Macromolecular complexes and interactions

Outline

- ☐ Macromolecular complexes
- ☐ Structure of complexes
- ☐ Prediction of 3D structures of complexes
- ☐ Analysis of macromolecular complexes

Biological relevance

- Types of biologically relevant complexes
    - Protein – small molecule ☑
    - Protein – protein
    - Protein – nucleic acids
    - Nucleic acids – small molecule ☒

} Macromolecular complexes

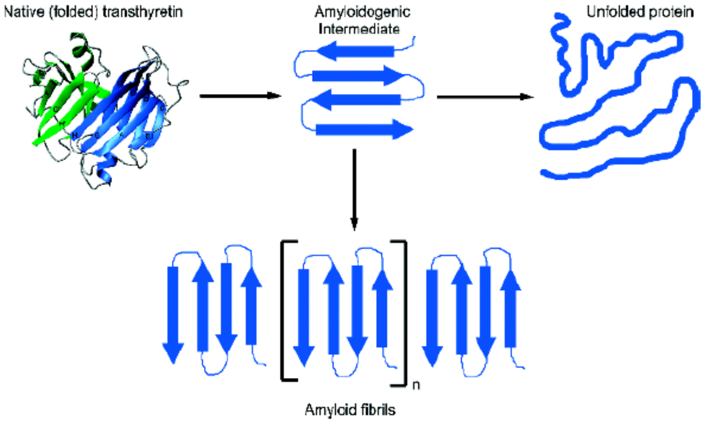
Biological relevance

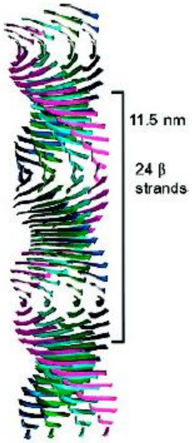
- Many proteins are formed by two or more polypeptide chains (protomers) interacting with each other
- Protein-protein and protein-nucleic acid interactions have central importance for virtually every process in a living cell (molecular recognition)
    - Regulation
    - Transport
    - Signal transduction
    - Genetic activity (transcription, translation, replication, repair, ...)
    - ...

Protein-protein complexes

- **Oligomerization**
  - Native interactions between proteins in native conditions

- **Aggregation**
  - Interactions between native proteins at extreme conditions
  - Interactions between misfolded/partially folded proteins → disease

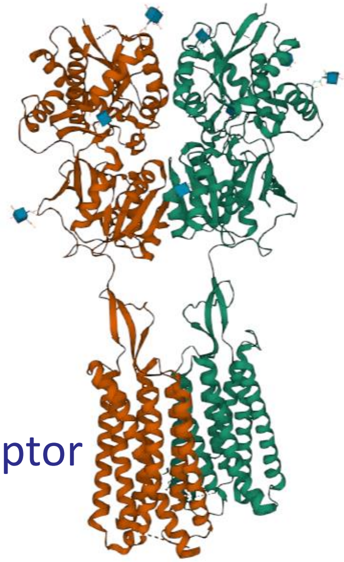




Protein-protein complexes

☐ Obligate complexes

- Protomers (individual polypeptides) do not function as independent structures, only when associated
- Examples: GABA receptors, ATP synthase, many ion channels, ribosome, etc.



GABA$_B$ receptor

☐ Non-obligate complexes

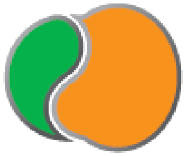
- Protomers can exist and be functional as independent structures
- Examples: hemoglobin, beta-2 adrenergic receptor, insulin receptor, etc.

Protein oligomerization

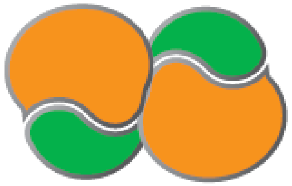
- ❑ Oligomerization is common
  - More than 35 % of proteins in a cell are oligomers
  - Tetramer is the average oligomeric state of proteins in *E. coli*
  - Homo-oligomers – the most common
  - Some proteins exists solely in the oligomeric state
- ❑ Oligomers are often symmetric
- ❑ Oligomerization interfaces are complementary
- ❑ Oligomerization is favored by evolution

homodimer: a2

heterodimer: ab



heterotetramer: a2b2



heteropentamer a2bcd

Advantages of oligomerization

- ❑ Why do proteins form oligomers?

Advantages of oligomerization

- **Morphological function**
  - More complex structures are often required for multiple functions
- **Cooperative function**
  - Allostery
  - Multivalent binding
- **Enhanced stability**
  - Smaller surface area
  - More interactions
- **… (ex. Translation error control)**

Oligomerization interface

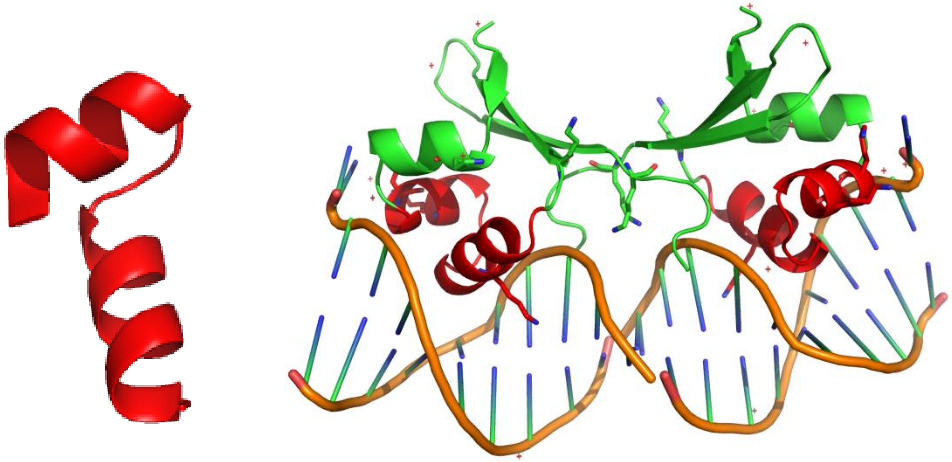
- **Characteristics of oligomeric interface**
    - Large surface area (> 1400 Å^2)
    - Tendency to circular and planar shape (not for obligates)
    - Some residues protrude from the surface
    - More non-polar residues (about 2/3) than in other parts of surface
    - More polar residues (about 1/5) than in protein cores
    - About 1 H-bond per 200 Å^2
- **Hot-spot residues**
    - Responsible for most of the oligomeric interactions
    - More evolutionary conserved than other surface residues
    - Frequently polar residues, located about the center of the interface

Protein-nucleic acids complexes

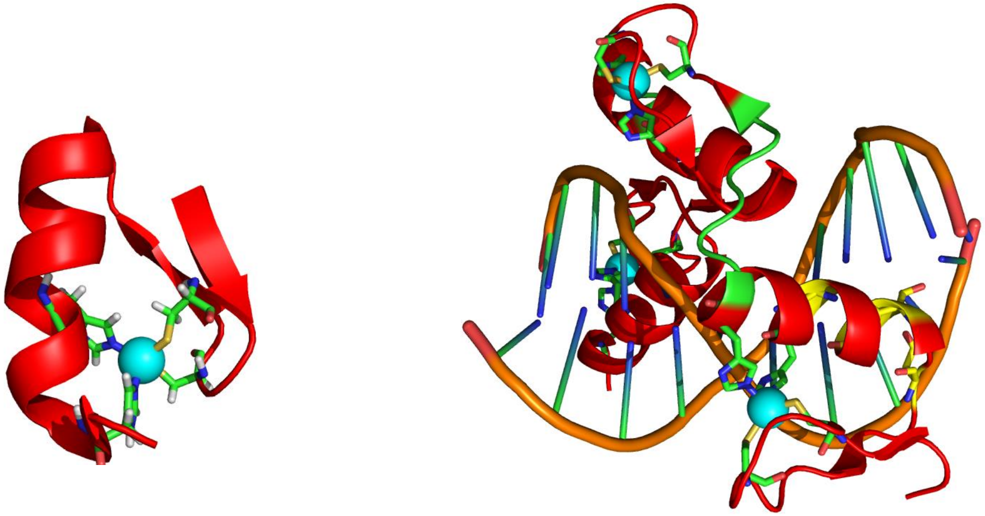
- **Protein-nucleic acid interactions**
  - Non-specific – electrostatic interactions with negative charge on the backbone of nucleic acid -> Lys and Arg residues
  - Specific – recognition of particular nucleotide sequences
    - Major groove – B-DNA
    - Minor groove – A-DNA or A-RNA
    - Single strand RNA
- **Typical interfaces/motifs**
  - DNA binding proteins
  - RNA binding proteins

Protein-nucleic acids complexes

- **DNA binding proteins**
  - Helix-turn-helix

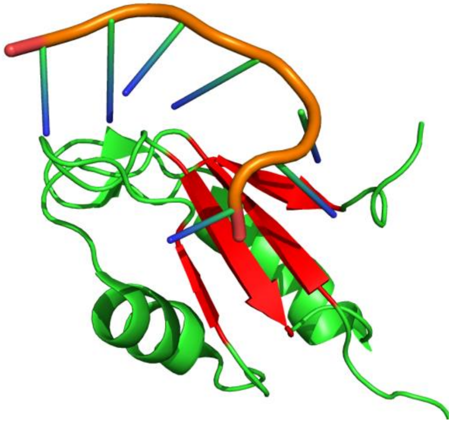


  - Zinc finger

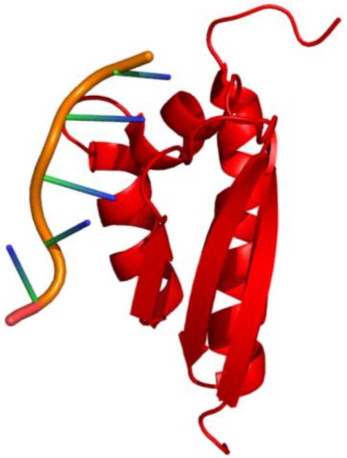


Protein-nucleic acids complexes

- **RNA binding proteins**
  - Recognition is often also governed by particular structures of RNA
  - Many motifs employed



RNA recognition motif



K-homology domain

Pumilio repeat domain

Structure of complexes

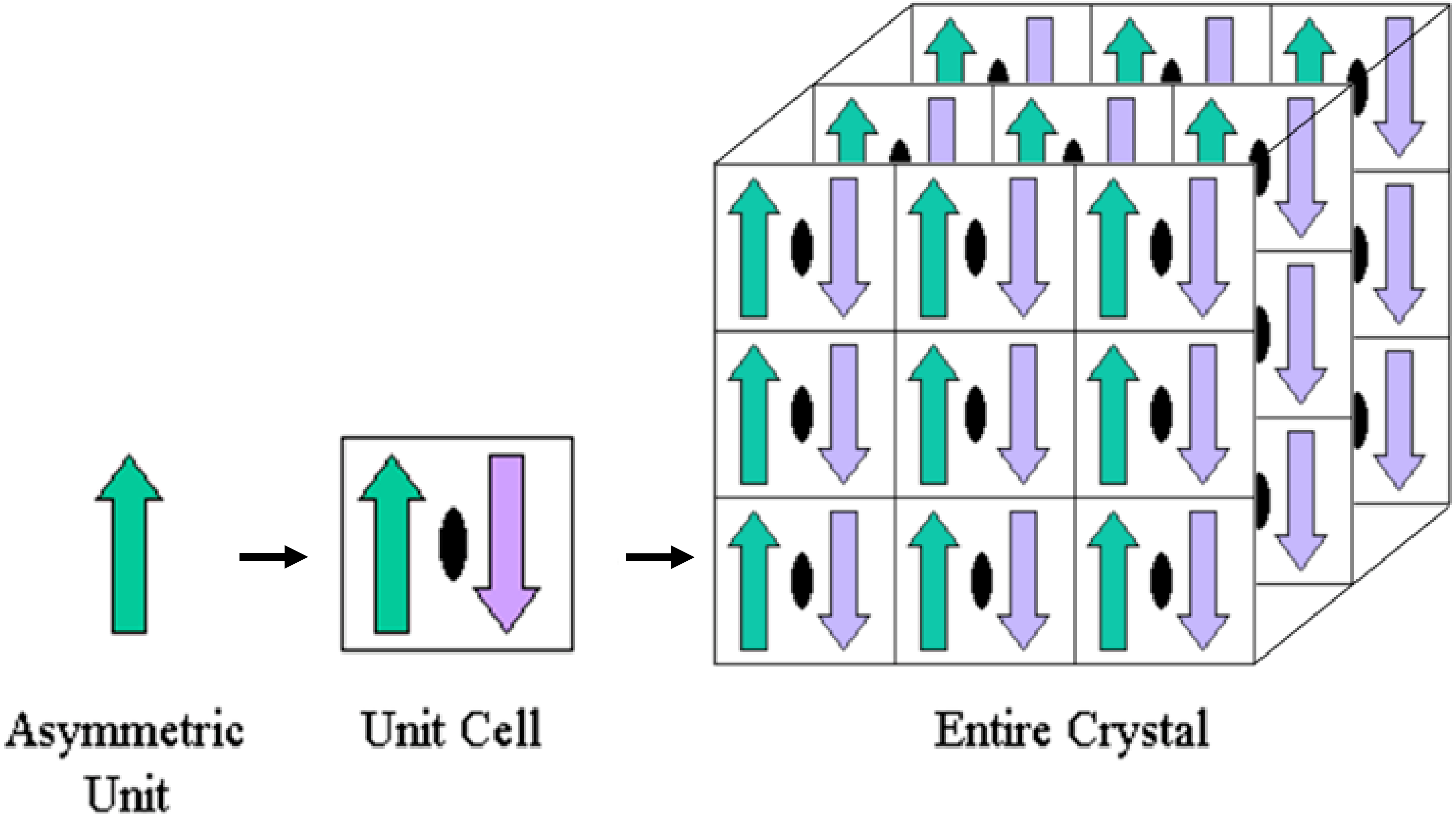
- Quaternary structure in PDB database
- Complex or crystallization artifact?

Quaternary structure in PDB database

- **Asymmetric unit (ASU)**
  - Macromolecular structures from X-ray crystallography deposited to PDB as a single asymmetric unit
  - The smallest portion of a crystal structure to which symmetry operations can be applied in order to generate the unit cell

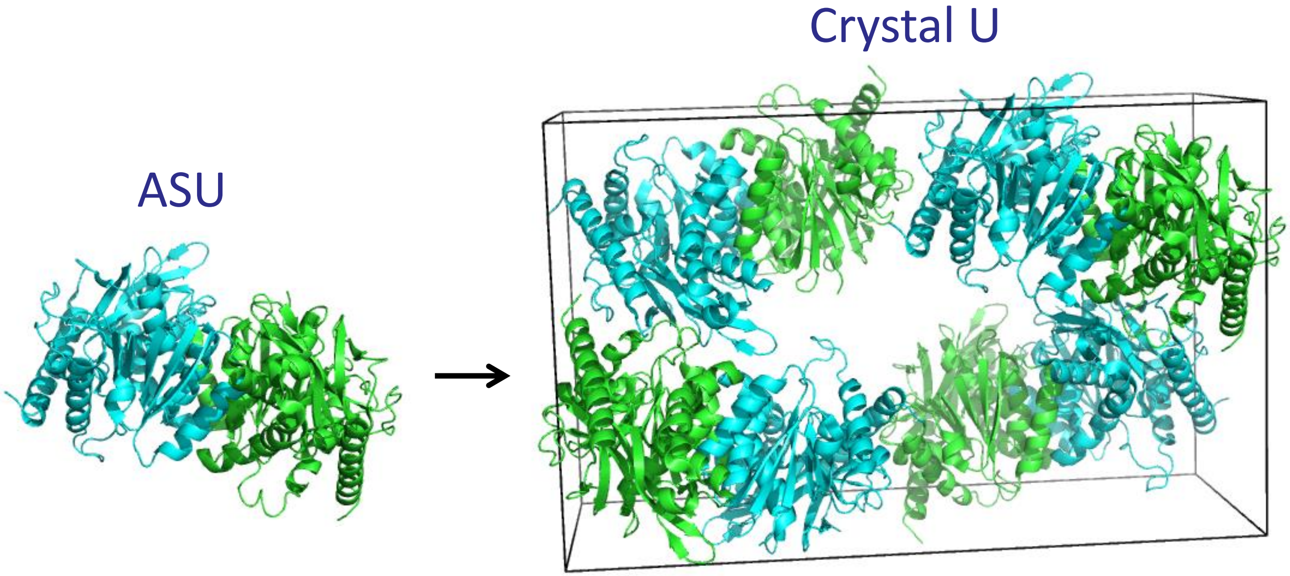
- **Unit cell (crystal unit)**
  - The basic unit of a crystal that, when repeated in three dimensions, can generate the entire crystal

Quaternary structure in PDB database



Crystalline environment

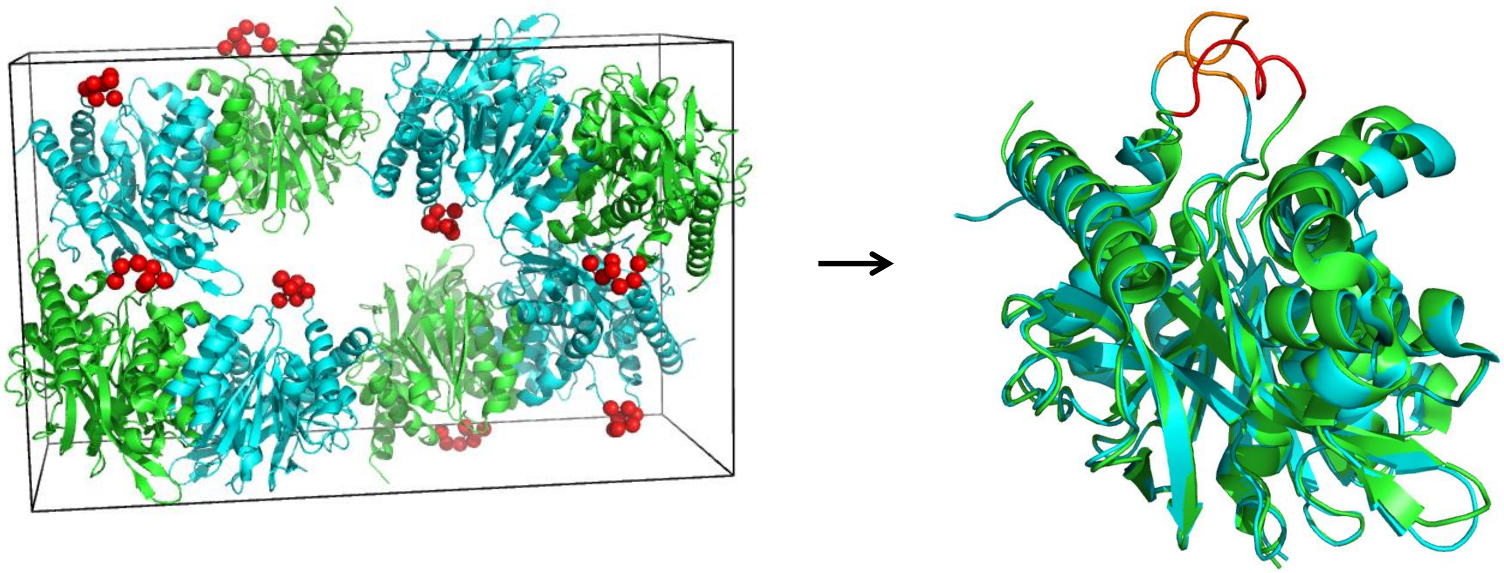
- **Crystal contacts**
  - Intermolecular contacts solely due to protein crystallization
    - Causes artifacts of crystallization
    - Crystal packing - complicates identification of native quaternary structure



Crystalline environment

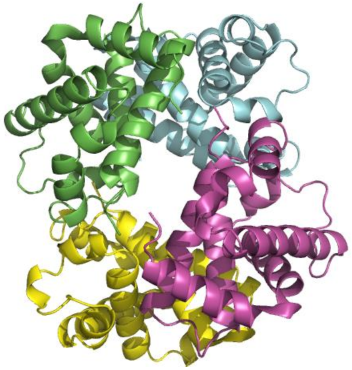
Artifacts of crystallization

- Concerns about conformation of some surface regions
- Often loops or side chains are affected
- Can complicate the evaluation of the effects of mutations



Quaternary structure in PDB database

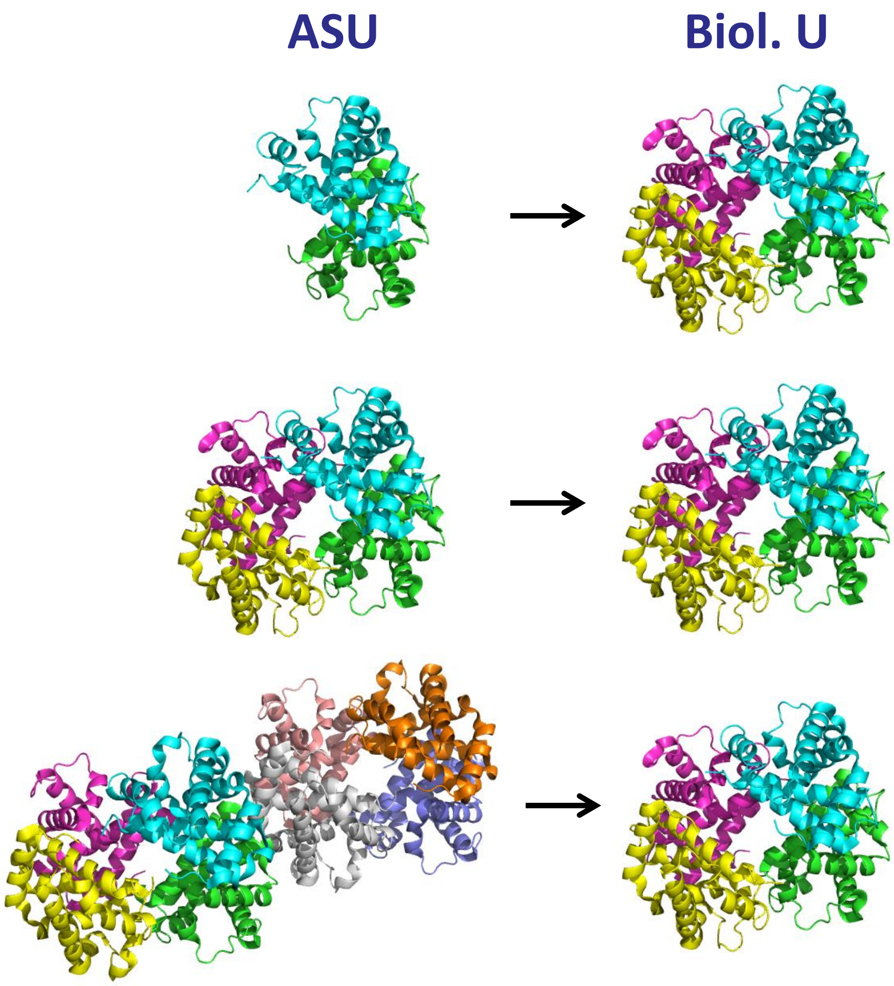
- **Biological unit**
  - The functional form of a protein in nature
  - Also called: functional unit, biological assembly, quaternary structure
  - Can depend on the environment, post-translational modifications of proteins and their mutations



Homotetramer hemoglobin

Biological *versus* asymmetric unit

- **Biological unit can consist of:**
  - Multiple copies of the ASU
  - One copy of the ASU
  - A portion of the ASU

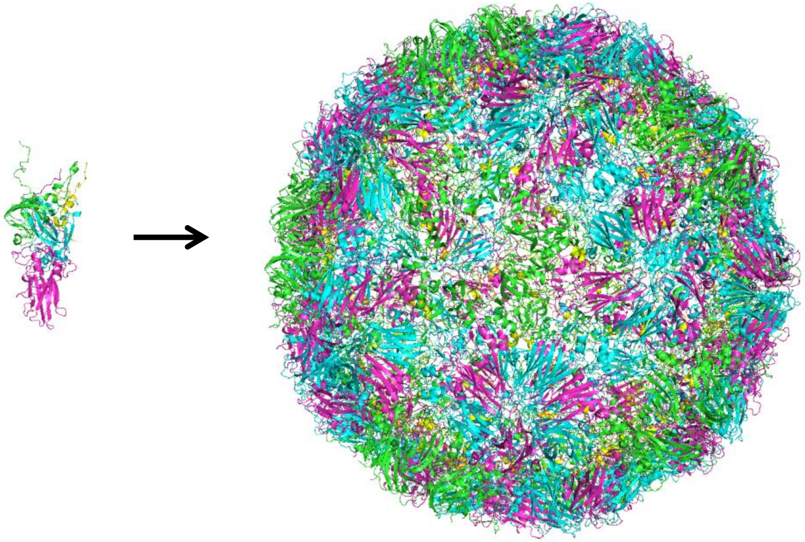


Biological *versus* asymmetric unit

- **Large assemblies**
  - Viral capsid

ASU

Biol. U



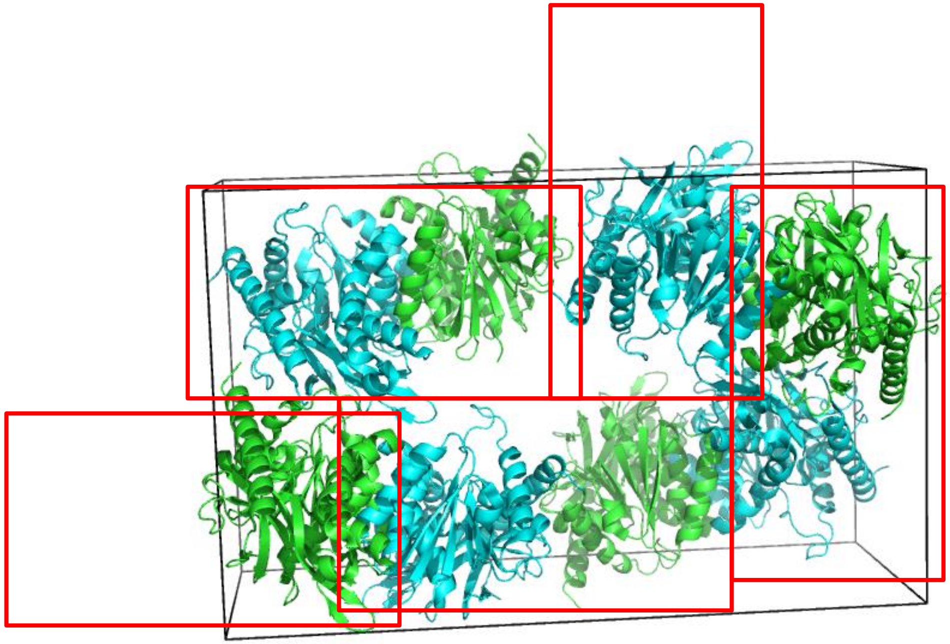
  - Filamentous bacteriophage PF1

ASU

Biol. U

Complex or artifact?

- **Problem**
  - Most proteins in the PDB have three or more crystal contacts that sum up to 30% of the protein solvent accessible surface area
  - How to recognize biologically relevant contacts from crystal one?



Complex or artifact?

- **Experimental knowledge of oligomeric state helps with identifying of the structure of native complex**
  - Search literature
  - Experimental methods
    - Gel filtration, static or dynamic light scattering, analytical ultracentrifugation, native electrophoresis, …
- **How to get the structure of a biological unit?**
  - Author-specified assembly
  - Databases
  - Predictive tools

Author-specified assembly

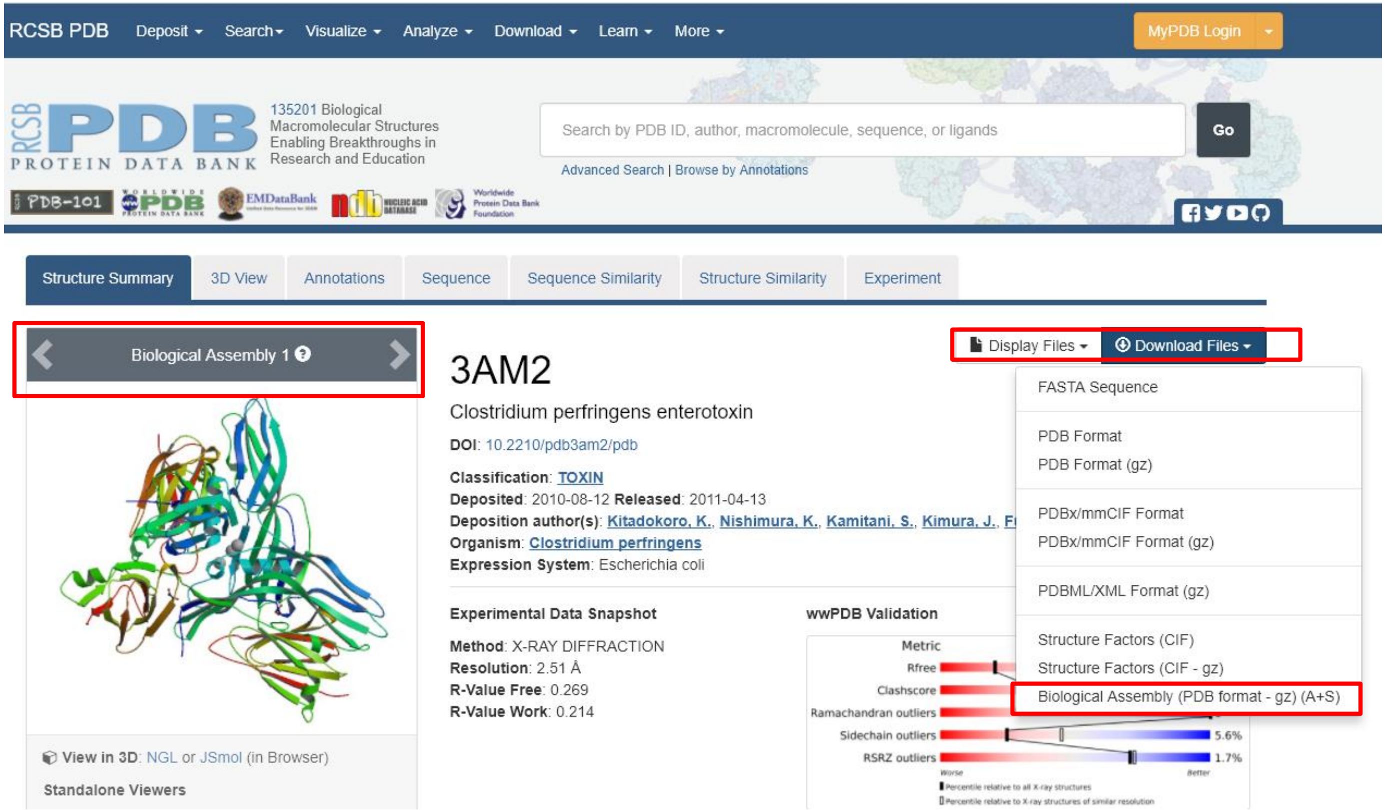
- **REMARK 350 in headers of PDB file**
  - Contains symmetry operations to reconstruct biological unit, but…

  → Verify author-proposed biological unit by other means
    - Sometimes the specific oligomers were not known at the time the ASU was published
    - Some authors may have failed to specify the biological unit even when it was known
    - Rarely, the specified biological unit might be incorrect
- **Employed by**
  - RCSB PDB and other tools

Author-specified assembly

RCSB PDB

- Generates a PDB file in which all protein chains are as separate models → complicates visualization and analysis



3AM2

Clostridium perfringens enterotoxin

DOI: 10.2210/pdb3am2/pdb

Classification: TOXIN
Deposited: 2010-08-12 **Released**: 2011-04-13
Deposition author(s): Kitadokoro, K., Nishimura, K., Kamitani, S., Kimura, J.,
Organism: Clostridium perfringens
Expression System: Escherichia coli

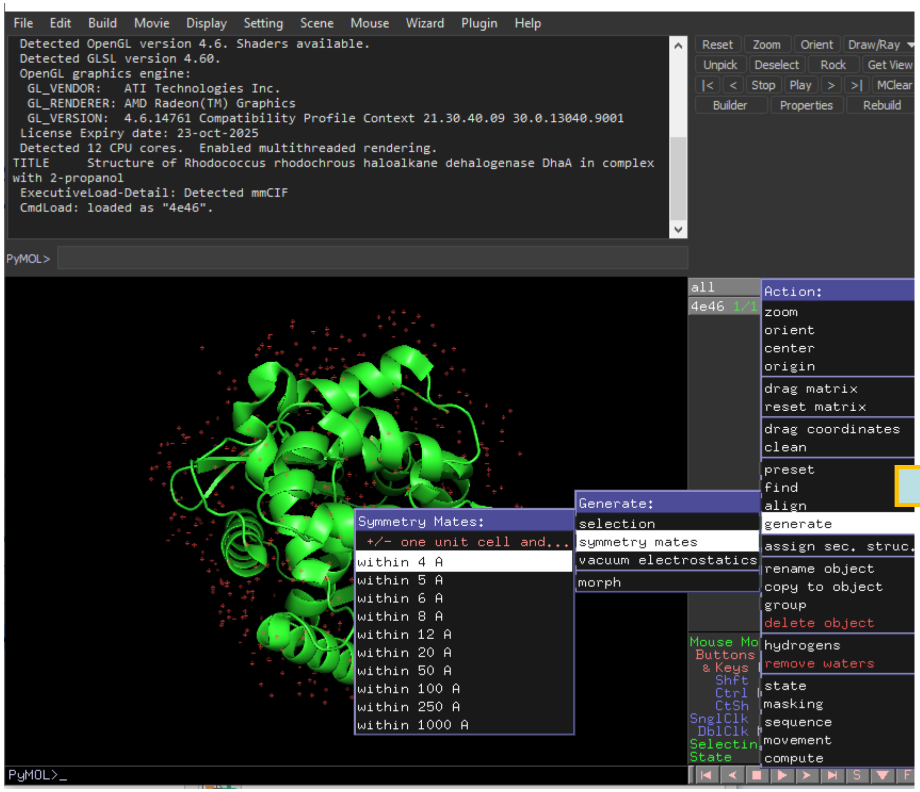
Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 2.51 Å
R-Value Free: 0.269
R-Value Work: 0.214

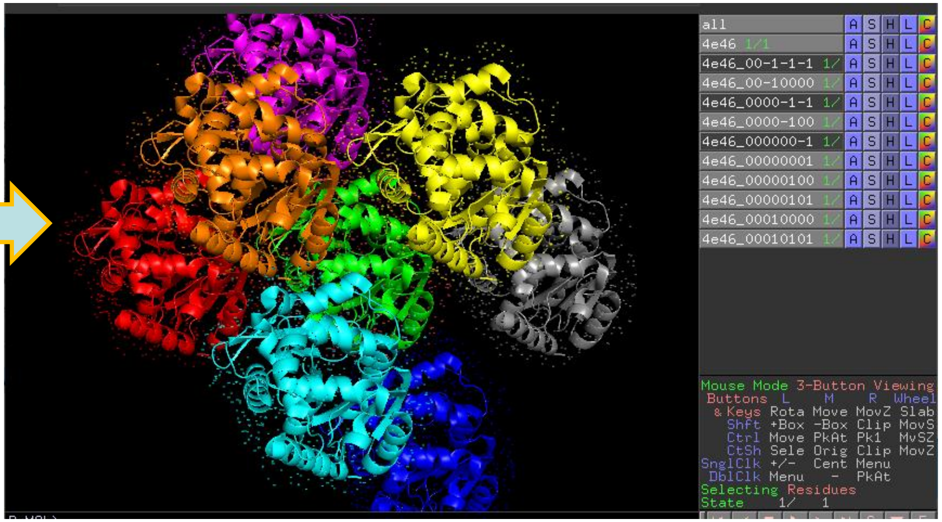
Crystal lattice

❑ PyMOL

- Generate > Symmetry mates → to visualize nearest partners
- You can select some and combine them in a PDB file



```
 Detected OpenGL version 4.6. Shaders available.
 Detected GLSL version 4.60.
 OpenGL graphics engine:
  GL_VENDOR:   ATI Technologies Inc.
  GL_RENDERER: AMD Radeon(TM) Graphics
  GL_VERSION:  4.6.14761 Compatibility Profile Context 21.30.40.09 30.0.13040.9001
 License Expiry date: 23-oct-2025
 Detected 12 CPU cores.  Enabled multithreaded rendering.
TITLE     Structure of Rhodococcus rhodochrous haloalkane dehalogenase DhaA in complex
with 2-propanol
 ExecutiveLoad-Detail: Detected mmCIF
 CmdLoad: loaded as "4e46".
```



Prediction of 3D structure of complexes

- How can we predict macromolecular complexes?

Prediction of 3D structure of complexes

- Homology-based methods
- Machine learning-based threading
- Macromolecular docking

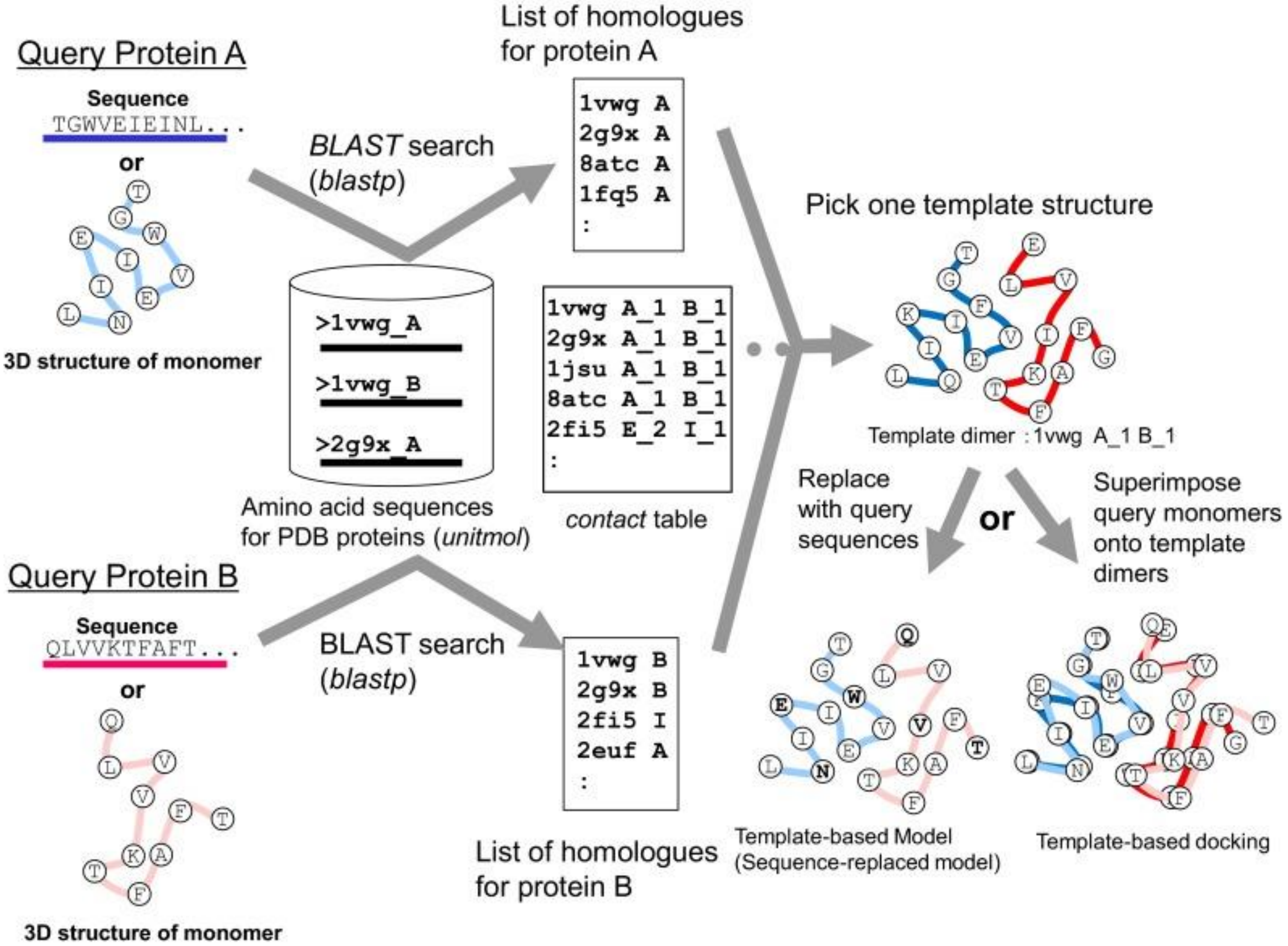
Homology based methods

- The model of a protein complex is built based on a similar protein complex with a known 3D structure
- Assumes that the interaction information can be extrapolated from one complex structure to close homologs of interacting proteins
  - Close homologs (≥ 40% sequence identity) almost always interact in the same way (if they interact with the same partner)
  - Sequence similarity is only rarely associated with a similarity in interactions
- Limited applicability (low number of templates)

Homology based methods

- HOMCOS (**Ho**mology **M**odeling of **Co**mplex **S**tructure)
  - https://homcos.pdbj.org/
  - Predicts 3D structure of homodimers and heterodimers by homology modeling
  - Optionally, identifies potentially interacting proteins
  - Steps:
    1. BLAST search to identify homologous templates in the latest representative dataset of heterodimer (homodimer) structures
    2. Evaluation of the model validity by the combination of sequence similarity and knowledge-based contact potential energy
    3. Generation of a script for building full atomic model by MODELLER

Homology based methods



Query Protein A

Sequence

TGWVEIEINL...

or

3D structure of monomer

BLAST search (*blastp*)

List of homologues for protein A

```
1vwg A
2g9x A
8atc A
1fq5 A
:
```

```
>1vwg_A
>1vwg_B
>2g9x_A
```

Amino acid sequences for PDB proteins (*unitmol*)

```
1vwg A_1 B_1
2g9x A_1 B_1
1jsu A_1 B_1
8atc A_1 B_1
2fi5 E_2 I_1
:
```

contact table

Pick one template structure

Template dimer : 1vwg A_1 B_1

Replace with query sequences

or

Superimpose query monomers onto template dimers

Template-based Model (Sequence-replaced model)

Template-based docking

Query Protein B

Sequence

QLVVKTFAFT...

or

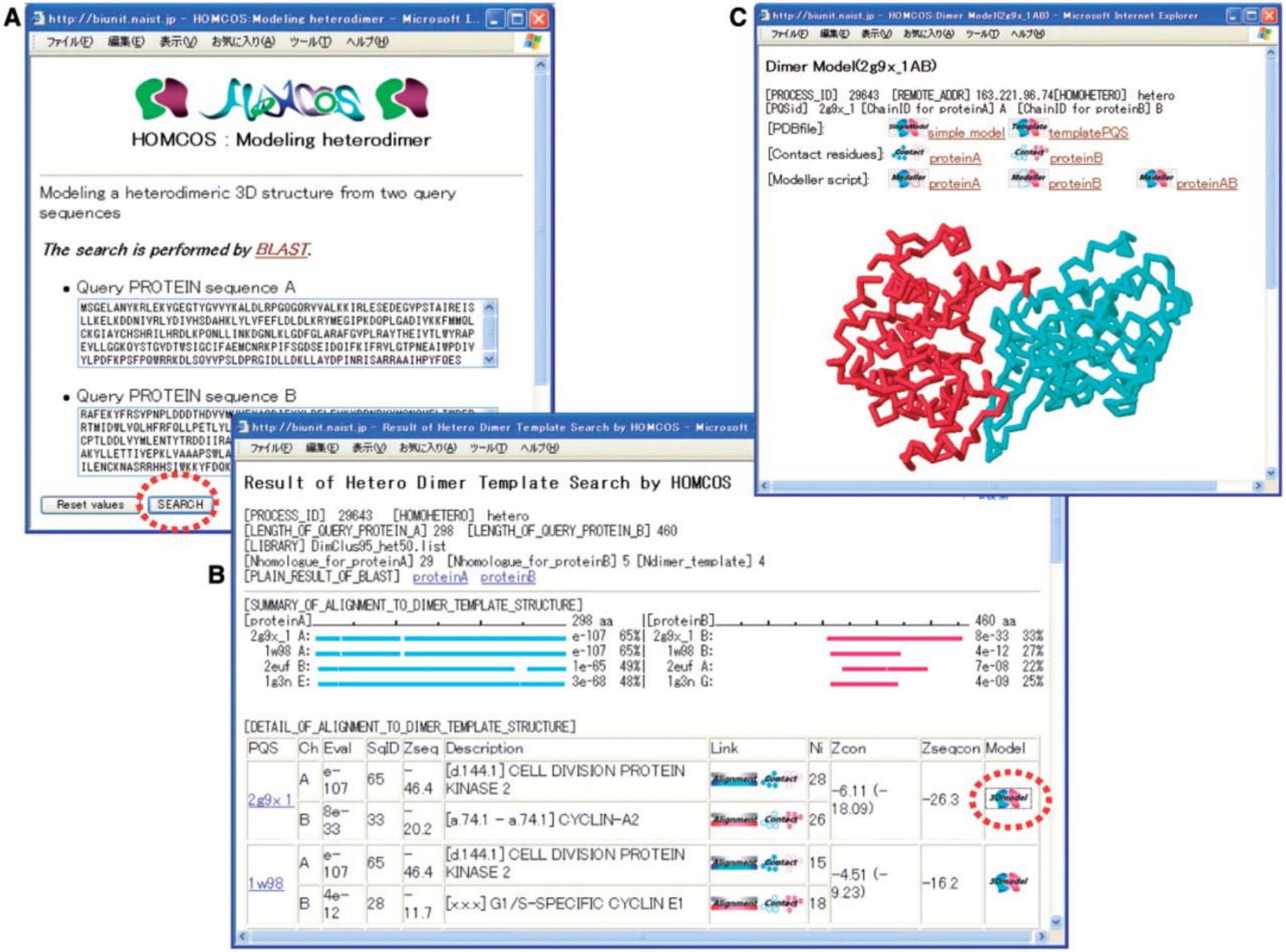
3D structure of monomer

BLAST search (*blastp*)

```
1vwg B
2g9x B
2fi5 I
2euf A
:
```

List of homologues for protein B

Homology based methods



A http://biunit.naist.jp - HOMCOS:Modeling heterodimer - Microsoft I...

HOMCOS : Modeling heterodimer

Modeling a heterodimeric 3D structure from two query sequences

The search is performed by BLAST.

- Query PROTEIN sequence A

```
MSGELANYKRLEKVGEGTYGVVYKALDLRPGOGORVVALKKIRLESEDEGVPSTAIREIS
LLKELKDDNIVRLYDIVHSDAHKLYLVFEFLDLDLKRYMEGIPKDOPLGADIVKKFMMOL
CKGIAYCHSHRILHRDLKPONLLINKDGNLKLGDFGLARAFGVPLRAYTHEIVTLWYRAP
EYLLGGKOYSTGVDTWSIGCIFAEMCNRKPIFSGDSEIDOIFKIFRVLGTPNEAIWPDIV
YLPDFKPSFPOWRRKDLSOVVPSLDPRGIDLLDKLLAYDPINRISARRAAIHPYFOES
```

- Query PROTEIN sequence B

```
RAFEKYFRSVPNPLDDDTHDVVM
RTMIDWLVOLHFRFOLLPETLYL
CPTLDDLVYMLENTYTRDDIIRA
AKYLLETTIVEPKLVAAAPSWLA
ILENCKNASRRHHSIWKKYFDOK
```

B http://biunit.naist.jp - Result of Hetero Dimer Template Search by HOMCOS - Microsoft

Result of Hetero Dimer Template Search by HOMCOS

```
[PROCESS_ID]  29643   [HOMOHETERO]  hetero
[LENGTH_OF_QUERY_PROTEIN_A] 298  [LENGTH_OF_QUERY_PROTEIN_B] 460
[LIBRARY] DimClus95_het50.list
[Nhomologue_for_proteinA] 29  [Nhomologue_for_proteinB] 5 [Ndimer_template] 4
[PLAIN_RESULT_OF_BLAST]  proteinA  proteinB
```

```
[SUMMARY_OF_ALIGNMENT_TO_DIMER_TEMPLATE_STRUCTURE]
[proteinA]                298 aa   |[proteinB]              460 aa
 2g9x_1 A:               e-107  65%| 2g9x_1 B:              8e-33  33%
   1w98 A:               e-107  65%|   1w98 B:              4e-12  27%
   2euf B:               1e-65  49%|   2euf A:              7e-08  22%
   1g3n E:               3e-68  48%|   1g3n G:              4e-09  25%
```

[DETAIL_OF_ALIGNMENT_TO_DIMER_TEMPLATE_STRUCTURE]

PQS	Ch	Eval	SqID	Zseq	Description	Link	Ni	Zcon	Zseqcon	Model
2g9x_1	A	e-107	65	-46.4	[d.144.1] CELL DIVISION PROTEIN KINASE 2	Alignment Contact	28	-6.11 (-18.09)	-26.3	3Dmodel
	B	8e-33	33	-20.2	[a.74.1 - a.74.1] CYCLIN-A2	Alignment Contact	26			
1w98	A	e-107	65	-46.4	[d.144.1] CELL DIVISION PROTEIN KINASE 2	Alignment Contact	15	-4.51 (-9.23)	-16.2	3Dmodel
	B	4e-12	28	-11.7	[x.x.x] G1/S-SPECIFIC CYCLIN E1	Alignment Contact	18			

C http://biunit.naist.jp - HOMCOS:Dimer Model(2g9x_1AB) - Microsoft Internet Explorer

Dimer Model(2g9x_1AB)

```
[PROCESS_ID]  29643  [REMOTE_ADDR] 163.221.96.74[HOMOHETERO]  hetero
[PQSid]  2g9x_1 [ChainID for proteinA] A  [ChainID for proteinB] B
```

[PDBfile]: simple model templatePQS

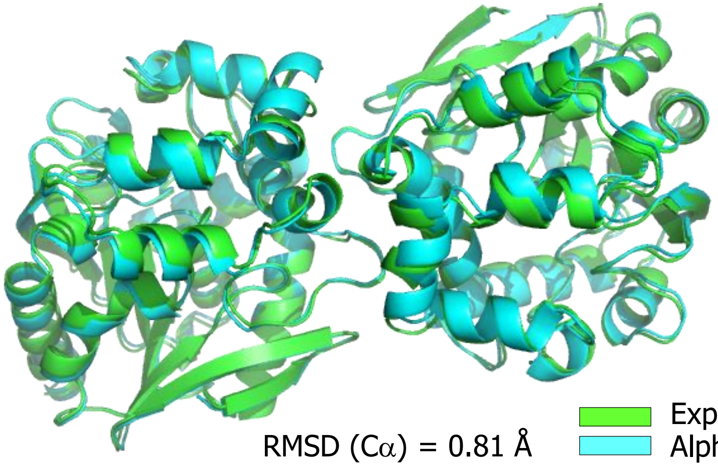
[Contact residues]: proteinA proteinB

[Modeller script]: proteinA proteinB proteinAB

Machine learning-based

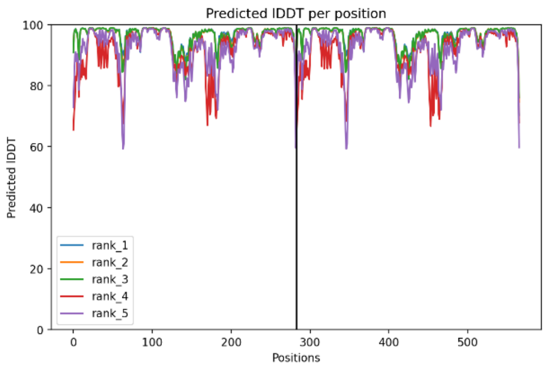
- AlphaFold-Multimer
  - Predicts 3D structure of multimers; similar to AlphaFold

```
>ChainA
MKRVDVLDSAMSYIDVGQGDPIVFLHGNPTSSYLWRNVI
PHLSDVGRCLAPDLIGMGASGTSPTFSYRFADHVRYLDA
WFEAVGITENVVLVVHDWGSALGFYRALRYPEQIAGIAY
MDALVQPRTWAGFTDYEPLMRALRTEQGERMALAENVFV
EKVVPGGVQRQLTEEEMAVYRTPYPTPQSRIPTLLWARE
IPVEGEPADVQAMVQEYADFLSRSDIPKLLIVAEPGAIL
HEGGSELDFARSWPNQREVKVAGRHFLQEDSPDAIGAAV
RAFVLDVRER
>ChainB
MKRVDVLDSAMSYIDVGQGDPIVFLHGNPTSSYLWRNVI
PHLSDVGRCLAPDLIGMGASGTSPTFSYRFADHVRYLDA
WFEAVGITENVVLVVHDWGSALGFYRALRYPEQIAGIAY
MDALVQPRTWAGFTDYEPLMRALRTEQGERMALAENVFV
EKVVPGGVQRQLTEEEMAVYRTPYPTPQSRIPTLLWARE
IPVEGEPADVQAMVQEYADFLSRSDIPKLLIVAEPGAIL
HEGGSELDFARSWPNQREVKVAGRHFLQEDSPDAIGAAV
RAFVLDVRER
```



RMSD (Cα) = 0.81 Å

🟩 Experimental
🟦 AlphaFold-multimer



Macromolecular docking

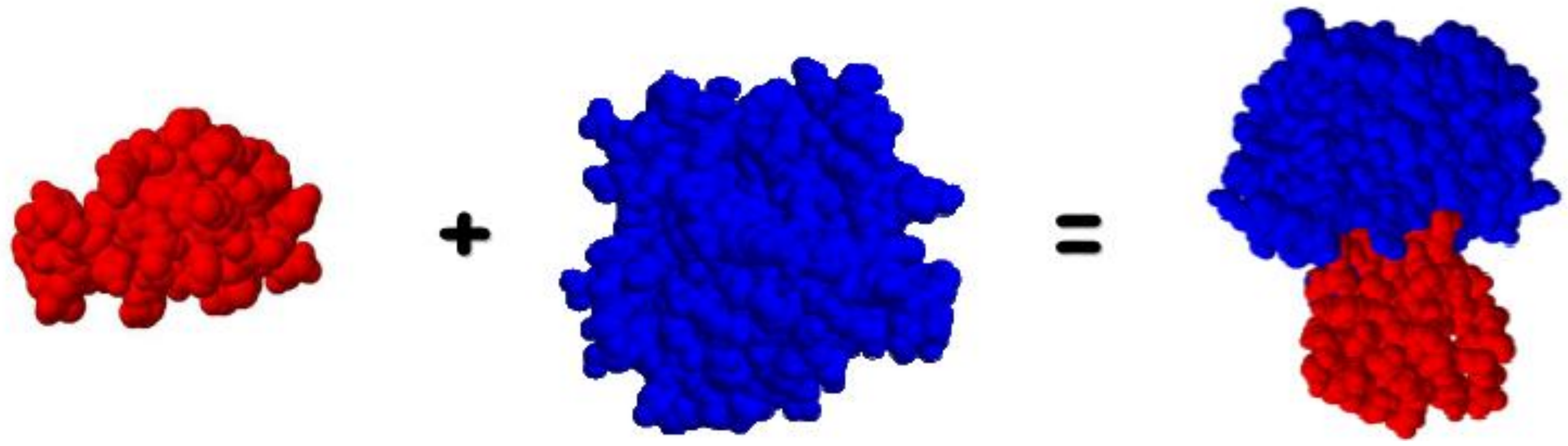
- Prediction of the best bound state for given 3D structures of two or more macromolecules
- Difficult task
  - Large search space - many potential ways in which macromolecules can interact
  - Flexibility of the macromolecular surface and conformational changes upon binding
- Can be facilitated by prior knowledge
  - Ex: known binding site → significant restriction of the search space
  - Distance constraints on some residues

Macromolecular docking

☐ Macromolecule representation

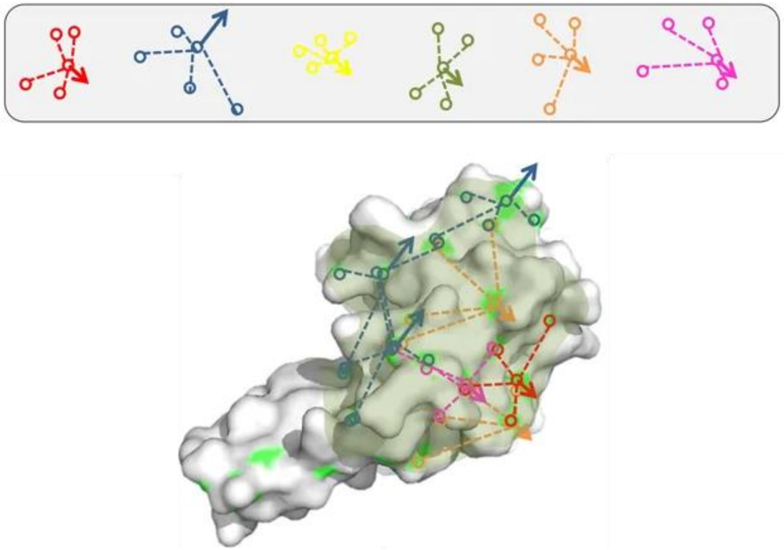
☐ Search algorithm

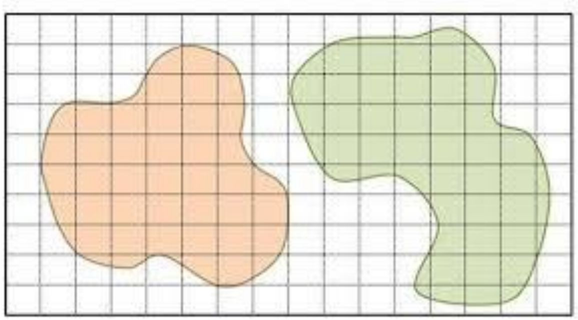
☐ Scoring function



Macromolecule representation

- Representation of the macromolecular surface (applicable to both receptor and ligand)
  - Geometrical descriptors of shape (set of spheres, surface normals, vectors radiating from the center of the molecule,...)
  - Discretization of space: grid representation





Macromolecule representation

- Macromolecule flexibility
  - Fully rigid approximation
  - Soft docking – employs tolerant “soft” potential scoring functions to simulate plasticity of otherwise rigid molecule
  - Explicit side-chain flexibility – optimization of residues by rotating part of their structure or rotation of whole side-chains using predefined rotamer libraries
  - Docking to molecular ensemble of protein structure – composed from multiple crystal structures, from NMR structure determination or from trajectory produced by MD simulation

Macromolecule representation

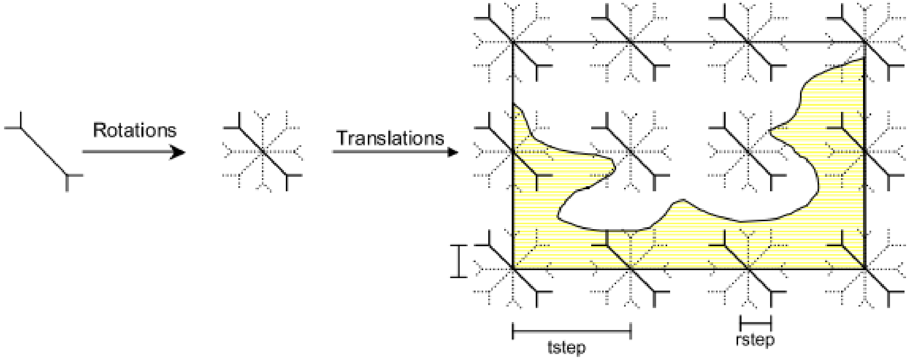
- Macromolecule flexibility
  - Rigid body docking – basic model that considers the two macromolecules as two rigid solid bodies
  - Semiflexible docking – one of the molecules is rigid, and one is flexible (typically the smaller one)
  - Flexible docking – both molecules are considered flexible

Macromolecular docking - search

- Generally based on the idea of complementarity between the interacting molecules (geometric, electrostatic or hydrophobic contacts)
- The main problem is the dimension of the conformational space to be explored:
  - Rigid docking: 6D (hard)
  - Flexible docking: 6D + N_{fb} (impossible!)
- Information on the rough location of the binding surface (experimental or predicted) → reduction of the search space

Macromolecular docking - search

- ❑ Exhaustive search
  - Full search of the conformational space: try every possible relative orientation of the two molecules
  - Computationally very expensive – 6 degrees of freedom for rigid molecules (translations + rotations)
  - Grid approaches



Macromolecular docking - search

- Stochastic methods
  - Monte Carlo
  - Genetic algorithms
  - Brownian dynamics
  - ...

Macromolecular docking - scoring

- Scoring functions
  - Evaluation of a large number of putative solutions generated by the search algorithms

- Methods often use a two-stage ranking
  1. Approximate and fast-to-compute function – used to eliminate very unlikely solutions
  2. More accurate function – used to select the best among the remaining solutions

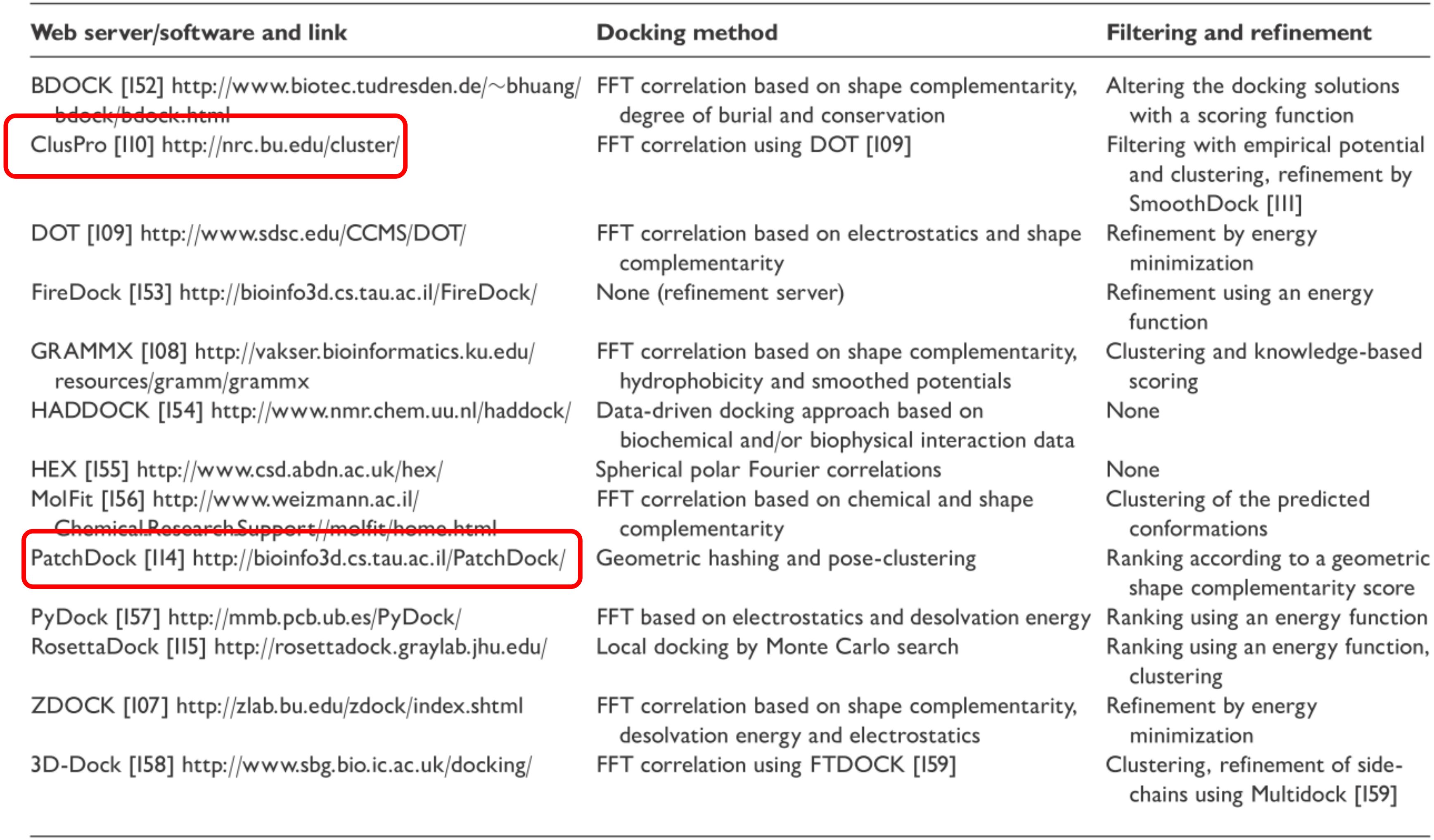
Macromolecular docking - scoring

- Scoring functions
  - Empirical
  - Knowledge-based
  - Force field-based
  - Clustering-based – the presence of many similar solutions is taken as an indication of correctness (all solutions are clustered, and the size of each cluster is used as a scoring parameter)

Macromolecular docking - scoring

- Good scores – a combination of several parameters:
  - Low free energy or pseudo-energy based on force field functions
  - Large buried surface area
  - Good geometric complementarity
  - Many H-bonds
  - Good charge complementarity
  - Polar/polar contacts favored
  - Polar/non-polar contacts are disfavored
  - Many similar solutions (large clusters)
  - …

Macromolecular docking - programs



Web server/software and link	Docking method	Filtering and refinement
BDOCK [152] http://www.biotec.tudresden.de/~bhuang/bdock/bdock.html	FFT correlation based on shape complementarity, degree of burial and conservation	Altering the docking solutions with a scoring function
ClusPro [110] http://nrc.bu.edu/cluster/	FFT correlation using DOT [109]	Filtering with empirical potential and clustering, refinement by SmoothDock [111]
DOT [109] http://www.sdsc.edu/CCMS/DOT/	FFT correlation based on electrostatics and shape complementarity	Refinement by energy minimization
FireDock [153] http://bioinfo3d.cs.tau.ac.il/FireDock/	None (refinement server)	Refinement using an energy function
GRAMMX [108] http://vakser.bioinformatics.ku.edu/resources/gramm/grammx	FFT correlation based on shape complementarity, hydrophobicity and smoothed potentials	Clustering and knowledge-based scoring
HADDOCK [154] http://www.nmr.chem.uu.nl/haddock/	Data-driven docking approach based on biochemical and/or biophysical interaction data	None
HEX [155] http://www.csd.abdn.ac.uk/hex/	Spherical polar Fourier correlations	None
MolFit [156] http://www.weizmann.ac.il/ChemicalResearchSupport//molfit/home.html	FFT correlation based on chemical and shape complementarity	Clustering of the predicted conformations
PatchDock [114] http://bioinfo3d.cs.tau.ac.il/PatchDock/	Geometric hashing and pose-clustering	Ranking according to a geometric shape complementarity score
PyDock [157] http://mmb.pcb.ub.es/PyDock/	FFT based on electrostatics and desolvation energy	Ranking using an energy function
RosettaDock [115] http://rosettadock.graylab.jhu.edu/	Local docking by Monte Carlo search	Ranking using an energy function, clustering
ZDOCK [107] http://zlab.bu.edu/zdock/index.shtml	FFT correlation based on shape complementarity, desolvation energy and electrostatics	Refinement by energy minimization
3D-Dock [158] http://www.sbg.bio.ic.ac.uk/docking/	FFT correlation using FTDOCK [159]	Clustering, refinement of side-chains using Multidock [159]

Macromolecular docking - programs

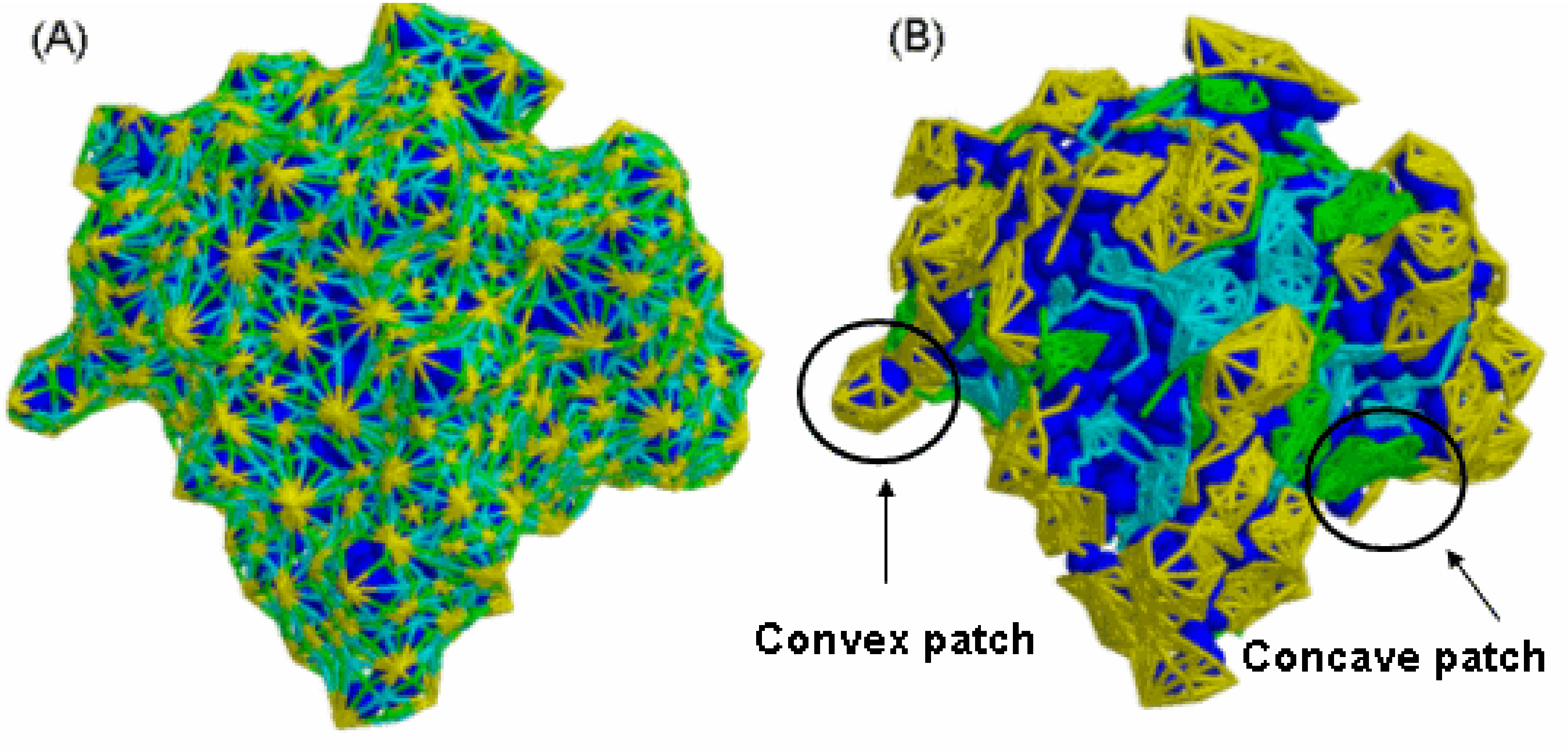
- ClusPro 2.0
  - http://cluspro.bu.edu/
  - Performs a global soft rigid-body search using PIPER docking program; employs knowledge-based potential
  - The top 1,000 structures are retained and clustered to isolate highly populated low-energy binding modes
  - A special mode for prediction of molecular assemblies of homo-oligomers

Macromolecular docking - programs

- PatchDock
  - http://bioinfo3d.cs.tau.ac.il/PatchDock/index.html
  - Performs a geometry-based search for docking transformations that yield good molecular shape complementarity (driven by local feature matching rather than brute force searching of the 6D space):
    1. The molecular surface is divided into concave, convex and flat patches
    2. Complementary patches are matched → candidate transformations
    3. Evaluation of each docking candidate by a scoring function considering both geometric fit and atomic desolvation energy
    4. Clustering of the candidate solutions to discard redundant solutions
  - Results can be redirected to FireDock for refinement and re-scoring

Macromolecular docking - programs

- PatchDock



Macromolecular docking - programs

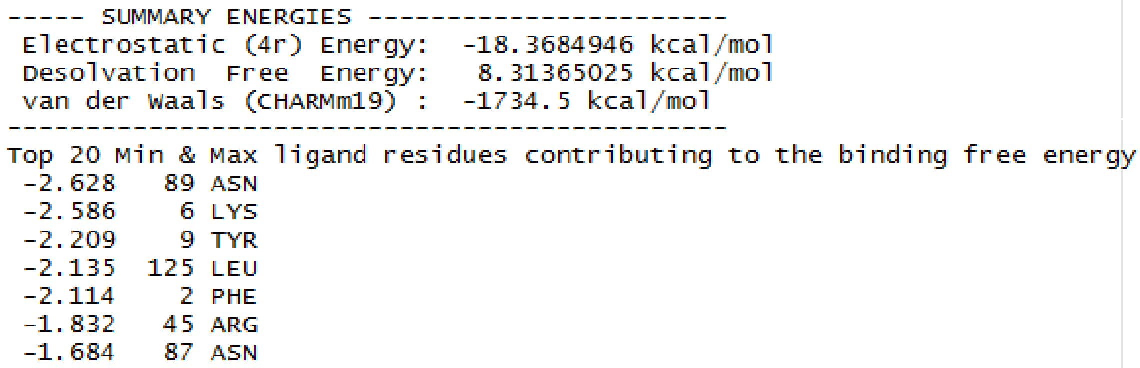
- FireDock
  - http://bioinfo3d.cs.tau.ac.il/FireDock/index.html
  - Refines and re-scores solutions produced by fast rigid-body docking algorithms
  - Optimizes the binding of each candidate by allowing flexibility in the side-chains and adjustments of the relative orientation of the molecules
  - Scoring of the refined candidates is based on softened van der Waals interactions, atomic contact energy, electrostatic, and additional binding free energy estimations

Analysis of macromolecular complexes

- ❑ Binding energy
- ❑ Macromolecular interface
- ❑ Interaction hot spots

Binding energy

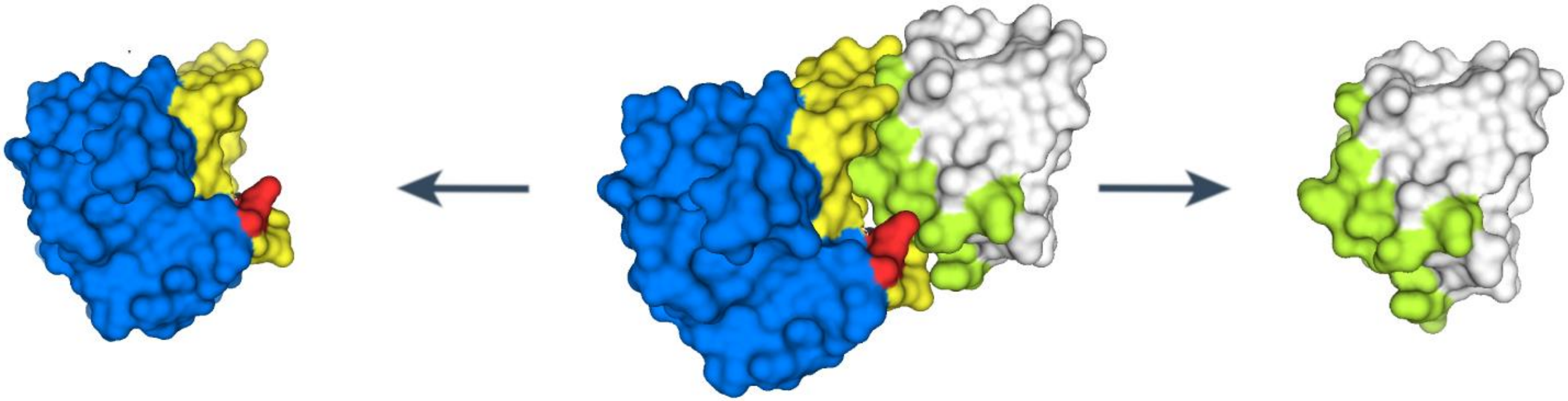
- FastContact
  - http://structure.pitt.edu/servers/fastcontact/
  - Rapidly estimates the electrostatic and desolvation components of the binding free energy between two proteins
  - Additionally, evaluates the van der Waals interactions using CHARMM and reports contribution of individual residues and pairs of residues to the free energy → highlight the interaction hot spots



```
----- SUMMARY ENERGIES -----------------------
 Electrostatic (4r) Energy:  -18.3684946 kcal/mol
 Desolvation  Free  Energy:   8.31365025 kcal/mol
 van der Waals (CHARMm19) :  -1734.5 kcal/mol
----------------------------------------------
Top 20 Min & Max ligand residues contributing to the binding free energy
 -2.628   89 ASN
 -2.586    6 LYS
 -2.209    9 TYR
 -2.135  125 LEU
 -2.114    2 PHE
 -1.832   45 ARG
 -1.684   87 ASN
```

Macromolecular interface

- The region where two protein chains or protein and nucleic acid chain come into contact
- Can be identified by the analysis of the 3D structure of the macromolecular complex



Interface analysis

- Provides information about basic features of macromolecular complexes interactions (e.g., shape complementarity, chemical complementarity,...)
- Provides information about interface residues
- Acquired information is useful for a wide range of applications
    - Design of mutants for experimental verification of the interactions
    - Development of drugs targeting macromolecular interactions
    - Understanding the mechanism of the molecular recognition
    - Computational prediction of interfaces and complex 3D structures
    - ...

Interface analysis

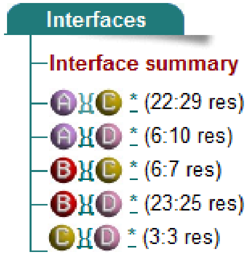
- Most common approaches for the definition of interfaces:
  - Methods based on the distance between interacting residues
  - Methods based on the change in the solvent accessible surface area (ASA) upon complex formation
  - Computational geometry methods (using Voronoi diagrams)
- All three approaches provide very similar results

Interface analysis - databases

- PDBsum (Pictorial database of 3D structures in the Protein Data Bank)
  - http://www.ebi.ac.uk/pdbsum/
  - Provides numerous structural analyses for all PDB structures and AlphaFold DB (human proteins), including information about protein-protein and protein-nucleic acid interfaces
  - Protein-protein interactions – schematic diagrams of all protein-protein interfaces and corresponding residue-residue interactions
  - Protein-nucleic acid interactions – schematic diagrams of protein-nucleic acid interactions generated by NUCPLOT

Interface analysis - databases

- PDBsum

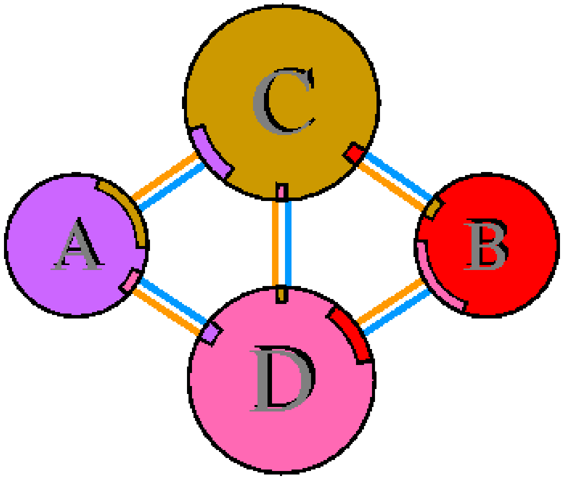


Interfaces

Interface summary

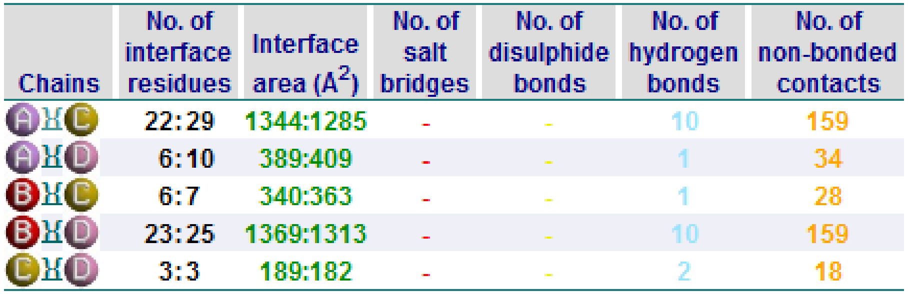
- A–C * (22:29 res)
- A–D * (6:10 res)
- B–C * (6:7 res)
- B–D * (23:25 res)
- C–D * (3:3 res)

Interfaces summary for 1fq9



Key: 🟥 Salt bridges 🟨 Disulphide bonds 🟦 Hydrogen bonds 🟧 Non-bonded contacts

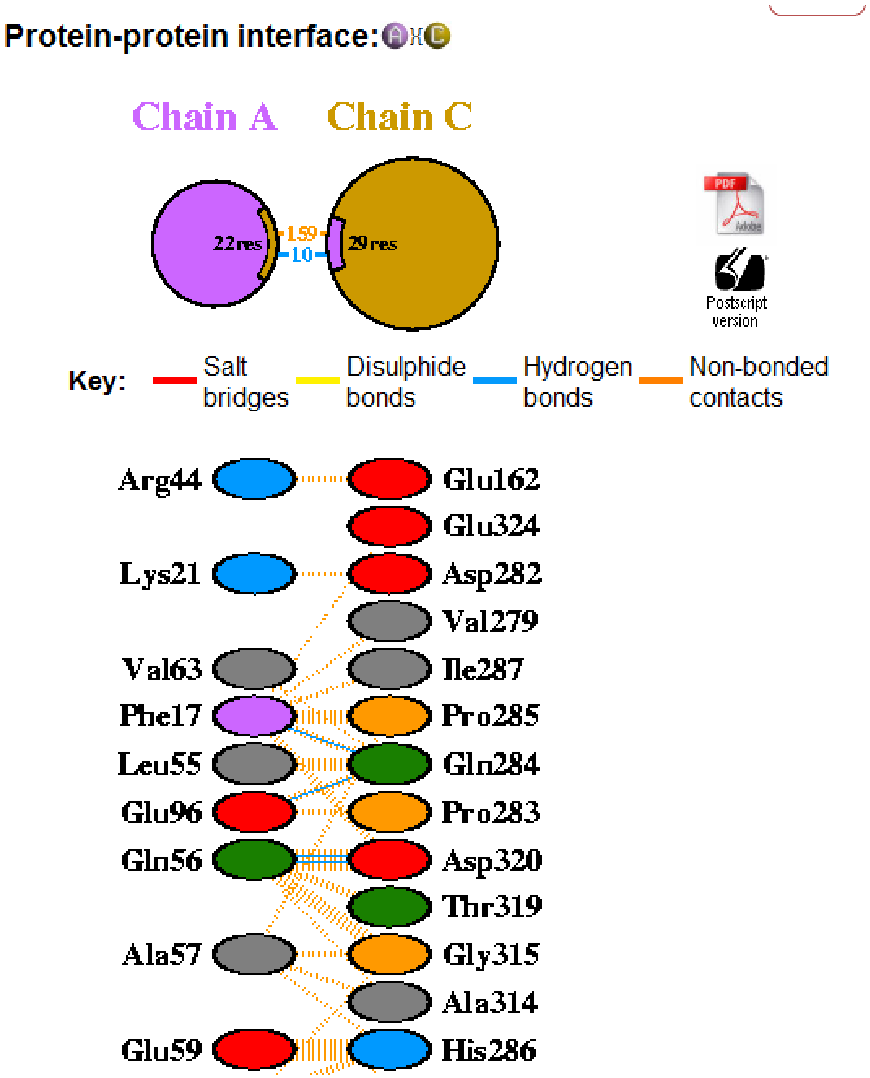
Interface statistics



Chains	No. of interface residues	Interface area (Å^2)	No. of salt bridges	No. of disulphide bonds	No. of hydrogen bonds	No. of non-bonded contacts
A–C	22:29	1344:1285	-	-	10	159
A–D	6:10	389:409	-	-	1	34
B–C	6:7	340:363	-	-	1	28
B–D	23:25	1369:1313	-	-	10	159
C–D	3:3	189:182	-	-	2	18

Interface analysis - databases

- PDBsum



Interface analysis - tools

- Analyze interface of a given macromolecular complex
    - PISA (Protein Interfaces, Surfaces and Assemblies)
    - MolSurfer
    - Contact Map WebViewer
    - PIC (Protein Interaction Calculator)
    - …

Interface analysis - tools

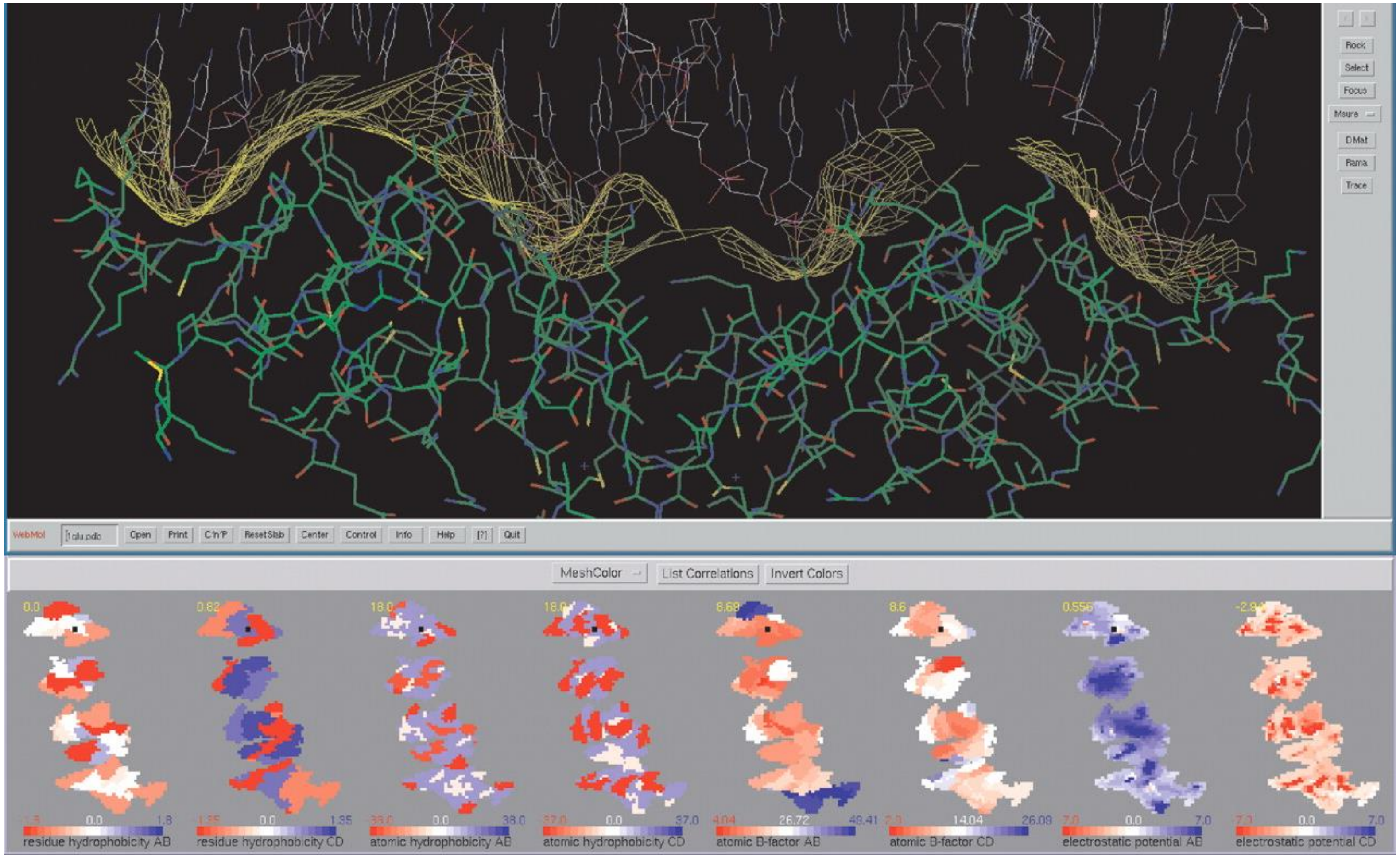
- PISA (**P**rotein **I**nterfaces, **S**urfaces and **A**ssemblies)
    - www.pdbe.org/pisa
    - An interactive tool for the exploration of macromolecular interfaces (protein, DNA/RNA and ligands), prediction of probable quaternary structures, database searches of structurally similar interfaces and assemblies
    - Overview and detailed characteristics of all interfaces found within a given structure (including those generated by symmetry operations)
    - Provides interface area, $\Delta^{i}G$, potential hydrogen bonds and salt bridges, interface residues and atoms, ...

Interface analysis - tools

- MolSurfer
  - http://projects.villa-bosch.de/dbase/molsurfer/index.html
  - Visualization of 2D projections of protein-protein and protein-nucleic acid interfaces as maps showing a distribution of interface properties (atomic and residue hydrophobicity, electrostatic potential, surface-surface distances, atomic distances,...)
  - 2D maps are linked with the 3D view of a macromolecular complex
  - Facilitates the study of intermolecular interaction properties and steric complementarity between macromolecules

Interface analysis - tools

- ❑ MolSurfer



Interface analysis - tools

- Contact Map WebViewer
  - http://cmweb.enzim.hu/
  - Represents residue-residue contacts within a protein or between proteins in a complex in the form of a contact map

- PIC (**P**rotein **I**nteraction **C**alculator)
  - http://pic.mbu.iisc.ernet.in/
  - Identifies various interactions within a protein or between proteins in a complex

Interaction hotspots

- Hot spots: the residues contributing the most to the binding free energy of the complex
- Knowledge of hot spots has important implications to:
  - Understand the principles of protein interactions (an important step to understand recognition and binding processes)
  - Design of mutants for experimental verification of the interactions
  - Development of drugs targeting macromolecular interactions
  - ...

Interaction hotspots

- Hot spots are usually conserved and appear to be clustered in tightly packed regions in the center of the interface
- Experimental identification by alanine scanning mutagenesis → if a residue has a significant drop in binding affinity when mutated to alanine it is labeled as a hot spot
- Experimental identification of hot spots is costly and cumbersome → the computational predictions of hot spots can help!

Prediction of hotspots - tools

- Most of the available methods are based on the 3D structure of the complex
- Knowledge-based methods
  - Combination of several physicochemical features
  - Evolutionary conservation, ASA, residue propensity, structural location, hydrophobicity,...)
- Energy-based methods
  - Calculation of the change in the binding free energy ($\Delta\Delta G_{bind}$) of the complex upon *in silico* modification of a given residue to alanine

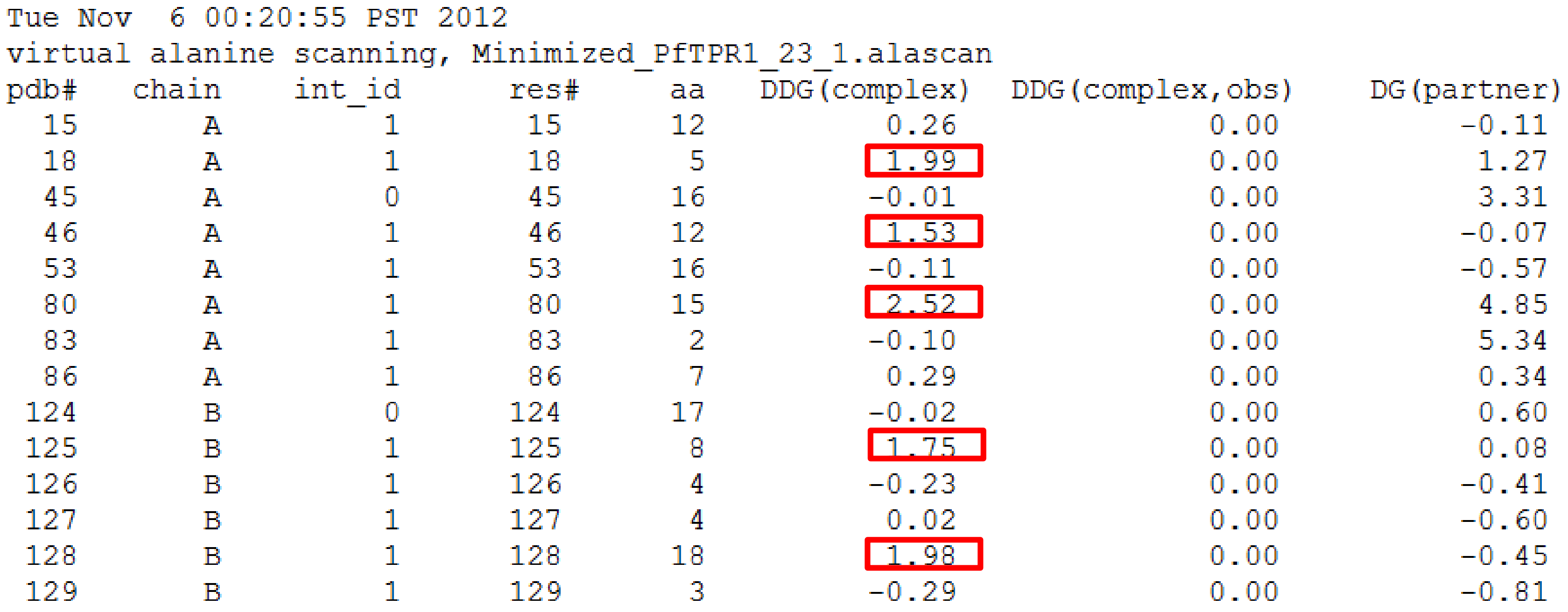
Prediction of hotspots - tools

- ☐ Robetta
  - http://old.robetta.org/alascansubmit.jsp
  - Energy-based method
  - Performs *in silico* alanine scanning mutagenesis of protein-protein or protein-DNA interface residues

1. The side chain of each interface residue is mutated to methyl
2. All side chains within 5 Å radius sphere of the mutated residue are repacked; the rest of the protein remains unchanged
3. For each mutant, $\Delta\Delta G_{bind}$ is calculated (residues with predicted $\Delta\Delta G_{bind} \geq +1$ kcal/mol = hot spot)

Prediction of hotspots - tools

- Robetta



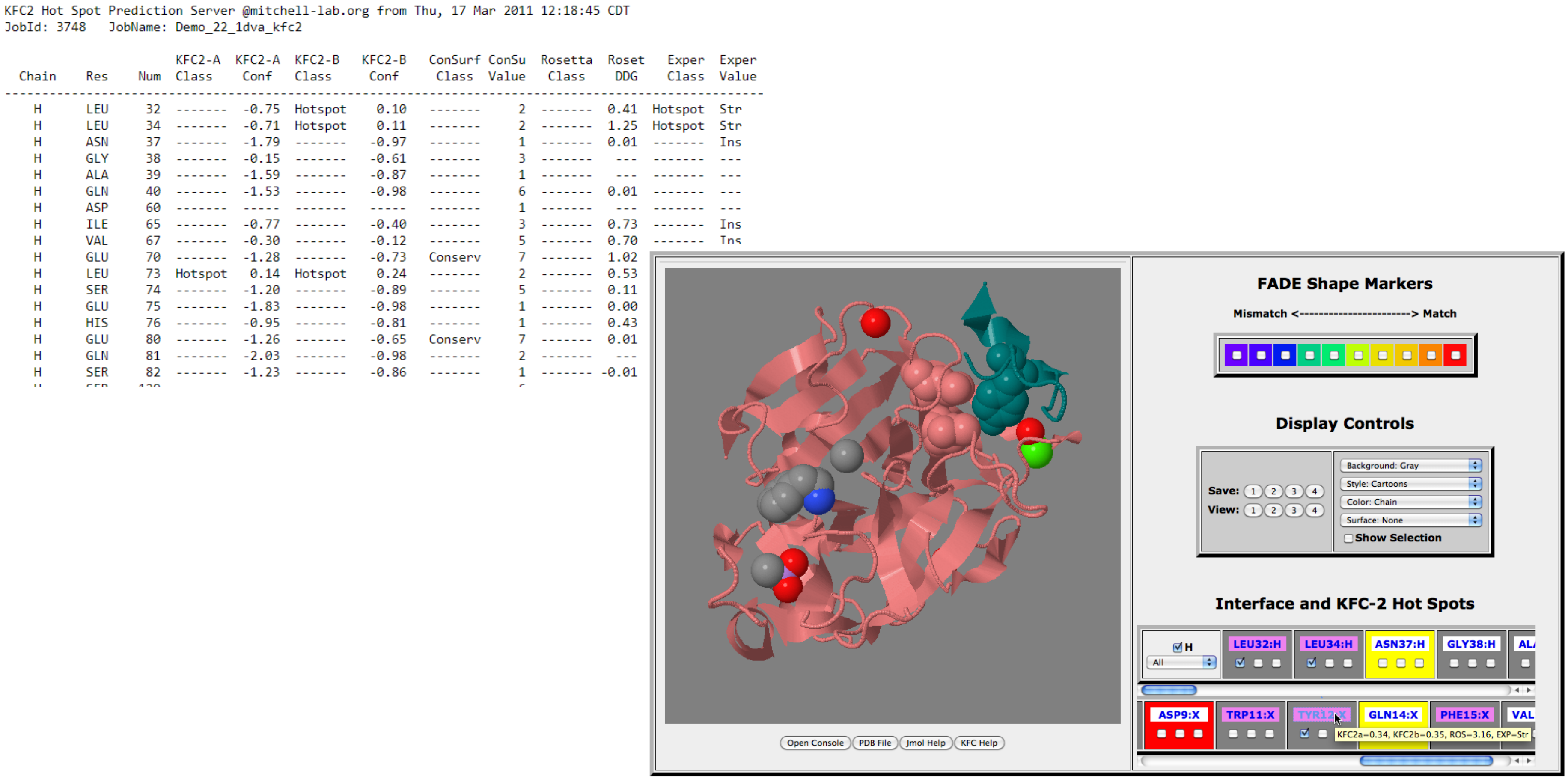
```
Tue Nov  6 00:20:55 PST 2012
virtual alanine scanning, Minimized_PfTPR1_23_1.alascan
pdb#    chain     int_id       res#      aa    DDG(complex)  DDG(complex,obs)     DG(partner)
  15        A          1         15      12            0.26              0.00           -0.11
  18        A          1         18       5            1.99              0.00            1.27
  45        A          0         45      16           -0.01              0.00            3.31
  46        A          1         46      12            1.53              0.00           -0.07
  53        A          1         53      16           -0.11              0.00           -0.57
  80        A          1         80      15            2.52              0.00            4.85
  83        A          1         83       2           -0.10              0.00            5.34
  86        A          1         86       7            0.29              0.00            0.34
 124        B          0        124      17           -0.02              0.00            0.60
 125        B          1        125       8            1.75              0.00            0.08
 126        B          1        126       4           -0.23              0.00           -0.41
 127        B          1        127       4            0.02              0.00           -0.60
 128        B          1        128      18            1.98              0.00           -0.45
 129        B          1        129       3           -0.29              0.00           -0.81
```

Prediction of hotspots - tools

- KFC2 (**K**nowledge-based **F**ADE and **C**ontacts)
  - https://mitchell-web.ornl.gov/KFC_Server/
  - Knowledge-based method utilizing machine learning
  - Predicts hot spots in protein-protein interfaces by recognizing features of important binding contacts – solvent accessibility, residue position within the interface, packing density, residue size, flexibility and hydrophobicity of residues around the target residue
  - Optionally, user can provide data to improve the prediction (ConSurf conservation scores, Rosetta alanine scanning results or experimental data)

Prediction of hotspots - tools

- KFC2 (**K**nowledge-based **F**ADE and **C**ontacts)



```
KFC2 Hot Spot Prediction Server @mitchell-lab.org from Thu, 17 Mar 2011 12:18:45 CDT
JobId: 3748   JobName: Demo_22_1dva_kfc2

                          KFC2-A  KFC2-A  KFC2-B   KFC2-B   ConSurf ConSu  Rosetta  Roset   Exper  Exper
  Chain    Res     Num    Class    Conf   Class     Conf     Class  Value   Class    DDG    Class  Value
-----------------------------------------------------------------------------------------------------------
    H      LEU      32   -------  -0.75  Hotspot   0.10    -------     2   -------  0.41  Hotspot  Str
    H      LEU      34   -------  -0.71  Hotspot   0.11    -------     2   -------  1.25  Hotspot  Str
    H      ASN      37   -------  -1.79  -------  -0.97    -------     1   -------  0.01  -------  Ins
    H      GLY      38   -------  -0.15  -------  -0.61    -------     3   -------   ---  -------  ---
    H      ALA      39   -------  -1.59  -------  -0.87    -------     1   -------   ---  -------  ---
    H      GLN      40   -------  -1.53  -------  -0.98    -------     6   -------  0.01  -------  ---
    H      ASP      60   -------  -----  -------  -----    -------     1   -------   ---  -------  ---
    H      ILE      65   -------  -0.77  -------  -0.40    -------     3   -------  0.73  -------  Ins
    H      VAL      67   -------  -0.30  -------  -0.12    -------     5   -------  0.70  -------  Ins
    H      GLU      70   -------  -1.28  -------  -0.73    Conserv     7   -------  1.02
    H      LEU      73   Hotspot   0.14  Hotspot   0.24    -------     2   -------  0.53
    H      SER      74   -------  -1.20  -------  -0.89    -------     5   -------  0.11
    H      GLU      75   -------  -1.83  -------  -0.98    -------     1   -------  0.00
    H      HIS      76   -------  -0.95  -------  -0.81    -------     1   -------  0.43
    H      GLU      80   -------  -1.26  -------  -0.65    Conserv     7   -------  0.01
    H      GLN      81   -------  -2.03  -------  -0.98    -------     2   -------   ---
    H      SER      82   -------  -1.23  -------  -0.86    -------     1   -------  -0.01
```

Background: Gray
Style: Cartoons
Color: Chain
Surface: None
☐ Show Selection

KFC2a=0.34, KFC2b=0.35, ROS=3.16, EXP=Str

References I

- Liljas, A. *et al*. (2009). **Textbook Of Structural Biology**, World Scientific Publishing Company, Singapore.
- Goodsell, D. S. & Olson, A. J. (2000) Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure* **29**: 105-153.
- Demachenko, A. P. (2001). Recognition between flexible protein molecules: induced and assisted folding. *Journal of Molecular Recognition* **14**: 42-61.
- Ali, M. H. & Imperiali, B. (2005) Protein oligomerization: How and why. *Bioorganic & Medicinal Chemistry* **13**: 5013-5020.
- Jahn, T. R. & Radford, S. E. (2008) Folding *versus* aggregation: Polypeptide conformations on competing pathways. *Archives of Biochemistry and Biophysics* **469**: 100-117.
- Csermely, P. *et al.* (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **35**: 539-546.

References II

- Bujnicki, J. (2009). **Prediction of Protein Structures, Functions, and Interactions**, John Wiley & Sons, Ltd., Chichester, p. 302.
- Tramontano, A. (2005). **The Ten Most Wanted Solutions in Protein Bioinformatics**, CRC Press UK, London, p. 186.
- Tuncbag, N., *et al.* (2009). A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Briefings in bioinformatics* **10**: 217-232.
- Ezkurdia, I., *et al.* (2009). Progress and challenges in predicting protein-protein interaction sites. *Briefings in bioinformatics* **10**: 233-246.
- Fernández-Recio, J. (2011). Prediction of protein binding sites and hot spots. *Computational molecular science* **6**: 680-698.
- Szilagyi, A., *et al.* (2005). Prediction of physical protein–protein interactions. *Physical biology* **2**: S1-S16.
- Moreira, I. S., *et al.* (2010). Protein-protein docking dealing with the unknown. *Journal of computational chemistry* **31**:317-342