Docking and macromolecular complexes

Structural Bioinformatics

Outline

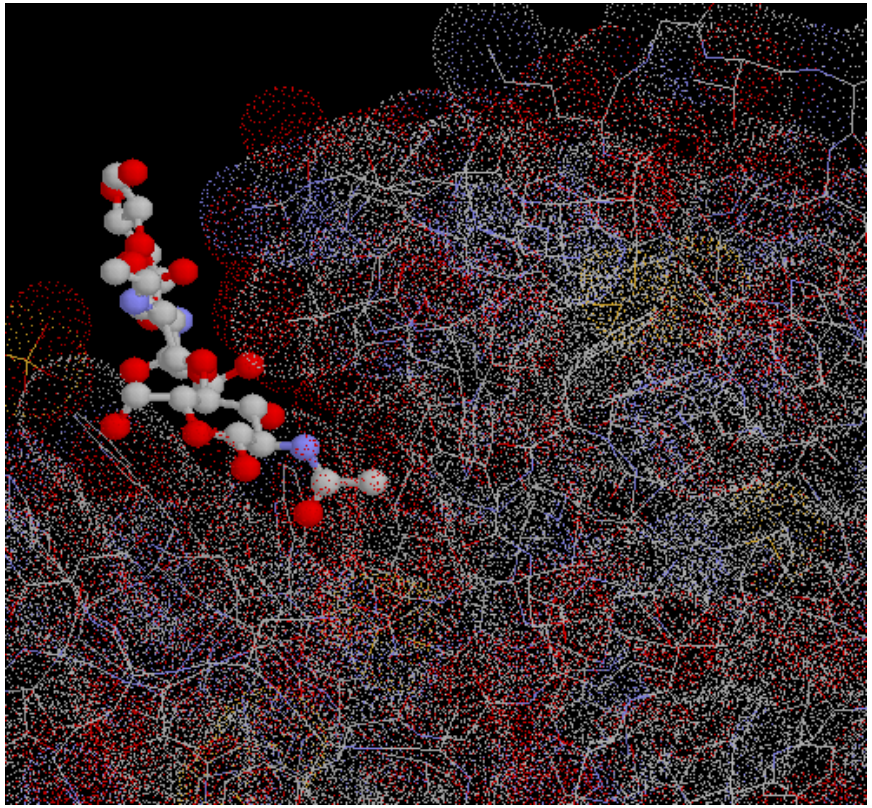
- Some Reminders
- Concepts, definitions
- Small-molecule docking
- Protein docking
  - Ab-initio docking
  - Data-driven docking
  - Assessment

Molecular recognition

- All macromolecules work through recognition processes
  - Protein-ligand.
    - Enzymes, membrane receptors, transport proteins, Drugs
  - Protein-protein
    - Regulation of enzyme activity (signal transduction), multi-subunit protein and complexes
  - Protein-NA
    - NA metabolism, gene regulation
  - NA-Ligand
    - Drugs
  - NA-NA
    - Transcription, replication, protein synthesis, ...
- Recognition is selective. It depends on the participating groups.
- Recognition is dynamic
  - Induced fit / Conformational selection
  - Complexes can be permanent, but most of them are transient

Energy considerations

- Entropic
  - Conformational, hydrophobic
- Enthalpic
  - Vdw
    - Shape and contacts
  - Hbond
    - Define geometry, must be complete
  - Electrostatic
    - Severe solvation penalty



Lisozyme + $(NAG)_3$

- **Structural complementarity**

- Complex formation implies to bury new interactions
  - Unstable (hydrophobic) surfaces in water may indicate binding regions.

Proteins do not act alone

- Protein – protein complexes:
  - Permanent associations (“quaternary structure”)
  - Enzyme – substrate (transient)
  - Regulatory associations
  - Multi protein clusters/groups
    - (Nuclear porus, cytochroms, …)
- Protein – NA complexes
  - Transcription factors
  - Replication, Splicing, Transcription machineries
  - Ribosomes
- …

Concepts, definitions

- Docking: Prediction of the structure of complexes
  - Ligand-protein docking, protein docking
- Receptor, ligand: the actors
- Pose: Binding mode
  - how the ligand positions on the receptor site
- Interface: Contact region in a complex

Concepts, definitions

- Binding energy (does not need presentation)
- Scoring: How the binding is evaluated (tries to approach to the binding energy)
- (Virtual | Reverse) screening: test for the feasibility of binding among a high number of ligands/receptors
- Docking evaluation: How to test the success of the docking

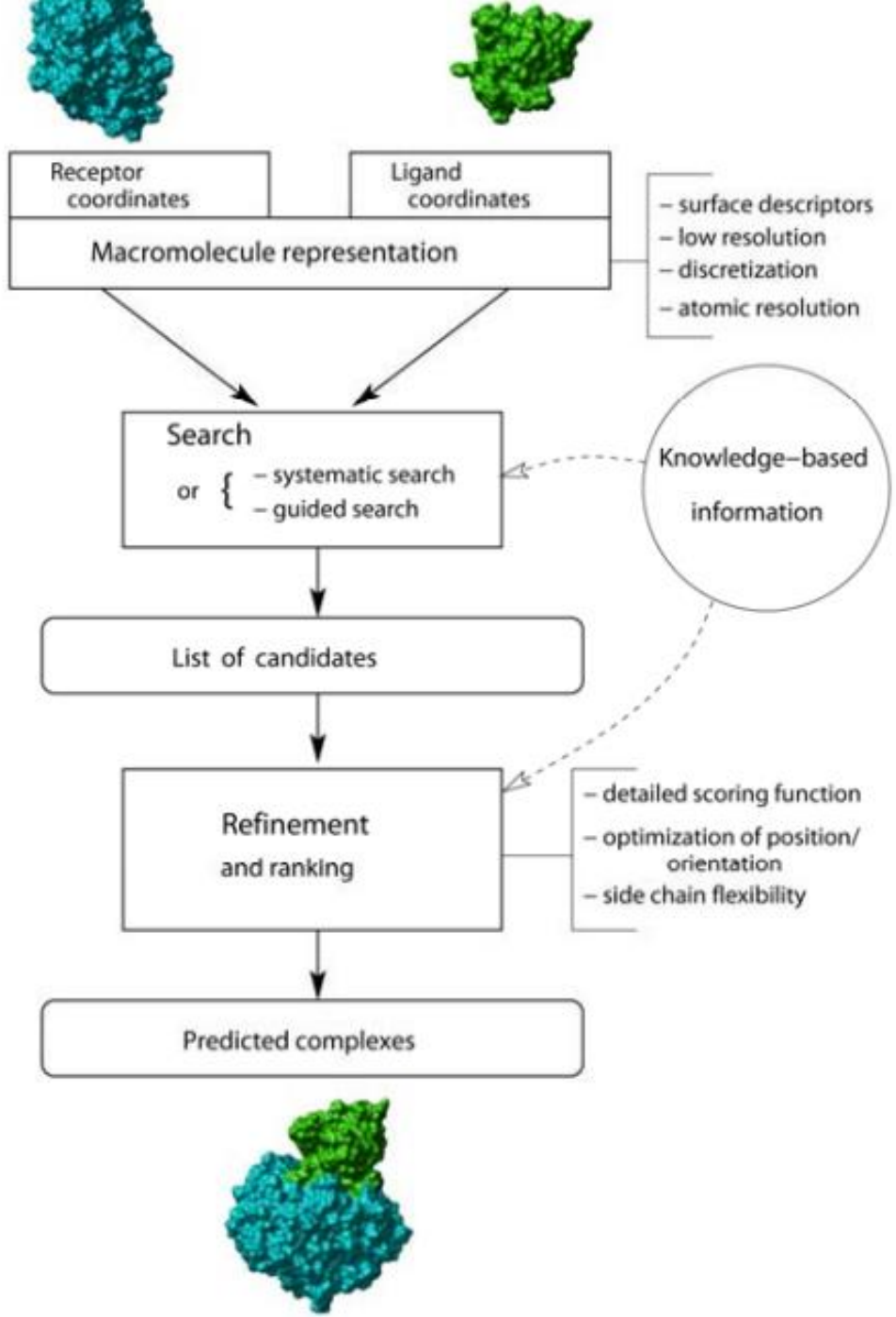
Protein docking terms

- Interaction prediction
  - Whether two protein interact to each other (no 3D structure)
- Interface prediction
  - Determine regions where complex components interact
- Protein docking
  - Predict poses (structure of the complex)

- Bound *vs.* Unbound docking
  - Docking using conformations in the complex (bound) or free (unbound)
- Flexible *vs.* Rigid docking
  - Whether proteins' flexibility is taken into account
- Local *vs.* global docking
  - Whether binding site is roughly known

Docking stategy

- Protein representation
- Search method
- Scoring method
- Refinement?

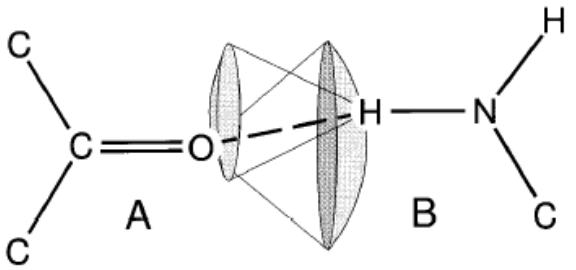


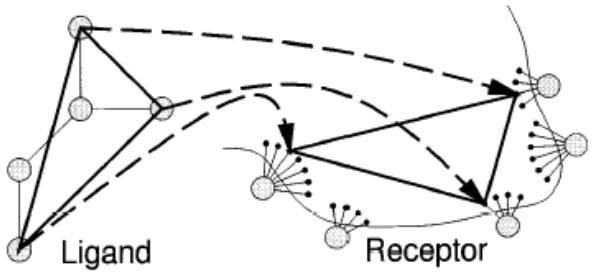
Ligand-Protein Docking

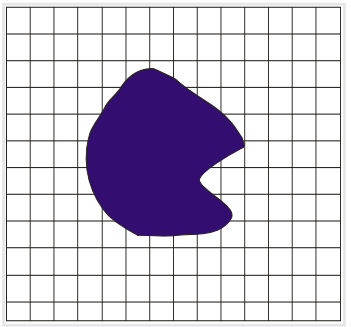
- **Molecular Docking**
  - Prediction of 3D structure of ligand-protein or protein-protein complexes.
  - One receptor – one or few ligands
  - **Quality** of the structure is the main objective. Realistic binding energies
  - Usually combined with other techniques, as MD
  - Experimental information can be considered
- **Virtual screening**
  - Identification of possible ligands from compound databases
  - One receptor – multiple ligands (> 10^6)
  - Calculation should be fast >10000 ligand-receptor dockings / day / proc.)
  - The main objective is to **select "some" ligands**, that can be optimized with other methods
- **Reverse Screening**
  - Identification of possible receptors for a known ligand
  - One ligand - multiple receptors
  - Points to possible side effects
- **Active site prediction**
  - Identification of binding regions

Ligand-Protein Docking

- Complete atomic representation
  - Large cost, high resolution
- Simplified representations
  - Quick and robust. Low resolution.
- 3D Grid representations (receptor)
  - Easier energy calculation
  - Definition of “pharmacophores” (MIPs)
- Flexibility (Ligand and receptor)
  - Ensembl docking







Scoring

- When to score
  - Score associated to the search process
  - **Scoring *a posteriori***
- Structural complementarity
  - Robust, low resolution
- Classical force-fields, Statistical Pot.
  - High resolution
  - Easy to transfer
- Empirical functions
  - ΔG_{bind} obtained from function fitted to experimental data

$$\Delta G = \Delta G_0 + \Delta G_{rot} \times N_{rot} \quad (1)$$

$$+\Delta G_{hb} \sum_{\text{neutral H-bonds}} f(\Delta R, \Delta\alpha) \quad (2)$$

$$+\Delta G_{io} \sum_{\text{ionic int.}} f(\Delta R, \Delta\alpha) \quad (3)$$

$$+\Delta G_{aro} \sum_{\text{aro int.}} f(\Delta R, \Delta\alpha) \quad (4)$$

$$+\Delta G_{lipo} \sum_{\text{lipo. cont.}} f^*(\Delta R) \quad (5)$$

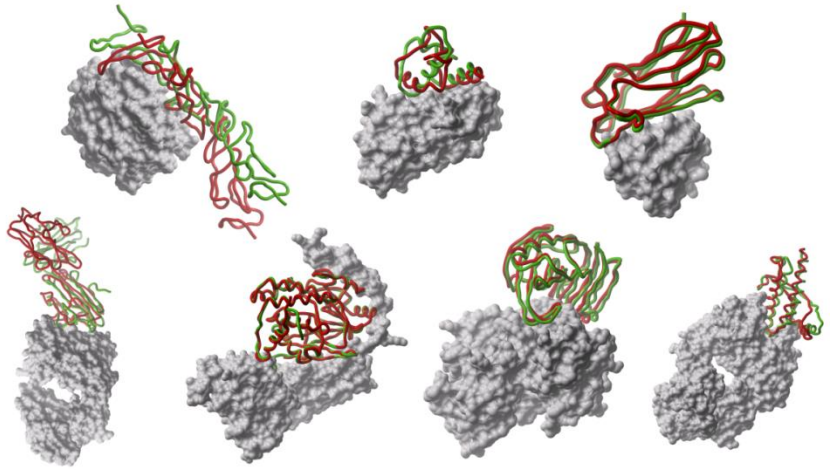
Ligand protein software

- https://en.wikipedia.org/wiki/List_of_protein-ligand_docking_software

- Most popular:
  - DOCK
  - AutoDock (Vina)
  - GOLD
  - FlexX
  - Glide (Commercial)

Protein Docking

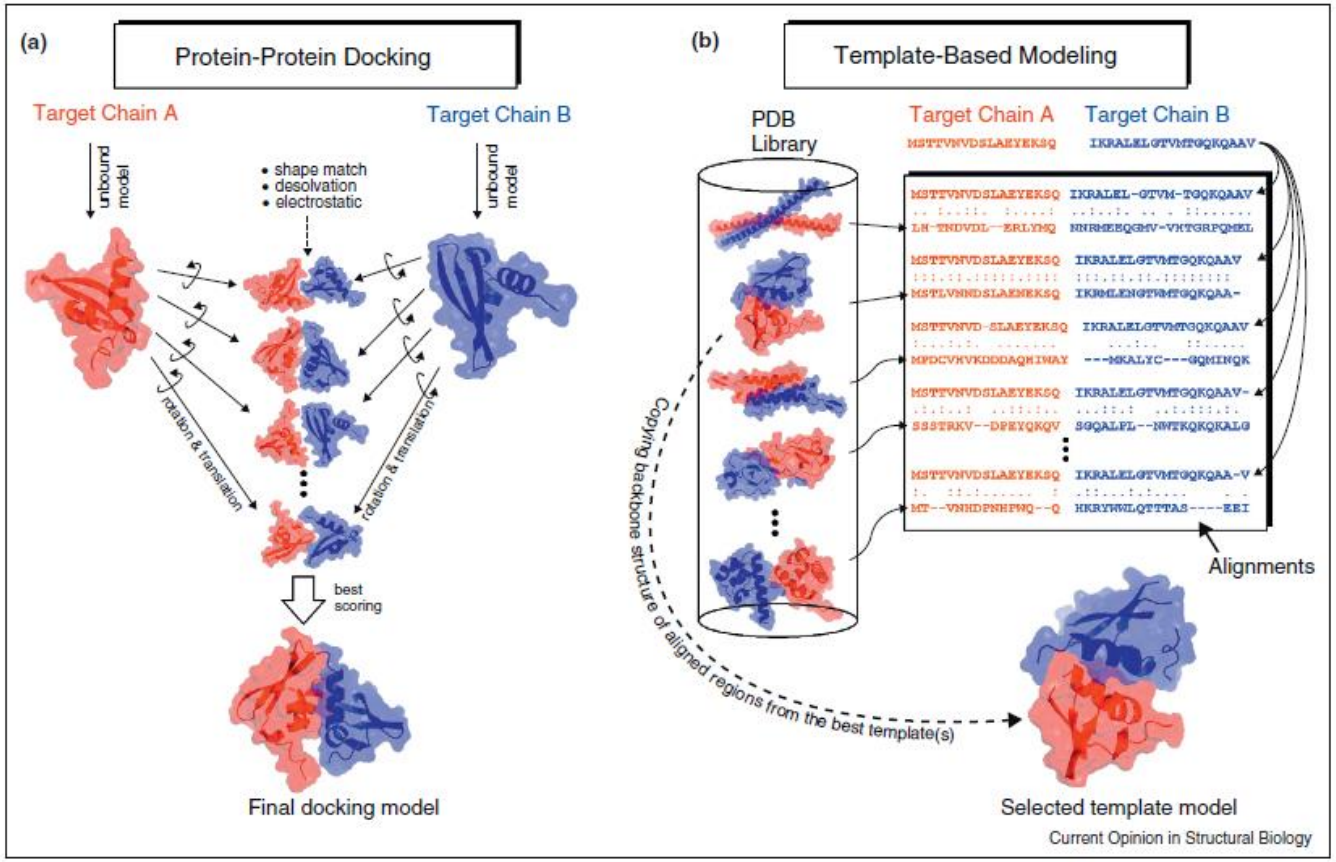
- Large systems
  - Typical contact area: 1,500-3,000 Å^2
- Large number of degrees of freedom and large conformational flexibility
- Few “easy” interactions
  - Hydrophobic contacts
  - Average 1 H-bond / 170 Å^2 (*)
  - 1 water / 100 Å^2 (*)



(*) Lo Conte et. al. *J. Mol. Biol.* (1999) 285, 2177

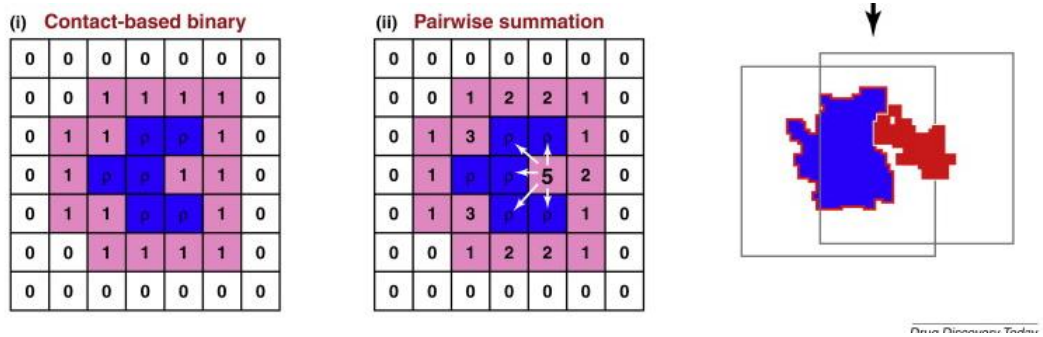
Basic strategies

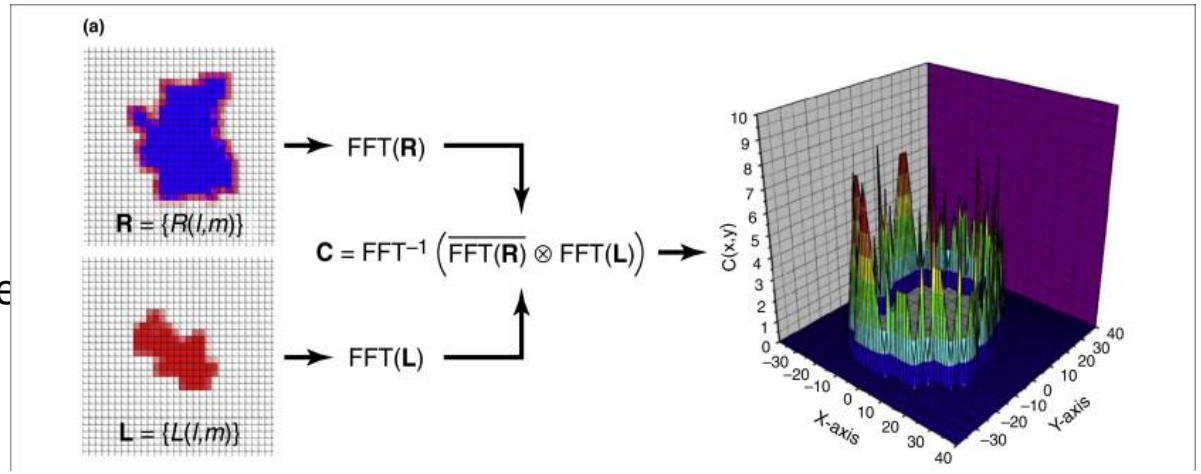
- "Pure" *Ab Initio* docking
  - Only information about ligand and receptor structures is known
    - Pseudo-random approaches (simulation, optimization)
    - Directed search (Geometric hashing)
    - Brute-Force approaches (Grid-based, FFT)
- Data driven docking (template based)
  - Experimental, homology data is used
  - Machine-learning methods
  - Co-evolution methods
  - Can be combined to help *ab initi*o approaches

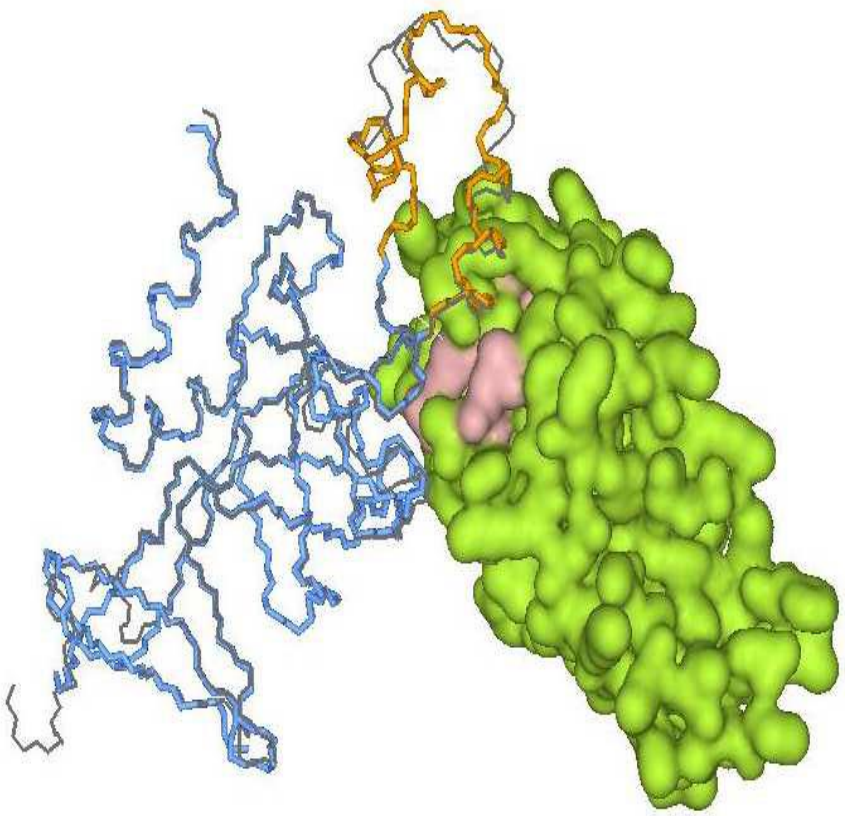


Ab initio Rigid-Body docking (1)

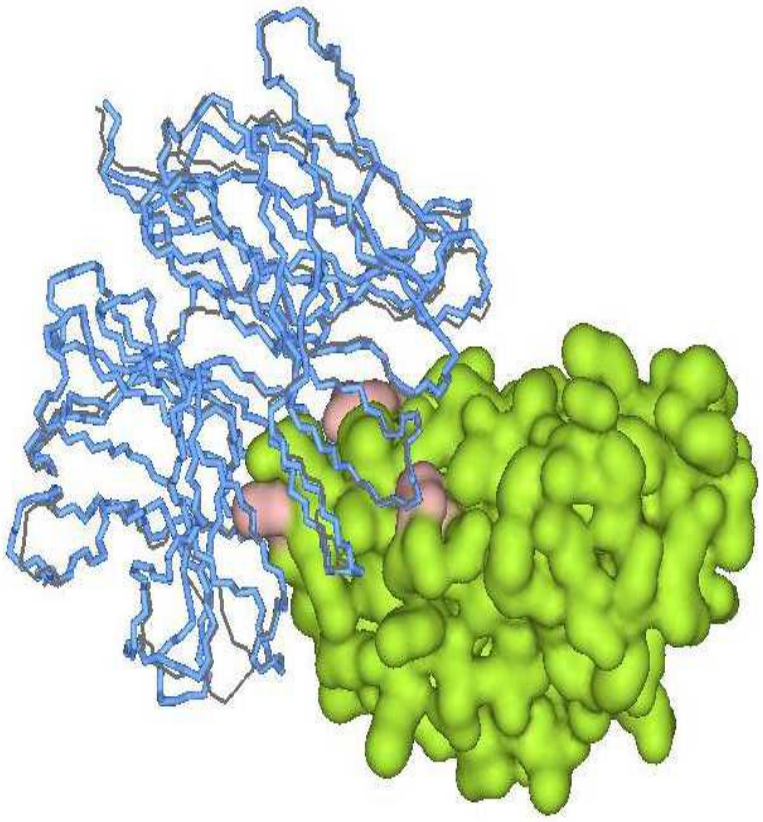
- Proteins are mapped onto 3D grids. Each grid point is evaluated as inner (<0 for receptor, >0 for ligand), Surface (1), outer (0).
- Blind 6Dim (3 translations x 3 rotations) search
- Score is based in 3D complementarity: i.e. matches among the "Surface" points (calculated with from the product of grid points)
- Fast Fourier Transforms to speed up translational (Fast Fourier Transform, FFT) or rotational (Spherical Polar Fourie SPF) searches.
- Computational cost can decrease by >10^4 (from N^6 to $N^3 \ln N^3$)







Orc1/Sir1 “Difficult” , Interface: 1,300Å^2

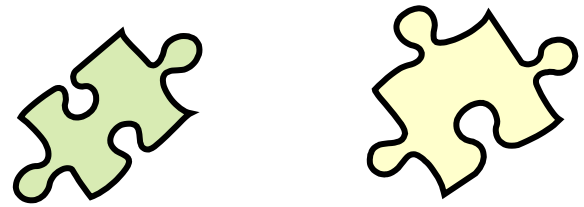


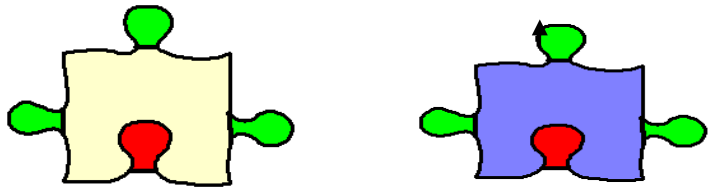
TolB/Pal “Easy” Interface: 2,600Å^2

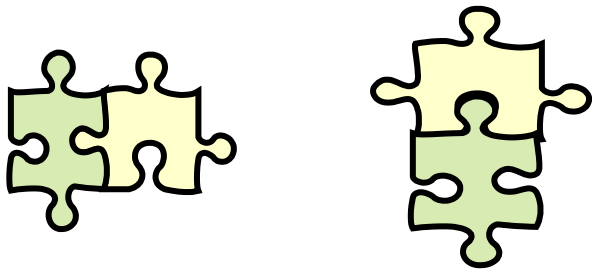
Complexes with larger buried interfaces are easier to predict

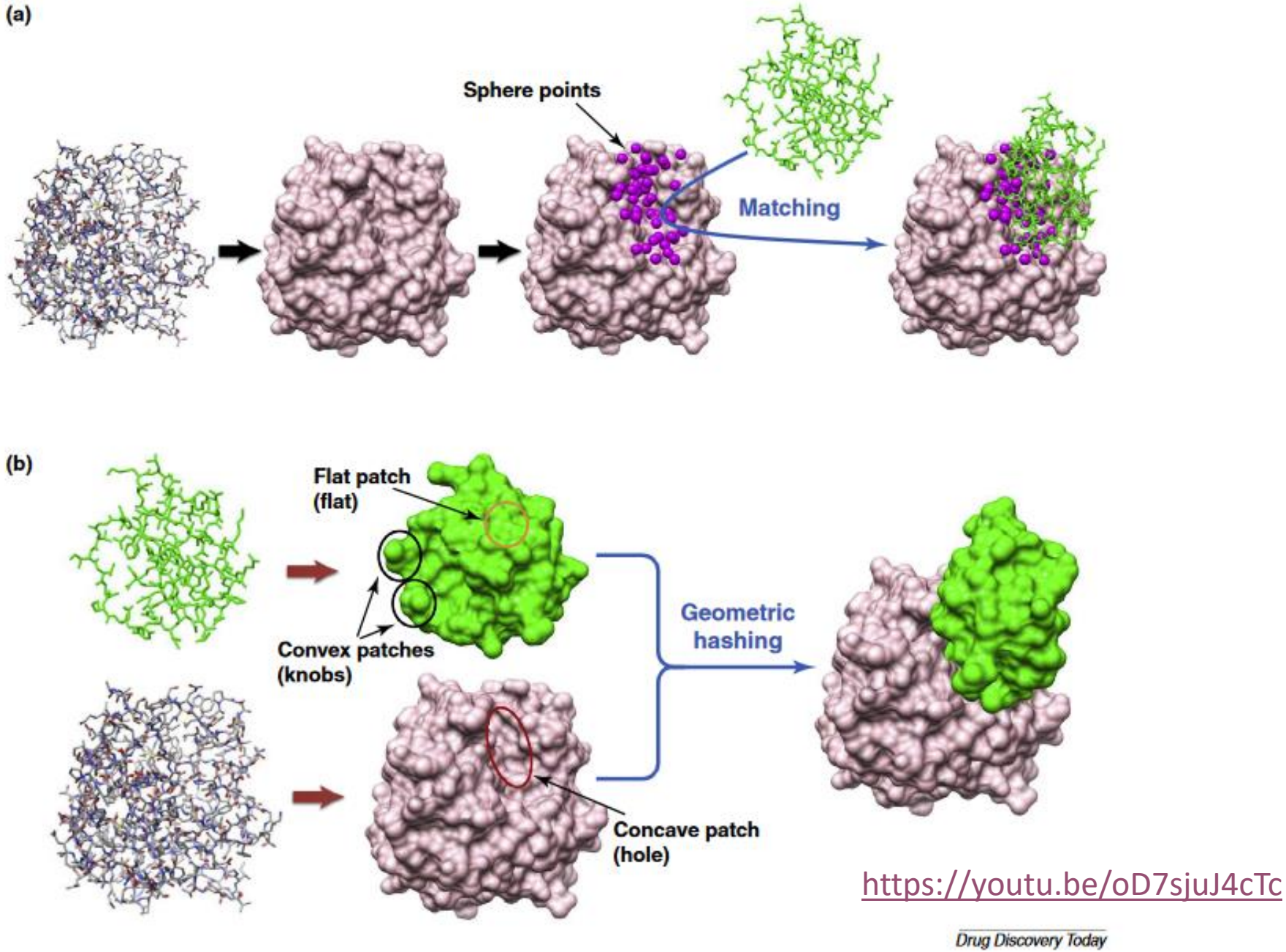
Ab initio Rigid-Body docking (2)

- Geometric hash, Surface are pre-processed to detect possible matching regions
- Solvation/desolvation can be mapped into Surface properties
- Most favourable poses are re-scored with better scoring functions
- Less efficient for pure blind docking (better with additional information)





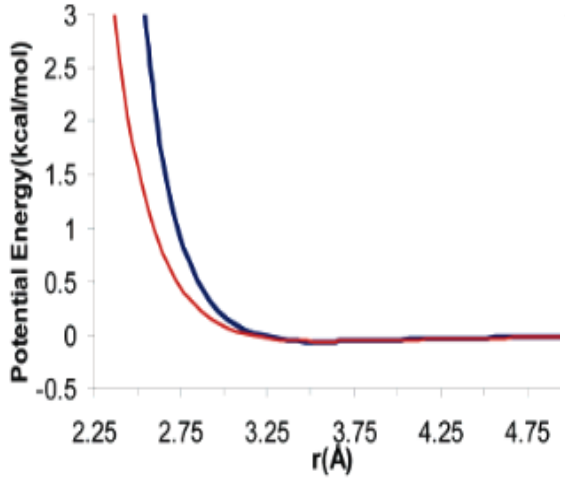


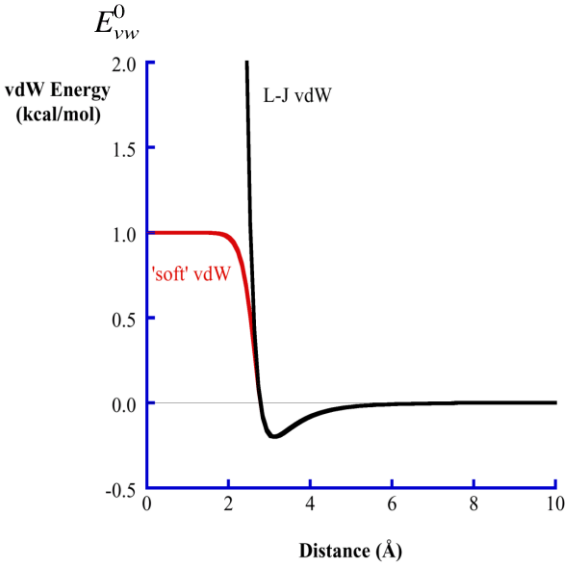


https://youtu.be/oD7sjuJ4cTc

Additional tricks

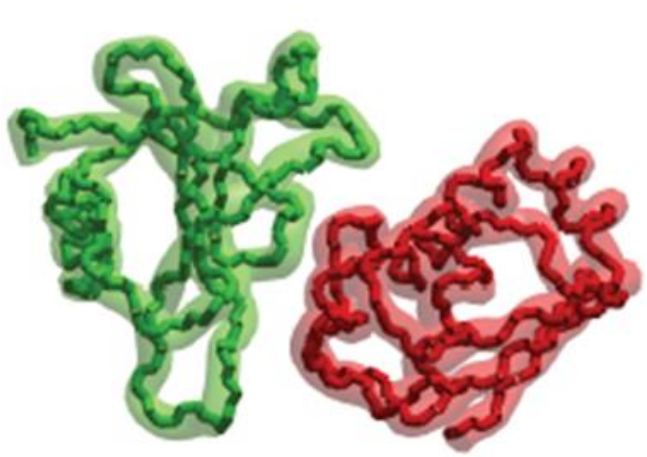
- Flexible docking
  - Traditional MD
    - Too costly (limited sampling)
    - Used to refine structures after docking
  - Conformational sampling
    - Rigid docking with set of posible conformations (experimental or produced from PCA)
    - Conformational search added to position and orientation (usually MC)
  - Combined cycles of docking and simulation
  - RosettaDock combines rigid body MonteCarlo for orientation/traslation + MonteCarlo among rotamer libraries (very expensive)

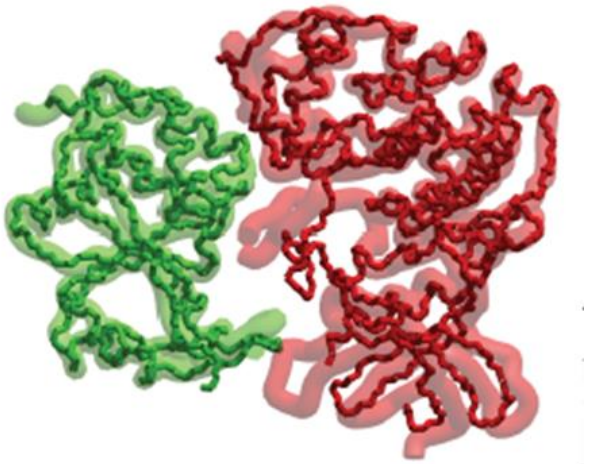


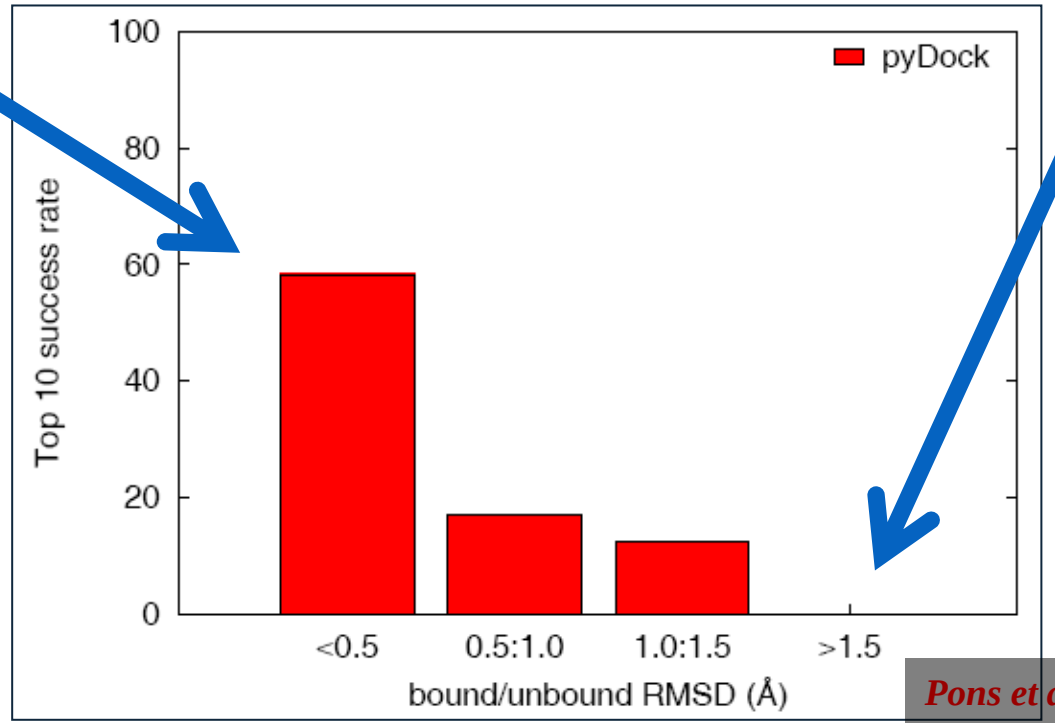


- Soft docking
  - Backbone is still rigid
  - Sidechain flexibility is mimicked using “soft” VdW potentials, and/or coarser FFT grids

Rigid-body limitations: flexible cases



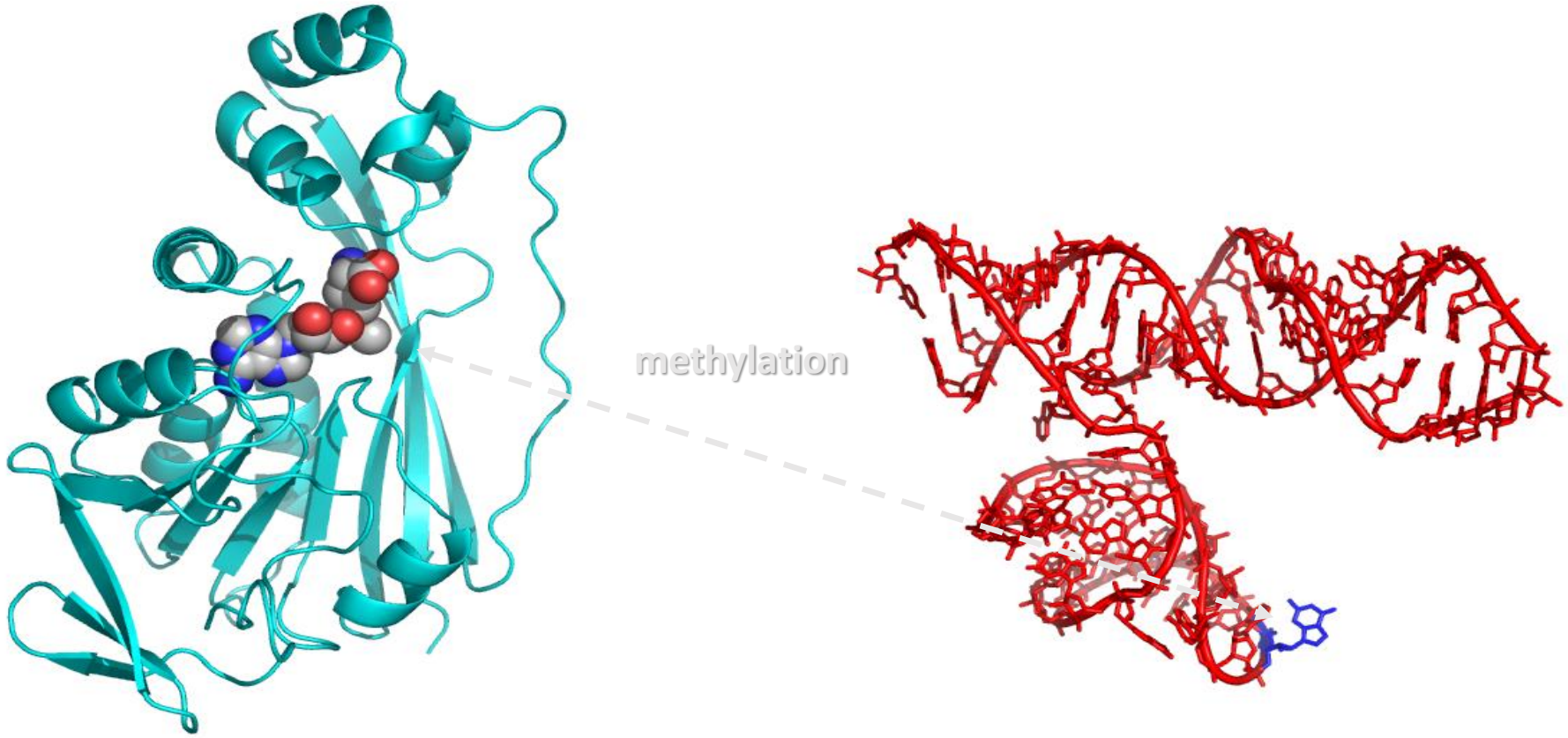




Pons et al. (2010) Proteins 78, 95-108

Docking with Distance Constraints

A complex of Rlma2 methyltransferase of *S. pneumoniae* and a 74 nucleotide RNA transcript



Ab-Initio Protein Docking Software

- FFT-based search
  FTDock, GRAMM, DOT, **ZDOCK**, MolFit, PIPER, F2DOCK, SDOCK, ASPDock, Cell-Dock
- Spherical Fourier transform-based search
  HEX, FRODOCK
- Direct search in Cartesian space
  SOFTDOCK, BIGGER, SKE-DOCK
- Local shape feature matching Distance geometry algorithm
  DOCK
- Geometric hashing
  PatchDock, SymmDock, LZerD
- Genetic algorithm
  GAPDOCK
- Randomized search Monte Carlo search
  RosettaDock, ICM-DISCO, ATTRACT, **HADDOCK**
- Particle swarm optimization
  SwarmDock, LightDock
- Genetic algorithm
  AutoDock
- Post-docking approach using advanced scoring functions
  RPScore, ZRANK, **PyDock**, EMPIRE, DARS, DECK, SIPPER, PIE, MDockPP, etc.
- Considering protein flexibility
  MultiDock, SmoothDock, **RDOCK**, FireDock, FiberDock, EigenHex

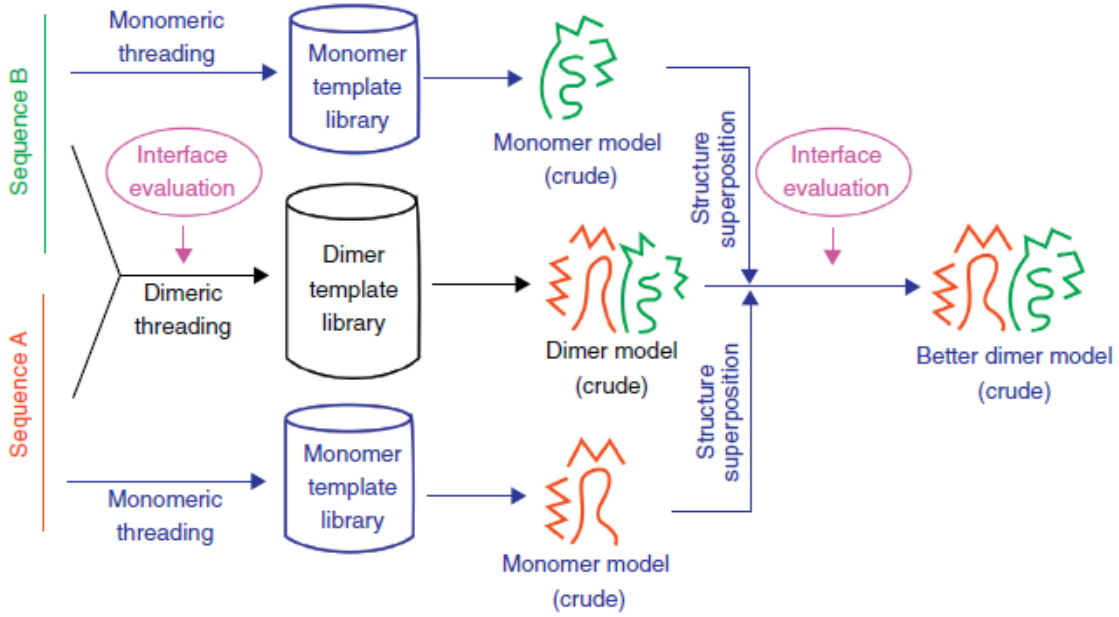
Data – driven methods

- Homology/threading based methods
  - Template based, use data from homologues

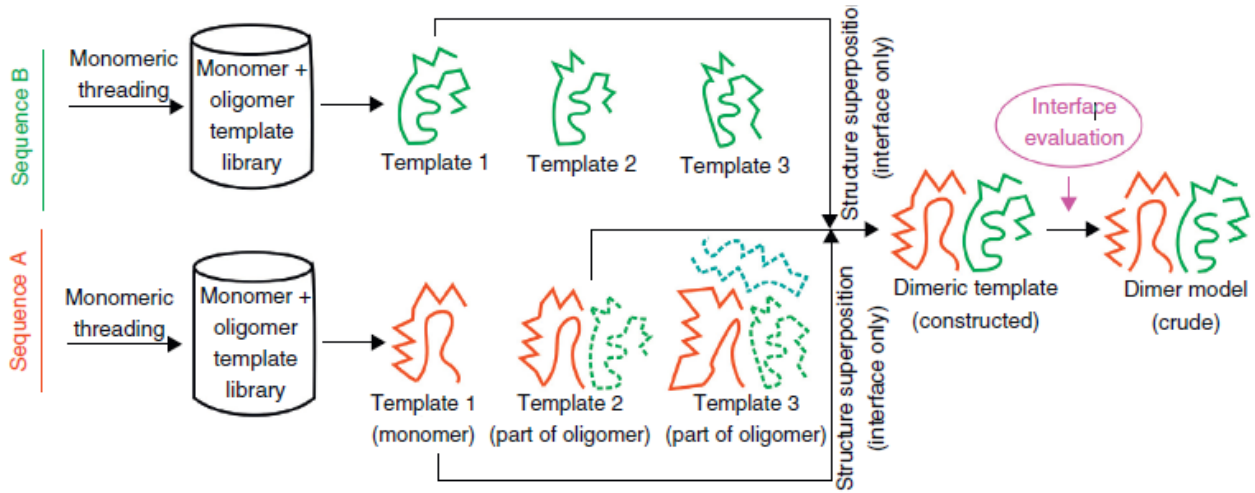
- Co_evolution methods
  - Growing popularity in protein structure prediction (Alphafold)
  - Uses data from “massive” multiple sequences alignment

- Interface prediction
  - To reveal interfaces without structure prediction

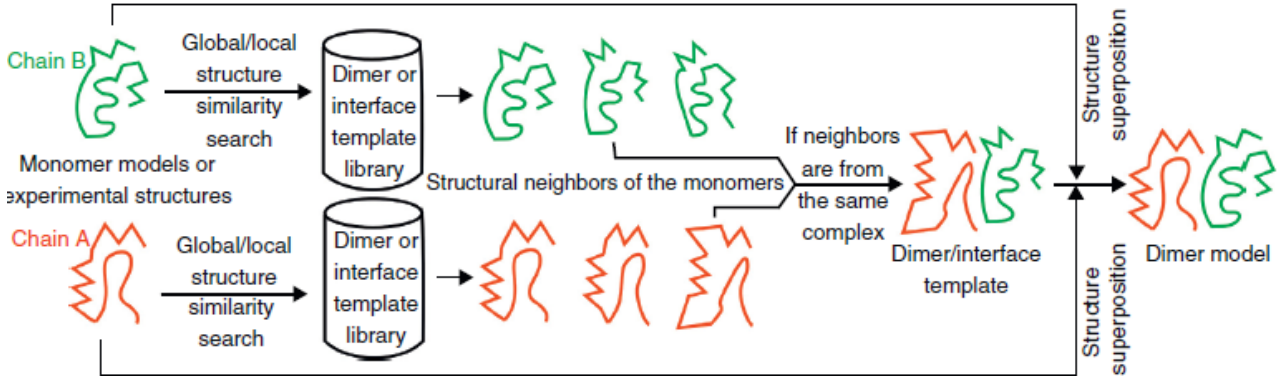
Dimeric threading



Monomer threading and oligomer mapping



Template-based docking



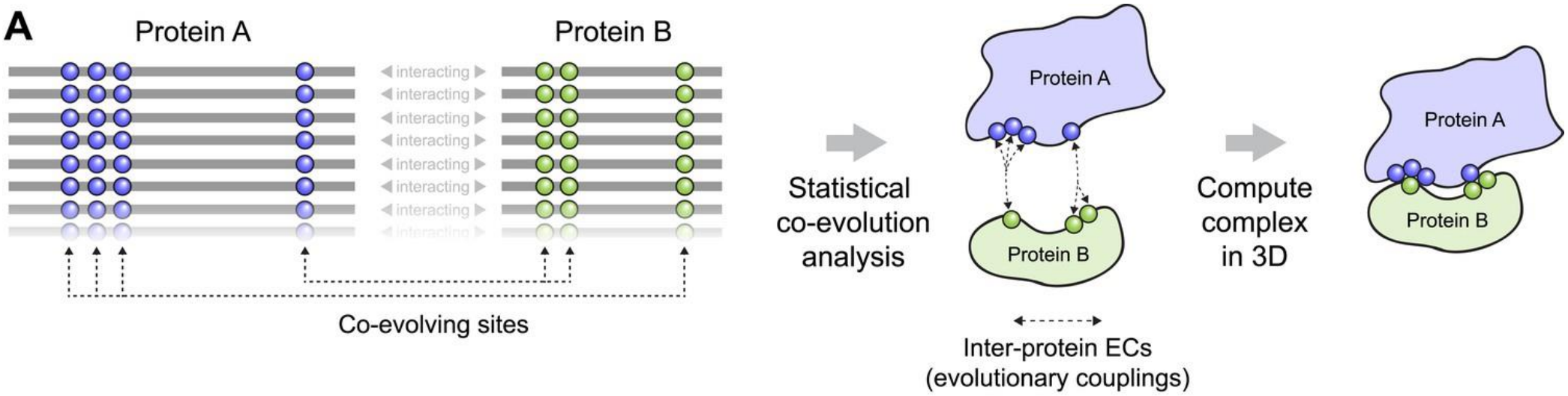
Data-driven Protein docking software

- Dimeric Threading
  Interactome3D (R), InterPreTS, Multiprospector, Coev2Net, Struct2Net, iWrap, HOMBACOP (R), HOMCOS(R), THSWP
- Dimeric Threading and Template Based
  ABCLM (M), KA
- Monomer threading and oligomer Mapping
  SPRING
- Dimeric Threading and Full complex simulation
  TACOS(R), M-TASSER(R)
- Template Based
  PrePPI, PRISM (M), SKV (M)

M: Complete models unrefined
R: Refined Structure

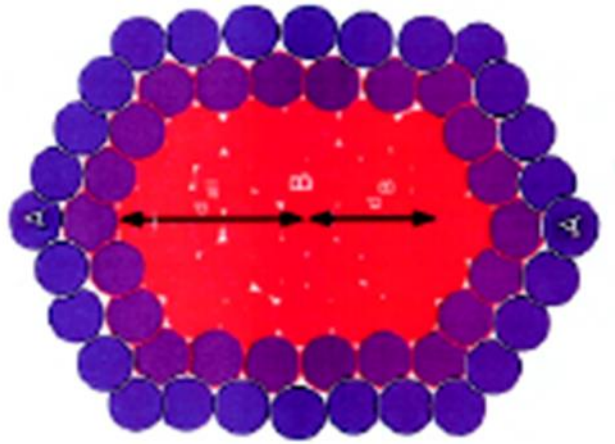
Co-evolution methods

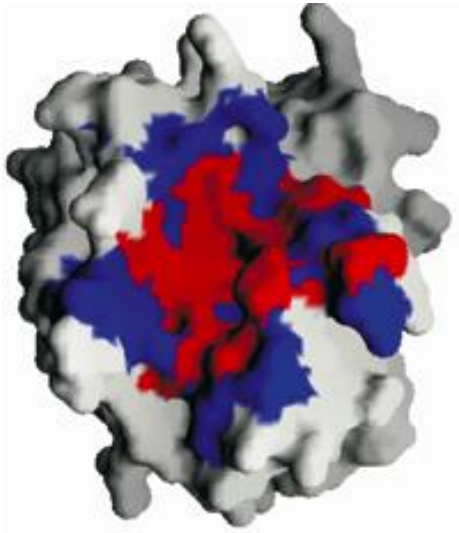
- Rely in multiple sequence alignments and correlated mutations to define correlated positions > 3D contacts
- Predicted contacts become restrains

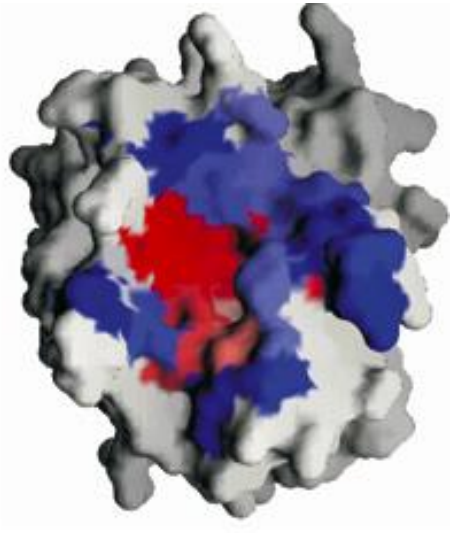


Interface/hot-spot prediction

- The hot-spot/O-ring model. Hot-spots make the real interactions, O-Ring (residues sourrounding the hot-spot) help to exclude solvent
  - Detecting solvent excluding (hydrophobic) regions help to detect interfaces
- Residue packing at the interface is similar to protein core.
- Residue conservation does not help (except for specific interactions)
  - Amino acid composition
  - Interface propensities (statistical models)
  - Machine learning approaches
- Definition of interface residues on training structures is key
  - Geometric (Distance based)
  - Optimal Docking Area (ODA), based in desolvation energy (ASA)





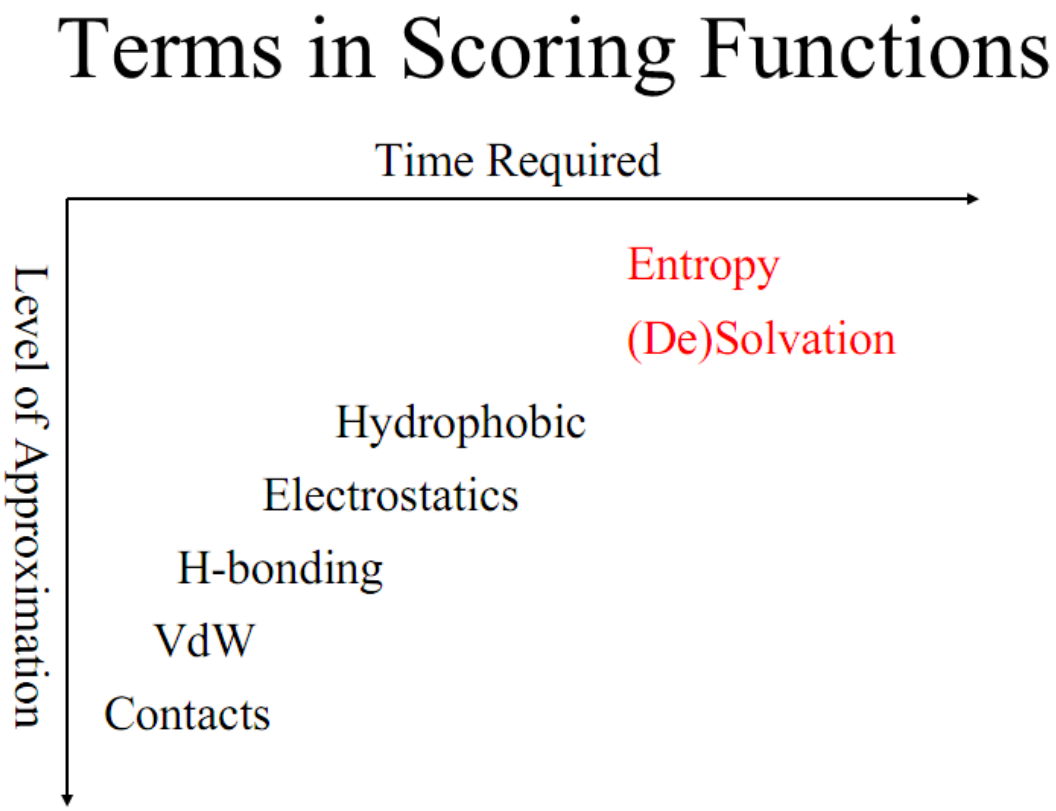


“Good docking” criteria

- Low Free Energy (!?)
- Low pseudo-energy (Scoring)
- Large Surface burial (~ 1,600 ± 400 $Å^2$)
- Low VdW overlaps
- No large cavities on the interface
- Good H-bonding (~1HB / 100 $Å^2$), Polar-polar contacts
- Good charge complementarity

Scoring functions

- Free energy (Forcefields)
- Solvation Score
  - Optimize hydrophobic solvation
- Statistical potentials
- Gometric scores
  - Buried surface
  - Surface shape complementarity
  - Volume of intersection
- Phylogenetic scores
  - Based on conservation
  - Correlated mutations - >contacts
- Re-scoring and consensous



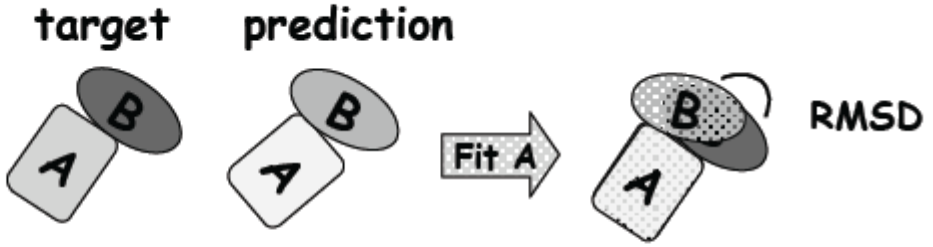
Assessment of docking performance

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

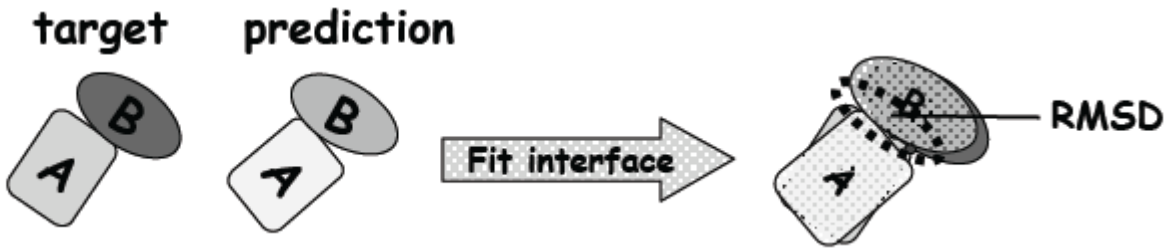
$$F_{nat} = \frac{\text{Correctly predicted contacts}}{\text{Total number of contacts in the target}}$$

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

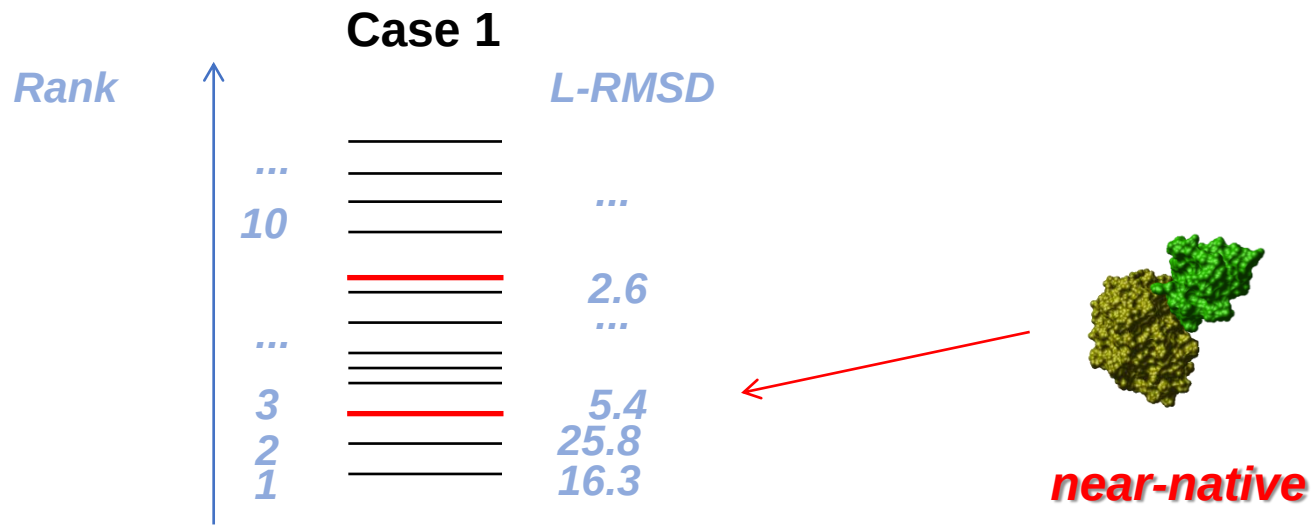
near-native: L-RMSD < 10Å



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

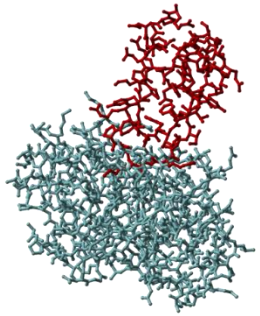


Assessment of docking performance



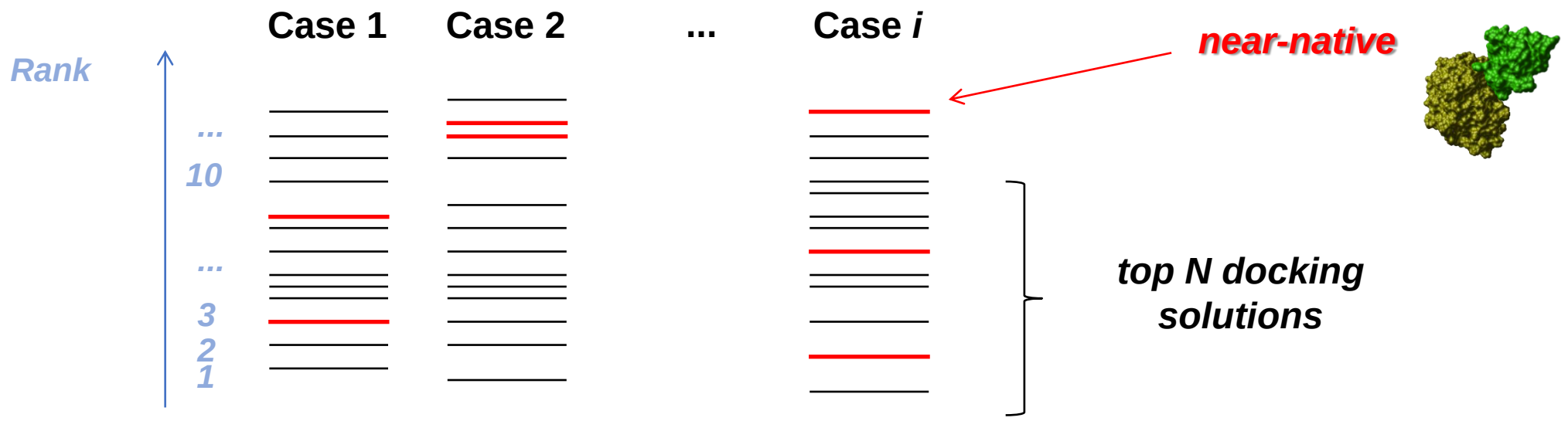
near-native

Best rank NN **3**

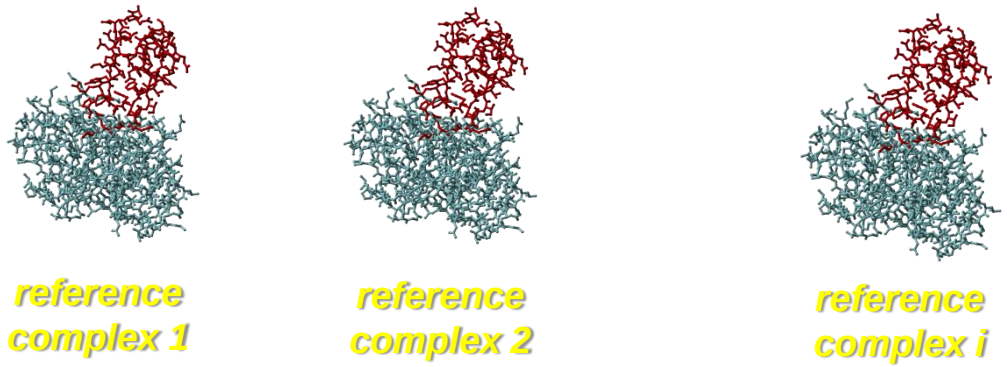


reference complex

Assessment of docking performance



Best rank NN **3** **11** **...** **2**



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

Protein-protein docking benchmark

- 176 protein-protein complexes
  - http://zlab.umassmed.edu/benchmark/
- Unique structural family combinations
- Diverse biological roles

	I-RMSD	fnat	fnon-nat	N
Rigid-body	0.9	79%	21%	121
Medium	1.76	63%	35%	30
Difficult	3.76	51%	51%	25

I-RMSD: RMSd bound-unbound. fnat, fnon-nat : Fraction of conserved contacts

CAPRI HOME

https://www.ebi.ac.uk/msd-srv/capri/

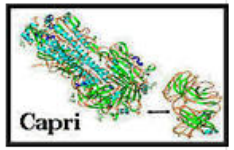
capri protein

Microsoft Office Home | Documentos de Google | Hojas de Cálculo de G... | EndNote | Aul@-ESCI: Entrar al si... | Calendario y horarios | Campus Virtual de la ... | MICINN - Sede Electró...

EMBL-EBI

Services | Research | Training | About us

CAPRI: Critical Assessment of PRediction of Interactions



- PDB idcodes for past targets
- Call For Targets
- Capri Rules 2007
- Original Capri Rules 2001
- Management
- Formats
- ROUND 44
- ROUND 43
- ROUND 42
- ROUND 41
- ROUND 40
- ROUND 39
- ROUND 38
- ROUND 37
- ROUND 36
- ROUND 35
- ROUND 34
- ROUND 33
- ROUND 32
- ROUND 31
- ROUND 30

Home > Databases > PDBe > Services > Capri-Home

CAPRI communitywide experiment on the comparative evaluation of protein-protein docking for structure prediction

Hosted by the Protein Data Bank in Europe (PDBe) Group

We are announcing the launch of CAPRI Round 44. For more details on registration please see New CAPRI ROUND 44

To download the target coordinates you must agree to the following conditions at the time of download.

Agreement (pdf)

To Register please email
Shoshana Wodak - shoshana.wodak AT gmail.com, Sameer Velankar - sameer AT ebi.ac.uk, Nurul Nadzirin - nurul AT ebi.ac.uk
Send
Team Leader Name:
e-Mail Address:
Staff Position At your Institute:
If a Team is submitting
For each member please send Name: & email:

Protein-protein interactions and other interactions between macromolecules are essential to all aspects of biology and medical sciences, and a number of methods have been developed to predict them. CAPRI is a community wide experiment designed to assess those that are based on structure. If we know the 3D structure of two components of a complex and build a model of their assembly, how reliable and accurate is that model likely to be?

CAPRI is a blind prediction experiment managed by capri management. Its targets are unpublished crystal or NMR structures of complexes, communicated on a confidential basis by their authors to the CAPRI management. Participant predictor groups are given the atomic coordinates of two proteins that make biologically relevant interactions. They model the target complex with the help of the coordinates and other publicly available data (sequence, mutations etc~), and submit sets of ten models for assessment on the CAPRI Web site. In addition, the predictors are invited to upload larger sets that are communicated to scorer groups who evaluate and rank them, and make a separate ten-model submission. After the prediction round is completed, the CAPRI assessors compare the submissions to the experimental structure, et evaluate the models on criteria that depend on the geometry and biological relevance of the predicted interactions.

Since CAPRI began in 2001, the experiment has had two to four prediction rounds each year, with one or a few targets per round. Each round is announced a week in advance, it lasts 3-6

Conclusions

- Protein docking works
  - (Much less efficient than ligand docking)
- Lots of methods exists, no clear winner
- Data-driven methods can generate better models if data is available
- Flexibility, conformational changes are the major problems
- Interface and interaction predictions (without docking) are possible and useful