THE UNIVERSITY of TEXAS
HEALTH SCIENCE CENTER AT HOUSTON
SCHOOL of HEALTH INFORMATION SCIENCES

Ligand Binding, Docking, and Screening

For students of HI 6327 “Biomolecular Modeling”

Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/modeling/10.html

Biological Importance of Ligand Binding

- binding of stubstrates
  → enzymatic reaction

- binding of inhibitors and effectors
  → metabolic regulation
  → signal transduction

- binding of effectors to regulators
  → genetic regulation

- binding of ions to nucleic acids and proteins
  → control of conformations

The Problem of Drug Design

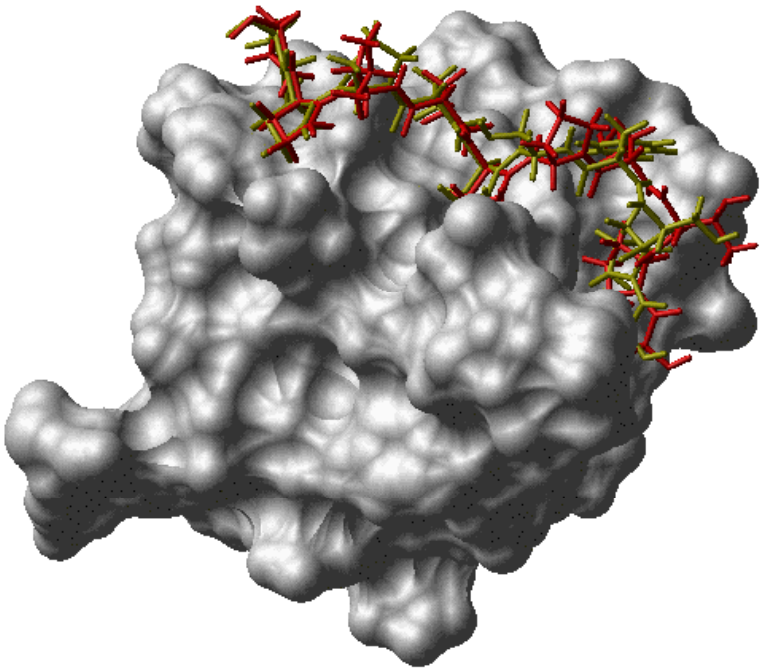
1. Where is the ligand binding pocket?

2. How to obtaining a structure for the ligand docked to the receptor?

3. How to calculate binding energies and binding constants?

4. How to construct an optimal ligand?

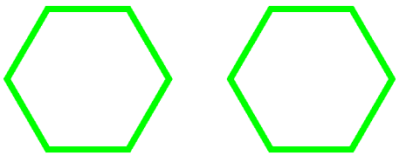
5. What can you do if you don't know the structure of the receptor?



Ligand Binding Energy Contributions

hydrophobic contacts

hydrogen bonds



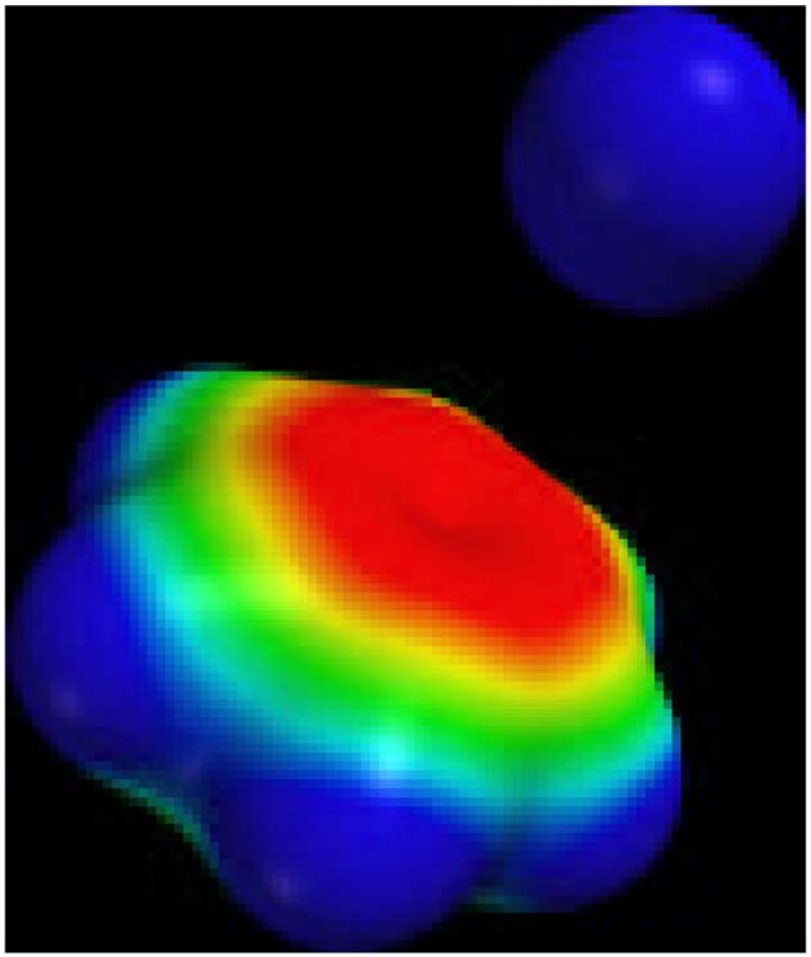
$\pi - \pi$-contacts

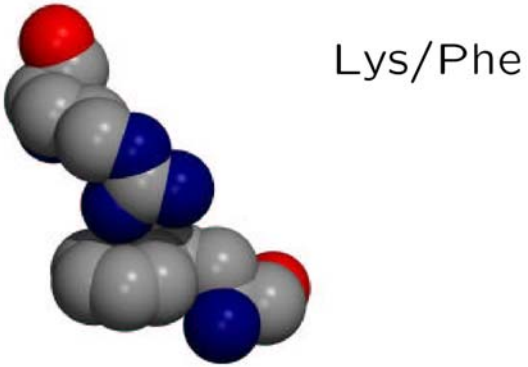
cation-π contacts

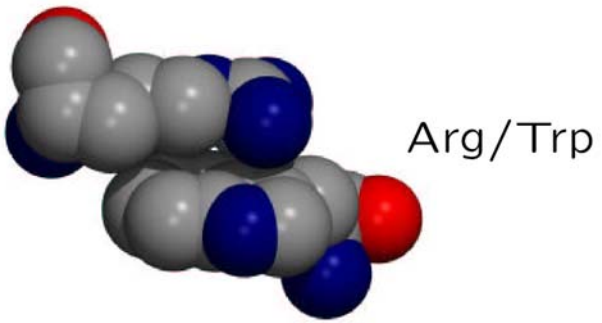
salt bridges

metal ion-ligand bonds

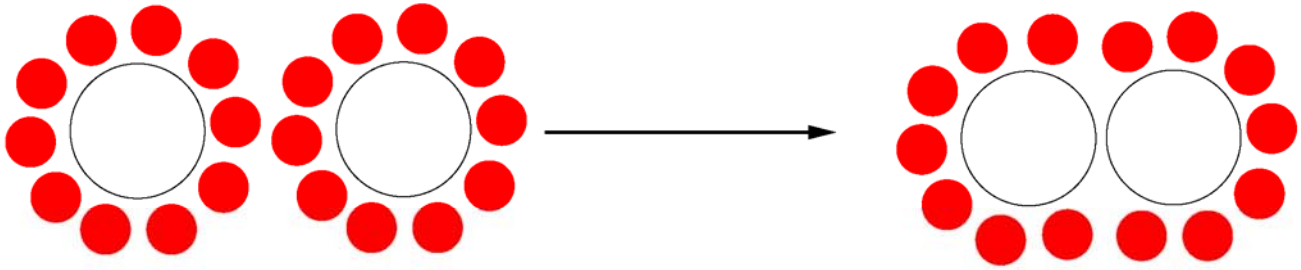
Cation –π Interactions







Hydrophobic Interactions



Entropic Effect: Less water needs to structure around the solutes

$$G_{non-polar} = \gamma A$$

A: Solvent Accessible Surface (computation: see below)

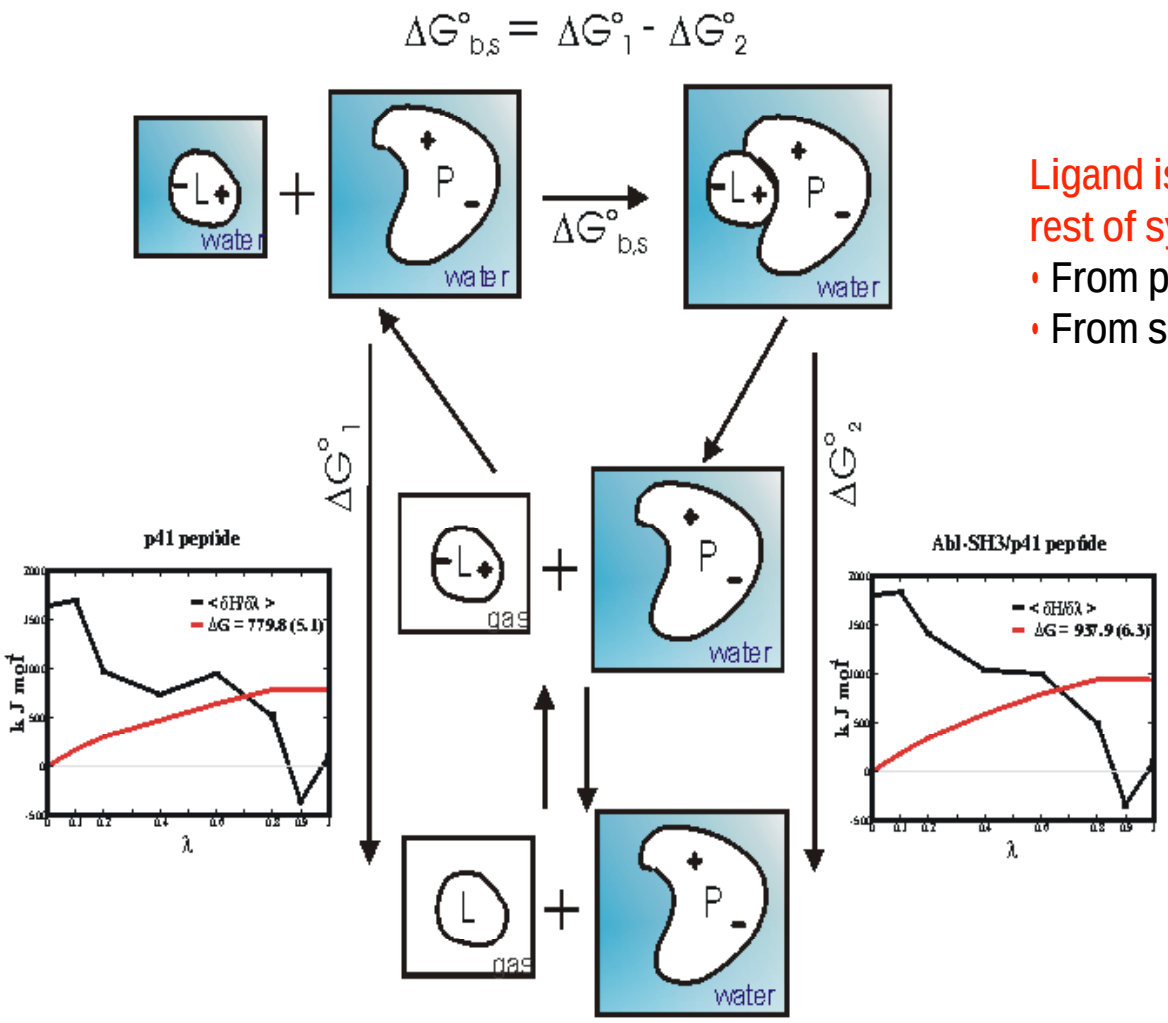
γ: 5 ... 20 cal/Å^2

Methods of Binding Energy Calculation

- Simulation methods:
  - complex and slow
  - clear physical and chemical concepts:
    - no *a priori* assumptions

- Continuum approach:
  - simple and fast
  - physics and chemistry not well founded:
    - various contributions to the free energy are assumed to be independent.

Simulation Method: MD



Ligand is slowly decoupled from the rest of system in two steps:

- From protein + solvent
- From solvent

Thermodynamic integration

$$\Delta G_b^{\Theta} = \int_0^1 \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

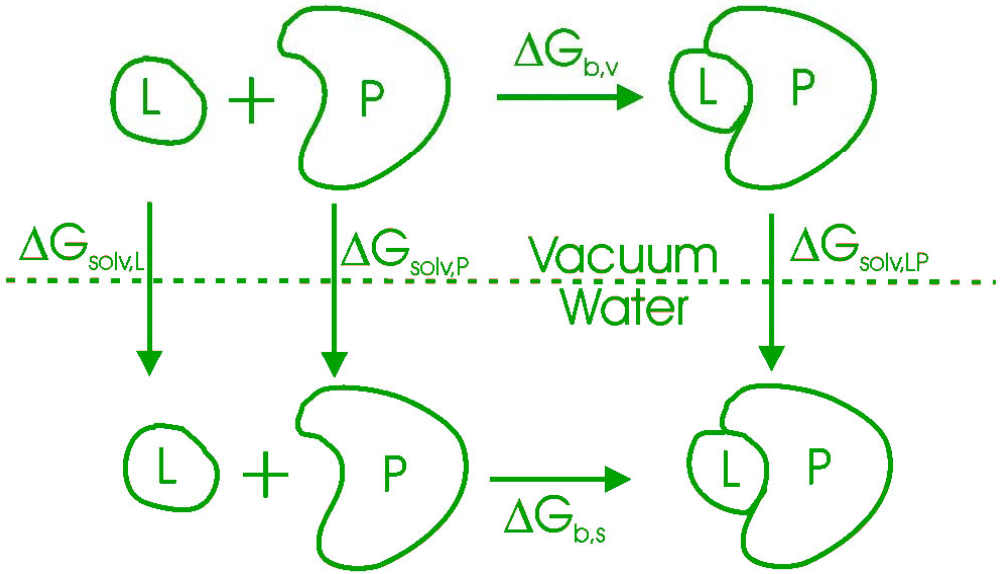
Both charged and uncharged ligands

Simulation Method: MD

PDB	Protein	Peptide ligand	ΔG^{Θ} calc.	ΔG^{Θ} exp.
1bbz	Abl-SH3	NH_3^+-APSYSPPPPP-COO^-	-168.0 (8.1)	-33.4 (0.2)
1abo	Abl-SH3	NH_3^+-APTMPPPLPP-COO^-	-135.1 (10.8)	-25.5 (0.2)
1lkk	P56-LCK SH2	Ace-pYEEI-COO^-	-247*	-38.9 (0.1)
1shd	C-SRC SH2	Ace-pYEEI-COO^-	-257*	-39.7 (0.1)

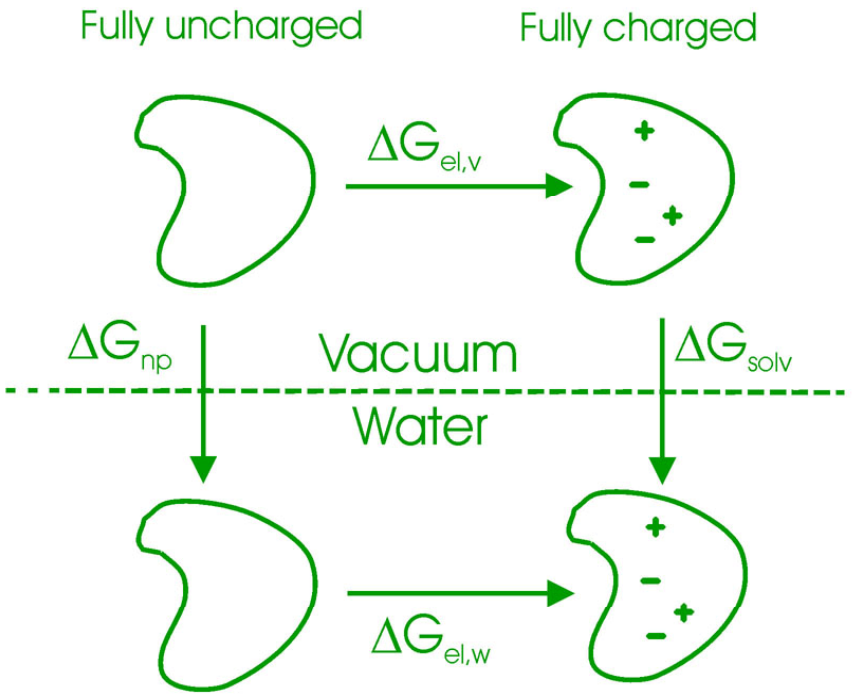
Correlation: 0.96

Continuum Approach

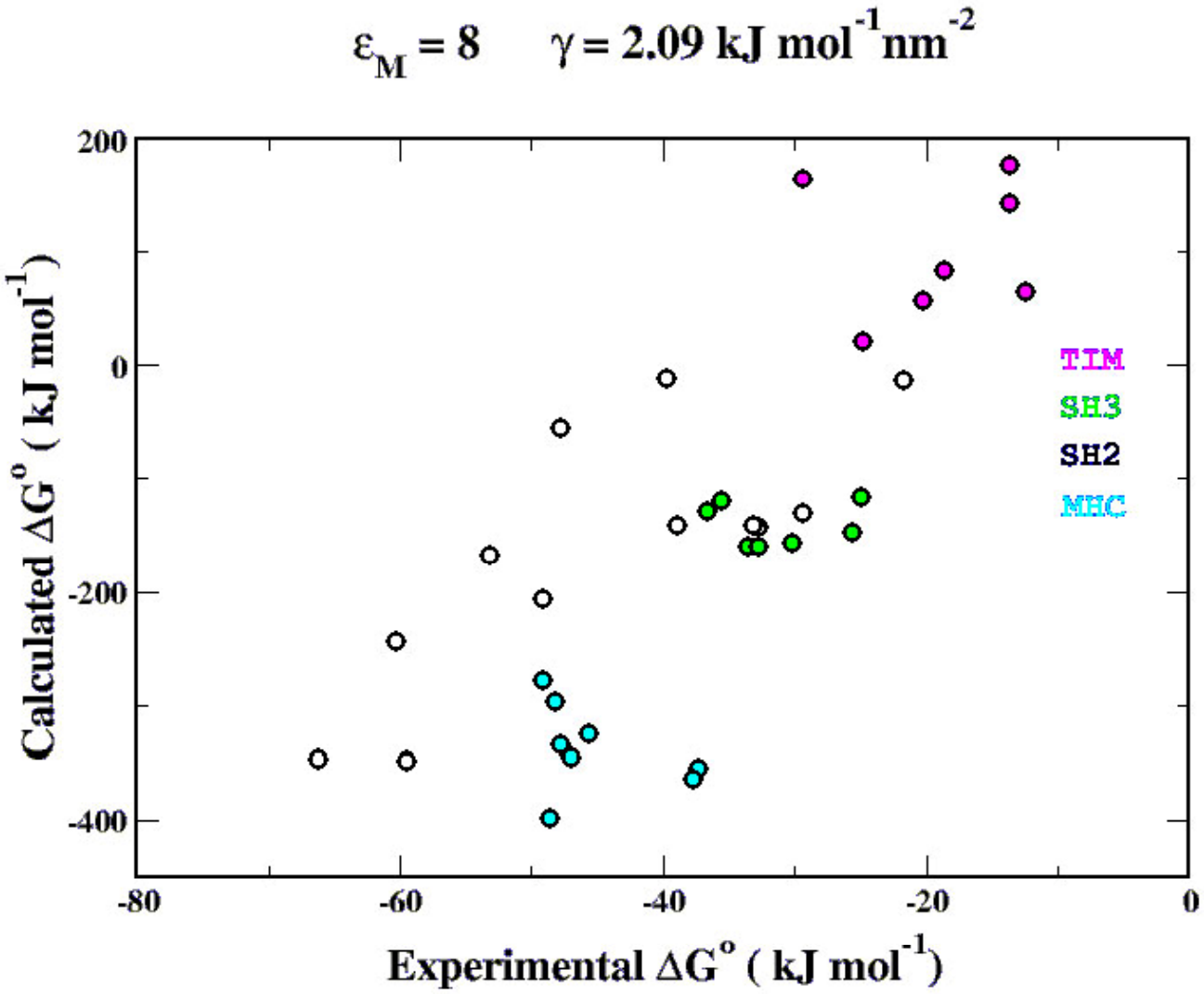


Thermodynamic cycle

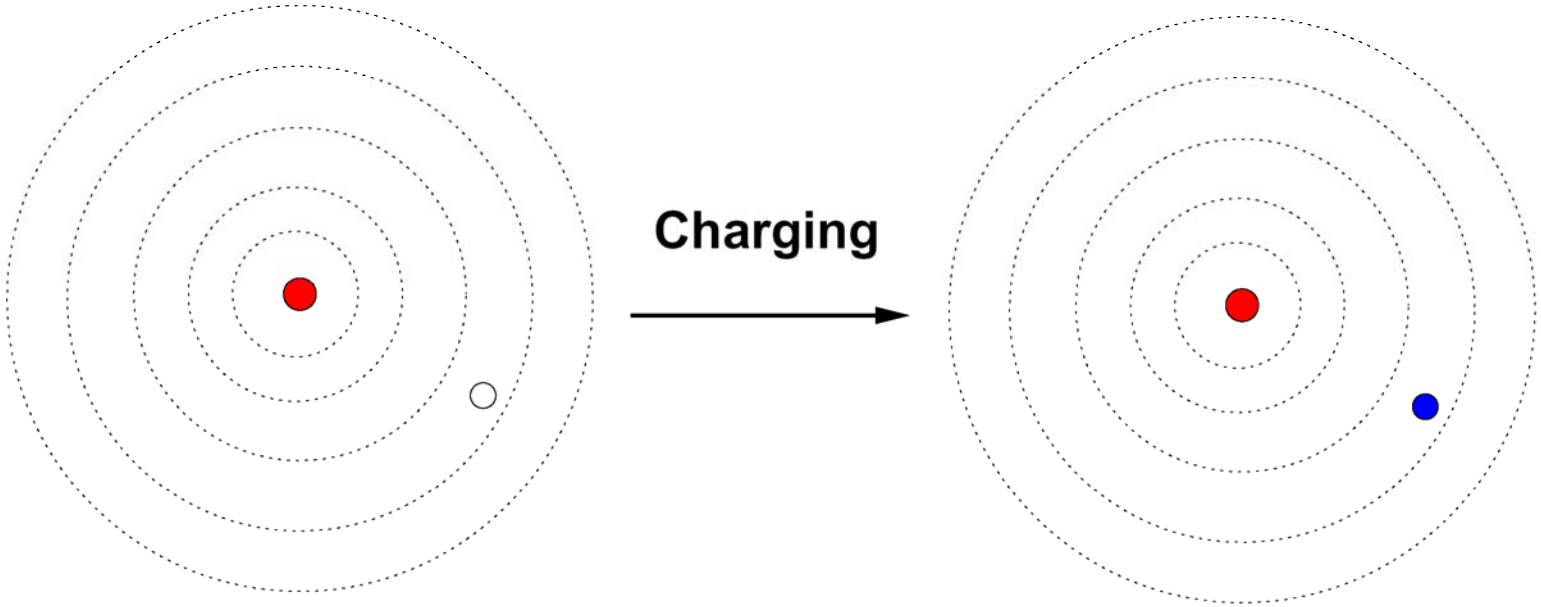
Solvation free energy



Continuum Approach



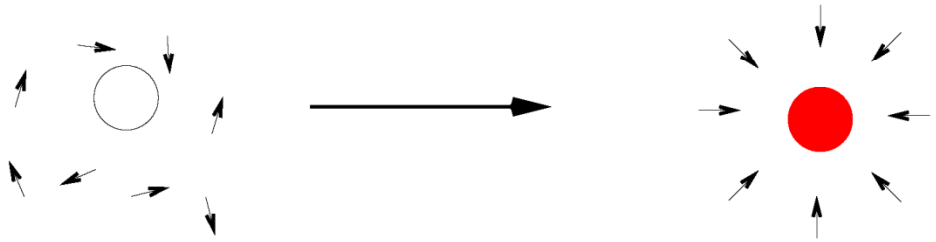
Purely Electrostatic Contributions



$$W = \int_0^{q_f} \phi \, dq = \phi \, q_f$$

Interaction energy of the two charges.

Reaction Field Energy



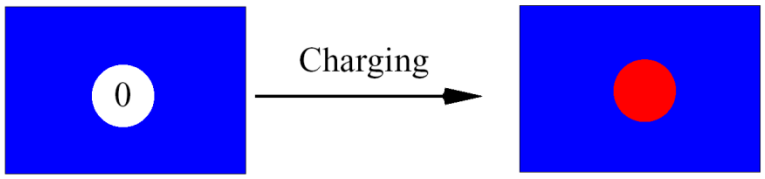
Total Potential:

$$\phi_{tot} = \phi(\rho) + \phi_{rf}$$

$$\phi_{rf} = C\ q; \quad C < 0$$

What is the work required to polarize the environment?

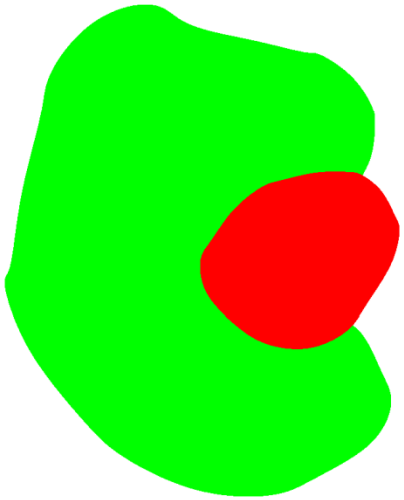
Self Energy: work to charge q from 0 to q_f



$$W = \int_0^{q_f} \phi_{rf}\ dq = \int_0^{q_f} Cq\ dq$$

$$= \frac{1}{2}\ C\ q_f^2 = \frac{1}{2}\ \phi_{rf}\ q_f$$

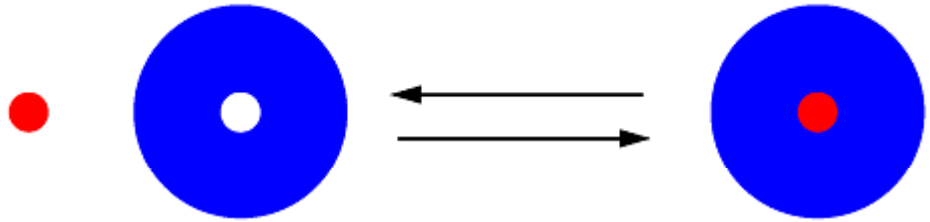
Electrostatic Energy of a Ligand in a Protein



$$E = \frac{1}{2} \sum_{i=1}^{Ligand} q_i \phi(\rho_{Ligand}) + \sum_{j=1}^{Protein} q_j \phi(\rho_{Ligand})$$

Self Energy Interaction Energy

Binding of a Single Ligand



$$\mathrm{L} + \mathrm{R} \overset{K}{\rightleftharpoons} \mathrm{RL}; \qquad K = \frac{[\mathrm{RL}]}{[\mathrm{R}][\mathrm{L}]}$$

free energy of ligand binding:

$$G = G^\circ + k_B T \ln K \quad \text{(follows from Boltzmann)}$$

At equilibrium: $G = 0 \rightarrow G^\circ = -k_B T \ln K$

chemical potential μ of a compound

$$\mu_\mathrm{A} = \mu_\mathrm{A}^\circ + k_B T \ln [\mathrm{A}]$$

A is either R, L, or RL

Chemical Potential

The free energy is often considered to be the most important quantity in thermodynamics, which is usually expressed as *Gibbs free energy*, G. The Gibbs free energy is appropriate for constant number of particles, temperature and pressure (constant *NPT*). Most experiments are conducted under conditions of constant N, T and P.

The **chemical potential** can be defined as: $\mu = \left(\frac{\partial G}{\partial N}\right)_{T,P}$

Which is just the Gibbs free energy per mole of substance.

It can be thought of in many ways:

1. A measure of the “escaping tendency” for a component in a solution;
2. A measure of the reactivity of a component in a solution;
3. In equilibrium it is equal in two systems placed in diffusive contact;
4. Particles move from a region of high chemical potential to a region of low chemical potential;
5. The chemical potential of ethanol increases in the following series of solutions:

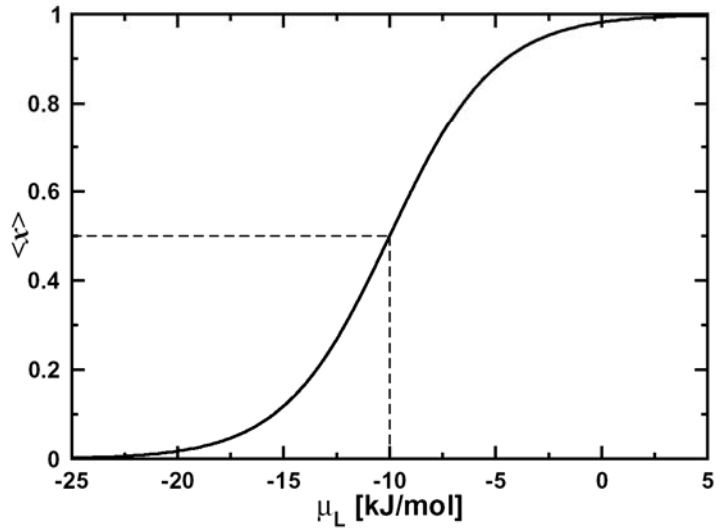
beer < wine < scotch whisky

Ligand Titration Curve

$$K = \frac{[RL]}{[R][L]} \quad \overset{\ln;\, k_B T}{\longrightarrow} \quad k_B T \ln K = k_B T \ln \frac{[RL]}{[R]} - k_B T \ln [L]$$

$$-G^\circ = k_B T \ln \frac{[RL]}{[R]} - \mu_L$$

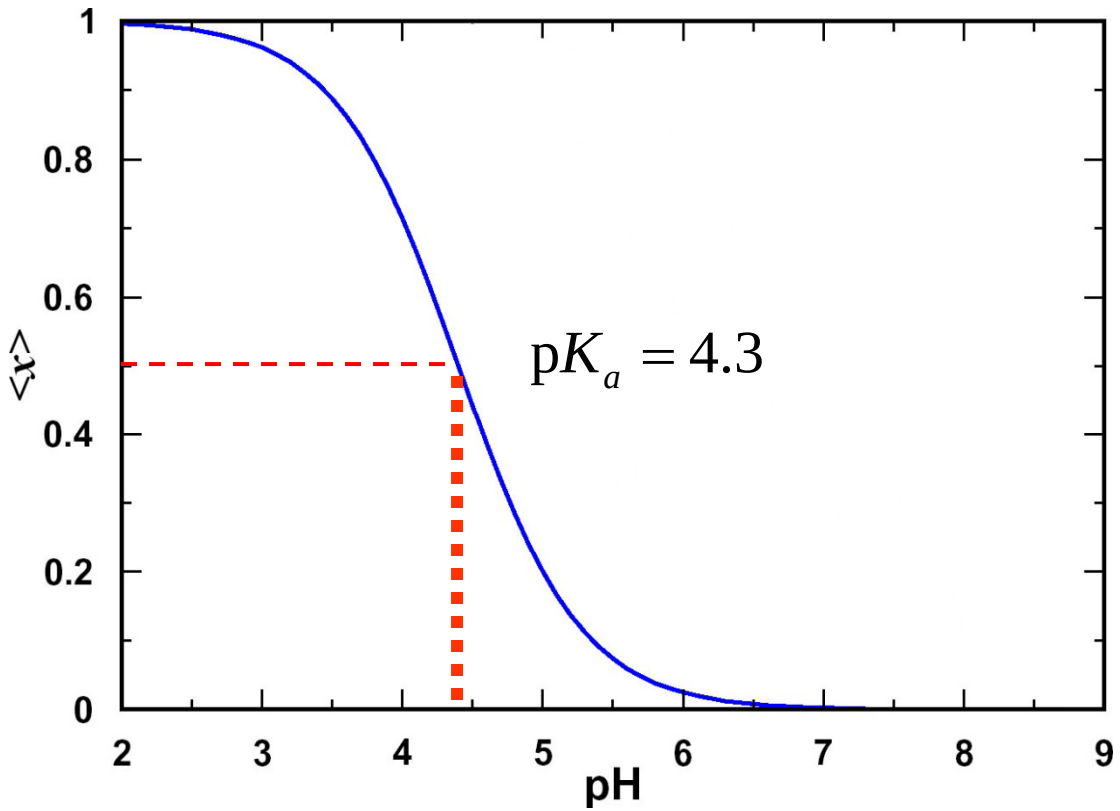
$$\langle x \rangle = \frac{[RL]}{[RL] + [R]} = \frac{e^{-(G^\circ - \mu_L)/k_B T}}{1 + e^{-(G^\circ - \mu_L)/k_B T}}$$



energy for ligand binding: $G(\mu_L) = G^\circ - \mu_L$

if $\langle x \rangle = 0.5 \longrightarrow G^\circ = \mu_L$

Example: pKa Calculations



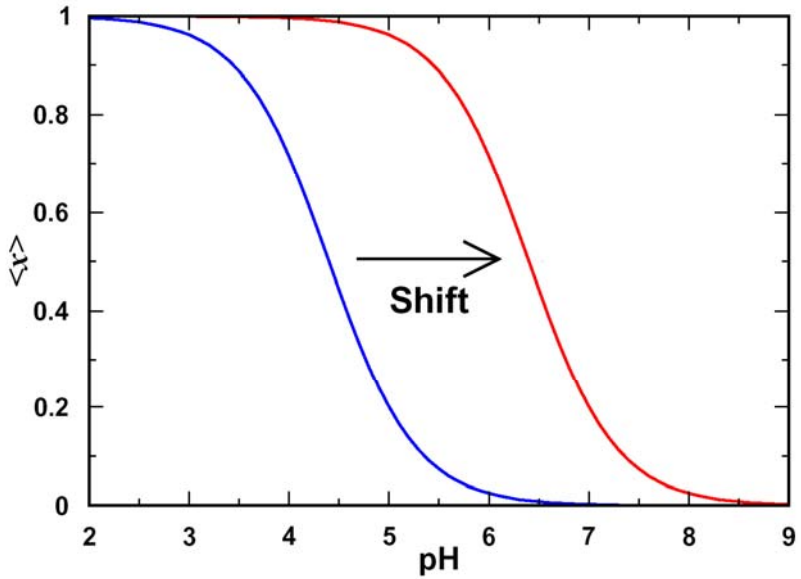
$$\langle x \rangle = \frac{1}{1 + e^{-\ln 10 (pK_a - pH)}}$$

Occupancy

Computer simulation:
Calculation of <x> as a function of pH

Example: pKa Calculations

Consider proton as a ligand
Assume: single titratable site

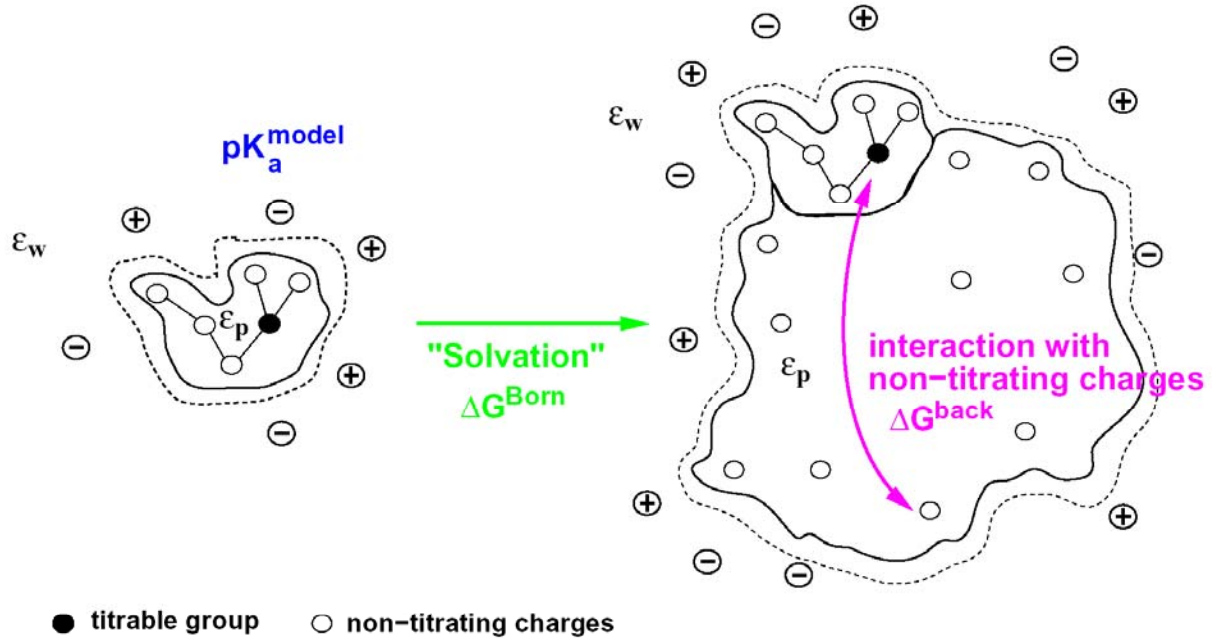


$$pK_a^{prot} = pK_a^{model} - \frac{1}{\ln 10 k_B T}(\Delta\Delta G^{Born} + \Delta\Delta G^{back})$$

from table

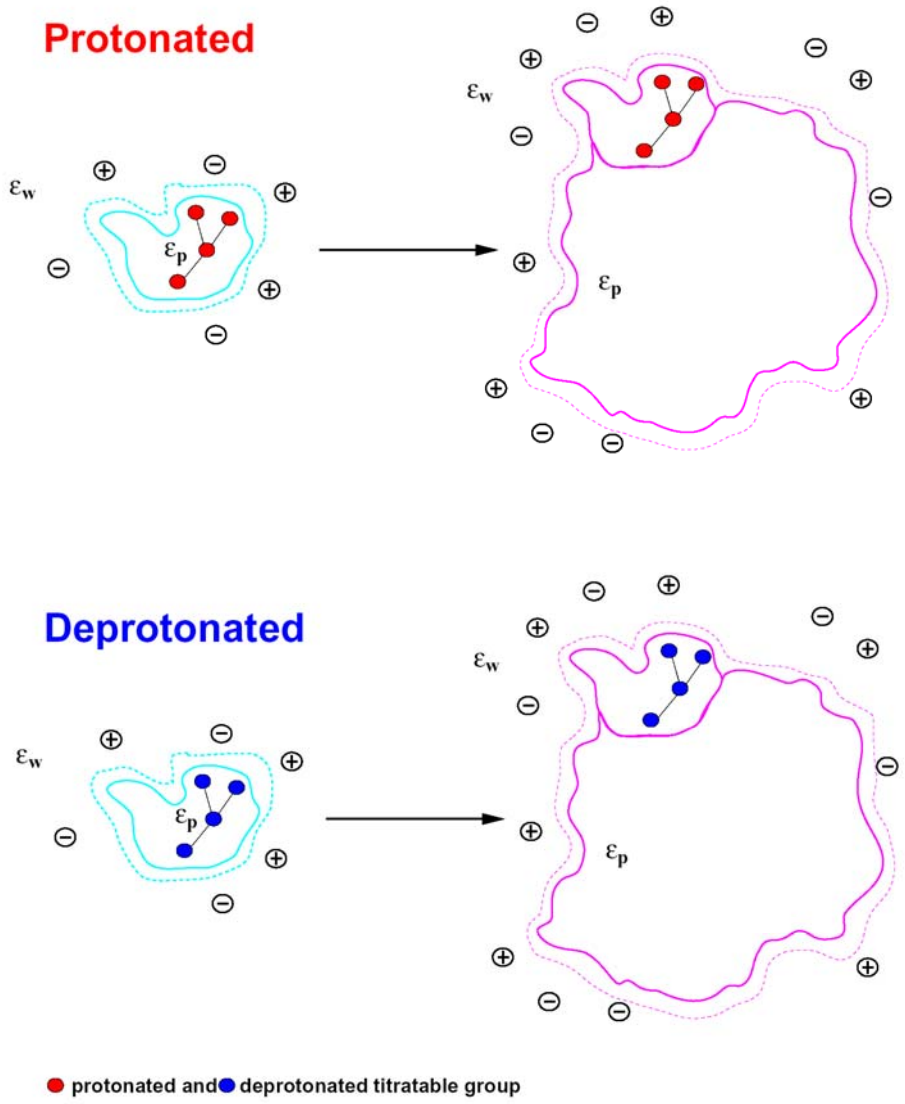
Example: pKa Calculations

Two "receptors": free aminoacid and protein



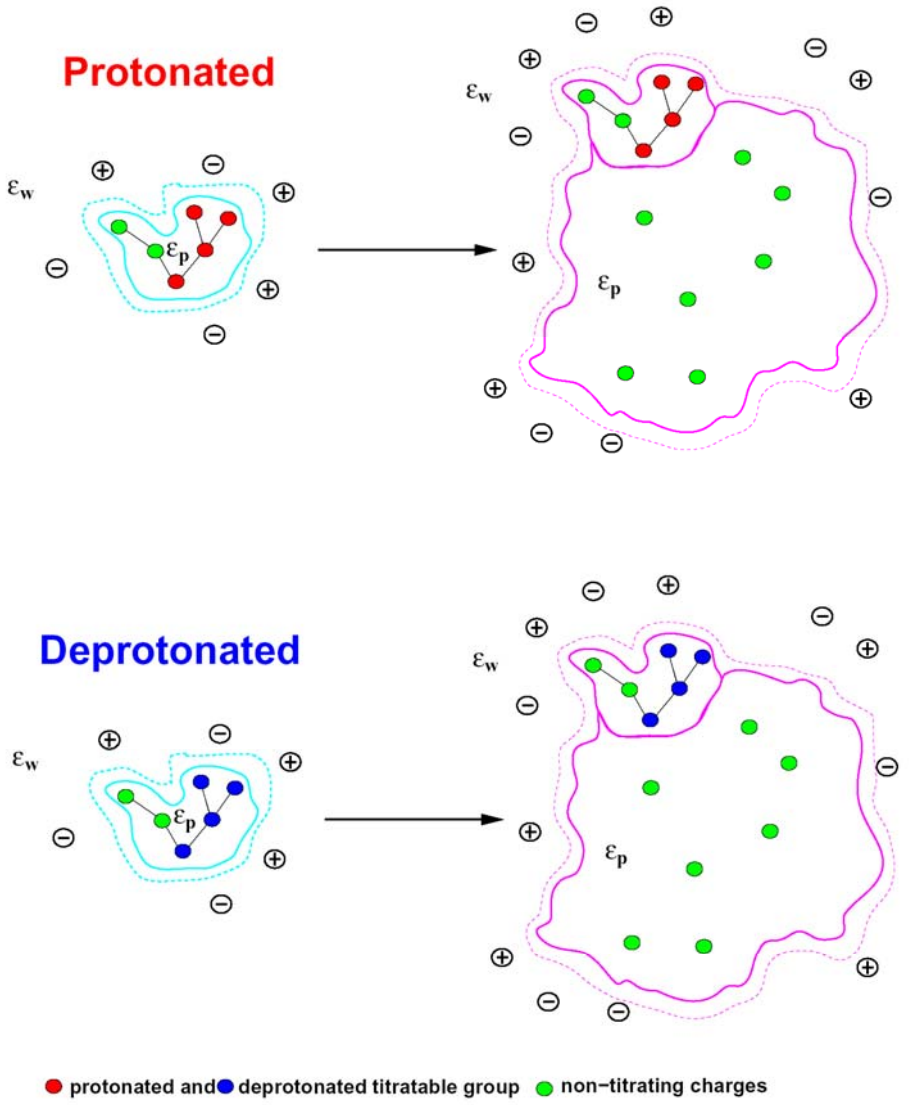
Change in Solvation Energy

$$\Delta\Delta G_{\text{Born}} = \frac{1}{2}\sum_{i=1}^{N_Q} Q_i^h[\phi_p(Q^h) - \phi_m(Q^h)] - \frac{1}{2}\sum_{i=1}^{N_Q} Q_i^d[\phi_p(Q^d) - \phi_m(Q^d)]$$



Change in Background Energy

$$\Delta\Delta G_{\text{back}} = \sum_{i=1}^{N_p} q_i[\phi_p(Q^h) - \phi_p(Q^d)] - \sum_{i=1}^{N_m} q_i[\phi_m(Q^h) - \phi_m(Q^d)]$$



Applications of pKa Calculations

- pK_a in the active site of a protein – implications for catalysis
- protonation change upon binding or conformational change – pH dependence of equilibirium properties
- proton transfer paths – proton transfer in bioenergetic systems like bacteriorhodopsin, the photosynthetic reaction center or cytochrome c oxidase

Questions: Ligand Binding Energies?

Biomolecular Interactions

•Biomolecular interactions are the core of all the regulatory and metabolic processes that together constitute the process of life

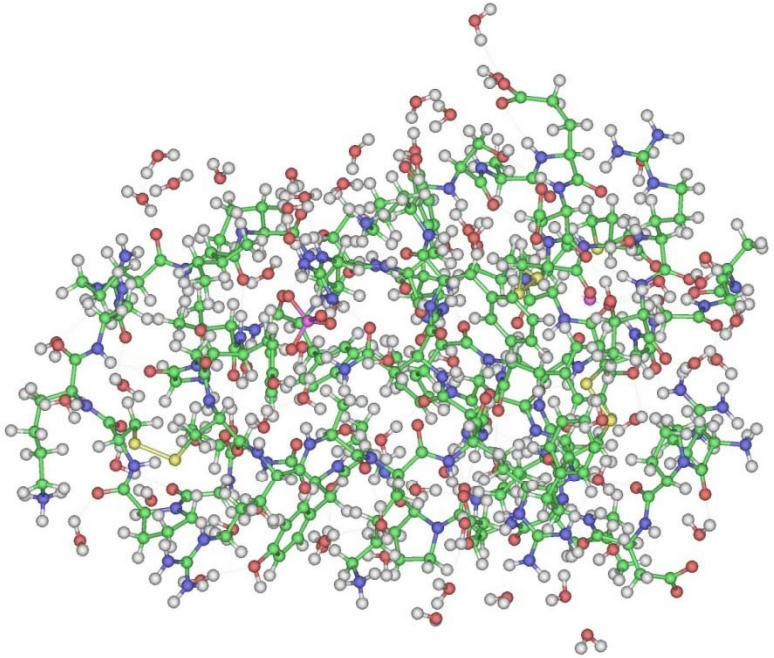
• Computer-aided analysis (e.g. docking) of these interactions is becoming increasingly important as the database of known biomolecular structures continues to grow

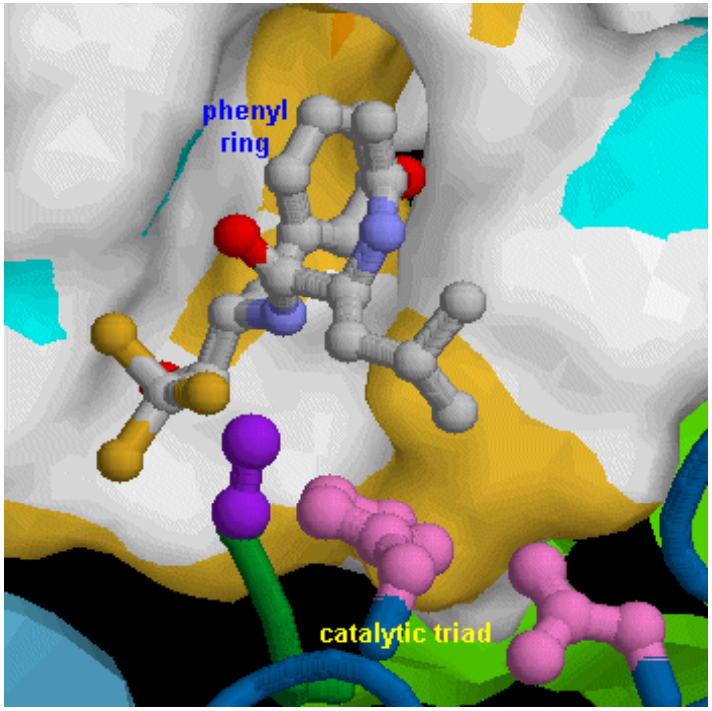
• Increasing processing power makes the analysis and prediction of molecular interaction more tractable

• Automated prediction of molecular interactions is the key to structure-based (rational) drug design

Structure-Based Drug Design

SBDD is a method that shortens the time for discovery of new drugs from 12-15 years to 4-6 years, and reduces the cost considerably





SBDD Achievements

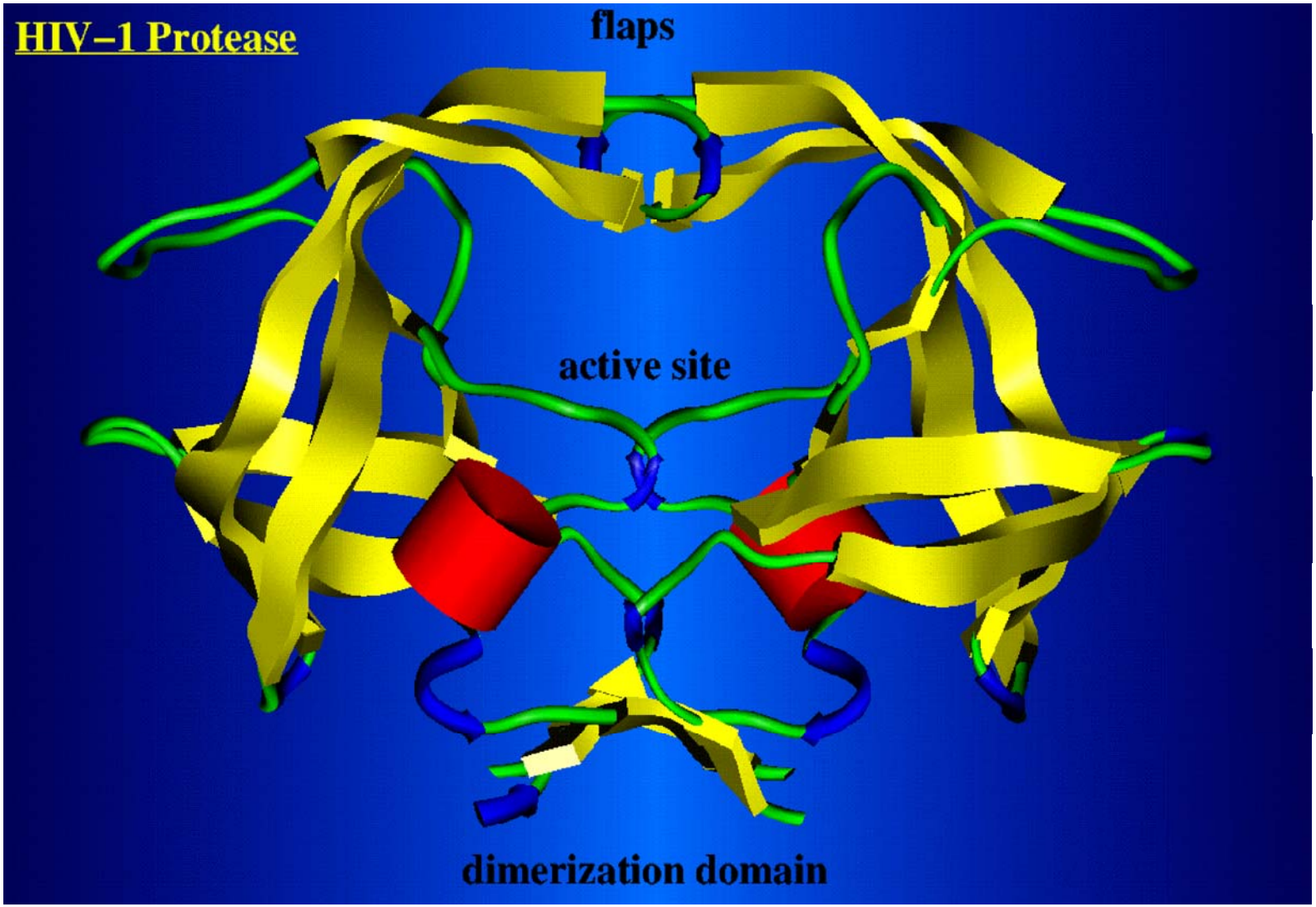
Structure based drug design is a relatively new field. It has recently proven to be successful in the design of the following drugs:

- Dorzolamide (Glaucoma treatment)
- Saquinavir, Indinavir, Ritonavir, Nelfinavir (HIV therapy)

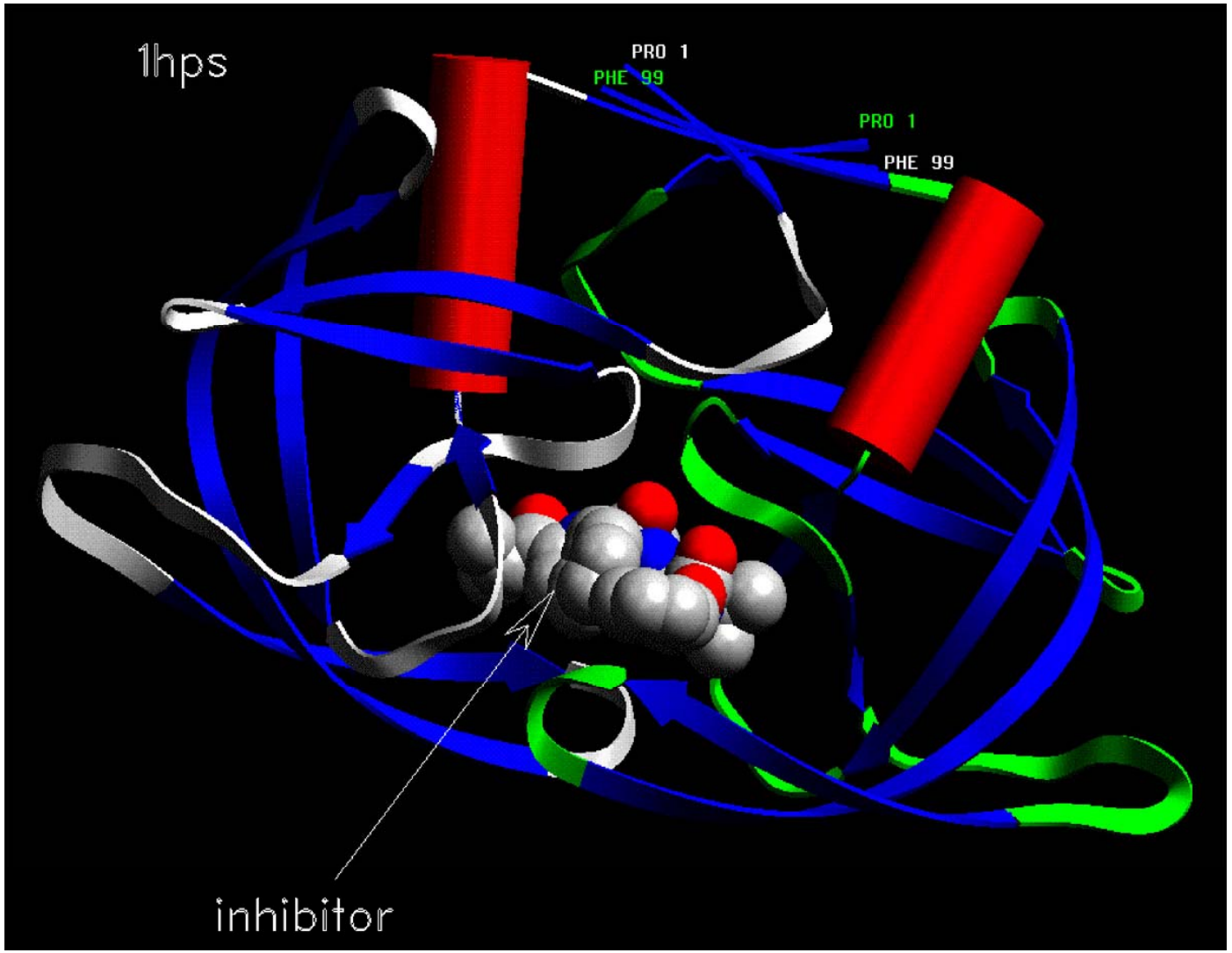
Currently In clinical trials:

- Neuroaminidase inhibitors (anti-influenza)
- COX-2 inhibitors (anti inflammatory)
- Phospholipase A(2) inhibitors (anti inflammatory)

An Example: HIV-1 Protease



An Example: HIV-1 Protease



The Problem

Given two biological molecules determine:

• Whether the two molecules “interact”, ie. is there an energetically favorable orientation of the two molecules such that one may modify the other’s function (do the two molecules fit together in any energetically favorable way).

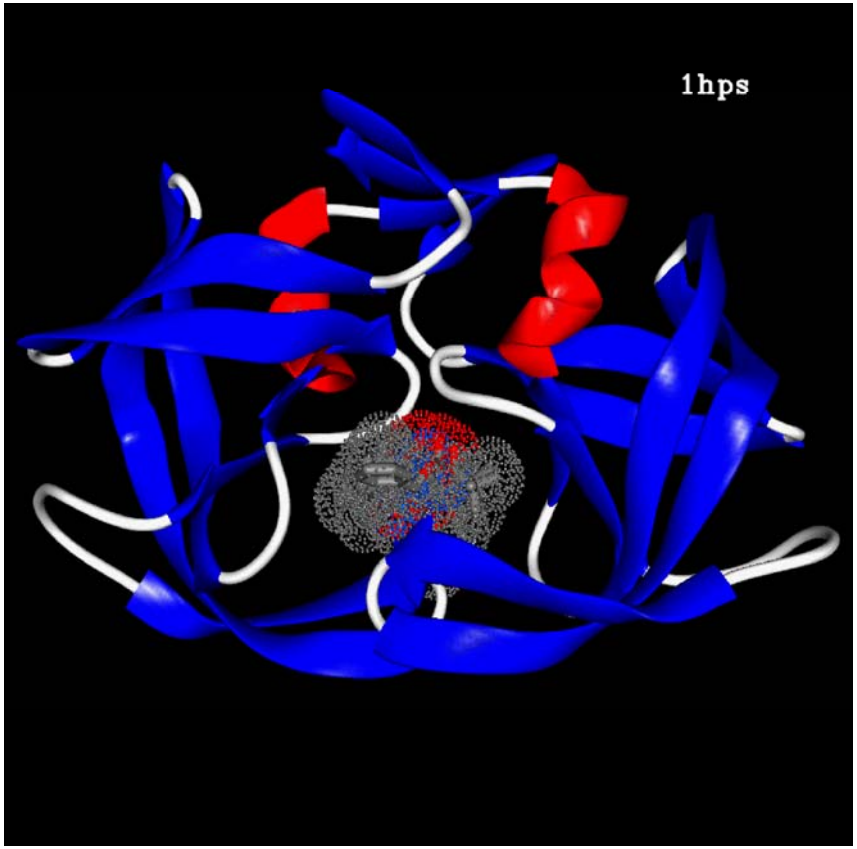
• If so, what is the orientation that maximizes the “interaction” while minimizing the total “energy” of the complex.

• Goal (rational drug design): To be able to search a database of molecular structures and retrieve all molecules that can interact with the query structure (*virtual ligand screening*).

Why is This Difficult?

Both molecules are flexible and may alter each other's structure:

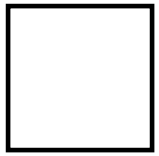
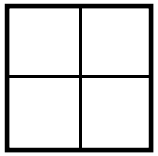
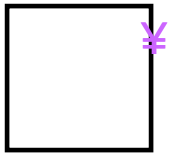
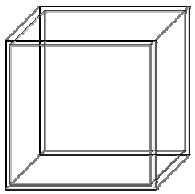
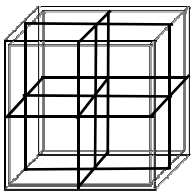
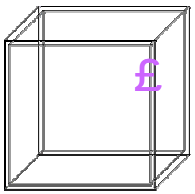
- Hundreds to thousands of degrees of freedom
- Total possible conformations are astronomical



Dimensionality of Docking Search

- Degrees of Freedom (DOF)
- Position or Translation
  - $(x,y,z) = 3$
- Orientation (Euler angles)
  - $(\Phi,\Theta,\Psi) = 3$
- Rotatable Bonds or Torsions (Flexible Ligands)
  - $(tor_1, tor_2, \ldots tor_n) = n$
- Total DOF, or Dimensionality,
  $D = 3 + 3 + n$

Multidimensional Treasure Hunt

Dimensions	Landscape	Divide into 2	Treasure	Chances?
1				0.5
2				0.25
3				0.125

Sampling Hyperspace

- Say we are hunting in *D*-dimensional hyperspace…
- We want to sample each of the D dimensions *N* times.
- The number of "evals" needed, *e*, is:

  $e = N^D$

  $\Rightarrow N = e^{1/D}$

  For example,

  - $e=10^6$, $D=6$, $N=10.0$ samples
  - $e=10^6$, $D=36$, $N=\sim 1.5$ samples

- More dimensions, tougher it gets.

Types of Docking Problems

Protein-Protein Docking

- both molecules usually considered rigid
- 6 degrees of freedom, 3 for rotation, 3 for translation
- first apply only steric constraints to limit search space
- then 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 (reduces conformational search space)
- Or search the conformational space using MM (Monte Carlo / MD)

Types of Docking Problems

Rough Docking

- Search a database of potential ligands to select lead compounds for drug design
- Often based on quick geometrical algorithms combined with heuristic functions to predict binding energy

Detailed Docking

- Accurate analysis of a single instance of docking
- To compute thermodynamic and kinetic properties of binding (free energy, rates of binding and dissociation)
- Computing free energy of binding requires models of both enthalpic and entropic contributions (last week!)
- Large amount of conformational sampling required to compute the entropy of the ligand in the binding site

Types of Docking Problems

Systematic

- Exhaustive.
- Deterministic.
- Dependent on granularity of sampling.
- Feasible only for low-dimensional problems.

Stochastic

- Random.
- Outcome varies.
- Repeat to improve chances of success
- Feasible for higher-dimensional problems.

Search Breadth and Detail

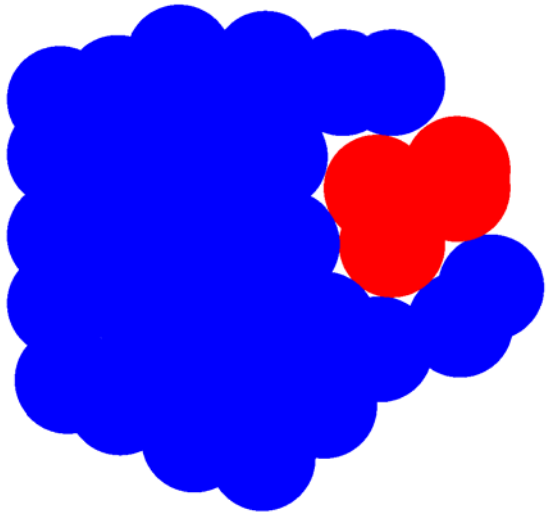
Search Breadth

- Local
  - –Molecular Mechanics
- Short - Medium
  - –Monte Carlo Simulated Annealing
  - –Brownian Dynamics
  - –Molecular Dynamics
- Global
  - –Docking

Level-of-Detail

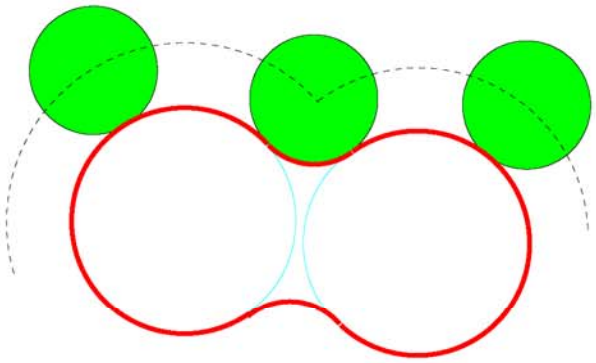
- Atom types
- Terms of force field
  - –Bond stretching
  - –Bond-angle bending
  - –Torsional potentials
  - –Polarizability terms
  - –Implicit solvation

Shape Complementarity



1. Looking for cavities big enough to hold the ligand
2. Fit the ligand in different orientations (and conformations) into the binding pocket
3. Calculate the binding energy for the different orientations and conformations to identify which orientation is the best.

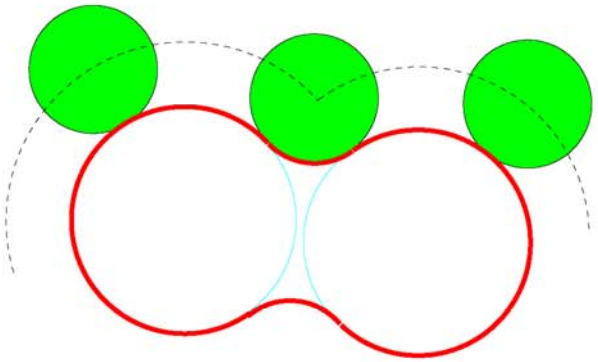
Solvent-Accessible Surface

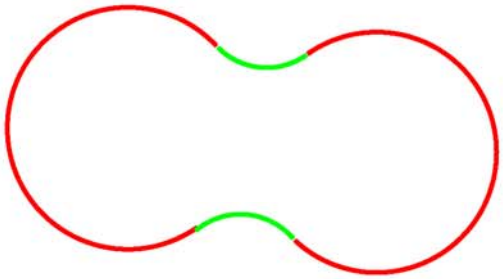


Rolling a Probe Sphere over the Van der Waals Surface leads to the Solvent Accessible Surface.

The probe sphere has the raduis of a solvent molecule (1.4 Å for water).

Solvent-Accessible Surface



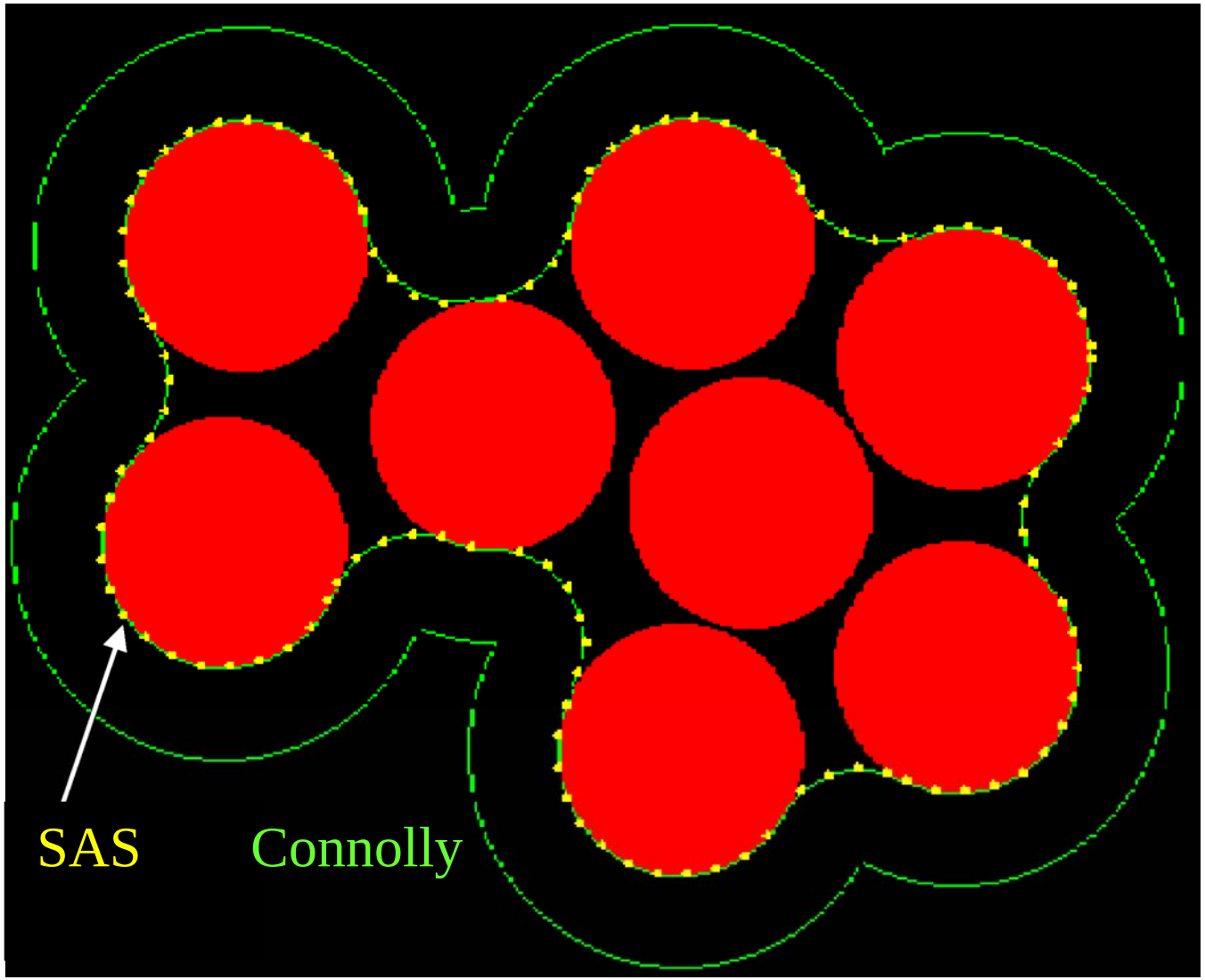


Rolling a Probe Sphere over the Van der Waals Surface leads to the Solvent Accessible Surface.

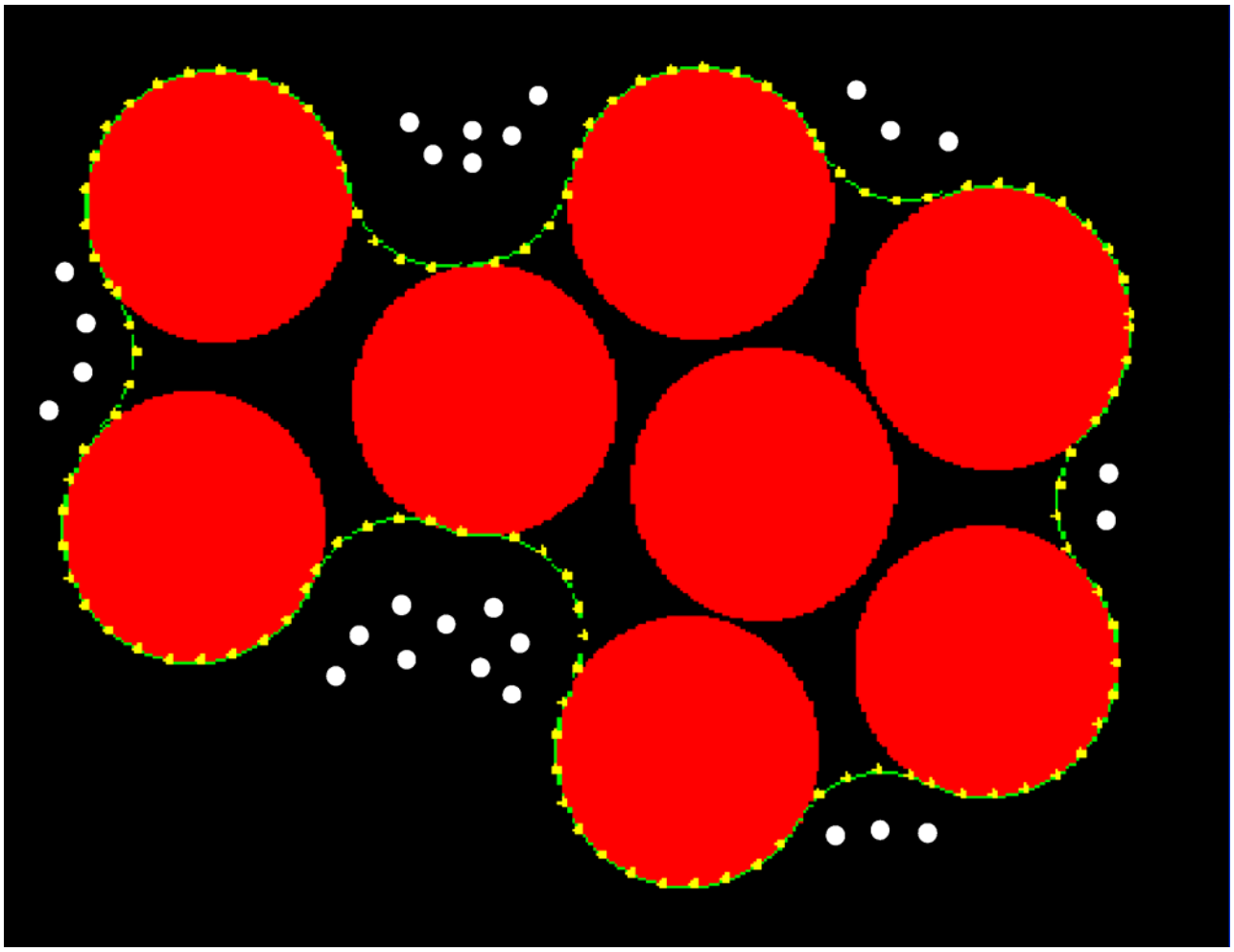
The probe sphere has the raduis of a solvent molecule (1.4 Å for water).

The Solvent Accessible Surface Area is then calculated from the parts forming the surface. (Conceptionally easy, but mathematically complicated).

Connolly Surface



Lenhoff “Surface”



Lenhoff “Surface”

- A “complementary” surface for the ligand instead of the Connolly surface of the receptor

- Possible positions (near the surface of the receptor) for the atom centers of the ligand

- Based on the contact-score of uniformly distributed points on probe spheres

Nussinov and Wolfson

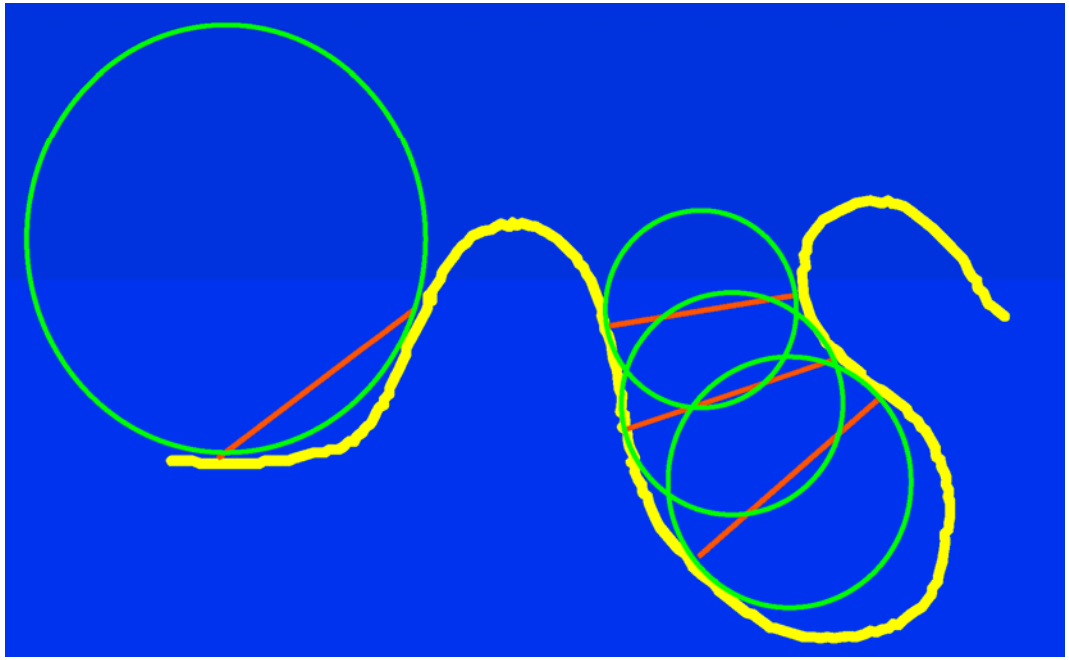
- Each concave, convex, and saddle face of the Connolly surface is replaced by a single critical point

- 44 atoms -> 5,355 Connolly Points -> 326 critical points

- Reduced complexity

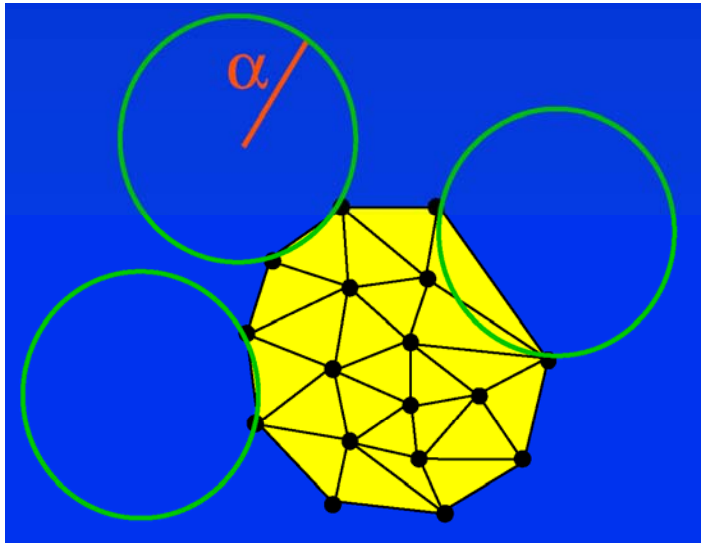
Kuntz

- 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)

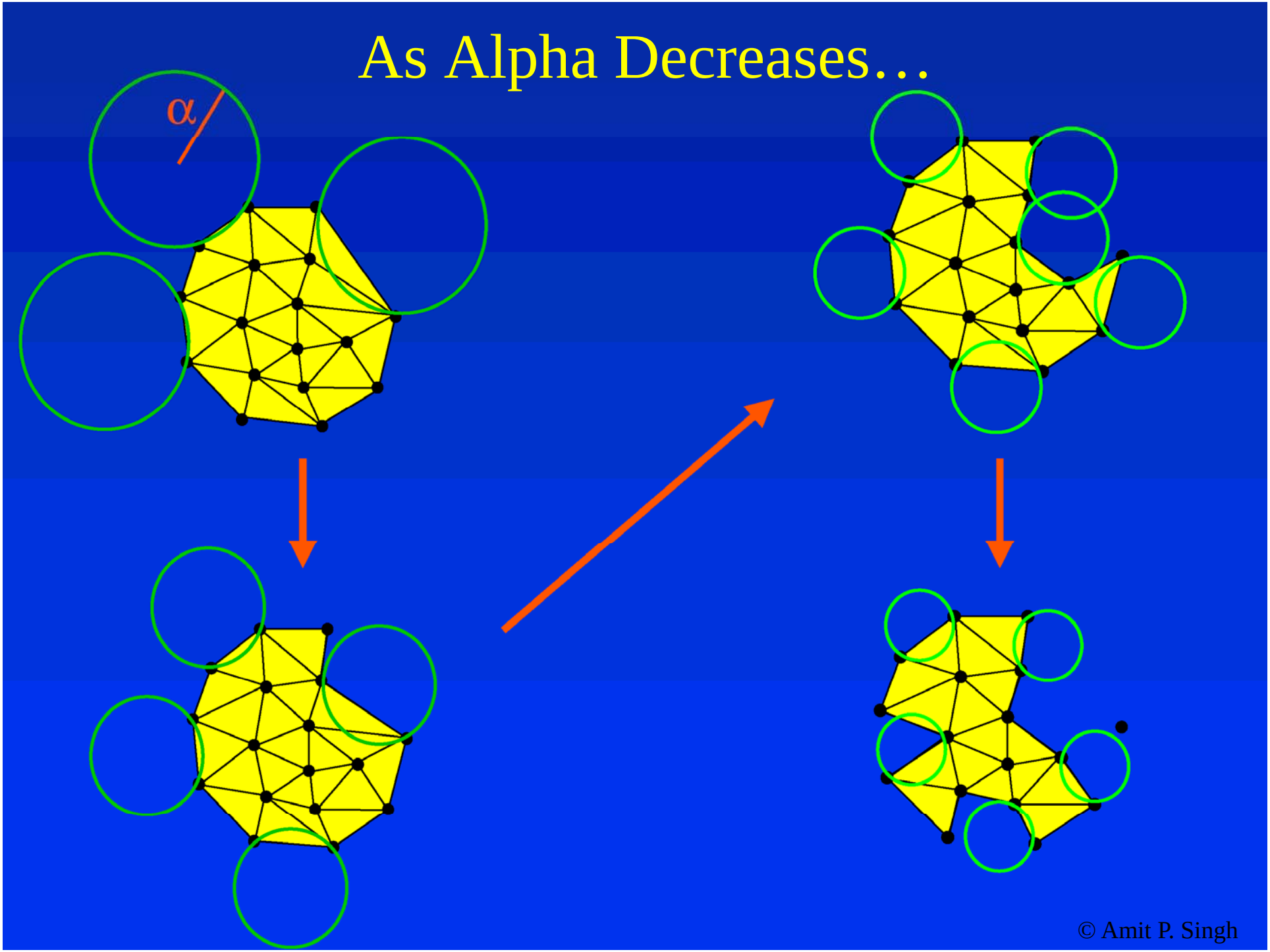


Alpha Shapes

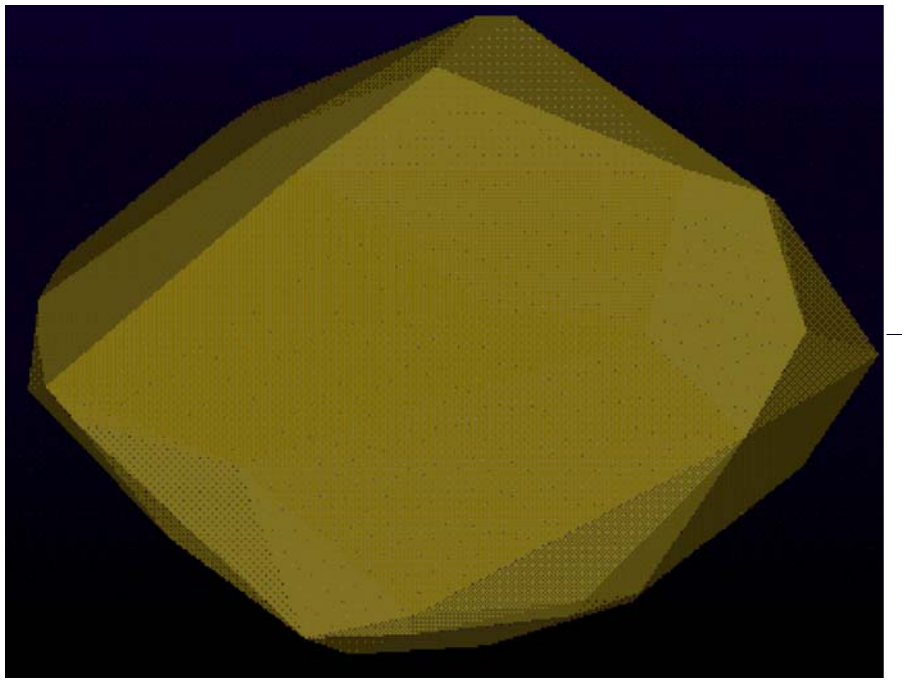
• Formalizes the idea of “shape”
• Captures the entire range of “crude” to “fine” shape representations of a point set
• In 2-dimensions:
• An edge between two points is “alpha-exposed” if there exists a circle of radius alpha such that the two points lie on the surface of the circle and the circle contains no other points from the point set.



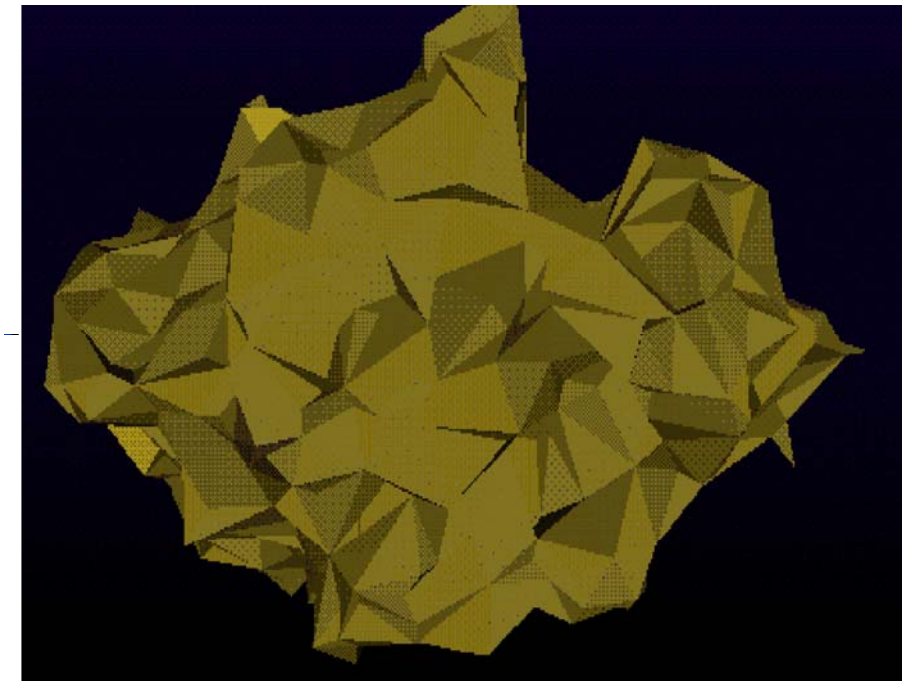
As Alpha Decreases…



Example: Trypsin

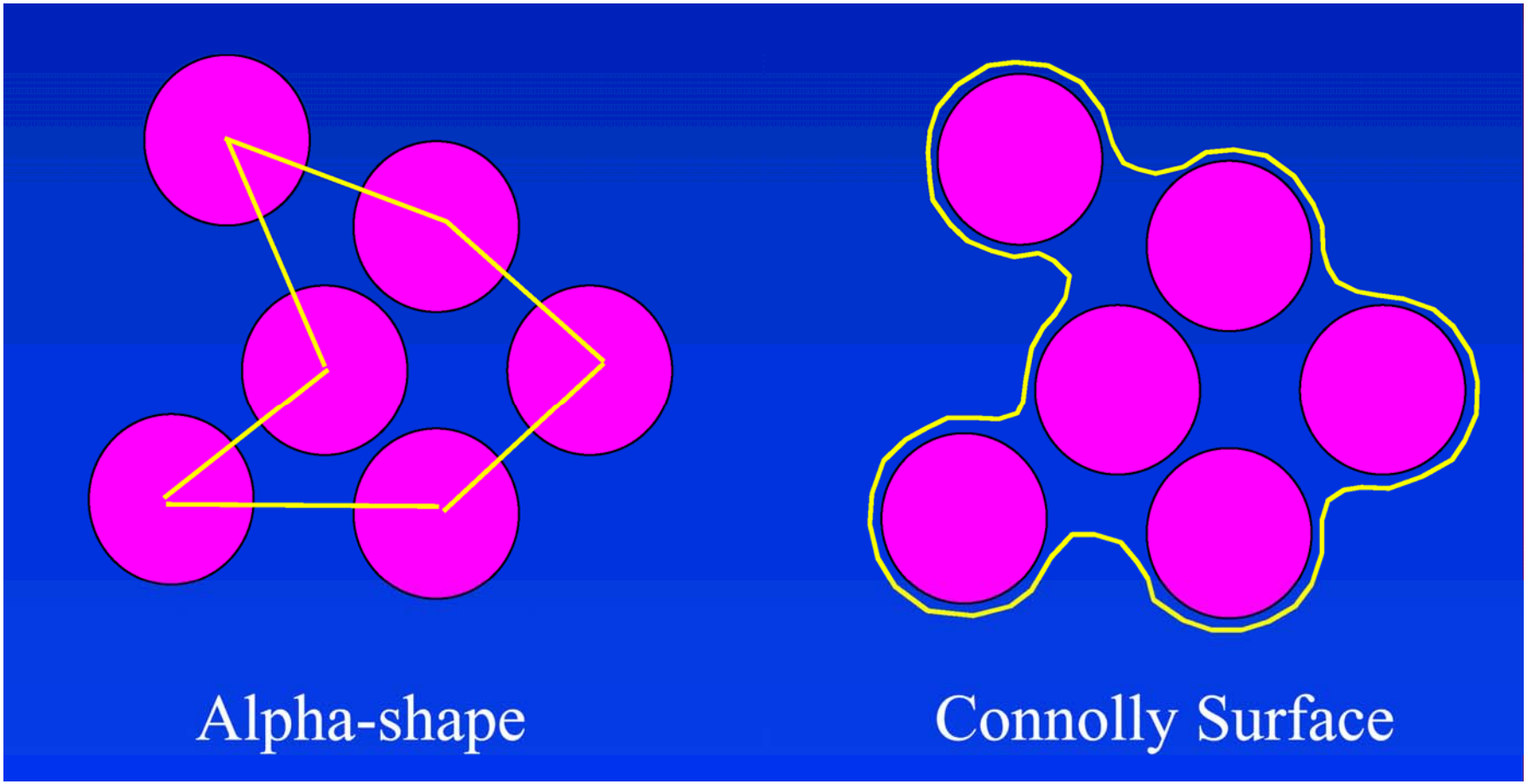


alpha = infinity
“convex hull”



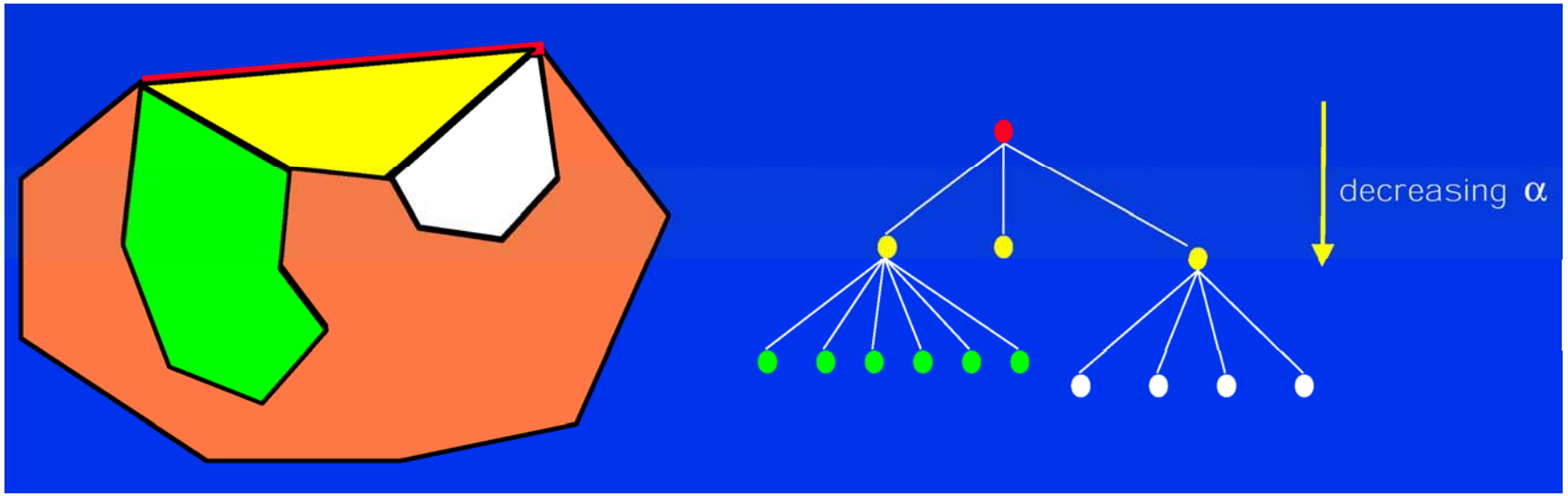
alpha = 3.0 Å

Alpha Shape vs. Connolly Surface



Identifying Cavities

• As alpha decreases, edges appear on the surface and then disappear (as alpha gets even smaller)
• We can compute a hierarchy of cavities by following edges as the appear and then disappear



Surface Matching

First satisfy steric constraints…

- Find the best fit of the receptor and ligand using only geometrical constraints
- Compute scores based on RMSD (or number of contact points) instead of MM energy

…then use energy calculations to refine the docking

- Compute the energy of interaction for each geometrically feasible docking pattern
- Select the fit that has the minimum energy

Surface Matching

The Problem:

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

A Solution: Geometric Hashing

• Compute all possible triangles formed by selecting triplets of atoms from the ligand and from the receptor

• Compare all receptor triangles to all ligand triangles using a hash table

• Use the set of triangles with the maximum number of matches to find the transformation matrix

• Advantage: local coordinate system frames independent of rotation

Geometric Hashing

Building the table:

• For each triplet of points from the ligand, generate a unique coordinate system

• Record the position and orientation of all remaining points in this coordinate system in an index table

Searching the table:

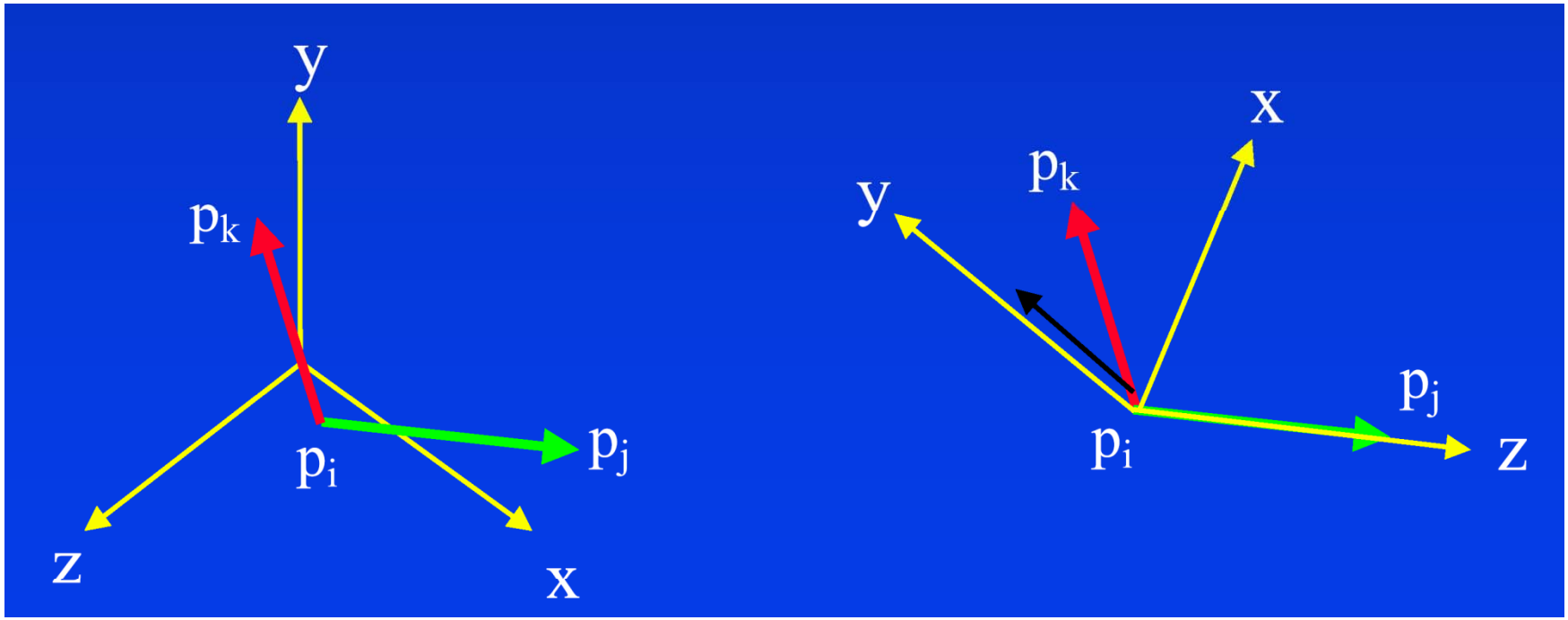
• For each triplet of points from the receptor, generate a unique coordinate system

• Search the table of ligand points to find the receptor coordinate system that results in the maximum number of similar points

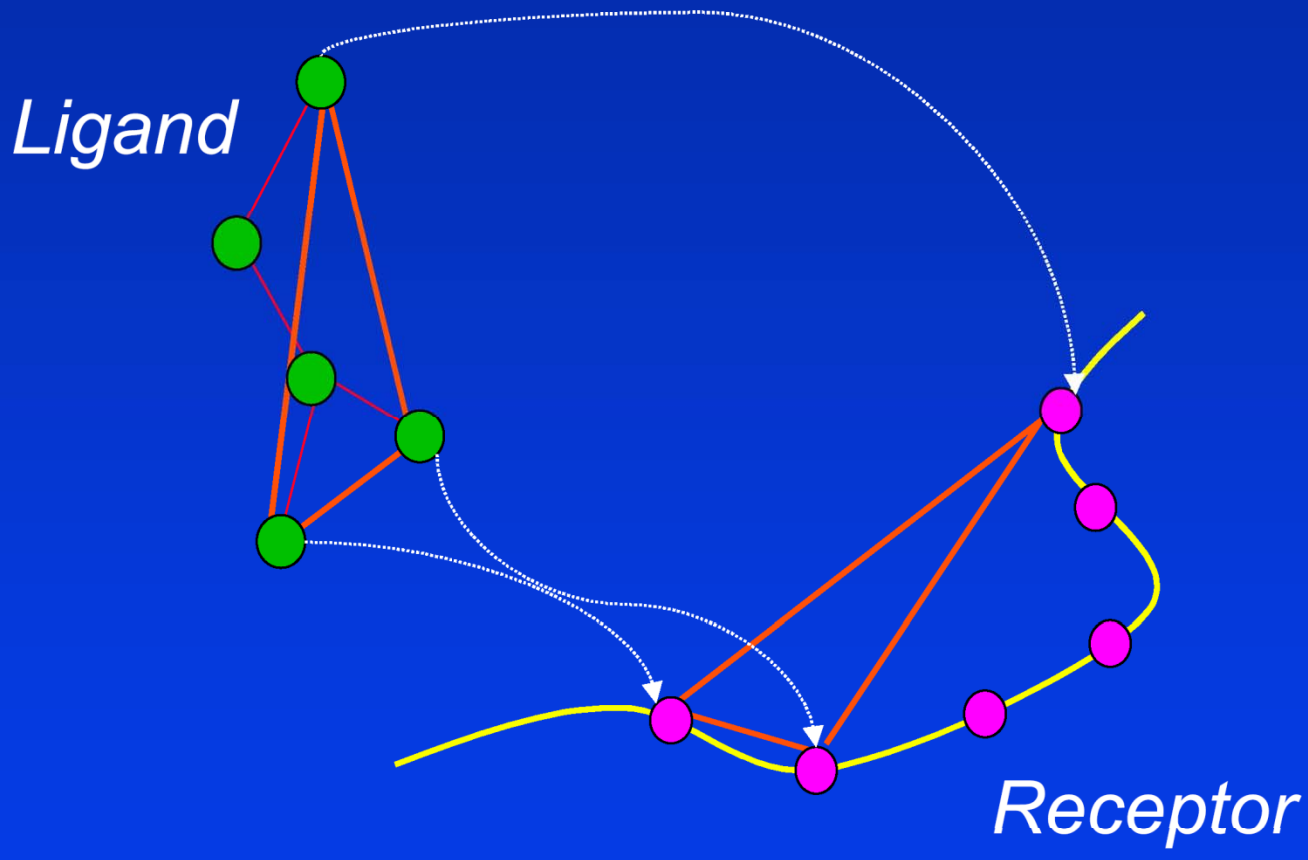
Generating a Coordinate System

- For each triplet of points (p_i,p_j,p_k):
- Transform the coordinates such that vector(p_i p_j) lies

on the Z-axis and the projection of vector(p_j p_k) on to the X-Y plane is parallel to the Y-axis



Matching Surfaces



Stochastic Search Methods

- Simulated Annealing (SA)
- Evolutionary Algorithms (EA)
  - Genetic Algorithm (GA)
- Others
  - Tabu Search (TS)
- Hybrid Global-Local Search
  - Lamarckian GA (LGA)

How SA Works

- One copy of the ligand (population = 1)
- Starts from a random or specific postion / orientation / conformation (=state)
- Constant temperature annealing cycle (accepted & rejected moves)
- Temperature reduced before next cycle
- Stops at maximum cycles

How a GA Works

- Start with a random population (50-200)

- Perform **Crossover** (reproduction: two parents give 2 children) and **Mutation** (*cosmic rays*, one individual gives 1 mutant child)

- Compute **fitness** of each individual

- **Proportional Selection** & **Elitism**

Search Parameters

Simulated Annealing

- Initial temperature
- Temperature reduction factor
- Termination criteria:
  - accepted moves
  - rejected moves
  - cycles

GA & Lamarckian GA

- Population size
- Crossover rate
- Mutation rate
- Local search
  - energy evals
- Termination criteria:
  - energy evals
  - generations

Docking Programs

General principles are illustrated with examples drawn from several programs.

- –DOCK (I. D. Kuntz, UCSF)
- –AutoDock (Arthur Olson, The Scripps Research Institute)
- –DOT (Lynn Ten Eyck, San Diego Supercomputer Center)
- –ICM (Ruben Abagyan, The Scripps Research Institute)

First Example: DOCK

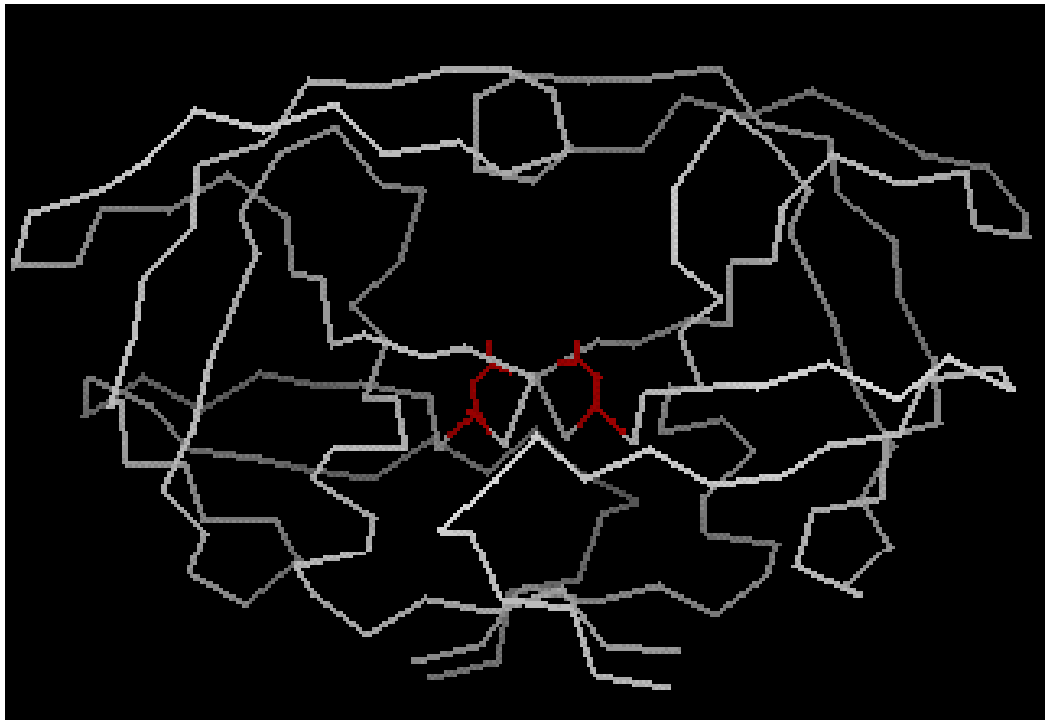
•DOCK was designed primarily to screen for possible lead compounds in the drug discovery process.

•The process of preparing your data is common to all docking problems, but differs in details.

- –Locate or build a model of the receptor
- –Locate or build a model of the ligand
- –Locate or compute all parameters required for an energy calculation or scoring function

How DOCK Works

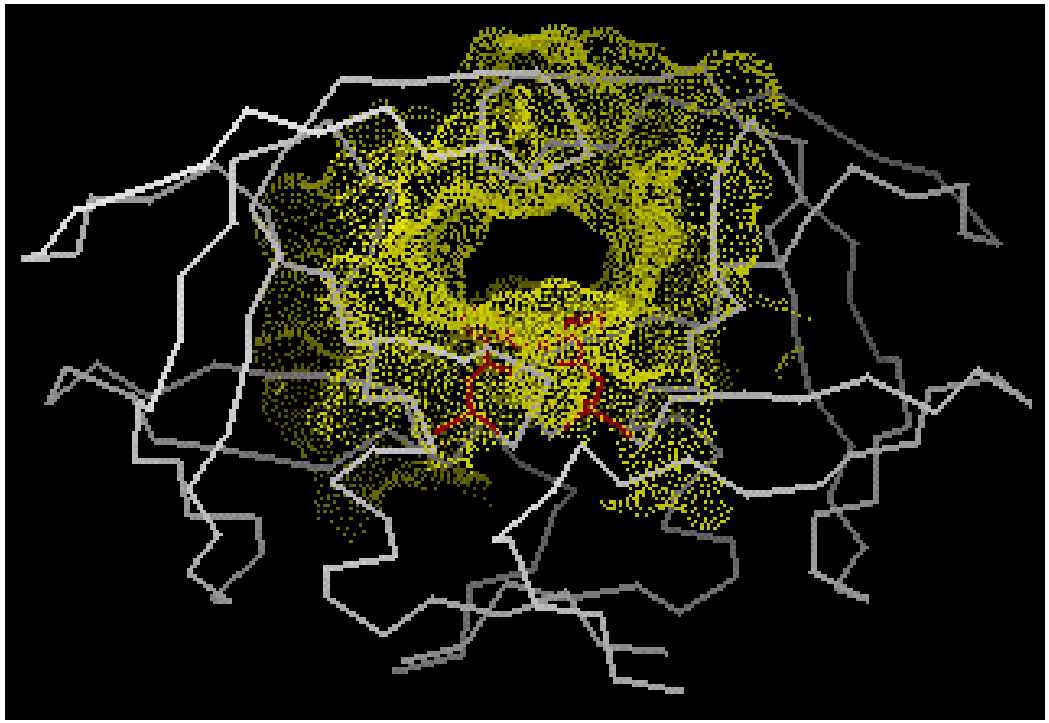
Step 1: Start with crystal coordinates of target receptor



In this example, HIV-1 protease is the target receptor, with its active site aspartyl groups identified in red.

How DOCK Works

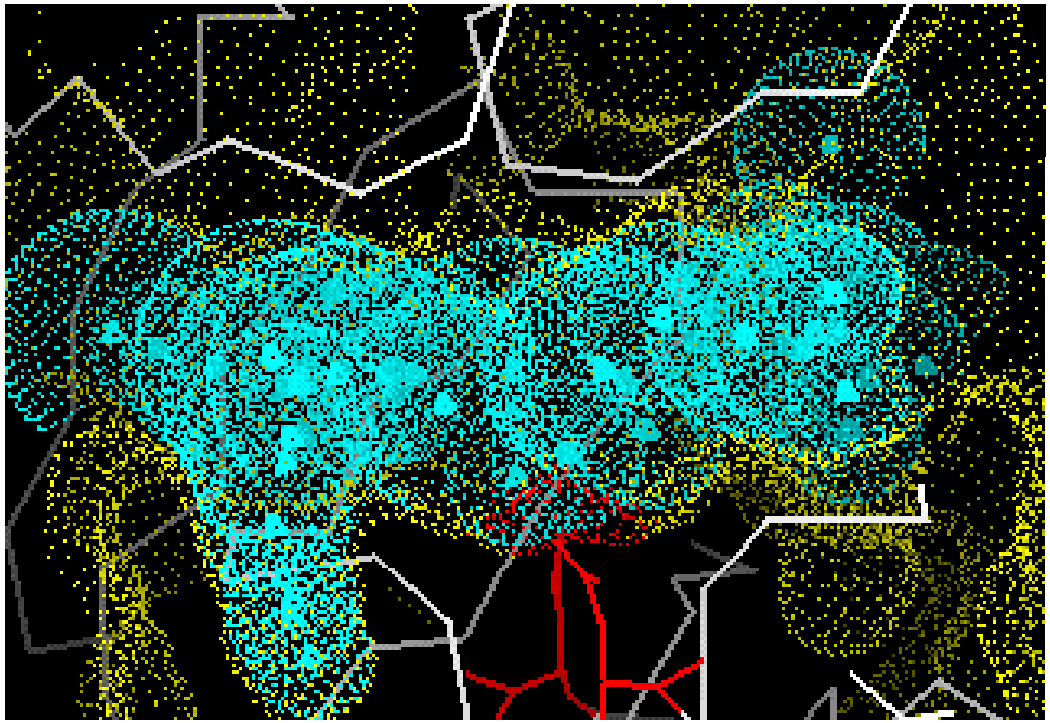
Step 2: Generate molecular surface for receptor



This is performed using Mike Connolly's ms program. Note that only the surface for the active site needs to be generated.

How DOCK Works

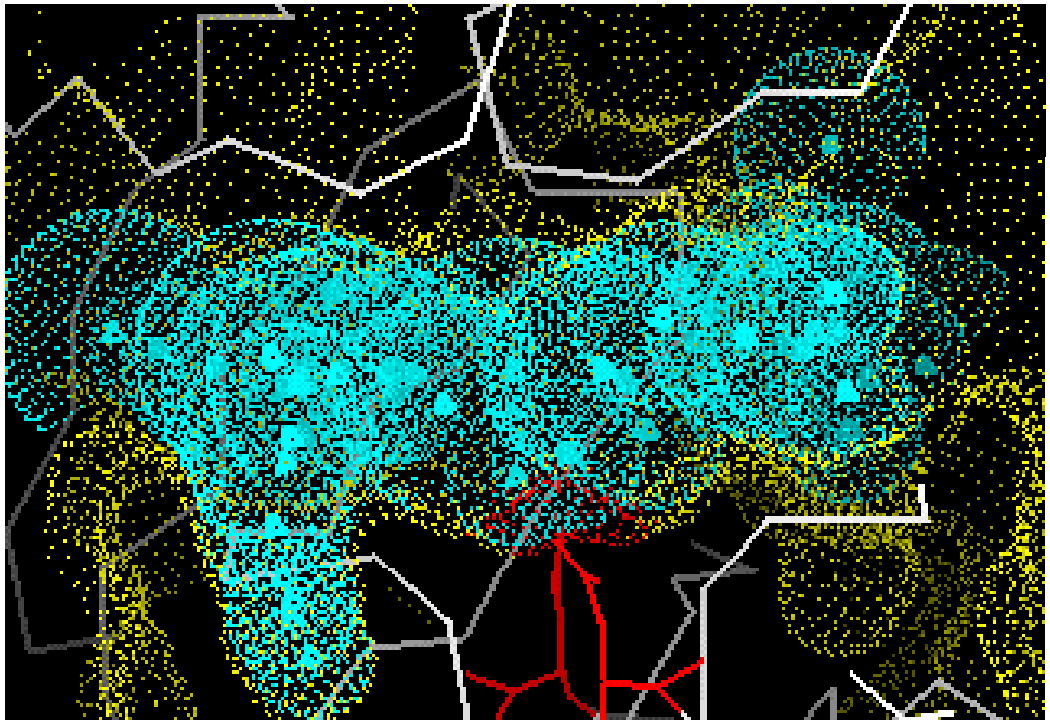
Step 3: Generate spheres to fill the active site



The shape of cavities in the receptor is used to define spheres; the centers of the spheres become potential locations for ligand atoms.

How DOCK Works

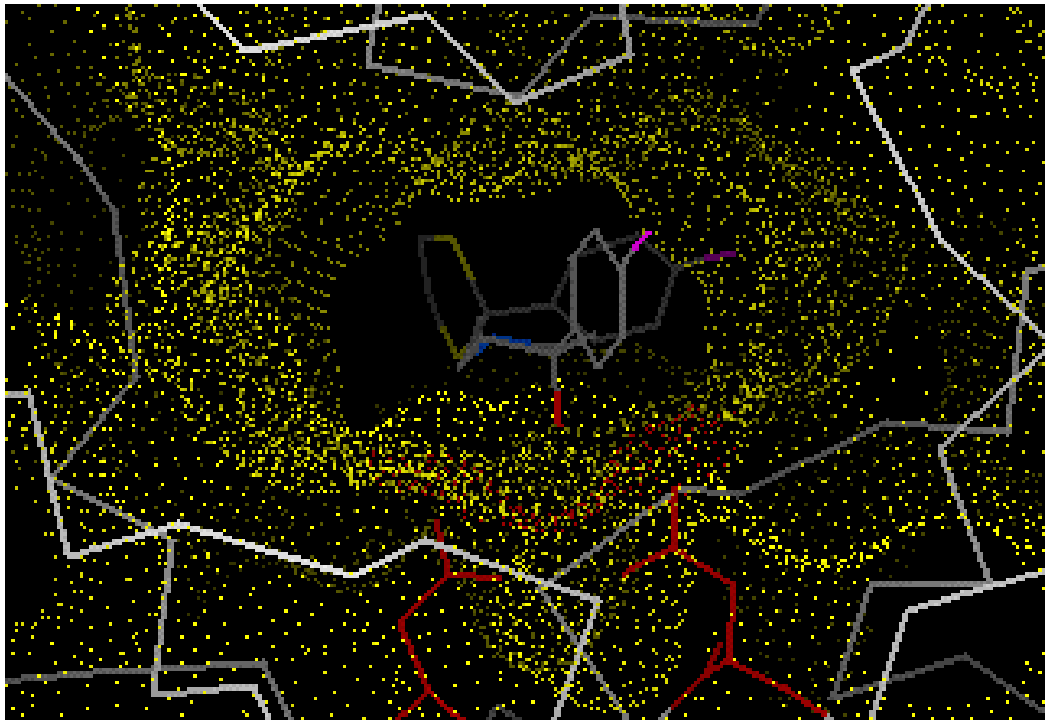
Step 4: Matching



Sphere centers are then matched to the ligand atoms, to determine possible orientations for the ligand. Typically on the order of tens of thousands of orientations are generated for each ligand molecule.

How DOCK Works

Step 5: Scoring



This is the top-scoring orientation for the molecule thioketal in the HIV1-protease active site, using force-field scoring.

There are currently 3 scoring schemes:

- **Shape scoring**, approximation to the Lennard-Jones potential
- **Electrostatic scoring**, using DELPHI to calculate electrostatic potential
- **Force-field scoring**, using the AMBER MM potential

How DOCK Works

Reality Check

Here is a comparison of the top scoring orientation of the molecule thioketal with the orientation found in the crystal structure.

Features of DOCK

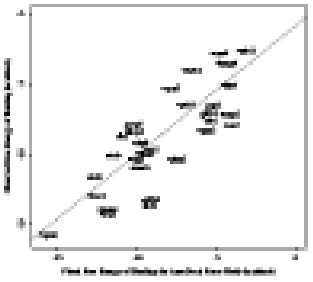
- Characterization of the binding site
  - Packing spheres to describe shape
- Geometric matching by binding site descriptors
- Direct search for optimum fit
  - Energy evaluation by table lookup on a grid
  - Search localized to binding site
- Optimization for searching compound libraries

Second Example: AutoDock

- AutoDock was designed to dock flexible ligands into receptor binding sites.
  - The AutoDock home page is at http://autodock.scripps.edu
- Essential features:
  - Energy calculation on grid
  - Global optimization by genetic algorithm or simulated annealing
  - Explicit ligand flexibility
- The strongest feature of AutoDock is the range of powerful optimization algorithms available.

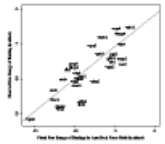
AutoDock Introduction

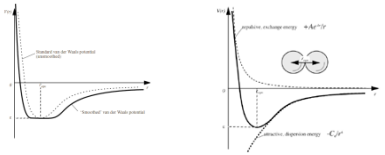
- Automated docking of flexible ligands to proteins.
- Global search algorithms:
  - –Simulated Annealing (Goodsell *et al.* 1990)
  - –Distributed SA (Morris *et al.* 1996)
  - –Genetic Algorithm (Morris *et al.* 1998)
- Local search algorithm:
  - –Solis & Wets (Morris *et al.* 1998)
- Hybrid global-local search algorithm:
  - –Lamarckian GA (Morris *et al.* 1998)
- Empirical free energy function estimates K_i *(Std. dev. ~2 Kcal mol^{-1})*



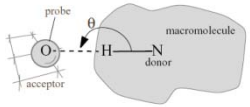
AutoDock Scoring Function

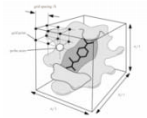
$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{elec} + \Delta G_{hbond} + \Delta G_{desolv} + \Delta G_{tors}$$

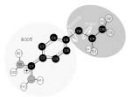




$$\varepsilon(r) = A + \frac{B}{1 + ke^{-\lambda Br}}$$







- ΔG_{vdW}

  12-6 Lennard-Jones potential

- ΔG_{elec}

  Coulombic with Solmajer-dielectric

- ΔG_{hbond}

  12-10 Potential with Goodford Directionality

- ΔG_{desolv}

  Stouten Pairwise Atomic Solvation Parameters

- ΔG_{tors}

  Number of rotatable bonds

Features of AutoDock

- What problem does AutoDock solve?
  - *Flexible* ligands (4.0 *flexible* receptor).
- What range of problems is feasible?
  - Depends on the search method:
    - LGA > GA >> SA >> LS
    - SA : trajectories, *D* = ~8 torsions.
    - LGA : *D* = ~20-30 torsions.
- When is AutoDock not suitable?
    - No 3D structures available;
    - Modelled structure of poor quality;
    - Too many (torsions, atoms, types);
    - Target protein too flexible.

Third Example: DOT

- The basic paper on DOT can be found at
  http://www.sdsc.edu/CCMS/Papers/DOT_sc95.html.

- Essential features of DOT:
  - Rigid body docking
  - "Near" Poisson-Boltzmann electrostatic energy model
  - Simplified contact potentials
  - Exhaustive search (Fourier-accelerated)

Electrostatic Energy as a Correlation

Given a charge distribution $Q(\mathbf{r})$ in a potential field $V(\mathbf{r})$, the electrostatic energy is given by

$$E = \int V(\mathbf{r}) Q(\mathbf{r}) d\mathbf{r}$$

If the charge distribution is rotated by an angle θ and translated to a position $\mathbf{r_0}$,

$$E_\Theta(\mathbf{r}_0) = \int V(\mathbf{r}) Q_\Theta(\mathbf{r} - \mathbf{r}_0) d\mathbf{r}.$$

Collisions as Correlations

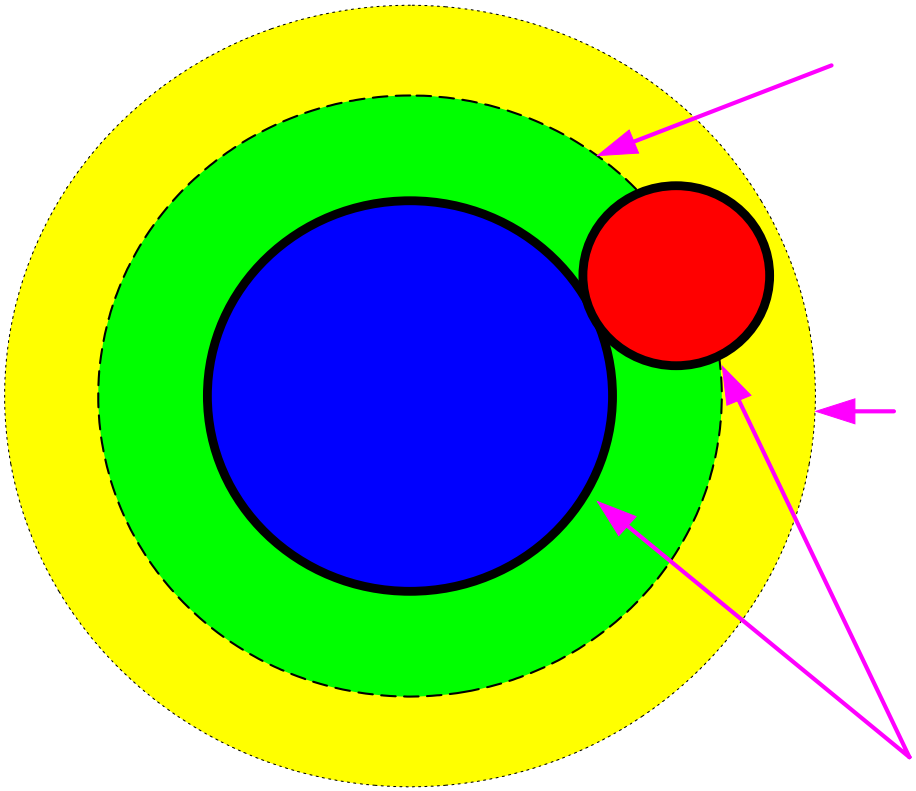
Define a fixed mask such that

$$G(\mathbf{r}) = \begin{cases} M \text{ if } \mathbf{r} \text{ is inside molecule} & 1 \\ 1 \quad \text{if } \mathbf{r} \text{ is in the surface layer} & \\ 0 \quad \text{otherwise} & \end{cases}$$

Given a set of points $A_\theta(\mathbf{r} - \mathbf{r_0})$, the overlap is given by

$$F_\Theta(\mathbf{r}_0) = \int G(\mathbf{r}) A_\Theta(\mathbf{r} - \mathbf{r}_0) d\mathbf{r}$$
$$= jM + k$$

The DOT Stereochemical Model

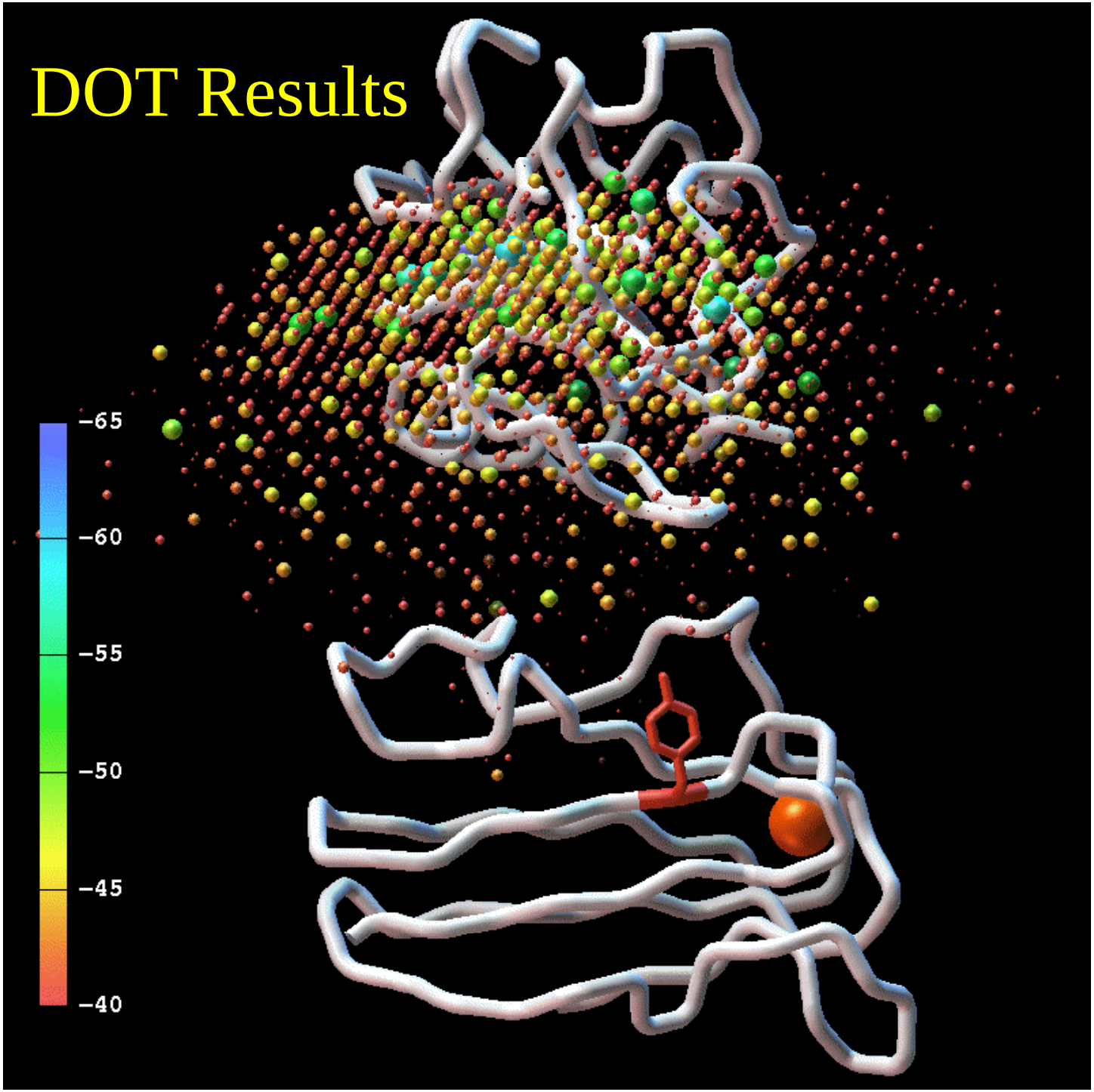


Solvent Excluded Volume

“Close” Volume

Molecular Volumes

DOT Results



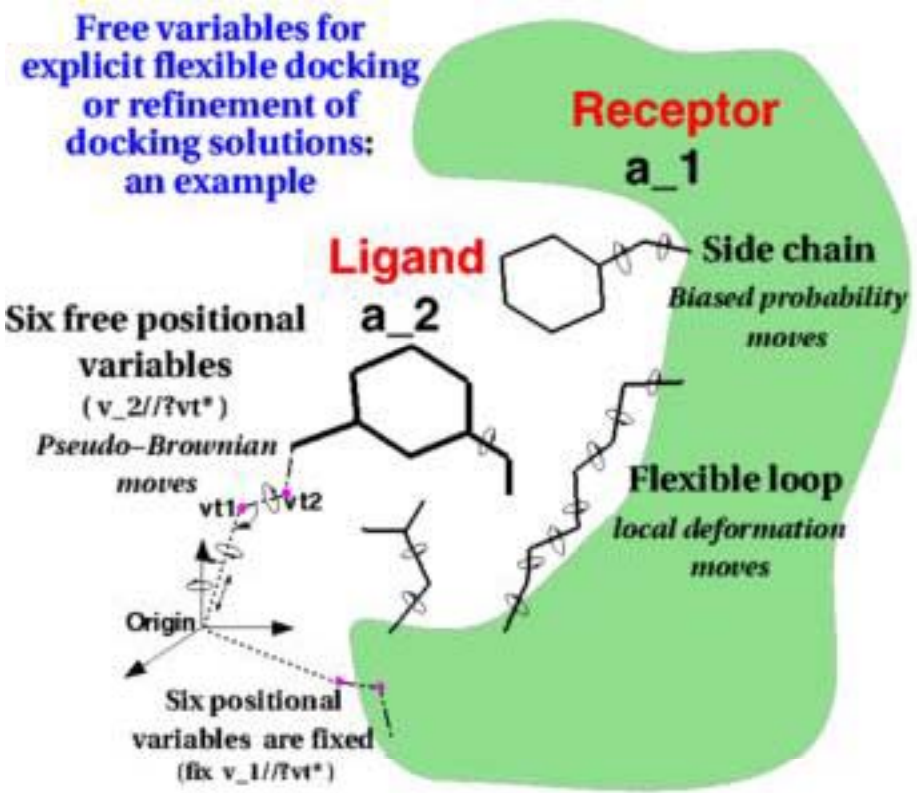
This figure shows the full backbone representation of the cytochrome c molecule in the configuration found as the second most favorable energy. DOT retains the rotation of the moving molecule, as well as the energy, so that the complete molecular complex can be rapidly regenerated for viewing and analysis.

Fourth Example: ICM

Internal Coordinate Mechanics

- Full atom, internal coordinates
- Gradient local minimization after random moves
- Optimally biased, designed, continuous group moves:

- Double energy scheme
- Reactive history mechanism, stack
  Not simulated annealing (T=const), *Not* Monte Carlo (RHM, no local balance)



ICM References:

Abagyan et al. (1994) "ICM - a new method for protein modeling.." J. Comp. Chem. 15, 488-506

Abagyan, and Totrov, (1994)."Biased Probability Monte Carlo searches …" J. Mol. Biol. 235, 983-1002

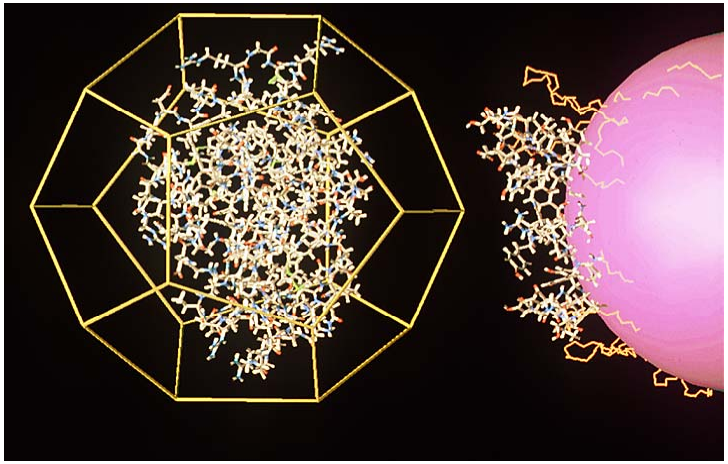
ICM Protein-Protein Docking

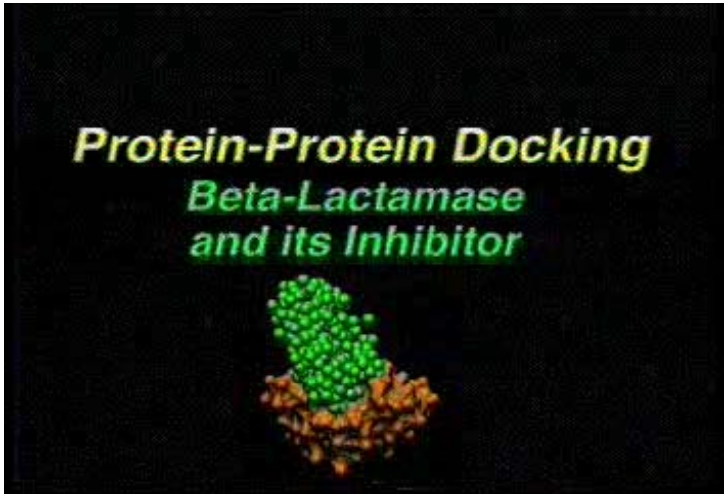
Procedure:
Multistart Global Stochastic Free-energy optimization with pseudo-Brownian moves and Biased Probability Monte Carlo (JCC, 1994).

Explicit All Atom Flexible docking and refinement
Lysozyme-Antibody (Nature SB, 1994)
Docking challenge (Nature SB 1995,96)

Grid Docking and refinement
24 known protein-protein complexes
(Protein Sci. 2002)

Global Grid Docking and refinement
CAPRI docking competition

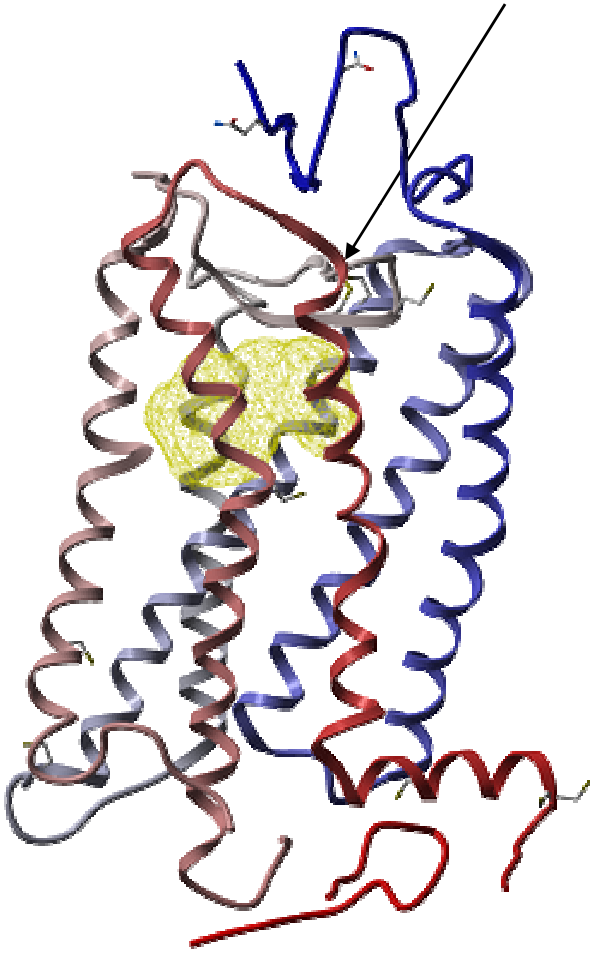




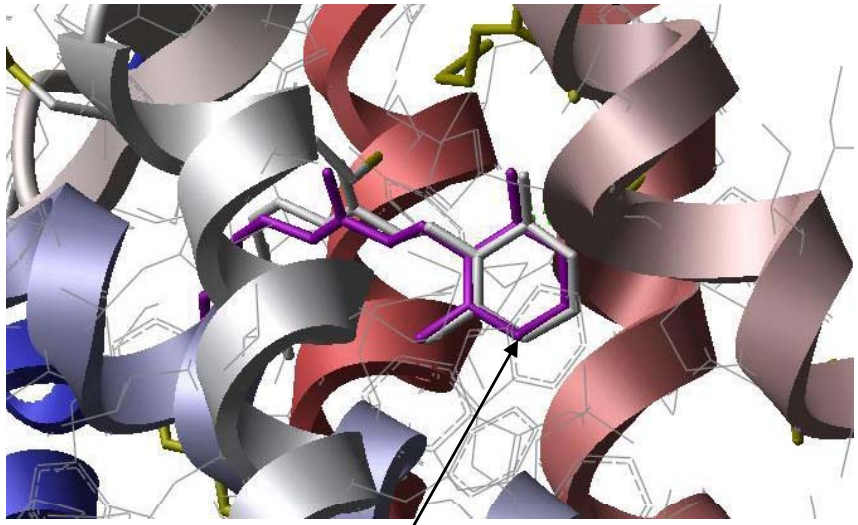
Success Story

G-protein coupled receptors

ligand binding pocket identified



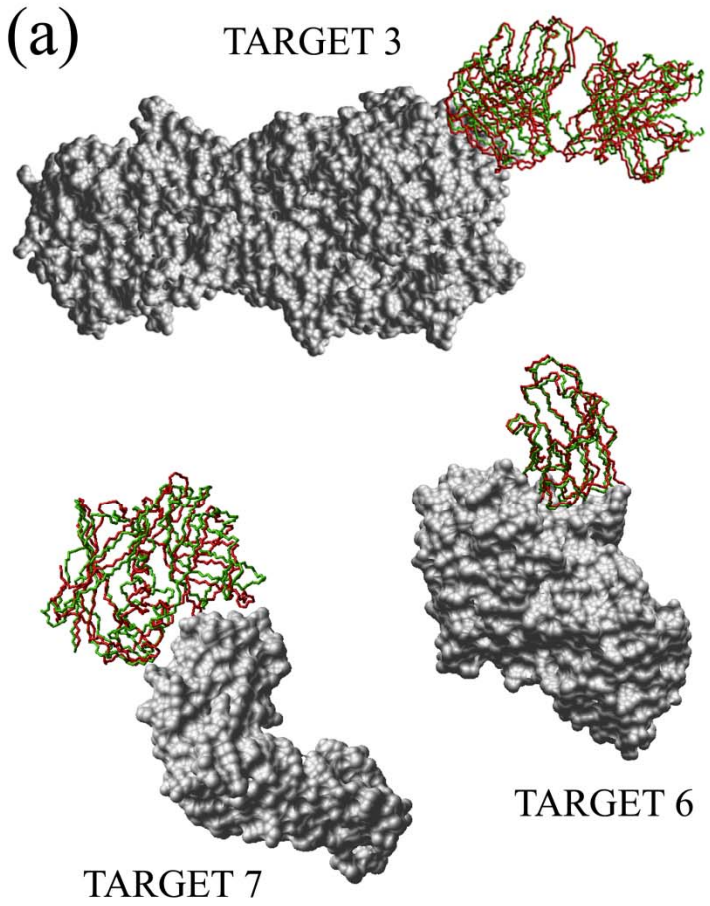
Cavasotto, Orry et al., Proteins, 2003



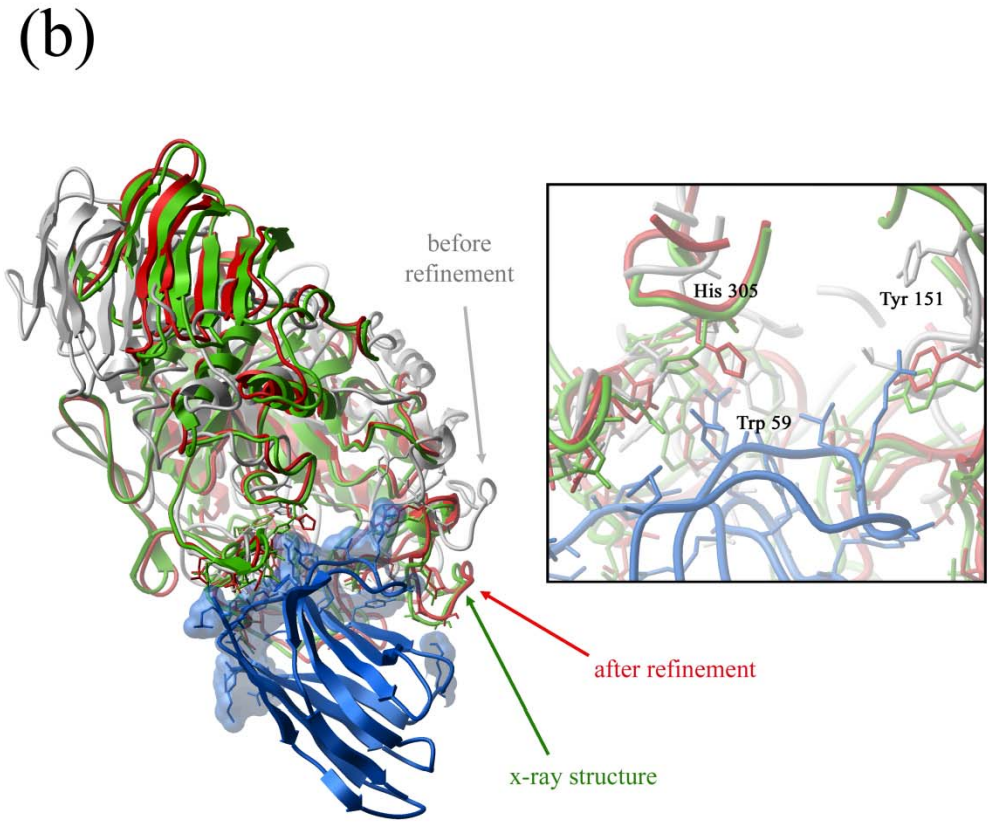
flexible docking accurately predicts the ligand binding pose

CAPRI: The Worldwide Docking Competition

CAPRI 1,2 (2002)



Abagyan Group predictions for Three protein docking targets

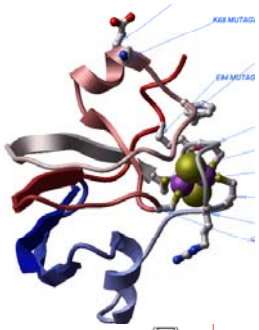


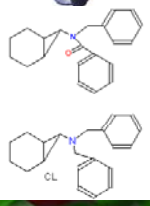
Refinement of Target 6
Dramatically improves the near-native solution

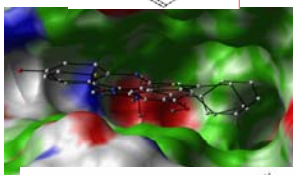
Mendez et al., Proteins 52, *51-53,* ***2003***

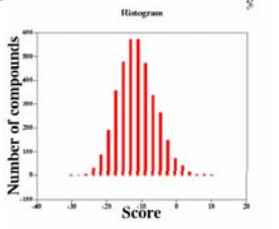
Summary: Virtual Ligand Screening

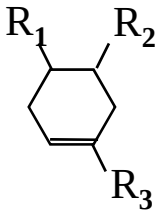
Abagyan, Totrov 2001 Current Opinion in Structural Biology











	Procedure	***Goal***	*Alternatives*	*Pitfalls*
1	**Receptor Modeling**	Correct receptor pocket model(s)	Sources: X-ray, NMR, or homology modeling. Apo-form or liganded-form. Alternative conformations predicted by simulations	Receptor model does not reflect the induced fit. Alternative conformations are missed
2	**Library Generation**	Sufficiently large and diverse set of relevant compounds	In-house collection, HTS hits, commercially available compounds, virtual libraries computed from accessible scaffolds and side-chains	The library is too restricted, molecules are not chemically feasible or not drug-like
3	**Flexible Docking**	Correct prediction of the binding geometry	MC or GA, Stochastic global optimization with gradient minimization, Incremental construction, grid or explicit receptor representations, etc.	Inaccurate energy function, poor optimization algorithm. Wrong receptor model, inadequate ligand flexibility.
4	**Ligand Scoring**	Maximal separation between binders and non-binders	Weighted interaction terms, Statistical potentials, combination of binding score with QSAR if binders are known	Poorly predicted binding geometries, score over-training to a particular case/family, large number of false positives.
5	**Hit List Post-Processing**	The best task for the chemist, screener or compound vendor	Clustering, diversity, selection of scaffolds and/or side-chains for a small combinatorial library of parallel synthesis	Domination of one chemical family, lack of chemical availability, or ADME-tox and patent considerations.

Points 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?

- Shape matching is an intuitive notion that does not have a formal definition based on first principles of physics. Discuss how shape complementarity relates to van-der-Waals interaction and steric repulsion. Do you think we should define shape complementarity in these terms?

- Why is it difficult to add flexibility of the receptor to any of these programs?

Resources and Further Reading

WWW:

- –Alpha shapes: http://biogeometry.duke.edu/software/alphashapes/
- –DOCK: http://dock.compbio.ucsf.edu/
- –AutoDock: http://autodock.scripps.edu
- –DOT: http://www.sdsc.edu/CCMS/Papers/DOT_sc95.html
- –Abagyan Lab: http://abagyan.scripps.edu

Textbooks:
Bourne & Weissig, Chapter 22, 23

Acknowledgement:
Ruben Abagyan, The Scripps Research Institute
Julie Mitchell, University of Wisconsin, Madison
Lynn Ten Eyck, UCSD
G. Matthias Ullmann, Uni Heidelberg
Amit P. Singh, Stanford U.
A. H. Juffer, University of Oulu, Finland