Schrödinger Workshop
March 25th, 2010
Focus on Advanced Structure Based Design

Introduction

In this workshop we will mainly focus on setting up and performing a virtual screening project: starting from choosing the target structure, prediction, validation, and characterization of the target structure. We will use factor Xa as our example. Factor Xa is a serine protease involved in blood coagulation. Direct inhibitors of factor Xa are being developed as anticoagulants. Most recently, Rivaroxaban developed by Bayer HealthCare has been approved by FDA. Other factor Xa inhibitors on the market, or in clinical phase are Apixaban by BMS, and Betrixaban by Portola Pharmaceuticals/Merck.

Please note, as this is an advanced workshop, the hands-on instructions assume basic knowledge of the Maestro GUI. While the calculations are running, you can look at the results that are stored in the prjzip files.

Target structure prediction and refinement (FXa_homology_model_refinement.prjzip)

Structure prediction

Before embarking on structure based design, we need to have a crystal structure, ideally of high resolution and high quality, carefully solved by a crystallographer. Often we don't have that luxury. When there is no crystal structure available for our target, we need to look for homologue protein structures, based on which we can generate a homology model. Here we are going to build a model of factor Xa based on either bovine factor Xa (1KIG) or human thrombin (2BVR).

The workflow can be summarized as follows:

1) Open Prime -> Structure prediction and import the fasta sequence
2) Either import the template structures or search in BLAST database. BLAST search will retrieve many hits since there are plenty of factor Xa crystal structures around.
3) Out of curiosity, please run Find Family. Look at the classification of the query sequence, does it make sense?
4) Choose for example the 1KIG template and then open the QuickNDirty run. 1KIG has 85% identity with human factor Xa and no gaps, yet it is a bad choice for a template.
5) In Edit Alignment step, use the scissor tool to clip away the part of protein that does not have a template aligned.

6) Go to the final step and click on Quick Build button and you should have a model within a minute.
7) If you choose the 2BVR template (run2 in the project), please run SSP. SSP gives you two secondary structure prediction (PSIPRED needs to be installed separately). When your query sequence and template structure have low sequence identity (<30%), you should consider click on Align, which takes into account the secondary structure predictions and protein family information. Here we don't need to do Align because sequence identity is ~40% and the alignment is reasonably clear.
8) Find the gaps and rearrange the gaps if needed.
9) Go to Build step and click on build.
10) Once the model is finished, export to project table and now you have your raw model of the target protein.

The model needs to be refined. Please use Ramachandran Plot and Protein Report in Tools menu to identity areas that need attention. Use Scripts -> Workspace Tools -> Select disallowed sidechain/backbone to select regions with problems and use Prime -> refinement to refine your model.

Loop refinement
In the alignment there is a quite large gap. Look at the alignment and the secondary structure prediction. Note that the loop was predicted to be a helix. We need to refine this loop as the coordinates for this loop are likely to be different.

Please open Prime -> Refinement and choose Task: Refine loops. Click on "Load from Workspace" button to load the loops. Find loop where the gap is located and toggle on "Run". Open the Options... We shall stick with Default protocol, but increase the "Maximum number of structures to return" to 10. We will also untoggle "Unfreeze side chains within". When you launch the job, Prime will warn that you might want to use a more extended protocol but please ignore.

Look at the results and compare the "real" factor Xa structure (1F0R in the project) with your predicted model.

Target structure preparation and characterization (FXa_virtual_screening.prjzip)

Preparation (Protein preparation wizard)
"Unfortunately" after you published your factor Xa model structure, new factor Xa crystal structures started to appear. For your virtual screening, the homology model was not good enough. Particularly in the loop region discussed earlier. For the virtual screening project you decided to go for 1F0R with a 2.1 Å resolution. Before you start to do a virtual screening, you need to prepare the protein structure.

Please open up the Workflow -> Protein preparation wizard and complete the workflow. It is important that you generate the different states for the ligand and try to think about which state you should choose. It is equally important to examine the results of hydrogen bonding network optimization. Note,

there is a water molecule that mediates hydrogen bond between the ligand and the Asp residue in the S1 pocket. Is it important to keep the water molecule?

Characterization of the binding site (SiteMap)
Once you have a prepared structure, one should ask the question "Is this target druggable?" and "What are the characteristics of the binding pocket?" The tool you will use is called SiteMap, which you can find under the Applications menu. Load the prepared 1F0R into Workspace and choose "evaluate a single binding site region" and click on the ligand. You can look for other possible sites, by first deleting the ligand (make a copy of your structure first) and run SiteMap in "Identify top-ranked potential receptor binding sites" mode. Look at the potential surfaces and evaluate whether the ligand matches well. The Site Score for a druggable site is normalized around 1.0.

Small molecule docking and virtual screening (FXa_virtual_screening.prjzip)

Self-docking: Reproduction of PDB binding mode
The first step in preparing for small molecule docking will be reproduction of the crystal ligand binding mode. In protein preparation you have noticed a water molecule in S1 pocket that mediates hydrogen bond to Asp residue. The decision to keep or discard the water for virtual screening is always difficult when there are no other crystal structures available. If you have time, try to dock 1F0R ligand and the 1FJS ligand into 1F0R with or without water. In the prjzip file, you will find the result for the water-less self-docking. Compare the results with 1F0R crystal structure and 1FJS crystal structure. The docked binding mode is similar, yet still different. Why?

Preparation of the ligand collection
Ligands need to be prepared by assigning ionization states and tautomer states. The penalties can be added to Glide Score and improve the enrichment. Please open LigPrep in the Applications menu, load in FXa_all.mae. Ionization states can be predicted using Epik. In our case we don't need to care about the stereoisomers, hence you can set LigPrep to "Determine chiralities from 3D structure". The result can be found in FXA_ligprep_out.mae.

Virtual screening and post docking processing.
Open Workflow->Virtual Screening Workflow, untoggle "Prepare ligands", and setup the rest of the panel. The final result from XP docking is loaded in the project. Use Workspace Styles to browse the docked poses. The active molecules have value 1 in the property "Active". How many actives are there in the top 95 hits? (Answer: there are 146 in total)

For post-docking processing you can try:

Scripts -> Docking Post Processing -> Pose filter

Scripts -> Cheminformatics -> Interaction fingerprints

Scripts -> Cheminformatics -> Canvas similarity and clustering

If time allows, please run Prime -> MM-GBSA. Are there any improvements on the ranking?

Fragment docking and structure based pharmacophores (FXa_fragment.prjzip)

To do fragment docking the Glide settings need to be changed.

Open the Glide Ligand Docking panel and there make sure that" XP mode" and "Write XP descriptor information" are toggled on. Go to "Advanced Settings… ". Change to "Keep 10000 poses per ligand", "Scoring window for keeping initial poses: 500.0" and "Keep best 1000 poses per ligand". Toggle on "Use expanded sampling". Go to output tab and request "Write out at most 100 poses per ligand".

Docking 400 fragments is going to take long time with these expanded sampling settings. The results are in the project. Quickly scroll through the docked fragments to get a sense how well the pocket is sampled. Go to Scripts->Docking post processing->Score Phase Hypothesis by XPDes. Point to the output PV file, toggle on "Fragments", "Perform clustering" and" Create receptor based excluded volumes". Click on Create hypothesis. This will take about 5 min. If you look at the features and their ranking, do they correspond to what you expect?

Instead of doing the fragment docking, take the 1F0R crystal ligand conformation and do a score in place with XP mode and write out the XP descriptor information. Use that output to generate a hypothesis, compare both hypothesis.