DATE: 04/10/2022

WEBLEM 7: CombiGlide

Aim:

CombiGlide: A building molecular database along with special emphasis on retrieval using structure input.

Introduction:

The human genome project and high-throughput crystallography efforts have led to an increase in the number of drug-discovery initiatives that have a high-resolution crystal structure of the receptor available in recent years, and this trend is predicted to continue. The pharmaceutical and biotechnology industries make substantial use of a number of dockings programmes, the most popular of which appear to be GOLD, FlexX, and DOCK.

The Glide software suite from FirstDiscovery now uses a new docking mechanism. Glide has been developed to screen enormous libraries with enough computing speed while doing as close to an exhaustive search of the positional, orientational, and structural space as is practical. This has been achieved by applying a number of hierarchical filters, as will be explained below.:

i. Docking times average less than 1 min for data sets having 0-10 rotatable bonds on an AMD Athelon MP 1800+ processor running Linux.

ii. Robustness in binding mode prediction is qualitatively superior to what is reported in the current literature fordocking methods in widespread use. For example, a comparison with results obtained by the developers of GOLD yields an average rmsd of 1.46 A for Glide compared with 2.56 À for GOLD for the 72 noncovalentlybound cocrystallized ligands of the GOLD test set15 that have 10 or fewer rotatable bonds. The comparison to FlexX is even more favorable. Comparisons for ligands having up to 20 rotatable bonds yield similar results.

iii. Binding affinity predictions, compared with experimental data for cocrystallized complexes, are reasonable (2.3 kcal/mol rmsd), though clearly subject to improvement.

iv. Results for library screening, reported in the following paper, 16 are very encouraging. Furthermore, databaseenrichment factors obtained using Glide 2.5 are significantly higher than those obtained using previous versions of Glide.

OVERVIEW OF DOCKING METHODOLOGY:

Glide looks for potential ligand sites in the receptor's active-site area using a hierarchy of filters. Different sets of fields on a grid that enable progressively more precise scoring of the ligand posture are used to describe the structure and characteristics of the receptor. Since these fields are produced during the calculation's preprocessing steps, they only need to be computed once for each receptor.

The next step generates a set of initial ligand structures. These conformations are selected from a detailed enumeration of the minimums in the ligand torsional angle space and represented in a compact combinatorial fashion. Given these ligand conformations, initial scans are performed over the entire phase space available for the ligand to find promising ligand positions.

This prescreening drastically reduces the region of phase space over which computationally expensive energy and gradient evaluations will later be performed while at the same time avoiding the use of stochastic methods; such methods can miss key phase-space regions a certain fraction of the time, thus precluding development of a truly robust algorithm. Glide is unique in its reliance on the techniques of exhaustive systematic search, though approximations and truncations are required to achieve acceptable computational speed.

Starting from the poses selected by the initial screening, the ligand is minimized in the field of the receptor using a standard molecular mechanics energy function in conjunction with a distance-dependent dielectric model.

Finally, a Monte Carlo approach is used to look for close torsional minima on the three to six lowest-energy positions that were found in this way. Sometimes such procedure is necessary to correctly orient peripheral groups, and it can also change internal torsion angles. Even in the absence of a solvent, the conventional molecular mechanics energy function provides a reliable model for predicting binding modes. It is not insufficient for ranking distinct ligands, such as ligands with varying net charges, nevertheless. To forecast binding affinity and rank-order ligands in database screens, a modified and enhanced version of the ChemScore! scoring tool, GlideScore, is built. The ideal docked configuration can be chosen by combining the GlideScore, the ligand-receptor molecular mechanics interaction energy, and the ligand strain energy.

The scoring algorithm needs to be changed to account for the fact that the protein structure used for docking is typically not tailored to fit a specific ligand, especially the molecular mechanics component. This is a final and crucial issue. When docking a library of ligands into a single rigid receptor structure, the biggest issue is that some actives cannot fit because the protein cavity is too small. To avoid this, commonly selected (e.g., nonpolar) protein and/or ligand atoms have their van der Waals radii scaled down to add more room to the binding pocket. Studies indicate that this strategy works well in the situation. Although changing the scale parameters for a certain receptor can frequently result in better enrichment, the default values are sufficient for everyday use.

SCORING FUNCTION:

The starting point for Glide scoring is the empirically based ChemScore function of Eldridge et al., which can be written as:

$$\Delta G_{\text{bind}} = C_0 + C_{\text{lipo}}\sum f(\text{r}_{\text{lr}}) + C_{\text{hbond}}\sum g(\Delta\text{r})\, h(\Delta\alpha) + C_{\text{metal}}\sum f(r_{\text{lm}}) + C_{\text{rotb}}H_{\text{rotb}}$$

The summation in the second term extends over all ligand-atom/receptor-atom pairs defined by ChemScore as lipophilic, while that in the third term extends over all ligand-receptor hydrogen-bonding interactions. In eq 1, *f*, *g*, and *h* are functions that give a full score (1.00) for distances or angles that lie within nominal limits and a partial score (1.00-0.00) for distances or angles that lie outside those limits but inside larger threshold values. For example, $g(\Delta\text{r})$ is 1.00 if the H---X hydrogen bond distance is within 0.25 Å of a nominal value of 1.85 Å but tails off to zero in a linear fashion if the distance lies between 2.10 and 2.50 Å. Similarly, $h(\Delta\text{R})$ is 1.00 if the Z-H---X angle is within 30° of 180° and decreases to zero between 150° and 120°.

GlideScore 2.5 modifies and extends the ChemScore function as follows:

$$\Delta G_{\text{bind}} = C_{\text{lipo-lipo}}\sum f(\text{r}_{\text{lr}}) + C_{\text{hbond-neut-neut}}\sum_{\text{g}}(\Delta r)\, h(\Delta\alpha) + C_{\text{hbond-neut-charged}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{hbond-charged-charged}}\sum g(\Delta r)\, h(\Delta\alpha) + C_{\text{max-metal-ion}}\sum f(r_{\text{lm}}) + C_{\text{rotb}}H_{\text{rotb}} + C_{\text{polar-phob}}V_{\text{polar-phob}} + C_{\text{coul}}E_{\text{coul}} + C_{\text{vdW}}EvdW + \text{solvation terms}$$

The lipophilic-lipophilic term is defined as in ChemScore. The hydrogen-bonding term also uses the ChemScore form but is separated into differently weighted components that depend on whether the donor and acceptor are both neutral, one is neutral and the other is charged, or both are charged. In the optimized scoring function, the first of these contributions is found to be the most stabilizing and the last, the charged-charged term, is the least important. The metal-ligand interaction term uses the same functional form as is employed in ChemScore but varies in three principal ways.

First, this term considers only interactions with anionic acceptor atoms (such as either of the two oxygens of a carboxylate group). This modification allows Glide to recognize the evident strong preference for coordination of anionic ligand functionality to metal centers in metalloproteases. In addition, Glide 2.5 counts just the single best interaction when two or more metal ligations are found. We set the coefficient to -2.0 kcal/mol, a value we believe to be reasonable, though the parameter refinement would have preferred an even more strongly negative value. Third, we assess the net charge on the metal ion in the unligated apo protein (generally straightforward via examination of the directly coordinated protein side chains). If the net charge is positive, the preference for an anionic ligand is incorporated into the scoring function.

DOCKING ACCURACY:

This section evaluates Glide's ability to replicate the cocrystallized ligand geometries of a large collection of 282 publically available PDB24 complexes. Most of the well-known GOLD and FlexX test set members are included in this set, along with about 50 PDB complexes used in prospective customers' evaluations of Glide and another 50 complexes whose experimental binding affinities were used to develop one or more of the empirical scoring functions described in the literature (e.g., ChemScore). These complexes, along with others found in the FlexX and GOLD test sets, were used to calibrate the GlideScore algorithm. Due to Glide's inability to handle ligands with more than 35 rotatable bonds and its inability to deal with ligands that are covalently bonded (seven cases: 1aec, 1ase, 1blh, 1tpp, 1lmp, and 4est), our coverage of the GOLD and FlexX sets is not entirely complete (one case: 2er6). In addition, one complex (6rsa) was disregarded because it contains a vanadium atomic species, which lacks parameters in the OPLS-AA force field employed by Glide.

CombiGlide includes the following capabilities and features:

Library enumeration: a. Enumerate complete combinatorial libraries. b. Untangles and minimizes structures

Interactive enumeration: a. Define and manage collections of fragments (R groups) b. Enumerate complete combinatorial libraries

Virtual combinatorial screening:

1. Performs rapid screening of large virtual combinatorial libraries against 3D targets
2. Are orders of magnitude faster than docking the entire library
3. Performs flexible docking using the standard and extra precision (XP) modes of Glide
4. Provides multiple post-docking library selection strategies and options
5. Allows for incorporation of predicted ADME properties into selection process
6. Analyzes selected libraries for enrichment of actives and chemical features

All workflows:

Provides automated reagent file preparation: 2D to 3D conversion, generation of reasonable ionizationand tautomeric states, stereoexpansion, assignment of attachment points

↓

Uses a “core plus side chains” approach

↓

Offers extensive flexibility in initial core placement

↓

Uses an intuitive, user-friendly wizard-based GUI for setting up and monitoring jobs and for visualization of docked poses

REFERENCES:

1) Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, *47*(7), 1739–1749. https://doi.org/10.1021/jm0306430

2) *CombiGlide 2.8: User Manual*. (n.d.). Retrieved October 3, 2022, from http://gohom.win/ManualHom/Schrodinger/Schrodinger_2012_docs/combiglide/combiglide_user_manual.pdf