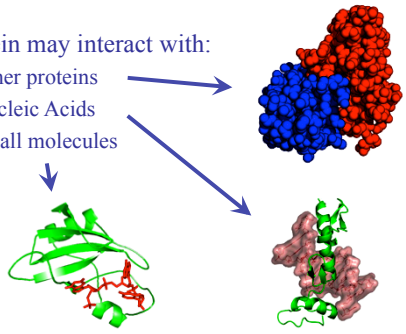
Protein-Protein Interactions

Protein Interactions

A Protein may interact with:

- Other proteins
- Nucleic Acids
- Small molecules



Protein-Protein Interactions: The "Interactome"

Experimental methods:

Mass Spec, yeast 2-hybrid system, microarrays,…

Computational techniques:

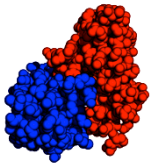
phylogenic profiles, sequence analysis,…

2 challenges:

- find which proteins interact (the partners)

- find which residues participate in the interactions

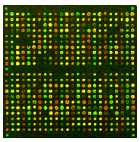
Finding Protein Partners

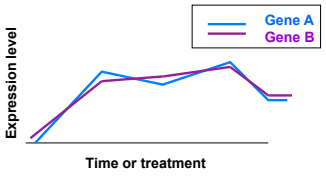


Experimental technique: co-expression

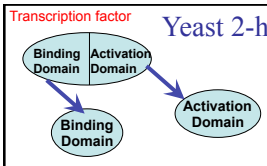
Microarray: study the expression of genes as a a function of time, or following treatment with a drug, …

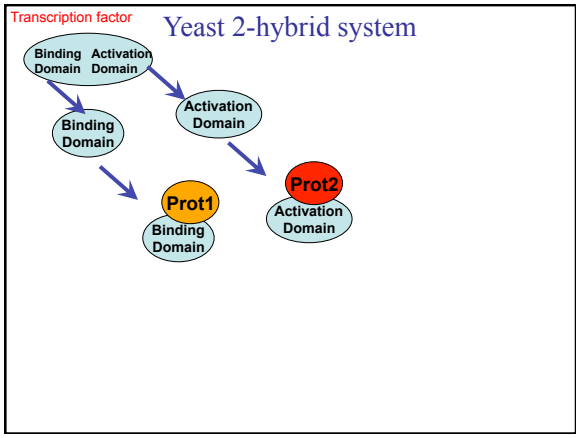
Co-expression of genes are usually a sign that the two proteins interact.

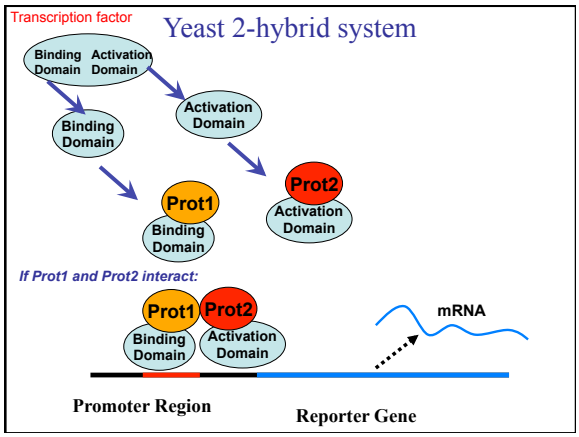




Yeast 2-hybrid system







Yeast 2-hybrid system

In words:

- A transcription factor is split into 2 domains
- 2 hybrid proteins are designed, each containing one of the two proteins that are tested
- If the two proteins interact, the two domains from the transcription factor will interact, causing expression of a (detectable) reporter gene
- The reporter can be:
  - essential, in which case the yeast colony dies if the 2 proteins do not interact
  - reversely, the reporter gene can be attached to a green fluorescent protein

Unfortunately, the rate of false positive is high (estimated > 45%)

Protein Interactions

- ***Databases of experimental protein interaction data***
  - MIPS: http://mips.helmholtz-muenchen.de/proj/yeast/CYGD/interaction/
    (protein-protein interactions in saccharomyces cerevisae)
  - InterAct: http://www.ebi.ac.uk/intact/index.html
    (protein interactions from literature curation)
  - DIP: http://dip.doe-mbi.ucla.edu/

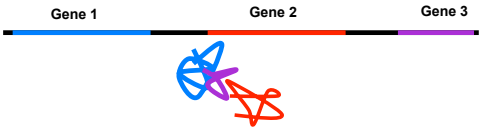
Predicting Protein-Protein Interactions

Genome-based approach

(1) Proximity of genes on chromosome

Genes that appear near each other on a chromosome are often expressed together.
They may interact (need confirmation from biology, or annotation)

Example: *operons*



Predicting Protein-Protein Interactions

Genome-based approach

(2) Homology

If A and B interact and C is homologous to A and D is homologous to B
Then do C and D interact?

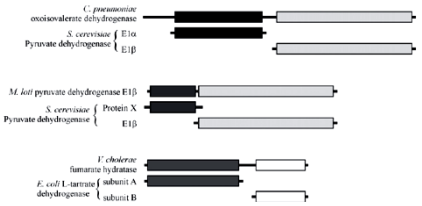
They may, if

- C&D are from the same species (unless host-pathogen interaction)
- The binding surface is conserved
- The binding residues are conserved

Predicting Protein-Protein Interactions

Genome-based approach

(2) Gene fusion (Rosetta Stone Protein)

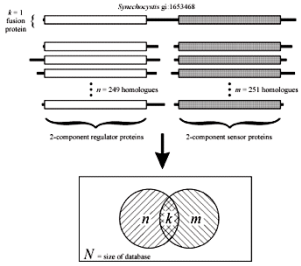


(Marcotte and Marcotte, 2002)

Predicting Protein-Protein Interactions

Genome-based approach

(2) Gene fusion: significance



n: number of A homologues
m: number of B homologues
k: number of observed fusion
N: size of the database
p: probability of fusion

Hypergeometric distribution:

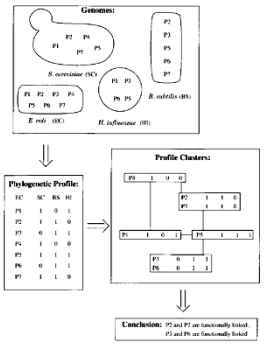
$$p(k \mid n,m,N) = \frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$$

(Marcotte and Marcotte, 2002)

Predicting Protein-Protein Interactions

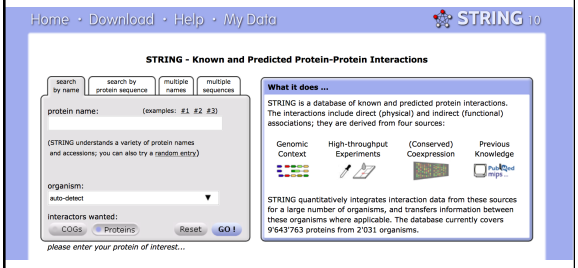
Genome-based approach

(2) Phylogenetic profiles



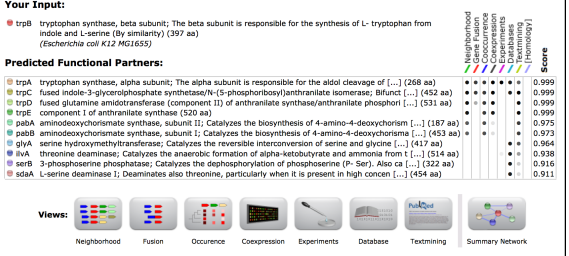
(Pellegrini et al, PNAS, 1999)

A web portal to Protein-Protein interactions

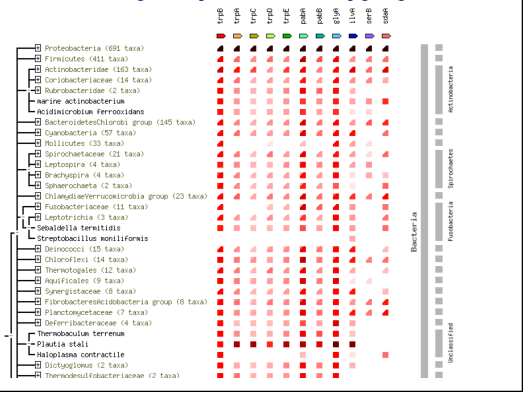


http://string.embl.de

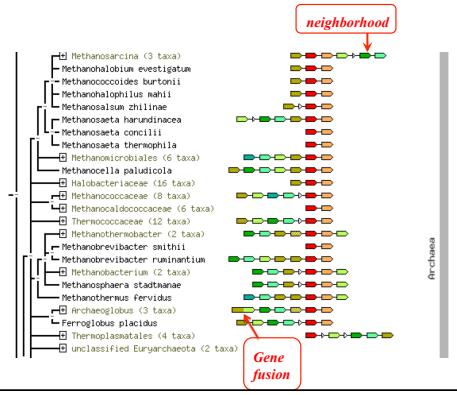
PPI: example: synthesis of tryptophan



PPI: example: synthesis of tryptophan

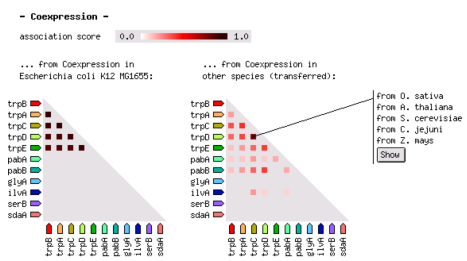


PPI: example: synthesis of tryptophan

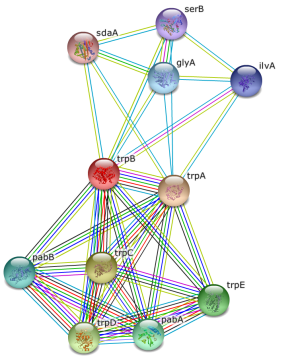


PPI: example: synthesis of tryptophan

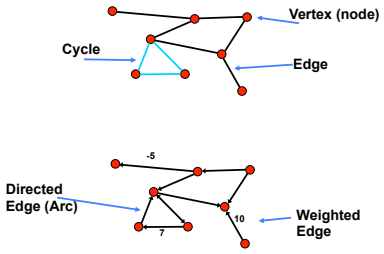
Co-expression



PPI: example: synthesis of tryptophan

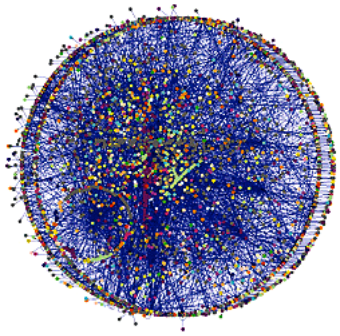


Graph Theory



Molecular interaction networks are mapped as graphs

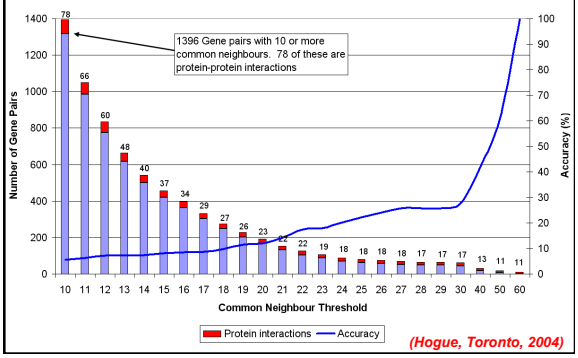
The yeast interaction network…



Useful Operation on Graphs

- A graph can be treated as a set of vertices and edges; we can then compute intersection, difference, union…
  - What is the intersection of my interaction set with all known published interactions?
- Filtering
  - Find all protein interactions where at least one partner is localized in the nucleus
- Overall statistics
  - Find the average number of interactions for cell cycle proteins

Predicting protein interactions from networks



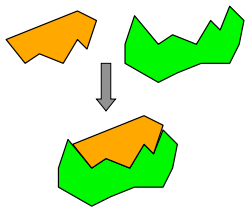
(Hogue, Toronto, 2004)

Summary

- To understand the function of a protein, we need to find its interacting partners
- Experimental methods for protein-protein interactions: Mass spec, 2-hybrid yeast system, structural study
- Predicting protein-protein interactions: sequence analysis (neighborhood, gene fusion, co-occurrence), text mining,…
- Graph theory to mine protein interaction networks

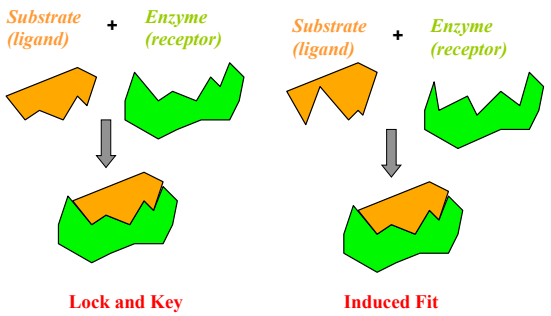
Enzyme – substrate binding: 2 models

Substrate (ligand) + *Enzyme (receptor)*



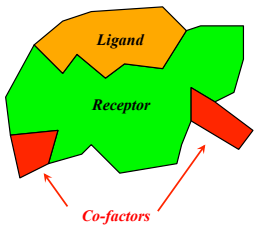
Lock and Key

Enzyme – substrate binding: 2 models



One receptor, several ligands…

The enzyme (receptor) may bind several ligands in different places:

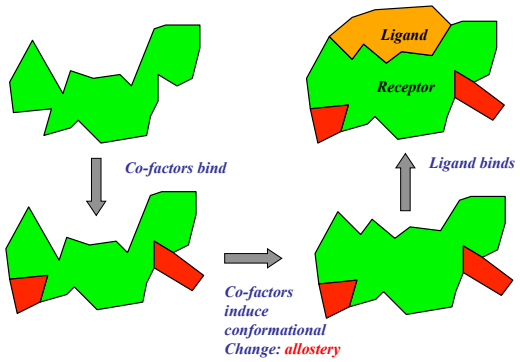


- If the protein binds to several ligands, and the affinity for binding ligands 2,3,… changes after the first ligand is bound, the binding is said to be **cooperative**:

- **Positive cooperativity**: the affinity increases
- **Negative cooperativity**: the affinity decreases

- If there is no change, the binding is said to be non cooperative

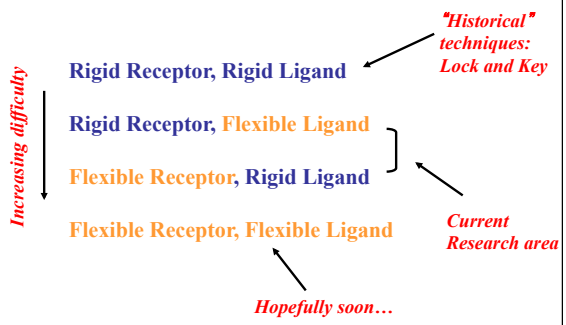
Co-factors may induce the fit: allostery



Predicting binding

- Computationally, Lock and Key is the simplest case to predict:
  - Little or no *flexibility* need be modeled
  - 6 degrees of freedom (DOF)
- Induced fit is much more difficult
  - >> 6 degrees of freedom (3 rotations, 3 translations)
  - Algorithms may need to model the movements of
    - *Side chains and backbone of the receptor*
    - *Ligands*

"Docking" scenarios



What is docking?

Docking is finding the binding geometry of two interacting molecules with known structures

The two molecules ("Receptor" and "Ligand") can be:

- **two proteins**
- **a protein and a drug**
- **a nucleic acid and a drug**

Two types of docking:

- **local docking**: the binding site in the receptor is known, and docking refers to finding the position of the ligand in that binding site
- **global docking**: the binding site is unknown. The search for the binding site and the position of the ligand in the binding site can then be performed sequentially or simulaneously

What is docking?

Some more definitions:

Two types of docking:

- **rigid docking**: both the receptor and ligand are kept rigid.

DOF = 6 (3 position + 3 orientation)

- **flexible docking**: flexibility is allowed for the receptor, or the ligand, or both

DOF = 6 + Nfree (3 position + 3 orientation + Nfree bonds)

What is docking?

Two types of docking:

- **bound docking**: the goal is to reproduce a known complex, where the starting structures for the receptor and ligand are taken from the structure of the complex
  (testing docking method)

- **unbound docking:** the structures of the receptor and ligand are taken from data on the unbound molecules
  (actual docking)

Docking Scoring Criteria

- ***Geometric match:***
  - Prevent overlap between atoms of the receptor and ligand
  - Maximum shape compatibility
  - Large surface burial
  - No large cavity at interface
- ***Energetic Match (Force-field + Statistical potential)***
  - Good hydrogen bonding
  - Good charge complementarity
  - Polar/polar contacts favored, polar / non polar contacts disfavoured
  - Low "free energy"

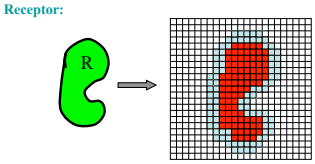
Docking Search Strategies

- Full search
  - Grid approaches (FFT…)
- Directed search
  - Spherical harmonics surface triangle
  - Geometric hashing
- Pseudo Random
  - Simulated annealing / Monte Carlo
  - Genetic algorithms

Global Rigid Docking: a FFT approach

1. Representation:

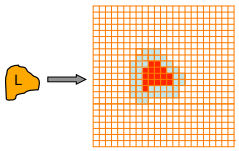
Receptor:



Assign value to each cell:

- Exterior: a(i,j) = 0
- Surface: a(i,j) = +1
- Interior: a(i,j) = -15

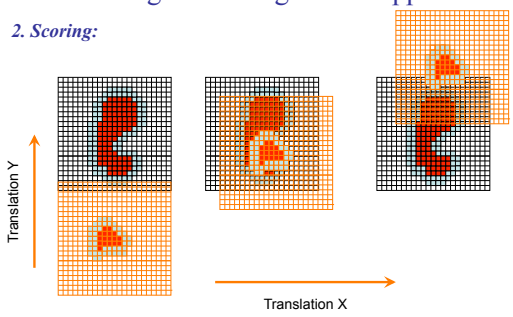
Ligand:



- Exterior: b(i,j) = 0
- Surface: b(i,j) = +1
- Interior: b(i,j) = -15

Global Rigid Docking: a FFT approach

2. Scoring:



$$Score = \sum_i \sum_j a(i,j)b'(i,j)$$

where b' is the grid for the ligand after rotation and translation

Global Rigid Docking: a FFT approach

2. Scoring:

Test all possible positions of ligand on receptor:
- *Test all rotations of ligand*
- *For each rotation, test all translations of ligand grid over receptor grid*

*Score(i,j) = Receptor(i,j)*Ligand(i,j)*

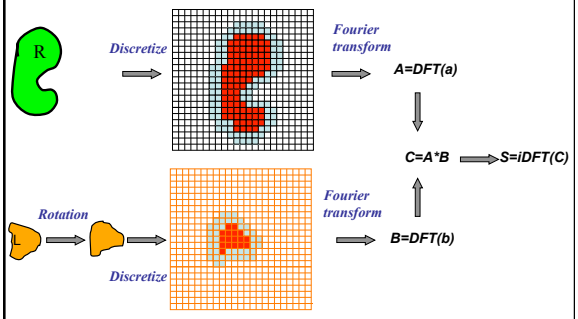
Rotation R; Translation: $T = T_x + T_y + T_z$:

$$S(R,T) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a(i,j,k)b'\left(i+T_x, j+T_y, k+T_k\right)$$

For each R, this requires N^6 operations…

But, for a given rotation, this is a correlation product, that can be computed in Fourier Space!

Global Rigid Docking: a FFT approach



Computing cost: $N^3\log(N^3)$!

DARWIN: An Example of Flexible Docking Program

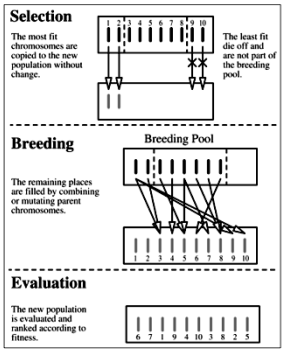
DARWIN uses a force field (CHARMM) for scoring, and a genetic algorithm for searching

Genetic algorithm:

- **Every "solution" is represented by a binary string.**
  - 3 genes describe the position (with 0.5 Å resolution)
  - 3 genes describe the orientation (11.25° resolution)
  - Each flexible bond is described by one parameter (60° resolution)
- **The population size is 100-1000 and the number of generation is 10% the population size**
- **The basic operations are:**
  - mutation (P = 0.2)
  - recombination with one cut (P = 0.4)
  - recombination with two cuts (P=0.4)
  - the "death rate" is 5% and the survival rate is 10-30%

*Taylor, J.S. and Burnett, R.M. (2000). DARWIN: A program for docking flexible molecules. Proteins **41**, 173-191*

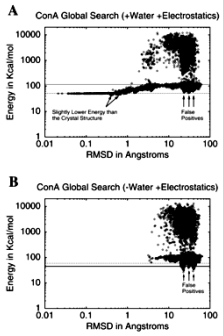
DARWIN: An Example of Flexible Docking Program



*Taylor, J.S. and Burnett, R.M. (2000). DARWIN: A program for docking flexible molecules. Proteins **41**, 173-191*

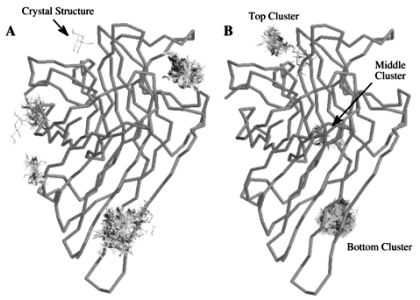
DARWIN: An Example of Flexible Docking Program

A test case: binding of Mannopyranoside in Concanavalin A (ConA)



*Taylor, J.S. and Burnett, R.M. (2000). DARWIN: A program for docking flexible molecules. Proteins **41**, 173-191*

DARWIN: An Example of Flexible Docking Program



No water in docking experiment ***With water in docking experiment***

*Taylor, J.S. and Burnett, R.M. (2000). DARWIN: A program for docking flexible molecules. Proteins **41**, 173-191*

Table 1

Programs for protein–protein docking.

Program*	Algorithm	Laboratory	URL	Details
Downloadable				
3D-Dock	Global: FFT; rescoring: residue potentials; refinement: mean-field sidechain multicopy	Imperial Cancer Research Fund/Imperial college (Sternberg)	www.bmm.icnet.uk/docking/	Free to academic; distribution mostly source code (C); refinement sgi/linux executables
HEX	Global: Fourier correlation of spherical harmonics	Aberdeen University (Ritchie)	www.biochem.abdn.ac.uk/hex/	Free to academic; sgi/sun/linux executables
GRAMM	Global: FFT clustering and rescoring decoys also available	SUNY/MUSC (Vakser)	reco3.ams.sunysb.edu/gramm/	Free to academic; sgi/sun/ibm/dec/linux/win32 executables
PPD	Global: geometric hashing; rescoring: multiple filters	Columbia (Honig)	ftp://flash62.bioc.columbia.edu/pub/other	Free to academic; sgi executables
DOT	Global: FFT for shape complementarity and approximate Poisson–Boltzmann electrostatics	University of California San Diego (Ten Eyck)	www.sdsc.edu/CCMS/DOT	Free to academic; parallelised under mpi; source (C/fortran) and sun/sgi/dec/ibm executables
BIGGER (Chemera)	Global: bit mapping; rescoring: multiple filters	Universidade Nova de Lisboa	www.dq.fct.unl.pt/bioin/chemera/www.tinkercnol.com/Paginas/Chemera.htm	Free to academic; win32 executables
MERL refinement protocol	Constrained minimisation desolvation	University of Boston (MERL) (Vajda)	engpub1.bu.edu/bioinfo/MERL/software/decoys.html	Free; Charmm input scripts, source code (f77)
DOCK	Global: grid-based energy function; flexible docking: random search plus incremental construction	University of California San Francisco (Kuntz)	www.cmpharm.ucsf.edu/kuntz/dock.html	Free to academic; sgi executables
AutoDock	Grid-based empirical potential flexible docking via Monte Carlo search and incremental construction	Scripps Institute (Olson)	www.scripps.edu/pub/olson-web/download.html	Free to academic; source code and executables for sgi/dec/sun
FlexX	Fragment assembly energy function: (Boehm potential)	GMD-SCAI (Lengauer)	cartan.gmd.de/flexx/	Licence required from Tripos.com

Contact authors

Program	Algorithm	Laboratory	Contact
DARWIN	GA Charmm force field	University of Pennsylvania	burnett@wistar.upenn.edu
ZDOCK	FFT for complementarity, electrostatics and residue potential	University of Boston (Weng)	sullivan.bu.edu/~rong/dock/download.shtml

*Programs given in italics are designed principally for protein–ligand docking, but may be usable in the protein–protein case. Programs for MD, BD and continuum electrostatics calculations are also useful.

(Smith and Sternberg, COSB, 2002)