the stacking feature search is exquisitely sensitive to movements away from ideal stacking.

Note here that very few, very low RMSD hits are to be found, but this is to be expected, as a cone correction of some type (i.e. a multiplication by $A/(RMSD)^N$ ($N \geq 2$)) is required to convert the histogram frequencies into a radial RMSD probability distribution. Very low RMSD matches are expected to be extremely rare.

Metal Coordination: Both the highest (pose A), and the third highest ranked pose (pose B) from the docking, give rise to RMSD distributions at the low end of the RMSD scale. However, the distribution for the poorer scoring pose is centred at 0.25, whereas that for the best pose is centred at 0.5. This indicates that the poorer scoring pose is in fact likely to be the correct pose, which it is, in fact.

The binding pose found in the 1hp0 structure gives ride to a very similar histogram to pose B, although, it appears slightly shifted to the left, which may again indicate a high sensitivity of this technique to small changes in pose.

Conclusions

These two examples show that it is simple to identify a binding pose close to the correct binding mode, should such a pose exist in the docking set; by comparing a small number of non-hydrogen-bonding interaction motifs between protein and ligand against similar interactions in the CSD. The technique appears to be very sensitive to relatively small deviations away from ideal interaction and this reflects the fact that all the 3-dimensional information is used in validating the interaction. Setting up the search for the interaction of choice is very easily done using the Packing Feature Tool in *Materials* Mercury.

We conclude therefore that this tool may prove to be of value to those modellers wishing to identify the ligand binding pose most likely to be correct, from a set of poses arising from a modelling procedure. One particular application is in fragment-based virtual screening where standard scoring functions may not be adequate enough to separate good and bad binding poses.

This module, now adapted for use by drug discovery researchers, is currently available as a beta release add-on module to the Cambridge Structural Database System. We are currently looking for beta testers for this module. If you are interested in evaluating this drug design software please contact support@ccdc.cam.ac.uk.

References

1. a) P. S. Charifson, J. J. Corkery, M. A. Murcko, W. P. Walters, *J. Med. Chem.,* **42**, (1999), 5100-5109. b) M. Feher, *Drug Discovery Today*, **11**, (2006), 421-428.

2. a) P. D. Lyne, M. L. Lamb, J. C. Saeh, *J. Med. Chem.,* **49**, (2006), 4805-4808. b) M. P. Gleeson, D. Gleeson. *J. Chem. Inf. Model.*, **49**, (2009), 1437-1448.

3. a) C. A.Sotriffer, H. Gohlke, G. Klebe, *J. Med. Chem.,* **45**, (2002), 1967-1970) W. T. M. Mooij, M. Verdonk, *Proteins: Structure, Function and Bioinformatics*, **61**, (2005), 272-287.

4. C. A. Hunter and J. K. M. Sanders, *J. Am. Chem. Soc.*, **112**, (1990), 5525-5534.

5. M. Luhmer, K. Bartik, A. Dejaegere, P. Bovy, J. Reisse, *Bull. Soc. Chim. Fr.,* **131**, (1994), 603-606.

Products

CSD – the world's only comprehensive, fully curated database of crystal structures, containing over 500,000 entries

Materials module of Mercury – a powerful exploration and comparison tool for solid state structures

Mercury – a versatile and feature-rich visualisation tool for molecular structures

GOLD – an accurate and reliable protein-ligand docking program

For further information please contact Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK. Tel: +44 1223 336408, Fax: +44 1223 336033, Email: admin@ccdc.cam.ac.uk