M1 ISDD - BI (Feb 2018)

PROTEIN DOCKING

Lesson 2

Geometry-based docking methods

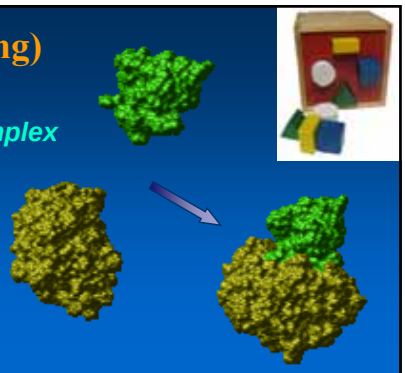
Juan Fernández-Recio

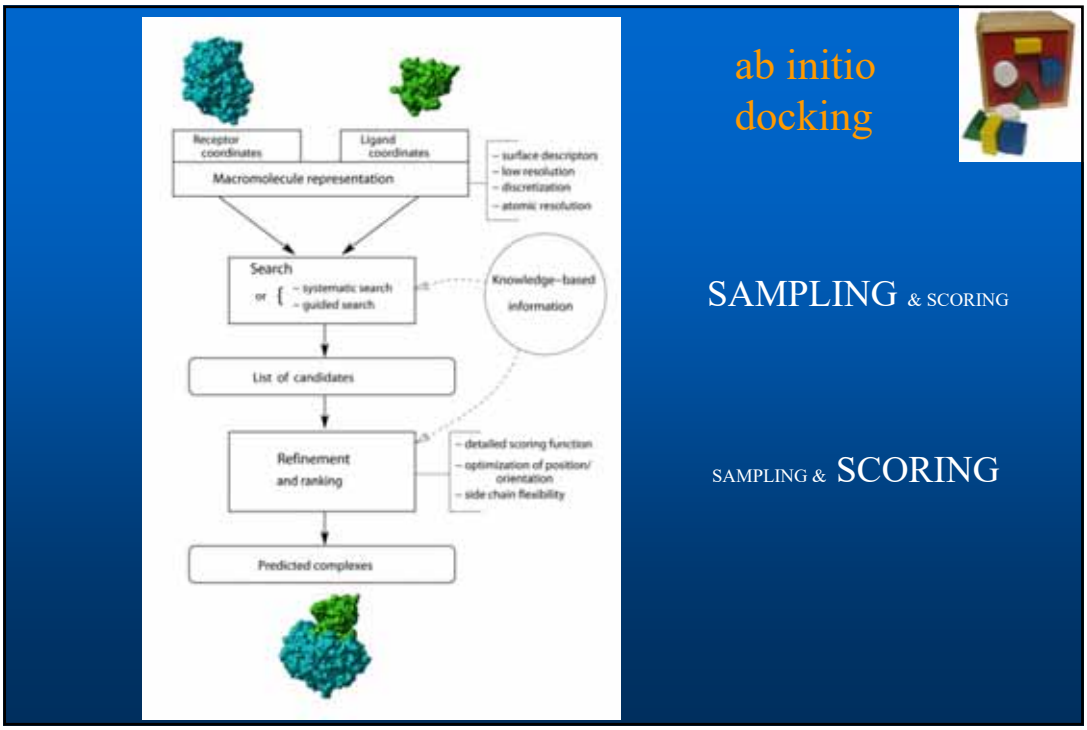
CSIC, BSC

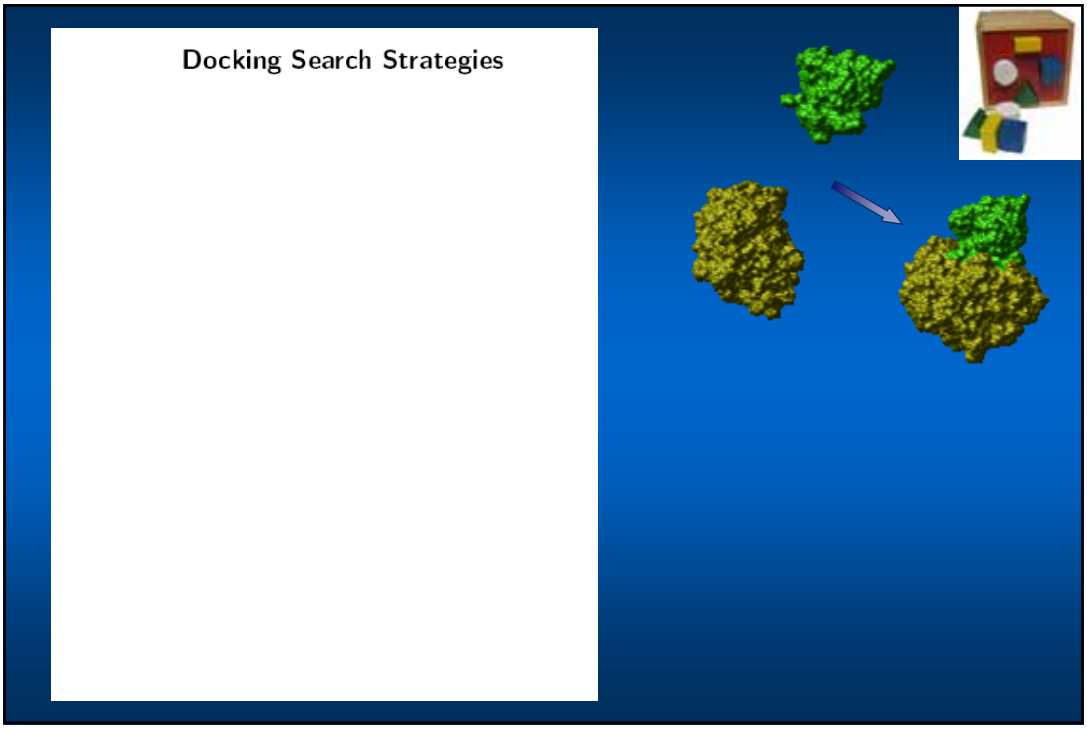
juanf@bsc.es

Complex structure prediction (docking)

Generation of the structure of a protein-protein complex from the individual protein structures

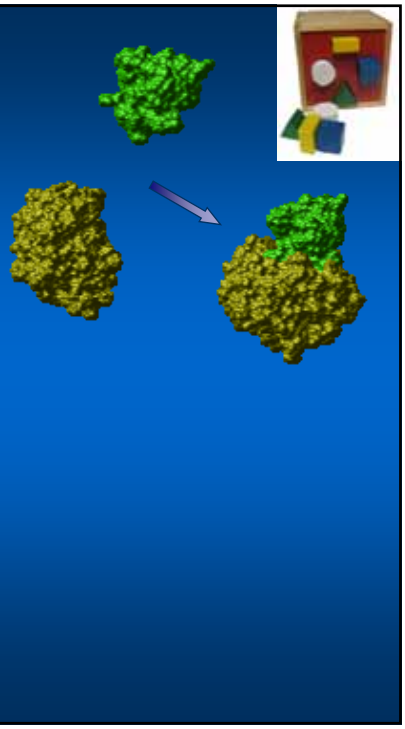



Receptor coordinates
Ligand coordinates
Macromolecule representation
– surface descriptors
– low resolution
– discretization
– atomic resolution
Search
or { – systematic search
– guided search
Knowledge-based information
List of candidates
Refinement and ranking
– detailed scoring function
– optimization of position/ orientation
– side chain flexibility
Predicted complexes
ab initio docking
SAMPLING & SCORING
SAMPLING & SCORING

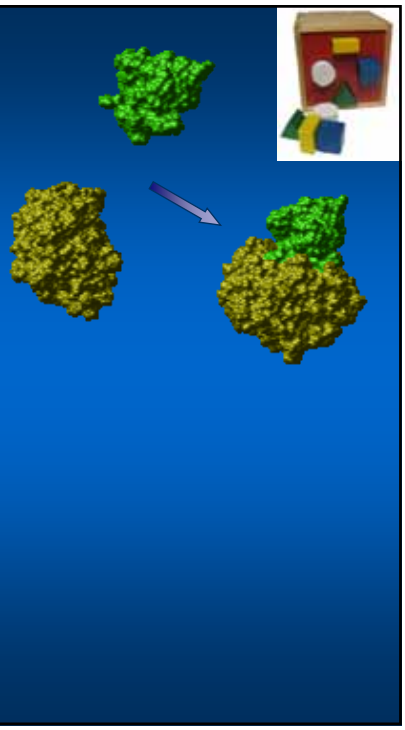

Docking Search Strategies

Docking Search Strategies

- **Pseudo Random**
  - Simulated Annealing / Monte Carlo
  - Genetic Algorithms
- **Directed Search**
  - Geometric Hashing
  - Spherical Harmonic Surface Triangles
- **Brute-Force Search**
  - Explicit Grid Correlations
  - Fast Fourier Transform (FFT) Correlations
  - Spherical Polar Fourier Correlations
- **Refinement Phase**
  - Classical or Soft Potentials (+/- Electrostatics)
  - Desolvation, Solvent Dipoles...
  - Visual Inspection!!

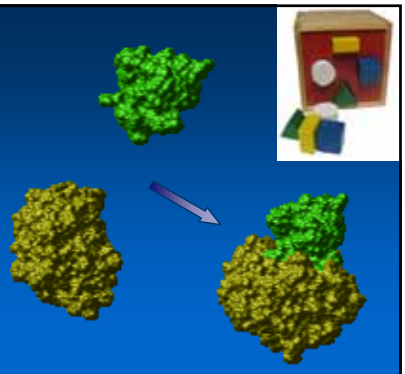


Criteria for Good Docking Orientations



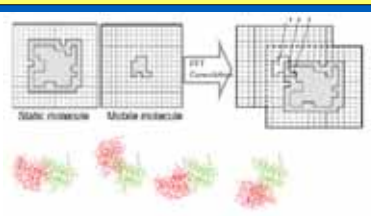
Criteria for Good Docking Orientations

- Low Free Energy (Difficult!)
- Low Pseudo-Energy (Easy) Based On...
- Large Surface Burial: $\sim 1600 \pm 400 Å^2$
- Small van der Waals Overlaps
- No Large Cavities in Interface
- Good H-Bonding: $\sim$ 1 HB/100 $Å^2$
- Good Charge Complementarity
- Polar/Polar Contacts Favoured
- Polar/Non-Polar Contacts Disfavoured



Protein-protein docking methods

Exhaustive search (FFT, surface-based)

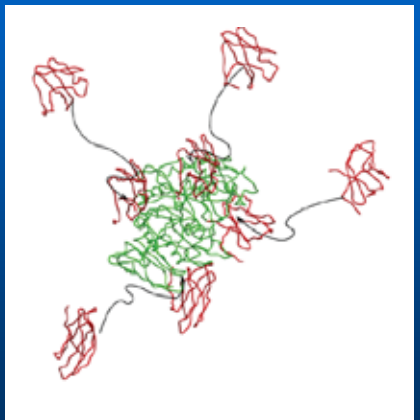


Scoring function

- Energy terms: ES, H-bonds, ...
- Conservation
- Statistical potentials

Geometry-based docking

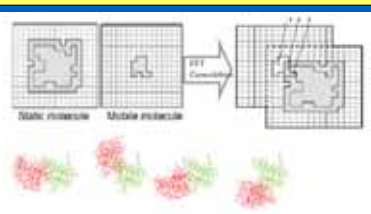
Stochastic sampling (Monte-Carlo, minimization)



Energy-based docking

Protein-protein docking methods

Exhaustive search (FFT, surface-based)



Scoring function

- Energy terms: ES, H-bonds, ...
- Conservation
- Statistical potentials

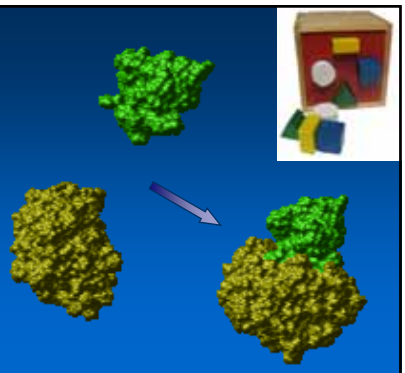
Geometry-based docking

Stochastic sampling (Monte-Carlo, minimization)

Energy-based docking

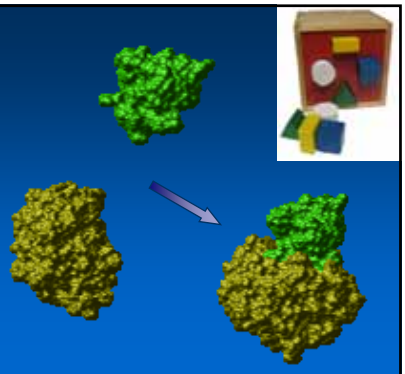
Docking Search Strategies

- **Pseudo Random**
  - Simulated Annealing / Monte Carlo
  - Genetic Algorithms
- **Directed Search**
  - Geometric Hashing
  - Spherical Harmonic Surface Triangles
- **Brute-Force Search**
  - Explicit Grid Correlations
  - Fast Fourier Transform (FFT) Correlations
  - Spherical Polar Fourier Correlations
- **Refinement Phase**
  - Classical or Soft Potentials (+/- Electrostatics)
  - Desolvation, Solvent Dipoles...
  - Visual Inspection!!



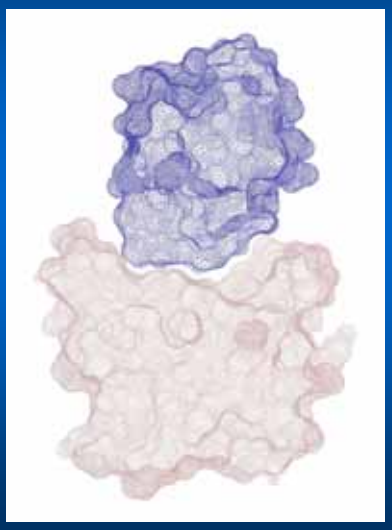
Criteria for Good Docking Orientations

- Low Free Energy (Difficult!)
- Low Pseudo-Energy (Easy) Based On...
- Large Surface Burial: $\sim 1600 \pm 400 Å^2$
- Small van der Waals Overlaps
- No Large Cavities in Interface
- Good H-Bonding: ~ 1 HB/100 $Å^2$
- Good Charge Complementarity
- Polar/Polar Contacts Favoured
- Polar/Non-Polar Contacts Disfavoured

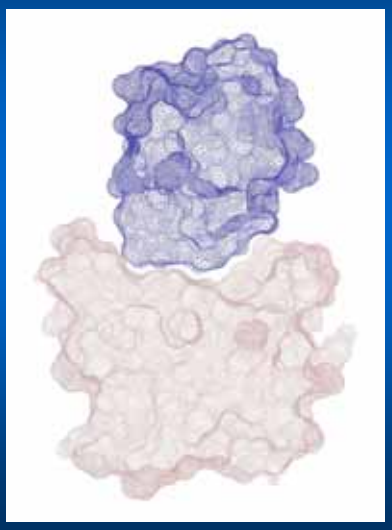


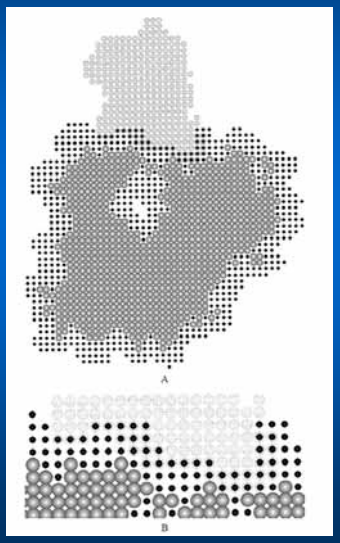
- **FFT-based grid search**
- **Geometric Hashing**
- **Adding distance-, symmetry- constraints**

Rigid-Body Docking: Geometry Approach

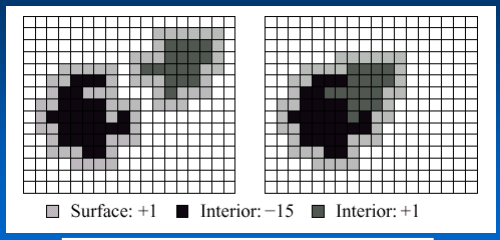


Rigid-Body Docking: Geometry Approach





MolFit



$$\bar{a}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

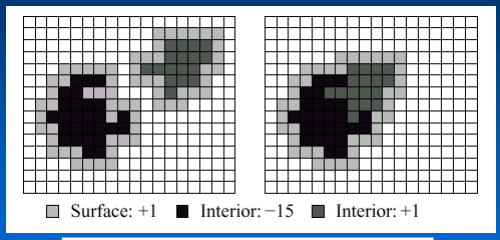
$$\bar{b}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$l, m, n = (1, \ldots, N)$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl Acad. Sci. USA 89 (1992) 2195–2199.

MolFit



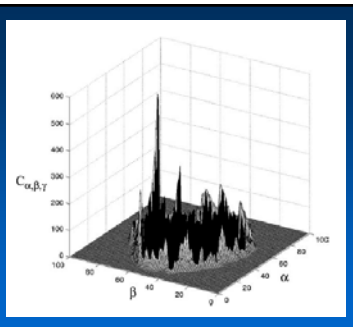
$$\bar{a}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$\bar{b}_{l,m,n} = \begin{Bmatrix} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{Bmatrix}$$

$$l, m, n = (1, \ldots, N)$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl Acad. Sci. USA 89 (1992) 2195–2199.



a

b

Corr(a,b)

c

Fourier Transform

It re-express a function in terms of sinusoidal basis functions

Fourier transform equations:

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i k x} dx$$

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2\pi i k x} dk$$

$F(k)$ = FT($f(x)$)

$f(x)$ = IFT($F(k)$)

Fourier Transform

It re-express a function in terms of sinusoidal basis functions

Fourier transform equations:

$$F(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i k x} dx$$

$$f(x) = \int_{-\infty}^{\infty} F(k) e^{2\pi i k x} dk$$

$F(k)$ = FT($f(x)$)

$f(x)$ = IFT($F(k)$)

Correlation function:

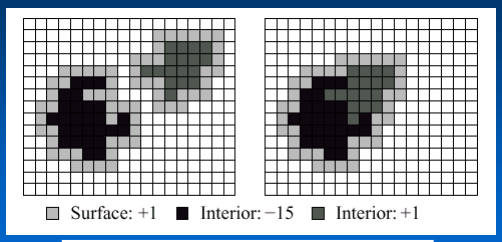
$$c(t) = \mathrm{Corr}(g, h) \equiv \int_{-\infty}^{\infty} g(\tau + t) h(\tau)\, d\tau$$

"Correlation Theorem":

$$FT(c) = \mathrm{FT}(g) \cdot [\mathrm{FT}(h)]^*$$

$$c = \mathrm{IFT}(\ \mathrm{FT}(g) \cdot [\mathrm{FT}(h)]^*\ )$$

MolFit



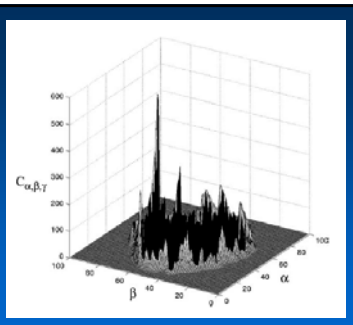
$$\bar{a}_{l,m,n} = \begin{cases} 1 \text{ on the surface of the molecule} \\ \rho \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{cases}$$

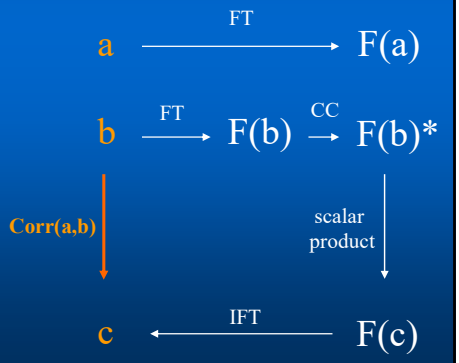
$$\bar{b}_{l,m,n} = \begin{cases} 1 \text{ on the surface of the molecule} \\ \delta \text{ inside the molecule} \\ 0 \text{ outside the molecule} \end{cases}$$

$$l, m, n = (1, \ldots, N)$$

$$\bar{c}_{\alpha,\beta,\gamma} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} \bar{a}_{l,m,n} \cdot \bar{b}_{l+\alpha,m+\beta,n+\gamma}$$

E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl Acad. Sci. USA 89 (1992) 2195–2199.





FFT (Fast Fourier Transform)

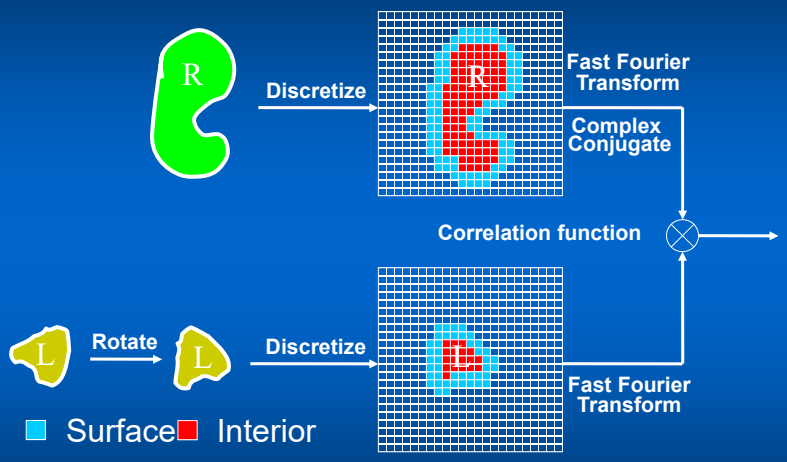
Algorithms for efficient calculation of FT and IFT

- Most common: Cooley-Tukey FFT (divide and conquer)
- Other: Prime-factor, Bruun's, Rader's, Bluestein's

Fourier transform timing: N^2 **(if** $N = 10^6$ **1MHz CPU time ~2 weeks)**
Fast Fourier transform: $N\log_2 N$ **(if** $N = 10^6$ **1MHz CPU time ~ 30 sec)**

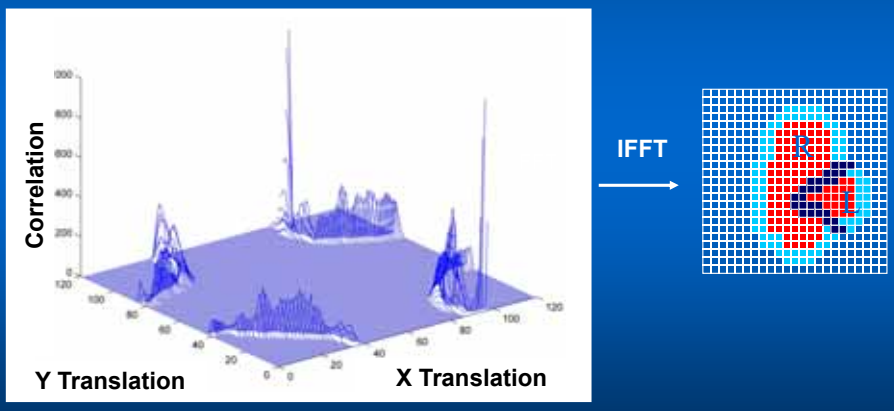
http://www.fftw.org/

Protein Docking Using FFT



Protein Docking Using FFT

Comp. cost can decrease by >10^4 (from N^6 to $N^3 \ln N^3$)



FTDOCK

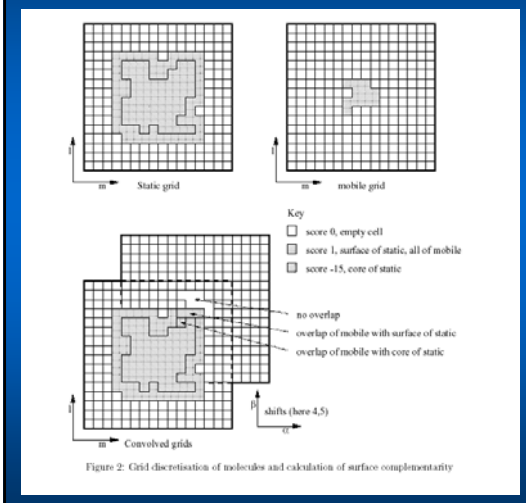
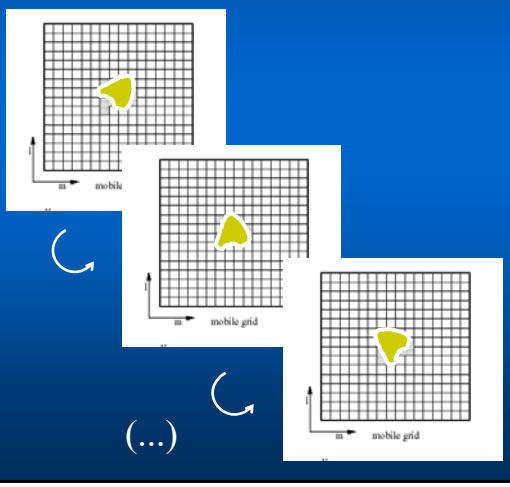


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

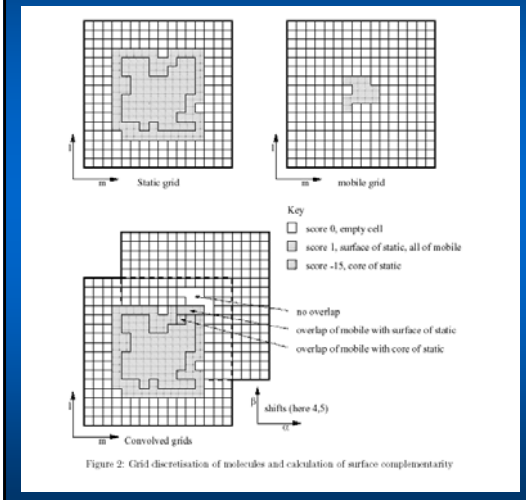
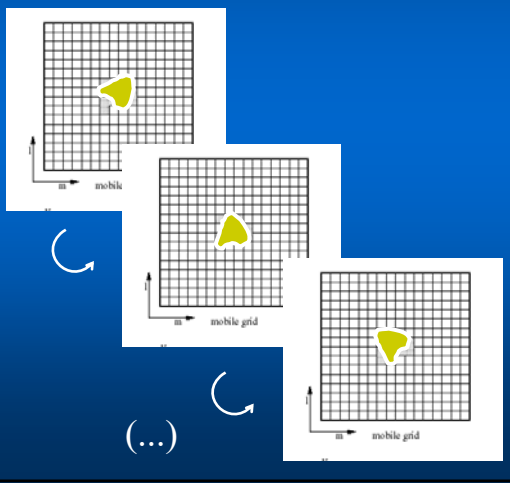


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

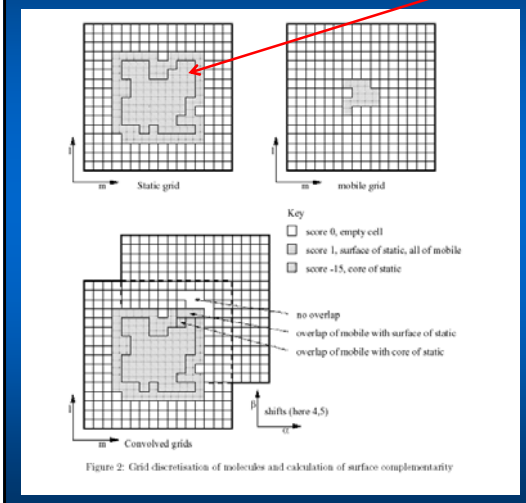
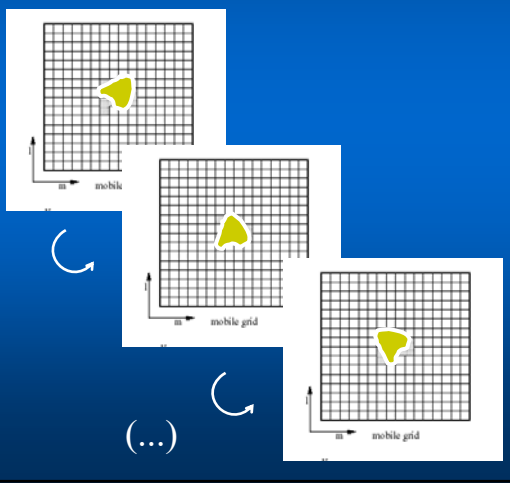


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



(…)

FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

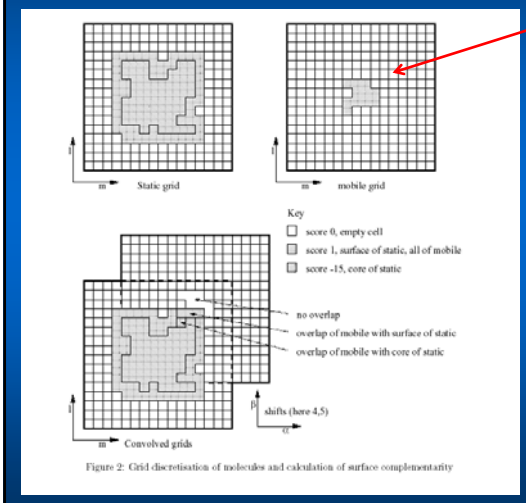
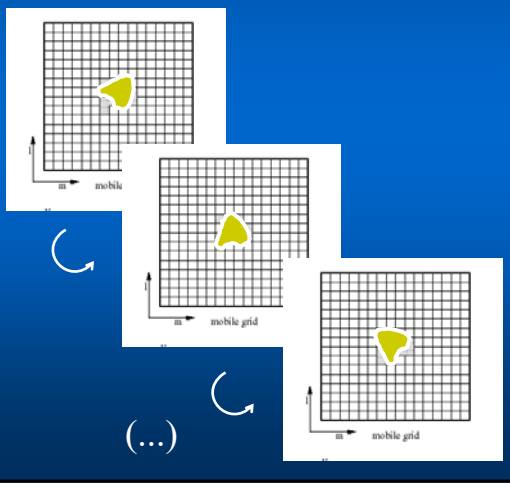


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



(…)

FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

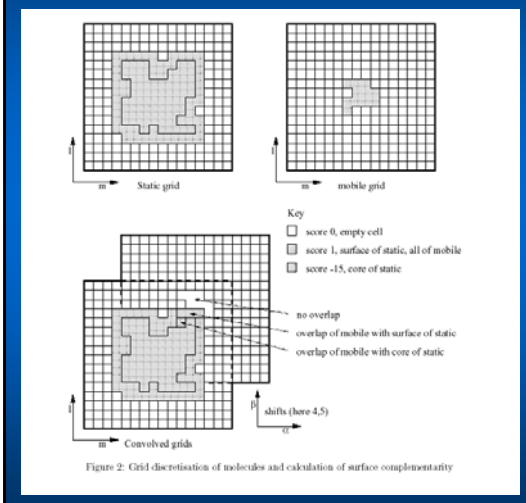
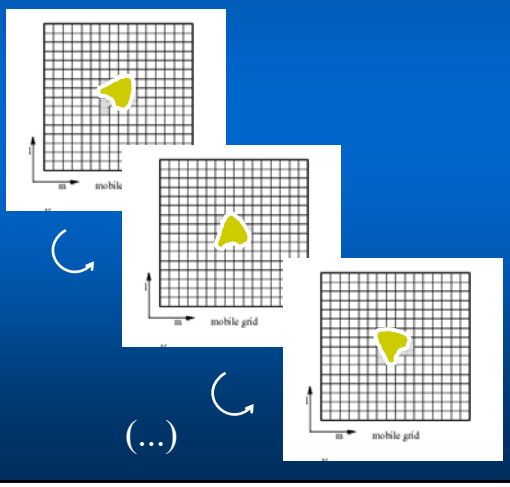


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



(…)

FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

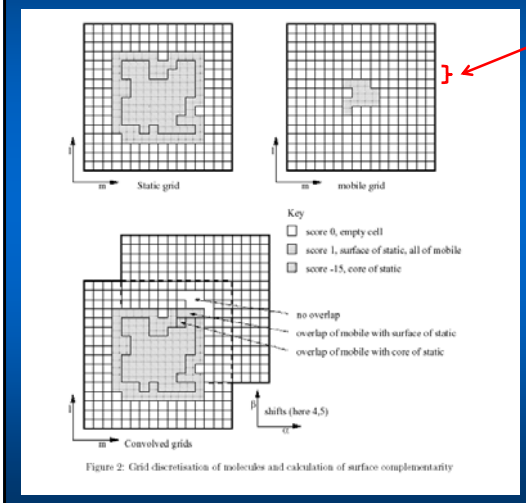
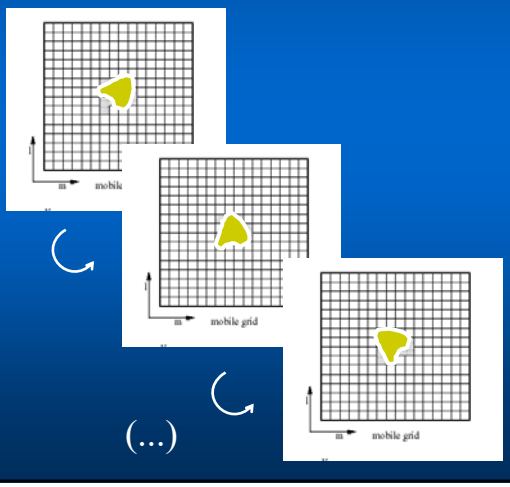


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



(…)

FTDOCK

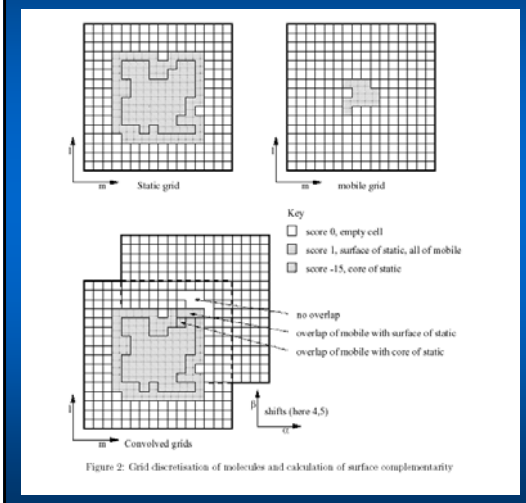
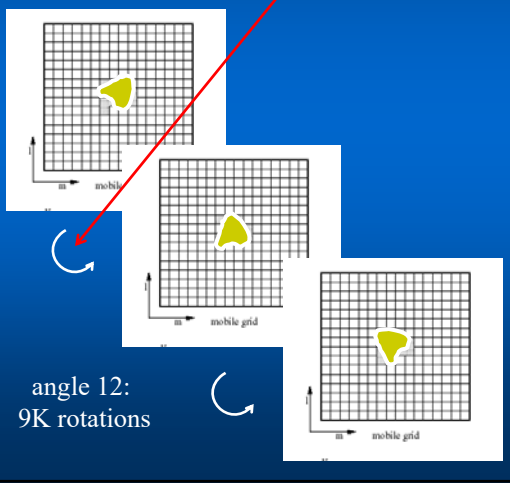


Figure 2: Grid discretisation of molecules and calculation of surface complementarity

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log



angle 12:
9K rotations

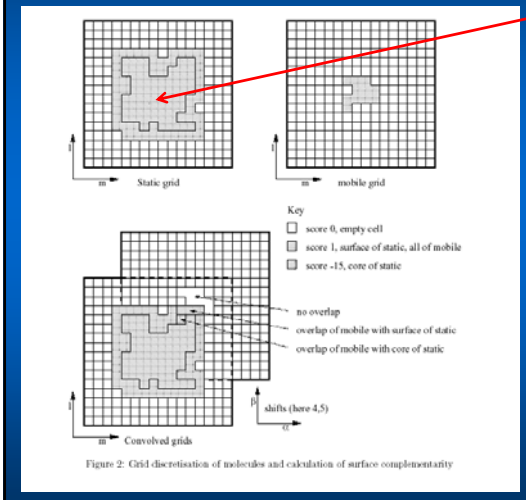
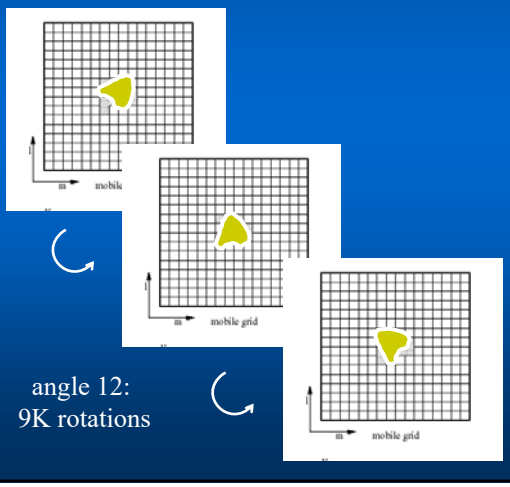


Figure 2: Grid discretisation of molecules and calculation of surface complementarity

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log



angle 12:
9K rotations

FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

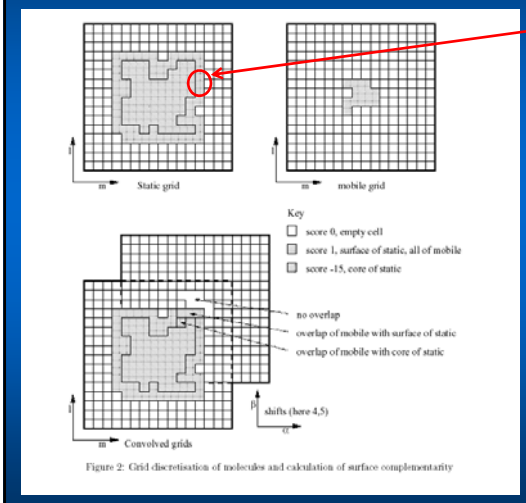
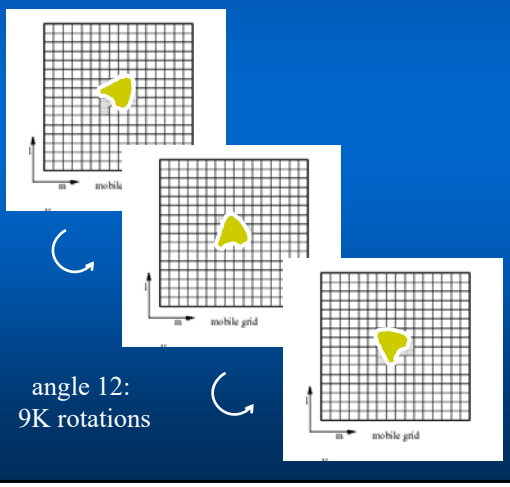


Figure 2: Grid discretisation of molecules and calculation of surface complementarity



angle 12:
9K rotations

FTDOCK

ftdock -static 1A2P_A.pdb -mobile 1A19_A.pdb
-noelec -calculate_grid 1.2 -angle_step 12
-internal -15 -surface 1.3 -keep 3
-out 1BRS.ftdock > 1BRS.ftdock.log

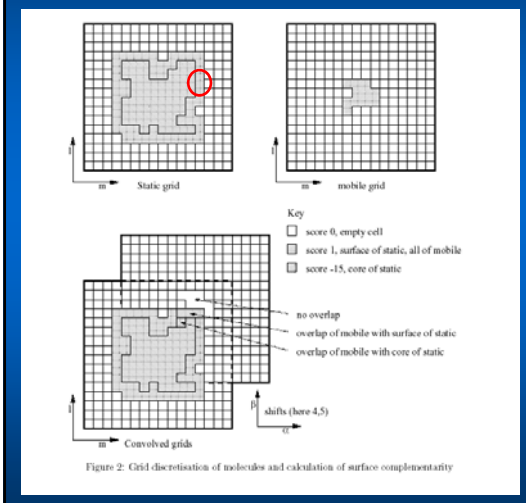
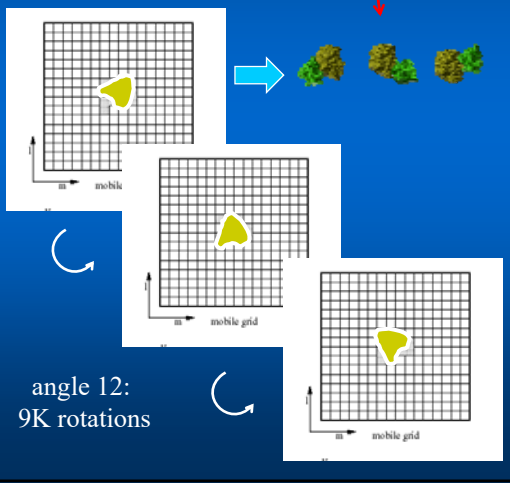
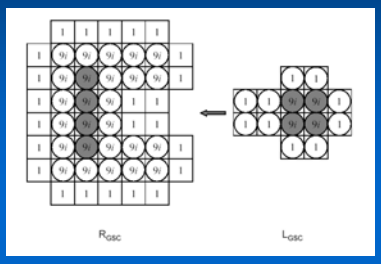


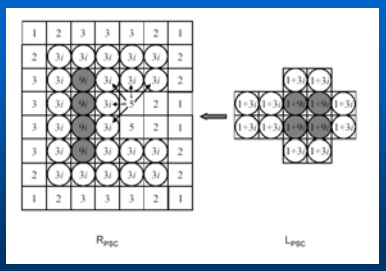
Figure 2: Grid discretisation of molecules and calculation of surface complementarity



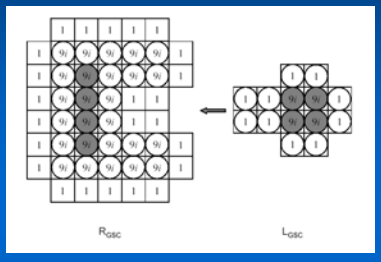
angle 12:
9K rotations

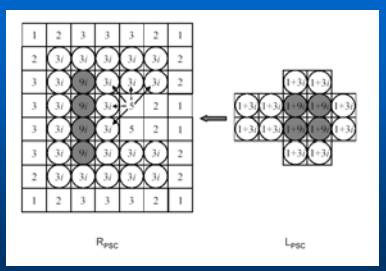
ZDOCK





ZDOCK

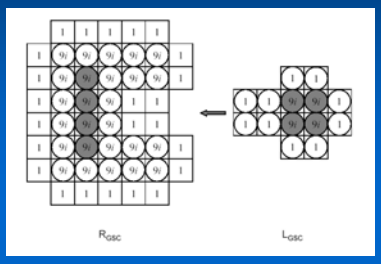




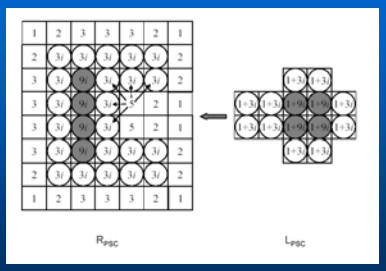
mark_sur 1A2P_A.pdb 1BRS_r.pdb
mark_sur 1A19_A.pdb 1BRS_l.pdb

```
ATOMS LIST
ATOM      1  N   VAL A   3      16.875  37.901  43.478  1     1 1.63    -0.15
ATOM      2  CA  VAL A   3      17.997  38.577  42.795  2     1 2.03     0.10
ATOM      3  C   VAL A   3      18.147  39.997  43.309  3     1 1.67     0.60
ATOM      4  O   VAL A   3      17.180  40.763  43.281  4     1 1.38    -0.55
ATOM      5  CB  VAL A   3      17.786  38.593  41.279  6     1 2.03     0.00
ATOM      6  CG1 VAL A   3      18.621  39.606  40.573 17     1 1.94     0.00
ATOM      7  CG2 VAL A   3      18.024  37.190  40.696 17     1 1.94     0.00
ATOM      8  N   ILE A   4      19.355  40.377  43.690  1     0 1.63    -0.15
ATOM      9  CA  ILE A   4      19.568  41.805  44.109  2     1 2.03     0.10
ATOM     10  C   ILE A   4      20.355  42.488  42.992  3     0 1.67     0.60
ATOM     11  O   ILE A   4      21.489  42.115  42.701  4     0 1.38    -0.55
ATOM     12  CB  ILE A   4      20.430  41.800  45.396  6     0 2.03     0.00
ATOM     13  CG1 ILE A   4      19.774  40.976  46.525 16     1 1.99     0.00
ATOM     14  CG2 ILE A   4      20.779  43.192  45.823 17     1 1.94     0.00
ATOM     15  CD1 ILE A   4      20.598  40.921  47.787 17     1 1.94     0.00
```

ZDOCK

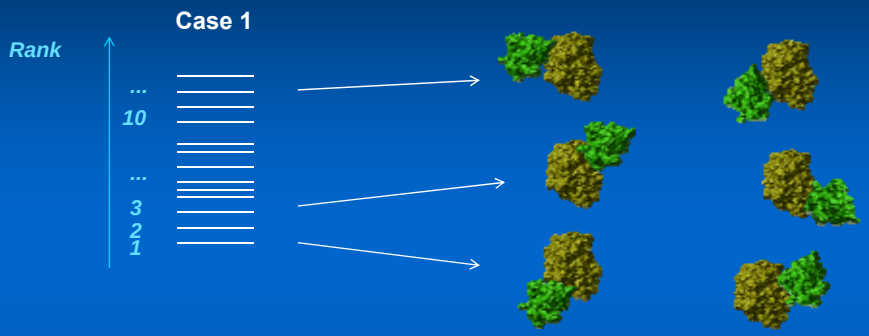


mark_sur 1A2P_A.pdb 1BRS_r.pdb
mark_sur 1A19_A.pdb 1BRS_l.pdb

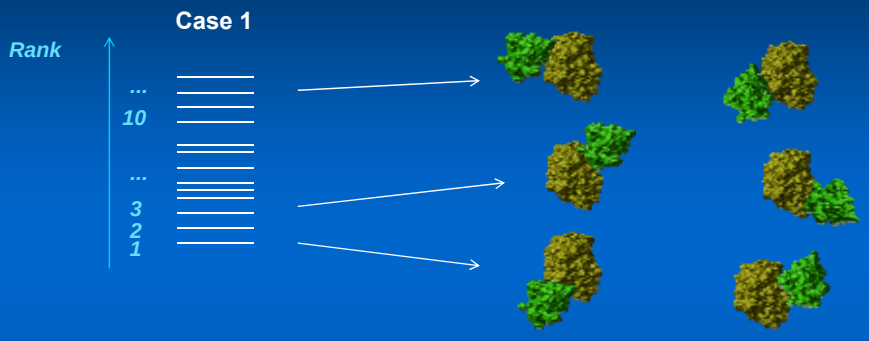


zdock -R 1BRS_r.pdb -L 1BRS_l.pdb -o 1BRS_zdock.out

Assessment of docking performance

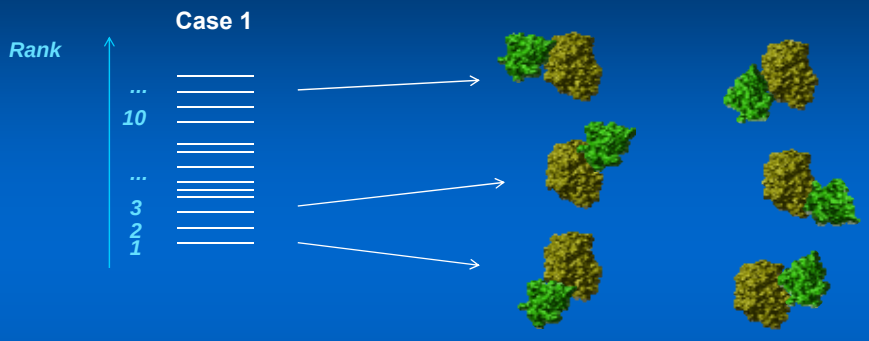


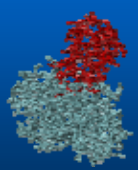
Assessment of docking performance



good solution ??

Assessment of docking performance





reference complex

good solution ??

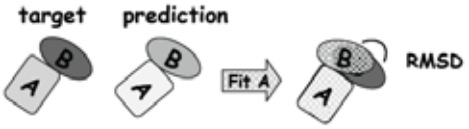
Assessment of docking performance

- F_{nat}: fraction of native contacts (within 5Å)

$$F_{nat} = \frac{\text{Correctly predicted contacts}}{\text{Total number of contacts in the target}}$$

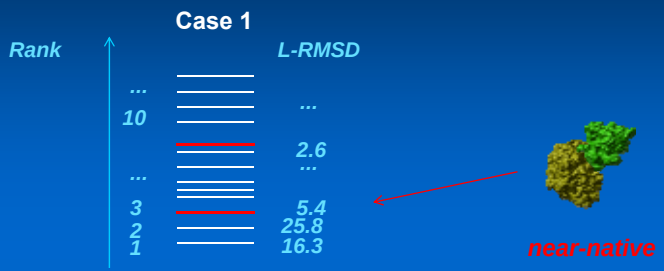
- l-RMSD: RMSD on second protein after superposition on first

near-native: L-RMSD < 10Å

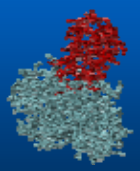


- i-RMSD: RMSD on interface residues (within 10 Å)

Assessment of docking performance

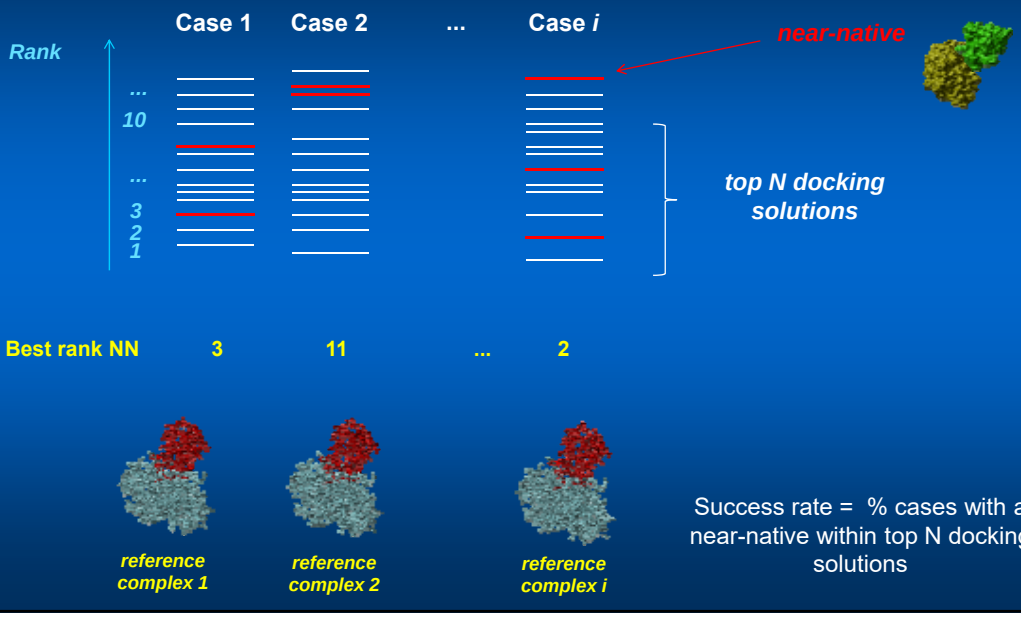


Best rank NN **3**



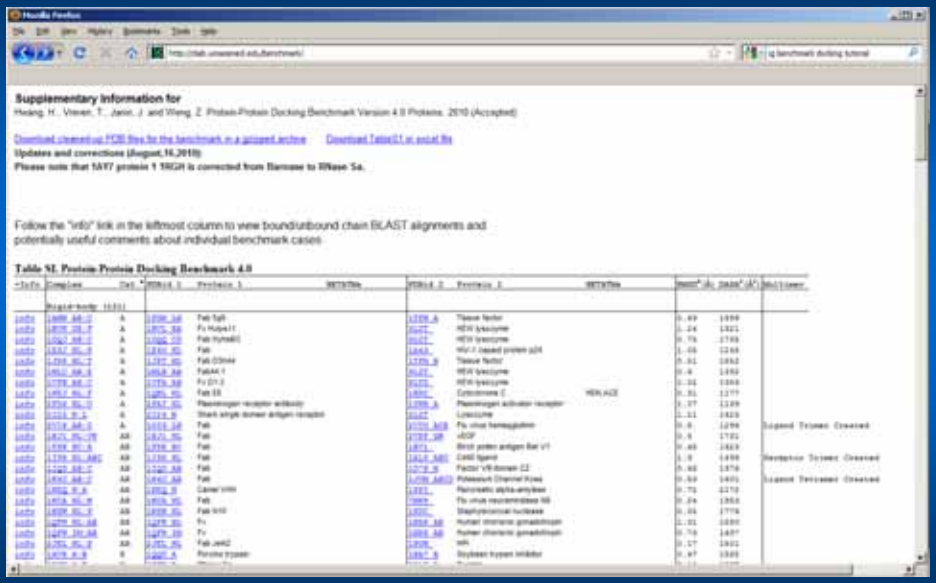
reference complex

Assessment of docking performance



Success rate = % cases with a near-native within top N docking solutions

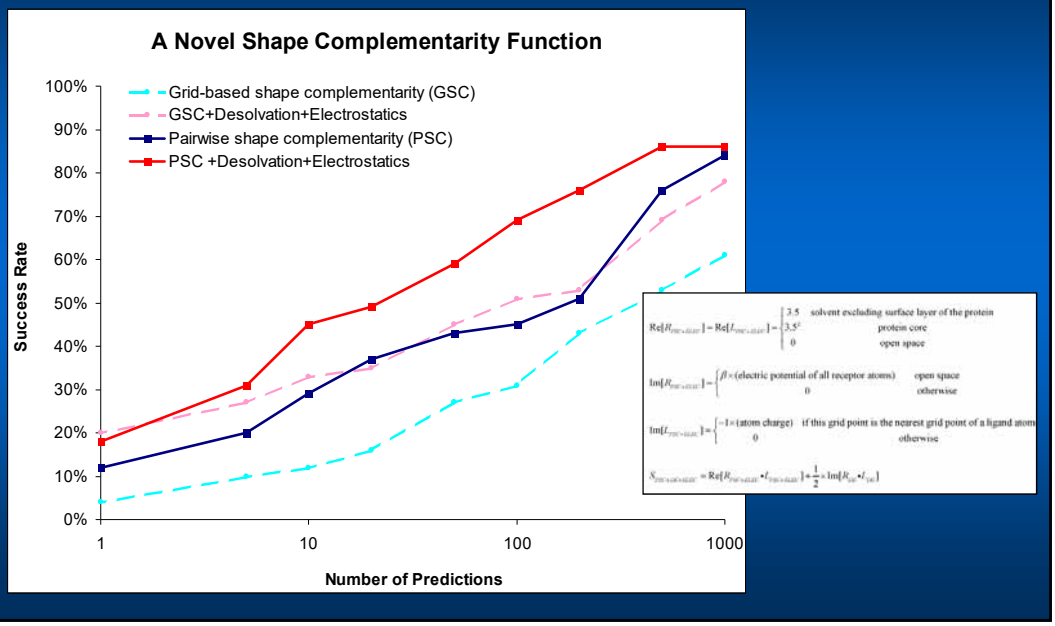
Protein-protein Docking Benchmark 4.0 - Weng



176 cases

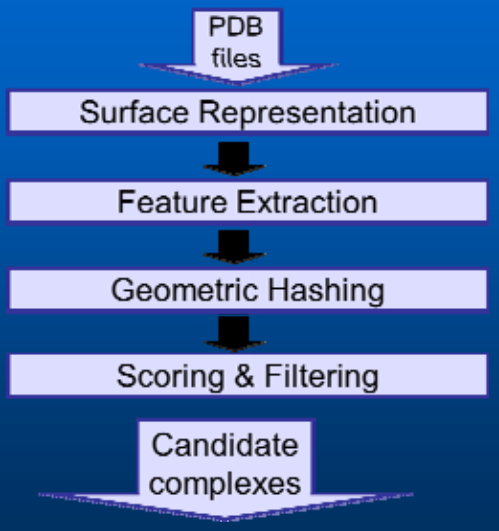
http://zlab.umassmed.edu/benchmark/

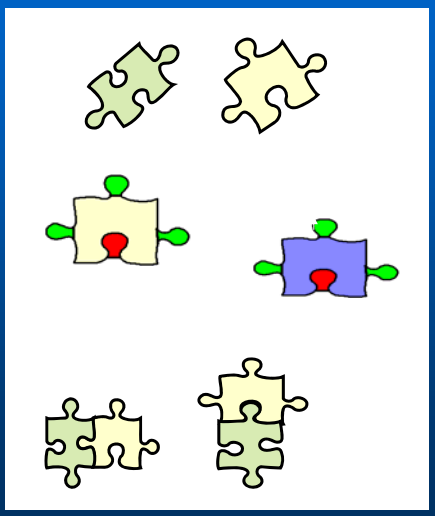
ZDOCK performance



- FFT-based grid search
- **Geometric Hashing**
- Adding distance-, symmetry- constraints

Geometrical hashing - PatchDock



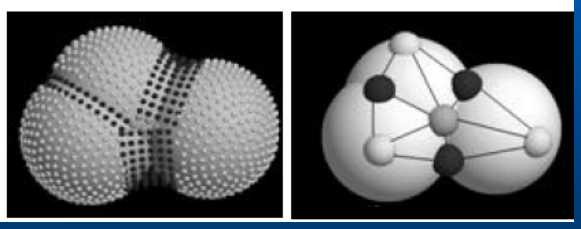


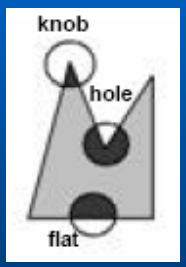
Schneidman-Duhovny et al. Proteins 2003
Duhovny (Schneidman), D., Nussinov, R. Wolfson,HJ. WABI 2002

Surface Representation

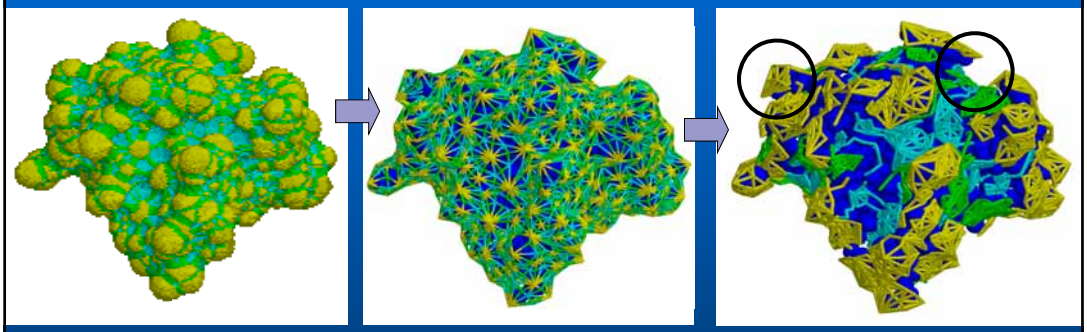
Surface Representation

- Dense MS surface (Connolly)
- Sparse surface (Shuo Lin et al.)





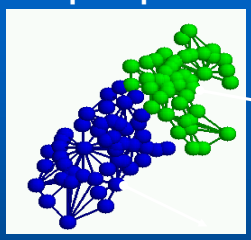
Feature Extraction



- We focus on sparse surface features, preserving the quality of shape representation.
- The sparse features reduce the complexity of the matching step.

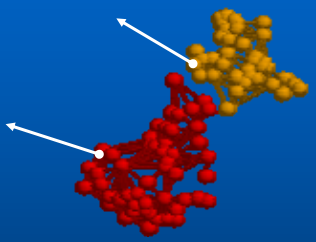
Patch-Pair Matching

Receptor patches



Transformation

Ligand patches

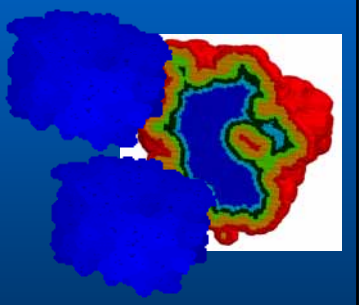


Base: 1 critical point with its normal from one patch and
1 critical point with its normal from a neighboring patch.

- Match every base from the receptor patches with all the bases from complementary ligand patches.
- Compute the transformation for each pair of matched bases.

Geometric Scoring

- The surface of the **receptor** is divided into five shells according to the distance function: **S1-S5**
- The number of **ligand** surface points in every shell is counted.
- The geometric score is a weighted sum of the number of ligand surface points inside every shell.

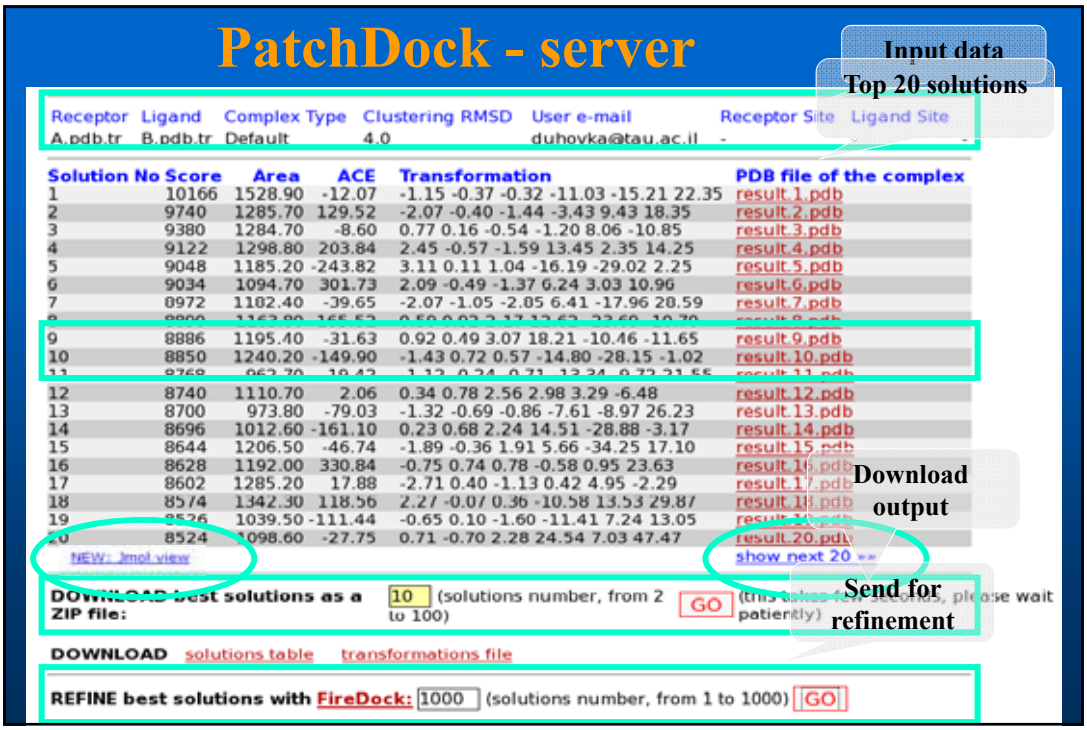


- **Multi-resolution surface** data structure was developed to speed up this stage.

http://bioinfo3d.cs.tau.ac.il/PatchDock

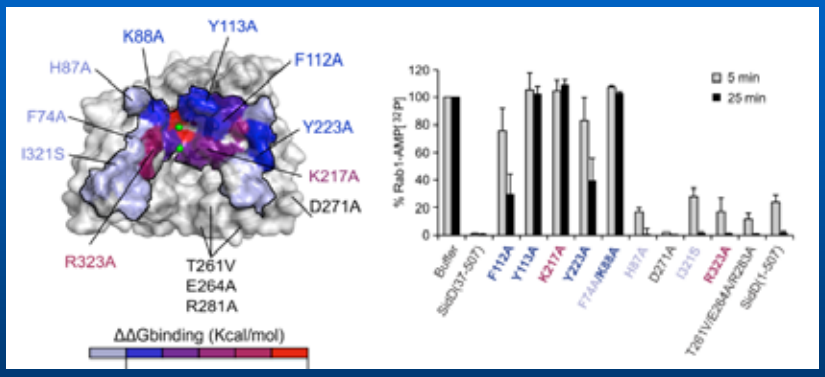
Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. NAR 2005

PatchDock - server



- FFT-based grid search
- Geometric Hashing
- Adding distance-, symmetry- constraints

Docking with known interface residues



Optional parameters: binding site

Advanced Options:
[Show][Hide]

Receptor Binding Site: Browse... upload receptor binding site file

Ligand Binding Site: Browse... upload ligand binding site file

Distance Constraints: Browse... upload distance constraints file

Submit Form Clear

Optional parameters: binding site

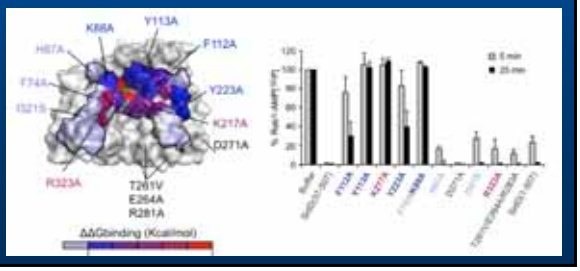
Advanced Options:
[Show][Hide]

Receptor Binding Site: Browse... upload receptor binding site file
Ligand Binding Site: Browse... upload ligand binding site file
Distance Constraints: Browse... upload distance constraints file
Submit Form Clear

Binding Site – list the residues of the receptor potential binding site. Only patches that include these residues will be used for matching.

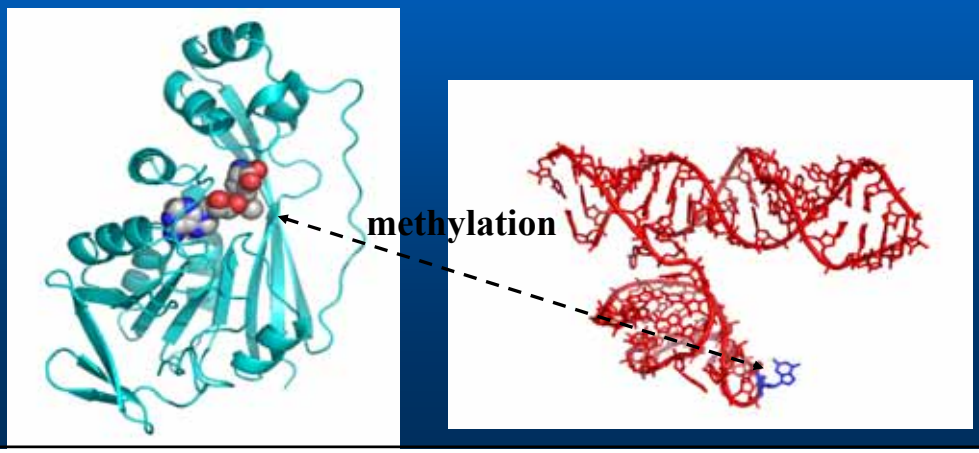
Format: ***[residue index] [chain ID]***
Example:

74 A
88 A
112 A
113 A
217 A
223 A



Docking with Distance Constraints

A complex of RIma2 methyltransferase of S. pneumoniae and a 74 nucleotide RNA transcript



Optional parameters

Advanced Options:
[Show][Hide]

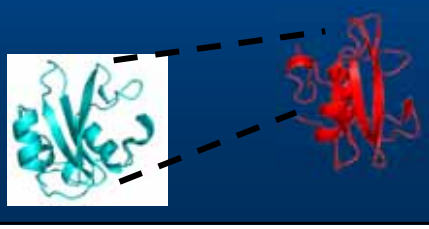
Receptor Binding Site: Browse... upload receptor binding site file
Ligand Binding Site: Browse... upload ligand binding site file
Distance Constraints: Browse... upload distance constraints file
Submit Form Clear

Distance Constraints –between pairs of atoms, one in the receptor and one in the ligand.

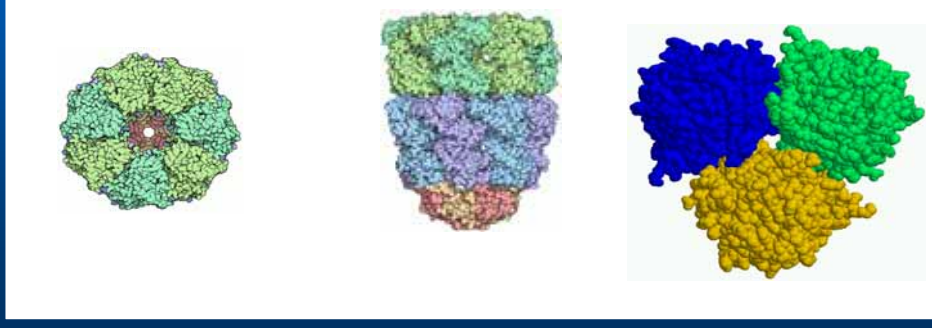
Format: ***[receptor_atom_index] [ligand_atom_index] [max_dist]***

For example:

25 377 5.0
340 5603 10.0



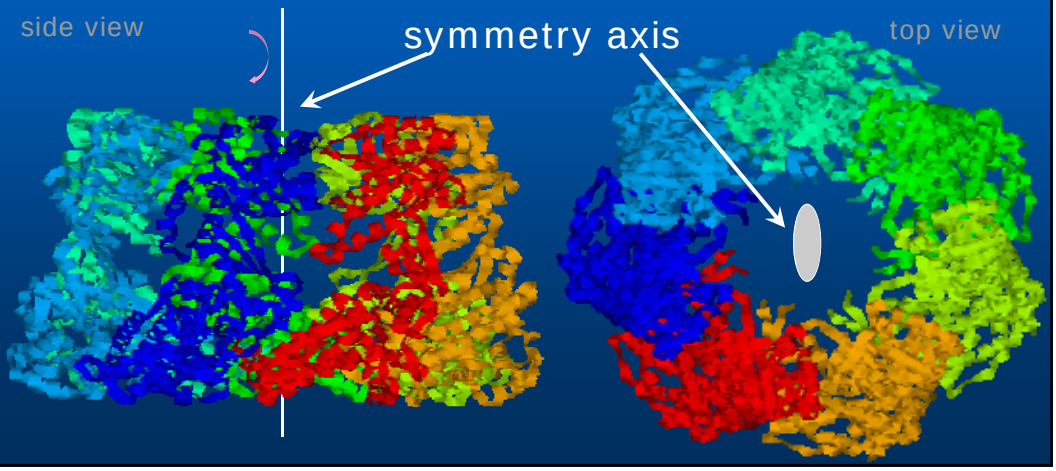
Symmetric Docking



Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ Proteins 05

Cyclic Symmetry

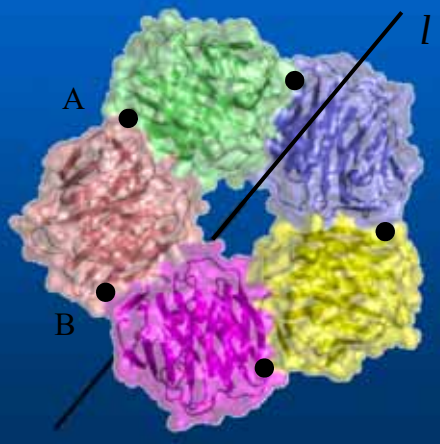
- Cyclic symmetry is defined by rotation of a single unit around an **axis**.
- The angle is determined by a number of units **n**.



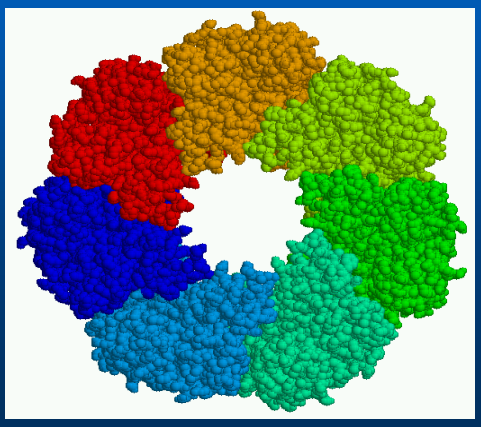
Geometrical view

$T(A) = B$

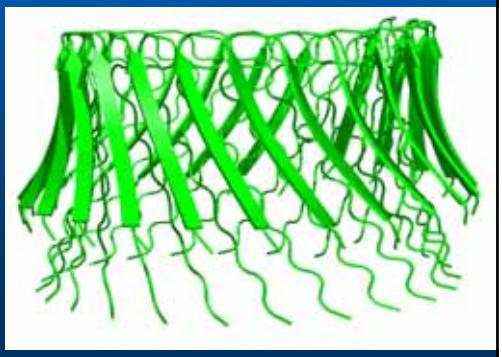
$l \cdot \overrightarrow{AB} = 0$



Chaperon: 2.5 Å RMSD prediction for the homo-heptamer.



Alzheimer's disease amyloid protein. Prediction of the membrane channel



N=24

WWW server: http://bioinfo3d.cs.tau.ac.il/SymmDock

SymmDock

3 mandatory fields

Unit Molecule: the asymmetric unit of the multimer complex (PDB format)

Type the PDB code [ ] (PDB-code:chain Id, e.g. 1f23:A)

or upload a file [ ] Browse...

Symmetry Order: [ ] (any number from 2 to 20: 3 for trimer, 4 for tetramer etc.)

e-mail address [ ] (the results are sent to this address)

Binding Site: [ ] Browse... binding site file (optional)

Submit Form Clear

Option to specify interface residues
(file format as in PatchDock)

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. NAR 2005

SymmDock Output Page

Top 20 solutions

Solution No	Score	Area	ACE	Transformation	PDB file of the complex
1	11858	1635.30	370.62	1.41 1.39 0.37 54.14 103.65 17.77	result.1.pdb
2	10512	1545.10	292.99	2.13 0.50 -1.39 42.10 38.18 -0.03	result.2.pdb
3	10468	1347.40	359.91	1.69 1.36 0.06 52.63 102.43 19.92	result.3.pdb
4	10284	1731.50	118.78	0.33 1.04 1.71 67.49 68.41 -23.84	result.4.pdb
5	9994	1808.40	228.47	1.59 1.28 0.26 45.22 104.90 15.19	result.5.pdb
6	9748	1420.40	641.49	2.17 -0.52 0.70 -54.29 101.23 -6.25	result.6.pdb
7	9478	1411.20	561.82	2.18 -0.53 0.68 -53.30 101.31 -2.72	result.7.pdb
9	9212	1528.30	761.87	2.03 0.43 0.17 9.12 119.03 -11.22	result.9.pdb
10	9186	1312.60	36.15	0.58 0.96 1.60 57.31 63.89 -28.67	result.10.pdb
12	9130	1290.40	221.23	1.40 1.13 0.66 24.37 95.76 8.39	result.12.pdb
13	9084	1105.10	319.00	2.41 0.83 -1.82 45.17 23.56 20.96	result.13.pdb
14	8922	1070.60	306.21	2.11 0.43 -1.24 40.73 44.84 -2.10	result.14.pdb
15	8888	1137.10	166.91	2.07 -0.14 0.72 -21.26 50.08 0.49	result.15.pdb
16	8818	1190.60	101.84	-1.01 -1.29 -2.49 -27.92 40.88 20.18	result.16.pdb
17	8798	1347.60	476.51	1.11 0.42 1.60 -1.25 53.16 -18.73	result.17.pdb
18	8784	1452.70	375.98	-0.53 0.79 2.18 86.39 90.79 21.56	result.18.pdb
19	8452	1410.70	499.66	2.08 -0.35 1.17 -52.48 58.60 19.25	result.19.pdb
20	8452	1134.10	342.38	0.31 0.04 2.07 81.88 66.44 -18.55	result.20.pdb

show next 20 »»

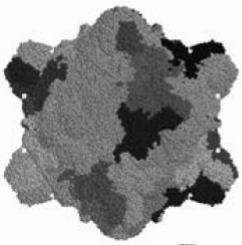
how many solutions to zip? 10 (any number from 2 to 100)

GO (this takes few seconds, please wait patiently)

Download output

Rigid-body docking: symmetry constraints

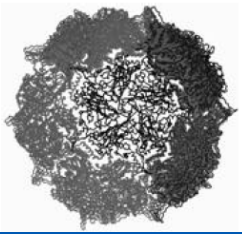
Virus structure prediction



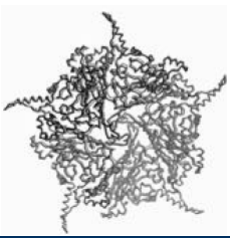
Icosahedron:
12 nodes,
20 faces,
30 edges
60 monomeric units

5-fold, 3-fold, 2-fold cyclic symmetry

Cyclic symmetry n=3



Cyclic symmetry n=5



Geometry-based docking: conclusions

- *Many available methods*
- *Very fast!*
- *In most of the cases, reasonable success rates*
- *Unbound much worse than bound*
- *Some times, missing native (even with bound subunits)*
- *Good as first docking step before refinement*