Structural Analysis and Virtual Screening for Membrane Fusion Inhibitors targeting Group 2 Influenza A Hemagglutinin

Initial Analysis of Crystal Structure

The co-crystal of HA X-31 and a promising small molecule inhibitor, tert-butyl hydroquinone (TBHQ)[1], was solved back in 2008 (PDBID: 3EYM)[8]. Visual inspection of the 3EYM's TBHQ binding pocket via Chimera[7] reaffirmed TBHQ's drug-like potential and the interactions noted in Russel et al. The structural differences between Group 1 and Group 2 HA were confirmed using the matchmaker function provided in Chimera. Two hemagglutinin subtypes from the two structurally distinct phylogenetic groups, 2009 H1 (PDBID: 3LZG) and H3 (PDBID: 3EYM), were selected for structural comparison. Using chains B and F of 3EYM as the reference structure with chains H and J of 3LZG, the two HA were overlaid with a RMSD value of 0.618 between 111 atom pairs. The orientation of key residue LYS 58 of HA2 in the RMSD fit structure coincides with past observations[8]. The intermonomer salt bridge between LYS 58 (HA2) and GLU 59 (HA2) occupies TBHQ binding mode in 3LZG.

The 2008 paper also mentions the importance of two key Hydrogen Bonds interacting with both the hydroxyl groups on TBHQ, although upon closer inspection the distance of the supposed hydrogen bond is 3.5 Angstroms long. This distance is too long under most circumstances to be considered a hydrogen bond. In addition, B-factors of two key interacting residues ranged from 80-90, suggesting that the given PDB structure represents one of many possible rotomers. TBHQ's were RMSD fit to each other using Chimera's match command with a value of <0.1 Angstroms which indicates that the conformations of the three TBHQ are similar.

Docking Experiments

With the crystal structures from the Russel et al. paper analyzed, Autodock Vina[6] and Autodock 4 (AD4)[4] were to be used consecutively to screen a large library of potential drug candidates. Control Dockings conducted first to validate the use of Vina and AD4 in a virtual

screening. Producing acceptable poses with both programs proved to be more difficult than anticipated, especially for the 3EYM structure. The following conditions were modified in an attempt to optimize the results: docking parameters (.dpf), TBHQ hydroxyl group orientation, grid box size, and protonation states[2]. Even with all of these changes, none of AD4 control experiments were able to correctly predict the TBHQ crystal pose. The majority of the clusters were situated on the opposite side of this large binding pocket (greater than 6 angstroms RMS). Vina control experiments were moderately better than the AD4 experiments. Positive Control experiments with H14 and H3 were able to calculate the correct orientation and hydrogen bonding, although the H3 positive control poses were noticeably incorrect in orientation.

It must be noted, however, that results for all the negative controls were poor. Predicted Negative Control binding energies for both H3 and H14 were lower than those from the positive controls. In addition, docked conformations were for Amantadine and Rimantadine matched well with TBHQ binding mode and mimicked several key TBHQ interactions. Docked conformations from the control experiments are shown below.

Fig 1: H14 (3EYK) Best Poses from Vina Controls

Fig 2: H3 (3EYM) Best Poses from Vina Controls

An earlier paper determined that escape mutants from Amantadine selection acquire a mutation at K58I (HA2)[9]. This particular residue is one of the key residues in the TBHQ pocket which suggests the existence of a secondary binding site for Amantadine. This mutation was modeled using one of Chimera's Rotomer Libraries[3]. The ILE rotomer with chi angles 121.641, -169.142, and 7.514 was selected due to its high calculated probability. Subsequent Vina control docking with this rotomer, however, did not significantly affect the docked Rimantadine poses.

For the preliminary screening, Vina was used to screen 3EYM against the following prepared ligand databases: NCDIS II (1485 ligands), FDA approved molecules (3176 ligands), and two drugbanks nutracuetics (79 ligands) and small molecules (1541 ligands). Visualization of the top 10 by binding energy revealed that the majority of the top compounds were large hydrophobic

molecules with some hydrophilic functional groups and several rotatable bonds.

Alternative Methods to produce the TBHQ pose

Due to the difficulty of reproducing the crystal pose within AD4, Autodock2MMGBSA (A2M) and AD4 with Restricted Electrostatic Potential (RESP) charges were explored. Both of which have been proven to produce robust results. RESP charges were assigned to TBHQ using Gaussian 03, and receptor charges were prepared using parameter set 99. AD4 parameters were also modified using parameters derived from two training sets. The resulting conformations from the 3EYM positive control docking were similar to the original AD4 redocking experiments. The two major clusters had an RMS value around 6 angstroms.

The A2M webservice on the magnetite cluster was also used to rescore some of the top hits. Top Vina poses from the Amantadine and Rimantadine control experiments and the TBHQ crystal pose were rescored. 3EYM was prepared with the PDBPQR Service[2]. The webservice was not able to rescore Amantadine, but resulting energies from the Rimantadine and TBHQ run were similar to the earlier control experiments with AD 4 and Vina. A2M predicted that Rimantadine's binding energies was lower than TBHQ's. Since this webservice is limited to one protein and one ligand at a time and has problems with Amantadine, use of this rescoring function was postponed.

In addition to the alternative scoring functions, minimizations were also performed to refine the 3EYM structure and produce better control results. With residues more than 5 angstroms of TBHQ binding mode restricted, one-thousand minimization steps were performed for 3EYM were performed at 300K using an iBelly minimization. The resulting minimized structure underwent dockings similar to earlier control experiments. Results for these control dockings were much better, although the predicted binding energies for the negative controls were still quite low. This minimized structure was used in a Vina screening with the same 4 ligand databases detailed above. Some of the top molecules from this screening matched the original Vina screening; however, there were still significant differences in energy ranking. Large hydrophobic molecules with some

hydrophobic character still dominated the top 10 across all these databases.

A 200 ps simulation of the 3EYM complex at pH 8.5 was then conducted to evaluate the TBHQ binding mode. The simulation utilized the implicit solvent model, Generalized Born surface, and AMBER's sander module. Again one-thousand normal minimization steps were performed. Another short simulation was performed, limiting movement only to hydrogen. After equilibration, all restraining masks were removed, and then the system was allowed to run for 200ps at 300K. The resulting simulation shows the movement of TBHQ away from the original binding mode, and suggests a problem with the crystal binding mode.

Future Directions

The poor results from the control docking suggest that the TBHQ binding required some type of induced fit. Two additional 200ps MD simulations will be needed using different velocity assignments in order to definitively draw conclusions. These future simulations will have to be analyzed for key hydrogen bond distances and overall energy change of the complex.

It is also necessary to find an alternate scoring function to rescore the top hits from the Vina Screening, as AD4 is currently unable to correctly redock TBHQ. X-Score or DrugScore are good potential candidates for this as they have been known to produce fairly robust results. In addition, A2M can also be used to rescore the top hits, though the calculation time may not be suited for such a large ligand set.

With the excellent Vina control experiments for 3EYK (H14), it also maybe worthwhile to screen ligand databases against this receptor. These results can be compared with 3EYM Vina screening to find compounds that can target multiple Hemagglutinin subtypes. Although again, an alternate scoring function will have to be used to rescore those results as well.

Bibliography

1. Bodian, D.L. et al. Inhibition of the fusion-inducing conformational change of influenza hemagglutinin by benzoquinones and hydroquinones. Biochemistry 32, 2967-2978 (1993).

2. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 32 W665-W667 (2004).

3. Dunbrack, R.L. Rotamer libraries in the 21st century. Curr. Opin. Struct. Biol 12, 431-440 (2002).

4. Sanner M.F., Python: A Programming Language for Software Integration and Development. J. Mol. Graphics Mod., 1999, Vol 17, February. pp 57-61.

5. Sanner M.F., Stoffler D. and Olson A.J. (2002). ViPEr, a visual Programming Environment for Python. In Proceedings of the 10th International Python conference. 103-115. February 4-7, 2002. ISBN 1-930792-05-0.

6. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry 31 (2010) 455-461*

7. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605-1612 (2004).

8. Russell, R.J. et al. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. Proc Natl Acad Sci U S A 105, 17736-17741 (2008).

9. Steinhauer, D.A., Wharton, S.A., Skehel, J.J., Wiley, D.C. & Hay, A.J. Amantadine selection of a mutant influenza virus containing an acid-stable hemagglutinin glycoprotein: evidence for virus-specific regulation of the pH of glycoprotein transport vesicles. Proc. Natl. Acad. Sci. U.S.A 88, 11525-11529 (1991).

Acknowledgements

Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

This research would not have been possible without support from the following people and institutions.

National Biomedical Computation Resource (NBCR): Dr. Wilfred Li and Jane Ren

National Taiwan University Pharmacoinformatics Lab: Dr. Jung Hsin Lin, Yu-Hsuan Chen, Jui-Chih Wang, Jing Yeh Chen, and Che-Chia

UCSD PRIME: Dr. Gabriele Wienhausen, Dr. Peter Arzberger, Teri Simas, Jim Galvin, Tricia Taylor Oliveria, and Jennifer Choy

Funding from the National Science Foundation (NSF) - IOSE-0710726
California Institute for Telecommunications and Information Technology (Calit2)
Pacific Rim Applications and Grid Middleware Assembly (PRAGMA)