Sampling experimentally observed ring conformations during protein-ligand docking

Aim

Ring structures commonly form the core of small molecule drugs and biological substrates. The problem of exploring conformational space of flexible rings in protein-ligand docking is examined using the program GOLD.

Introduction

Successful protein-ligand docking depends upon two factors: exhaustive exploration of search space and effective scoring. In this case study we explore the improvements that can be gained in both pose prediction and scoring by applying two different algorithms to explore conformational space of flexible rings. The first algorithm, ring corner flipping[1, 2] has been part of GOLD[3] since its inception. The second algorithm, which makes use of ring templates from the CSD[4] (Cambridge Structural Database), is a new feature of GOLD 4.1. One area where the ring template matching methodology is envisaged to be of particular use is in the docking of ligands containing macrocycles.

Method

The ring template matching algorithm relies on finding a matching template in the ring template library. The ring template library is derived from the CSD. During ligand initialisation each flexible ring is checked against the ring template library. If there is a match in the library the templates for that ring will be sampled during the genetic algorithm.

In this study protein-ligand complexes, where the ligand had a ring capable of adopting different conformers, were obtained from Relibase+ (PDB codes 1s19, 2pcp, 1ki8 and 1tt1). Non-native conformations of the ligands were generated using Corina 3.2. The non-native ligands were then docked back into the protein binding site using four different protocols:

1. Without either the ring corner flipping or the ring template options turned on
2. With the ring corner flipping option turned on
3. With the ring template matching option turned on
4. With both the ring corner flipping and the ring template matching options turned on

In order to ensure exhaustive exploration of conformational space each ligand was docked 25 times using 10 genetic algorithm runs, resulting in 250 poses for each ligand. In all figures the best scoring poses are illustrated.

Results

One example where sampling ring conformational space proves to be crucial for obtaining the correct pose is calcipotriol binding to the vitamin D3 receptor (1s19). In this case the flexible ring does not form part of the central core of the ligand, but is located at one of the ends of the elongated ligand, see figure 1. In this case the ring conformation generated by Corina is not able to form key hydrogen bonds and as such the wrong pose is obtained when not sampling ring conformational space. The correct pose is obtained by either using the ring corner flipping or the ring template methodology.

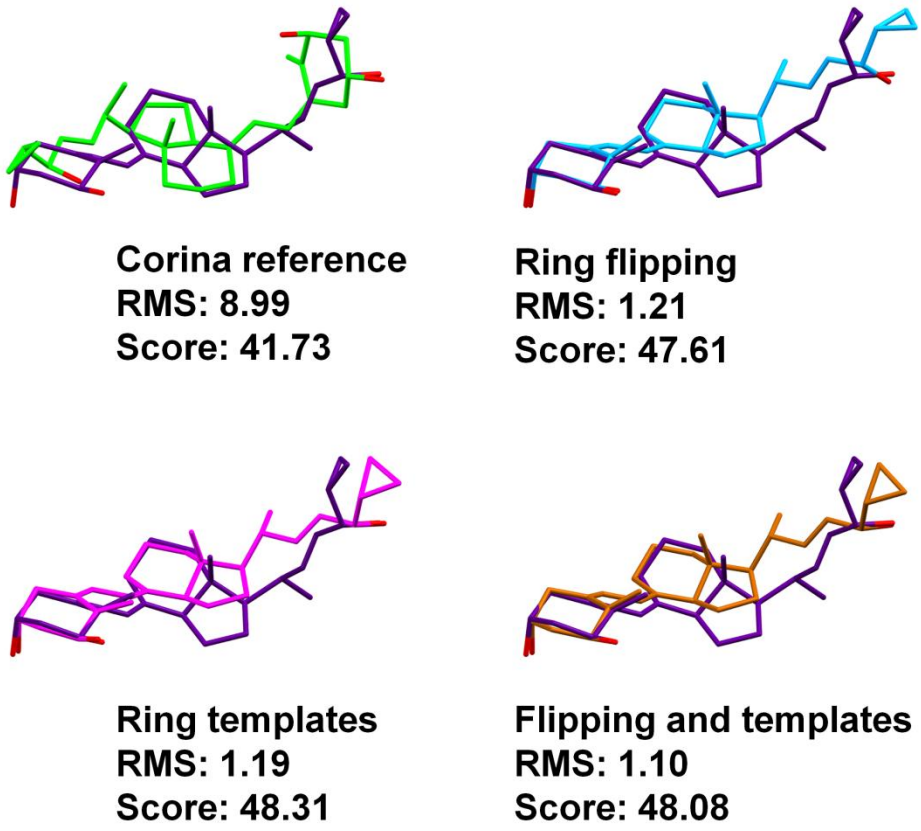

Figure 1 - Best poses of 1s19 based upon the fitness score using the ASP scoring function. In this example the Corina generated ligand conformation (green, top left hand corner) is not compatible with the binding mode of the ligand in the protein binding site as the conformation of the terminal ring does not allow it to make key hydrogen bonds. This is corrected by GOLD using either the ring corner flipping algorithm (blue ligand, top right hand corner), the ring template matching algorithm (magenta ligand, bottom left hand corner) or a combination of both (brown ligand, bottom right hand corner). In all cases the purple ligand represents the crystallographic pose.

Another example that illustrates how a flexible ring conformation can drive the overall pose is 2pcp. In this case a simple inversion of a six membered chair conformation is all that is required to obtain

the correct binding mode. As such using either the ring corner flipping or the ring template matching algorithms work equally well, see figure x6.

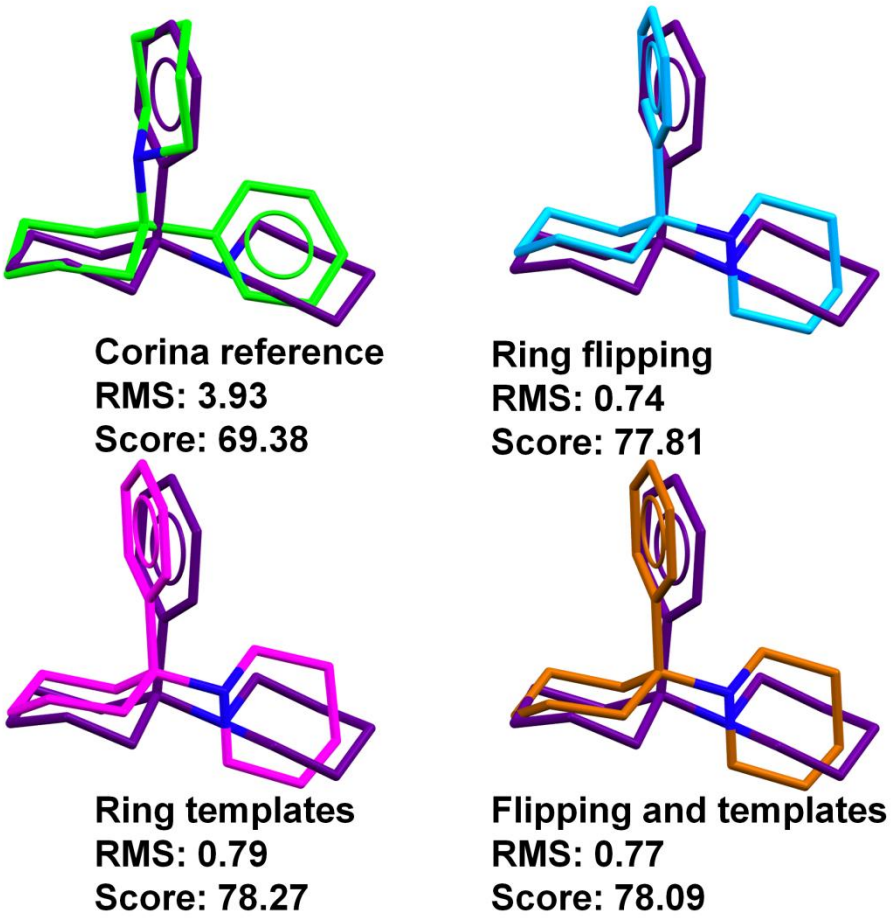

Figure 2 - Best poses of 2pcp based upon the fitness score using the GoldScore scoring function. In this example the Corina generated ligand conformation (green, top left hand corner) is not compatible with the binding mode of the ligand in the protein binding site as the left hand side ring is in an alternate conformation. This is easily corrected by GOLD using either the ring corner flipping algorithm (blue ligand, top right hand corner), the ring template matching algorithm (magenta ligand, bottom left hand corner) or a combination of both (brown ligand, bottom right hand corner).

However, in some cases the value of sampling ring conformational space is not necessarily apparent when using the overall RMS as a measure of success. Take for example 1ki8, docked using the PLP scoring function, in this case the use of the ring corner flipping algorithm results in an average fitness score increase of over 7 units. In this case this can be attributed to improved Van der Waals interactions (the shape fitting component of PLP) and the formation of an additional hydrogen bond. This is in contrast to the improvement in the RMS, which is relatively small 0.47 Å, see figure 3.

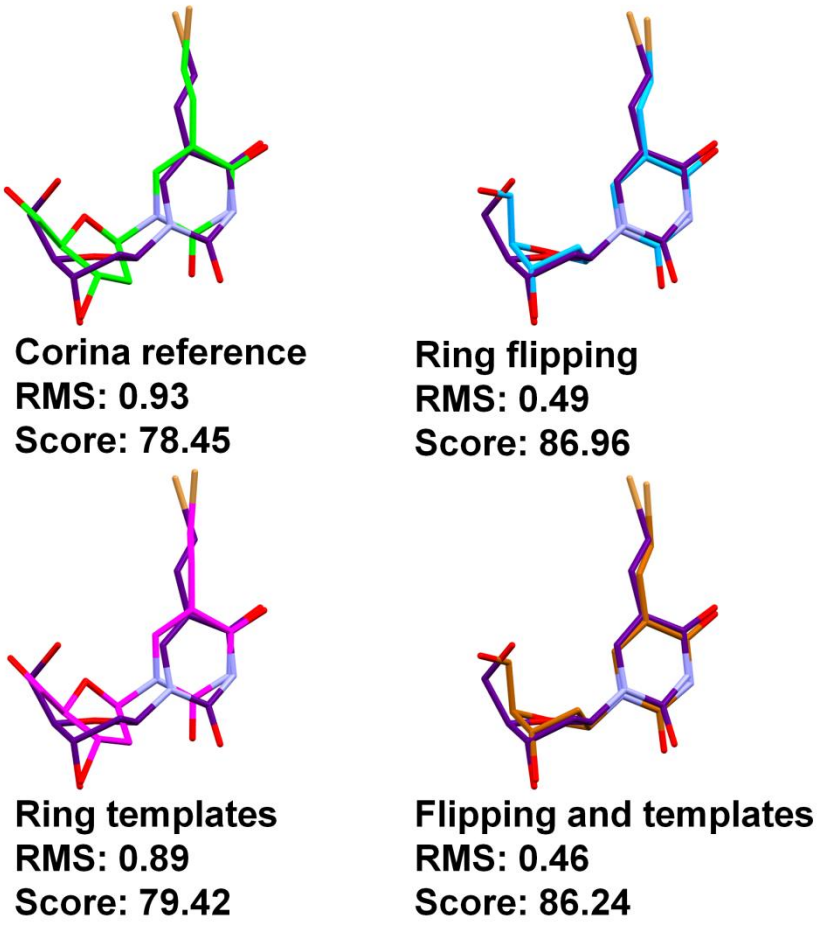

Figure 3 - Best poses based upon the fitness score for 1ki8 using the ChemPLP scoring function. Docking the Corina generated ligand without any ring flexibility (green ligand, top left hand corner) resulted in a good pose. Using the ring corner flipping methodology resulted in an even better pose (blue ligand, top right hand corner), with an astonishing improvement in the fitness score. The ring template methodology did not perform as well as the ring corner flipping (magenta ligand, bottom left hand corner). When the ring corner flipping and the ring template matching methodologies were used in conjunction (brown ligand, bottom right hand corner) GOLD successfully selected a pose generated using the ring corner flipping algorithm. Purple ligands represent the crystallographic pose.

What is also demonstrated by the 1ki8 example is the behaviour observed when using both the ring corner flipping and the ring template matching algorithms in conjunction. In the case of 1ki8 the improvement in the fitness score is much lower when using ring template matching than when using the ring corner flipping algorithm. When using the two algorithms in conjunction the appropriate pose is retrieved.

Another example highlighting the dramatic changes that can be achieved in the fitness score by small variations in the binding pose is provided by kainate binding to the glutamate receptor (1tt1). In this case the improvements in GoldScore with respect to the reference are 12.48, 14.58 and 20.50 using the ring corner flipping, the ring template and the combination of the flipping and template methodologies respectively. However, the differences in the RMSs are never greater than 0.45 Å,

see figure 4. In this case the improvement in the fitness score, with respect to the reference, is due to the hydrogen bonding term. Interestingly, the combined use of the ring corner flipping and the ring template methodologies does in this instance result in a ring conformation that was impossible to retrieve using either of the two methodologies independently.

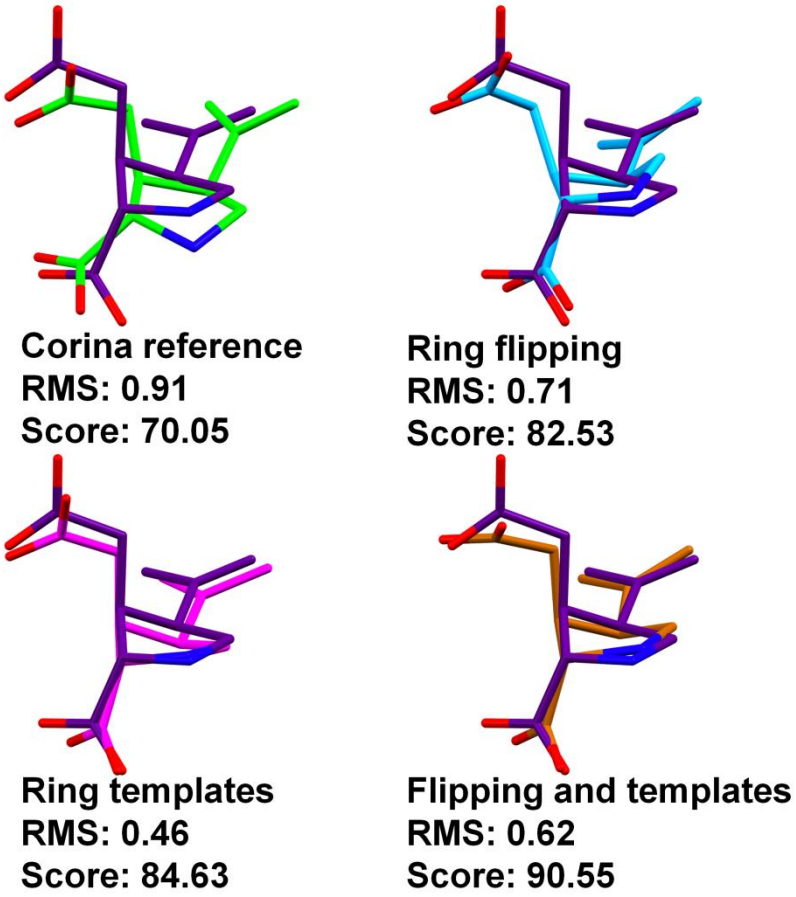

Figure 4 - Best poses of 1tt1 based upon the GoldScore fitness function. Docking the Corina generated ligand resulted in a good pose (green ligand, top left hand corner). However, both the ring corner flipping (blue ligand, top right hand corner) and the ring template matching (magenta ligand, bottom left hand corner) methodologies resulted in improved poses with greater fitness scores. Further using the ring corner flipping and the ring template matching methodologies in conjunction resulted in a pose with an even lower RMS and a fitness score 20 scoring function units greater. This increase in the fitness score can be explained by the charged ring nitrogen atom becoming able to donate two hydrogen bonds to carboxylate acceptors in the protein binding site.

Conclusions

In this study the ring corner flipping and the ring template matching algorithms present in the GOLD docking program have been investigated. The algorithms provide different approaches to the problem of sampling conformational space of flexible rings. The ring corner flipping algorithm has the advantage of being able to provide a systematic, although limited, search of ring conformational space. The ring template matching algorithm is to an extent less systematic in that it relies upon pre-

existing conformations being available. However it has the advantage that it can produce small subtle changes in the ring conformation, which may be required to optimize the interactions of protruding functional groups. The two algorithms can be used in conjunction.

In virtual screening experiments one is to a large extent dependent upon the fitness score generated by the docking program. As such the observation that sampling ring conformational space can lead to large improvements in the fitness scores, sometimes whilst the overall RMS does not change much, is of particular importance in virtual screening.

The ring templates included in GOLD, derived from the CSD, do not include any rings with more than seven atoms. Further, for a particular ring system the CSD might not have any representative structures. In such cases it is possible to point GOLD at ring template libraries derived using other methodologies. It is therefore envisaged that the ring template functionality will be of particular use for people trying to dock large macrocycles, such as macrolides. However, more work is still required in this area as the ring template matching algorithm in GOLD is not yet capable of dealing with fused rings.

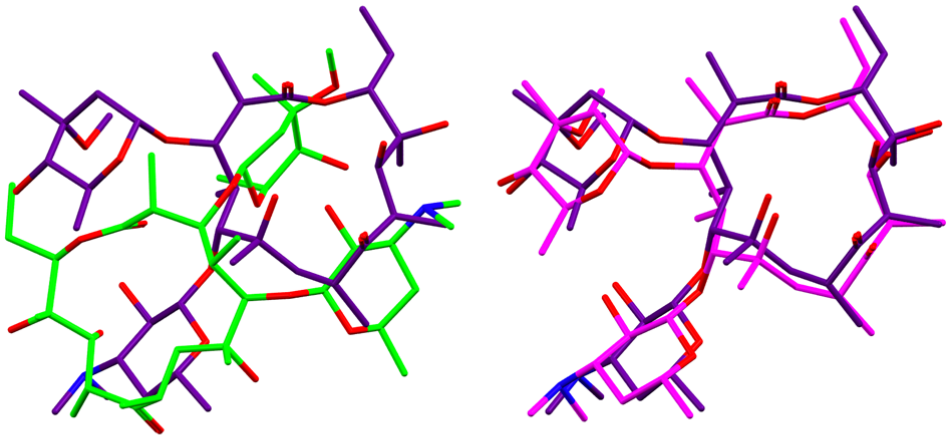

Figure 5 – Docking erythromycin into the the macrolide biosensor protein MPHR (3frq) using the ChemPLP scoring function. The initial conformation of the ligand was taken from the ideal model of erythromycin (http://ligand-expo.rcsb.org/reports/E/ERY/index.html). The macrocyle template used in this 'proof of concept' docking experiment was simply taken from the 3frq ligand structure. The green structure on the left hand side represents the pose obtained when not using any ring conformational sampling during the docking. The magenta structure on the right represents the best scoring pose obtained when using the ring template matching methodology. The purple structures represent the crystallographic pose.

References

1. H. Goto, E. Osawa. *J. Am.Chem.Soc., 1989,* **111**, 8950-8951
2. A.W.R. Payne, R.C. Glen. *J.Mol.Graph., 1993,* **11**, 74-91
3. G. Jones, P. Willett, R.C. Glen. *J.Mol.Biol., 1995,* **245**, 43-53
4. F.H. Allen. *Acta.Cryst.B.* 2002, **58**, 380-388

Products

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

Relibase+ - an essential tool for searching, exploring and comparing all protein-ligand data from public and in-house data sources.

GOLD – an accurate and reliable protein-ligand docking program

Hermes – CCDC's life science visualiser, used by GOLD, GoldMine, Relibase+ and SuperStar

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