



[Artificial Receptors](#)

[GOLD SAMPL](#)

[GOLD Validation:
Virtual Screening](#)

[Optimum Docking &
Rescoring Protocols](#)

[Purchasing Information](#)

[GOLD Validation Tests](#)

[GOLD References](#)

Docking with flexible side chains in GOLD: Cross-docking and Virtual Screening

Background

A challenge recently taken up in the development of protein/ligand docking tools, is to take account of conformational changes of the protein. It is well understood that the protein and ligand each influence the conformation of the other, a phenomenon referred to as induced fit. Even very small movements in the active site such as a single side-chain rotation can have large consequences for ligand-protein docking and hence for structure based drug design. Many of the current docking programs such as GOLD are known to perform well in so called self-docking experiments, where a ligand is extracted from its co-crystal, its conformation randomised, and it is then docked back into its native protein structure. However, when the extracted ligand is docked into the same protein but using a different crystal structure, the performance drops dramatically. The success rate for these types of dockings, commonly referred to as cross-dockings, is substantially lower than for self-dockings. Since one of the most common application of docking programs is to dock several different ligands to the same target and rank them in order of binding, most routine dockings are essentially cross-dockings. This decrease in docking accuracy is often caused by slightly different conformations of the protein and could be addressed by using a less rigid model of the protein structure. Allowing for different conformations of the protein is also a way to deal with other problems as well. For instance they may be useful for dealing with poor resolution models and the uncertainty of side-chain conformations in homology models. In this case-study we will describe a cross-docking experiment that incorporates a conformational change of the protein in the form of side-chain rotations in the active site. We will also show the effect of protein flexibility in a virtual screening experiment. It is important to note that VS-runs using a single protein to represent the target structure are essentially cross-dockings.

In GOLD there are several ways to treat conformational changes of the receptor. These comprise, the use of 'soft' potentials for vdW-interactions, allowing specified side-chains to rotate, or docking the same ligand to several different target structures and analysing the resulting data. The latter option is not an automatic procedure and several individual docking runs with different protein structures are required. This case-study will focus on the use of flexible side-chains in the active site of the target. Our test-set contains two well-known therapeutic targets, the serine protease factorXa (fXa) and the kinase CDK2. Six structures of fXa from the PDB were used for a cross-docking experiment and one structure of CDK2 was used for a virtual screening experiment using the decoy/active set from the DUD database.

Cross-docking FactorXa

Six structures of the serine protease factorXa were downloaded from the PDB: 1EZQ, 1KSN, 1NFW, 1NFX, and 1NFY. The six structures were overlayed and all waters except key waters in the S1 and S4 pocket were removed prior to docking. The retained waters were specified as active in GOLD so that they could be handled automatically during docking (for more information about docking with waters, please see the GOLD manual). The six different ligands were extracted from their native crystal structures and prepared for docking. There are a few key issues with docking and they can roughly be divided into three areas:

1. The docking problem: Ensuring that the binding pose of the ligand will be sampled and found
2. The scoring problem: There exists no exact energy function to evaluate the ligand-protein interaction so the correct solution is not necessarily the best scoring
3. The protein flexibility problem: The rigid receptor is only an approximation to a plastic protein

Since we want to address only the final problem in the present case study we have endeavoured to limit the influence of the docking and scoring problem as much as possible. In order to allow enough time to do an exhaustive conformational search of the ligand we allow 100 GAs per ligand instead of the default 10. We have also relaxed the scoring criteria, allowing the 'correct' solution to be within the top 10 ranked solutions. Studying the overlayed structures of the active site of fXa

(See Fig. 1) it is apparent that there is very little movement in the active site, we should thus expect a high success rate for the cross-docking even when using a rigid model of the protein, the results of the rigid cross-dockings are presented in Table 1.

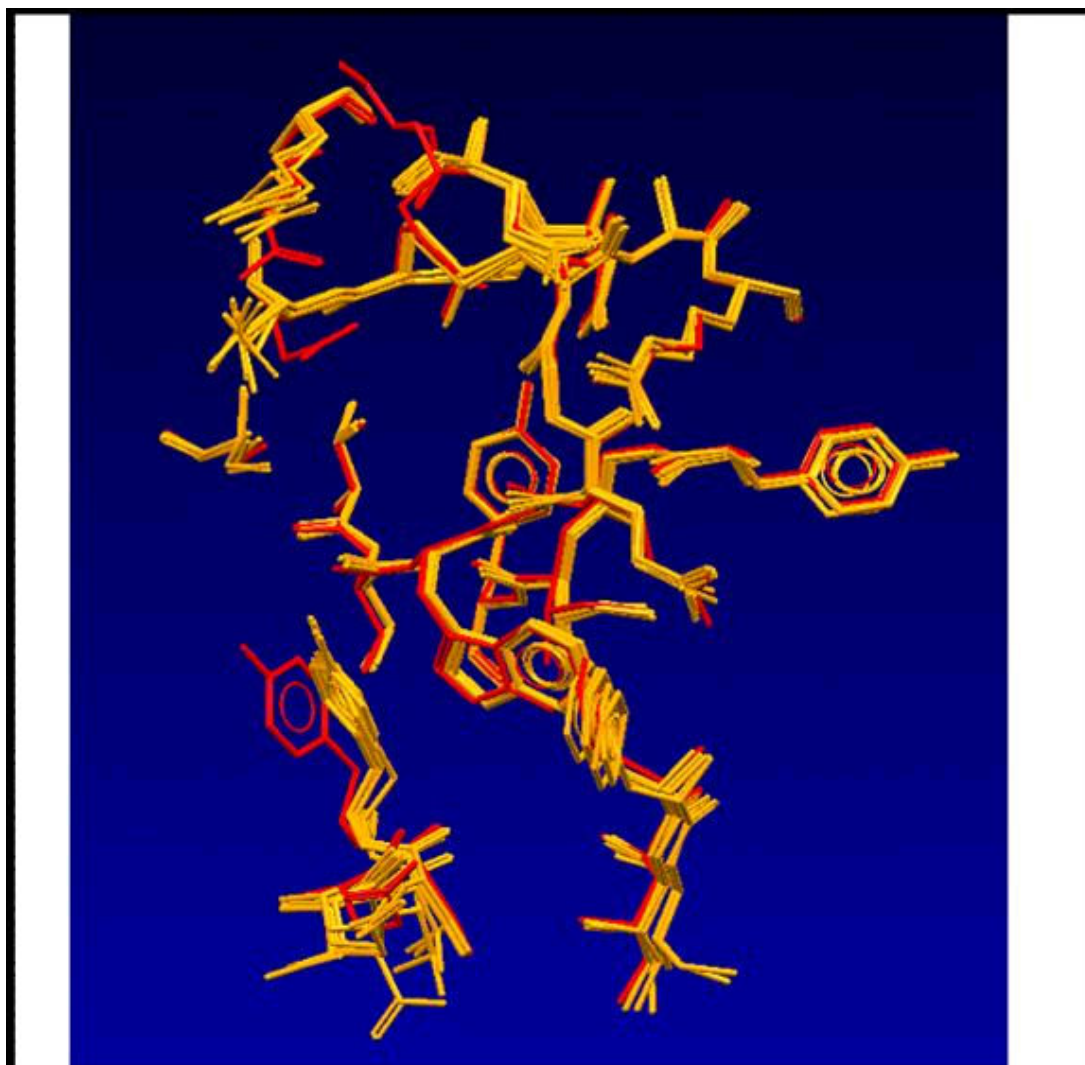


Figure 1. The active site of fXa overlaid, with 1XKA in red.

	1EZQ	1KSN	1XKA	1NFW	1NFX	1NFY
<i>lezq</i>	0.5	0.6	2.8	0.7(2)	2.0(2)	0.8
<i>lksn</i>	1.0	1.0	2.5	0.9	1.1	0.7
<i>lxka</i>	1.7	1.7	1.6	1.1(2)	1.8	1.6
<i>lnfw</i>	1.6	1.0	1.7	1.1	1.1	1.0
<i>lnfx</i>	1.3	1.5(6)	2.0	1.2	1.1	1.0
<i>lnfy</i>	1.0	1.9(6)	2.0(3)	0.9	0.9	0.9

Table 1. RMSD compared to crystal conformations for the rigid cross-docking to factorXa. Grey solutions have the correct solution ranked as 1, the other rankings are given in parenthesis. Failed dockings are given in bold.

As a general rule, a correct docked structure should not have an RMS deviation from the reference structure of more than 2Å. In the case of fXa there are only two docking solutions that fail to meet this criterion and these are the docking solutions of the ligand from 1EZQ and 1KSN to the protein structure of 1XKA. None of the 100 solutions show an RMSD lower than 2Å, with the closest solutions having an RMSD of 2.8 and 2.5Å, respectively. The overall success of the rigid

cross-docking indicates that there is not much movement of the active site in fXa, however this could be due to structural similarity of the ligands. Going back to the overlay of the protein structures it can be seen that 1XKA (red structure in Fig. 1) deviates from the other 5 crystal conformations. The side-chains of Gln192 and Tyr99 occupy a different rotamer, using this information we can set up flexible docking in GOLD (for more information of how to set up flexible docking in GOLD, please see the manual). GOLD uses a pre-defined rotamer library (ref) for the different conformations of side-chains, user defined conformations can be added to or replace the default rotamers. In this example we will specify the two side-chains Gln192 and Tyr99 to be flexible and we will use the default rotamer library. All of the cross-dockings were re-done using the flexible docking (see Table 2).

	1EZQ	1KSN	1XKA	1NFW	1NFX	1NFY
<i>1ezq</i>	0.6	0.8	0.9(3)	0.9	2.0	0.8
<i>1ksn</i>	1.0	0.9	0.9	0.7	0.8	1.2
<i>1xka</i>	2.0	1.8	1.4	1.9	1.5(2)	1.3
<i>1nfw</i>	1.7(9)	2.0(3)	0.8	1.5	1.0	0.8
<i>1nfx</i>	1.1	1.1	0.8	1.0	1.1	1.1
<i>1nfy</i>	1.1(6)	1.1(4)	0.6	0.8	1.0	1.0

Table 2. RMSD compared to crystal conformations for the flexible cross-docking to factorXa. Grey solutions have the correct solution ranked as 1, the other rankings are given in parenthesis.

Incorporating several rotamers for Gln192 and Tyr99 allow us to find the correct solution for 1XKA in the top 10 ranked solutions. Comparing Table 1 and 2 shows a slight difference in the rank of correct solutions, however the overall docking success has increased. Allowing side-chains to rotate will increase the conformational search space for the docking, and it is encouraging that previously successful dockings do not fail with the use of flexible side-chains.

Virtual Screening CDK2

Kinases are important for signal transduction pathways involved in cellular regulation, leading to their interest as therapeutic targets for the treatment of cancer. CDK2 in its monomeric form is inactive and is activated in two steps, each step leading to conformational changes of the binding site. The conformational changes will also lead to backbone movements, especially important in this regard, is the b-strand of residues 81-83 linking the two domains of CDK2, commonly referred to as the hinge region. The adenine base of ATP forms several protein bonds to this portion of the backbone, and most if not all inhibitors have a double-ring system in order to mimic these interactions. Since CDK2 structures show significant movements of this b-strand we cannot expect that our flexible docking, which only includes side-chain movements, will be as successful as for fXa cross-docking example. The highly flexible side-chains Lys33 and Lys89 have been identified to be important for ligand binding with Lys33 interacting directly with ATP and facilitating the change in conformation between in-activated and activated CDK2. We used here the structure 1JSV from the PDB, this structure is crystallised with a small fragment and the side-chain of Lys33 is extended into the ATP binding pocket and will probably prevent larger ligands to bind properly if the structure is used for rigid docking.

We have chosen to perform a virtual screening experiment using the popular DUD ligand set. First we performed two virtual screens with the rigid structure of 1JSV, the first VS run uses GOLDScore with 10% automatic GA settings, the second VS run uses the kinase ChemScore energy function with hydrogen bond constraints to the hinge region of CDK2. The second run was done to provide a result an estimate of the most efficient kinase virtual screen that might be possible with the rigid 1JSV structure. This was followed by a virtual screen where we allow the side-chain of Lys33 to occupy different rotamers, again using GOLDScore without constraints. The results are presented in Table 3.

	ROC_AUC	ER(1%)	ER(10%)
CDK-2 – Rigid	0.56	8.0	2.2
CDK-2-CS-Rigid Constraint Leu83	0.59	-	-
CDK-2 – Flexible Lys33	0.73	10.2	3.2

Table 3. Virtual screening of CDK2 with DUD.

Using the area under a receiver operating characteristic (ROC) curve as the measure of success it is clear that the flexible docking outperforms both rigid runs. The ROC curve can be represented by plotting the fraction of true positives fraction of false positives and is a way to select optimal models, a value of 0.5 indicates a random result. As can be seen from Table 3, both rigid models are slightly better than random while our flexible VS experiment improves the area to 0.73 and also improves the enrichment. An important goal in early stages of drug discovery is to find diverse ligands that bind to a target, leading to several series of compounds that can be further developed to lead compounds. A quick visual comparison of the results for the rigid and flexible dockings show that the diversity of ligands are much improved for the flexible docking (see Fig. 2a and b).

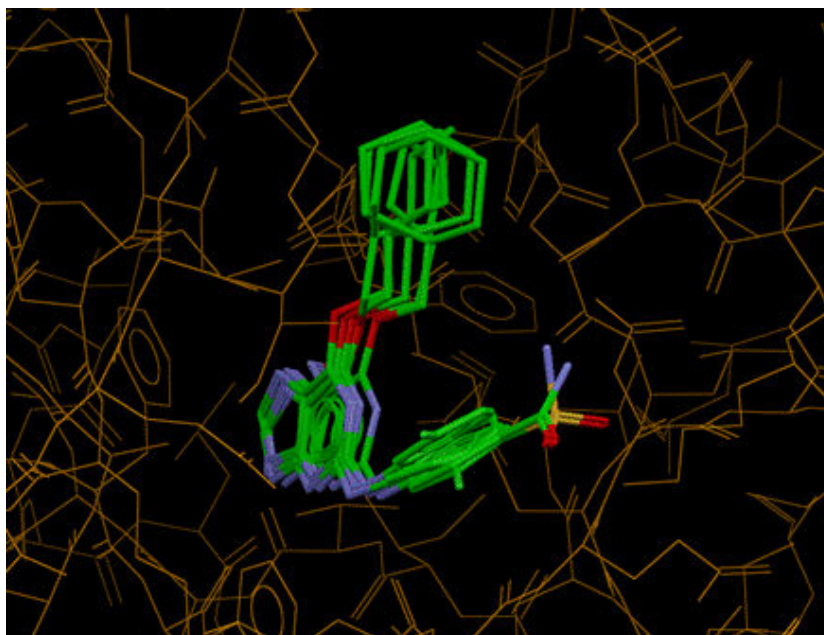


Figure 2a. Active ligands in top 1% for the rigid docking.

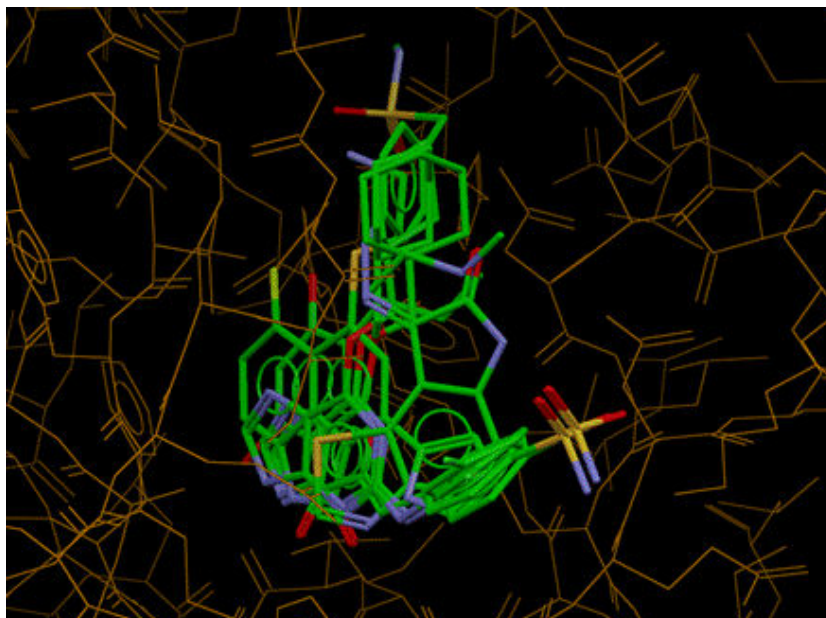


Figure 2b. Active ligands in top 1% for the flexible docking.