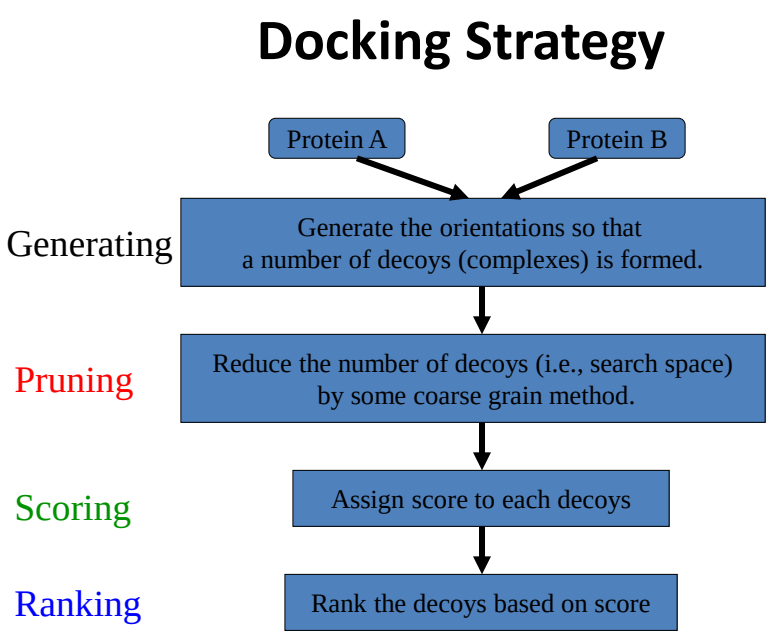
Lecture 19-21

Protein-protein docking



Scoring methods

Ab initio scoring

- Contact Area
- Contact Packing
- Non-bonded interactions
- Solvation Energy
- Etc.

Template based

Ab initio method

- Interface area (IA)
- Normalized interface packing (NIP)
- Normalized surface complementarity (NSc)
- Non-bonded energy (NE):

$$NE = \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{4\Pi\varepsilon R_{ij}}\right)$$

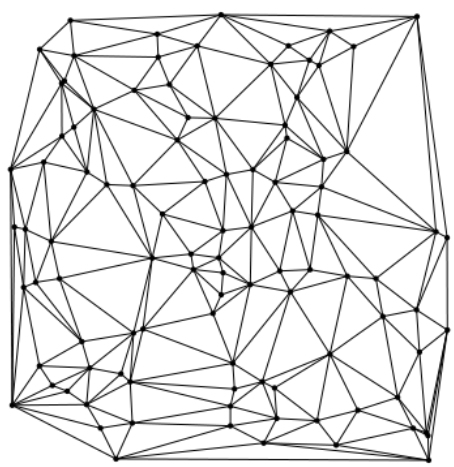
- Solvation energy (SE):*

$$SE = \sum_{\text{interface atoms}} \Delta\sigma(\text{Atom Type}) \times \Delta\text{ASA}$$

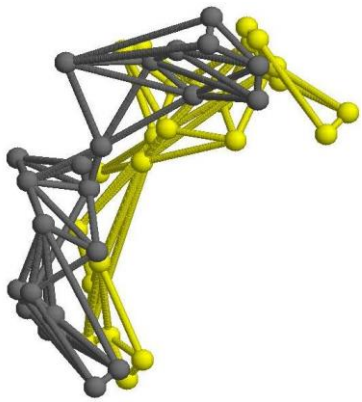
Interface area

- Accessible surface area:
  - The surface area of a molecule which is accessible by the solvent molecules (mostly water, the universal solvent molecule)
- Interface atom:
  - An atom is called as an interface atom if it loses its accessible surface area (ASA) by more than 0.1Å^2
- Interface area/Interacting surface:
  - The amount of ASA loses by all the interface atom is the interface area of a complex.

Surface complementarity

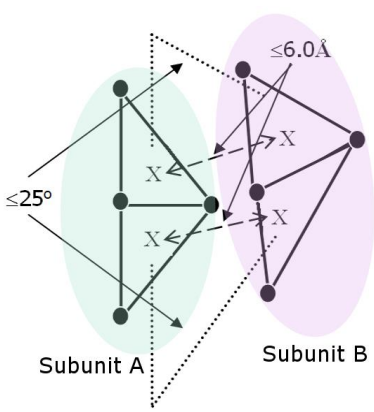


Surface complementarity



- Select the interface atoms
- Apply transformation
- Delineate interface atoms of each subunits by Delaunay triangulation

Surface complementarity

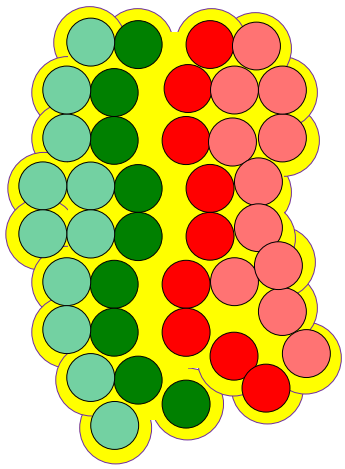


- Compute the complementarity between the triangles of the different subunits.

Surface complementarity = Complemented Area/Total Triangle area

Surface complementarity is divided by interface area to get **Normalized surface complemenarity** (***NSc***)

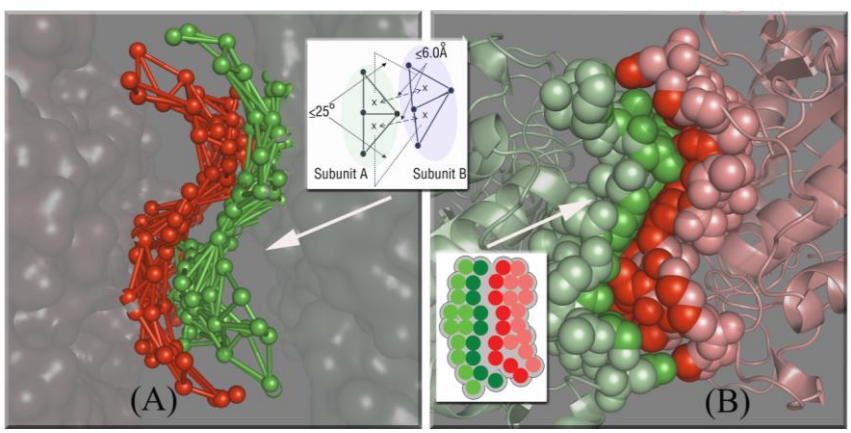
Packing at interfaces



Interface packing =
Actual Volume/Enclosed Volume

Interface packing is divided by interface area to get **Normalized interface packing** (***NIP***)

NSc and NIP at protein interface*



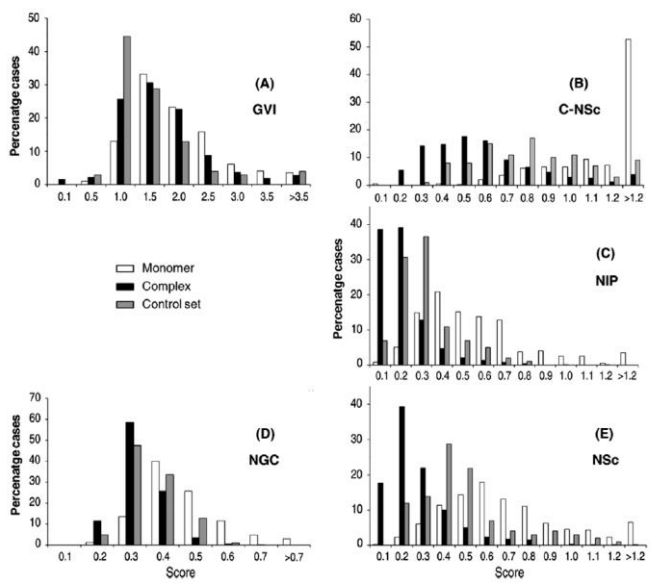
Correlation coefficient of NIP and NSc is **+0.95**

Data set

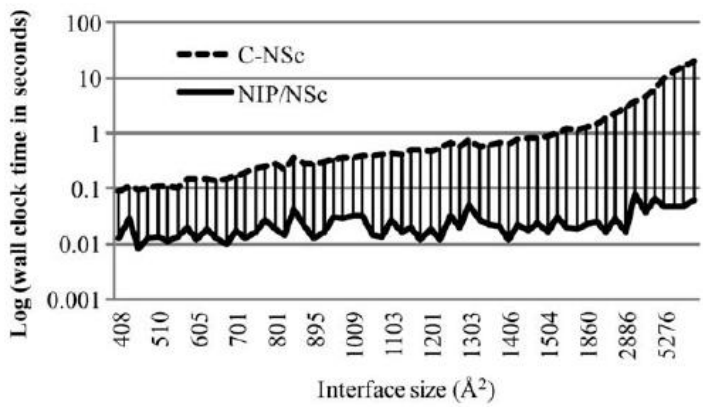
- The complex set
  - 906 unique protein dimers at <90% sequence identity was selected of which 800 were homodimers and the rest were heterodimers.
  - Literature, PiQSi database, and PQS database was used to verify that the interfaces chosen were biological.
  - Interface area range 240–7659 $Å^2$
  - More stringent non-redundant subsets at 60%, 40% and 35% sequence identity at the interface contained 855, 640, 118 numbers of complexes, respectively.
- The monomer set
  - 386 monomers was chosen from PiQSi with crystal contact area range 188-2111$Å^2$
- The control set
  - 100 protein complexes were arbitrarily chosen from the above *complex* set.

Resolution ≤ 2.5 Å and R-factor ≤0.2

Geometric Fitting

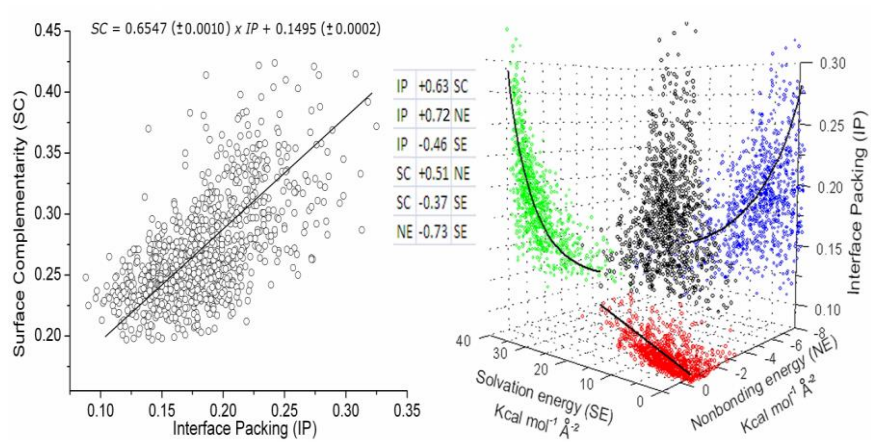


Efficiency



Scoring methods

Correlation among the four physico-chemical properties at the protein interfaces



Scoring and Ranking

Compute *IP*, *SC*, *NE* and *SE* at the decoy interface

↓

Group the decoys such that all decoys with RMSD<1.0Å and difference in SP<0.04 is in a group G, where SP = |SC-IP×0.6547-0.1495|

↓

Nonbonded energy for a group G: $NE^G = \overline{NE} - \sigma(NE)$
Solvation energy for a group G: $SE^G = \overline{SE} + \sigma(SE)$

↓

NE_i^G: NE^G bin number in all groups' *NE* histogram
SE_i^G: SE^G bin number in all groups' *SE* histogram

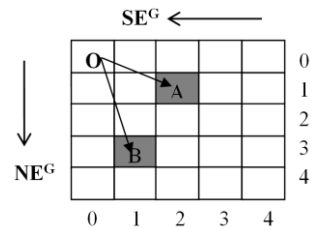
↓

$Score = \sqrt{((NE_i^G \times NE_i^G) + (SE_i^G \times SE_i^G))} + SP^G \times 10.0$
where, SP^G is minimum SP of the group G.

→ Sort (in ascending order) the group of decoys based upon their scores.

↑

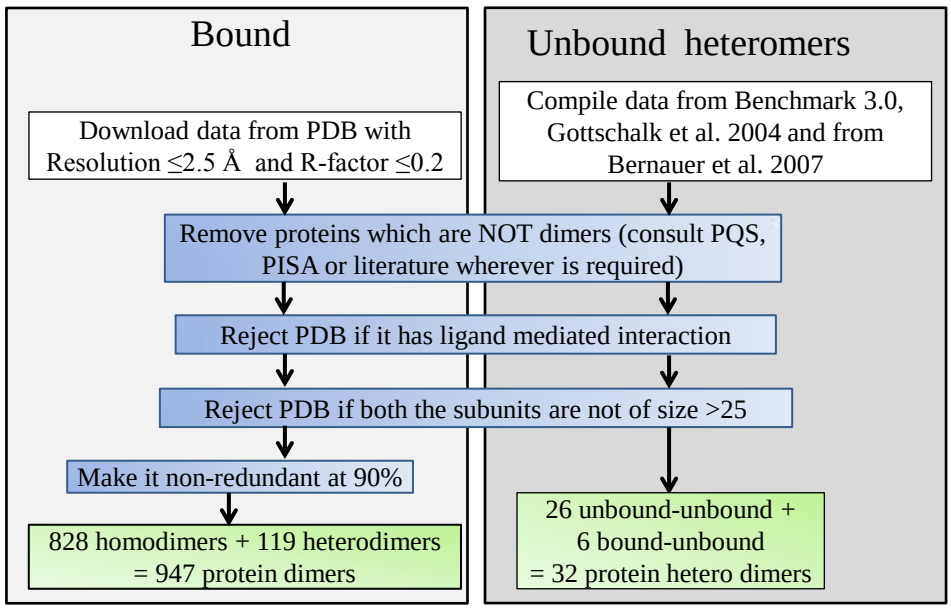
Rank of a decoy is its position in the sorted list



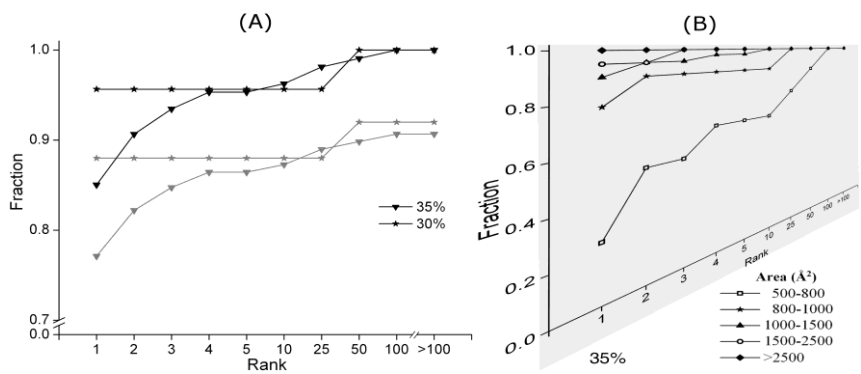
Docking types

- Bound docking
  - The crystal structure of complex is available. Interacting/docking partners are taken from that complex structure.
  - Easy to model since the side chain orientation is proper.
- Predictive/Unbound docking
  - The docking partners and complex structure is separately crystallized.
  - Side chain refinement is required

The Dataset

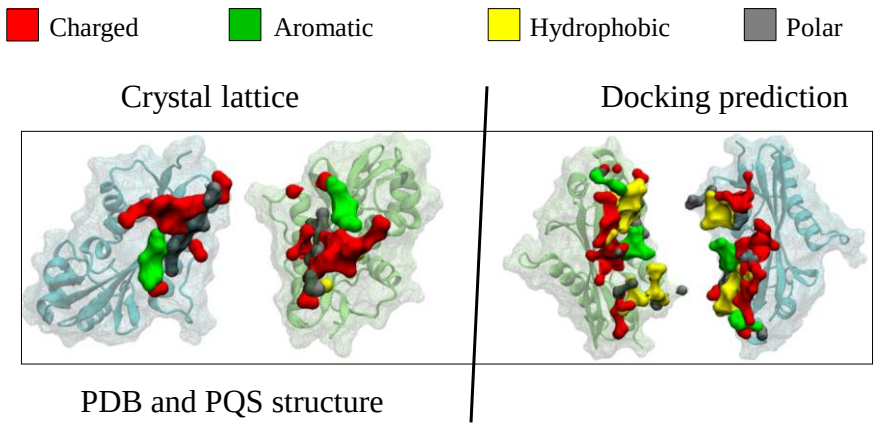


Evaluating bound dataset



(A) Variation of accuracy with rank . The darker curve shows the accuracy where the dimers could be successfully screened by IA filter. The lighter curve shows the accuracy over the whole dataset.
(B) Variation of accuracy with rank when the cases screened by IA filter was divided into various interface area categories.

Example prediction (PDB: 1EX2)



Residue property at the interface of the protein
- a conserved *Bacillus subtilis* protein Maf

ZRANK

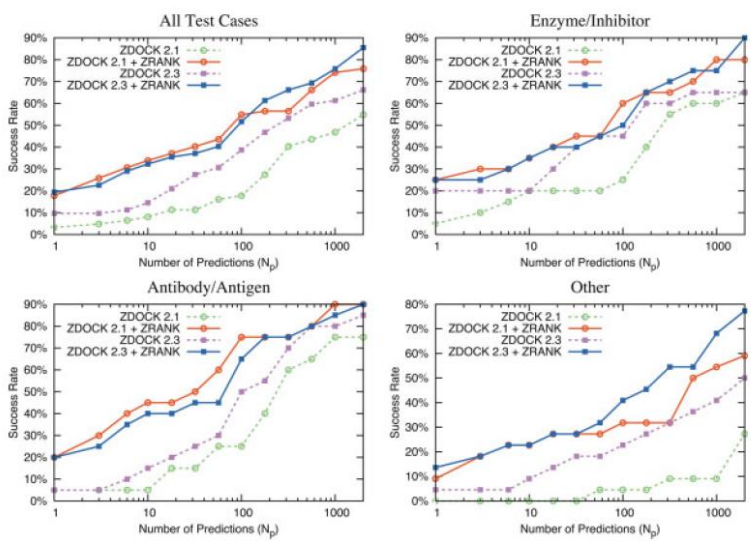
$$Score = w_{\text{vdW\_a}}E_{\text{vdW\_a}} + w_{\text{vdW\_r}}E_{\text{vdW\_r}} + w_{\text{elec\_sra}}E_{\text{elec\_sra}} + w_{\text{elec\_srr}}E_{\text{elec\_srr}} + w_{\text{elec\_lra}}E_{\text{elec\_lra}} + w_{\text{elec\_lrr}}E_{\text{elec\_lrr}} + w_{\text{ds}}E_{\text{ds}}$$

$$E_{\text{vdW}}(i,j) = \varepsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right)$$ Van der Wall interaction

$$E_{\text{elec}}(i,j) = 332\frac{q_i q_j}{r_{ij}^2}$$ Electrostatic Interaction

$$E_{\text{ds}}(i,j) = a_{ij}$$ Desolvation energy

ZRANK



PatchDock and FireDock

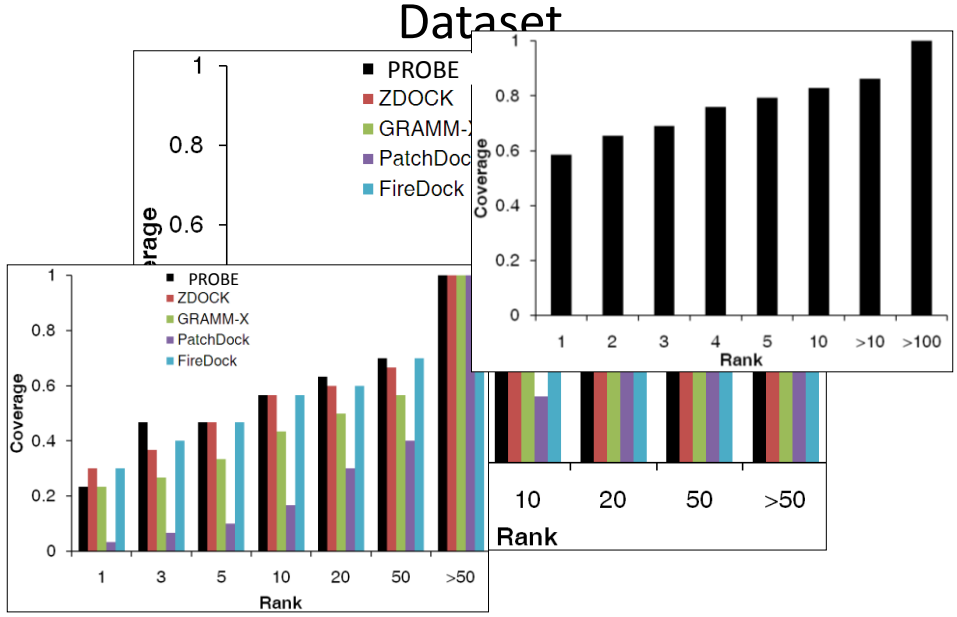
- **PatchDock:** Molecular Docking Algorithm Based On Shape Complementarity Principles
- **FireDock:** Includes three main steps:

(1) Side-chain optimization: The side-chain flexibility of the receptor and the ligand is modeled by a rotamer library. The optimal combination of rotamers for the interface residues is found by solving an integer LP problem.

(2) Rigid-body minimization: This minimization stage is performed by a MC technique that attempts to optimize an approximate binding energy by refining the orientation of the ligand structure.

(3) Scoring and ranking: This final ranking stage attempts to identify the near-native refined solutions. The ranking is performed according to a binding energy function that includes a variety of energy terms: desolvation energy, van der Waals interactions, partial electrostatics, hydrogen and disulfide bonds, *p*-stacking and aliphatic interactions, rotamer's probabilities and more.

Predictive Docking - The Unbound Dataset



Classification of Docking Algorithms According to Function Parameters

Algorithm name (reference)	Scoring stage in algorithm flow	Reference for the solution	Geometric complementarity	Hydrogen bonds	Contact area	Intramolecular overlap
Sobolev et al.[241]	Integrated	Self	+	+	+	−
SP-DOCK Fradera et al.[186]	Edge	Known structure	+	+	−	+
SG-DOCK Fradera et al.[186]	Integrated	Known structure	+	+	−	+
Norel et al.[230]	Edge	Self	+	−	−	−
FTDOCK, Katchalski-Katzir et al.[142]	Edge	Self	+	−	−	−
Fischer 1995	Integrated	Self	+	−	−	−
DARWIN, Burnett and Taylor [34]	Integrated	Self	−	+	−	−
PUZZLE, Helmer-Citterich et al.[205]	Edge	Self	−	−	+	−
Hybrid algorithm, Hou et al.[242]	Integrated	Self	+	+	−	−
Gardiner et al.[77]	Integrated	Self	+	+	−	−
Jackson et al.[115]	Edge	Self	+	−	−	−
Norel et al.[40]	Edge	Self	+	−	−	−
ESCHER, Ausiello et al.[38]	Integrated	Self	+	+	−	−
Camacho et al.[148]	Integrative	Self	−	−	−	−
BiGGER, Palma et al.[113]	Integrative	Self	+	−	−	−

Classification of Docking Algorithms According to Function Parameters

Algorithm name (reference)	Intermolecular overlap	Pairwise amino acid contacts	Electrostatic interactions	Solvation energy	Active site residues	Free energy
Sobolev et al.[241]	+	−	+	−	−	−
SP-DOCK Fradera et al.[186]	+	−	+	−	−	−
SG-DOCK Fradera et al.[186]	+	−	+	−	−	−
Norel et al.[230]	−	−	+	−	−	+
FTDOCK, Katchalski-Katzir et al.[142]	−	−	−	−	−	−
Fischer 1995	+	−	−	−	−	−
DARWIN, Burnett and Taylor [34]	+	−	+	+	−	+
PUZZLE, Helmer-Citterich et al.[205]	+	−	−	−	−	−
Hybrid algorithm, Hou et al.[242]	+	−	+	−	−	−
Gardiner et al.[77]	+	−	−	−	−	−
Jackson et al.[115]	+	+	−	−	+	−
Norel et al.[40]	+	−	+	−	−	−
ESCHER, Ausiello et al.[38]	+	−	+	−	−	−
Camacho et al.[148]	−	−	+	+	−	+
BiGGER, Palma et al.[113]	+	+	+	+	−	−

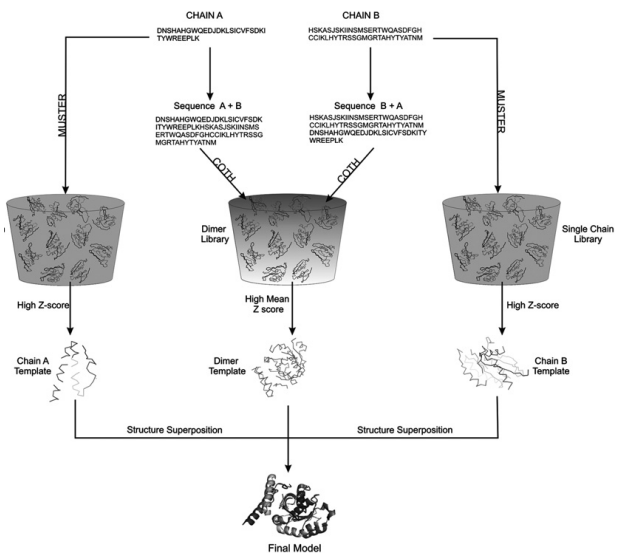
Critical Assessment of PRediction of Interactions (CAPRI)

Critical Assessment of PRediction of Interactions (CAPRI)

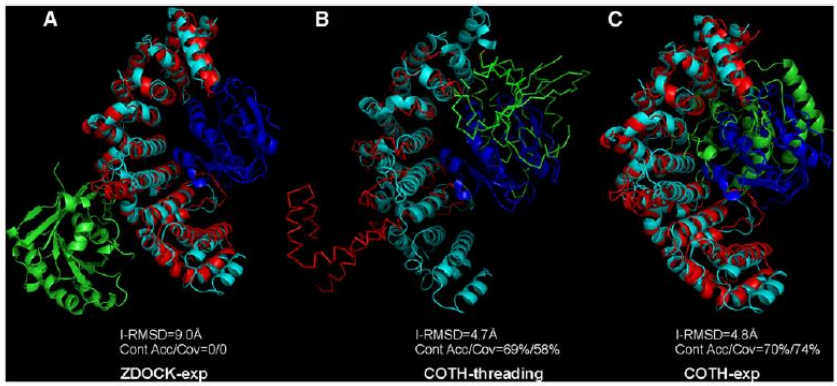
Predictor	*Affiliation*	*Software*	*Algorithm*
Abagyan	Scripps	ICM	Force Field
Camacho/Vajda	Boston	CHARMM	Force Field Refinement
Gardiner	Sheffield	GAPDOCK	Shape+Area GA
Sternberg/Smith	Imperial	FTDOCK	FFT
Bates/Fitzjohn	ICRF	Guided Docking	Force Field
Ten Eyck/Mitchell	SDSC	DOT	FFT
Vakser/Tovchigrechko	SUNY/MUSC	GRAMM	FFT
Olson	Scripps	Harmony	Spherical Harmonics
Weng/Chen	Boston	ZDOCK	FFT
Eisenstein	Weizmann	MolFit	FFT
Wolfson/Nussinov	Tel Aviv	BUDDA/PPD/FireDock	Geometric Hashing
Iwadate	Kitasato	TSCF	Force Field+Solvent
Ritchie/Mustard	Aberdeen	Hex	Spherical Polar Fourier
Palma	Lisbon	BIGGER	Geometric+Electrostatic
Gray/Baker	Washington/JHU	RosettaDock	Monte Carlo+Flexibility
Mitra and Pal	IISc, Bangalore	PROBE/PRUNE	FFT

Docking from sequence
Application to Genome-wide scale

COTH – docking from sequence



COTH – docking from sequence



The native complex (Ran-Importin β complex) is represented in cyan.

Parallel Implementation

- At the generation phase:
  - The protein can be divided into different parts that are mutually exclusive.
- At the scoring phase:
  - All the decoys are mutually independent; thus they can be processed separately on different processors.

Summary

✓ The bound test set is easy to predict, but the real benchmark set is unbound data set.

✓ Refining the side chain of the unbound docked complexes are still an active area of research.

✓ Computationally flexible docking is more challenging than rigid body docking.

✓ Evolutionary information can be integrated to improve the performance of the method.