Docking small ligands to proteins is a way to find potential drugs \(\Leftrightarrow\) Industrially important

Rapid docking algorithm is required to dock large sets of small molecules into a protein

Scoring function needs to
- 1: distinguish correct binding pose from incorrect binding pose
- 2: rank ligands by their affinity for target

A wide variety of scoring functions exist. Most perform well on (1) and poorly on (2)

A small region of interest (pharmacophore) can be identified, reducing computation
Assessing Scoring Functions for Protein-Ligand Interactions

An assessment of nine scoring functions commonly applied in docking using a set of 189 protein–ligand complexes is presented. The scoring functions include the CHARMM potential, the scoring function DrugScore, the scoring function used in AutoDock, the three scoring functions implemented in DOCK, as well as three scoring functions implemented in the CScore module in SYBYL (PMF, Gold, ChemScore). We evaluated the abilities of these scoring functions to recognize near-native configurations among a set of decoys and to rank binding affinities. Binding site decoys were generated by molecular dynamics with restraints. To investigate whether the scoring functions can also be applied for binding site detection, decoys on the protein surface were generated. The influence of the assignment of protonation states was probed by either assigning “standard” protonation states to binding site residues or adjusting protonation states according to experimental evidence. The role of solvation models in conjunction with CHARMM was explored in detail. These include a distance-dependent dielectric function, a generalized Born model, and the Poisson equation. We evaluated the effect of using a rigid receptor on the outcome of docking by generating all-pairs decoys (“cross-decoys”) for six trypsin and seven HIV-1 protease complexes. The scoring functions perform well to discriminate near-native from misfolded conformations, with CHARMM, DOCK-energy, DrugScore, ChemScore, and AutoDock yielding recognition rates of around 80%. Significant degradation in performance is observed in going from decoy to cross-decoy recognition for CHARMM in the case of HIV-1 protease, whereas DrugScore and ChemScore, as well as CHARMM in the case of trypsin, show only small deterioration. In contrast, the prediction of binding affinities remains problematic for all of the scoring functions. ChemScore gives the highest correlation value with $R^2 = 0.51$ for the set of 189 complexes and $R^2 = 0.43$ for the set of 116 complexes that does not contain any of the complexes used to calibrate this scoring function. Neither a more accurate treatment of solvation nor a more sophisticated charge model for zinc improves the quality of the results. Improved modeling of the protonation states, however, leads to a better prediction of binding affinities in the case of the generalized Born and the Poisson continuum models used in conjunction with the CHARMM force field.
Ranking of Different Binding Poses for one Ligand

Here, we present an assessment of nine scoring functions, most of which are implemented in widely used docking programs. The scoring functions cover the three classes described above: CHARMM$^{22}$ and DOCK-chemical$^5$ represent force-field-based methods; ChemScore$^{23}$ and the potentials implemented in GOLD$^{3,24}$ and AutoDock$^{6,25,26}$ are empirical scoring functions; DrugScore$^{20}$ and PMF$^{19}$ are knowledge-based potentials. Finally, DOCK-contact counts the number of contacts between the ligand and the receptor. The study was performed on data from the Ligand–Protein Database (LPDB), which is World Wide Web accessible (http://lpdb.scripps.edu) and comprises 189 protein–ligand complexes.$^{27}$ This data set corresponds to 49 different receptors with both high-resolution structure (2.1 Å on average) and known experimental binding affinity. In this respect, the current study is the most comprehensive comparison of scoring functions reported so far.

Figure 3. Percentage of complexes for which the lowest energy decoy is within 2 Å from the crystal structure. The scoring functions are represented on the x-axis (CHARM, CHARMM-RDIE ($\epsilon(i) = 4\tau$); DNrg, DOCK-Energy; DChm, DOCK-chemical; DCnt, DOCK-contact; DrugS, DrugScore; ChemS, ChemScore; AutoD, AutoDock), and the various data sets are represented on the y-axis (All, whole set (189 complexes); All/NoTr., All without the complexes used to calibrate ChemScore and AutoDock; AspPr., aspartic protease; Oxido., oxidoreductase; SerPr., serine protease; MetPr., metalloprotease; Immu., immunoglobulin; Arab., L-arabinose binding protein; Mhc, major histocompatibility protein).
Discriminative Power in Ranking of Different Binding Poses

Here, we present an assessment of nine scoring functions, most of which are implemented in widely used docking programs. The scoring functions cover the three classes described above: CHARMM\textsuperscript{22} and DOCK-chemical\textsuperscript{5} represent force-field-based methods; ChemScore\textsuperscript{23} and the potentials implemented in GOLD\textsuperscript{3,24} and AutoDock\textsuperscript{6,25,26} are empirical scoring functions; DrugScore\textsuperscript{20} and PMF\textsuperscript{19} are knowledge-based potentials. Finally, DOCK-contact counts the number of contacts between the ligand and the receptor. The study was performed on data from the Ligand–Protein Database (LPDB), which is World Wide Web accessible (http://lpdb.scripps.edu) and comprises 189 protein–ligand complexes.\textsuperscript{27} This data set corresponds to 49 different receptors with both high-resolution structure (2.1 Å on average) and known experimental binding affinity. In this respect, the current study is the most comprehensive comparison of scoring functions reported so far.

\textbf{Figure 4.} Discriminative power. A discriminative power value of zero means no discriminative power, and the lower the value, the more discriminative is the scoring function. See Figure 3 for the definition of the scoring functions and the data sets.
3.3. Prediction of Binding Affinities. In this section, we analyze the ability of the scoring functions to rank binding energies. Figure 7 shows the square of the correlation coefficients ($R^2$ value) for the different data sets using the minimized crystal structure. ChemScore achieves the highest correlation with an $R^2$ value of 0.51 for set 1a. ChemScore also yields the highest $R^2$ value ($R^2 = 0.43$) for set 2, where the complexes used to calibrate this scoring function and AutoDock were removed. In particular, ChemScore outperforms all other scoring functions except AutoDock for the immunoglobulin set, although no immunoglobulin was included in the training set. It also gives higher $R^2$ values than most of the other scoring functions for the data sets for oxidoreductase, lyase, and others. The training set of ChemScore contains 6 oxidoreductase, 1 lyase, and 11 other complexes that belong also to the LPDB. Furthermore, ChemScore ranks binding affinities significantly better than the other regression-based scoring function AutoDock. Although the results always depend on the set of complexes, this suggests that ChemScore was more broadly parametrized; for instance, a larger di-

Figure 7. Square of the correlation coefficients ($R^2$) between experimental and calculated binding energies. Values were set to zero in the case of an anticorrelation. See Figure 3 for the definition of scoring functions and data sets.
Predicting Binding Energies for 229 Protein-Ligand Complexes (R=0.63)

- aspartic protease (R=0.41)
- hydrolase (R=0.2)
- immunoglobulin (R=0.5)
- L-arabinose bp (R=0.69)
- mhc (R=0.14)
- oxidoreductase (R=0.33)
- serine protease (R=0.67)
- transferase (R=0.65)
- other (R=0.56)
A critical barrier to entry into structure-based virtual screening is the lack of a suitable, easy to access database of purchasable compounds. We have therefore prepared a library of 727,842 molecules, each with 3D structure, using catalogs of compounds from vendors (the size of this library continues to grow). The molecules have been assigned biologically relevant protonation states and are annotated with properties such as molecular weight, calculated LogP, and number of rotatable bonds. Each molecule in the library contains vendor and purchasing information and is ready for docking using a number of popular docking programs. Within certain limits, the molecules are prepared in multiple protonation states and multiple tautomeric forms. In one format, multiple conformations are available for the molecules. This database is available for free download (http://zinc.docking.org) in several common file formats including SMILES, mol2, 3D SDF, and DOCK flexibase format. A Web-based query tool incorporating a molecular drawing interface enables the database to be searched and browsed and subsets to be created. Users can process their own molecules by uploading them to a server. Our hope is that this database will bring virtual screening libraries to a wide community of structural biologists and medicinal chemists.
An Analysis of ZINC’s compliance with Lipinski Rule of Five

Lipinski’s rule states that, in general, an orally active drug has no more than one violation of the following criteria:

- Not more than 5 hydrogen bond donors (nitrogen or oxygen atoms with one or more hydrogen atoms)
- Not more than 10 hydrogen bond acceptors (nitrogen or oxygen atoms)
- A molecular mass not greater than 500 daltons
- An octanol-water partition coefficient log P not greater than 5

Note that all numbers are multiples of five, which is the origin of the rule’s name. As with many other rules of thumb, there are many exceptions to Lipinski’s Rule. Further, many variations of that rule exist including restrictions on polar surface area and flexibility. – C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney; "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings"; Adv Drug Deliv Rev; 2001; Vol. 46 (1-3): p. 3-26.)
Inhibitors of HIV-1 Protease – A Major Success for Drug Design

- Combination of
  - Crystallography
  - NMR Spectroscopy
  - Computation
  - Chemical Synthesis

- 10 FDA-approved HIV-1 Protease Inhibitors are on the Market

ABSTRACT By using a structure-based computer-assisted search, we have found a butyrophenone derivative that is a selective inhibitor of the human immunodeficiency virus 1 (HIV-1) protease. The computer program creates a negative image of the active site cavity using the crystal structure of the HIV-1 protease. This image was compared for steric complementarity with 10,000 molecules of the Cambridge Crystallographic Database. One of the most interesting candidates identified was bromperidol. Haloperidol, a closely related compound and known antipsychotic agent, was chosen for testing. Haloperidol inhibits the HIV-1 and HIV-2 proteases in a concentration-dependent fashion with a $K_i$ of $\approx 100 \, \mu M$. It is highly selective, having little inhibitory effect on pepsin activity and no effect on renin at concentrations as high as 5 mM. The hydroxy derivative of haloperidol has a similar effect on HIV-1 protease but a lower potency against the HIV-2 enzyme. Both haloperidol and its hydroxy derivative showed activity against maturation of viral polypeptides in a cell assay system. Although this discovery holds promise for the generation of nonpeptide protease inhibitors, we caution that the serum concentrations of haloperidol in normal use as an antipsychotic agent are $< 10 \, \text{ng/ml} (0.03 \, \mu M)$. Thus, concentrations required to inhibit the HIV-1 protease are $> 1000$ times higher than the concentrations normally used. Haloperidol is highly toxic at elevated doses and can be life-threatening. Haloperidol is not useful as a treatment for AIDS but may be a useful lead compound for the development of an antiviral pharmaceutical.
**ΔΔG Prediction for HIV1-Protease Inhibitor Complexes at R=0.826**

\[ ΔΔG_{Prediction} \text{ (kcal)} \]

\[ ΔΔG_{Experiment} \text{ (kcal)} \]

- AcetylPepstatin
- Indinavir
- KNI-272
- KNI-764
- Lopinavir
- Nelfinavir
- Ritonavir
- Saquinavir
- AG1776
- Peptidomimetic

6 December 2011 © Jens Meiler
**ΔΔG Prediction for HIV1-Protease Inhibitor Complexes at SD=1.14kcal**

- 0 - 8 Mutations: $sd=1.08 \text{kcal}$
- 9 - 16 Mutations: $sd=1.19 \text{kcal}$
- 16 - $\infty$ Mutations: $sd=1.24 \text{kcal}$
Discovery of a Novel Binding Trench in HIV Integrase

Abstract: Docking of the 5CITEP inhibitor to snapshots of a 2 ns HIV-1 integrase MD trajectory indicated a previously uncharacterized trench adjacent to the active site that intermittently opens. Further docking studies of novel ligands with the potential to bind to both regions showed greater selective affinity when able to bind to the trench. Our ranking of ligands is open to experimental testing, and our approach suggests a new target for HIV-1 therapeutics.

Figure 1. The two predominant docking conformations of 5CITEP to an open MD snapshot of integrase. The ligand in green shows 5CITEP in the orientation similar to the crystal structure of the complex. The ligand in yellow shows 5CITEP in its “flipped” orientation. Residues lining both ligand positions are highlighted.

Figure 4. Compounds D (blue) and I (red) superimposed in the same open MD snapshot. Each ligand samples the active site and the trench for maximal binding energy.