Surflex-QMOD: Protein Pocket Modeling for Affinity Prediction

Alexander Steudle[1], Rocco Varela[2], Ajay N. Jain[3]

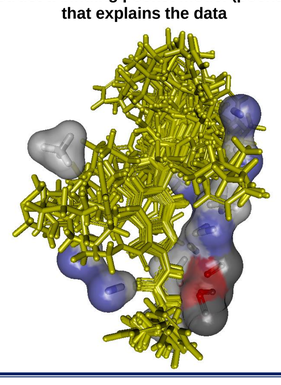
[1]Certara International, Martin-Kollar-Straße 17, 81829 München
[2]Certara™, 1699 S Hanley Rd, St. Louis, MO, 63144
[3]Dept. of Bioengineering and Therapeutic Sciences, University of California, San Francisco

INTRODUCTION

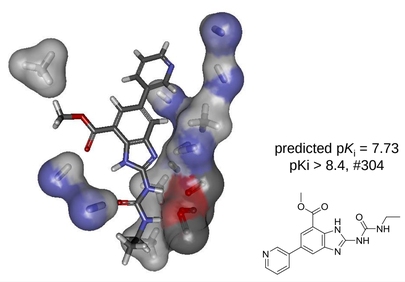
A common challenge faced during molecule design is the ability to make accurate predictions of a molecule's **effect** or **potency** when the structure of the therapeutic target is unavailable. The detailed structure of a protein binding pocket can be critical in rationalizing and guiding design efforts of small molecule compounds. Without a reliable protein structure, one must rely on information provided by known ligands to make inferences about the interaction between the molecule and target in question. What is needed is a physically sensible model for quantitative predictions of binding affinities that must: (1) be sensitive to conformations and alignments of ligands, (2) model non-additive effects (e.g. where two substituents make a molecule too big), (3) show a direct relationship to experimentally determined binding sites.

To address this challenge we have developed a ligand-based method (Surflex-QMOD: **q**uantitative **mod**eling) for modeling protein binding pockets and the quantitative prediction of small molecule binding activity. The QMOD method uses structure and activity data of known competitive ligands to construct a physically realistic model that captures the detailed shape and electrostatic characteristics of the protein binding site. Newly designed molecules can then be flexibly fit to the pocket to make predictions of effect or potency.

Construct a binding pocket model (pocketmol) that explains the data

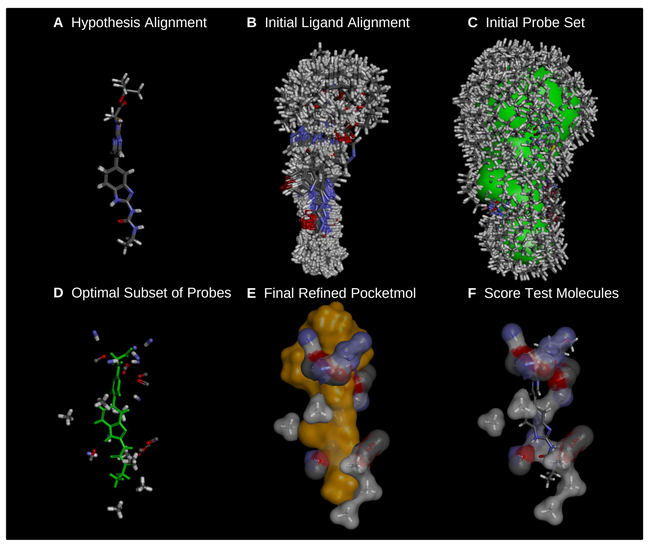


So that we can predict the activity and geometry of new ligands: predicted pKi = 7.73



SURFLEX-QMOD METHOD

QMOD uses multiple-instance learning to build a pocket model (**pocketmol**). The procedure builds an initial alignment (A), generates multiple poses for each ligand (B), tessellates the poses with pocket probes (C), selects an initial optimal probe set that fits the binding data (D), and iteratively refines the positions of the poses and probes until the calculated interactions converge to the binding data (E). New molecules are then be docked to the pocketmol to make predictions of binding potency and geometry (F).

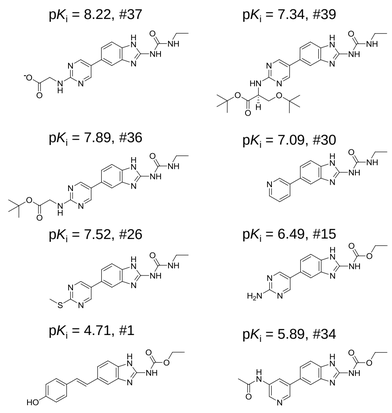


SIMULATED MEDICINAL CHEMISTRY PROJECT

Data

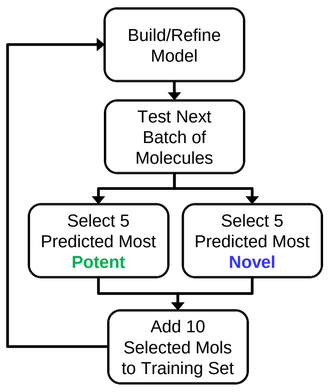
426 gyrase inhibitors from Vertex Pharm. Inc. (Pat Walters).Consists of dates of synthesis and binding data (pK_i). Activity range 4.7-8.4 pK_i

Input: Gyrase Ligand Structure-Activity Data



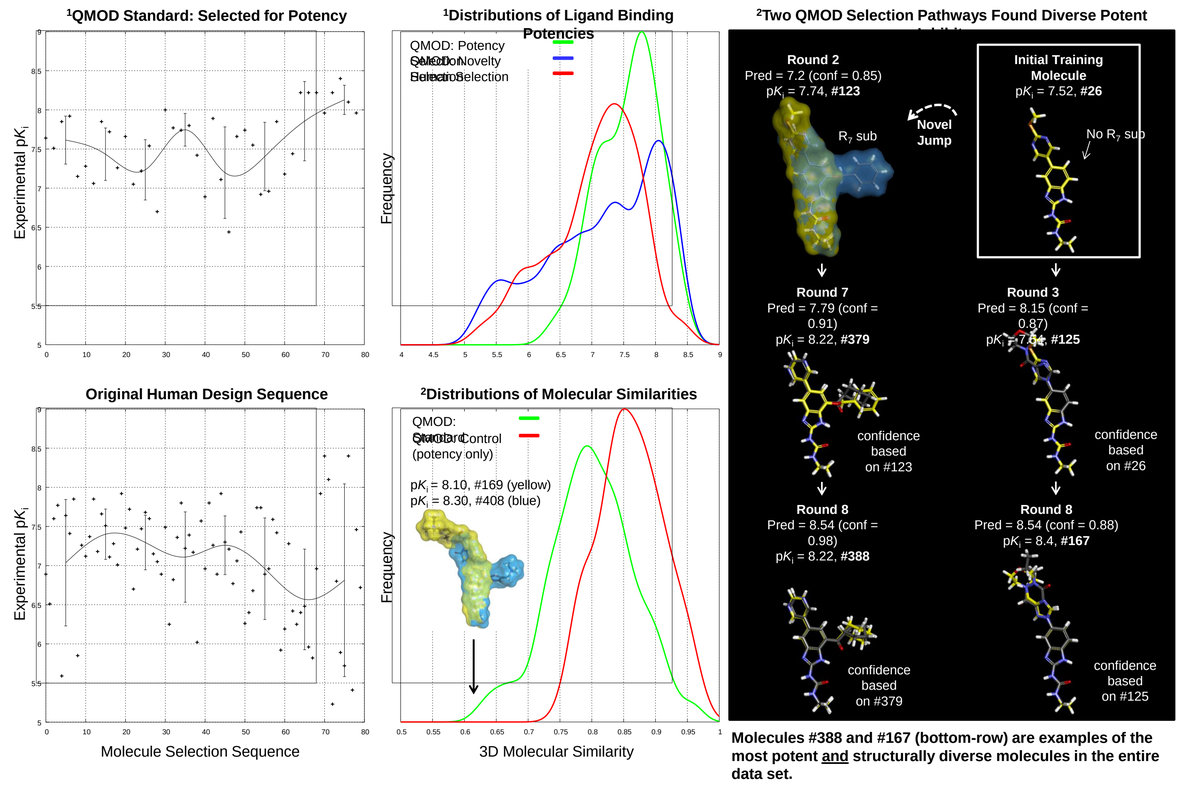
Strategy

- Train temporally
- Begin with the first 39 molecules
- Add 10 chosen (5 pred. potent + 5 pred. novel)
- Refine
- Repeat until all mols have been tested (8 rounds)



RESULTS

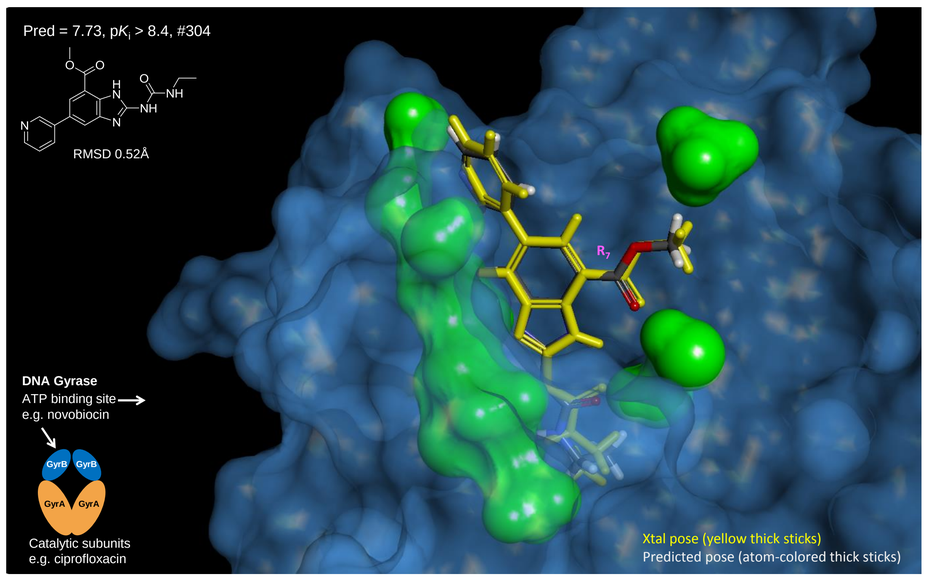
QMOD confidently converges on potent inhibitors[1] and identifies structurally diverse potent compounds[2]



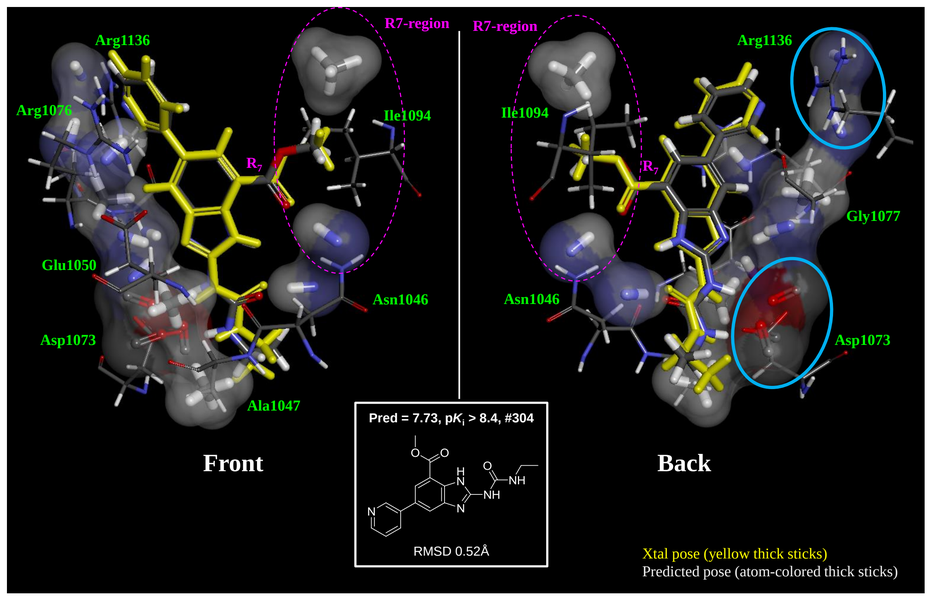
Molecules #388 and #167 (bottom-row) are examples of the most potent and structurally diverse molecules in the entire data set.

The final model actually looks like the crystallographically determined gyrase pocket.

The final pocketmol is physically representative and highly concordant with the gyrase pocket and makes accurate predictions of ligand activity. The predicted pose of #304 (atom-colored sticks) deviates from the crystal structure pose (yellow sticks) by only 0.52Å RMSD. Molecule #304 is an example of one of the most potent inhibitors in the entire data set that revealed the most structural divergence from any of the initial training ligands (examples in Data section). Probes of the final pocketmol providing measurable ligand interaction scores (green skin) capture the funnel-like concavity of the gyrase binding pocket while providing a physical representation of the R_7 cavity on the right-side of the pocket.



Pocketmol probes (atom-colored, skinned) are in good spatial agreement with protein residues responsible for ligand-binding (atom-colored thin sticks).. The pattern of contact between compound #304 and the pocketmol compared with the contacts made with the protein matched for 30/42 total atoms ($p < 0.01$ by permutation analysis).



CONCLUSIONS

Surflex-QMOD makes predictions using a physical model that closely approximates reality, so the predictions can be applied more accurately. With explicit guidance based on potency and novelty we can identify diverse potent molecules.

Acknowledgements:
Pat Walters and Brian Goldman for providing the gyrase data set

Refercences

- Varela, R.; Cleves, A. C.; Spitzer, R.; Jain, A. N. A Structure-Guided Approach for Protein Pocket Modeling and Affinity Prediction. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 917–934.
- Brown, S. P.; Muchmore, S. W. Large-scale application of high-throughput molecular mechanics with Poisson–Boltzmann surface area for routine physics-based scoring of protein–ligand complexes. *J. Med. Chem.* **2009**, *52*, 3159–3165.
- Varela, R., Walters, W. P., Goldman, B. B., Jain, A. N. Iterative refinement of a binding pocket model: active computational steering of lead optimization. *J. Med. Chem.* **2012**, *55*, 8926–8942.
- Charifson , P.S. et al. (2008) Novel Dual-Targeting Benzimidazole Urea Inhibitors of DNA Gyrase and Topoisomerase IV Possessing Potent Antibacterial Activity: Intelligent Design and Evolution through the Judicious Use of Structure-Guided Design and Structure-Activity Relationships. J Med Chem, 51, 5243-5263
- Langham, J. J.; Cleves, A. E.; Spitzer, R.; Kirshner, D.; Jain, A. N. Physical binding pocket induction for affinity prediction. *J. Med. Chem.* **2009**, *52*, 6107–6125.