Introduction to Molecular Docking

Edelmiro Moman
DEFINITION

Molecular docking tries to predict the structure of the intermolecular complex formed between two or more constituent molecules.
APPLICATIONS

- Virtual screening (hit identification)
- Drug Discovery (lead optimisation)
- Prediction of $K_A$ (biological activity ?)
- Binding-site identification (blind docking)
- De-orphaning of a receptor
- Protein – Protein (or Protein – Nucleic Acid) interactions
- Structure-function studies
- Enzymatic reactions mechanisms
- Protein engineering
The Protein Data Bank (PDB)

http://www.rcsb.org/
Most typical case: **Protein - Ligand docking**

- The final goal uses to be to predict the biological activity of a given ligand

Two different problems:

**POsing**

The process of determining whether a given conformation and orientation of a ligand fits the active site. This is usually a fuzzy procedure that returns many alternative results.

**SCORING**

The pose score is a measure of the fit of a ligand into the active site. Scoring during the posing phase usually involves simple energy calculations (electrostatic, van der Waals, ligand strain). Further re-scoring might attempt to estimate more accurately the free energy of binding ($\Delta G$, and therefore $K_a$) perhaps including properties such as entropy and solvation.
introduction to molecular docking

\[ [E]_{aq} \rightleftharpoons [E]_{aq} + [I]_{aq} \]  \hspace{1cm} (1)

\[ \Delta G_{\text{bind}} \]

\[ [E]_{aq} \hspace{1cm} [I]_{aq} \hspace{1cm} [E + I]_{aq} \]

The free energy of binding ($\Delta G$) is related to binding affinity by equations 2 and 3:

\[ \Delta G = -RT \ln K_A \]

\[ K_A = K_f^{-1} = \frac{[EI]}{[E][I]} \]

Prediction of the correct structure (posing) of the [E+I] complex does not require information about $K_A$. However, prediction of biological activity (ranking) requires this information; scoring terms can therefore be divided in the following fashion. When considering the term [EI], the following factors are important: steric, electrostatic, hydrogen bonding, inhibitor strain (if flexible) and enzyme strain. When considering the equilibrium shown in equation 1, the following factors are also important: desolvation, rotational entropy and translational entropy.

Molecular Mechanics Terms

**van der Waals**
\[ \Delta G_{vdW} = W_{vdW} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \]

**Hydrogen Bonding**
\[ \Delta G_{H-bond} = W_{H-bond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) \]

**Electrostatics**
\[ \Delta G_{elec} = W_{elec} \sum_{i,j} \left( \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \]

**Desolvation**
\[ \Delta G_{desolv} = W_{desolv} \sum_{i,j} \left( S_i V_j \exp\left(-\frac{r_{ij}^2}{2 \sigma^2}\right) \right) \]

**Change in Torsional Free Energy when the Ligand goes from Unbound to Bound**

**Torsional**
\[ \Delta G_{tor} = W_{tor} N_{tor} \]

**SCORING FUNCTION IN AUTODOCK**

\[
\Delta G = \Delta G_{vdW} \sum_{i,j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \\
+ \Delta G_{hbond} \sum_{i,j} E(t) \left( \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} + E_{hbond} \right) \\
+ \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \\
+ \Delta G_{tor} N_{tor} \\
+ \Delta G_{sol} \sum_{i,c,j} S_i V_j e^{-r_{ij}^2/2 \sigma^2}
\]
introduction to molecular docking

Scoring Functions

- Force-field-based
  - Dock
  - AutoDock
  - GOLD
  - D-Score
  - G-Score

- Empirical
  - Fresno
  - X-Score
  - LUDI
  - ChemScore
  - F-Score
  - SCORE

- Knowledge-based
  - PMF
  - SMoG
  - DrugScore
Posing and Scoring: Flexible docking algorithms

Flexible Docking Search Algorithms

Trial & Error
- Single Structure
  - Tabu Search
- Multiple Structures
  - Genetic Alg.

Systematic
- Systematic pose generation, stochastic optim.
- Incremental Construction
- Global Fit
  - Hyper Graph Cliqu.D.
  - Tree Search
  - Multiple Copy Rigid
  - Polyhed. Feature Graph Match

Simulation
- MD, QM/MM

Random
- Directed
  - Particle Swarm Optimisation
  - PSO@Autodock, SODOCK
introduction to molecular docking

Reconnected Ligand Pose:
MOLECULAR REPRESENTATIONS

Atomic

Surfaces

Grid
PROTEIN FLEXIBILITY

- Molecular Dynamics
- Energy Minimisation
- Monte Carlo
- Normal Modes
- Rotamer Libraries
- Protein Ensembles (NMR, MD, NMA) / Protein Ensemble Grids
- Soft Potentials
LIMITATIONS OF CURRENT DOCKING METHODOLOGIES

- Flexible ligands → Rotatable bonds → Combinatorial explosion
- Entropic effects → Rotatable bonds
- Solvation / desolvation → Accurate computation is expensive
- Water molecules (and ions)
- Tautomers
- Protein flexibility → Induced fit
- Specificity of binding → Understanding important interactions (currently larger ligands are favoured by the scoring functions)
- Pharmacokinetic effects, allosteric effects, biomolecule-biomolecule interactions (molecular context), etc.
DOCKING SOFTWARE
RESOURCES:

- Protein – Protein Interaction Website (docking software):
  
  http://www.imb-jena.de/jcb/ppi/jcb_ppi_software.html

- Structural Biology Software Database:
  
  http://www.ks.uiuc.edu/Development/biosoftdb/biosoft.cgi

- Molecular Docking Web:
  
  http://mgl.scripps.edu/people/gmm/

- Molecular Docking Servers:
  
  http://www.dockingserver.com/web
  
  http://bioinfo3d.cs.tau.ac.il/PatchDock/

- My Website:
  
  http://www.edelmiromoman.eu/

- AutoDock GUIs: AutoDockTools (ADT), BDT, DOVIS
CONCLUDING REMARKS

- Molecular modelling is about READING !!!
- Inspect all available structures of your protein: check for alternative conformations, R factors, ligands; become familiar with your binding site
- Carefully consider tautomerism, protonation, waters, X-ray resolution, etc.
- Also look to proteins of the same or related families / functions
- Collect all the relevant empirical data: site-directed mutagenesis, ligands, affinity / inhibition constants, etc.
- Critically analyse all this stuff in view of existing (and your own) hypotheses
- You are ready for modelling ! (docking is just the beginning)
- Find the right balance between simulations and experiments