Overview of the lecture

Introduction to molecular docking:

- Definition
- Types
- Some techniques
- Programs

"Protein-Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations"
What is Docking?

Docking attempts to find the “best” matching between two molecules
Given two biological molecules determine:

- Whether the two molecules “interact”
- If so, what is the orientation that maximizes the “interaction” while minimizing the total “energy” of the complex

**Goal:** To be able to search a database of molecular structures and retrieve all molecules that can interact with the query structure
Why is docking important?

- It is of extreme relevance in cellular biology, where function is accomplished by proteins interacting with themselves and with other molecular components.

- It is the key to rational drug design: The results of docking can be used to find inhibitors for specific target proteins and thus to design new drugs. It is gaining importance as the number of proteins whose structure is known increases.
Example: HIV-1 Protease

Active Site
(Aspartyl groups)
Example: HIV-1 Protease
Why is this difficult?

- Both molecules are flexible and may alter each other’s structure as they interact:
  - Hundreds to thousands of degrees of freedom (DOF)
  - Total possible conformations are astronomical
Types of Docking studies

- **Protein-Protein Docking**
  - Both molecules usually considered rigid
  - 6 degrees of freedom
  - First apply steric constraints to limit search space and the examine energetics of possible binding conformations

- **Protein-Ligand Docking**
  - Flexible ligand, rigid-receptor
  - Search space much larger
  - Either reduce flexible ligand to rigid fragments connected by one or several hinges, or search the conformational space using monte-carlo methods or molecular dynamics
Some techniques

- **Surface representation**, that efficiently represents the docking surface and identifies the regions of interest (cavities and protrusions)
  - Connolly surface
  - Lenhoff technique
  - Kuntz et al. Clustered-Spheres
  - Alpha shapes

- **Surface matching** that matches surfaces to optimize a binding score:
  - Geometric Hashing
Surface Representation

- Each atomic sphere is given the van der Waals radius of the atom.
- Rolling a Probe Sphere over the Van der Waals Surface leads to the Solvent Reentrant Surface or Connolly surface.
Lenhoff technique

- Computes a “complementary” surface for the receptor instead of the Connolly surface, i.e. computes possible positions for the atom centers of the ligand

![Diagram showing atom centers of the ligand and van der Waals surface](image)
Kuntz et al. Clustered-Spheres

- Uses clustered-spheres to identify cavities on the receptor and protrusions on the ligand
- Compute a sphere for every pair of surface points, $i$ and $j$, with the sphere center on the normal from point $i$
- Regions where many spheres overlap are either cavities (on the receptor) or protrusions (on the ligand)
Alpha Shapes

- Formalizes the idea of “shape”
- In 2D an “edge” between two points is “alpha-exposed” if there exists a circle of radius alpha such that the two points lie on the surface of the circle and the circle contains no other points from the point set.
Alpha Shapes: Example

Alpha=∞ Å

Alpha=3.0 Å
Surface Matching

- Find the transformation (rotation + translation) that will maximize the number of matching surface points from the receptor and the ligand

  - **First satisfy steric constraints…**
    - Find the best fit of the receptor and ligand using only geometrical constraints
  
  - **… then use energy calculations to refine the docking**
    - Select the fit that has the minimum energy
Docking programs

- **DOCK** (I. D. Kuntz, UCSF)

- **AutoDock** (Arthur Olson, The Scripps Research Institute)
  - AutoDock was designed to dock flexible ligands into receptor binding sites
  - The strongest feature of AutoDock is the range of powerful optimization algorithms available

- **RosettaDOCK** (Baker, Washington Univ., Gray, Johns Hopkins Univ.)
  - It models physical forces and creates a very large number of decoys
DOCK

DOCK works in 5 steps:

Step 1 Start with crystal coordinates of target receptor

Step 2 Generate molecular surface for receptor

Step 3 Generate spheres to fill the active site of the receptor: The spheres become potential locations for ligand atoms

Step 4 Matching: Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand

Step 5 Scoring: Find the top scoring orientation
- HIV-1 protease is the target receptor
- Aspartyl groups are its active side
DOCK

DOCK works in 5 steps:

- **Step 1** Start with crystal coordinates of target receptor

- **Step 2** Generate molecular surface for receptor

- **Step 3** Generate spheres to fill the active site of the receptor. The spheres become potential locations for ligand atoms

- **Step 4** Matching: Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand

- **Step 5** Scoring: Find the top scoring orientation
DOCK: Example

- HIV-1 protease is the target receptor
- Aspartyl groups are its active side
DOCK works in 5 steps:

Step 1 Start with crystal coordinates of target receptor

Step 2 Generate molecular surface for receptor

Step 3 Generate spheres to fill the active site of the receptor: The spheres become potential locations for ligand atoms

Step 4 Matching: Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand

Step 5 Scoring: Find the top scoring orientation
DOCK: Example

- HIV-1 protease is the target receptor
- Aspartyl groups are its active side
DOCK works in 5 steps:

- **Step 1** Start with crystal coordinates of target receptor
- **Step 2** Generate molecular surface for receptor
- **Step 3** Generate spheres to fill the active site of the receptor: The spheres become potential locations for ligand atoms
- **Step 4** Matching: Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand
- **Step 5** Scoring: Find the top scoring orientation
DOCK works in 5 steps:

- **Step 1** Start with crystal coordinates of target receptor
- **Step 2** Generate molecular surface for receptor
- **Step 3** Generate spheres to fill the active site of the receptor: The spheres become potential locations for ligand atoms
- **Step 4** Matching: Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand
- **Step 5** Scoring: Find the top scoring pose
Three scoring schemes: Shape scoring, Electrostatic scoring and Force-field scoring

Image 5 is a comparison of the top scoring orientation of the molecule thioketal with the orientation found in the crystal structure.
Other Docking programs

- FlexX
- GOLD
- Hammerhead
- FLOG
FLEXX

- Receptor is treated as rigid
- Incremental construction algorithm:
  - Break Ligand up into rigid fragments
  - Dock fragments into pocket of receptor
  - Reassemble ligand from fragments in low energy conformations
GOLD (Genetic Optimization for Ligand Docking)

Performs automated docking with full acyclic ligand flexibility, partial cyclic ligand flexibility and partial protein flexibility in and around active site.

Scoring: includes H-bonding term, pairwise dispersion potential (hydrophobic interactions), molecular and mechanics term for internal energy.

Analysis shows algorithm more likely to fail if ligand is large or highly flexible, and more likely to succeed if ligand is polar

- The GA is encoded to search for H-bonding networks first;
- Fitness function contains a term for dispersive interactions but takes no account of desolvation, thus underestimates The Hydrophobic Effect
Rigid-body docking algorithms

- Historically the first approaches.
- Protein and ligand fixed.
- Search for the relative orientation of the two molecules with lowest energy.
- FLOG (Flexible Ligands Oriented on Grid): each ligand represented by up to 25 low energy conformations.
Introducing flexibility: Whole molecule docking

- Monte Carlo methods (MC)
- Molecular Dynamics (MD)
- Simulated Annealing (SA)
- Genetic Algorithms (GA)

- Available in packages:
  - AutoDock (MC, GA, SA)
  - GOLD (GA)
  - Sybyl (MD)
Monte Carlo

- Start with configuration A (energy $E_A$)
- Make random move to configuration B (energy $E_B$)
- Accept move when:
  - $E_B < E_A$ or if
  - $E_B > E_A$ except with probability $P$:

$$P = \exp\left(-\frac{[E_A - E_B]}{kT}\right)$$
Molecular Dynamics

- force-field is used to calculate forces on each atom of the simulated system
- following Newton mechanics, calculate accelerations, velocities and new coordinates from the forces.
  - (Force = mass times acceleration)
- The atoms are moved slightly with respect to a given time step
Simulated Annealing

Finding a global minimum by lowering the temperature during the Monte Carlo/MD simulation
Genetic Algorithms

- Ligand translation, rotation and configuration variables constitute the genes
- Crossovers mixes ligand variables from parent configurations
- Mutations randomly change variables
- Natural selection of current generation based on fitness
- Energy scoring function determines fitness
Introducing flexibility: Fragment Based Methods

- build small molecules inside defined binding sites while maximizing favourable contacts.
- De Novo methods construct new molecules in the site.
- division into two major groups:
  - Incremental construction (FlexX, Dock)
  - Place & join.
Scoring Functions

- Shape & Chemical Complementary Scores
- Empirical Scoring
- Force Field Scoring
- Knowledge-based Scoring
- Consensus Scoring
Shape & Chemical Complementary Scores

- Divide accessible protein surface into zones:
  - Hydrophobic
  - Hydrogen-bond donating
  - Hydrogen-bond accepting
- Do the same for the ligand surface
- Find ligand orientation with best complementarity score
Empirical Scoring

- Scoring parameters fit to reproduce
- Measured binding affinities
  - (FlexX, LUDI, Hammerhead)
Empirical scoring

\[ \bigotimes G = \bigotimes G_0 + \bigotimes G_{rot} \cdot N_{rot} \]

\[ + \bigotimes G_{hb} \bigotimes f\left(\bigotimes R, \bigotimes \right) \]

\textit{neutral,H-bonds}

\[ + \bigotimes G_{io} \bigotimes f\left(\bigotimes R, \bigotimes \right) \]

\textit{ionic,int.}

\[ + \bigotimes G_{arom} \bigotimes f\left(\bigotimes R, \bigotimes \right) \]

\textit{arom.int}

\[ + \bigotimes G_{lipo} \bigotimes f\left(\bigotimes R, \bigotimes \right) \]

\textit{lipo.cont}

- Loss of entropy during binding
- Hydrogen-bonding
- Ionic interactions
- Aromatic interactions
- Hydrophobic interactions
Force Field Scoring (Dock)

\[ E_{\text{nonbond}} = \sum_{i < j} \left( \frac{A_{ij}}{r_{ij}^{12}} + \frac{B_{ij}}{r_{ij}^6} \right) + c \frac{q_i q_j}{r_{ij}^{\infty}} \]

Nonbonding interactions (ligand-protein):

- van der Waals
- electrostatics

Amber force field
Knowledge-based Scoring Function

Free energies of molecular interactions derived from structural information on Protein-ligand complexes contained in PDB

Boltzmann-Like Statistics of Interatomic Contacts.

\[ P(\mathcal{C}_p, \mathcal{C}_l) = P_{\text{ref}} \exp \left[ \mathcal{R} F(\mathcal{C}_p, \mathcal{C}_l) \right] \]
Distribution of interatomic distances is converted into energy functions by inverting Boltzmann’s law.
Potential of Mean Force (PMF)

\[ F_{ij}(r) = k_B T \ln \frac{n_{ij,\text{corr}}(r) \int_{ij,\text{seg}}(r)}{n_{ij,\text{bulk}}(r)} \]

Number density of atom pairs of type \( ij \) at atom pair distance \( r \)

Number density of atom pairs of type \( ij \) in reference sphere with radius \( R \)