The Application of COMBINE Analysis to Generate Target-Specific Scoring Functions

Stefan Henrich[1], Isabella Feierberg[2], Ting Wang[1,3], Niklas Blomberg[2], Rebecca Wade[1]

[1] EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany
[2] AstraZeneca R&D Mölndal, S-43183 Mölndal, Sweden
[3] present address: UC Davis Genome Center, 431 East Health Science Drive, University of California, Davis, CA 95616-8816, USA

Introduction

The family of trypsin-like serine proteases are important targets in drug design. Due to high structural similarity of its family members, it is difficult to find or design target-specific inhibitors. In this study we applied COMparative BINding Energy (COMBINE) analysis, a receptor-based 3D QSAR method, to a set of crystal structures of thrombin and the anti-target trypsin in complex with small inhibitors. In this method, experimental inhibitor constants are correlated with interaction energy terms derived from receptor-ligand-structures to describe the binding affinity. We compared the two independent COMBINE analyses of trypsin and thrombin and used the results to predict the binding affinity as well as the selectivity of ligands.

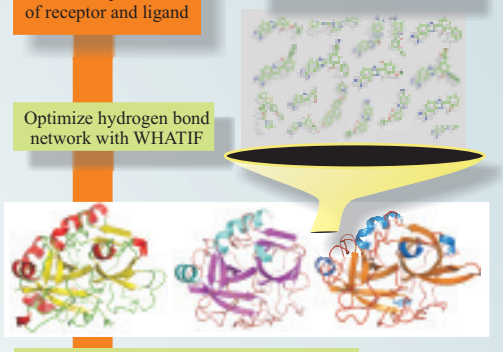
Method

The principal idea of COMBINE analysis is the assumption that the binding free energy ΔG is correlated with a subset of weighted interaction energy terms determined by receptor-ligand complexes. The energy minimized model structures are divided for energy calculations into parts according to their spatial location, normally its amino acid residues and the bound ligand. For each complex electrostatic and Lennard-Jones interaction energy terms as well as solvation energy terms between ligand and receptor are computed. These terms are analyzed by Principal Component Analysis (PCA) and are correlated to activity values by Partial Least Squares (PLS) coupled with variable selection and data pretreatment. This correlation results in a target-specific scoring function which can be used to highlight important residues for ligand binding and to predict the binding affinity of new, docked ligands (figure 1).

Crystal, NMR or modeled structures of complexes of receptor and ligand

Flexible docking of ligands into target receptor with GOLD 3.0

Optimize hydrogen bond network with WHATIF



Molecular mechanics energy minimization of complexes in AMBER 8

Experimental binding data of protein-ligand-interactions K_i or IC_{50} values converted to ΔG

1

Principle Component Analysis (PCA) in GOLPE 4.6

Correlate different components of binding energy ΔU with ΔG values by Partial Least Squares (PLS) in GOLPE 4.6 for generating target-specific scoring functions. Binding free energy can be modelled as

$$\Delta G = \sum_{i=1}^{n}\omega_i^{vdw}\Delta u_i^{vdw} + \sum_{i=1}^{n}\omega_i^{ele}\Delta u_i^{ele} + \sum_{i=1}^{n}\omega_i^{solv}\Delta u_i^{solv} + C$$

Calculate ligand-receptor binding energy ΔU for each complex with ANAL module of AMBER 8

Partition ΔU of the receptor and the ligand into several components on basis of location in the complex and physicochemical properties

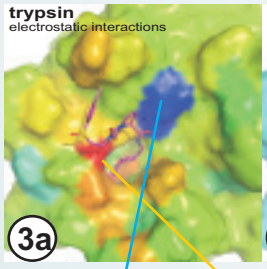
Δu_i^{vdw} ; Δu_i^{ele}

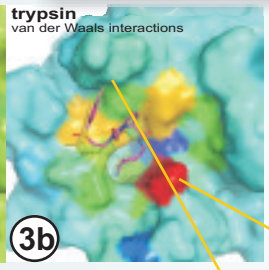
Calculate solvation energy term with UHBD for ligands, receptor, and complexes

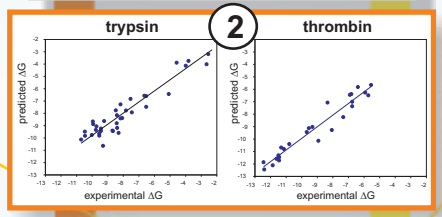
Δu_i^{solv}

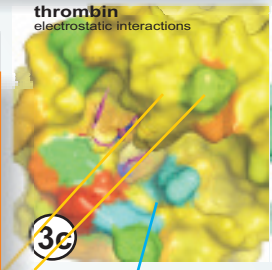
ω_i^{vdw} ; ω_i^{ele} ; ω_i^{solv}

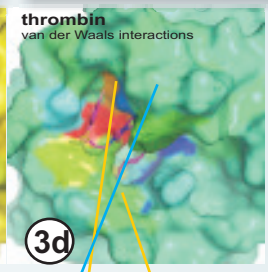
$$\sum_{i=1}^{n}\omega_i^{vdw}\Delta u_i^{vdw} + \sum_{i=1}^{n}\omega_i^{ele}\Delta u_i^{ele} + \sum_{i=1}^{n}\omega_i^{solv}\Delta u_i^{solv} + C$$

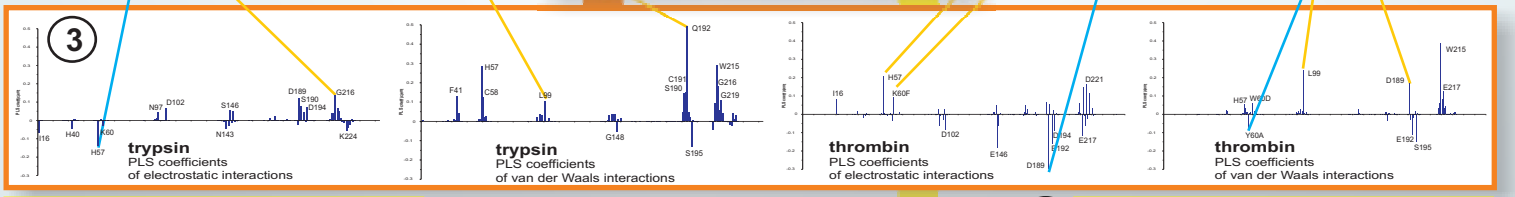


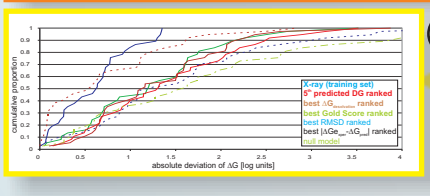




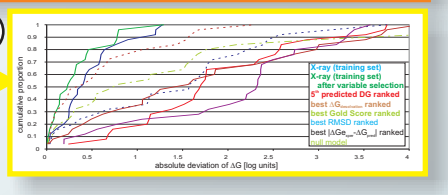








Predicted binding affinity ΔG for docked compounds



Model building

The COMBINE models for trypsin and thrombin based on 37 and 25 X-ray structures of the PDB and their published inhibitor constants. For both targets representative structures were selected and were minimized by molecular mechanics calculation in complex with the experimental determined conformations of the ligands. A correlation of calculated interaction and desolvation energy terms with published binding free energy values ΔG resulted in R^2 and Q^2 values for predicted versus experimental ΔG values of 0.90 and 0.82 for trypsin (at latent variable 3) and 0.93 and 0.81 for thrombin (at latent variable 4), respectively (figure 2).

In figure 3 the real PLS coefficients of the electrostatic (A, C) and van der Waals (B, D) interaction energy terms of the COMBINE models of trypsin and thrombin were plotted and were used for colouring the residues of the active-site clefts. The colours illustrate important parts for ligand binding (red: favoured parts, positive PLS coefficients; blue: unfavoured parts, negative PLS coefficents).

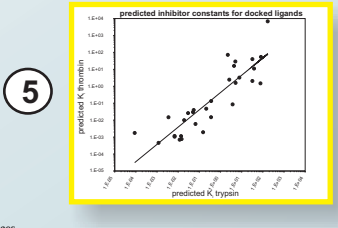
Docking

Ligands, which were already used as training set for COMBINE model building ('pseudo test set'), were docked ten times to the selected target structures of trypsin and thrombin using the program GOLD 3.0. The interaction and desolvation energy terms were calculated for all docking solutions and the binding free energy were predicted by the corresponding COMBINE models. The ten predicted ΔG values for each of the ligands were re-ranked according to RMSD, predicted ΔG values, the absolute difference between experimental and predicted ΔG values, GOLD Score Fitness, desolvation energy. The best predictions for the binding affinity could be yield by selecting the predicted ΔG values of the top desolvation energy ranked or 5th predicted ΔG ranked docking solution. Within an accuracy of 2 log units for more than 70 % of the ligands a correct binding free energy could be predicted (figure 4).

http://www.eml-research.de/english/research/mcm/index6.php
http://projects.villa-bosch.de/mcm

Selectivity

Ligands with known inhibitor constants Ki for trypsin and thrombin were docked ten times to the protein structures. The predicted ΔG values (5th predicted ΔG ranked docking solution) with an absolute error of less than 3 log units were computed back to Ki and were plotted on a logarithmic scale against each other. The R2 value of 0.77 shows a good selectivity in predicting the binding affinity of both proteases.



References

1. Wade, R.C., Henrich, S., Wang, T. (2004) Using 3D protein structures to derive 3D-QSARs. *Drug Discovery Today: Technologies* 1:241-246.
2. Ortiz, A.R., Pisabarro M.T., Gago F., Wade R.C. (1995) Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* 38:2681-2691.
3. Wade, R.C., Ortiz, A.R., Gago F. (1998) Comparative Binding Energy Analysis. *Persp. Drug Disc. and Des.* 9:19-34.

We gratefully acknowledge support of AstraZeneca, Sweden, and of the Klaus Tschira Foundation.