Structure Prediction for Macromolecular Interactions

Julie C. Mitchell
San Diego Supercomputer Center
University of California at San Diego

Lecture Overview

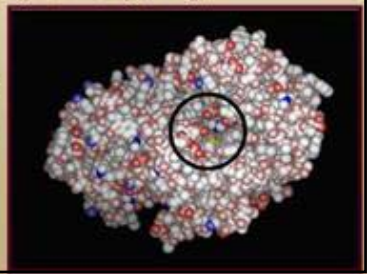
- *This lecture will discuss the binding of proteins to ligands. Ligands can be proteins, DNA or smaller molecules, such as phamaceutical compounds.*
- *The mechanisms underlying many biological processes involve the binding of two molecules.*
- *The physical laws that determine whether a pair of molecules can interact include both long and short-range forces that dictate biochemical function.*
- *The general principles for computer prediction of bound structures will be discussed, along with a survey of existing software.*

Molecular Docking

- *Molecular docking is often described in terms of a "lock and key" or "knobs and holes" model.*
- *Molecular docking is a matter of both geometry and physics. An optimal docking configuration should have good shape fit and be electrostatically favorable.*
- *Because interacting proteins can be highly flexible, or at least achieve some induced fit, predicting a bound complex can be as difficult as the sequence to structure prediction problem.*

Properties of Protein Active Sites

- *The region of a protein that interacts with a ligand is generally referred to as the "active site."*
- *Some general guidelines...*
  - *The active site generally lies on the surface of the protein. In some cases, the active site is buried within the protein.*
  - *Protein active sites tend to occur in large crevices along the surface. Some of these pockets have more hydrophobic residues than is typical.*
  - *Residues with reactive groups (Asp, Glu, Ser, Cys, His, Lys, Arg) tend to be abundant in protein active sites. The Ser-His-Asp (sometimes Ser-His-Glu) "catalytic triad" is a motif commonly found in enzyme active sites.*
  - *Some occur at the interface between two protein domains.*



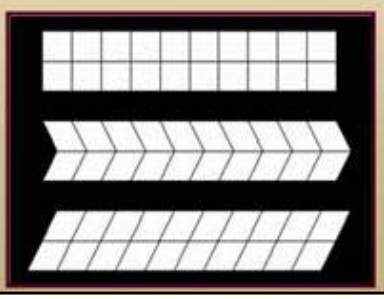
Docking Free Variables

- *The simplest model for molecular docking uses six free variables. The "receptor" is considered fixed in space while the ligand is free to move about it.*
- *A rigid motion of the ligand can be described in terms of six basic motions: translations along and rotations about the x/y/z axes.*

- *Many molecules undergo structural changes during interactions. This makes predicting the bound complex much more difficult.*

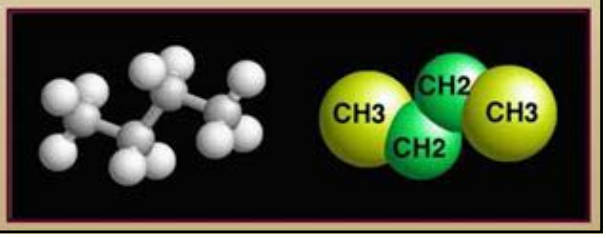
Protein Flexibility

- *There are three common types of motions seen in interacting proteins: sidechain rearrangement, hinge bending and shear.*
- *When modeling proteins for a docking prediction, setting variables to allow flexibility is more an art than a science. The goal is to have few free variables and still be accurate.*



United Atom Models

- *United atom models help simplify protein docking calculations in the same way residue-based models are used to speed protein folding computations.*
- *A united atom model combines hard atoms, such as carbon or nitrogen, and their bonded hydrogens into meta-atoms.*
- *Polar hydrogens should not be approximated using a united atom model.*

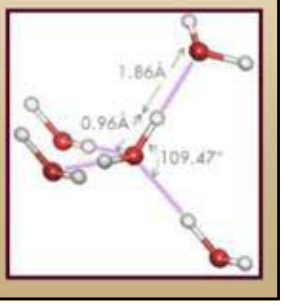


Long-range and Short-range Forces

- *Both short-range and long-range forces play a role in binding.*
- *Electrostatic interactions are the only long-range force involved in molecular docking.*
- *There are many short-range forces involved in protein binding.*
  - *Special types of non-covalent bonds, known as hydrogen bonds and salt bridges, are short-range electrostatic forces that contribute to the binding energetics.*
  - *Van der Waals interactions describe an attraction between nearby atoms.*
  - *Steric repulsion prevents two atoms from colliding.*
  - *The displacement of solvent near the binding site results in a desolvation (or Born) free energy.*

Hydrogen Bonding

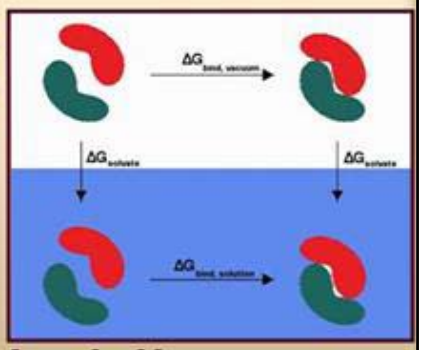
- *When a hydrogen atom is covalently bonded to an electronegative atom, such as nitrogen or oxygen, the hydrogen atom becomes very positively charged.*
- *A "hydrogen bond" occurs when a hydrogen atom covalently bonded to one electronegative atom is also attracted to another electronegative atom.*
- *Hydrogen bonds join DNA base pairs, and they hold together alpha helices and beta sheets in proteins.*
- *The formation of hydrogen bonds is an important part of many molecular interactions.*



Free Energy of Desolvation

- *Desolvation is a penalty for displacing water molecules from the binding site.*
- *The penalty for transferring an atom of charge Q and radius R from a medium of dielectric constant ε_1 to a medium of dielectric constant ε_2 is*

$$\triangle G = 166 \cdot \frac{Q^2}{R}\left(\frac{1}{\epsilon_1} - \frac{1}{\epsilon_2}\right) \quad \text{kcal/mol}$$



- *The desolvation free energy is described by a "thermodynamic cycle."*
- *Desolvation enhances electrostatic interactions.*

Direct Electrostatic Interactions

- *In vacuum, electrostatics are very long-range and follow Coulomb's Law, which is described by a Poisson Equation*

$$-\nabla\cdot(\epsilon\nabla\phi) = \rho$$

The dielectric constant, ε, measures the degree to which the atoms in a medium can be polarized.

- *Water is highly polarizable, and individual water molecules arrange themselves so as to counteract the protein's electrostatic force.*
- *In bound complexes, there can be strong non-covalent interactions ("salt bridges") that result from oppositely-charged residues in close contact.*

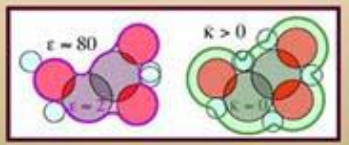
The Poisson-Boltzmann Equation

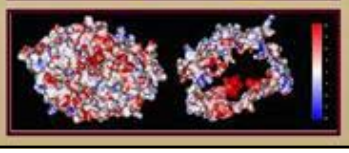
- *The Poisson-Boltzmann Equation describes an "implicit solvent" electrostatic model.*

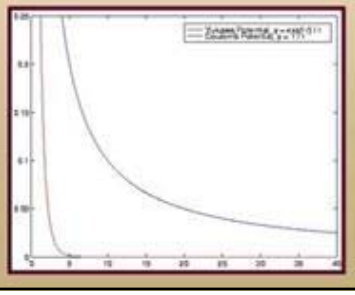
$$-\nabla\cdot(\epsilon\nabla\phi) + \bar{\kappa}^2\sinh(\phi) = \rho$$

- *The linearized Poisson-Boltzmann equation gives a good approximation in systems that are not highly-charged.*

$$-\nabla\cdot(\epsilon\nabla\phi) + \bar{\kappa}^2\phi = \rho$$







pK_a Values

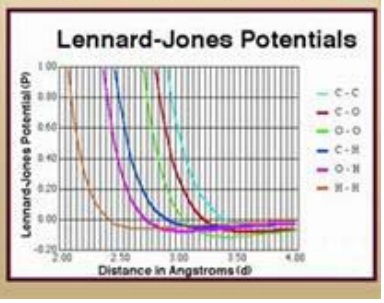
- The lower the pK_a of a sidechain, the more readily it acts as a proton donor for hydrogen bonds.
- When a residue is desolvated, its pK_a will shift so as to favor the uncharged form.
  - The pK_a values of Asp and Glu will shift upwards because of desolvation.
  - The pK_a values of Arg and Lys residues will shift downwards.
- When a residue forms salt bridges, its pK_a can shift to favor the charged form.
- The formula $pK_a = -\Delta G/RT$ relates the change in pK_a to a change in the free energy.
- The electrostatic properties of amino acid residues in the target depend very much on their local environment.

Lennard-Jones Potentials

- For protein-protein systems, the van der Waals interactions across a large binding interface can contribute a significant amount to the binding energy.
- When atoms get too close, steric repulsion occurs.
- The van der Waals interaction and steric repulsion are often modeled using a single function, the Lennard-Jones "6-12" potential

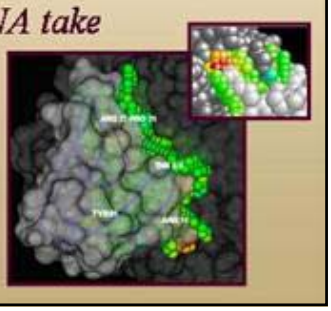
$$P(d) = A \cdot d^{-12} - B \cdot d^{-6}$$

Here, d is the distance between atom centers and A,B>0 depend on the atom types.



Shape Complementarity

- *The geometric match between a receptor and ligand is essential to most protein interactions.*
- *Interactions between two proteins often have large interfaces. Much of it consists of flattish regions in close contact, and there are a few "knobs and holes" that are key to providing a good shape fit.*
- *Interactions between proteins and DNA take place along saddle-shaped regions.*
- *Shape complementarity can be analyzed using surface matching, atomic density, buried surface area and alpha shapes.*



All Interactions Are Not Created Equal

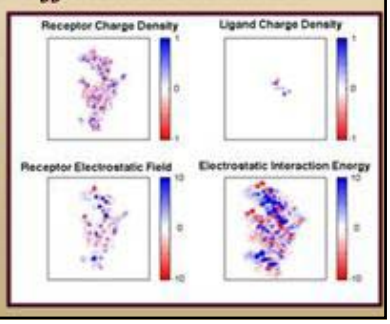
- *There are some generalities to be made regarding different types of protein interactions, although every system has unique features.*
  - *The interactions between proteins and small ligands are often driven electrostatically. Although the ligand must have a good geometric match at the binding site, van der Waals interactions may not contribute significantly to the interaction energy. Some hydrophobic "greasy" molecules may attach readily to most any site but use specificity to pick their receptor.*
  - *Interactions between two proteins often rely heavily on shape complementarity. In some systems (eg: antibody-antigen), van der Waals interactions are the only significant contribution, while in others (eg: enzyme-inhibitor) there can large electrostatic forces.*
  - *Protein-DNA interactions tend to have large electrostatic components and lots of hydrogen bonds. The reliance on shape complementarity varies from system to system.*

Predicting Molecular Interactions

- *Most currently available software for molecular docking uses either exhaustive search, global optimization or a combination of these methods.*
- *Every method relies on some sort of scoring function.*
  - *Many methods use potential energy models.*
  - *Some use purely geometric measures.*
- *The correct solution can often be predicted from shape match alone, but an energetic model provides a more biologically realistic analysis.*
- *The purely geometric methods are often faster, but the energy-based methods are more accurate in some circumstances.*

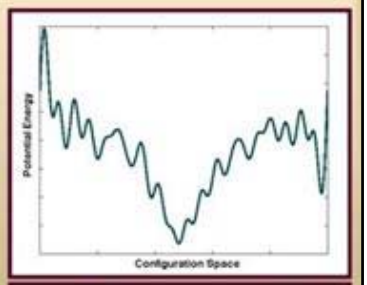
Exhaustive Search

- *It is impossible to examine every possible orientation between a receptor and ligand.*
- *A common approach to rigid docking problems is to search a large number of docking configurations.*
- *It is easy to distribute sample points for the three translational variables. It is more difficult to create a uniform sample of rotations.*
- *Exhaustive search methods are less effective for flexible docking.*
- *Most exhaustive search methods rely on Fourier correlation.*



Energy Minimization

- Energy minimization is another common method used in docking. Global minimization of a good energy model can often provide a good docking prediction.
- Finding the global minimum is complicated by the presence of many local minima.
- Global minimization methods used in molecular docking include:
  - Simulated Annealing
  - Convex Global Underestimation
  - Branch and Bound
  - Genetic Algorithms
  - Monte Carlo Minimization



Survey of Available Tools

Molecular Docking

DOCK	(E)
AutoDock	(G)
DOT	(E)
ICM	(G)
FlexX	(G)
GRAMM	(E)
FTDOCK	(E)
DOME*	(E,G)

PB Electrostatics

UHBD
DelPhi
APBS

Molecular Shape Analysis

FADE and PADRE
CAST
[B. Duncan & A. Olson]
[M. Connolly]

Charge Models

AMBER	(A,U)
CHARMM	(A)
PARSE	(A)

E	=	Exhaustive Search
G	=	Global Optimization
A	=	All Atom
U	=	United Atom
*	=	Not yet available at this writing

Topics For Discussion

- *In a weakly-charged system of two large molecules, what force is likely to dominate the interaction energy?*
- *Do you think a fast-acting enzyme must be highly charged?*
- *Assume that the electrostatic interaction between a ligand and receptor is favorable at all atoms in their interface. Would doubling the charges on all the ligand's atoms necessarily make it more favorable to the receptor? (Hint: what about desolvation?)*
- *Shape match is an intuitive notion that does not have a formal scientific definition. Discuss how shape complementarity relates to the van der Waals interaction and steric repulsion. Do you think we should define shape complementarity in these terms?*
- *Pick one of the programs listed on the previous page. Use a search engine to find the web home page for the project. Write and post a brief description of the program and the website link.*

Acknowledgements

- *Hydrogen bonding graphic*
  - http://webexhibits.org/causesofcolor
- *Protein motion movie*
  - http://molmovdb.org *(M Gerstein & WG Krebs (1998) Nuc. Acid. Res. 26:4280-4290)*
- *All materials produced using an Apple Titanium Powerbook*
  - *Rotating molecule movie created with MM2 Lite*
  - *Protein still images created with Rasmol, Molscript and WebLab Viewer*
  - *Potential function and shear motion graphics created with MatLab*
  - *Movie editing performed with MediaEdit*
  - *Image editing performed with GraphicConverter*