Assessment and Preparation of Crystal Structures for Drug Design

Andrew Henry, Hooman Shadnia, Alain Ajamian, Audrey Bonin, Mike Drummond, Chris Williams

Chemical Computing Group
St Johns Innovation Centre, Cowley Road, Cambridge CB4 0WS, UK, and 1010 Sherbrooke St. W., Suite 910, Montreal QC, Canada H3A 2R7

Chemical Computing Group

Abstract

The wide availability of X-ray protein-ligand crystal structures has made Structure-Based Drug Design (SBDD) an integral part of the drug development process. However, most crystal structures need to be examined and processed before they can be used in modeling studies. Aspects of crystal structures that commonly require inspection and processing include:

- Protonation states of the ligand and pocket residues
- Reconstruction of missing loops and atoms
- Confirming the conformation of the bound ligand
- Analysis of protein motion via temperature factors
- Comparison of model electron density vs experimental electron density
- Analysis of the crystal contacts
- Analysis of solvent location and energy

Here we describe the use of MOE as an integrated platform to address all of the above points and to produce an all-atom 'model' structure suitable for SBDD and molecular dynamics studies. The PDB entry 1XKK which contains the EGFR tyrosine kinase domain bound to a small-molecule inhibitor lapatinib will be used as the example system.

Protonation and Tautomerization States

Hydrogens are unresolved in most crystal structures so the protonation states of bound ligands need to be set. MOE uses a modified PROPKA [1] based method to analyze possible ligand protonation states in the context of the binding pocket residues and associated solvent molecules. **Figure 1** shows the possible protonation states of lapatinib and their corresponding populations at pH=7. Users can choose which protonated state to retain for modeling studies.

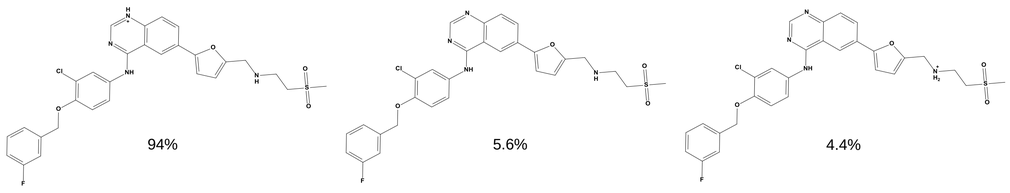


Figure 1. Protonation states of the ligand lapatinib at pH=7 (PDB code 1XKK).

Reconstruction of Missing Loops and Atoms

Crystal structures are often missing side-chain atoms and loops. These regions must be repaired in order to produce an all-atom model suitable for SBDD and molecular dynamics. Missing regions are reconstructed using the MOE **Structure Preparation** tool which uses the MOE **Loop Modeler** to reconstruct missing loops (using either existing crystal structures as templates, or via a *de novo* loop building algorithm [2-6]), and a side-chain packing algorithm to place missing side chains. Structures are scored with a GBVI potential energy function which evaluates interaction with the environment and internal strain [7].

The missing loops are shown in red on the 1XKK sequence in **Figure 2a** . **Figure 2b** shows the 3D structure with the added loops in red. **Table 1** shows the PDB code, chain letter, and residue numbers (UID) of the PDB templates used to build each loop. Due to low sequence similarity with the templates (less than 20%) the GBVI score was used to choose the best loop models. The repaired loops may not be close to the binding pocket but they are still required for holistic studies of the complex such as molecular dynamics simulations.

Loop Number	PDB Code	Chain Letter	Residue numbers
1	1H32	A	1-31
2	3P02	A	192-195
3	3N5K	A	23-26
4	2EV9	A	164-171
5	3FO3	A	81-87
6	3BOJ	A	54-58
7	3PS0	A	216-219

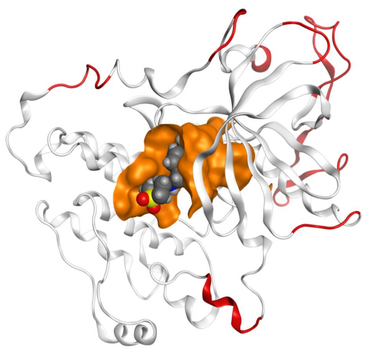


Figure 2a. The EGFR sequence from PDB:1XKK. The missing loops are highlighted in red.

Figure 2b. 3D structure of the PDB:1XKK with the missing loops built in place. Receptor backbone is shown in grey cartoons, with the reconstructed loops colored in red. Pocket surface is displayed in orange, and the ligand atoms are displayed in element colored space-filling model. None of the missing loops are close to the pocket.

Pocket Interactions and Conformations

Hydrogens are added to the system using the **Protonate3D** [8] application, which assigns ionization states based on pH, adds explicit hydrogens to the entire system and then optimizes the H-bond network. The protonated and optimized system is then examined to reveal important protein-ligand interactions. These interactions are abstracted into a convenient 2D depiction (**Figure 3**) which shows lapatinib forming an H-bond donor (Gln791), an H-bond acceptor (Met793) and a proton-pi stacking (Leu718) interaction with the pocket. The pocket wall is represented as a contour in 2D. The contour shows a tight fit of the pocket around the fluorophenyl tail. The purple halos indicate solvent exposure, and highlight the exposure of the sulfone tail (which also interacts with Asp800 via a H-bond network).

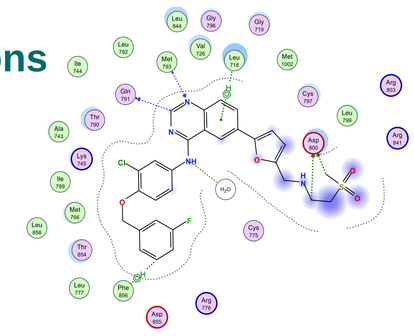


Figure 3. 2D Ligand Interactions. For details of legend and annotations, please see the MOE manual.

Analysis of Temperature Factors

The Debye-Waller temperature factors provided with crystal structures in PDB format reveal atoms with high degrees of motion corresponding to positional and conformational certainty. The conformations of these highly dynamic moieties can be further analyzed and improved using the **Torsion Analysis and Rotamer Explorer** applications.

As shown in **Figure 4** the red colored atoms have higher B-Factors, including the sulfonyl tail of the ligand as well as the pocket residues Met766, Leu858, and Met 1002.

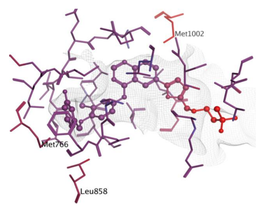


Figure 4. Temperature factors of pocket residues. Blue-purple colors are cooler i.e. less mobile and red colors indicate high temperature, i.e. higher mobility. The pocket surface is displayed in grey mesh. The ligand is rendered in ball and stick mode. The sulfone tail of the ligand as well as pocket residues Met766, Leu858, and Met 1002 exhibit higher than average temperature factors.

Pocket Side Chain Exploration

Once the highest mobility residues are located using the temperature factors, the **Rotamer Explorer** application can be used to identify optimum conformations. This application quickly scans through a pre-generated library of rotamers, evaluating their interactions with the environment atoms. An example is shown in **Figure 5a** where the crystal conformation of Met766 exhibits steric clashes with the ligand, shown by the orange disc. An alternative conformation found by the **Rotamer Explorer** in **Figure 5b** exhibits very favorable Van der Waals interaction energy of -10.9 kcal/mol with neighboring hydrophobic residues Ala:763, Leu:777, Phe:856, Leu:858.

Figure 5a. Met766 in crystal conformation. The side chain exhibits steric clashes with the ligand, shown by the orange disc.

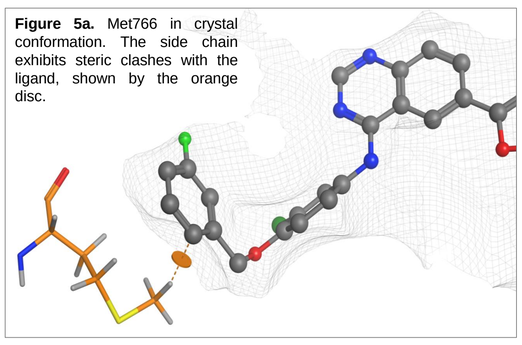
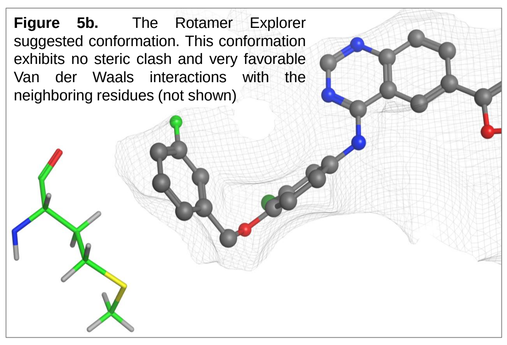


Figure 5b. The Rotamer Explorer suggested conformation. This conformation exhibits no steric clash and very favorable Van der Waals interactions with the neighboring residues (not shown)



Torsional Analysis

The highly mobile areas of the ligand and other biomolecules can be analyzed via the **Torsional Analysis** application. This application rotates selected bonds via rigid or flexible force-field search and can seamlessly take advantage of crystallographic data obtained from the Cambridge Crystallographic Database software Mogul [9] as well as quantum mechanical software Gaussian [10] where available.

Figure 6a demonstrates three bonds of the ligand tail subjected to torsional analysis. Here the rigid rotation energy of the free ligand is in dotted grey lines, that of the ligand in the context of the pocket is in solid grey lines and the crystallographic frequencies found via Mogul are in blue histograms. The green vertical line shows the current (crystallographic) torsion angle. It is clear that none of the three torsions are in their most highly observed configurations in crystal structures. By choosing the highly observed values for each bond, the conformation shown in cyan (**Figure 6b)** is generated, which is very close to the original conformation (grey carbons) with a heavy atom RMSD of 0.65Å.

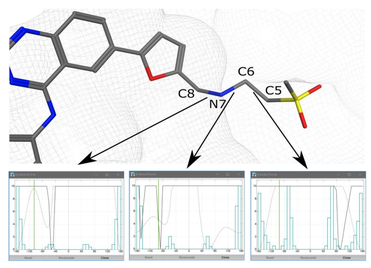


Figure 6a. Torsional analysis of the ligand tail. None of the selected bonds are in highly observed crystallographic conformations. See text for details.

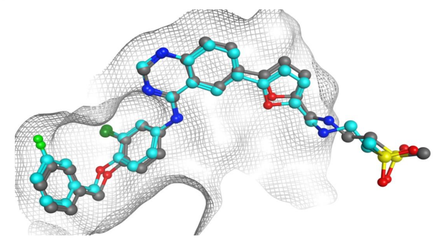


Figure 6b. The torsion improved conformation of the ligand. The modified ligand is shown with cyan carbons. The overall position of the ligand, including the sulfone tail is relatively unchanged.

Comparison of the Model vs Electron Densities

Visualization of electron densities along with the structure is a standard practice in quality control of crystal structures, and can be fully conducted using the graphical interface in MOE. **Figure 7a** demonstrates the F_o electron density map for ligand at 2σ level. Most of the core and the tight-fitting tail of the ligand i.e. the left hand side is well resolved as the blue density blobs cover the structure. The sulfone tail is less well resolved, suggesting the need for improvement. To facilitate visualization of the densities for the entire pocket, MOE can take advantage of a range of 3D hardware, including the inexpensive Anaglyph goggles [11]. This is shown in **Figure 7b** for the green-magenta 3D goggles.

The **Difference Maps** can be rendered by computing the simulated electron density for the current structure, and subtracting that from the experimental densities. **Figure 7c** displays this for the ligand, as well as pocket Asp800. Here a negative blob (pink) covering the sulphur and oxygen ligand atoms suggests that the density is unexpectedly low in this area, assuming that conformation is correct and rigid. This is simply due to the mobility of the ligand tail, as discussed previously. Another negative blob over the side chain carboxyl carbon of Asp800 is likely due to the same thing. Altogether these suggest that the ligand sulfone tail-Asp800 interaction is fairly strong but mobile, which is expected from solvent exposed hydrogen bonds.

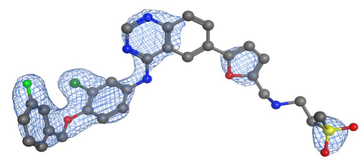


Figure 7a. Fo electron density of ligand. The sulfone tail is poorly resolved, due to high mobility, hence the need for alternative methods to improve this conformation.

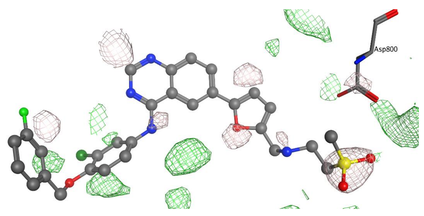


Figure 7c. The Difference ED Maps. The Asp800 carboxyl carbon and ligand tail show negative (pink) densities. This suggest the two are mobile and interact strongly.

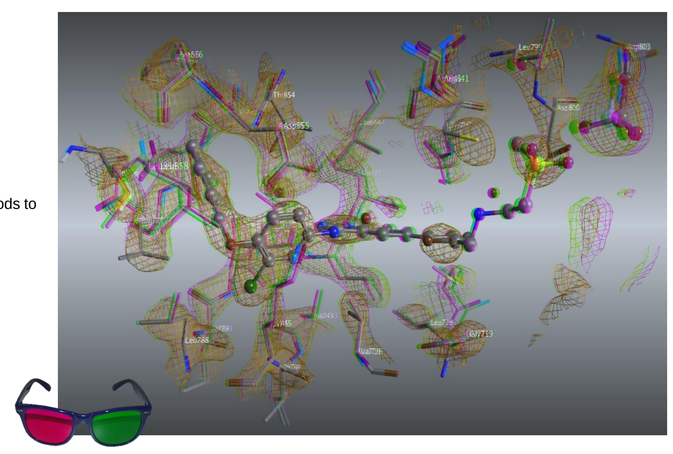


Figure 7b. A 3D image of the pocket Fo electron density. This image is rendered for green-magenta 3D goggles. Most pocket side chains and ligand atoms overlay the experimental densities. Exceptions include the sulfone tail of the ligand, as well as Asp800 Cβ.

Analysis of Crystal Contacts

Crystallographic atoms in the vicinity of neighboring asymmetric units incur interactions which are artefacts of crystallization. Thus, it is important to be mindful about their location when discussing ligand-protein interactions, especially if they are in the vicinity of the pocket or other areas of interest.

Figure 8a shows the **crystal contacts** in PDB:1XKK. Here the receptor surface for residues involved in crystal contacts is shown in green, while other receptor residues are rendered in gray and the ligand surface is displayed in magenta. Additionally, the molecular surface of the neighboring units is shown in red halo. **Figure 8b** focuses on the active site. Here Asp800, Asp803, Met1002, Leu718 and Gly719 are pocket residues involved in crystal contacts and Asp942, Arg 977 and Leu941 belong to the neighboring unit, and they are displayed in red.

Once again, these suggest that the conformation of the sulfone tail of the ligand and the residues at the mouth of the pocket may not pertain to the aqueous conformation.

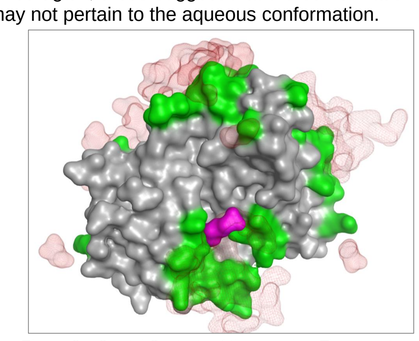


Figure 8a. Crystal Contacts of the complex. The receptor surface is displayed in gray. The crystal contacts are colored green. Ligand is shown in magenta. Residues of the neighboring unit cell are displayed in red halo.

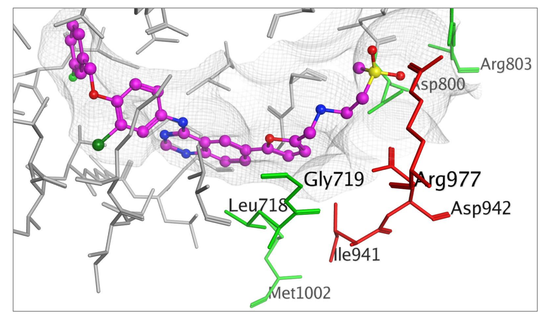


Figure 8b. Pocket Crystal contacts. Pocket crystal contacts are colored green. Residues of the neighboring unit cell are displayed in red. Ligand carbon atoms highlighted in magenta.

Analysis of Solvent Conformations and Energetics

Crystallographic waters can be crucial in understanding ligand-pocket interactions. However, uncertainty over their location (e.g. in low resolution crystal structures) as well as their energetics may result in ambiguity about their role. The **Solvent Analysis** application uses the 3D-RISM method [12-14] to assist with **water placement** and estimation of solvent effects and energetics in protein-ligand binding.

Table 2 lists the predicted water sites near crystal waters in the binding pocket. Here each site is identified with an identification number. Various measurable properties are displayed as described in the MOE manual. A 3D representation of these sites is shown in **Figure 9a**.

Here for example, the four water sites 292, 294, 295 and 379 are associated with a single resolved crystal water. The closest site, number 292 which is 0.26Å away from the resolved crystal structure, interacts with ligand N22 but it carries a unfavorable free energy of +6.3 kcal/mol and is computed to be of low occupancy of 0.5. This suggests that the interaction is likely not crucial, and this water molecule can be displaced.

The predicted sites 260, 269 and 357 are close to three resolved water molecules. Their low occupancies suggests that they comprise a water channel. This can be visualized using the oxygen densities shown as the blue surface in **Figure 9a**. A top view of the protein surface in **Figure 9b** confirms the existence of this tunnel containing resolved waters, leading to N20 of the ligand.

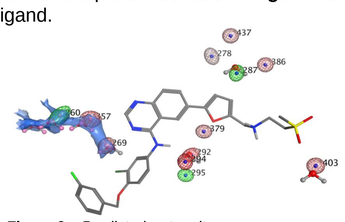


Figure 9a. Predicted water sites. See text for details.

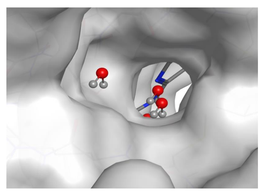


Figure 9b. Water channel. Protein surface of the water tunnel from **Figure 9a** containing waters leading to the ligand

ID	Occ	dG	Vol	Aniso	HB Tot	HB Lig	Xtal	Dist
260	0.83	-8.09	1.38	0.23	0	0	HOH27	0.77
269	0.35	4.48	1.09	0.09	-0.86	-0.86	HOH4	0.44
278	0.58	1.39	1.37	0.23	-0.27	-0.27	HOH29	3.21
287	0.64	-9.7	0.98	0.08	0	0	HOH29	0.55
292	0.52	6.25	1.09	0.12	-3.54	-3.54	HOH22	0.26
294	0.7	6.67	0.98	0.08	-0.08	-0.08	HOH22	2.56
295	0.42	-11.26	0.87	0.04	-0.1	-0.1	HOH22	1.87
357	0.53	5.75	1.53	0.13	-0.25	-0.25	HOH5	0.74
379	0.39	3.13	0.85	0.03	-0.16	-0.16	HOH22	2.89
386	0.9	3.88	1.06	0	-0.01	-0.01	HOH29	2.82
403	0.8	4.27	1.06	0	-1.92	-1.92	HOH71	1.17
437	0.52	2.67	1.06	0	0	0	HOH29	2.8

Table 2. Predicted water sites. See text and MOE manual for details.

References

1. Liu, P., Zhu, F., Rassokhin, D.N., Agrafiotis, D.K.; A self-organizing algorithm for modeling protein loops. PLoS Computational Biology, 5(8), (**2009**) e1000478.
2. Coutsias, E. A., Seok, C., Jacobson, M. P., Dill, K. A.; A kinematic view of loop closure. Journal of Computational Chemistry, 25(4) (**2004**) 510-528.
3. Canutescu, A.A., Shelenkov, A.A., Dunbrack, R.L. Jr; A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci. Sep;12(9) (**2003**) 2001-14.
4. Boomsma, W., Hamelryck, T.; Full cyclic coordinate descent: solving the protein loop closure problem in Cα space. BMC Bioinformatics, 6(1) (**2005**) 159.
5. Rotkiewicz, P., Skolnick, J.; Fast procedure for reconstruction of full-atom protein models from reduced representations. Journal of Computational Chemistry, 29(9), (**2008**) 1460-1465
6. Gront, D., Kmiecik, S., Kolinski, A.; Backbone building from quadrilaterals: A fast and accurate algorithm for protein backbone reconstruction from alpha carbon coordinates. Journal of Computational Chemistry, 28(9) (**2007**) 1593-1597.
7. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J., Honig, B., Shaw, D. E., Friesner, R. A.; A hierarchical approach to all-atom protein loop prediction. Proteins: Structure, Function, and Bioinformatics, 55(2), (**2004**) 351-367.
8. Labute, P.; Protonate3D: Assignment of Ionization States and Hydrogen Coordinates to Macromolecular Structures; Proteins 75 (**2008**) 187–205.
9. Mogul - A knowledge-based library of molecular geometry derived from the Cambridge Structural Database (**2018**) https://www.ccdc.cam.ac.uk
10. Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT (**2016**) www.gaussian.com
11. Rollmann, W. (**1853**) Zwei neue stereoskopische Methoden, Annalen der Physik (in German), 90: 186–187,
12. Beglov, D. and Roux, B.; An Integral Equation to Describe the Solvation of Polar Molecules in Liquid Water; J. Phys. Chem. B 101 (**1997**) 7821–7826.
13. Kovalenko, A. and Hirata, F.; Self-Consistent Description of a Metal-Water Interface by the Kohn-Sham Density Functional Theory and the Three Dimensional Reference Interaction Site Model; J. Chem. Phys. 110 (**1999**) 10095–10112.
14. Kovalenko, A. and Hirata, F.; Potential of Mean Force of Simple Ions in Ambient Aqueous Solution. I. Three-Dimensional Reference Interaction Site Model Approach; J. Chem. Phys. 112 (**2000**) 10391–10417.