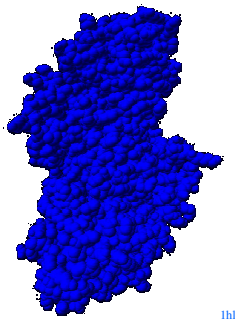
Protein-Ligand Docking Methods

Thomas Funkhouser
Princeton University
CS597A, Fall 2007

Review

Goal:
- Given a protein structure, predict its ligand bindings

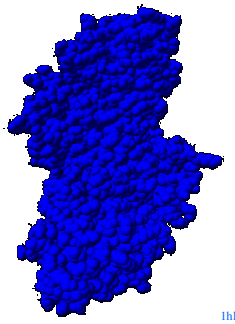
Applications:
- Function prediction
- Drug discovery
- etc.



1hld

Review

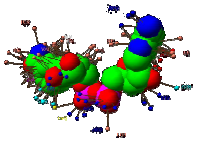
Questions:
- Where will the ligand bind?
- Which ligand will bind?
- How will the ligand bind?
- When?
- Why?
- etc.

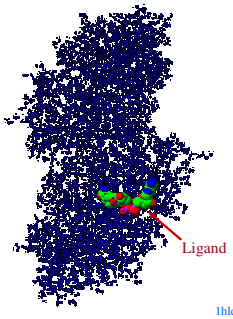


1hld

Protein-Ligand Docking

Questions:
- Where will the ligand bind?
- Which ligand will bind?
- How will the ligand bind?
- When?
- Why?
- etc.

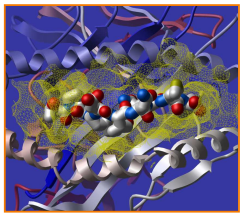




1hld

Protein-Ligand Docking

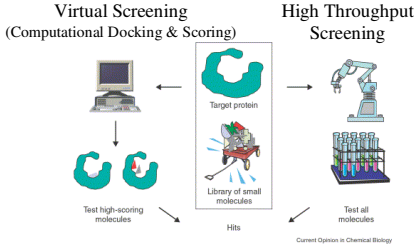
Goal 1:
- Given a protein and a ligand, determine the pose(s) and conformation(s) minimizing the total energy of the protein-ligand complex



http://www.molsoft.com/

Protein-Ligand Docking

Goal 2:
- Predict the binding affinity of any protein-ligand complex



Protein-Ligand Docking

Challenges:

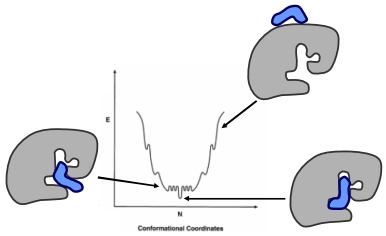
- Predicting energetics of protein-ligand binding
- Searching space of possible poses & conformations

Protein-Ligand Docking

Challenges:

Ø Predicting energetics of protein-ligand binding

- Searching space of possible poses & conformations



Protein-Ligand Docking

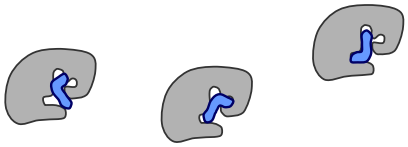
Challenges:

- Predicting energetics of protein-ligand binding
- Searching space of possible poses & conformations
  - Ø Relative position (3 degrees of freedom)

Protein-Ligand Docking

Challenges:

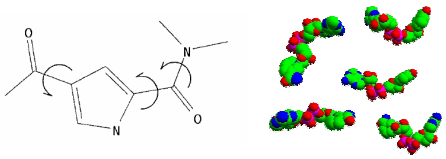
- Predicting energetics of protein-ligand binding
- Searching space of possible poses & conformations
  - § Relative position (3 degrees of freedom)
  - Ø Relative orientation (3 degrees of freedom)



Protein-Ligand Docking

Challenges:

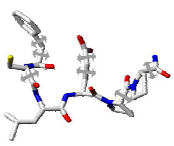
- Predicting energetics of protein-ligand binding
- Searching space of possible poses & conformations
  - § Relative position (3 degrees of freedom)
  - § Relative orientation (3 degrees of freedom)
  - Ø Rotatable bonds in ligand (n degrees of freedom)



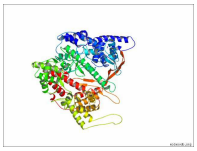
Protein-Ligand Docking

Challenges:

- Predicting energetics of protein-ligand binding
- Searching space of possible poses & conformations
  - § Relative position (3 degrees of freedom)
  - § Relative orientation (3 degrees of freedom)
  - § Rotatable bonds in ligand (n degrees of freedom)
  - Ø Rotatable bonds in protein (m degrees of freedom)



Side chain movements



Large-scale movements

Outline

Introduction

Scoring functions
- Molecular mechanics
- Empirical functions
- Knowledge-based

Searching poses & conformations
- Systematic search
- Molecular dynamics
- Simulated annealing
- Genetic algorithms
- Incremental construction
- Rotamer libraries

Results & Discussion

Outline

Introduction

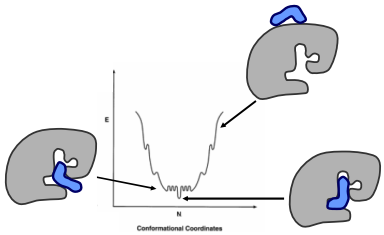
Scoring functions ←
- Molecular mechanics
- Empirical functions
- Knowledge-based

Searching poses & conformations
- Systematic search
- Molecular dynamics
- Simulated annealing
- Genetic algorithms
- Incremental construction
- Rotamer libraries

Results & Discussion

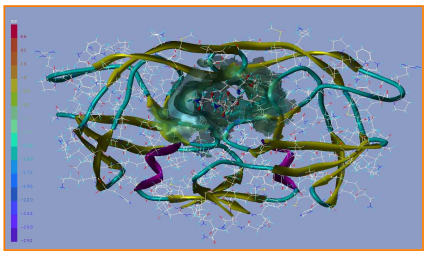
Scoring Functions

Goal: estimate binding affinity for given protein, ligand, pose, and conformations



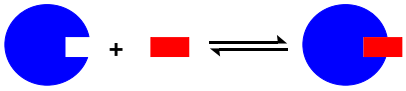
Scoring Functions

Goal: estimate binding affinity for given protein, ligand, pose, and conformations



HIV-1 protease complexed with an hydroxyethylene isostere inhibitor [Marsden04]

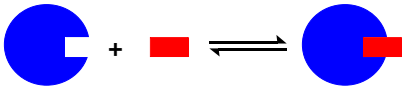
Scoring Functions



Binding Equations:

$R + L \underset{K_d}{\overset{K_a}{\rightleftharpoons}} R'L'$ $\qquad K_a = K_d^{-1} = \frac{[R'L']}{[R][L]}$

Scoring Functions



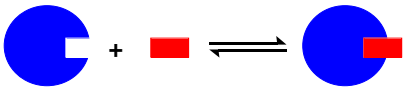
Binding Equations:

$R + L \underset{K_d}{\overset{K_a}{\rightleftharpoons}} R'L'$ $\qquad K_a = K_d^{-1} = \frac{[R'L']}{[R][L]}$

Free Energy Equations:

$\Delta G^o = -RT\ln(K_a)$ $\qquad \Delta G^o = \Delta H^o - T\Delta S^o$

Scoring Functions



Binding Equations:

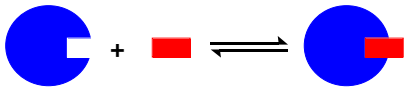
$R + L \underset{K_d}{\overset{K_a}{\rightleftharpoons}} R'L'$ $\qquad K_a = K_d^{-1} = \frac{[R'L']}{[R][L]}$

Free Energy Equations:

$\Delta G^o = -RT\ln(K_a)$ $\qquad \Delta G^o = \Delta H^o - T\Delta S^o$

enthalpy entropy

Scoring Functions



Binding Equations:

$R + L \underset{K_d}{\overset{K_a}{\rightleftharpoons}} R'L'$ $\qquad K_a = K_d^{-1} = \frac{[R'L']}{[R][L]}$

Free Energy Equations:

$\Delta G^o = -RT\ln(K_a)$ $\qquad \Delta G^o = \Delta H^o - T\Delta S^o$

K_a : 10^{-2} – 10^{-12} M
ΔG^o : -10 to -70 kcal/mol

enthalpy entropy

Factors Affecting ΔG^o

Intramolecular Forces (covalent)
- Bond lengths
- Bond angles
- Dihedral angles

Intermolecular Forces (noncovalent)
- Electrostatics
- Dipolar interactions
- Hydrogen bonding
- Hydrophobicity
- van der Waals

ΔH and ΔS work against each other in many situations.

Scoring Functions

Molecular mechanics force fields:
- CHARMM [Brooks83]
- AMBER [Cornell95]

Empirical methods:
- ChemScore [Eldridge97]
- GlideScore [Friesner04]
- AutoDock [Morris98]

Knowledge-based methods
- PMF [Muegge99]
- Bleep [Mitchell99]
- DrugScore [Gohlke00]

Scoring Functions

Molecular mechanics force fields: ←
- CHARMM [Brooks83]
- AMBER [Cornell95]

Empirical methods:
- ChemScore [Eldridge97]
- GlideScore [Friesner04]
- AutoDock [Morris98]

Knowledge-based methods
- PMF [Muegge99]
- Bleep [Mitchell99]
- DrugScore [Gohlke00]

Molecular Mechanics Force Fields

CHARMM:

$$U\left(\vec{R}\right) = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{UB}} K_{UB} (S - S_0)^2 + \sum_{\text{angle}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) + \sum_{\text{impropers}} K_{\text{imp}} (\phi - \phi_0)^2 + \sum_{\text{nonbond}} \epsilon \left[\left(\frac{R_{min_{ij}}}{r} \right)^{12} - \left(\frac{R_{min_{ij}}}{r} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}}$$

K_b, K_{UB}, K_θ, K_χ, K_{imp} are constants, b is the bond length, b_0 is the equilibrium bond length, S is the UB 1,3-distance, S_0 is the ideal UB 1,3-distance, θ is the angle value, θ_0 is the equilibrium angle value, χ is the dihedral angle value, n is the periodicity, ϕ is the improper angle value, ϕ_0 is the ideal improper angle value, ϵ is the Lennard-Jones well depth, $R_{min_{ij}}$ is the distance at the Lennard Jones minimum, q_i and q_j are the atoms' charge ϵ_1 is the effective dielectric constant, r_{ij} is the distance between the atoms

[Brooks83]

Molecular Mechanics Force Fields

AMBER:

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{\text{nonbond}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right] + \sum_{\text{H-bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right]$$

K_r, K_θ, V_n, A_{ij}, B_{ij}, C_{ij}, D_{ij} are constants, r is the bond length, r_{eq} is the equilibrium bond length, θ is the angle value, θ_{eq} is the equilibrium angle value, ϕ is the dihedral angle value, n is the periodicity, q_i and q_j are the atoms' charge ϵ is the effective dielectric constant, R_{ij} is the distance between the atoms

[Cornell95]

Scoring Functions

Molecular mechanics force fields:
- CHARMM [Brooks83]
- AMBER [Cornell95]

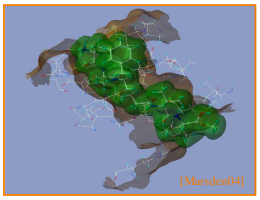
Empirical methods:
- ChemScore [Eldridge97]
- GlideScore [Friesner04]
- AutoDock [Morris98]

Knowledge-based methods
- PMF [Muegge99]
- Bleep [Mitchell99]
- DrugScore [Gohlke00]

Empirical Scoring Functions

ChemScore:

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{hbond} \sum_{il} g_1(\Delta r) g_2(\Delta\alpha) + \Delta G_{metal} \sum_{aM} f(r_{aM}) + \Delta G_{lipo} \sum_{iL} f(r_{iL}) + \Delta G_{rot} H_{rot}$$



[Marsden04]

[Eldridge97]

Empirical Scoring Functions

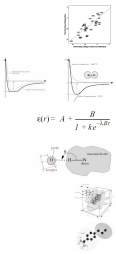
GlideScore:

$$\Delta G_{bind} = C_{hbond-neut-neut} \sum g(\Delta r) h(\Delta\alpha) + C_{hbond-neut-charged} \sum g(\Delta r) h(\Delta\alpha) + C_{hbond-charged-charged} \sum g(\Delta r) h(\Delta\alpha) + C_{mux-metal-ion} \sum f(r_{lm}) h(\Delta\alpha) + C_{lipo-lipo} \sum f(r_{lr}) + C_{rotb} H_{rotb} + C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} + solvation terms$$

[Friesner04]

Empirical Scoring Functions

AutoDock 3.0:



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

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

[Huey & Morris, AutoDock & ADT Tutorial, 2005]

Scoring Functions

Molecular mechanics force fields:
- CHARMM [Brooks83]
- AMBER [Cornell95]

Empirical methods:
- ChemScore [Eldridge97]
- GlideScore [Friesner04]
- AutoDock [Morris98]

Knowledge-based methods
- PMF [Muegge99]
- Bleep [Mitchell99]
- DrugScore [Gohlke00]

Knowledge-Based Scoring Functions

DrugScore:

$$\Delta W = \gamma\left[\sum_{k_i}\sum_{l_j}\Delta W_{i,j}(r)\right] + (1-\gamma)\times\left[\begin{array}{l}\sum_{k_i}\Delta W_i(SAS, SAS_0) + \\ \sum_{l_j}\Delta Wj(SAS, SAS_0)\end{array}\right]$$

Protein-Ligand Atom Distance Term

Solvent Accessible Surface Terms

[Gohlke00]

Scoring Functions

Molecular mechanics force fields:
- CHARMM [Brooks83]
- AMBER [Cornell95]

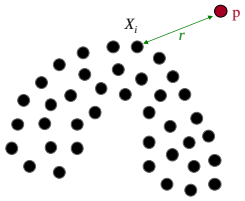
Empirical methods:
- ChemScore [Eldridge97]
- GlideScore [Friesner04]
- AutoDock [Morris98]

How to compute score efficiently?

Knowledge-based methods
- PMF [Muegge99]
- Bleep [Mitchell99]
- DrugScore [Gohlke00]

Computing Scoring Functions

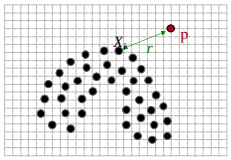
Point-based calculation:
- Sum terms computed at positions of ligand atoms (this will be slow)

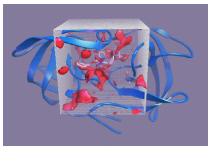


Computing Scoring Functions

Grid-based calculation:
- Precompute "force field" for each term of scoring function for each conformation of protein (usually only one)
- Sample force fields at positions of ligand atoms

ØAccelerate calculation of scoring function by 100X





[Huey & Morris]

Outline

Introduction

Scoring functions
- Molecular mechanics
- Empirical functions
- Knowledge-based

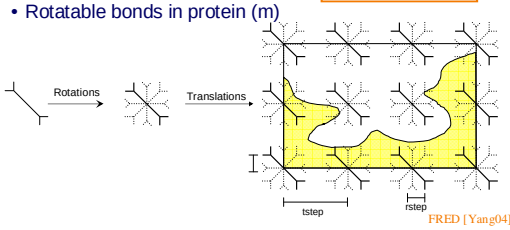
Searching poses & conformations
- Systematic search
- Molecular dynamics
- Simulated annealing
- Genetic algorithms
- Incremental construction
- Rotamer libraries

Results & Discussion

Systematic Search

Uniform sampling of search space
- Relative position (3)
- Relative orientation (3)
- Rotatable bonds in ligand (n)
- Rotatable bonds in protein (m)

The search space has dimensionality $3 + 3 + r^n + r^m$

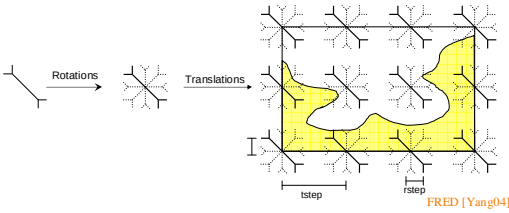


FRED [Yang04]

Systematic Search

Uniform sampling of search space

- Exhaustive, deterministic
- Quality dependent on granularity of sampling
- Feasible only for low-dimensional problems

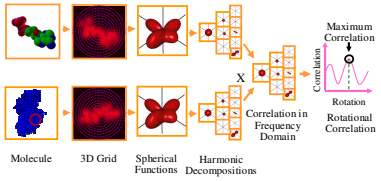


FRED [Yang04]

Systematic Search

Uniform sampling of search space

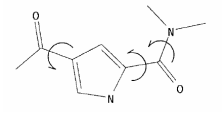
- Exhaustive, deterministic
- Quality dependent on granularity of sampling
- Feasible only for low-dimensional problems
  - § Example: search all rotations – Wigner-D^{-1}

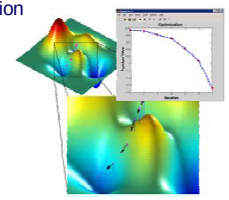


Molecular Mechanics

Energy minimization:

- Start from a random or specific state (position, orientation, conformation)
- Move in direction indicated by derivatives of energy function
- Stop when reach local minimum



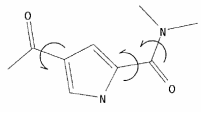


[Marai04]

Simulated Annealing

Monte Carlo search of parameter space:

- Start from a random or specific state (position, orientation, conformation)
- Make random state changes, accepting up-hill moves with probability dictated by "temperature"
- Reduce temperature after each move
- Stop after temperature gets very small

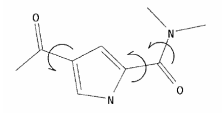


AutoDock 2.4 [Morris96]

Genetic Algorithm

Genetic search of parameter space:

- Start with a random population of states
- Perform random crossovers and mutations to make children
- Select children with highest scores to populate next generation
- Repeat for a number of iterations

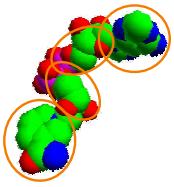


Gold [Jones95], AutoDock 3.0 [Morris98]

Incremental Extension

Greedy fragment-based construction:

- Partition ligand into fragments

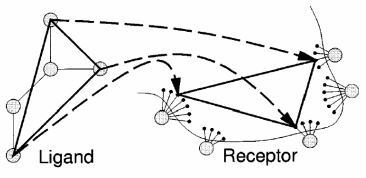


FlexX [Rarey96]

Incremental Extension

Greedy fragment-based construction:

- Partition ligand into fragments
- Place base fragment (e.g., with geometric hashing)

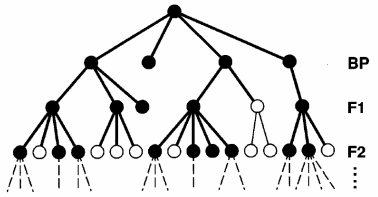


FlexX [Rarey96]

Incremental Extension

Greedy fragment-based construction:

- Partition ligand into fragments
- Place base fragment (e.g., with geometric hashing)
- Incrementally extend ligand by attaching fragments

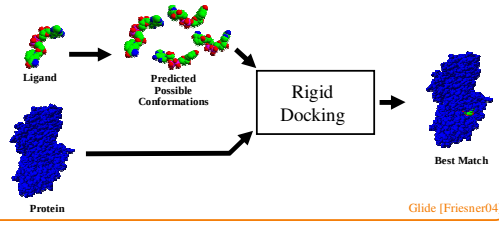


FlexX [Rarey96]

Rotamer Libraries

Rigid docking of
many conformations:

- Precompute all low-energy conformations
- Dock each precomputed conformations as rigid bodies

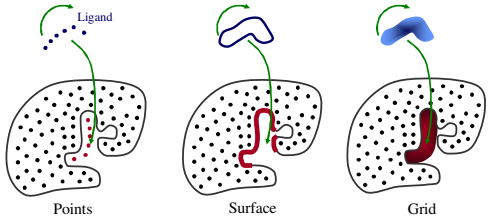


Glide [Friesner04]

Rigid Docking

This is just like matching binding sites (complement)

- Can use same methods we used for matching and indexing point, surface, and/or grid representations



Outline

Introduction

Scoring functions

- Molecular mechanics
- Empirical functions
- Knowledge-based

Searching poses & conformations

- Systematic search
- Molecular dynamics
- Simulated annealing
- Genetic algorithms
- Incremental construction
- Rotamer libraries

Results & Discussion

Discussion

?

References

[Brooks83] Bernhard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, Martin Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," J. Comp. Chem, 4, 2, 1983, pp. 187-217.

[Eldridge97] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, "Empirical scoring functions. I: The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes," J. Comput.-Aided Mol. Des., 11, 1997, pp. 425-445.

[Friesner04] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, "Glide: A New Approach for Rapid, Accurate Docking and Scoring.," J. Med. Chem, 47, 2004, pp. 1739-1749.

[Gohlke00] H. Gohlke, M. Hendlich, G. Klebe, "Knowledge-based scoring function to predict protein-ligand interactions," J. Mol. Biol., 295, 2000, pp. 337-356.

[Huey & Morris] Ruth Huey and Garrett M. Morris, "Using AutoDock With AutoDockTools: A Tutorial", http://www.scripps.edu/mb/olson/doc/autodock/UsingAutoDockWithADT.ppt

[Jones97] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, "Development and Validation of a Genetic Algorithm for Flexible Docking," J. Mol. Biol., 267, 1997, pp. 727-748.

[Marai04] Christopher Marai, "Accommodating Protein Flexibility in Computational Drug Design," Mol Pharmacol, 57, 2, 2004, p. 213-8.

[Marsden04] Marsden PM, Puvanendrampillai D, Mitchell JBO and Glen RC., "Predicting protein ligand binding affinities: a low scoring game?", Organic Biomolecular Chemistry, 2, 2004, p. 3267-3273.

[Mitchell99] J.B.O. Mitchell, R. Laskowski, A. Alex, and J.M. Thornton, "BLEEP - potential of mean force describing protein-ligand interactions: I. Generating potential", J. Comput. Chem., 20, 11, 1999, 1165-1176.

[Morris98] Morris, G. M., Goodsell, D. S., Halliday, R.S., Huey, R., Hart, W. E., Belew, R. K. and Olson, A. J. "Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function", J. Computational Chemistry, 19, 1998, p. 1639-1662.

[Muegge99] I. Muegge, Y.C. Martin, "A general and fast scoring function for protein-ligand interactions: A simplified potential approach," J. Med. Chem., 42, 1999, pp. 791-804.

[Rarey96] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, "A Fast Flexible Docking Method using an Incremental Construction Algorithm," Journal of Molecular Biology, 261, 3, 1996, pp. 470-489.