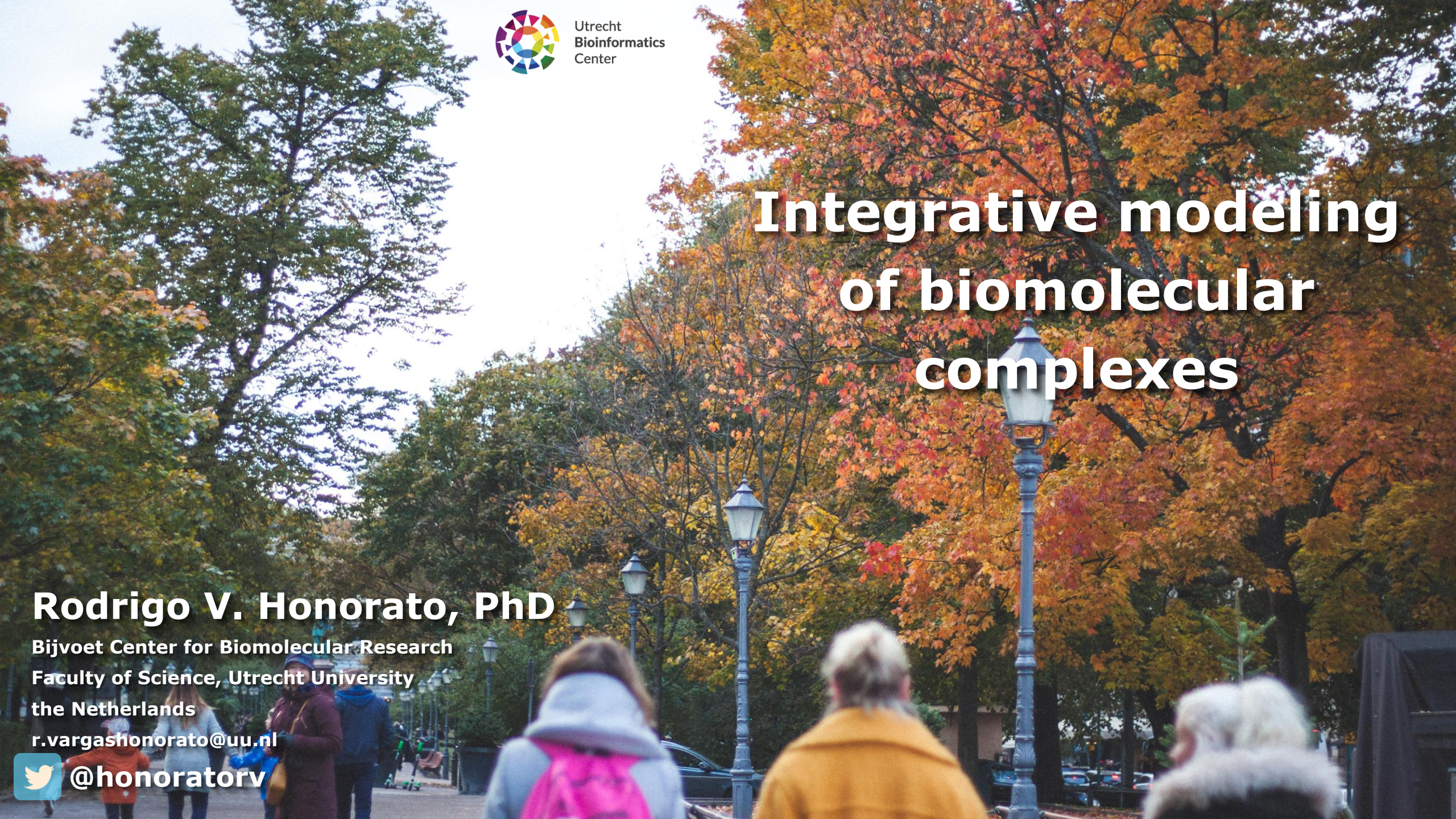


Utrecht Bioinformatics Center

Integrative modeling of biomolecular complexes

Rodrigo V. Honorato, PhD

Bijvoet Center for Biomolecular Research

Faculty of Science, Utrecht University

the Netherlands

r.vargashonorato@uu.nl

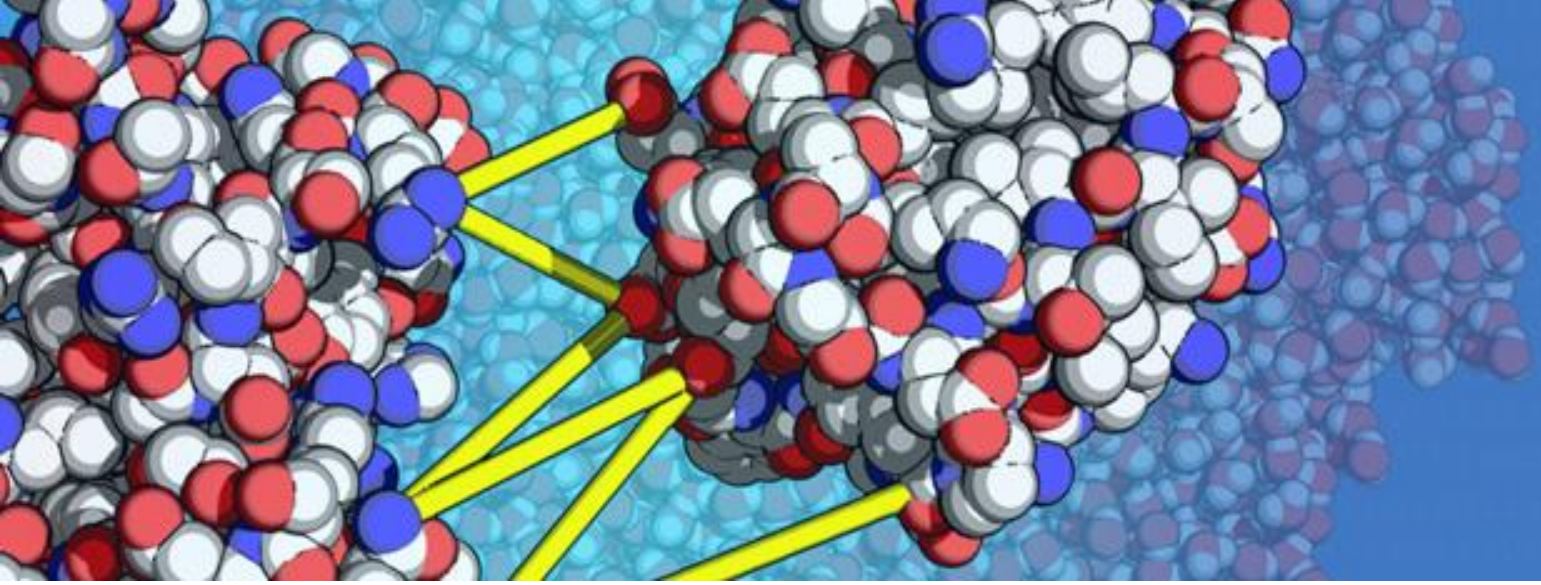
@honoratorv

Solution NMR: 950, 900-cryo, 750, 600-cryo, 600US, 2x500 MHz
Solid-state NMR: 800WB-DNP, 400WB-DNP, 700US, 500WB MHz

e-infrastructure: >1900 CPU cores + EGI grid (>110'000 CPU cores)

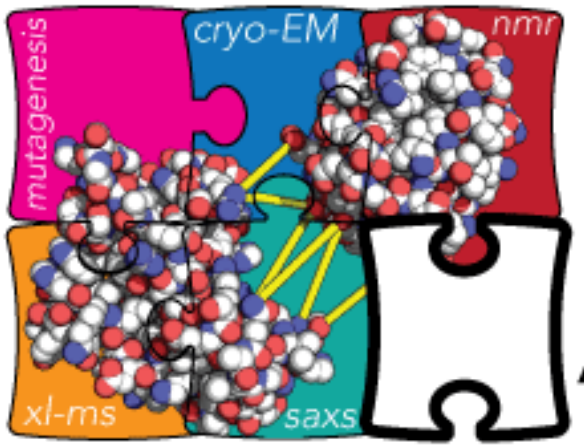
National uNMR-NL and European infrastructure iNEXT

MoBrain we-nmr West-Life bioexcel INDIGO - DataCloud EOSC-hub



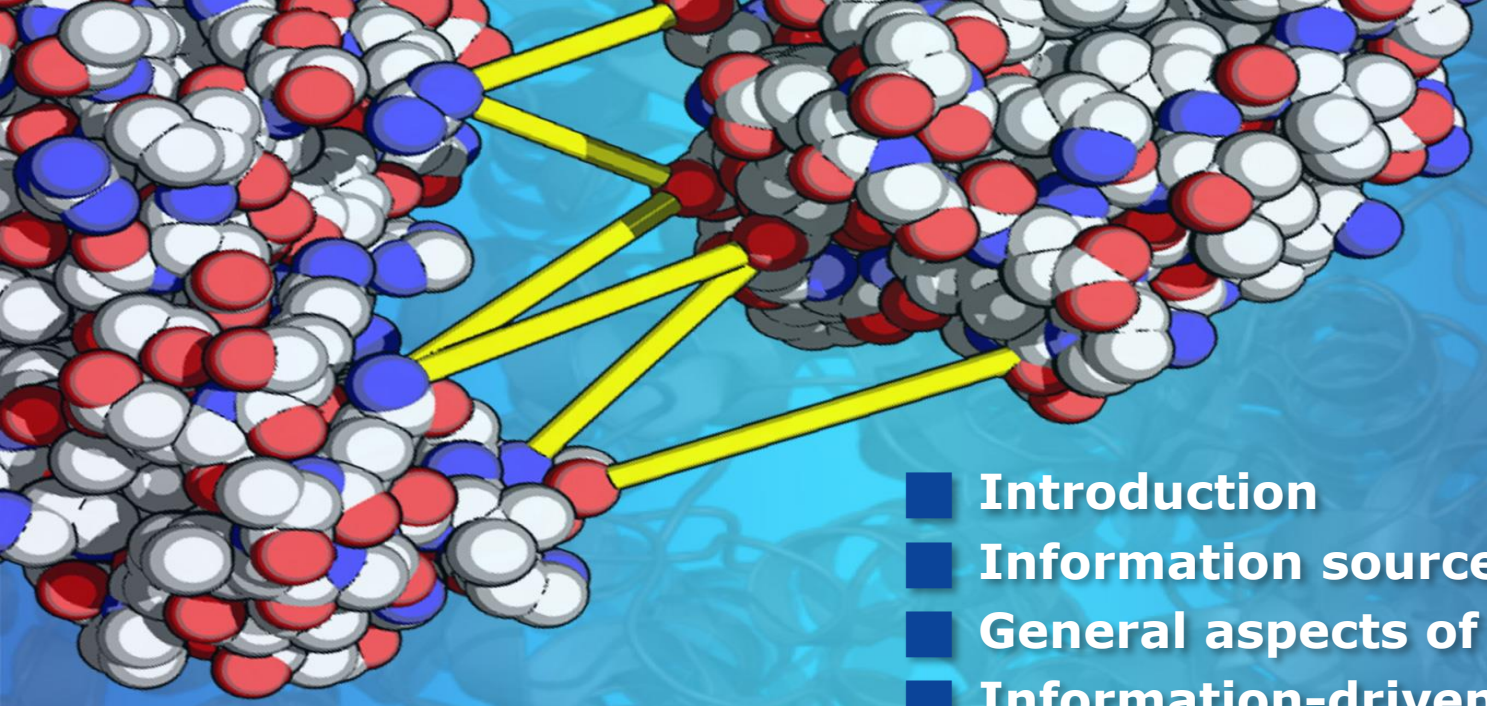
Bonvin Lab

computational structural biology



HADDOCK

High-Ambiguity Driven Docking



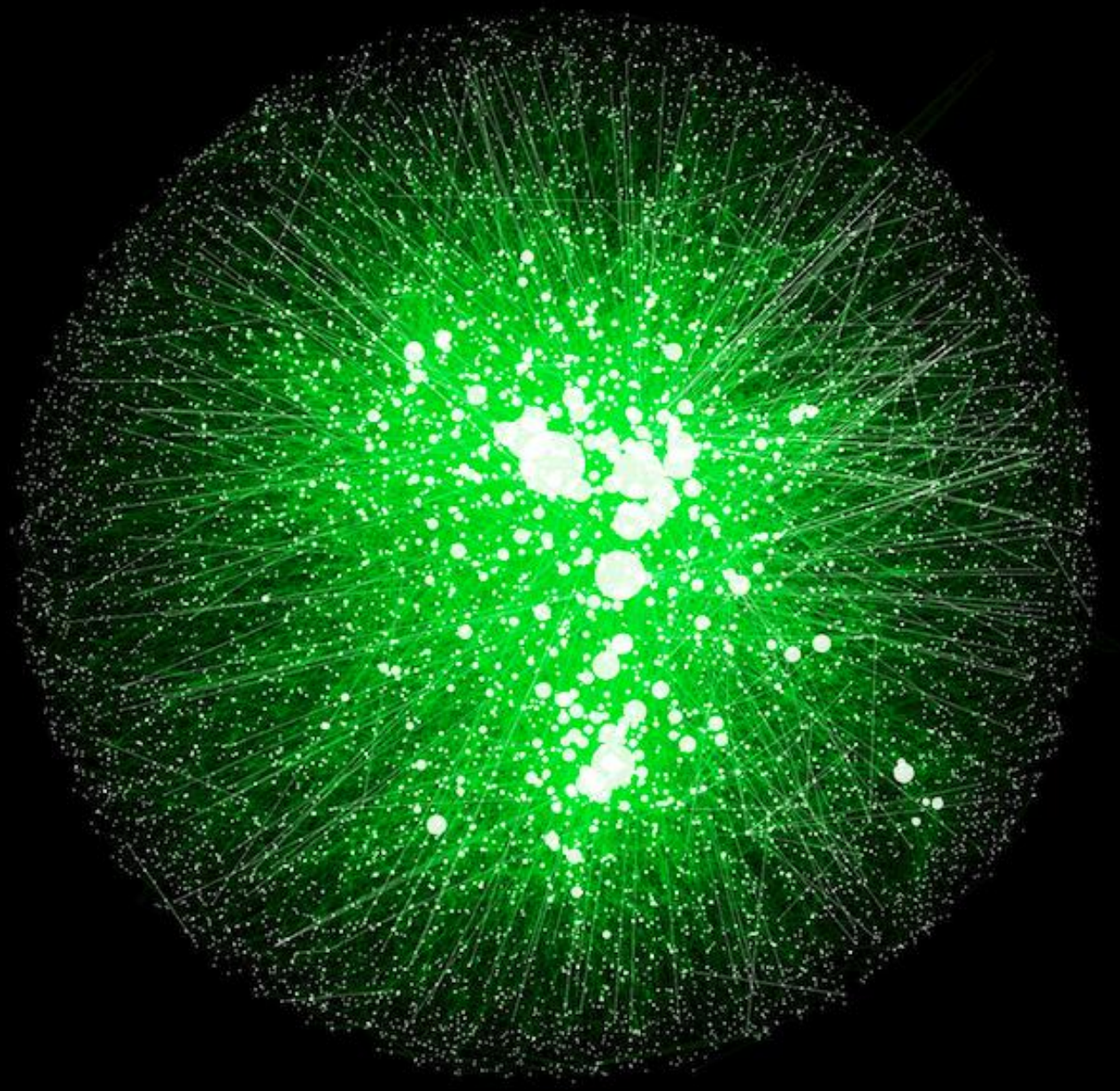
Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
- Assessing the interaction space
- Conclusions & perspectives

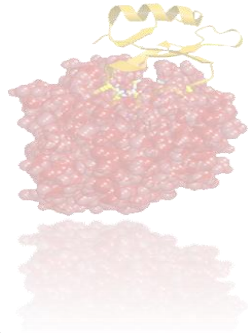
The social network of proteins

Majority of 'life' depends on interactions, particularly protein-protein

The protein-protein interaction Cosmos



Structural biology of interactions

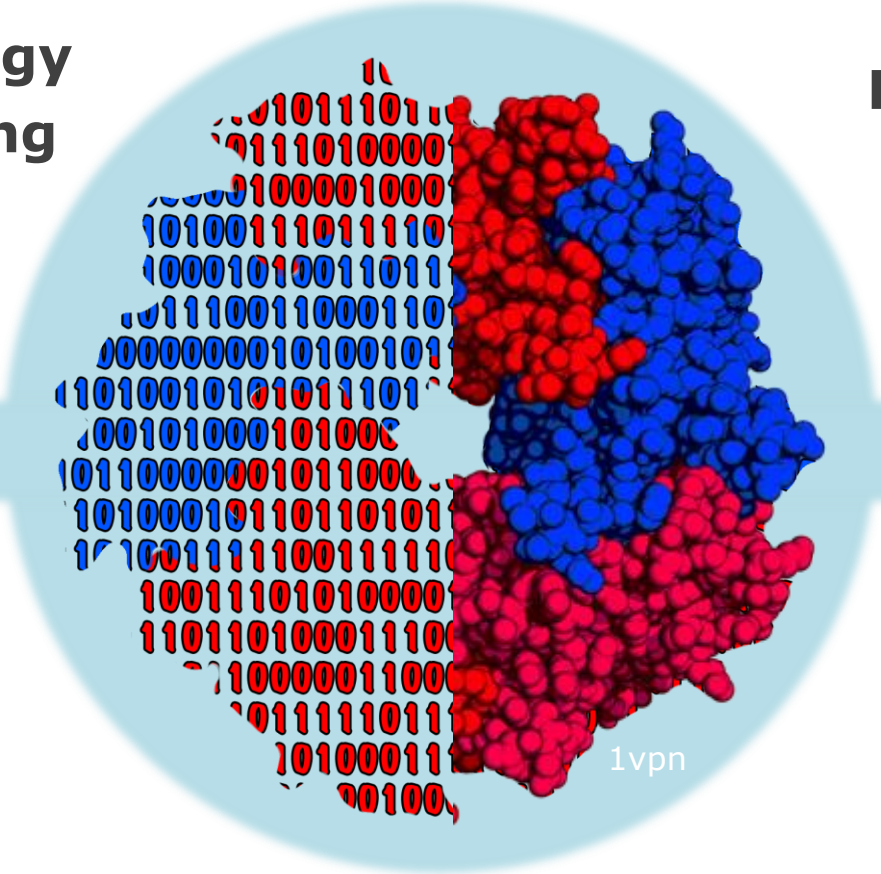


Docking

Homology Modeling

Molecular Dynamics

Threading



SAS

MS

FRET

X-Ray

EPR

Cryo-EM

NMR

Computation

Experiment

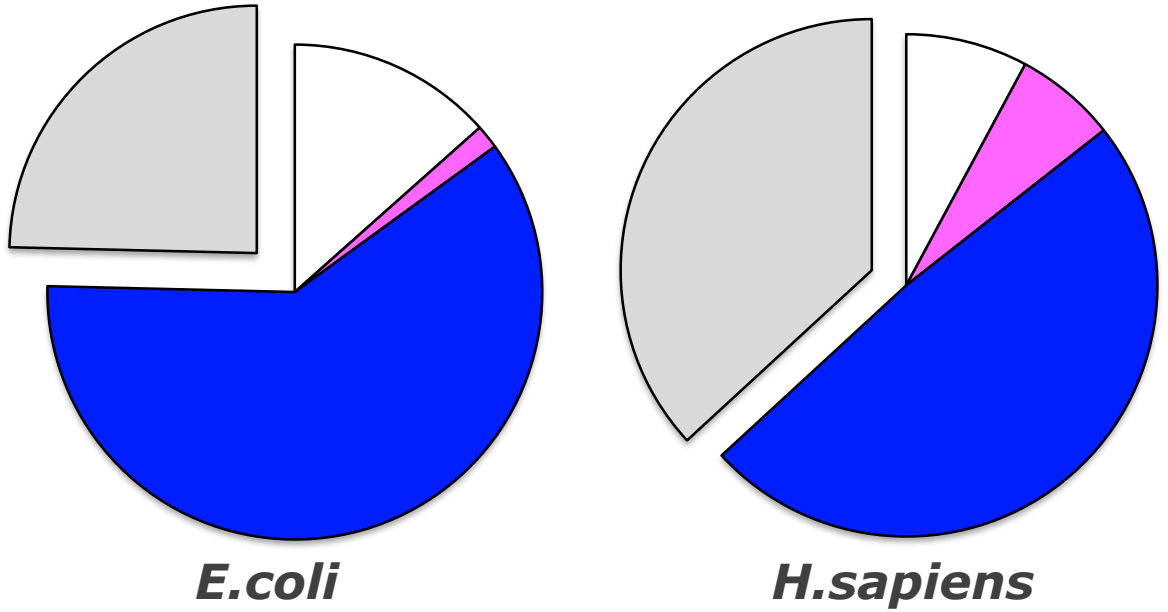
High-throughput computation vs. High-resolution experiments

computational models are often not trusted by the experimental community

Structural coverage of interactomes

Unique interactions in interactomes

- ~7,500 binary interactions in *E.coli*
- ~44,900 binary interactions in *H.sapiens*

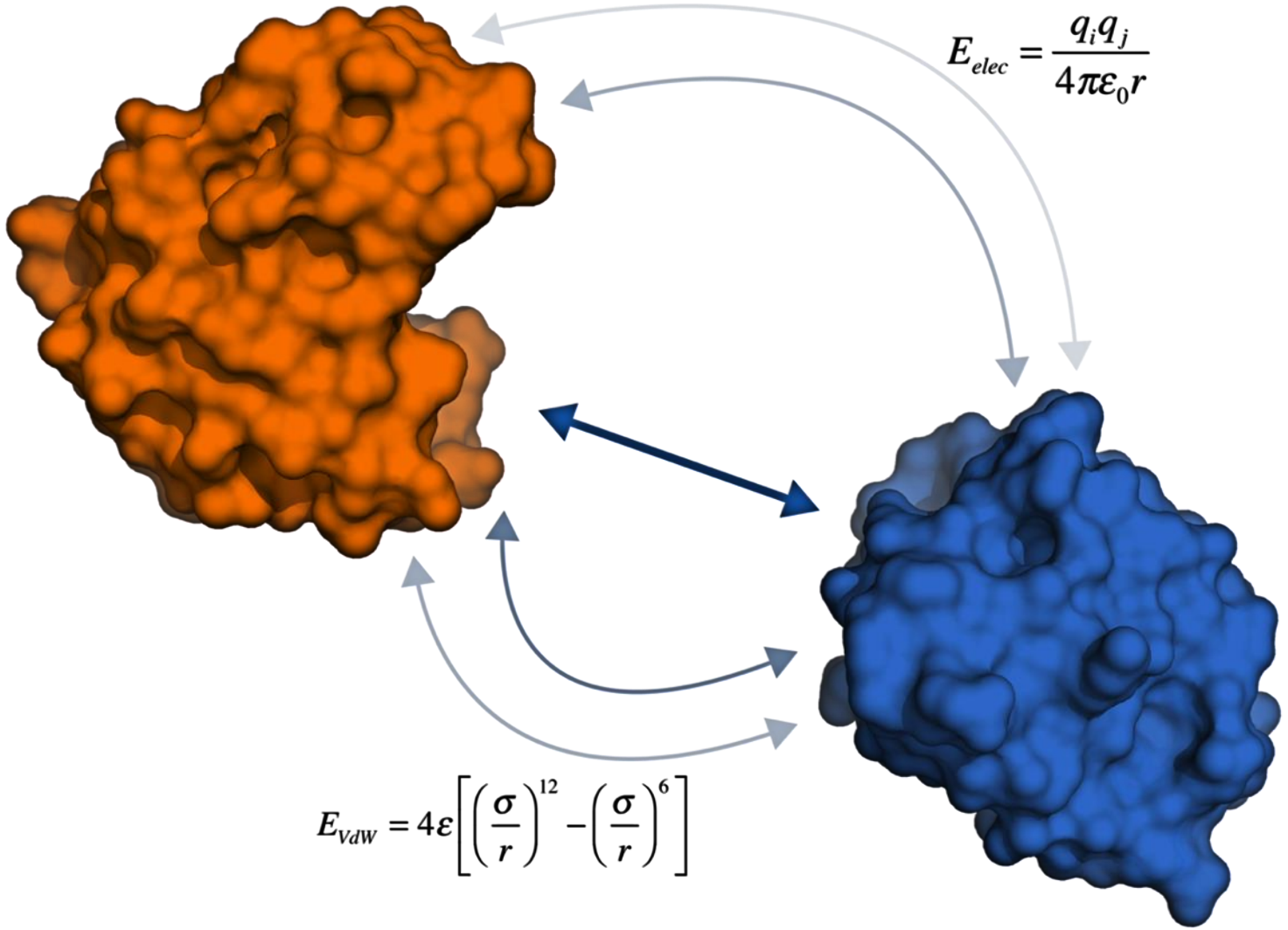


- with complete structures
- with partial (domain-domain) or complete models
- with structures for the interactors (suitable for docking)
- without structural data

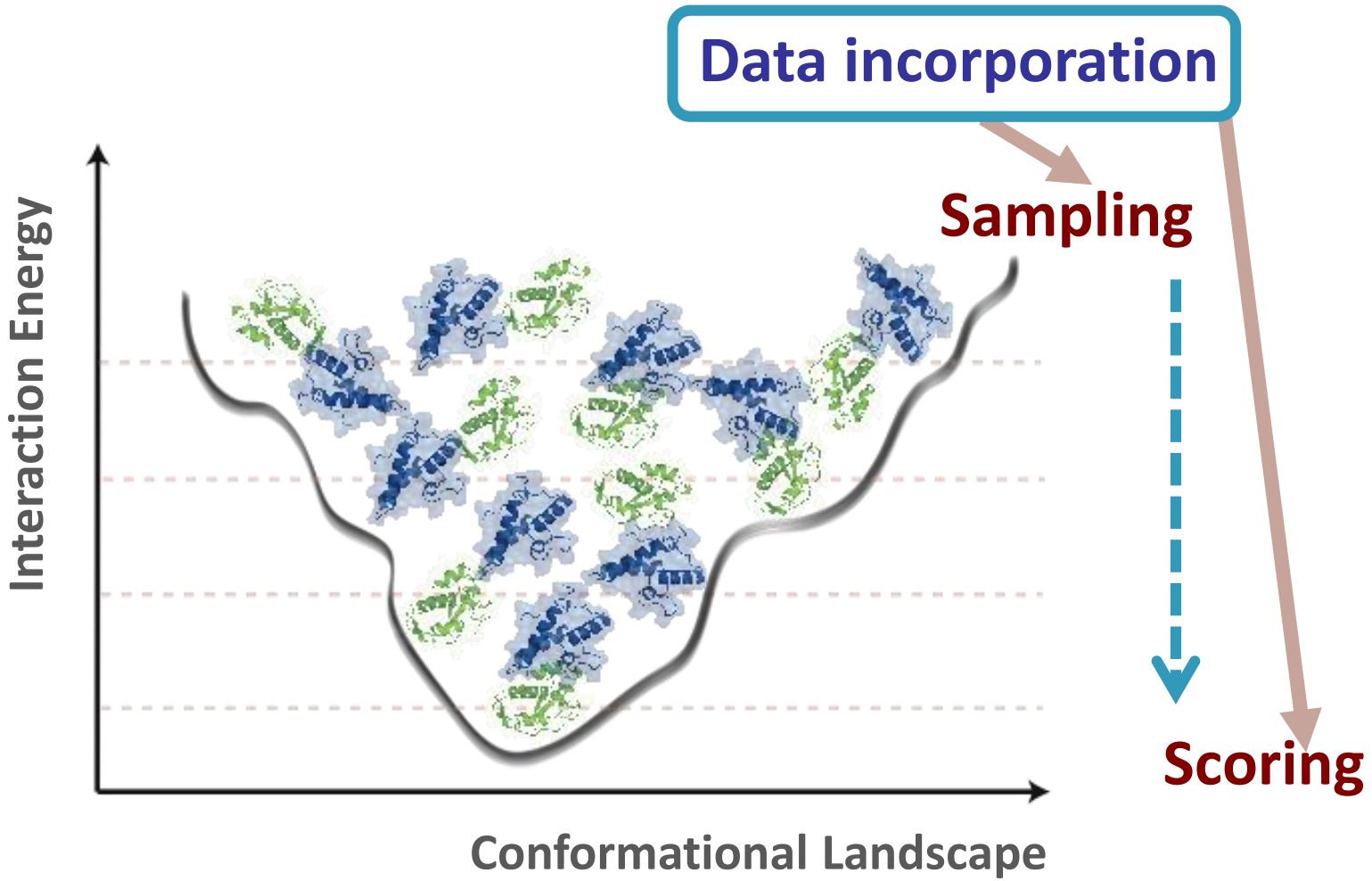
Statistics from Interactome3D (2013-01)

Mosca *et al.* Nature Methods 2013

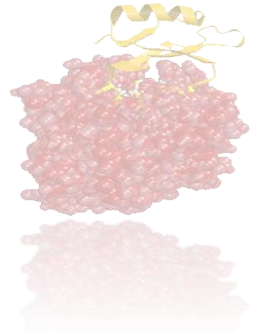
Molecular Docking



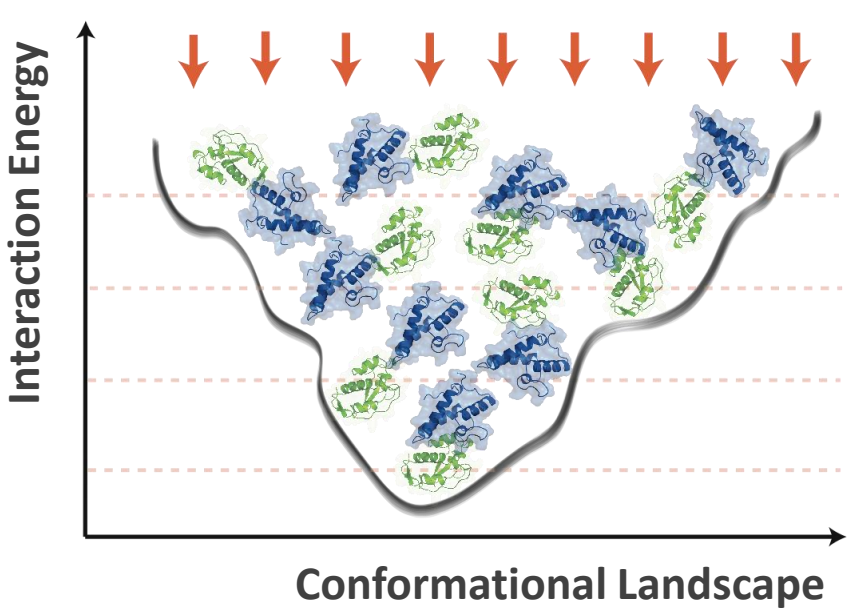
Methodology



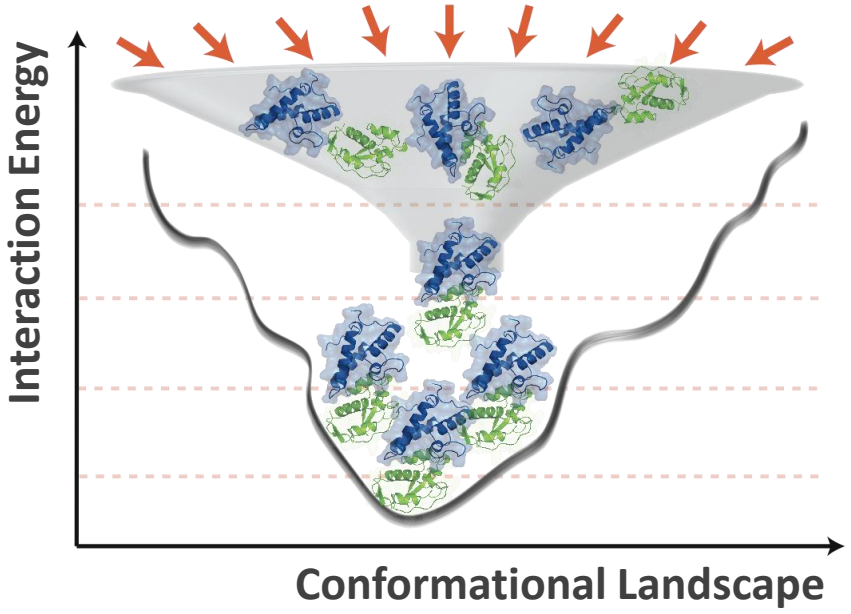
Data Integration during Sampling



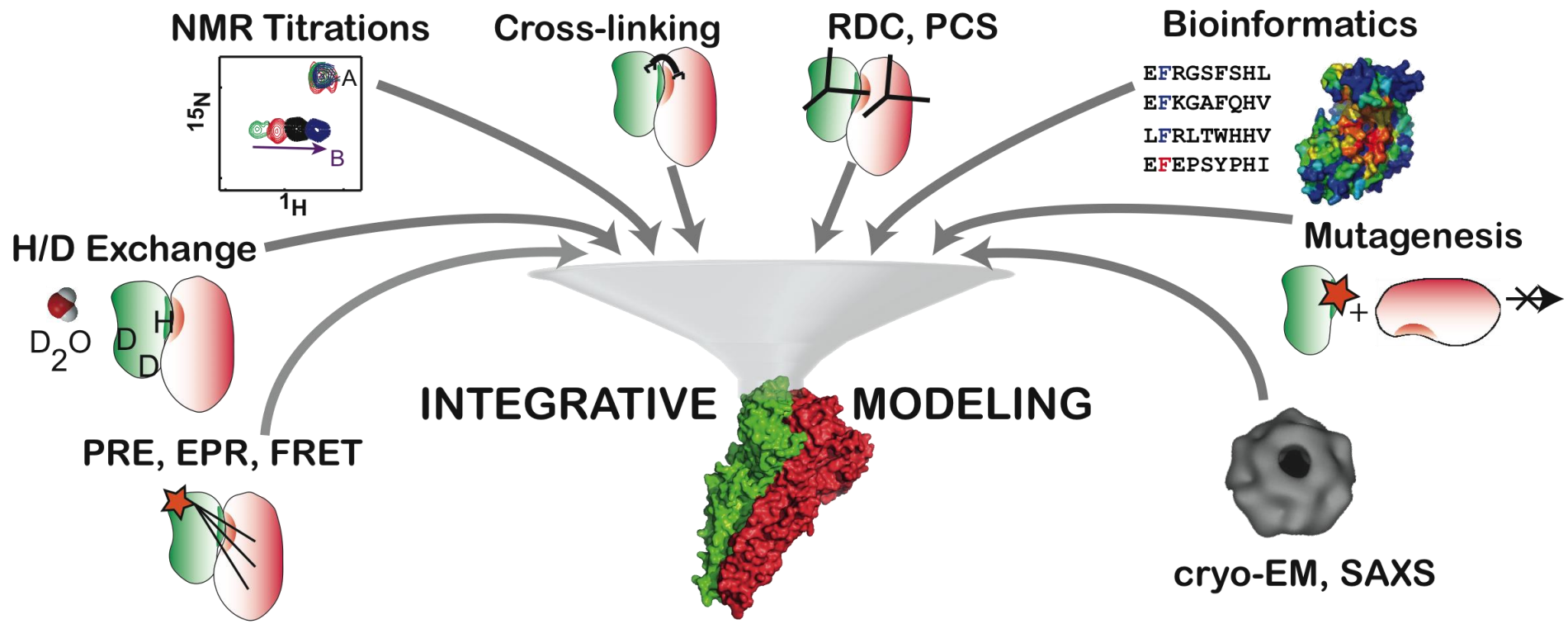
Global Search



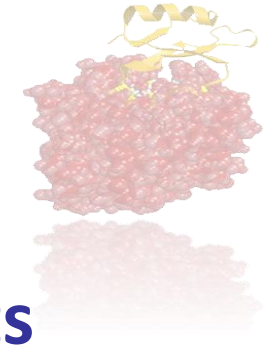
Information-driven Search



What is Integrative Modeling?



Why integrative modelling?



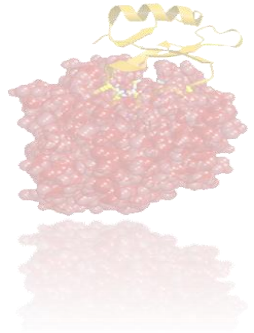
For Experimentalists

- ✓ New hypothesis to drive experiments
- ✓ Speed up structure determination
- ✓ Increase our understanding of function

For Modelers

- ✓ Decrease high false positive rate
- ✓ Ease accuracy assessment

Related reviews



- Halperin *et al.* (2002) **Principles of docking: an overview of search algorithms and a guide to scoring functions.** *PROTEINS: Struc. Funct. & Genetics* **47,** 409-443.

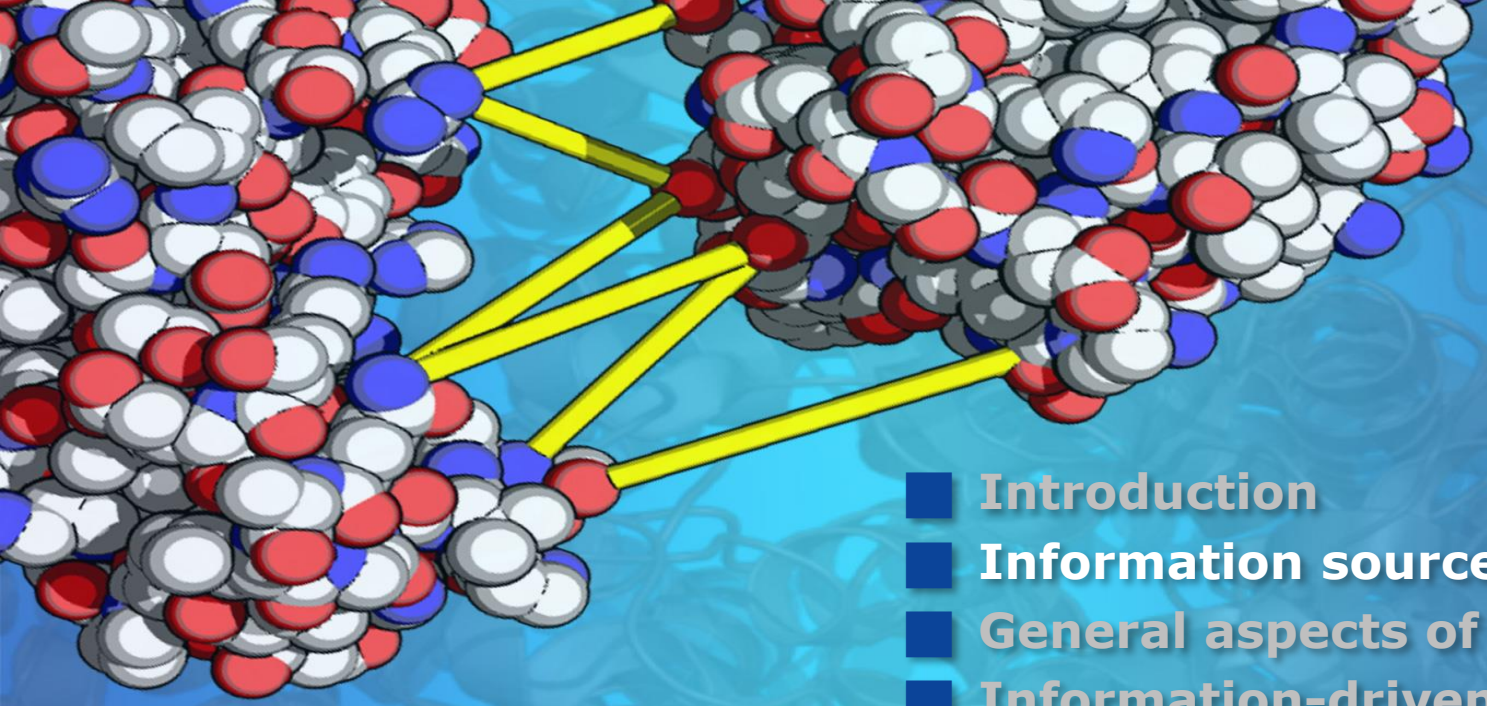
- **Special issues of *PROTEINS:*** (2003) (2005) (2007) (2010) (2013) and (2016), which are **dedicated to CAPRI.**

- de Vries SJ and Bonvin AMJJ (2008). **How proteins get in touch: Interface prediction in the study of biomolecular complexes**. *Curr. Pept. and Prot. Research* **9**, 394-406.

- Melquiond ASJ, Karaca E, Kastritis PL and Bonvin AMJJ (2012). **Next challenges in protein-protein docking: From proteome to interactome and beyond**. *WIREs Computational Molecular Science* **2**, 642-651 (2012).

- Karaca E and Bonvin AMJJ (2013). **Advances in integrated modelling of biomolecular complexes**. *Methods*, **59**, 372-381 (2013).

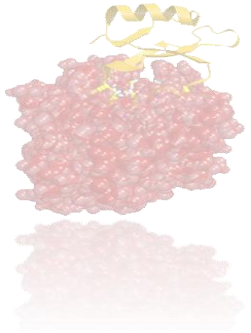
- Rodrigues JPGLM and Bonvin AMJJ (2014). **Integrative computational modelling of protein interactions**. *FEBS J.*, **281**, 1988-2003 (2014).

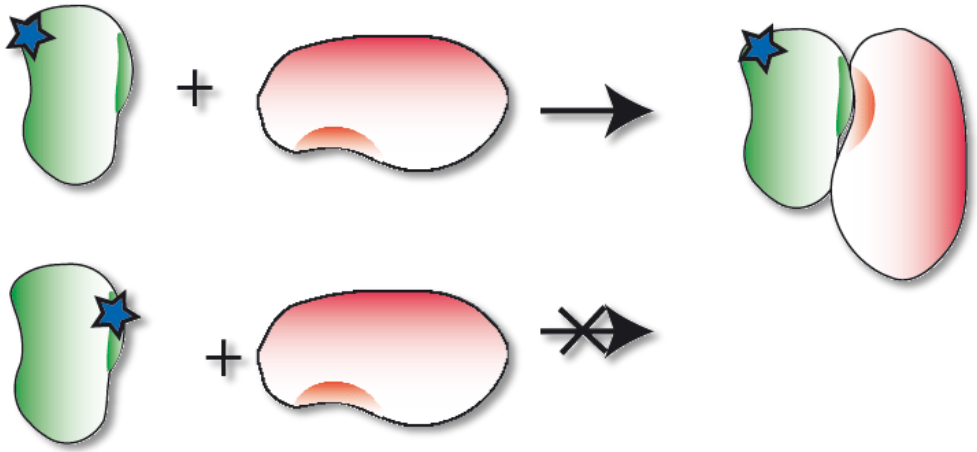


Overview

- Introduction
- **Information sources**
- General aspects of docking
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
- Assessing the interaction space
- Conclusions & perspectives

Experimental sources:
mutagenesis





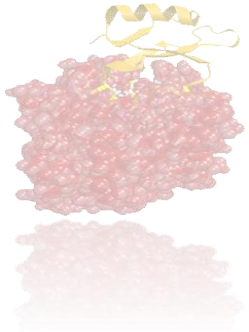
Advantages/disadvantages

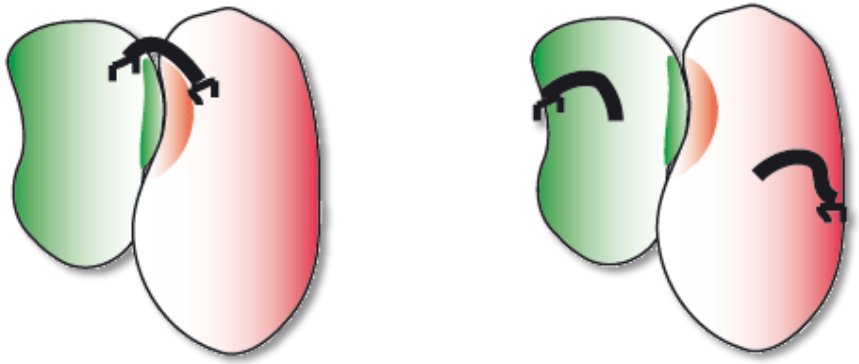
+ Residue level information
- Loss of native structure should be checked

Detection

- Binding assays
- Surface plasmon resonance
- Mass spectrometry
- Yeast two hybrid
- Phage display libraries, …

Experimental sources:
cross-linking and other chemical modifications





Advantages/disadvantages

+ Distance information between linker residues
- Cross-linking reaction problematic
- Detection difficult

Detection

- Mass spectrometry

Experimental sources:
H/D exchange

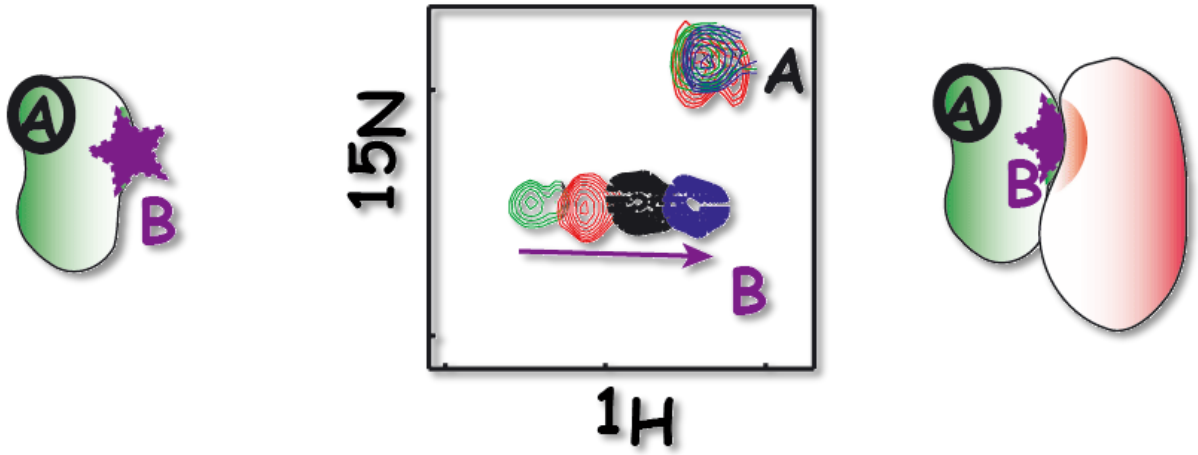
Advantages/disadvantages

+ Residue information
- Direct vs indirect effects
- Labeling needed for NMR

Detection

- Mass spectrometry
- NMR ^{15}N HSQC

Experimental sources:
NMR chemical shift perturbations



Advantages/disadvantages

+ Residue/atomic level

+ No need for assignment if combined with a.a. selective labeling

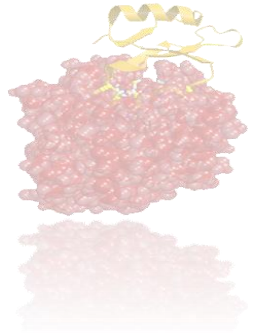
- Direct vs indirect effects

- Labeling needed

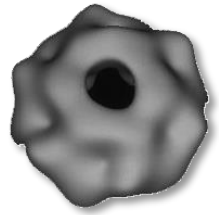
Detection

- NMR ^{15}N or ^{13}C HSQC

Other potential experimental sources

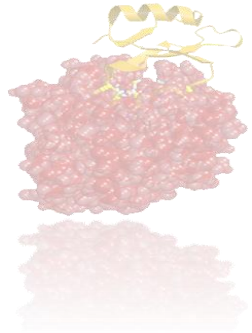


- **Paramagnetic probes** in combination with NMR
- **Cryo-electron microscopy** or tomography and **small angle X-ray scattering** (SAXS) ==> shape information



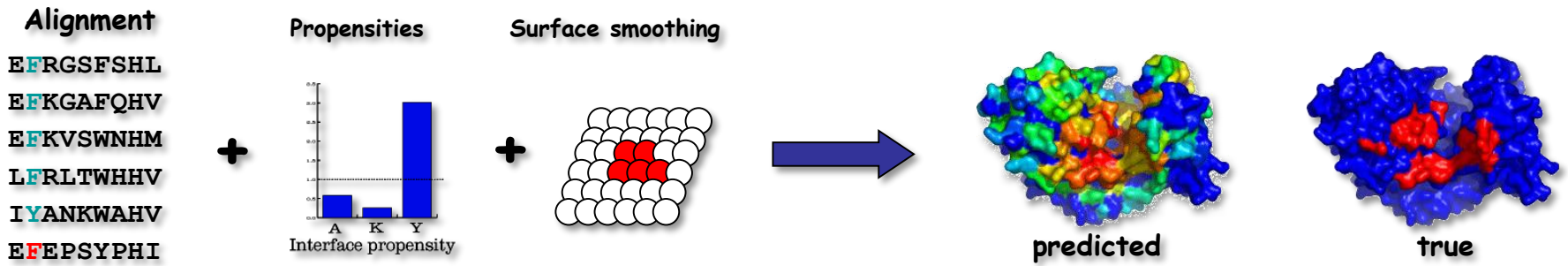
- Fluorescence quenching
- Fluorescence resonance energy transfer (FRET)
- Infrared spectroscopy combined with specific labeling
- …

Predicting interaction surfaces



- In the absence of any experimental information (other than the unbound 3D structures) **we can try to predict interfaces from sequence information?**
- **WHISCY:**

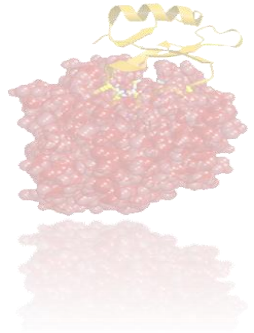
  WHat Information does Surface Conservation Yield?



http://www.nmr.chem.uu.nl/whiscy

De Vries, van Dijk Bonvin. *Proteins* 2006

Predicting interaction surfaces

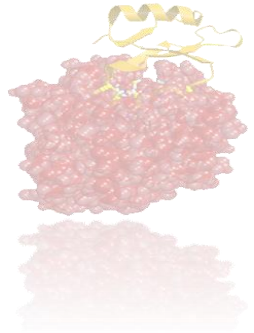


- Several other approaches have been described:
  - HSSP **(Sander & Schneider, 1993)**
  - Evolutionary trace **(Lichtarge *et al.*, 1996)**
  - Correlated mutations **(Pazos *et al.*, 1996)**
  - ConsSurf **(Armon *et al.*, 2001)**
  - Neural network **(Zhou & Shan, 2001) (Fariselli *et al.*, 2002)**
  - Rate4Site **(Pupko *et al.*, 2002)**
  - ProMate **(Neuvirth *et al.*, 2004)**
  - PPI-PRED **(Bradford & Westhead, 2005)**
  - PPISP **(Chen & Zhou, 2005)**
  - PINUP **(Liang et al., 2006)**
  - SPPIDER **(Kufareva et al, 2007)**
  - PIER **(Porolo & Meller, 2007)**
  - SVM method **(Dong et al., 2007)**
  - ... and many more since then

  - Our recent meta-server: **CPORT** **(de Vries & Bonvin, 2011)**

See review article (de Vries & Bonvin 2008)

Interface prediction servers

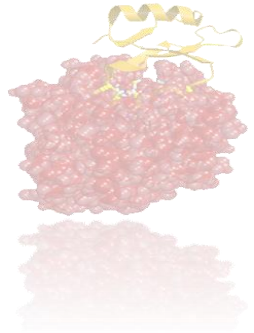


- PPISP (Zhou & Shan,2001; Chen & Zhou, 2005)
  http://pipe.scs.fsu.edu/ppisp.html
- ProMate (Neuvirth et al., 2004)
  http://bioportal.weizmann.ac.il/promate
- WHISCY (De Vries et al., 2005)
  http://www.nmr.chem.uu.nl/whiscy
- PINUP (Liang et al., 2006)
  http://sparks.informatics.iupui.edu/PINUP
- PIER (Kufareva et al., 2006)
  http://abagyan.scripps.edu/PIER
- SPPIDER (Porollo & Meller, 2007)
  http://sppider.cchmc.org

Consensus interface prediction (CPORT)

haddock.science.uu.nl/services/CPORT

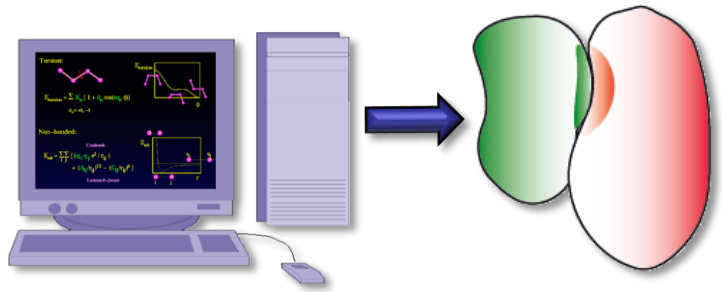
CPORT webserver

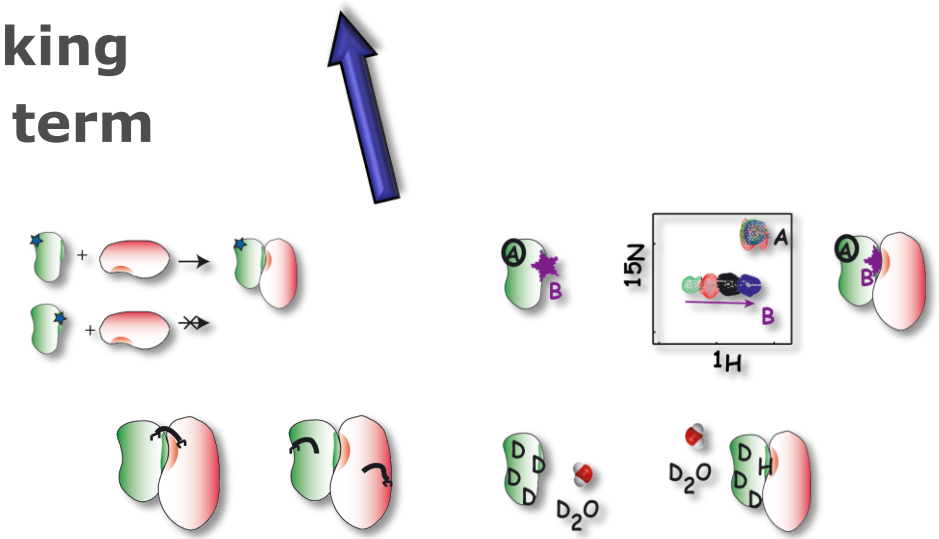


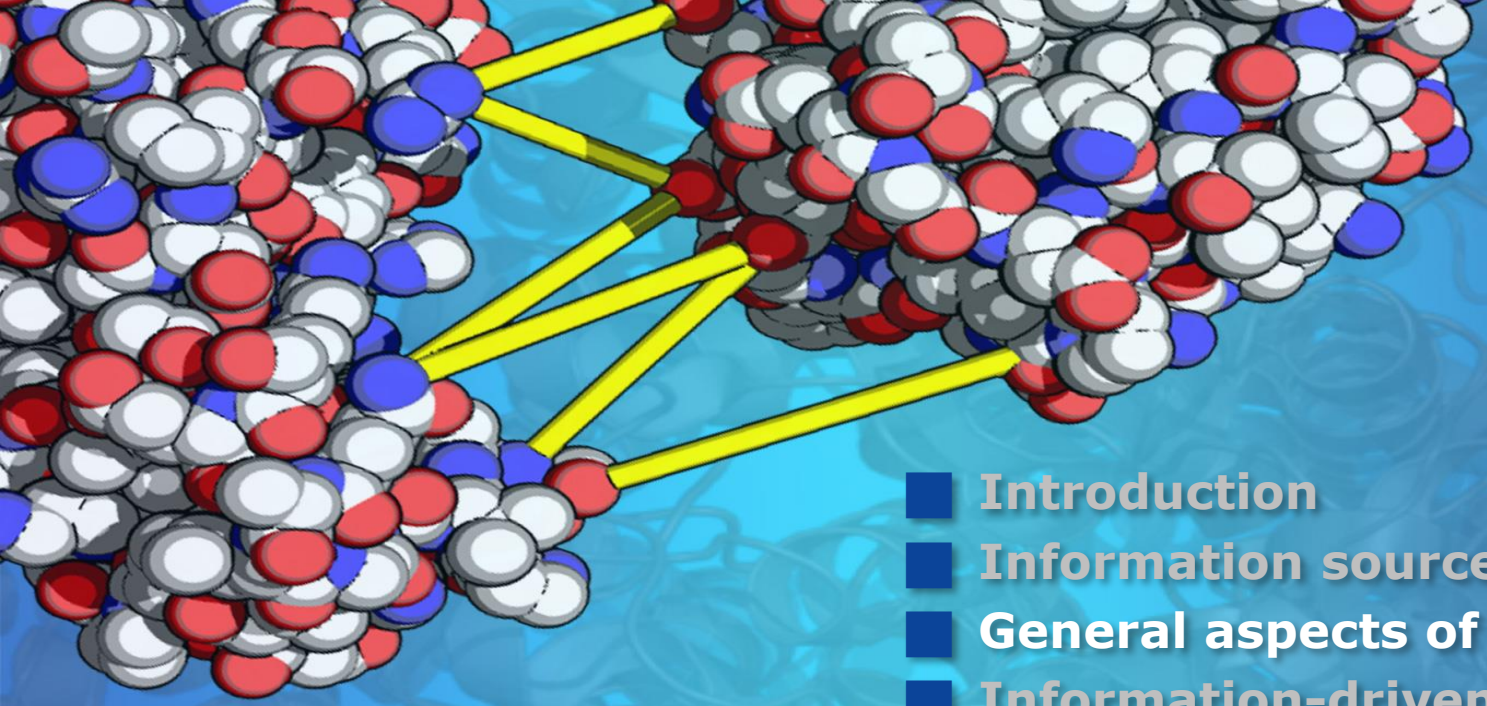
Combining experimental or predicted data with docking

- ***a posteriori*: data-filtered docking**
  - **Use standard docking approach**
  - **Filter/rescore solutions**

- ***a priori*: data-directed docking**
  - **Include data directly in the docking by adding an additional energy term or limiting the search space**



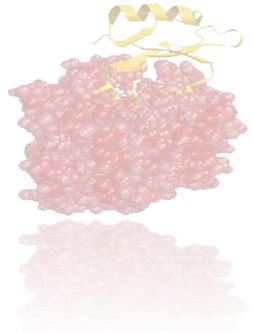




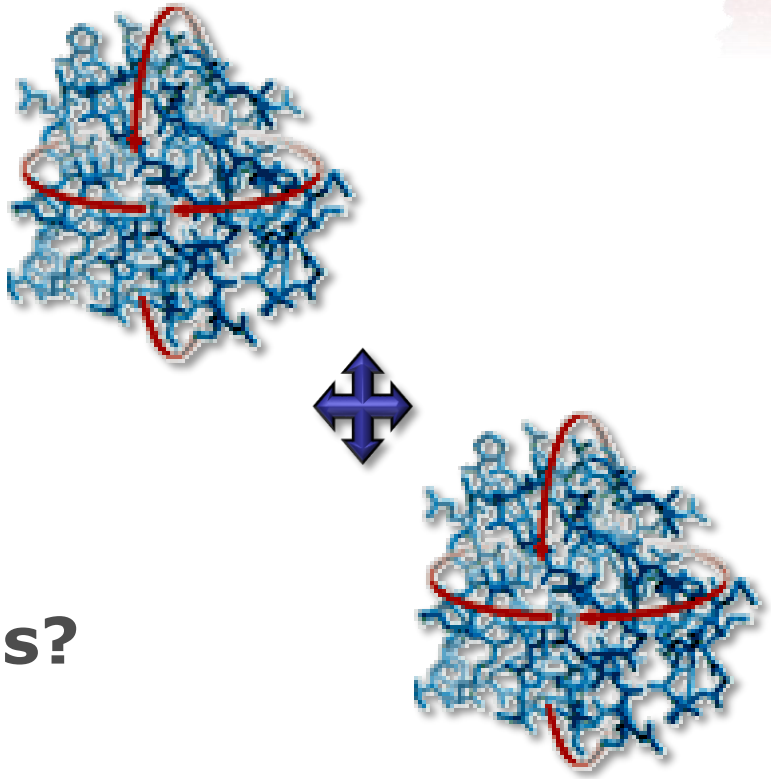
Overview

- Introduction
- Information sources
- **General aspects of docking**
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
- Assessing the interaction space
- Conclusions & perspectives

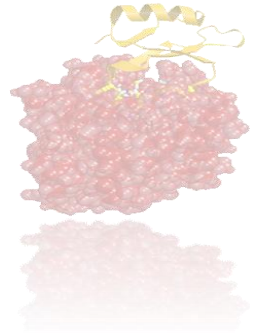
Docking



- **Choices to be made in docking:**
  - Representation of the system
  - Sampling method:
    - 3 rotations and 3 translations
    - Internal degrees of freedom?
  - Scoring
  - Flexibility, conformational changes?
  - Use experimental information?

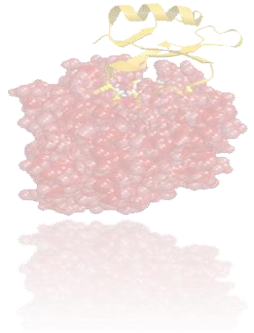


Systematic search



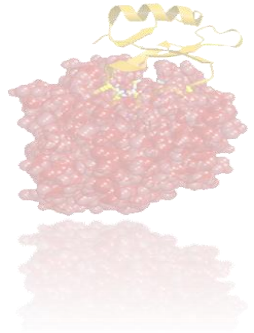
- Sample rotations (3) and translations (3)
- For each orientation calculate a score
- Can be very time consuming depending on scoring function
- Translational search often carried out in (2D or 3D) Fourier space by convolution of the grids
- Examples:
  - FFT methods: Z-DOCK, GRAMM, FTDOCK…
  - Direct search: Bigger (uses fast boolean operations)

“Energy-driven” search methods



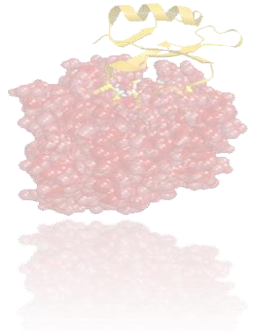
- Conformational search techniques aiming at minimizing some kind of energy function (e.g. VdW, electrostatic…):
  - Energy minimization
  - Molecular dynamics
  - Brownian dynamics
  - Monte-Carlo methods
  - Genetic algorithms
  - …
- Often combined with some simulated annealing scheme

Dealing with flexibility

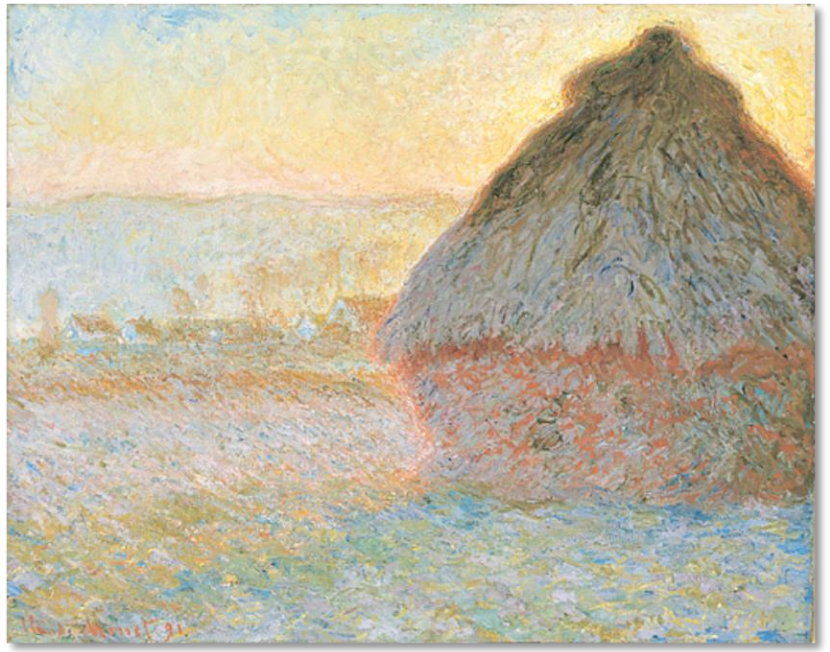


- Flexibility makes the docking problem harder!
  - Increased number of degrees of freedom
  - Scoring more difficult
- Difficult to predict a-priori conformational changes
- Current docking methodology can mainly deal with small conformational changes
- Treatment of flexibility depends on the chosen representation of the system and the search method

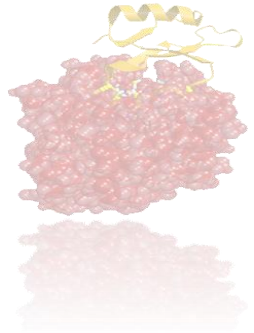
Scoring



- **The holy grail in docking!**
- **Depends on the representation of the system and treatment of flexibility**
- **Depends on the type of complexes**
  - **e.g. antibody-antigen might behave differently than enzyme-inhibitors complexes**

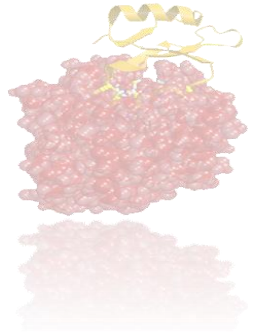


Scoring

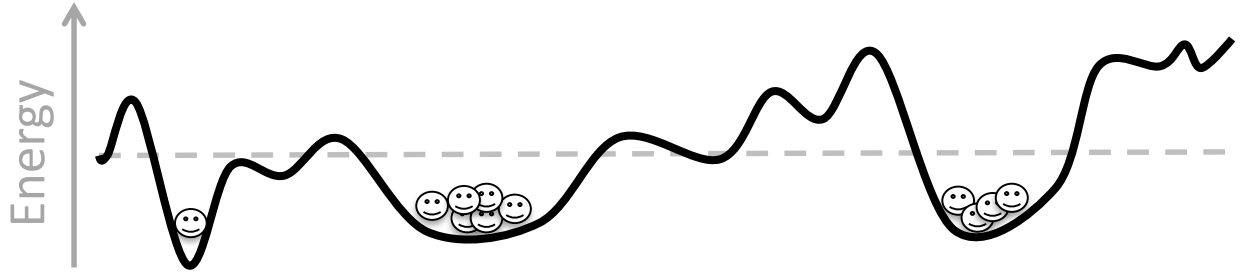


- Score is often a combination of various (empirical) terms such as
  - Intermolecular van der Waals energy
  - Intermolecular electrostatic energy
  - Hydrogen bonding
  - Buried surface area
  - Desolvation energy
  - Entropy loss
  - Amino-acid interface propensities
  - Statistical potentials such as pairwise residue contact matrices
  - ...
- Experimental filters sometimes applied *a posteriori* if data available (e.g. NMR chemical shift perturbations, mutagenesis,..)

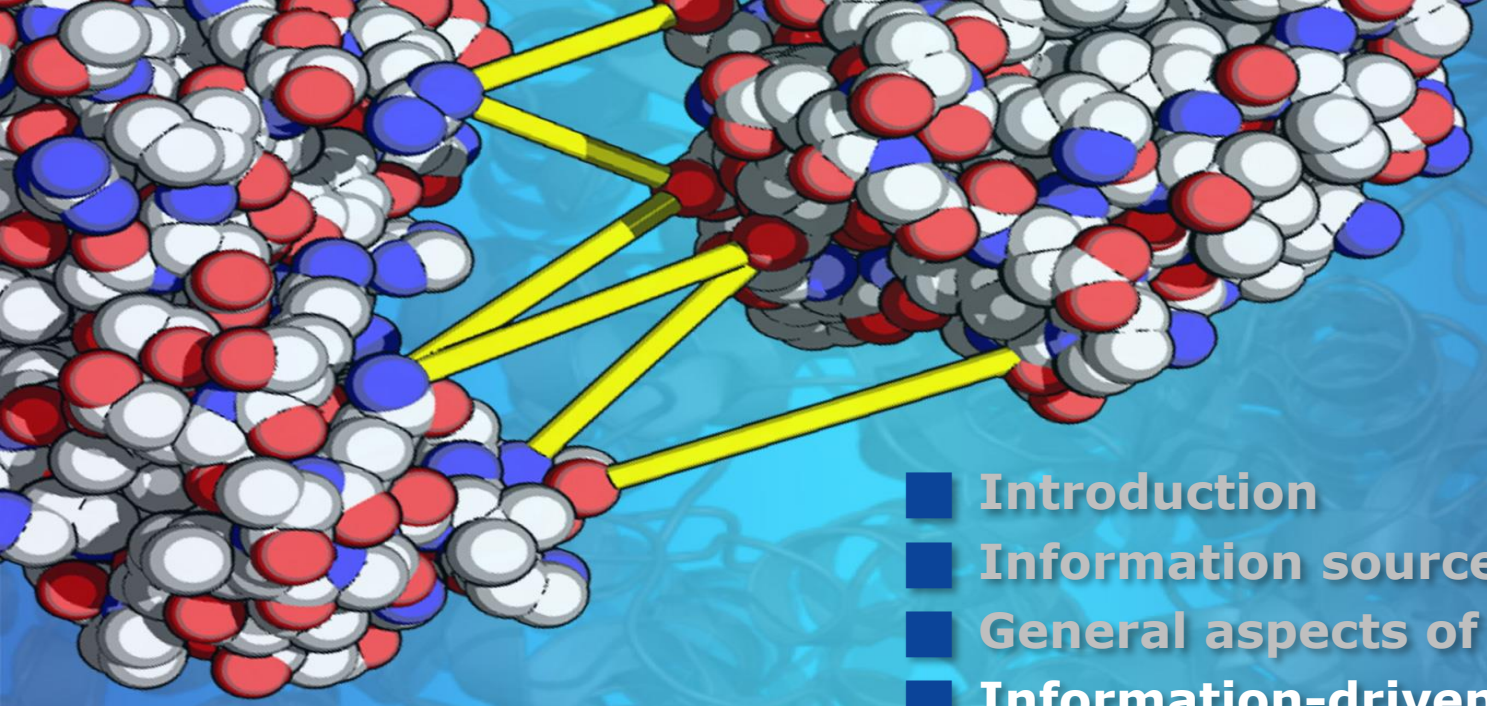
Clustering protein complexes



- Docking methods often produce thousands of models.
- Scoring functions do not perfectly describe the energy landscape.



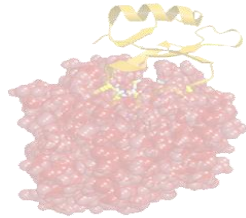
- Clustering groups similar structures together and allows **better analysis**.
- Similarity is defined by a specific measure (e.g. RMSD, interface RMSD, FCC)



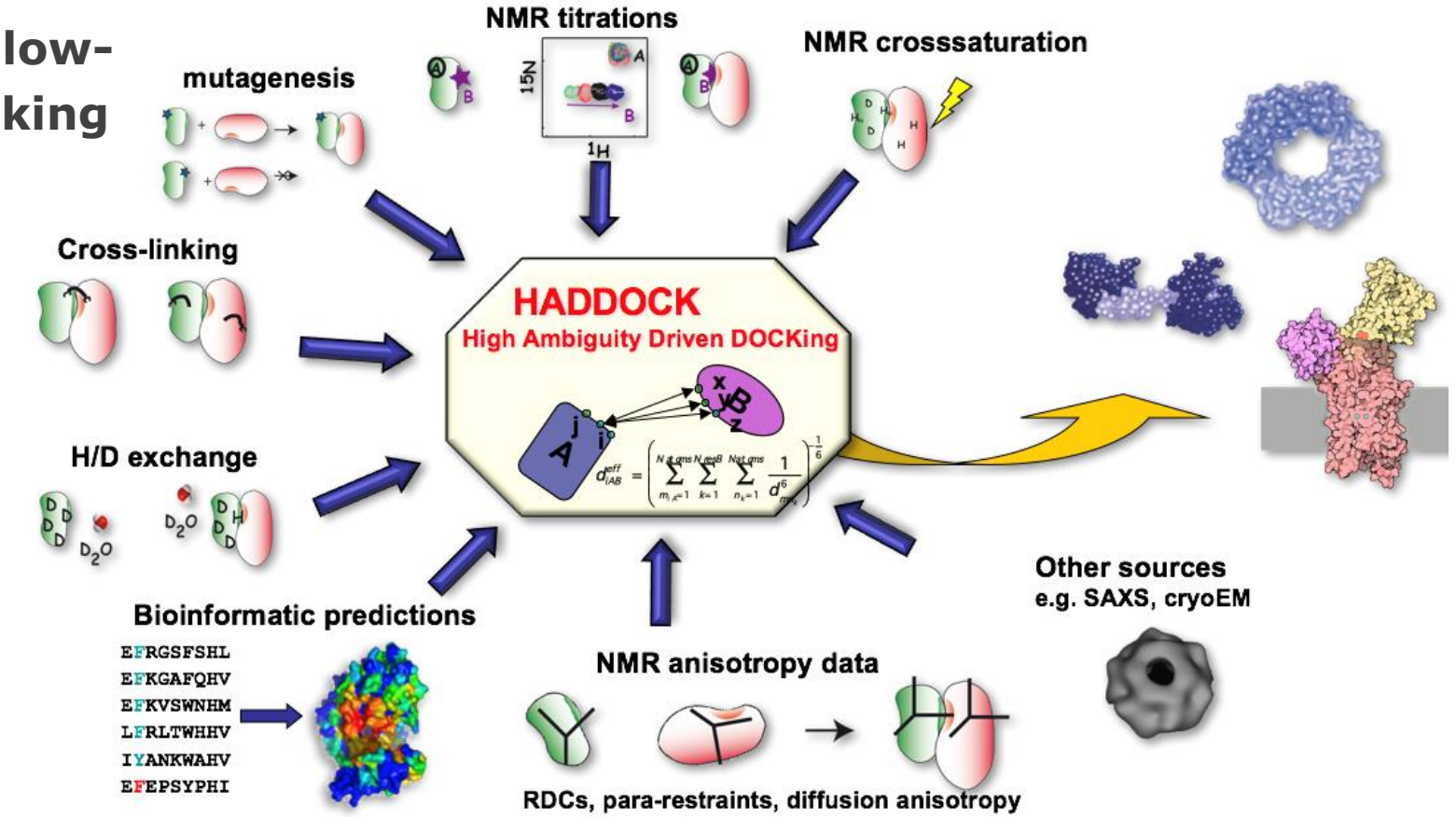
Overview

- Introduction
- Information sources
- General aspects of docking
- **Information-driven docking with HADDOCK**
- Incorporating biophysical data into docking
- Assessing the interaction space
- Conclusions & perspectives

HADDOCK:
An integrative modeling platform



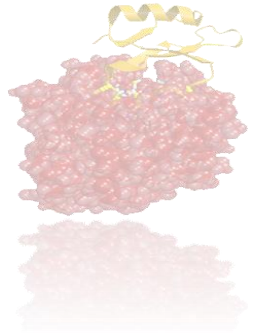
- Incorporates ambiguous and low-resolution data to aid the docking
- Capable of docking up to 20 molecules (new version)
- Symmetries can be leveraged
- Allows for flexibility at the interface
- Final flexible refinement in explicit solvent
- One of the best performing software in CAPRI



Dominguez, Boelens & Bonvin. *JACS* 125, 173 (2003).

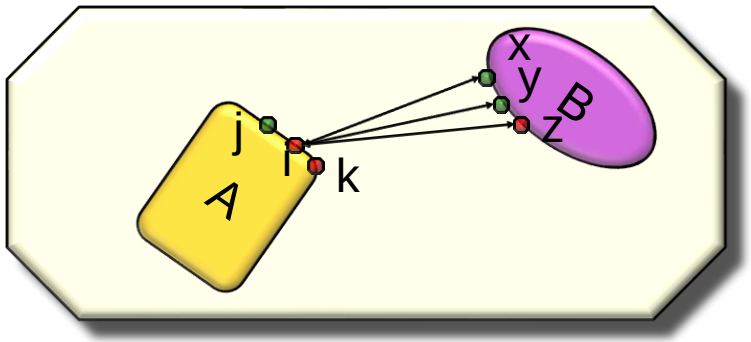
http://www.bonvinlab.org/software

Data-driven docking with HADDOCK



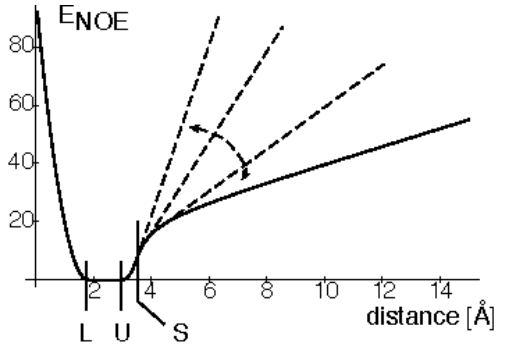
List of interface residues for protein A

List of interface residues for protein B



Effective distance d_{iAB}^{eff} calculated as

$$d_{iAB}^{eff} = \left(\sum_{m_{iA}=1}^{N\_atoms} \sum_{k=1}^{N\_resB} \sum_{n_k=1}^{Nat\_oms} \frac{1}{d_{mn_k}^{6}} \right)^{-\frac{1}{6}}$$



Ambiguous Interaction Restraint:

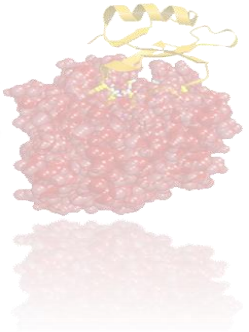
a residue must make contact with any residue from the other list

Different fraction of restraints (typically 50%) randomly deleted for each docking trial to deal with inaccuracies and errors in the information used

$$E_{NOE} = \begin{cases} (r - L)^2 & \text{if } r < L \\ 0 & \text{if } L < r < U \\ (U - r)^2 & \text{if } U < r < S \\ A(r - U)^{-1} + B(r - U) + C & \text{if } r > S \end{cases}$$

(Nilges & Brunger 1991)

Searching the interaction space in HADDOCK

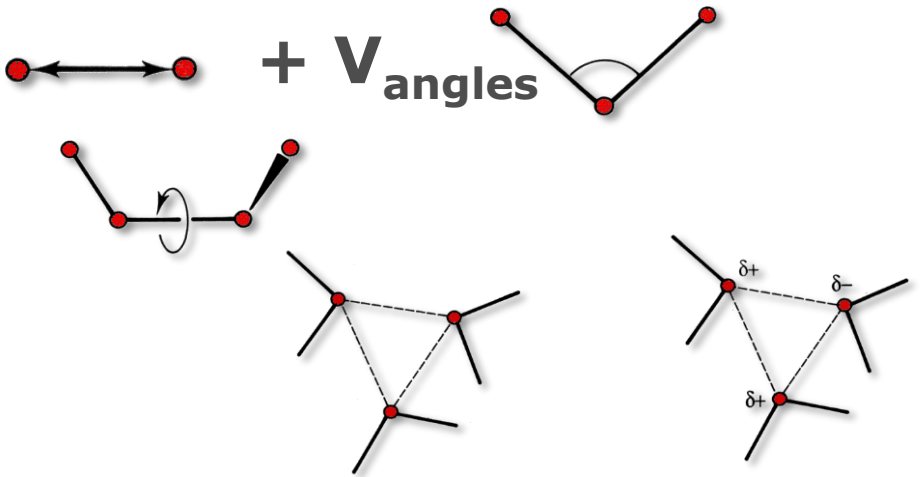


- Experimental and/or predicted information is combined with an empirical force field into an energy function whose minimum is searched for
- $V_{potential} = V_{bonds}$ + V_{angles}

  + $V_{torsion}$

  + $V_{non\text{-}bonded}$

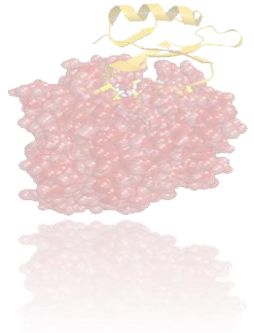
  + V_{exp}



Van der Waals electrostatic

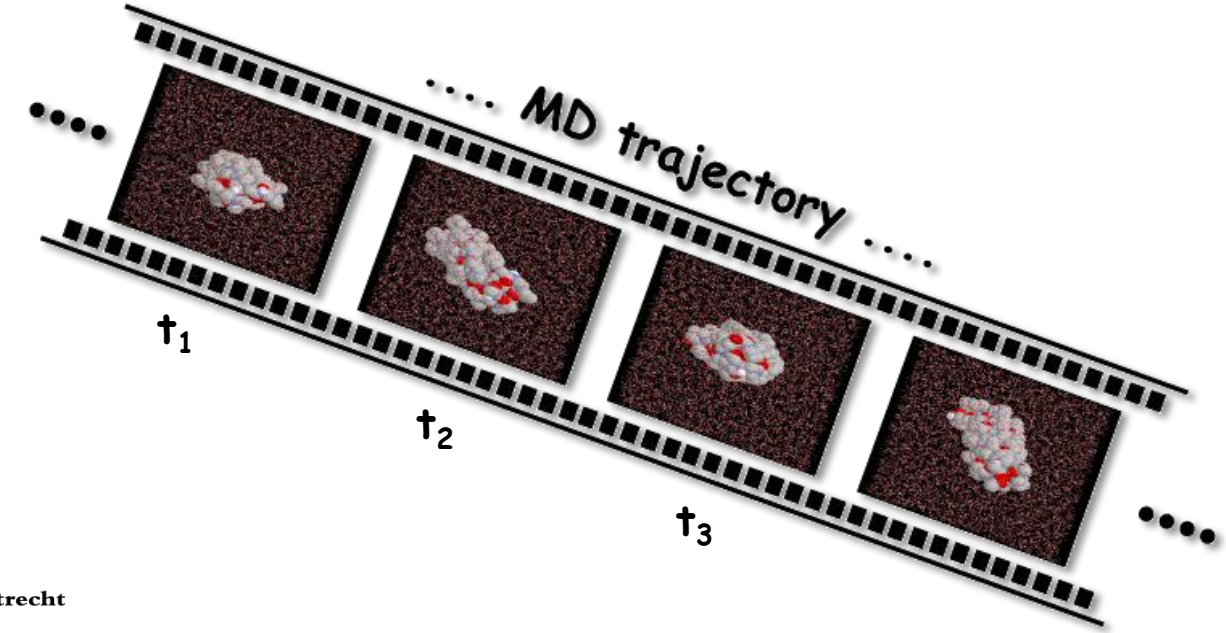
- Search is performed by a combination of gradient driven energy minimization and molecular dynamics simulations

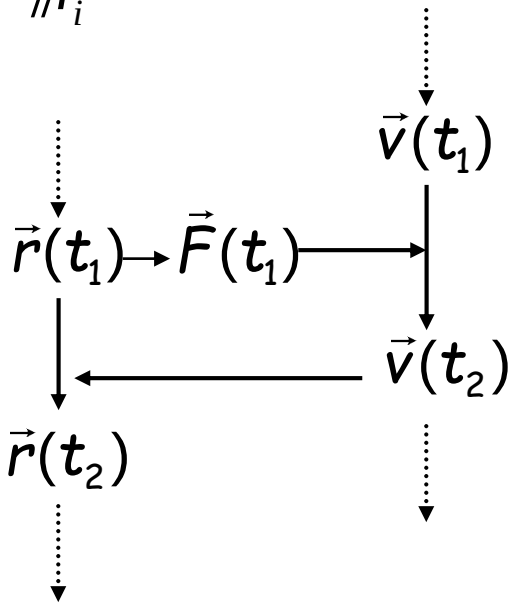
Classical mechanics



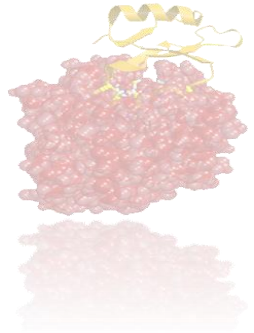
- **Molecular dynamics: generates successive configurations of the system by integrating Newton's second law**

$$\frac{d^2}{dt^2}\vec{r}_i = \frac{\vec{F}_i}{m_i} \qquad \text{with} \qquad \vec{F}_i = -\frac{\partial V}{\partial \vec{r}_i}$$





HADDOCK docking protocol

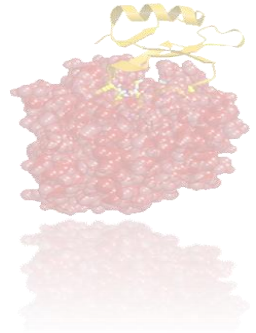


it0 it1 itw

Succession of energy minimization and molecular dynamics protocols

reminiscent of NMR structure calculations

HADDOCK docking protocol



it0

Rigid-body Energy Minimization

Rigid-body protocol allows generation of several thousand of models in a short period of time.

Simultaneous docking of max. 6 molecules, resembling *in vivo* complex assembly (vs. sequential docking)

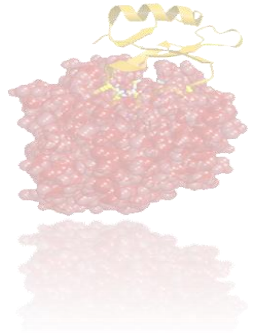
Typically, 10.000 conformations are sampled but only the best 1.000 are written to disk.

Rotational and translational optimization of the interacting partners, guided by the data-driven energy function.

Rigid-body energy minimization guided by restraints for fast sampling

in the absence of data, define restraints between centers of mass

HADDOCK docking protocol



it1

Semi-flexible simulated annealing

3-step process that increasingly allows more flexibility at the interface: rigid-body, side-chain, backbone + side-chain.

Torsion angle dynamics allows for faster integration time steps, while sampling relevant motions.

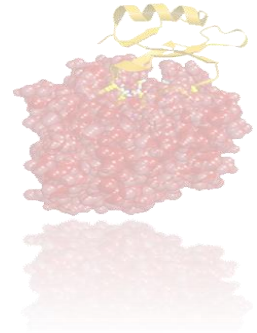
Flexibility reproduces conformation changes up to 2Å, typical of small induced fit.

Typically, the 200 best models of it0 undergo refinement.

Flexible simulated annealing in torsion angle space at the interface region

thorough optimization reproduces small conformational changes

HADDOCK docking protocol



Refinement in explicit solvent

Short molecular dynamics simulation in explicit solvent to refine residue-residue contacts, mainly electrostatics, at the interface.

Position restraints on backbone heavy atoms ensure conformation remains largely the same.

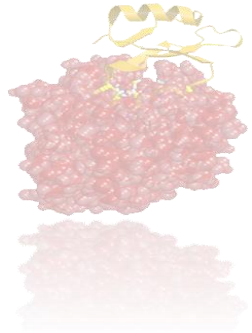
Explicit solvent models include TIP3P water and DMSO (membrane mimic).

Typically, all models of it1 are refined, i.e. there is no selection between it1 and itw.

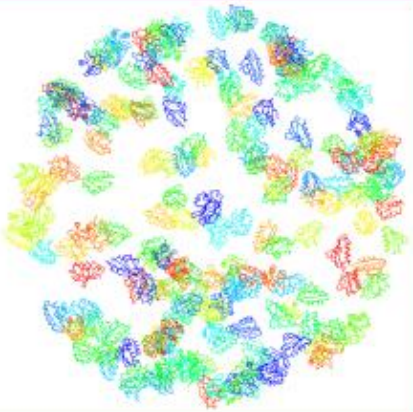
Refinement in explicit solvent to optimize the contacts at the interface

can be used in isolation to refine and score existing models

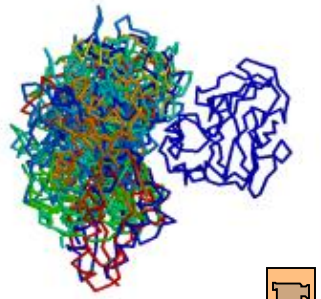
HADDOCK docking protocol



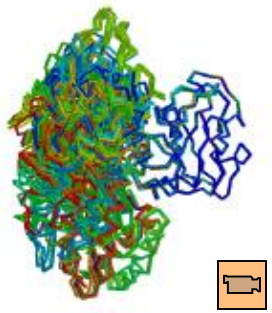
Separate proteins and apply random rotations



Rigid body energy minimization:
- first only rotations
- then rotations + translations

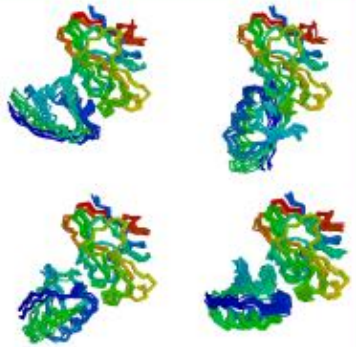


Semi-flexible simulated annealing in torsion angle space;
1) Rigid body search
2) SA with flexible side-chains at interface
3) SA with flexible side-chains and backbone at interface

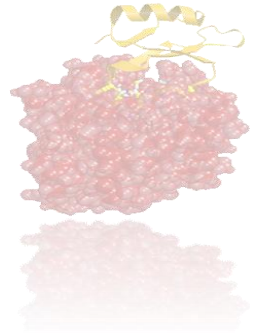


Final refinement in explicit water

Clustering and analysis
Sorting of clusters according to the HADDOCK score

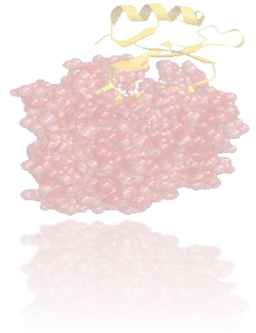


HADDOCK & Flexibility

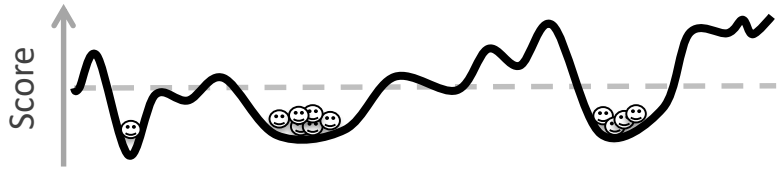


- Several levels of flexibility:
- **Implicit:**
  - docking from ensembles of structures
  - Scaling down of intermolecular interactions
- **Explicit:**
  - semi-flexible refinement stage with both side-chain and backbone flexibility during in torsion angle dynamics
  - Final refinement in explicit solvent

Energetics & Scoring



- **OPLS non-bonded parameters** (Jorgensen, *JACS* 110, 1657 (1988))
- **8.5Å non-bonded cutoff, switching function, $\sum$=10**
- **Clustering of solutions**

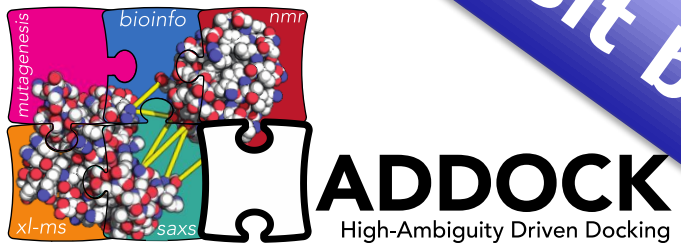


- **Ranking based on cluster-based HADDOCK score:**

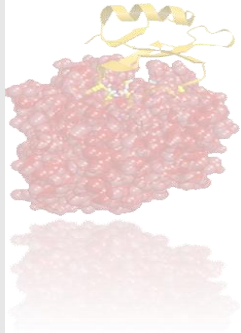
Rigid:	$Score = 0.01\,E_{air} + 0.01\,E_{vdW} + 1.0\,E_{elec} + 1.0\,E_{desolv} - 0.01\,BSA$
Flexible:	$Score = 0.1\,E_{air} + 1.0\,E_{vdW} + 1.0\,E_{elec} + 1.0\,E_{desolv} - 0.01\,BSA$
Water:	$Score = 0.1\,E_{air} + 1.0\,E_{vdW} + 0.2\,E_{elec} + 1.0\,E_{desolv}$

  - **E_{air}: ambiguous interaction restraint energy**
  - **E_{desolv}: desolvation energy using Atomic Solvation Parameters** (Fernandez-Recio et al *JMB* 335, 843 (2004))
  - **BSA: buried surface area**

Haddock web portal



Visit bonvinlab.org/software



> 14000 registered users

> 220000 served runs since June 2008

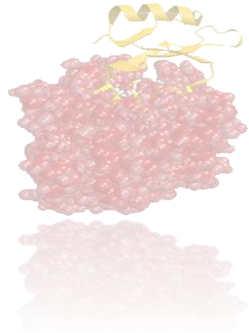
> 40% on the GRID

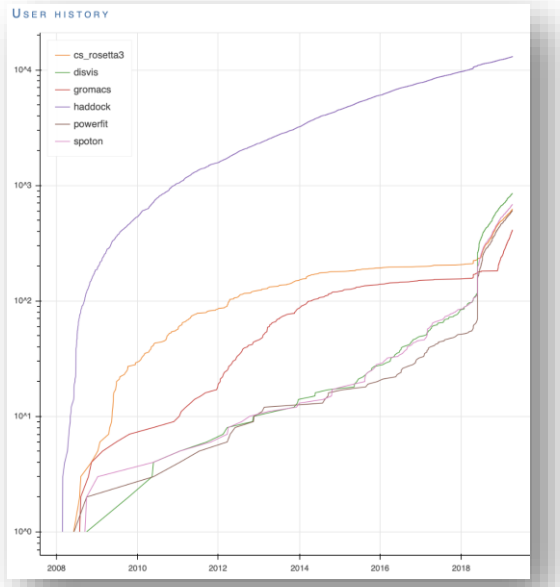
De Vries *et al.* Nature Prot. 2010

Van Zundert *et al.* J.Mol.Biol. 2016

https://wenmr.science.uu.nl/user_map

Haddock web portal

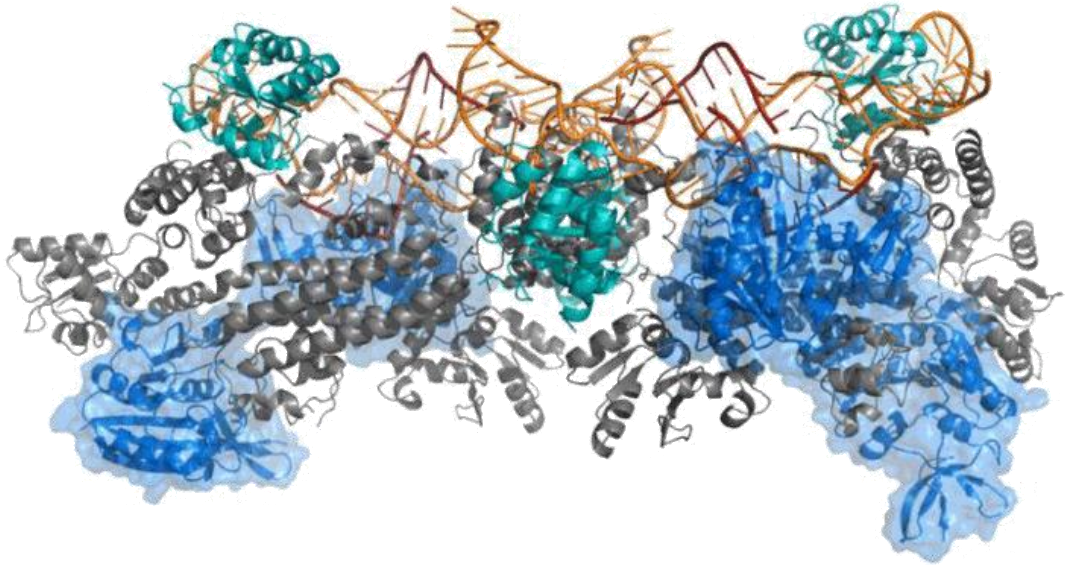




	Country	All_Users	HADDOCK	DISVIS	POWERFIT	SPOTON	CS_ROSETTA3	GROMACS
1	Total Users	14,447	13,853	1,102	804	926	777	599
2	EU Users	3,258	3,040	296	172	188	181	117
3	India	3,088	3,035	170	144	182	135	158
4	United States	2,269	2,180	178	106	138	109	67

HADDOCK development's highlights

- **Extension to up to 20 molecules**



Example of a complex protein structure calculated with the new HADDOCK framework: the box C/D enzyme for RNA methylation.

nature methods

Access provided by Utrecht University

Altmetric: 28 More detail »

Article

M3: an integrative framework for structure determination of molecular machines

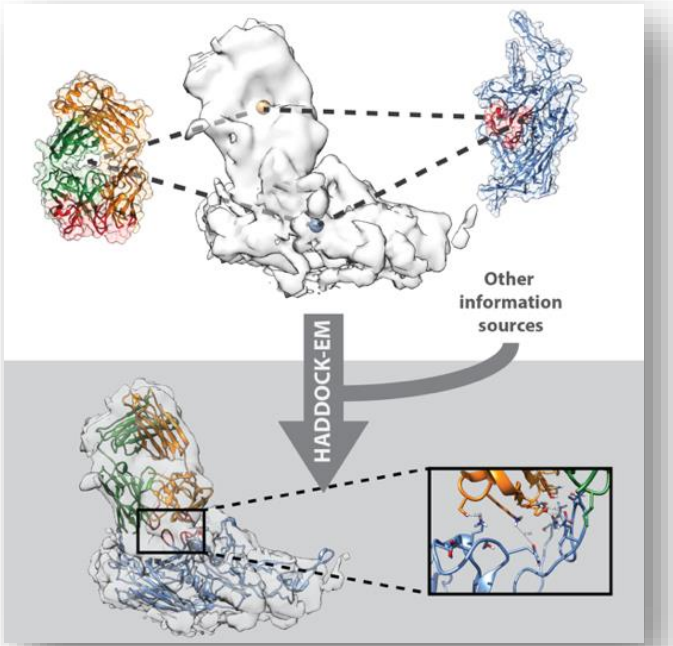
Ezgi Karaca, João P G L M Rodrigues, Andrea Graziadei, Alexandre M J J Bonvin & Teresa Carlomagno

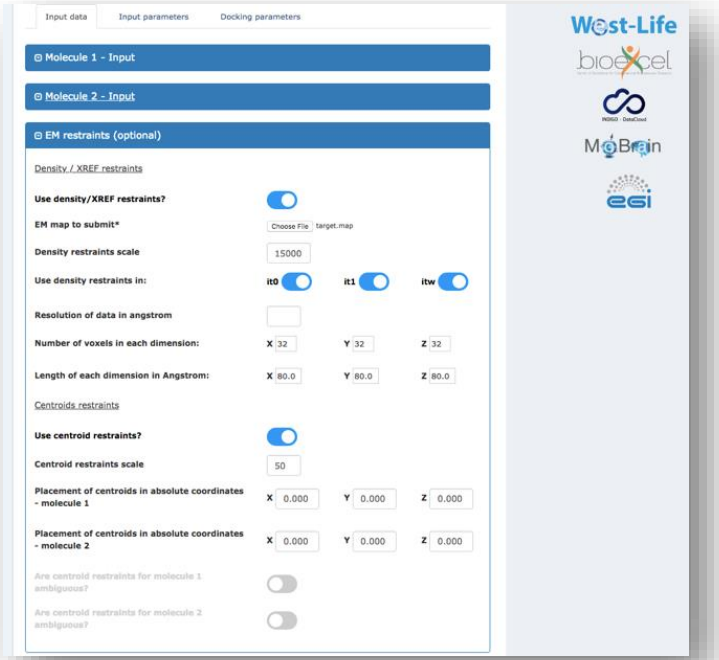
Nature Methods **14**, 897–902 (2017)
doi:10.1038/nmeth.4392

Received: 21 April 2017
Accepted: 05 July 2017

HADDOCK development's highlights

- **Complete rewrite of the portal (v2.4 to be released soon)**
- **Provides support for cryo-EM data, coarse-graining, …**





https://haddock.science.uu.nl/services/HADDOCK2.4

BioExcel Centre of Excellence

Driving and Supporting Computational Biomolecular Research in Europe

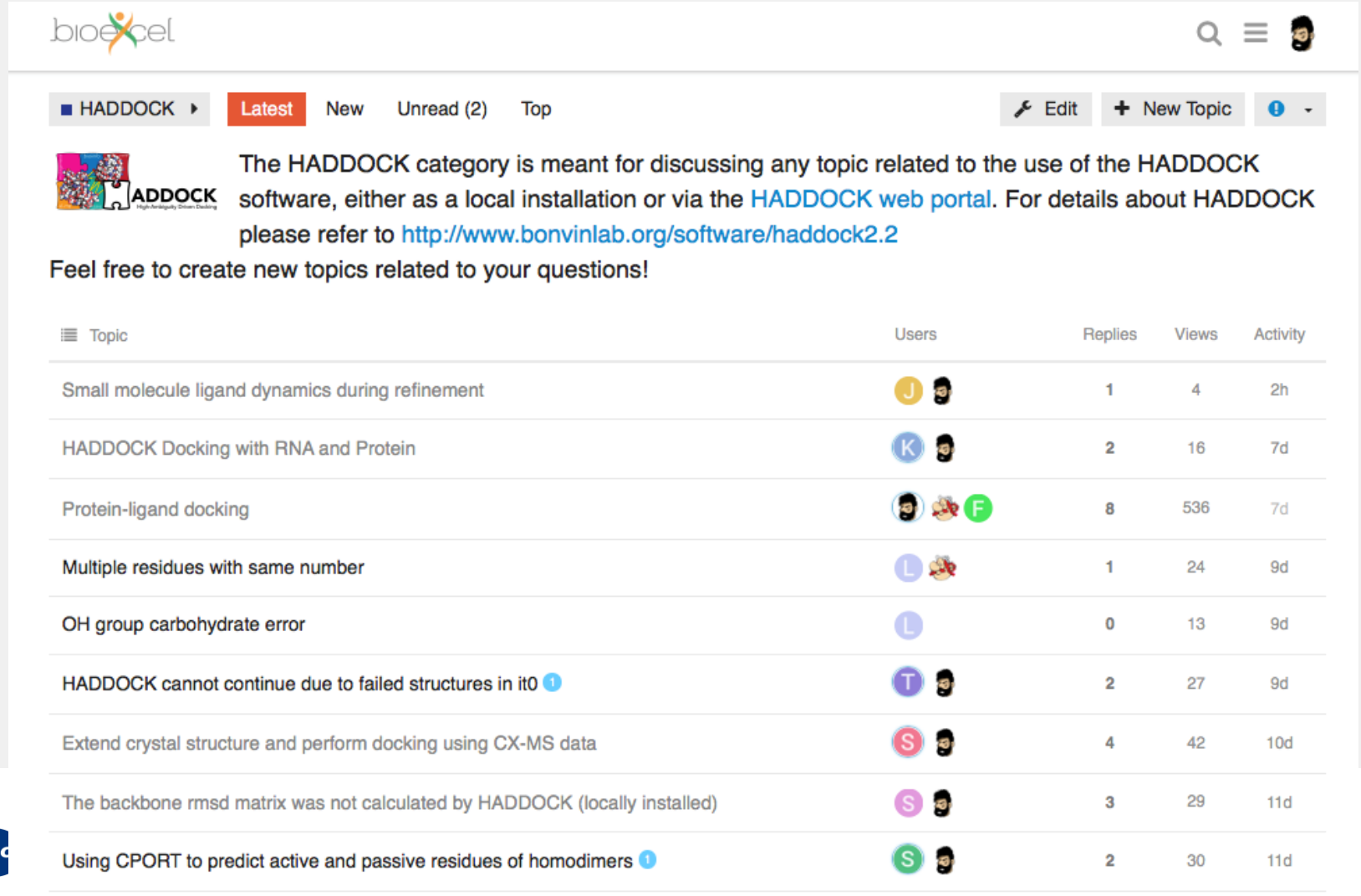
Partners

Funding

Horizon 2020
European Union Funding
for Research & Innovation

HADDOCK forum in BioExcel

ask.bioexcel.eu



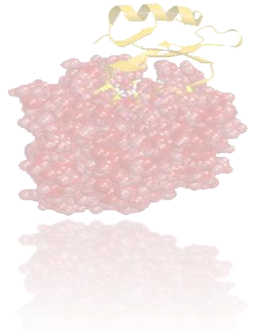
Partners

Funding

Horizon 2020
European Union Funding
for Research & Innovation

Bonvin Lab

Computational Structural Biology
@Utrecht University



DisVis

Restraints visualization

Prodigy

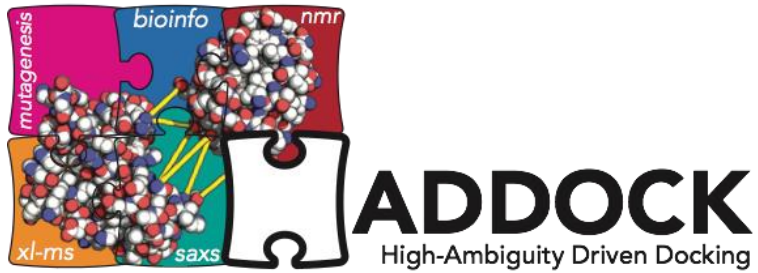
affinity prediction

CS-Rosetta

Chemical shift-based structure prediction

CPORT

Interface predicton



PowerFit

cryo EM map fitting

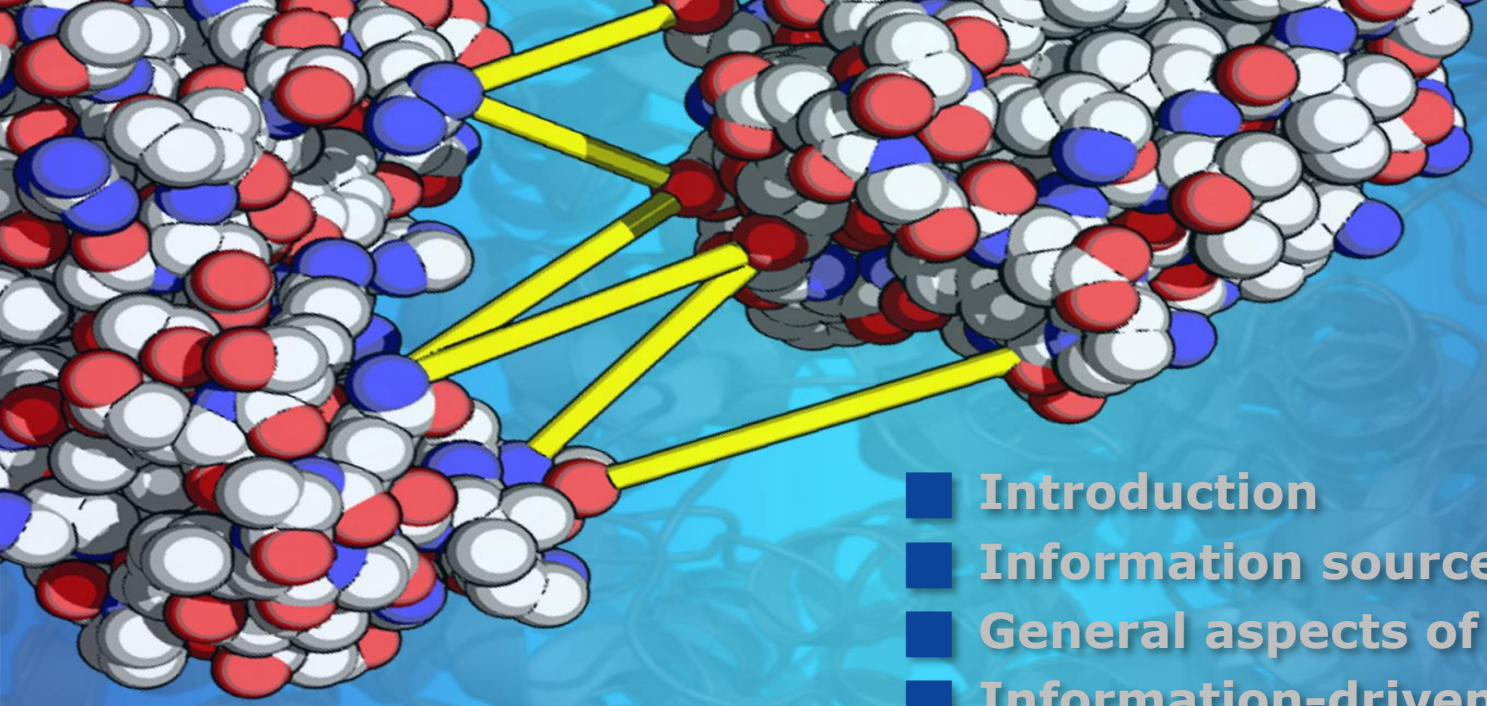
3D-DART

DNA structure modelling

SpotON

HotSpot predicton

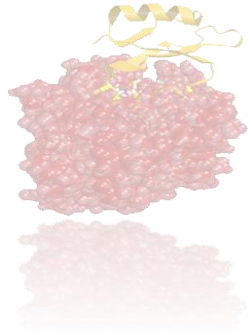
haddock.science.uu.nl



Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- **Incorporating biophysical data into docking**
- Assessing the interaction space
- Conclusions & perspectives

Iron Piracy:
NMR-based modelling of the FusA-ferredoxin complex



nature COMMUNICATIONS

ARTICLE

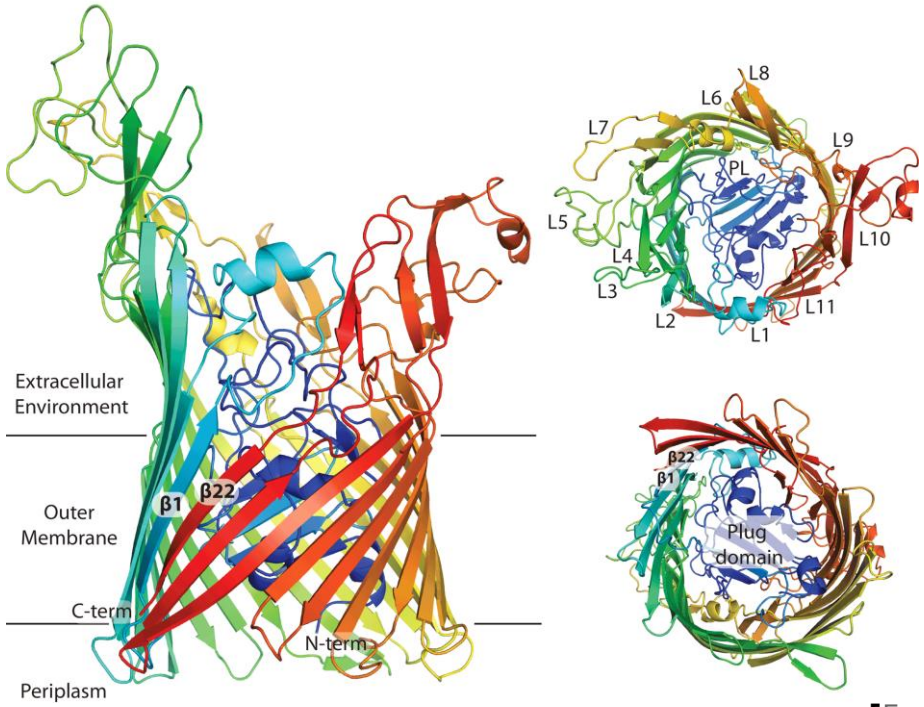
Received 21 Jan 2016 | Accepted 21 Sep 2016 | Published 31 Oct 2016

DOI: 10.1038/ncomms13308 OPEN

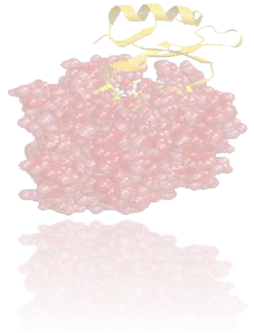
Structure of the bacterial plant-ferredoxin receptor FusA

Rhys Grinter[1,2,3], Inokentijs Josts[1], Khedidja Mosbahi[1], Aleksander W. Roszak[4], Richard J. Cogdell[4], Alexandre M.J.J. Bonvin[5], Joel J. Milner[6], Sharon M. Kelly[4], Olwyn Byron[6], Brian O. Smith[4] & Daniel Walker[1]

- **Iron import machinery in gram-negative bacteria**
- **First complete crystal structure of such a receptor**

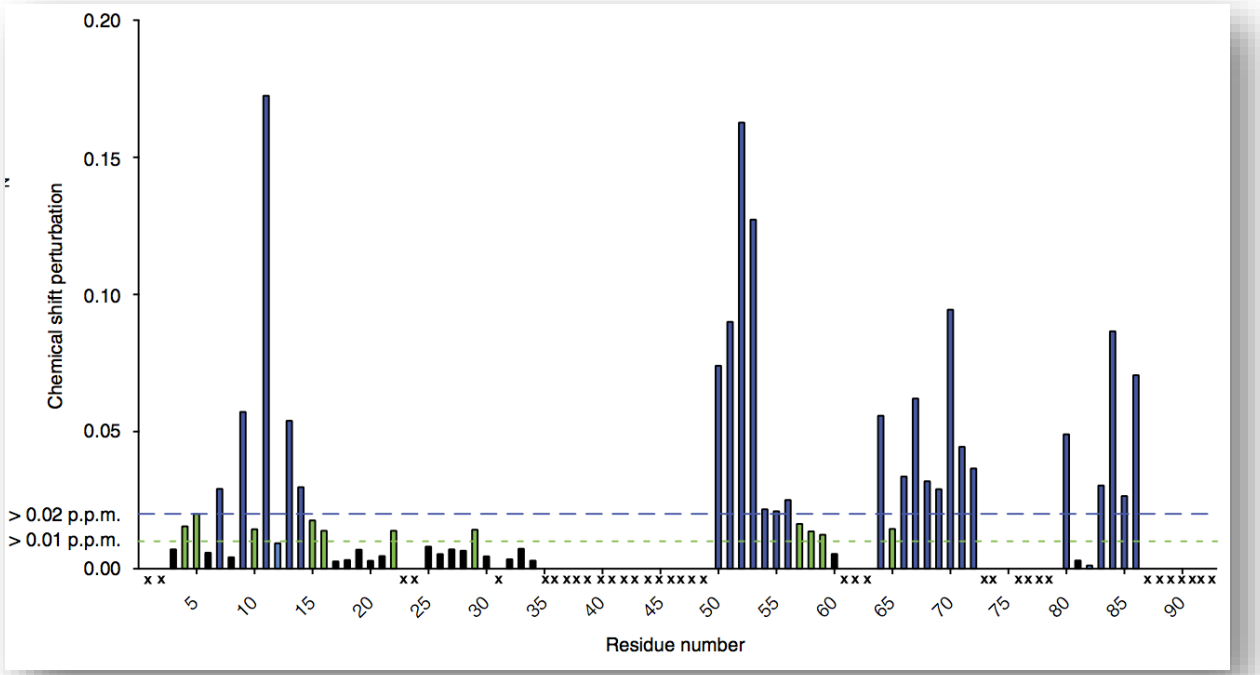


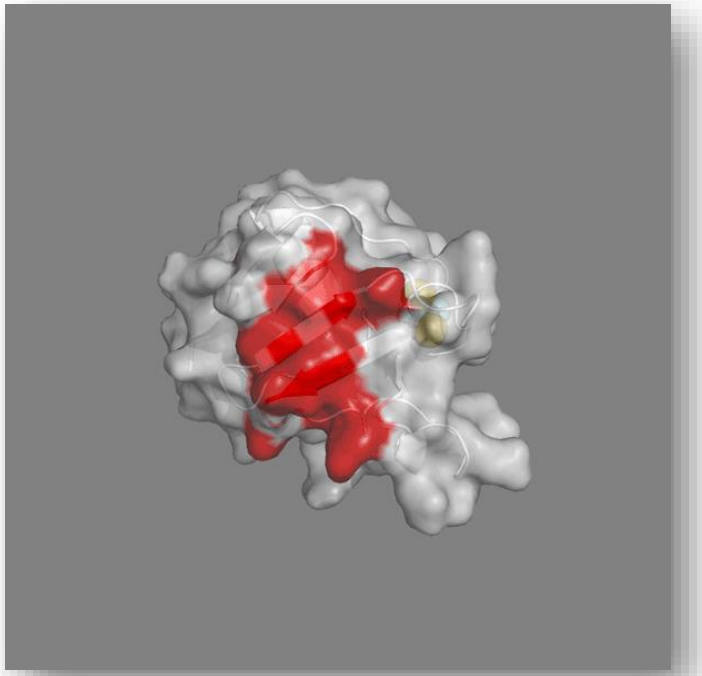
Docking strategy



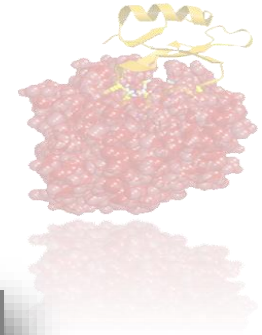
- **NMR chemical shift perturbation experiments define the binding site on ferredoxin** (which carries an iron-sulfur cluster)

  → **active residues in HADDOCK**





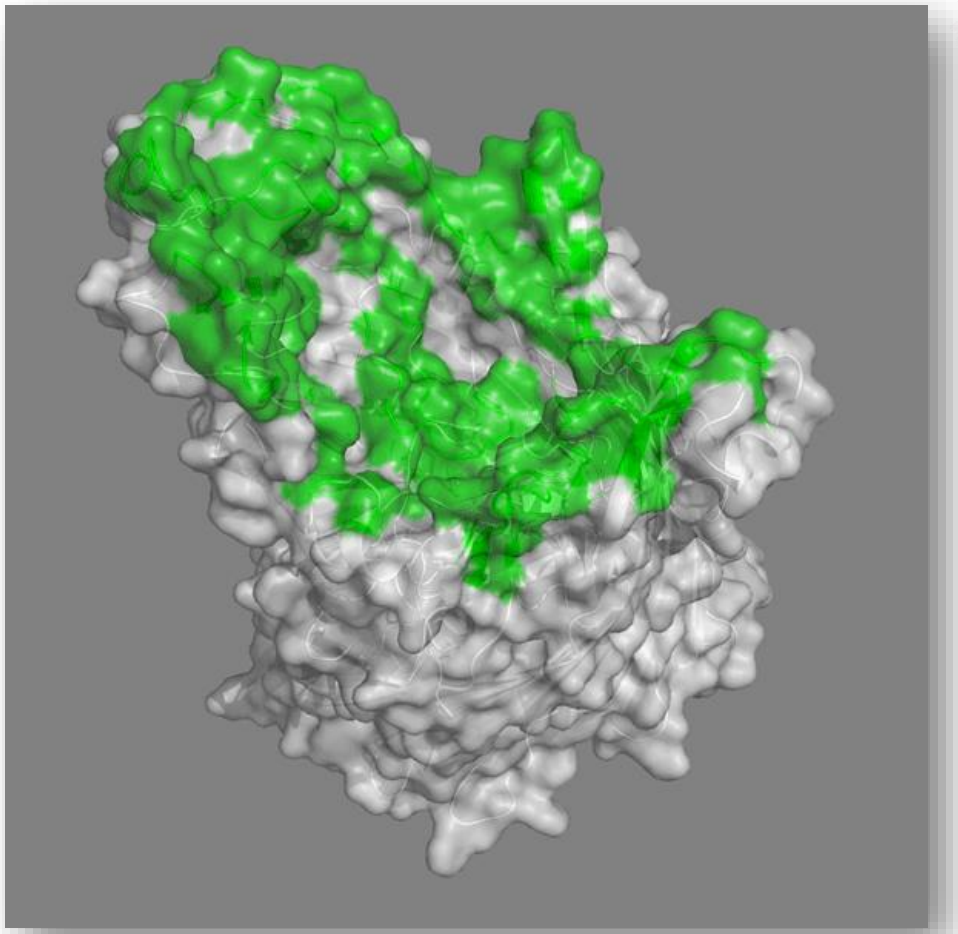
Docking strategy



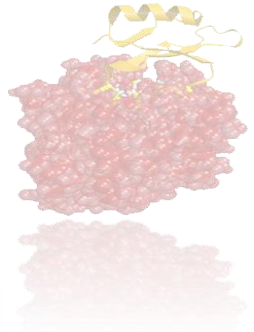
- **No info for FusA (expect that the binding occurs in the extracellular part)**

  → **extra cellular loops defined as passive** (which does not generate an energetic penalty if not contacted)

  → Definition of passive refined in a second docking run



Model of the FusA-ferredoxin complex

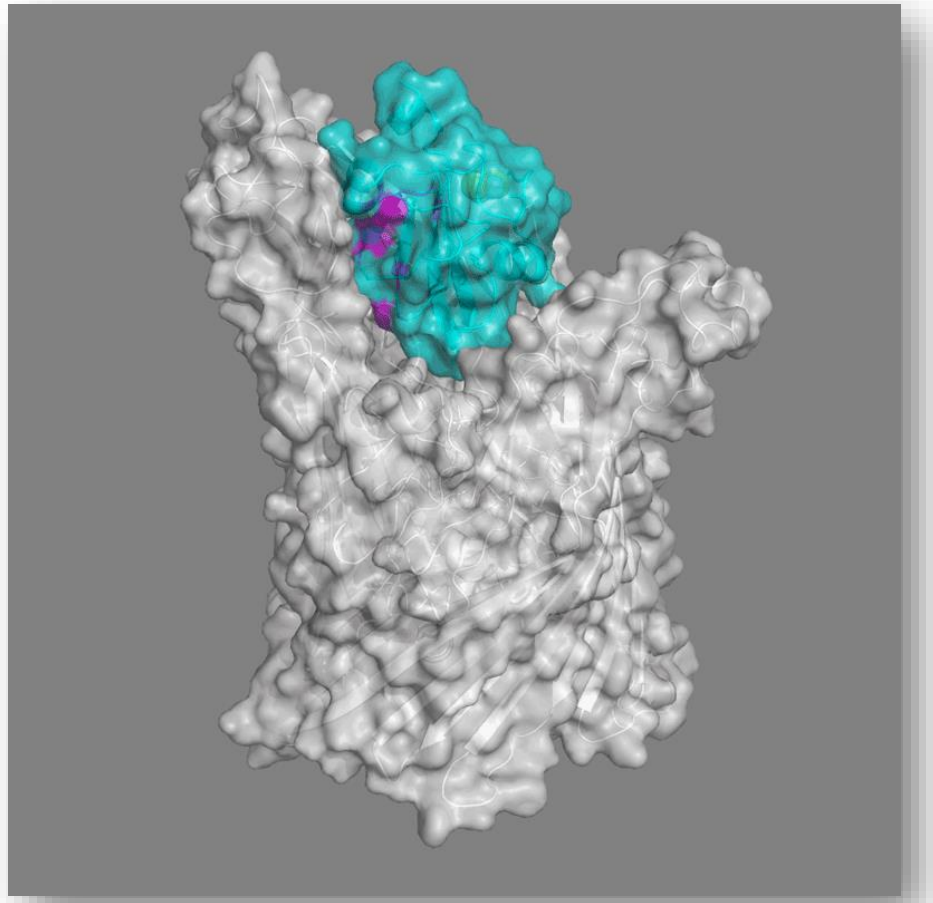


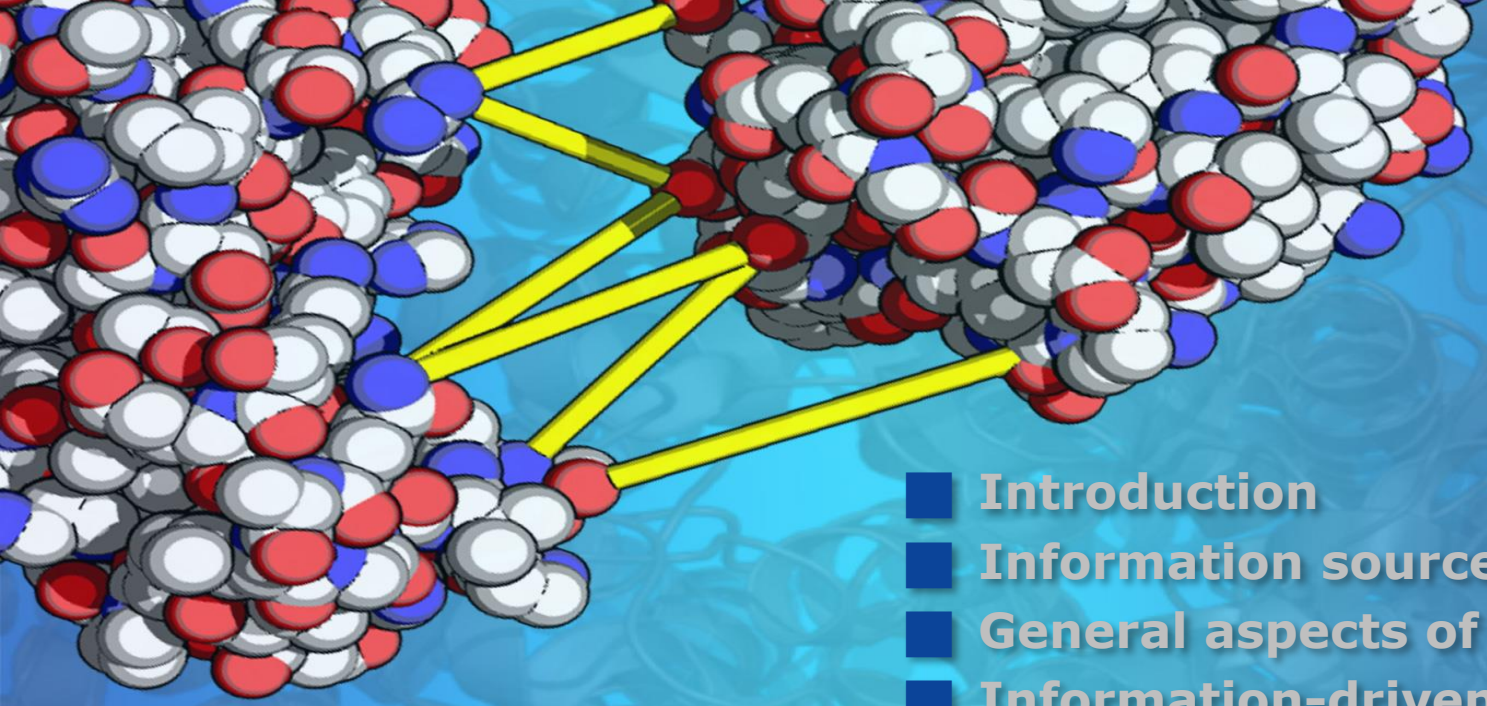
Cluster 1

HADDOCK score	-137.8 +/- 2.1
Cluster size	151
RMSD from the overall lowest-energy structure	5.8 +/- 0.1
Van der Waals energy	-72.5 +/- 10.5
Electrostatic energy	-476.2 +/- 66.5
Desolvation energy	28.9 +/- 10.0
Restraints violation energy	11.2 +/- 9.62
Buried Surface Area	2524.8 +/- 175.9
Z-Score	-1.3

Cluster 4

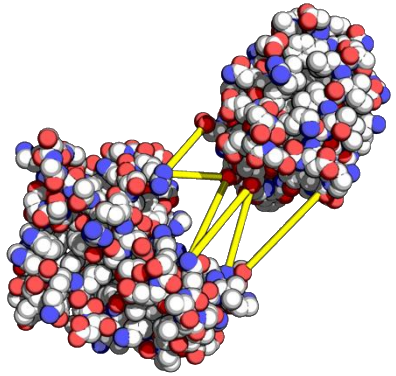
HADDOCK score	-130.8 +/- 20.3
Cluster size	7
RMSD from the overall lowest-energy structure	1.4 +/- 0.8
Van der Waals energy	-70.4 +/- 17.9
Electrostatic energy	-494.9 +/- 39.9
Desolvation energy	33.9 +/- 14.7
Restraints violation energy	47.2 +/- 29.71
Buried Surface Area	2728.9 +/- 345.5
Z-Score	-1.0

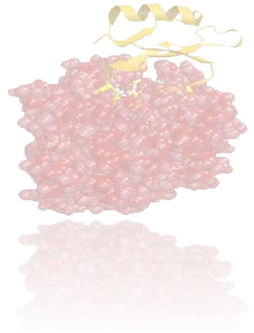




Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- **Incorporating biophysical data into docking**
  - Chemical shift perturbation data
  - **MS data as filters in docking**
- Assessing the interaction space
- Conclusions & perspectives

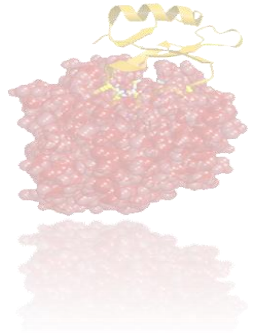




MS-based modelling of a bacterial circadian clock machinery

Adrien Melquiond

Insight into cyanobacterial circadian timing: the KaiB-KaiC interaction



Circadian clock controlled by the Kai system consisting of three proteins: KaiA, KaiB and KaiC

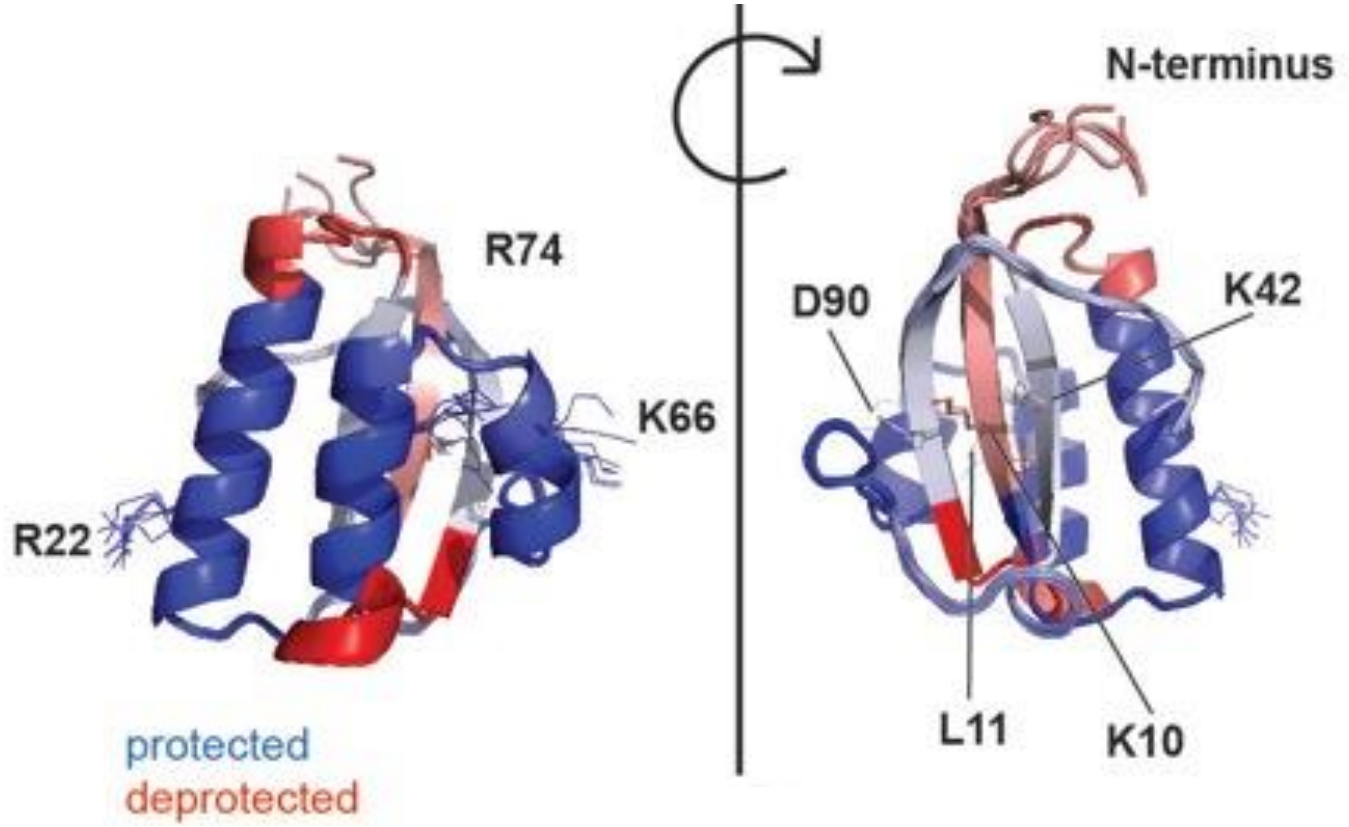
Interactions define the phosphorylation status of KaiC and control the phase of the cycle

Information from MS:

- **From native MS: Stochiometry of the KaiB-KaiC complex (6:1)**
- **From HD exchange: Binding interface and allosteric effects upon binding**

Snijder et al. PNAS 111, 1379 (2014)

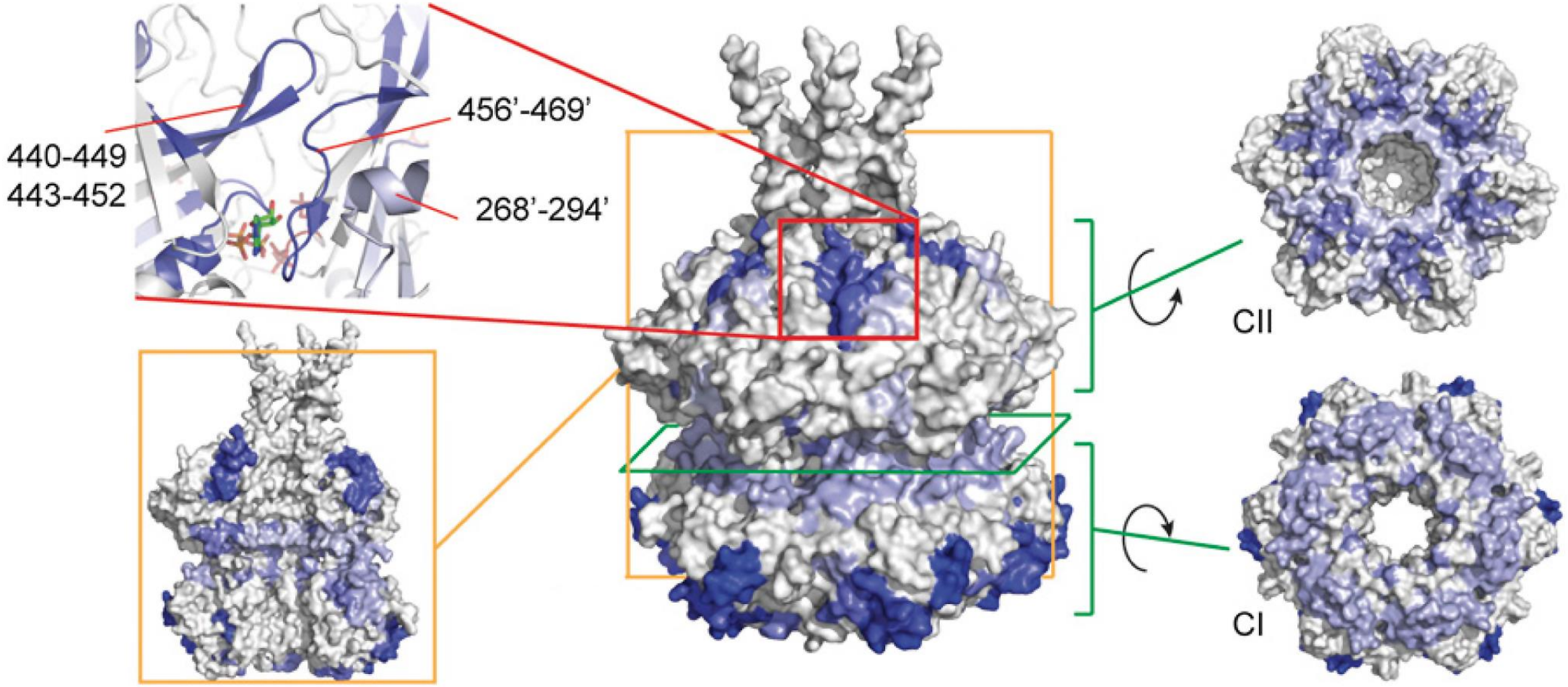
The KaiB-KaiC interaction: HDX



- **HDX-MS data reveal one protected face on KaiB**
- **Mutagenesis data show that R22, K67 and R74 abolish or alter the circadian rhythm when mutated**

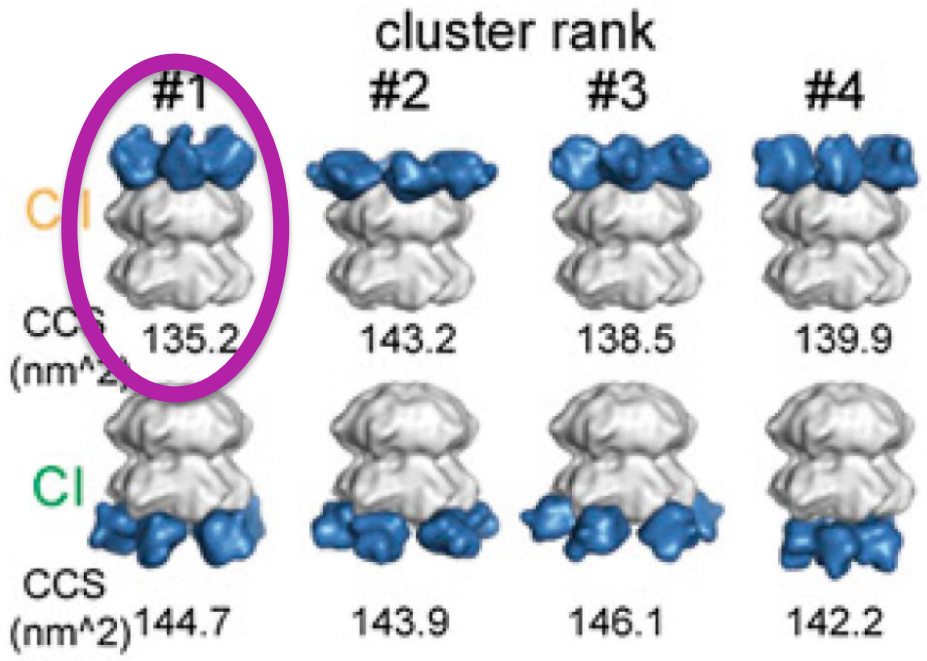
The KaiB-KaiC interaction: HDX

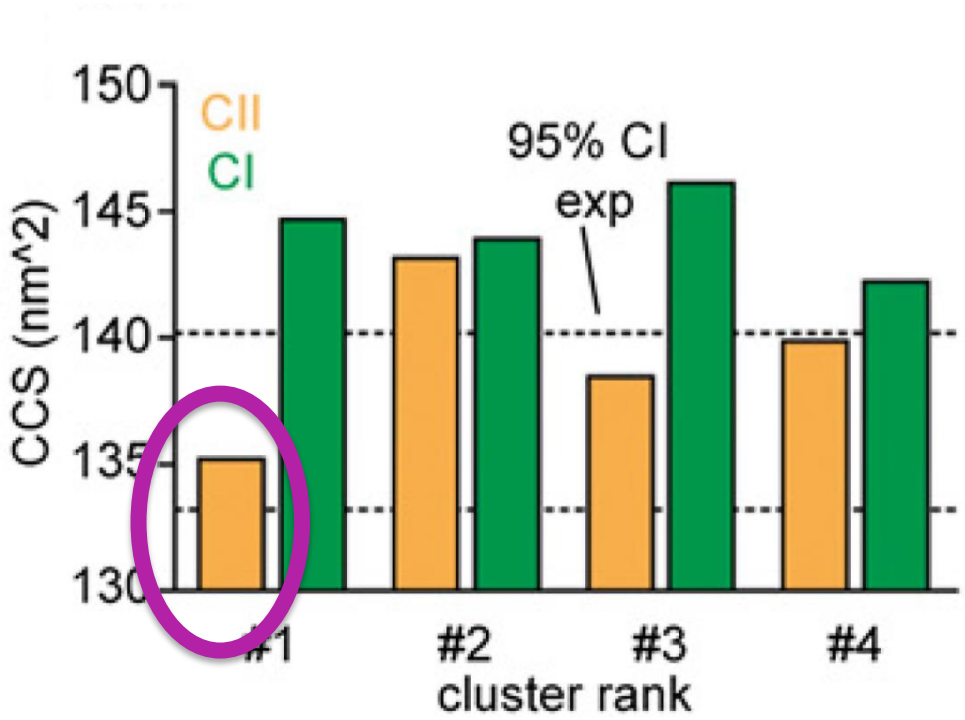
KaiC



The KaiB-KaiC interaction: CCS

Collision cross section from MS allows to filter the HADDOCKing solutions



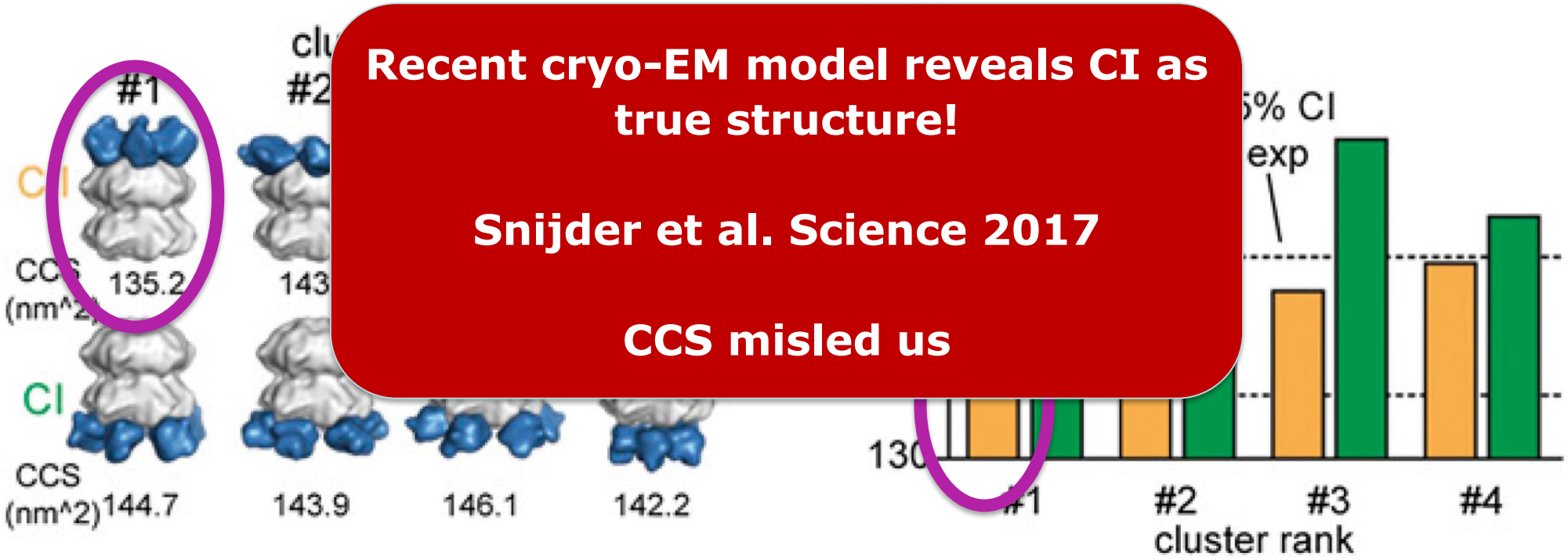


HADDOCK best scoring/most populated solution of CII

Snijder et al. PNAS 111, 1379 (2014)

The KaiB-KaiC interaction: CCS

Collision cross section from MS allows to filter the HADDOCKing solutions



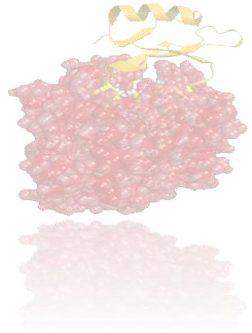
Recent cryo-EM model reveals CI as true structure!

Snijder et al. Science 2017

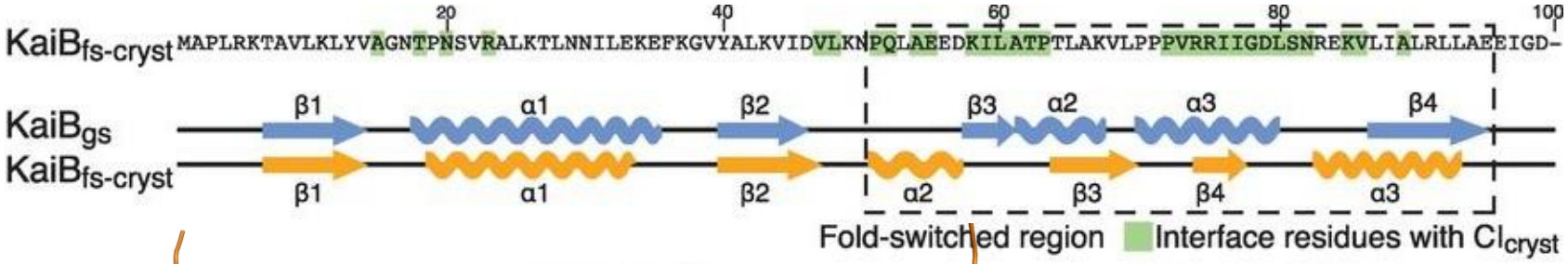
CCS misled us

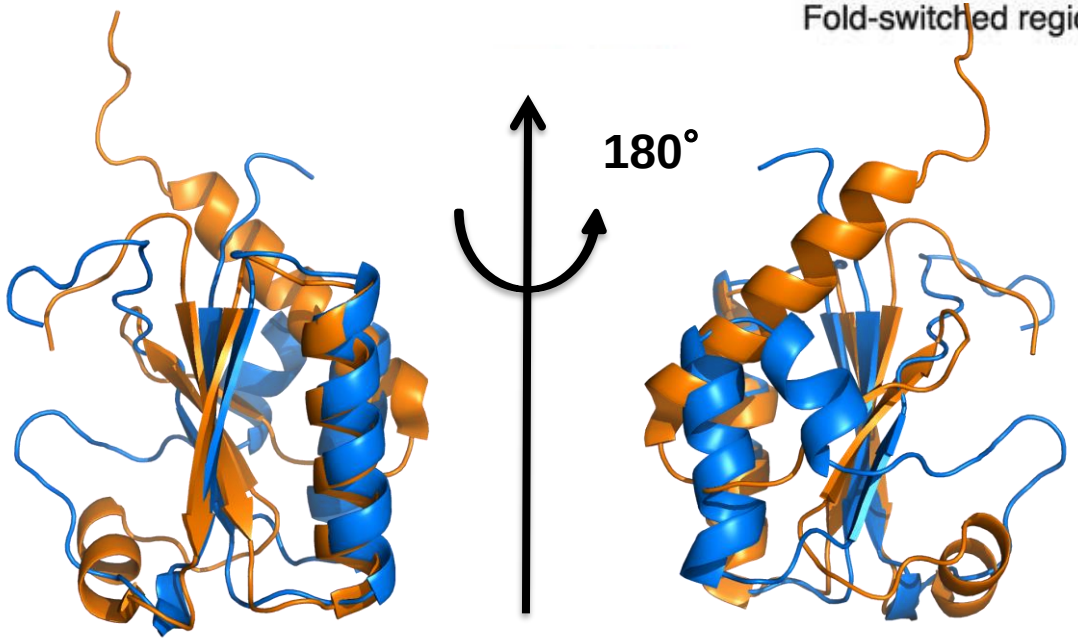
HADDOCK best scoring/most populated solution of CII

Fooled by KaiB!



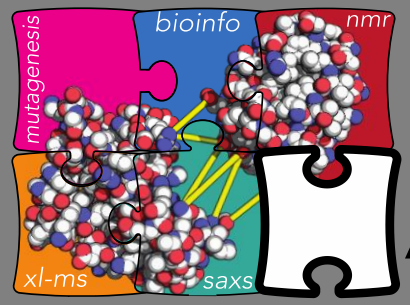
Recent structure of KaiB reveals a different fold for the low populated monomeric form





"KaiB belongs to a rare class of so-called metamorphic proteins, which reversibly switch between different folds under native conditions. KaiB transitions from a highly populated, inactive tetrameric ground-state fold ($KaiB_{gs}$) to a rare, active-state monomeric fold ($KaiB_{fs}$)"

INTERMEZZO



HADDOCK - CG
High-Ambiguity Driven Docking

MARTINI 2.2p
De Jong et al. JCTC 2013

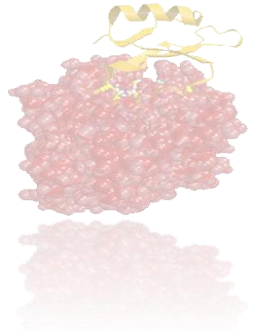
Jorge Roel-Touris

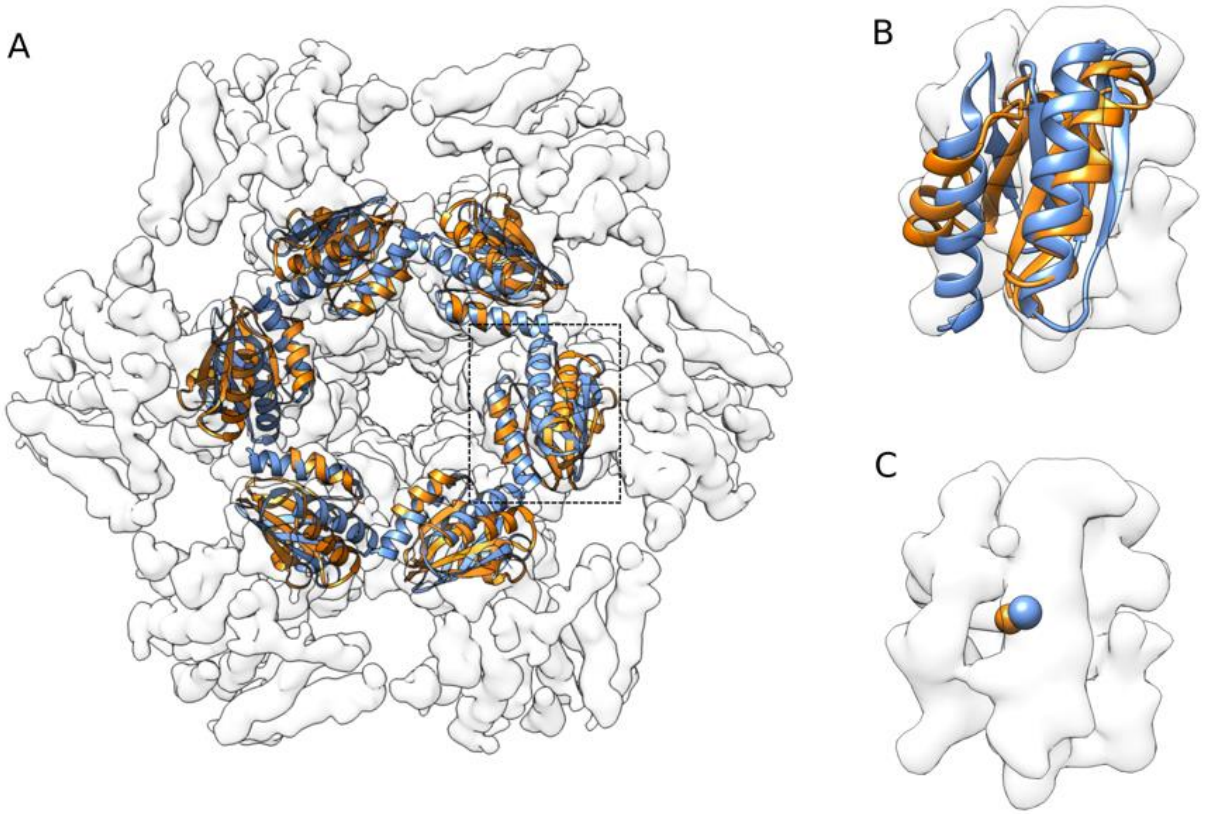
All Atom to Coarse Grain

DOCKING

Coarse Grain to All Atom

Full 7 body 6:1 KaiB:KaiC docking





- **HDX + mutagenesis**
- **C6 symmetry restraints**
- **7-body simultaneous docking with HADDOCK-CG**

Now consistent with cryo-EM

Haddock score

- best CI model -216 ± 13

- best CII model +45 ± 19

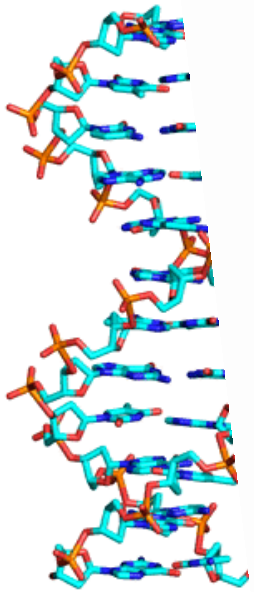
~7 fold speed-up

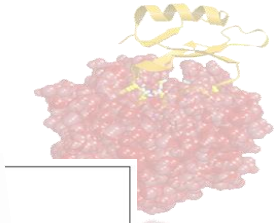
Independent validation:

- Fitting in cryo-EM map using Chimera
- Correlation score: 0.82 (vs 0.84 for EM model PDB-UD 5N8Y)

Coming soon:

- **Protein-DNA/RNA Coarse-grained docking**
- **Und**





frontiers in Molecular Biosciences

TECHNOLOGY AND CODE
published: 01 October 2019
doi: 10.3389/fmolb.2019.00102

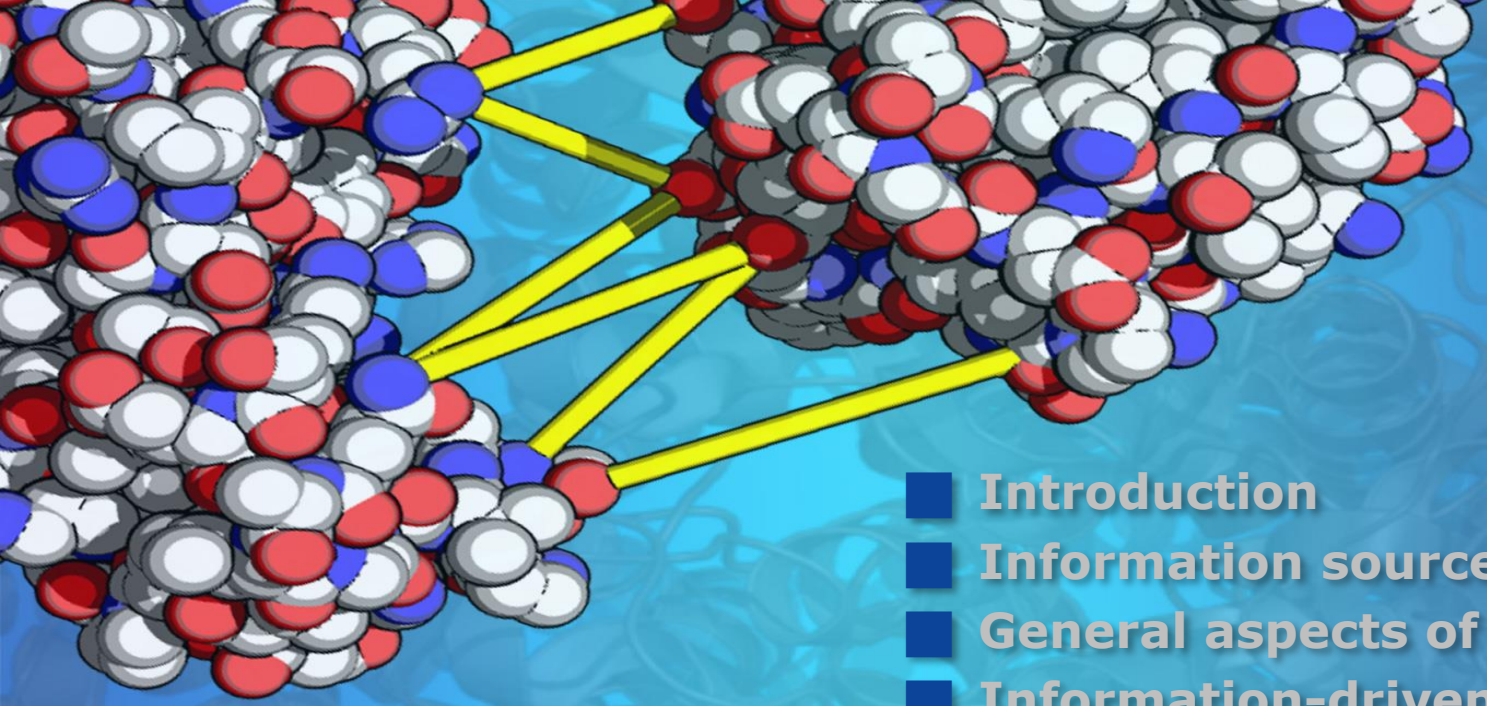
Check for updates

MARTINI-Based Protein-DNA Coarse-Grained HADDOCKing

Rodrigo V. Honorato [1,2†], Jorge Roel-Touris [1†] and Alexandre M. J. J. Bonvin [1*]

[1] Faculty of Science–Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, Netherlands, [2] Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas, Brazil

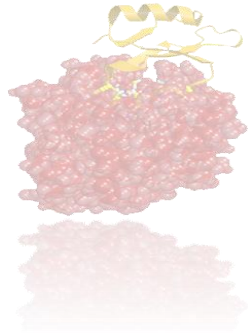
...ingle structure comparison of top-ranking models predicted by HADDOCK. Superimposition of the best models (top-ranked) predicted by HADDOCK using atomistic (blue) or coarse-grained (orange) docking onto the experimental crystal structure (PDB-ID 4r8p, green; McGinty et al., 2014). The two residues PRC1-Cys85 and H2A-Lys119 which are expected to form a covalent bond (Kerscher et al., 2006; an information used to guide the docking) are shown as spheres. The interface RMSD of the all-atom and coarse-grained top rankings models against the reference crystal structure are 3.23 and 3.0 Å, respectively.



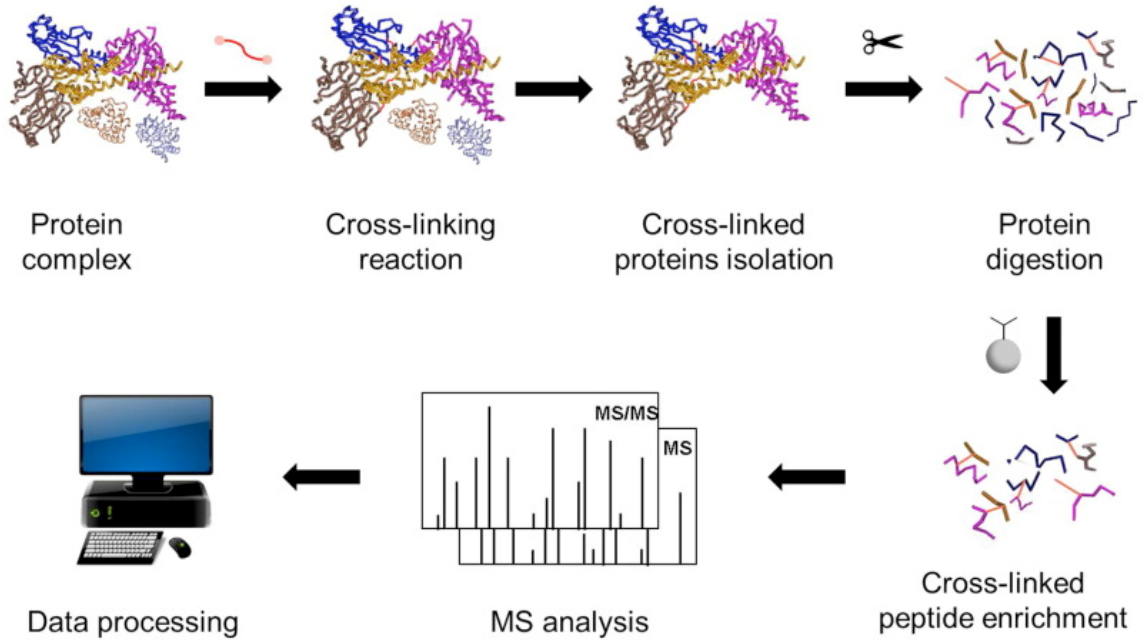
Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
- **Assessing the interaction space**
- Conclusions & perspectives

Distance-based information

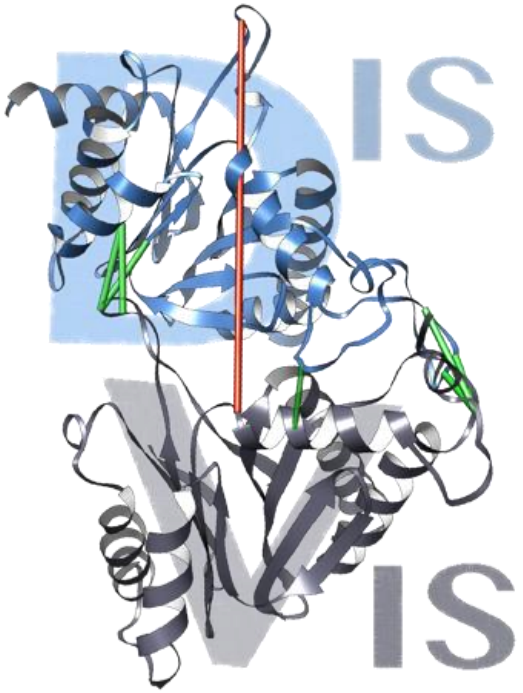


- **Many experimental methods can provide sparse and possibly ambiguous distance information for the modelling of complexes**
- **E.g. cross-links detected by MS provide distance restraints with an upper bound**



Gydo van Zundert, PhD

Defining the information content and consistency of distance restraints



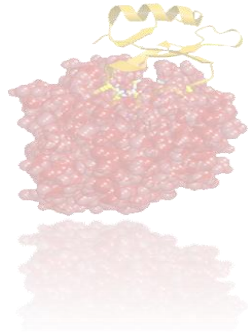
Given 2 interacting structures and a set of distance restraints between them, are there any solutions that satisfy *N* restraints?

A solution is a complex that satisfies all N distance restraints

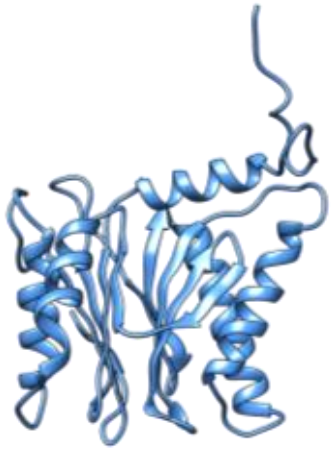
A complex is a conformation where:
The subunits are interacting
The subunits are not clashing

*The **accessible interaction space** is the set of all solutions satisfying at least N restraints*

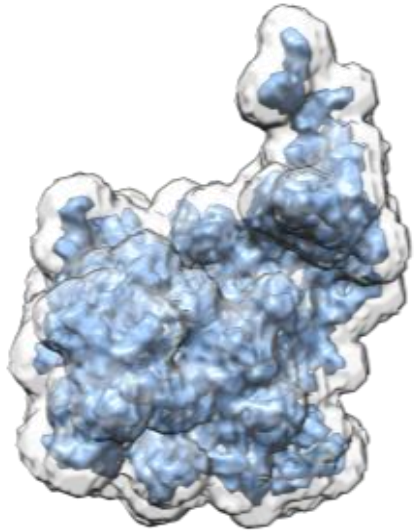
DisVis: re-using old tools to solve new problems



Sample many conformations, by a systematic 6D exhaustive search
(3 rotations and 3 translations)
(rigid-body FFT-docking)



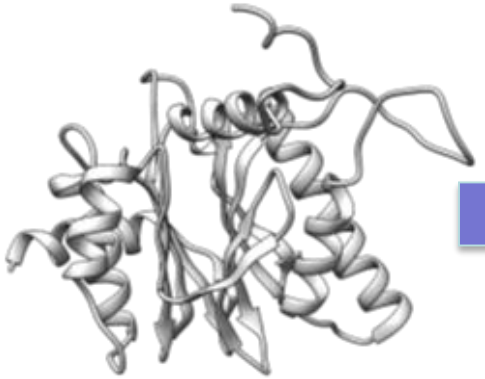
receptor



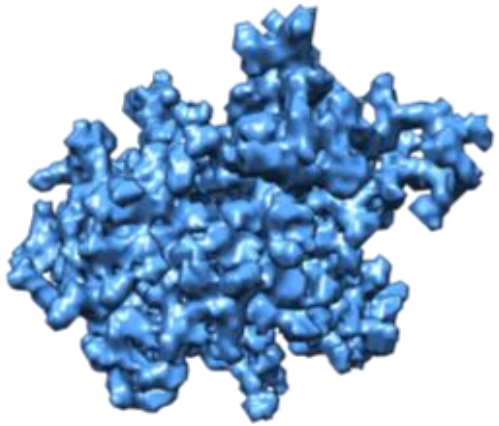
core region
interaction region

For each conformation check whether it is a complex (at least one contact), and count them

For each complex check how many and which restraints are obeyed, and count them

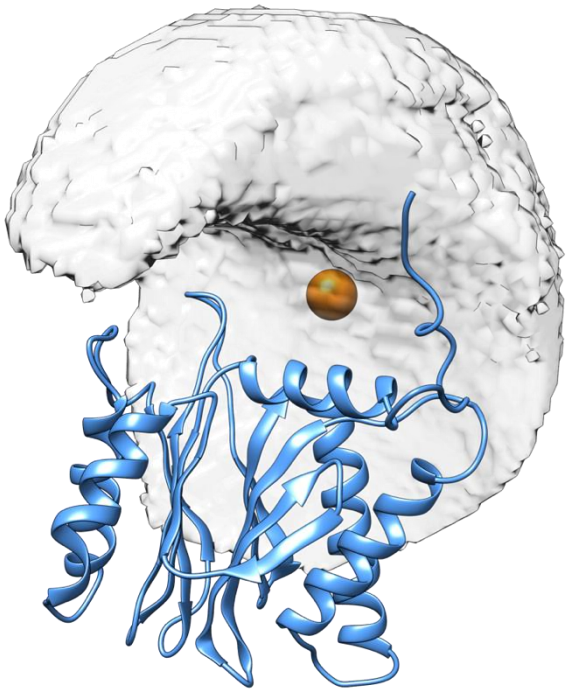


ligand

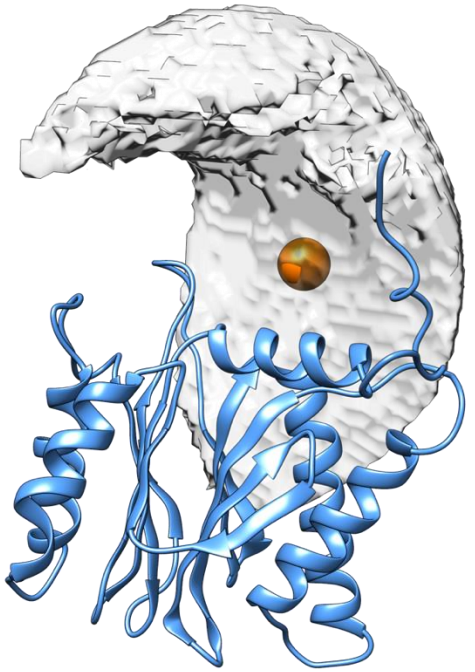


core region

Visualizing the accessible interaction space



Accessible interaction space consistent with at least 5 restraints

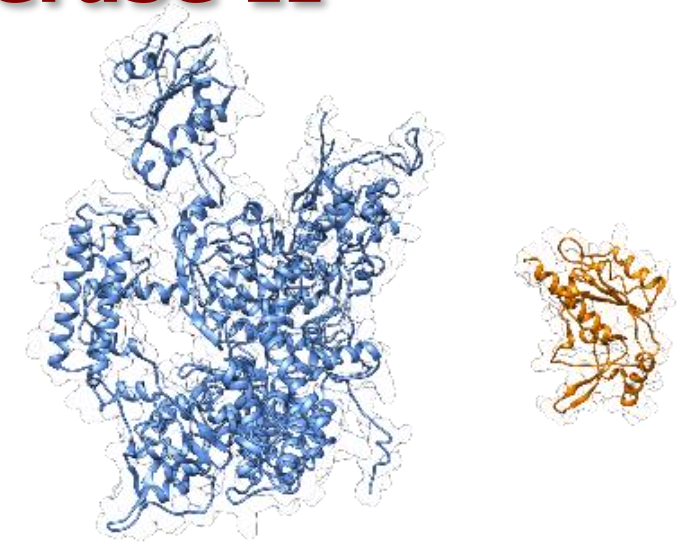


Accessible interaction space consistent with at least 7 restraints

At every grid position, save the maximum number of consistent restraints found during the 6D search

Case study: RNA-polymerase II

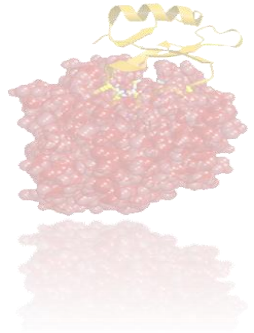
- **Two chains of RNA Polymerase II**
- **Crystal structure available**
- **6 cross-links available**
- **Molecular dynamics trajectory analysis:**
  - 30Å max Lys-Lys distance (C_b – C_b)
- **Added 2 false-positive restraints**



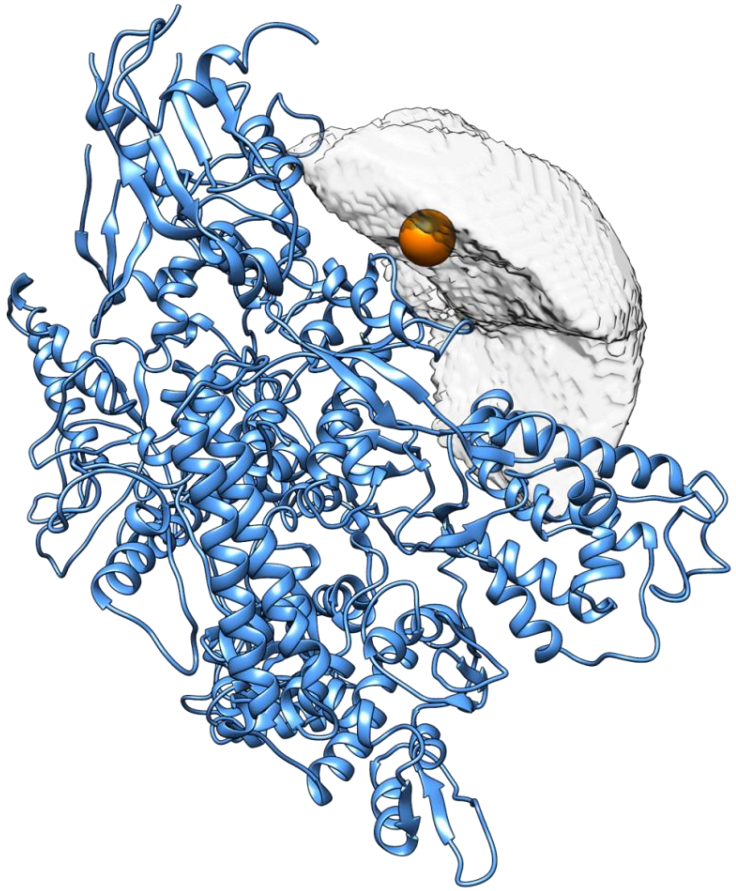
Cross-linker[a]	Distance in complex (Å)[b]
BS3	12.5
BS3	19.8
BS3	12.9
BS3	19.6
BS3	21.8
BS3	15.1
Virtual	35.7
Virtual	42.2

BS3: Bissulfosuccinimidyl suberate

RNA-polymerase II: Accessible interaction space

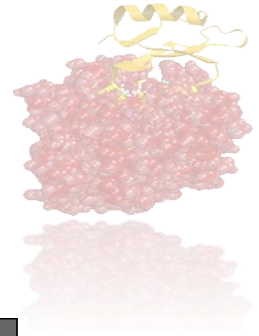


Number of consistent restraints (N)	Number of accessible complexes consistent with at least N restraints	Fraction of accessible complexes consistent with at least N restraints
0	18940752204	1.0000
1	2370295166	0.1251
2	977410985	0.0516
3	298922038	0.0158
4	92651659	0.0049
5	17687776	0.0009
6	5172437	0.0003
7	9716	0.0000
8	0	0.0000



DisVis 6D systematic search with a 1Å grid size and 5.27° interval

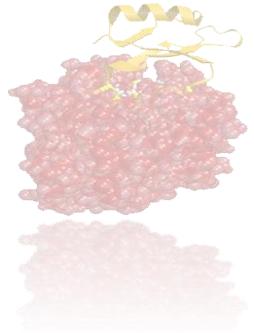
RNA-polymerase II: Detecting false-positive restraints



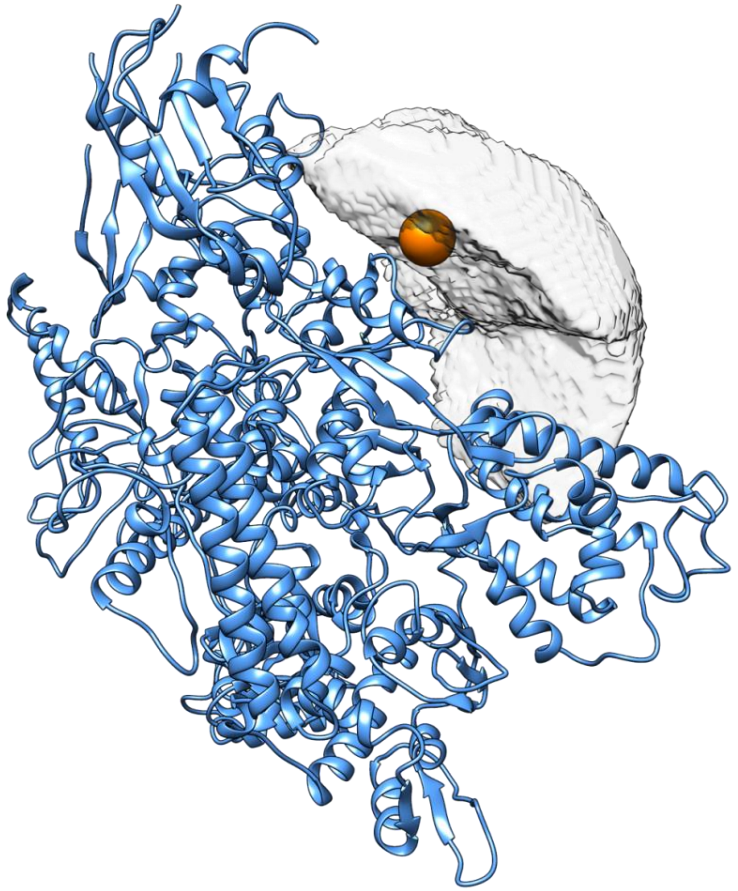
Number of consistent restraints (N)	Fraction of complexes consistent with N restraints in which a specific restraint is violated							
	Restraint 1	Restraint 2	Restraint 3	Restraint 4	Restraint 5	Restraint 6	Restraint 7	Restraint 8
1	0.731	0.813	0.781	0.813	0.742	0.780	0.772	0.981
2	0.676	0.617	0.586	0.725	0.504	0.497	0.974	0.996
3	0.308	0.344	0.285	0.434	0.654	0.622	0.970	0.996
4	0.080	0.151	0.057	0.238	0.653	0.607	0.968	1.000
5	0.015	0.140	0.001	0.371	0.180	0.061	0.940	1.000
6	0.000	0.000	0.000	0.000	0.001	0.000	0.997	1.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

DisVis 6D systematic search with a 1Å grid size and 5.27° interval

RNA-polymerase II: Accessible interaction space

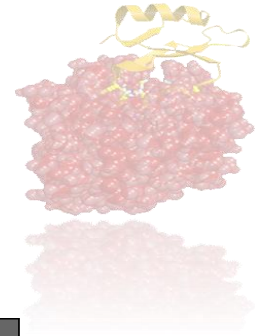


Number of consistent restraints (N)	Number of accessible complexes consistent with at least N restraints	Fraction of accessible complexes consistent with at least N restraints
0	18940752204	1.0000
1	2370295166	0.1251
2	977410985	0.0516
3	298922038	0.0158
4	92651659	0.0049
5	17687776	0.0009
6	5172437	0.0003
7	9716	0.0000
8	0	0.0000



DisVis 6D systematic search with a 1Å grid size and 5.27° interval

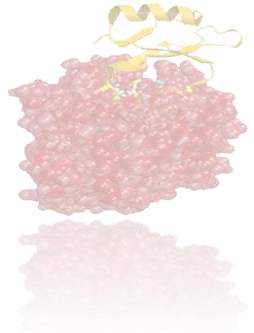
RNA-polymerase II: Detecting false-positive restraints

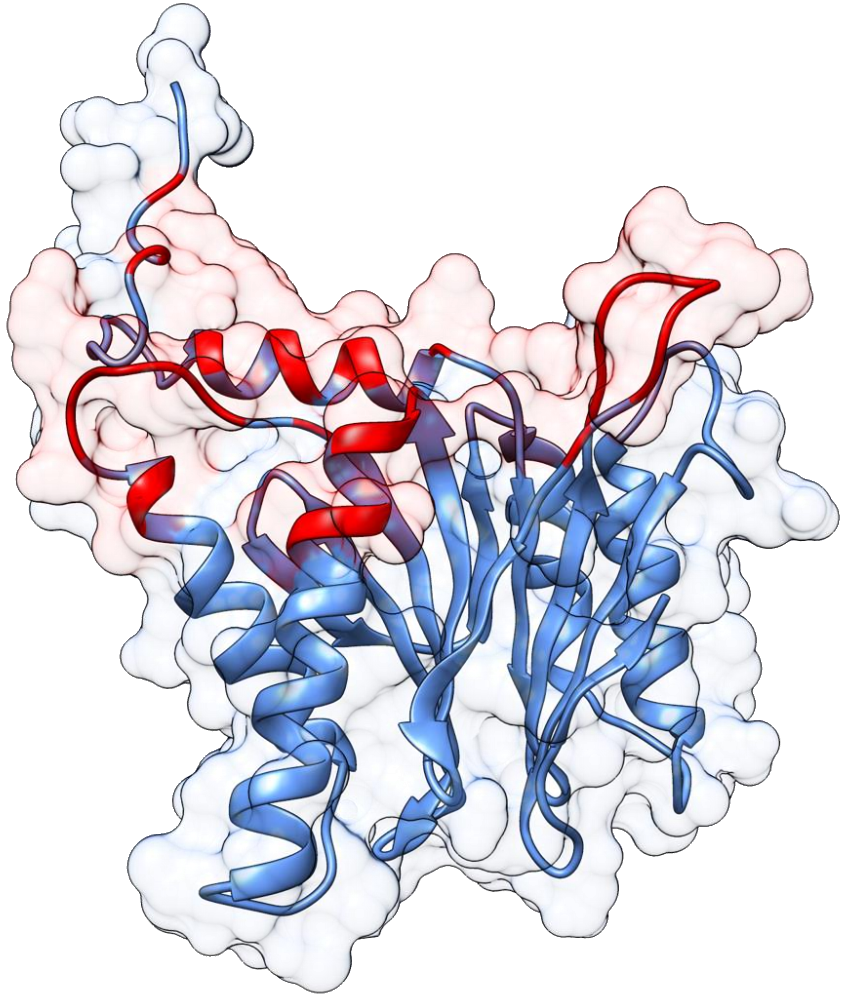


Number of consistent restraints (N)	Fraction of complexes consistent with N restraints in which a specific restraint is violated							
	Restraint 1	Restraint 2	Restraint 3	Restraint 4	Restraint 5	Restraint 6	Restraint 7	Restraint 8
1	0.731	0.813	0.781	0.813	0.742	0.780	0.772	0.981
2	0.676	0.617	0.586	0.725	0.504	0.497	0.974	0.996
3	0.308	0.344	0.285	0.434	0.654	0.622	0.970	0.996
4	0.080	0.151	0.057	0.238	0.653	0.607	0.968	1.000
5	0.015	0.140	0.001	0.371	0.180	0.061	0.940	1.000
6	0.000	0.000	0.000	0.000	0.001	0.000	0.997	1.000
7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

DisVis 6D systematic search with a 1Å grid size and 5.27° interval

Mapping the interface

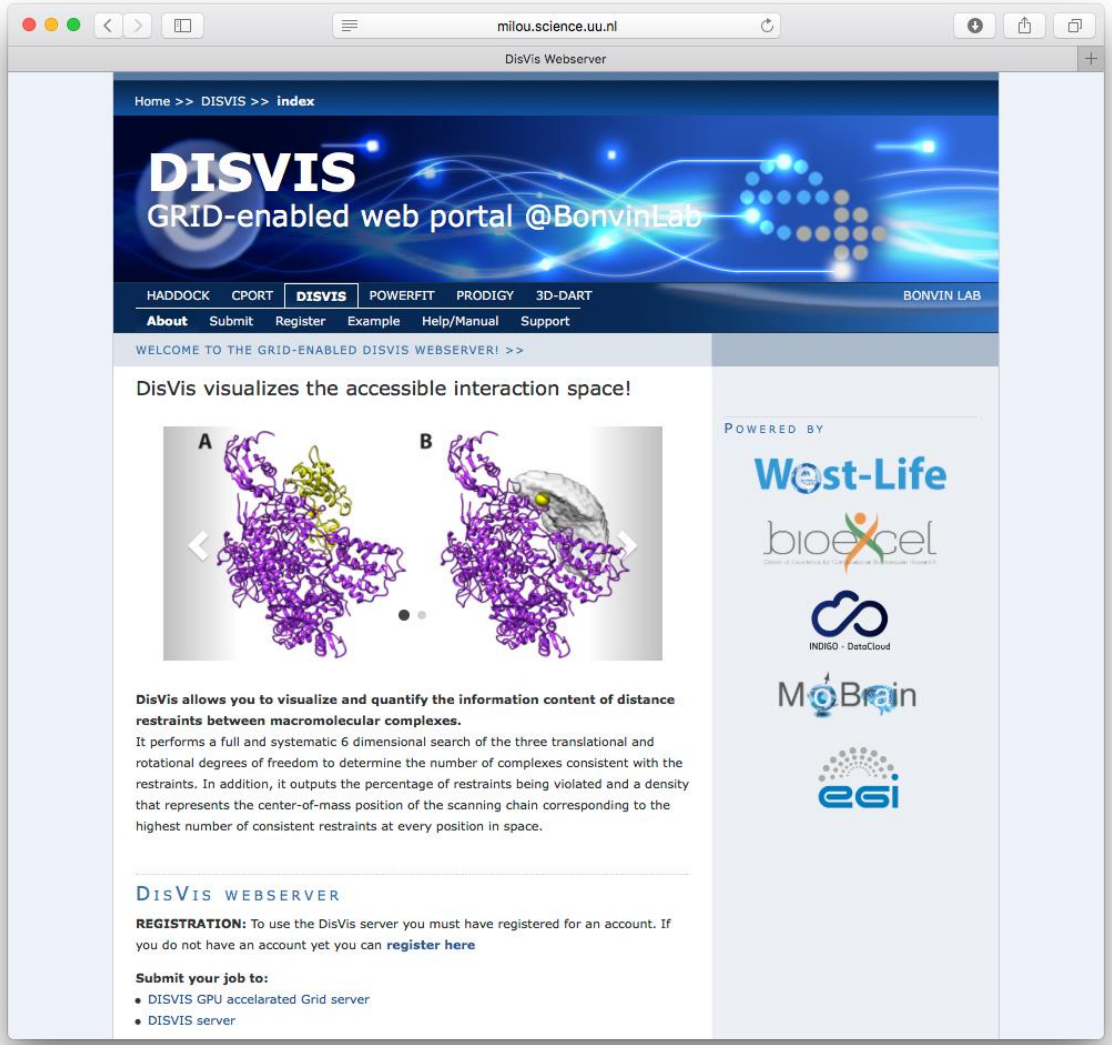




Interface residues from consistent solutions

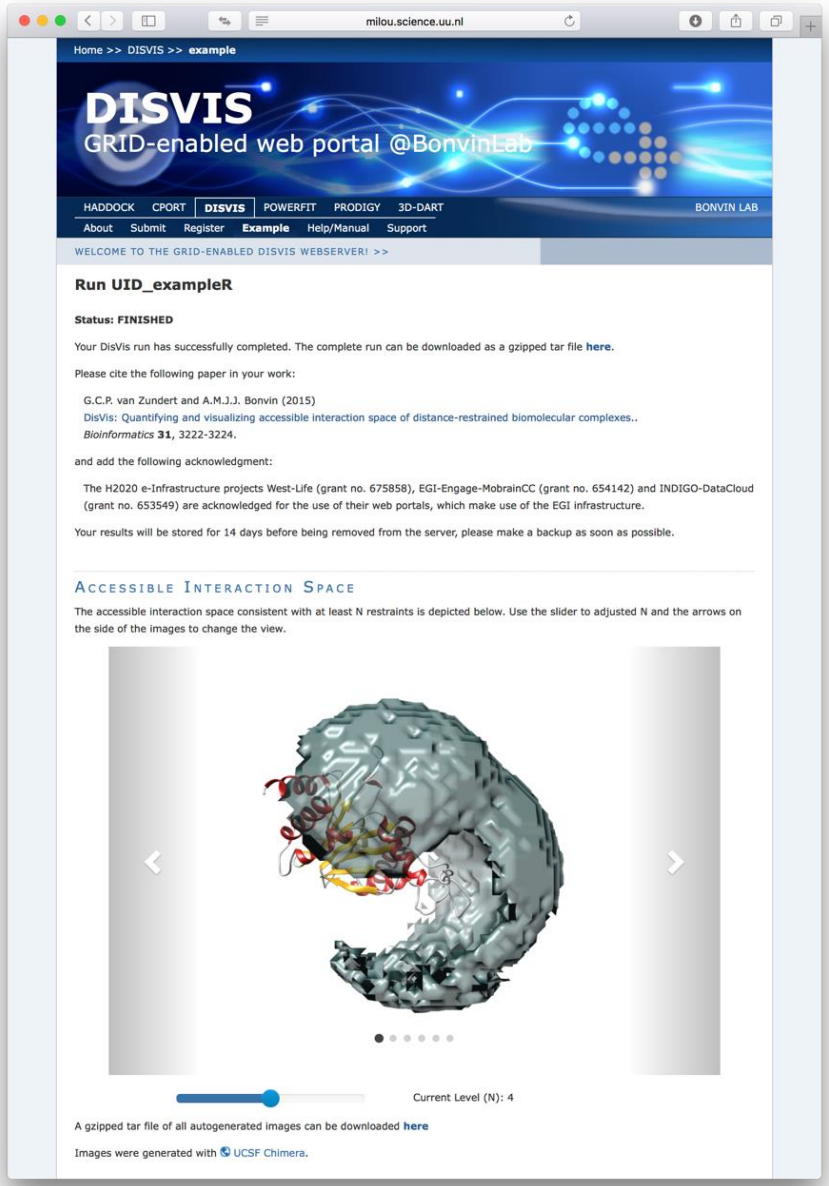
DISVIS: grid, GPGPU-enabled web portal

Mikael Trellet

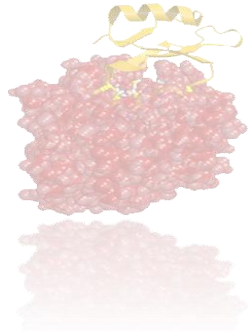


Van Zundert *et al.*, J. Mol. Biol. (2017)

Jörg Schaarschmidt



Guided interpretation of results

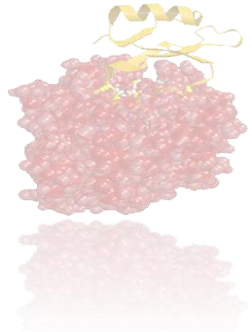


z-Score

The table below features the z-Score for each restraint. The higher the score, the more likely the restraint is a false-positive. Z-scores above 1.0 are explicitly mentioned in the output of DisVis.

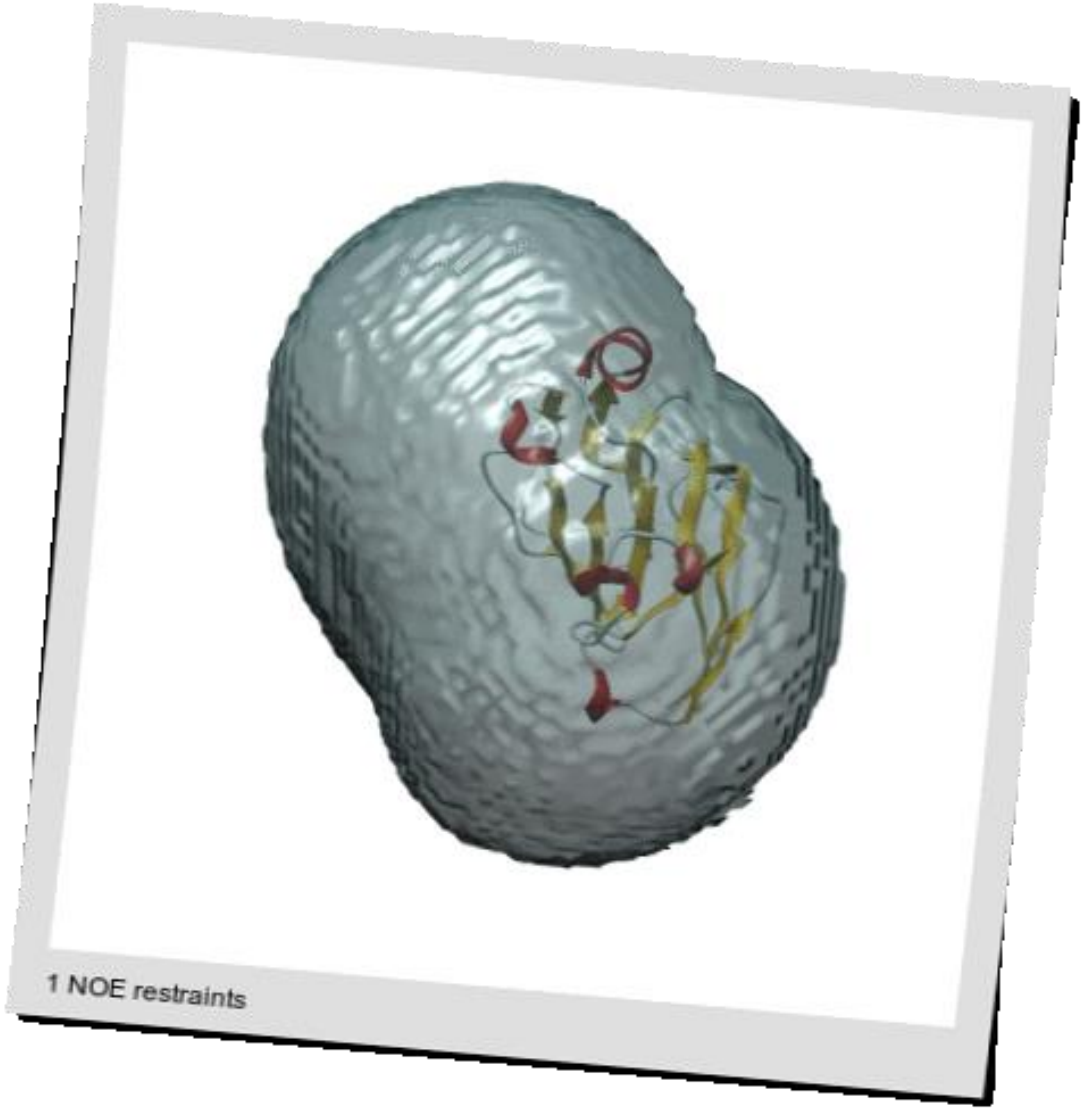
#	Restraint	Average violated fraction	Standard deviation	Z-score
8	**A1092(CB)-E152(CB)**	**1.00**	**0.01**	**2.05**
7	**A180(CB)-E122(CB)**	**0.80**	**0.33**	**1.29**
4	A15(CB)-E171(CB)	0.39	0.30	-0.29
5	A934(CB)-E201(CB)	0.38	0.29	-0.35
6	A938(CB)-E201(CB)	0.36	0.31	-0.39
2	A129(CB)-E161(CB)	0.29	0.29	-0.68
1	A1003(CB)-E166(CB)	0.25	0.30	-0.82
3	A129(CB)-E171(CB)	0.25	0.30	-0.82

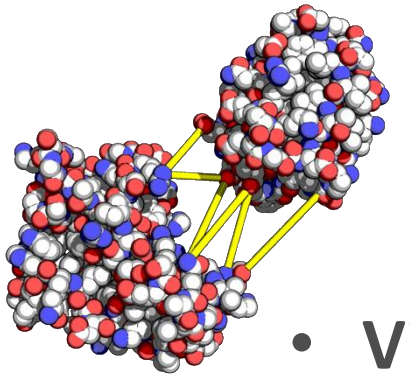
Not limited to MS cross-links



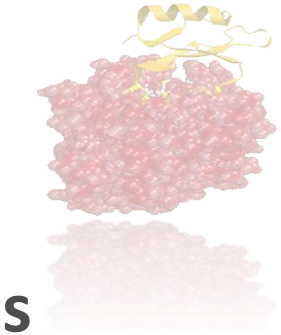
E2A-HPR mapping from unbound structures using 56 intermolecular NOEs

(Wang et al, EMBO J 2000)



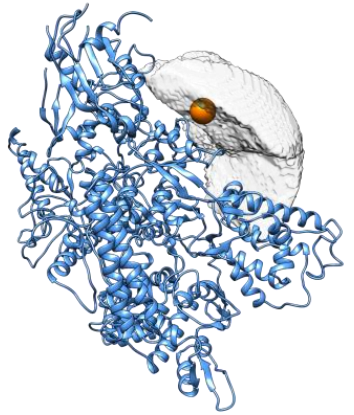


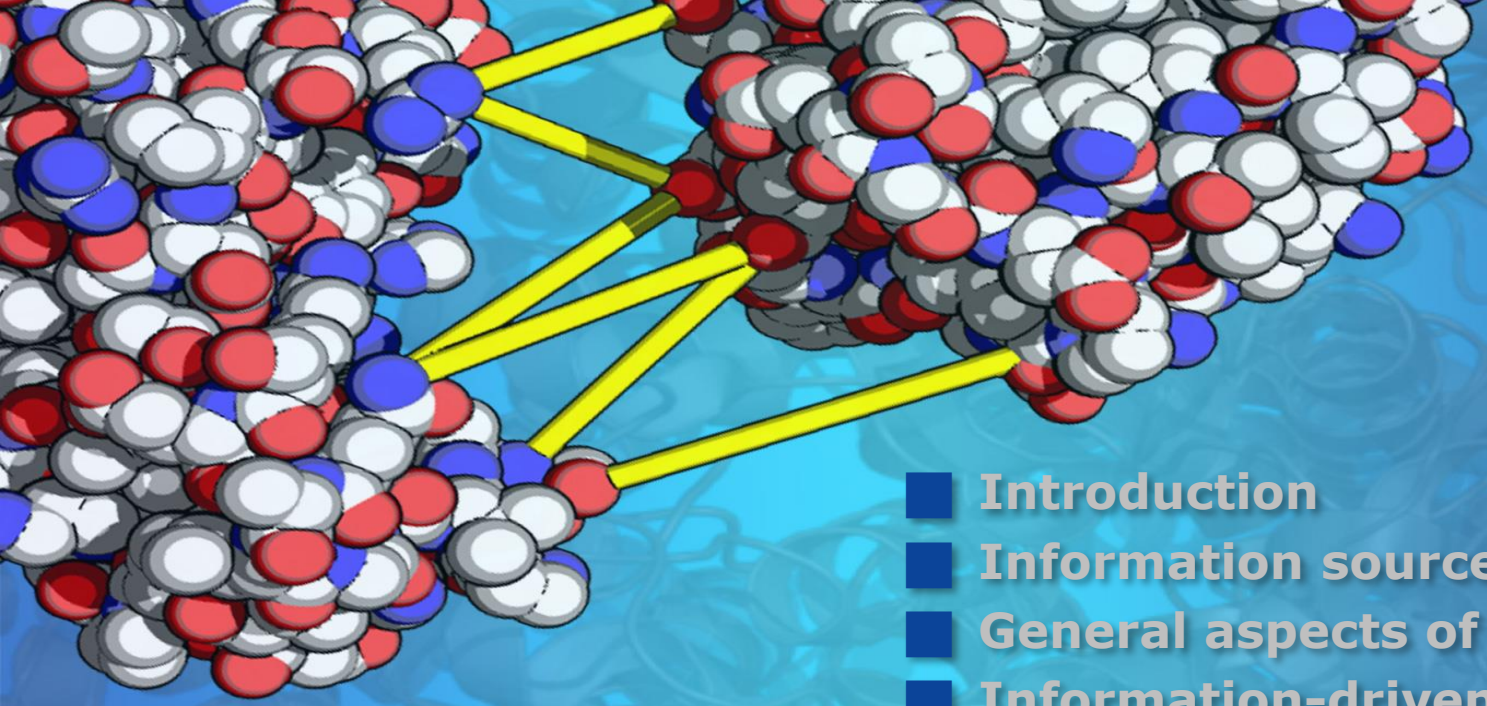
Conclusions - DISVIS



- Visualization the information content of distance restraints
- Solely based on geometric considerations
- Identification of possible false positives
- Provides information about possible interfaces, valuable information to guide modelling

- **BUT**: Does not account for conformational changes and energetics

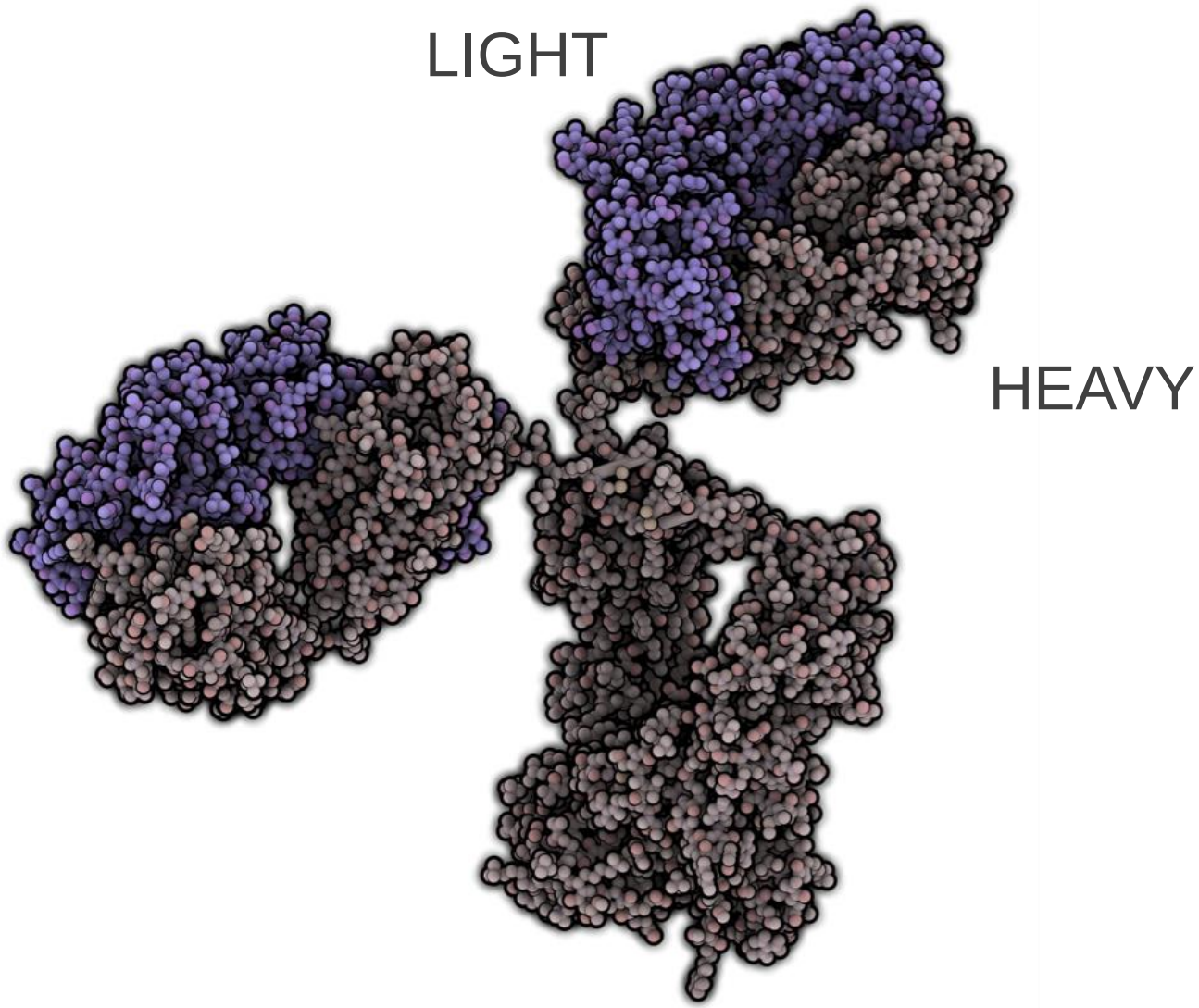




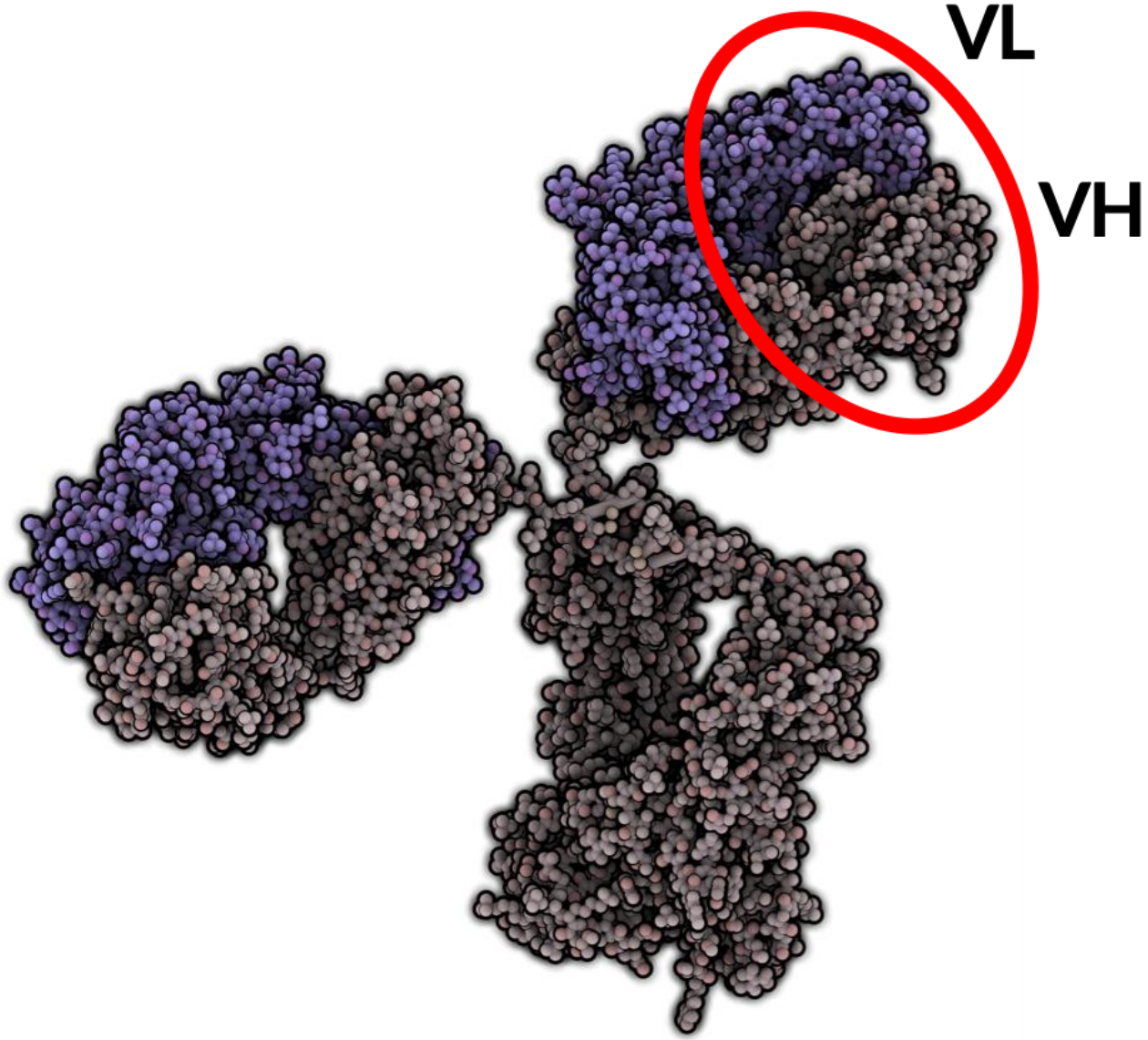
Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
- Assessing the interaction space
- **Bonus topic: Antibody antigen modelling**
- Conclusions & perspectives

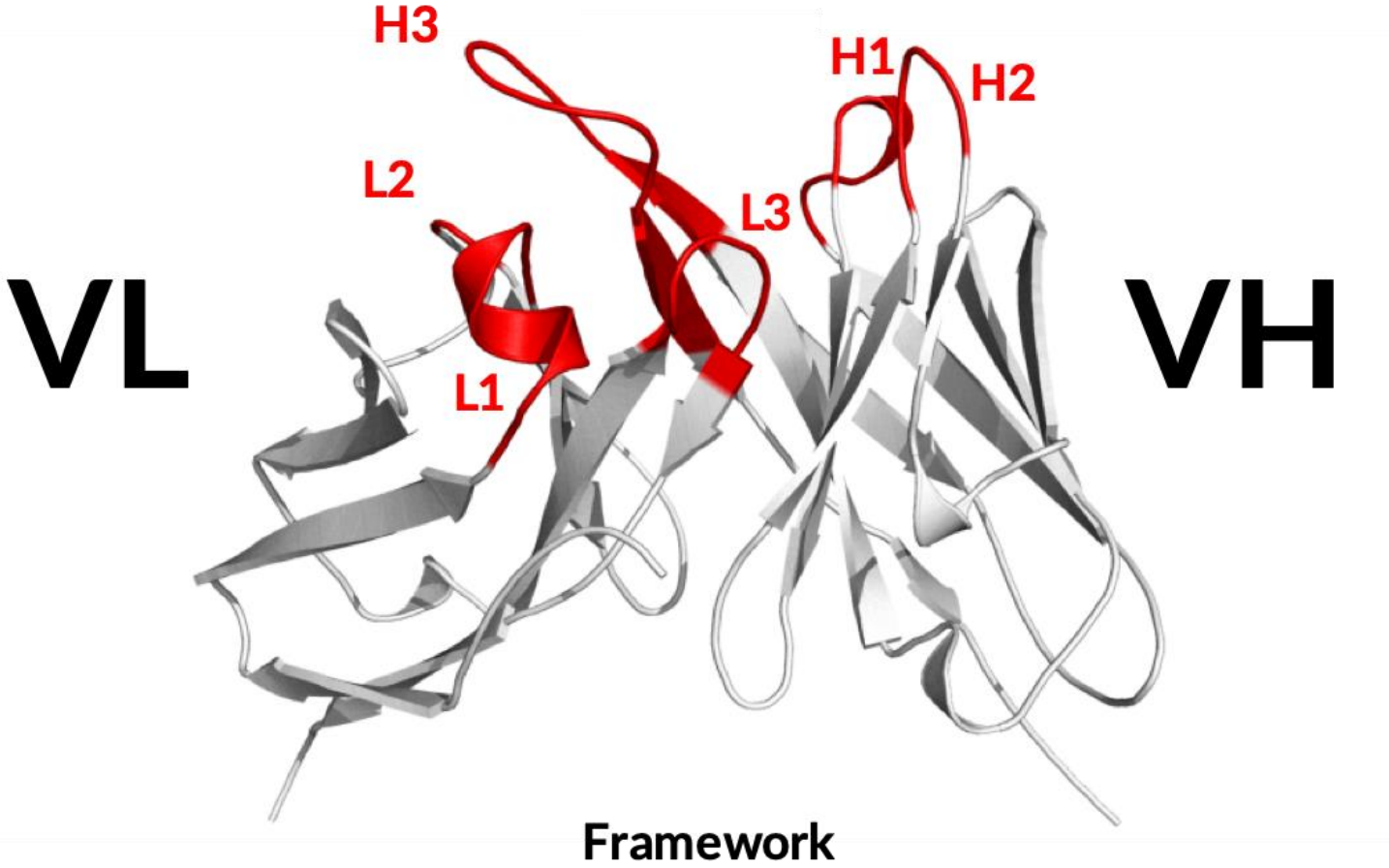
Antibody structure



Antibody structure

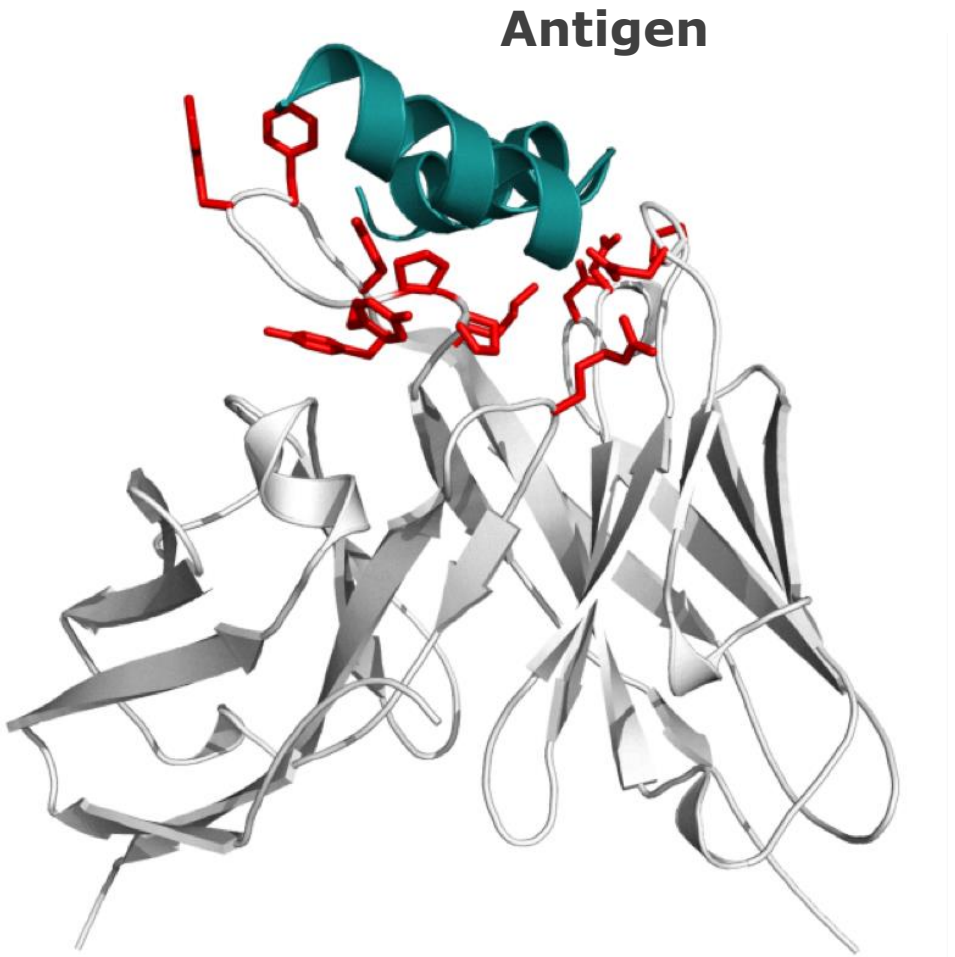


Variable domains: Complementarity Determining Regions

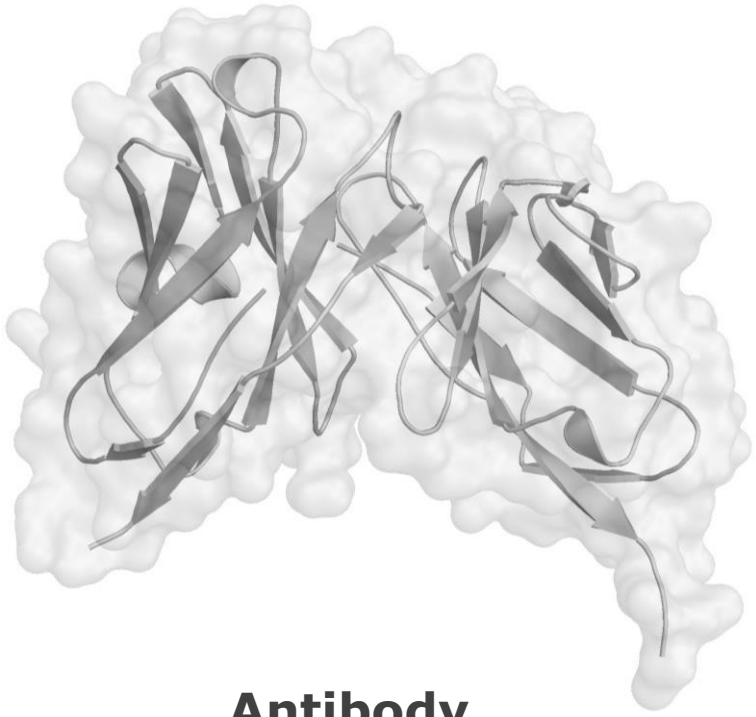


Antibody-Antigen binding

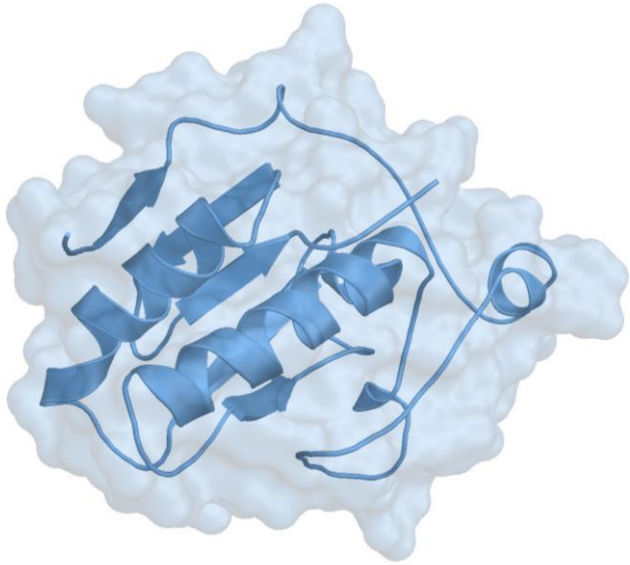
- The antibody region able to bind the antigen is named **paratope**
- The antigen region recognised by the antibody is called **epitope**



Antibody Docking Dataset



Antibody



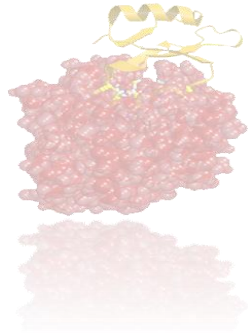
Antigen

16 complexes with unbound structures from docking benchmark 5

Vreven, T. et al. ***Updates to the Integrated Protein-Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2****.*
J. Mol. Biol. (2015). doi:10.1016/j.jmb.2015.07.016

Francesco
Ambrosetti

Antibody-antigen modelling

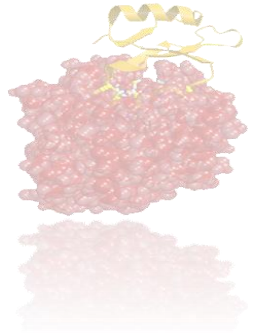


- 4 software with specific options for antibody docking considered

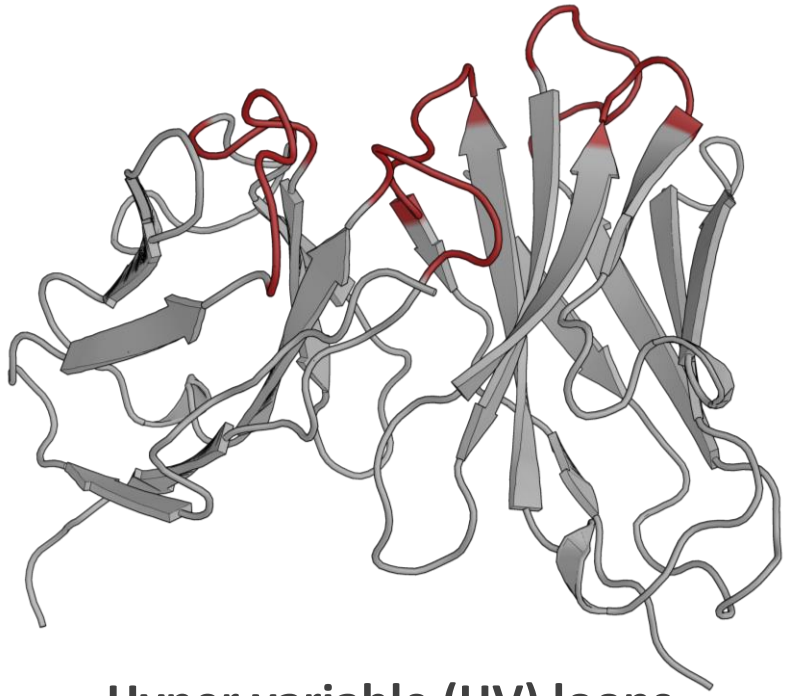
- Restraints used either in scoring (ClusPro, ZDOCK) or to drive the docking (HADDOCK, LightDock)

Antibody-antigen modelling:

Information used



Antibody

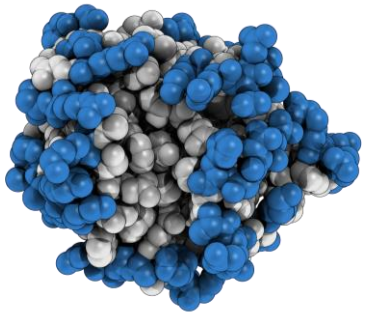


Hyper variable (HV) loops

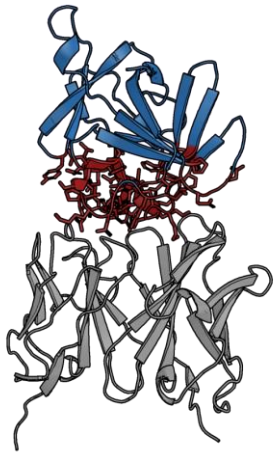
HV loops - Surface

HV loops - Epi 9

Antigen



Surface residues



Epitope defined at 9Å

+ true interface (at 4.5Å) as reference

Antibody Docking Evaluation criteria

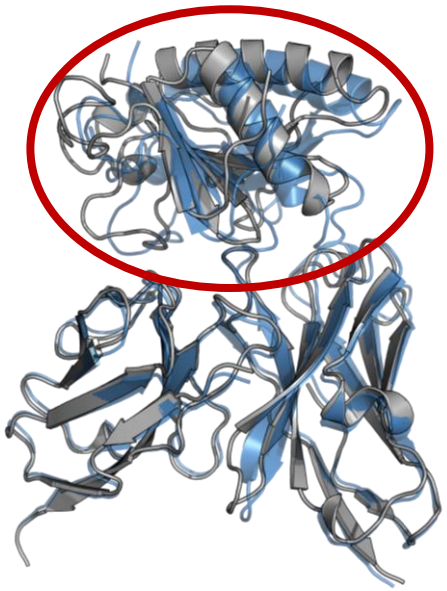
F_{nat}

Reference	
A:26	B:5
A:27	B:8
A:30	B:12
A:32	B:13
A:50	B:14
A:52	B:30
A:96	B:32
A:97	B:33
A:100	B:34
A:103	B:40

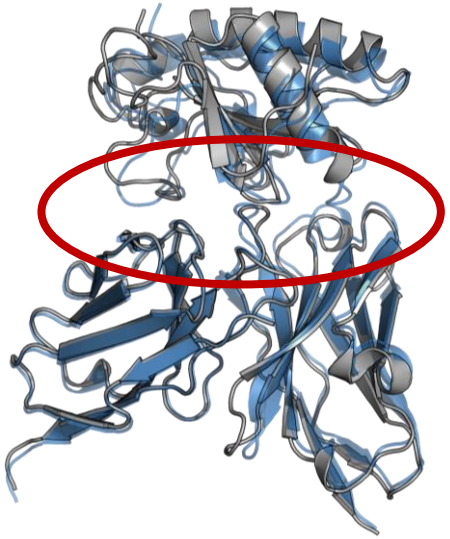
Docking model	
A:28	B:5
A:27	B:8
A:30	B:11
A:32	B:13
A:51	B:18
A:52	B:28
A:97	B:33
A:100	B:34
A:100	B:37
A:103	B:40

$$F_{nat} = \frac{number of common contacts}{number of reference contats}$$

L-RMSD



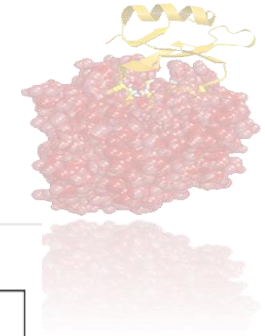
i-RMSD

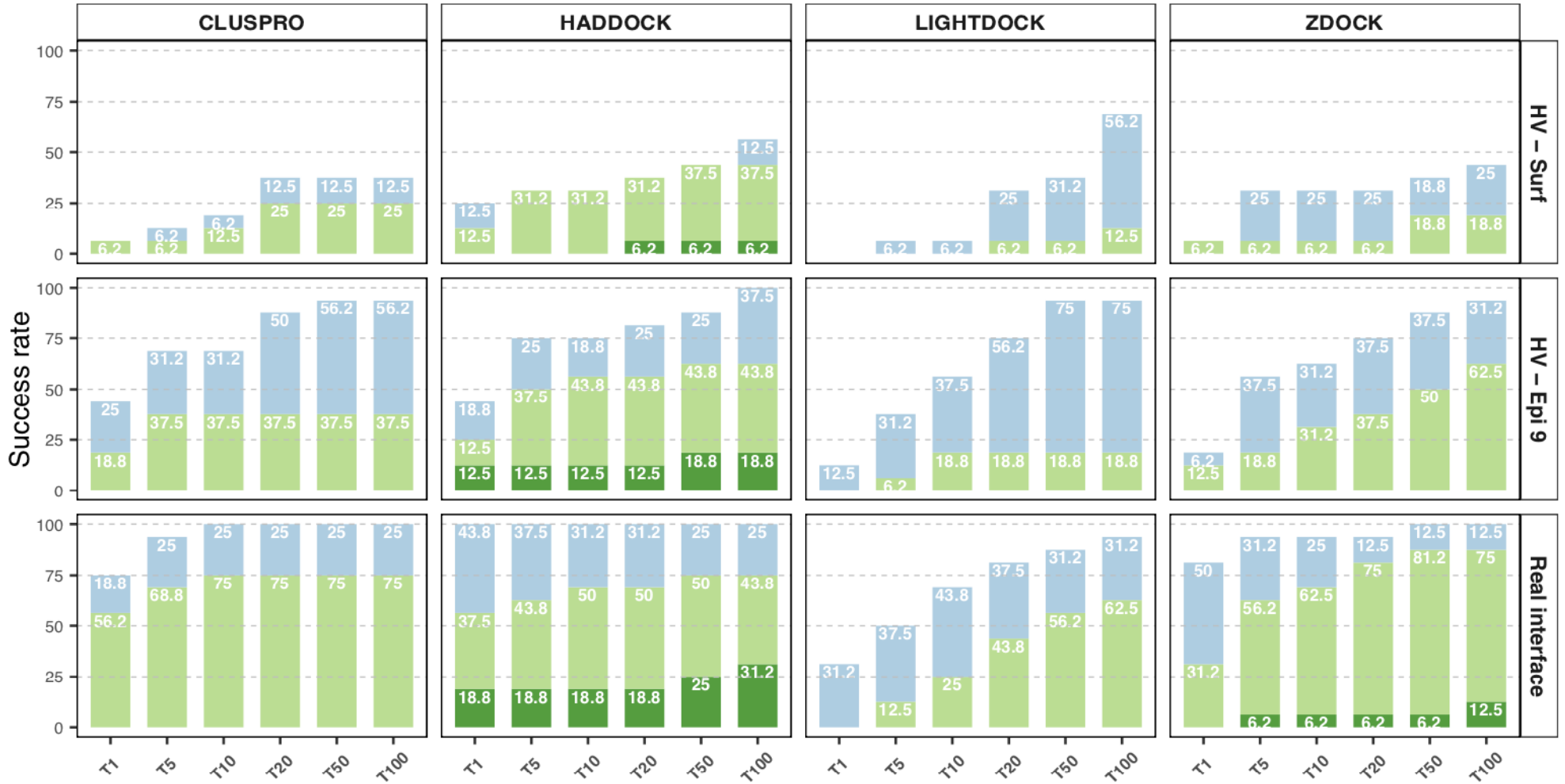


Class	F_{nat}	L-RMSD[Å]	i-RMSD[Å]
High (***)	≥ 0.5	≤ 1.0	or ≤ 1.0
Medium (**)	≥ 0.3	≤ 5.0	or ≤ 2.0
Acceptable (*)	≥ 0.1	≤ 10.0	or ≤ 4.0

Docking success rate

(single structure-based)

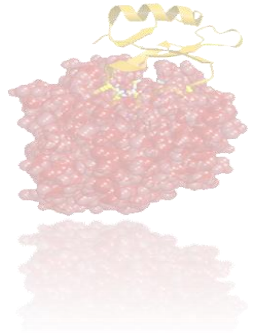


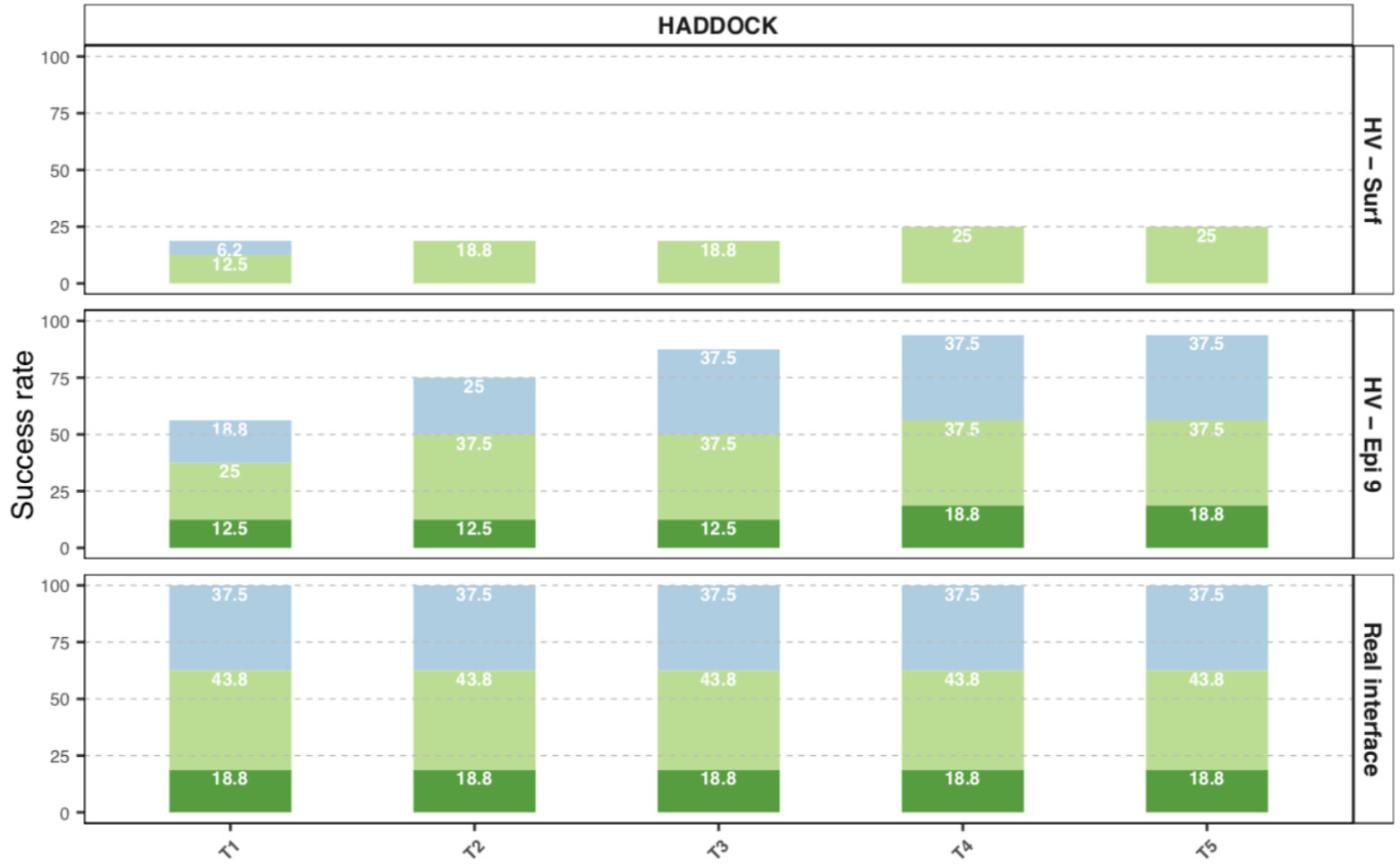


Quality: Acceptable Medium High

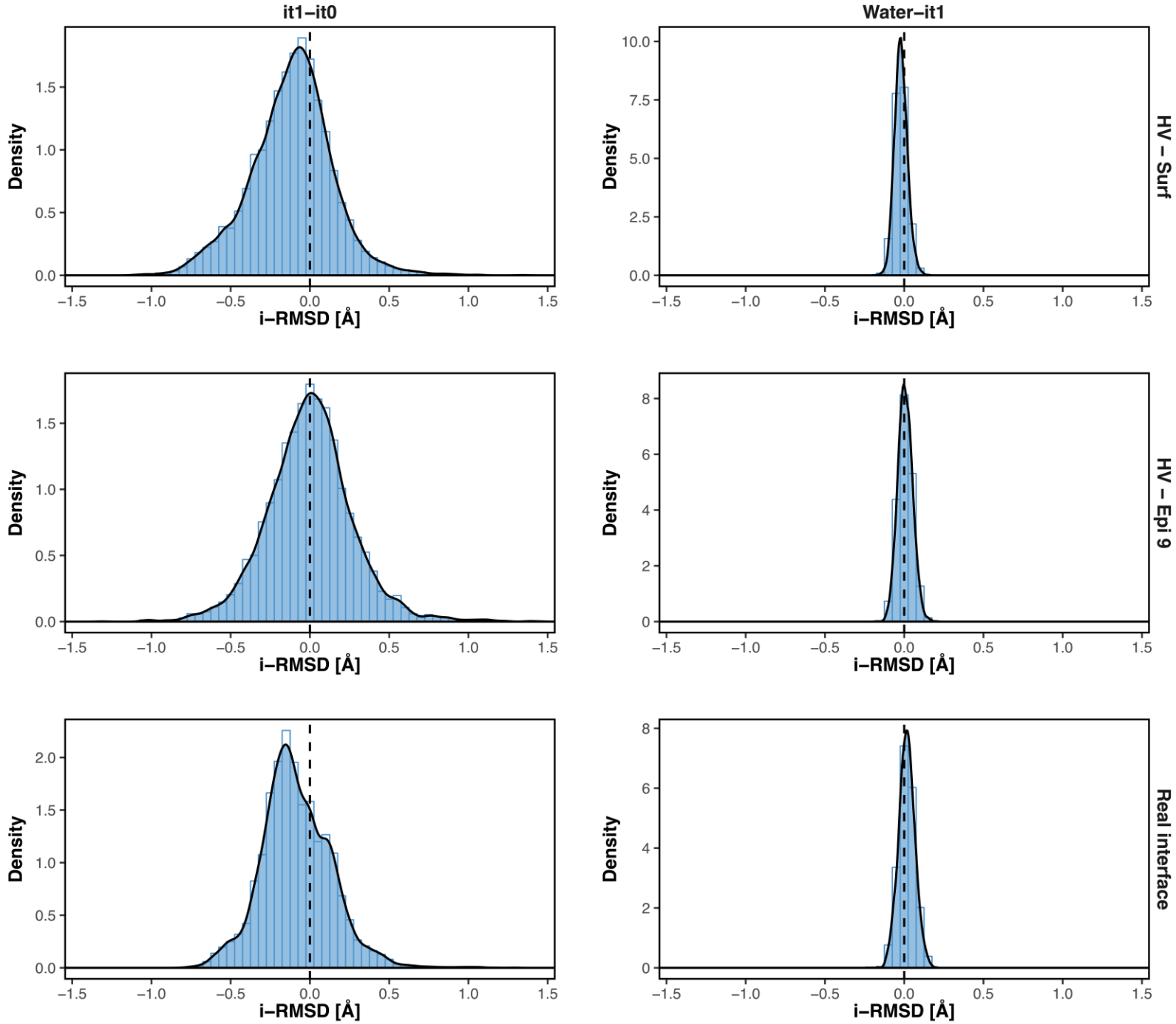
Docking success rate

(cluster-based)

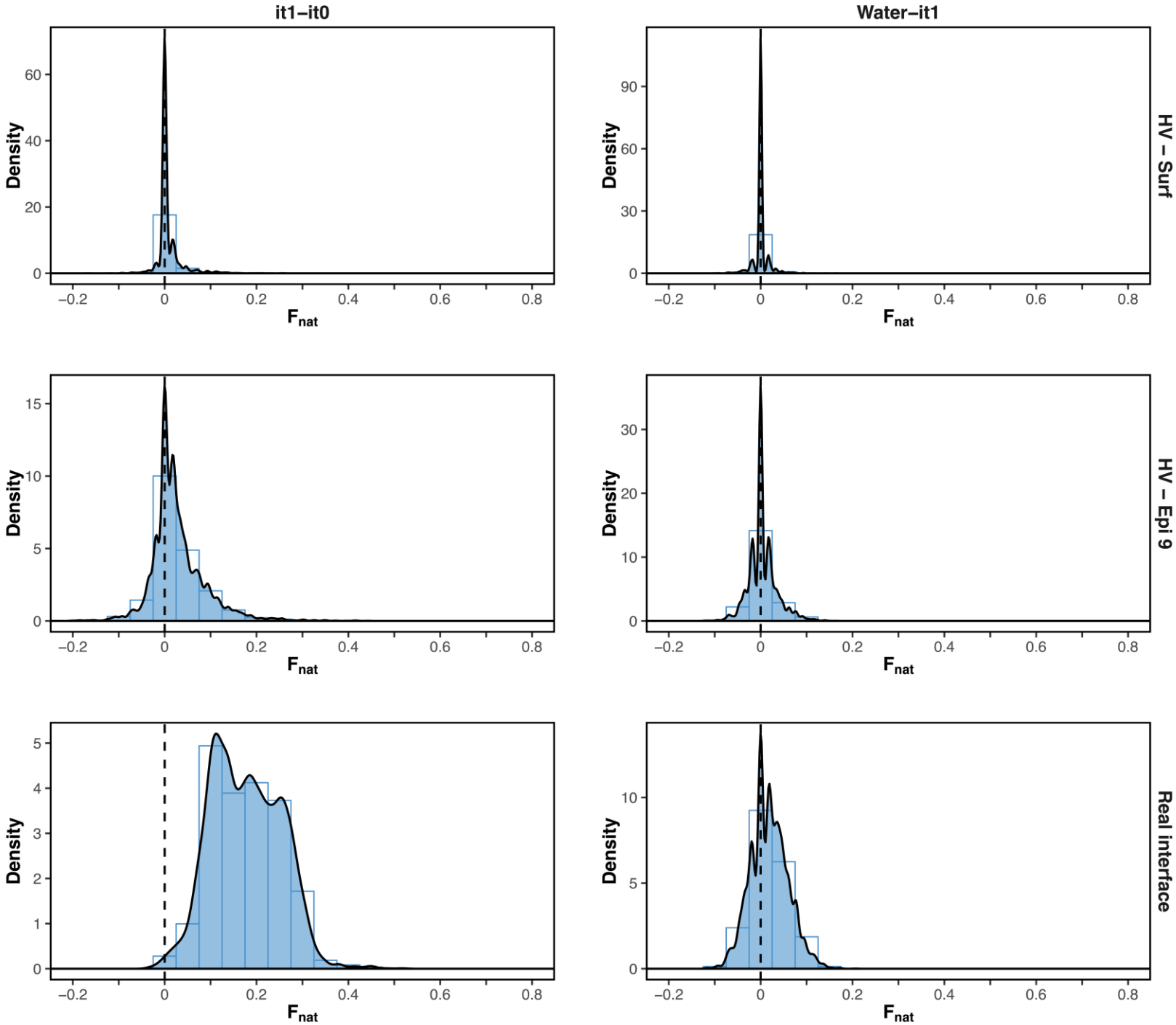




Impact of flexible refinement: i-RMSD distributions



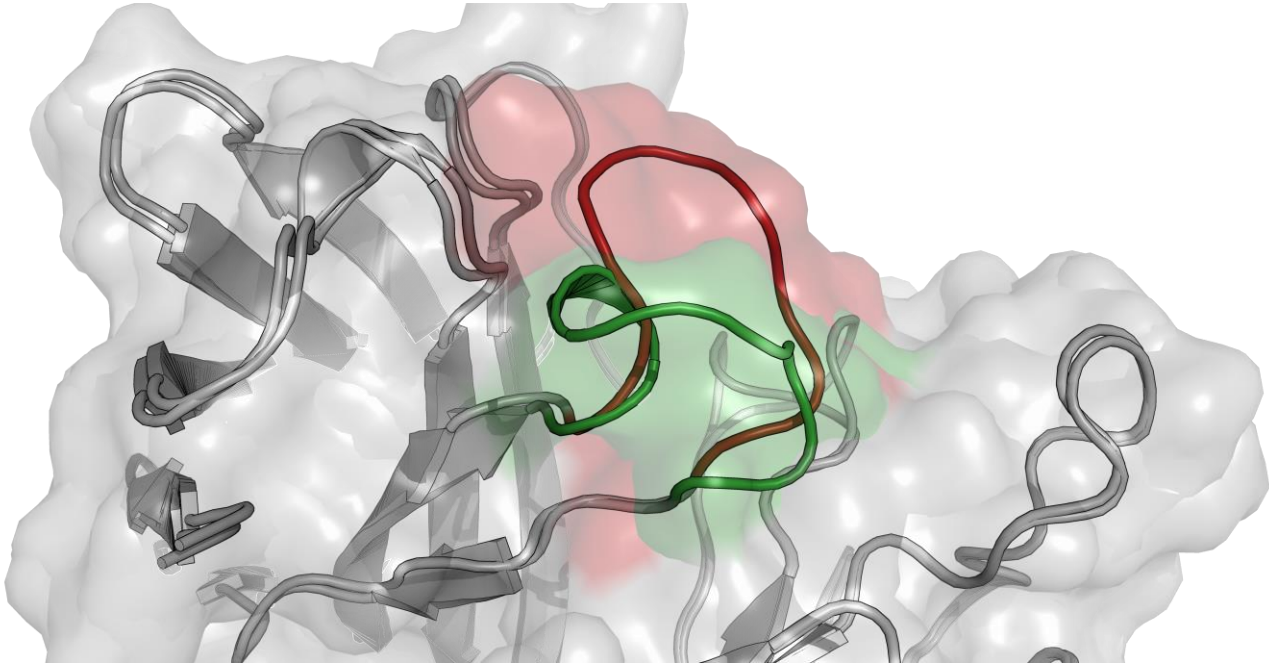
Impact of flexible refinement: Fnat distributions



H3 modelling

H3 is crucial for the antigen recognition

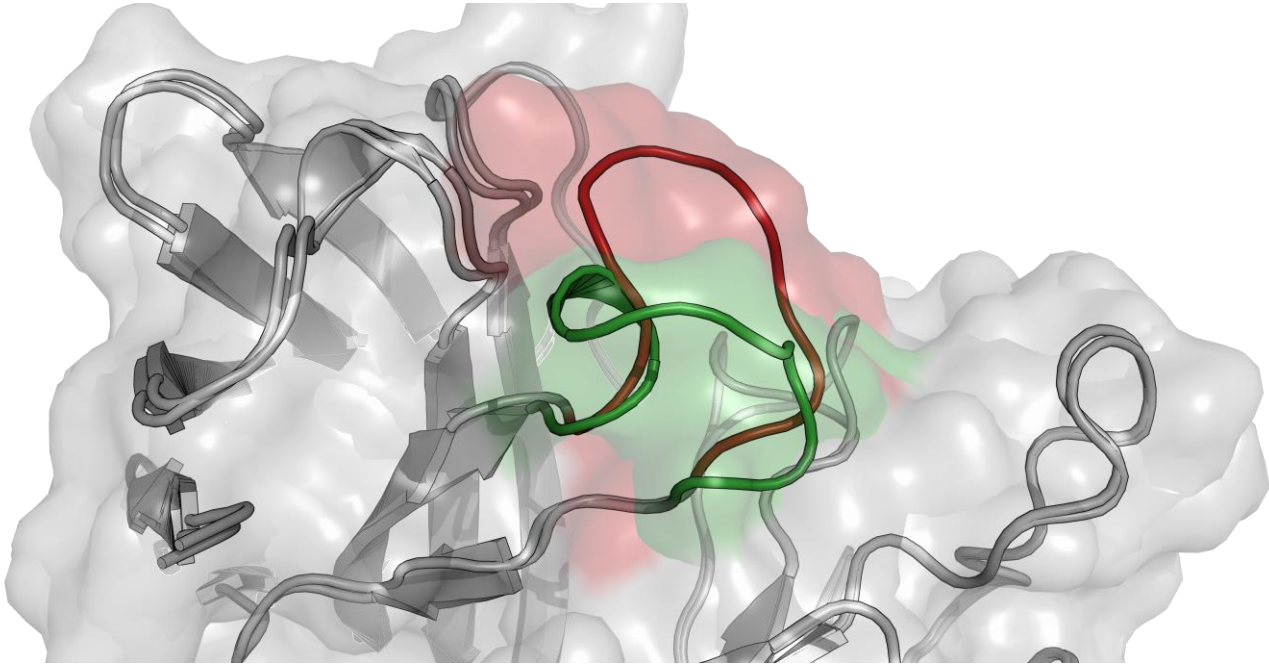
Modelling of the H3 loop of antibodies is still challenging



H3 modelling

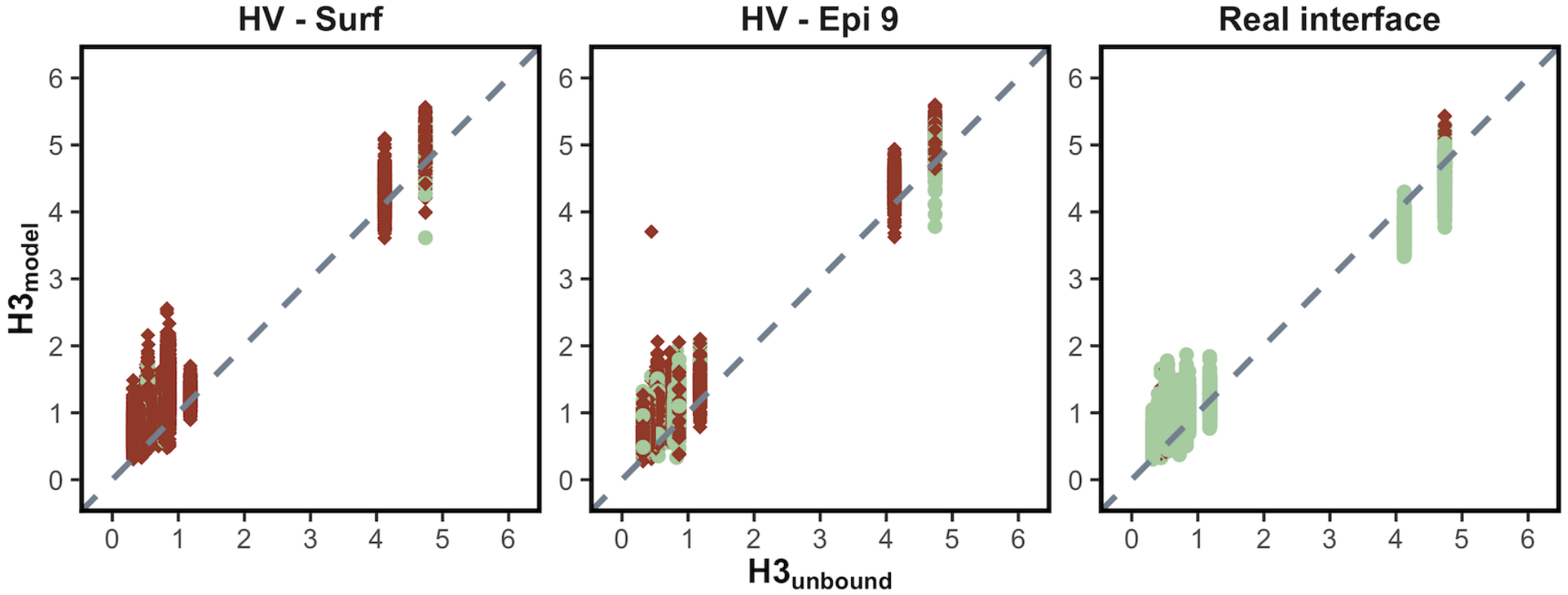
H3 is crucial for the antigen recognition

Modelling of the H3 loop of antibodies is still challenging



Is molecular docking able to correctly model H3?

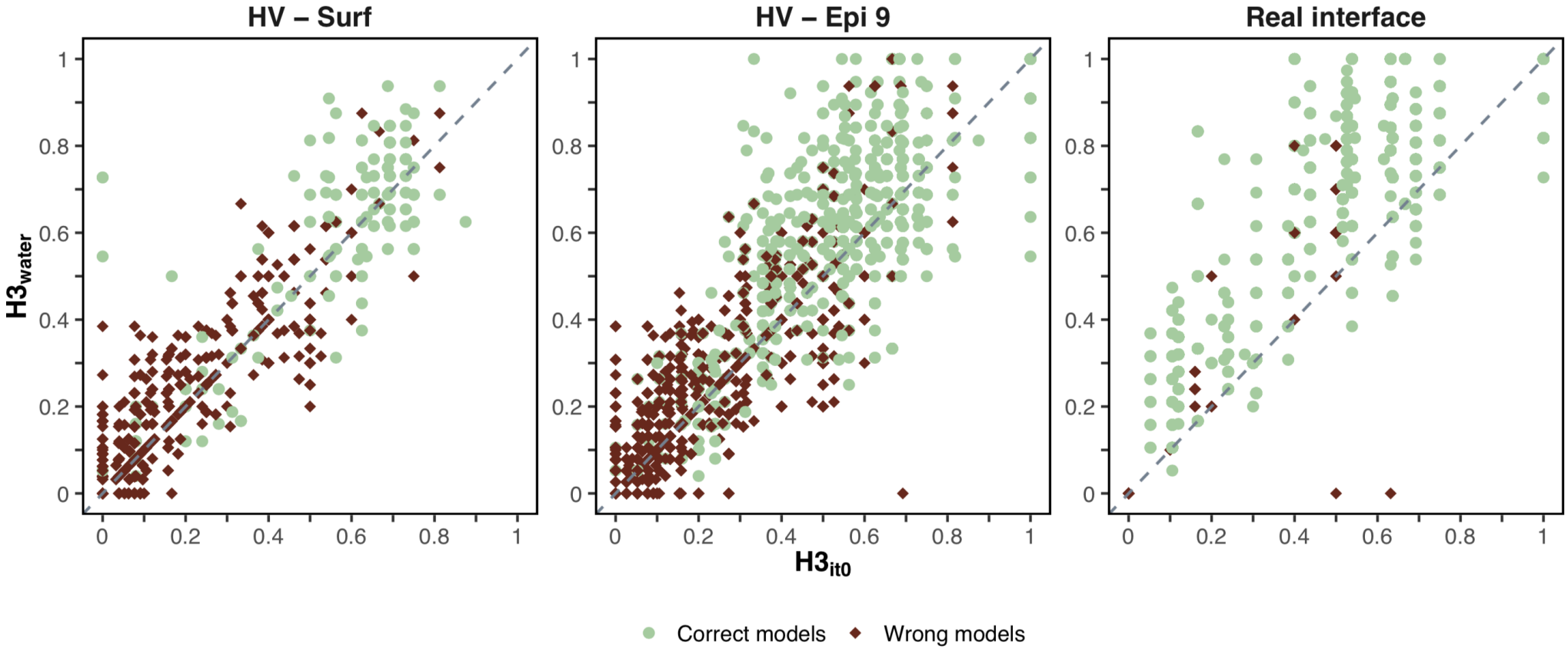
Does flexible docking improves H3?



RMSD [Å] H3 unbound vs complex – HADDOCK models

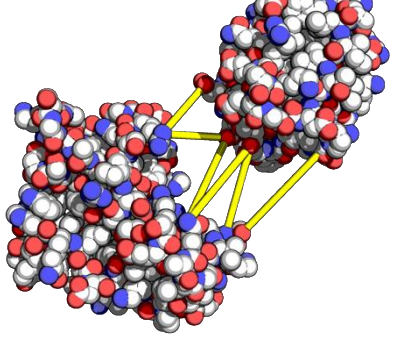
(points below the diagonal indicate improvement)

Does flexible docking improves H3?

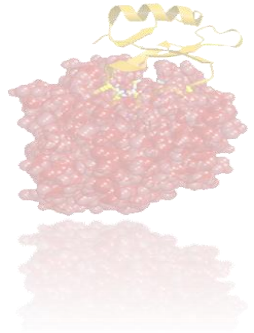


Fnat H3 unbound vs complex – HADDOCK models

(points above the diagonal indicate improvement)

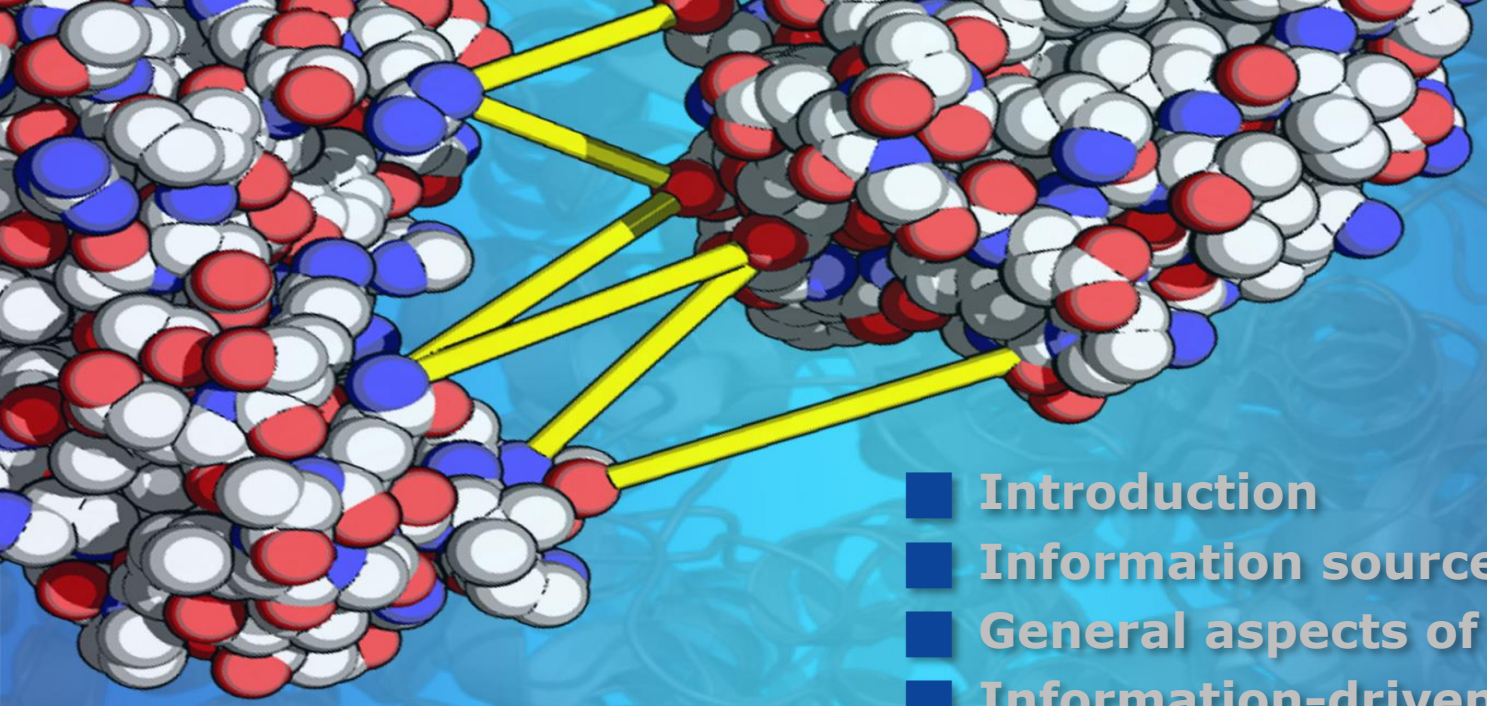


Conclusions - antibodies



- Using information to drive the modelling process improves antibody-antigen modelling as demonstrated by the top performance of HADDOCK.

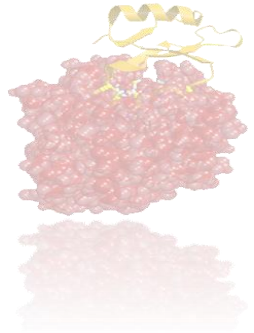
- Accurate modelling of H3 remains challenging, but contacts can be predicted more accurately



Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
- Modelling protein-ligand interactions
- Modelling from cryo-EM data
- Assessing the interaction space
- **Conclusions & perspectives**

Conclusions

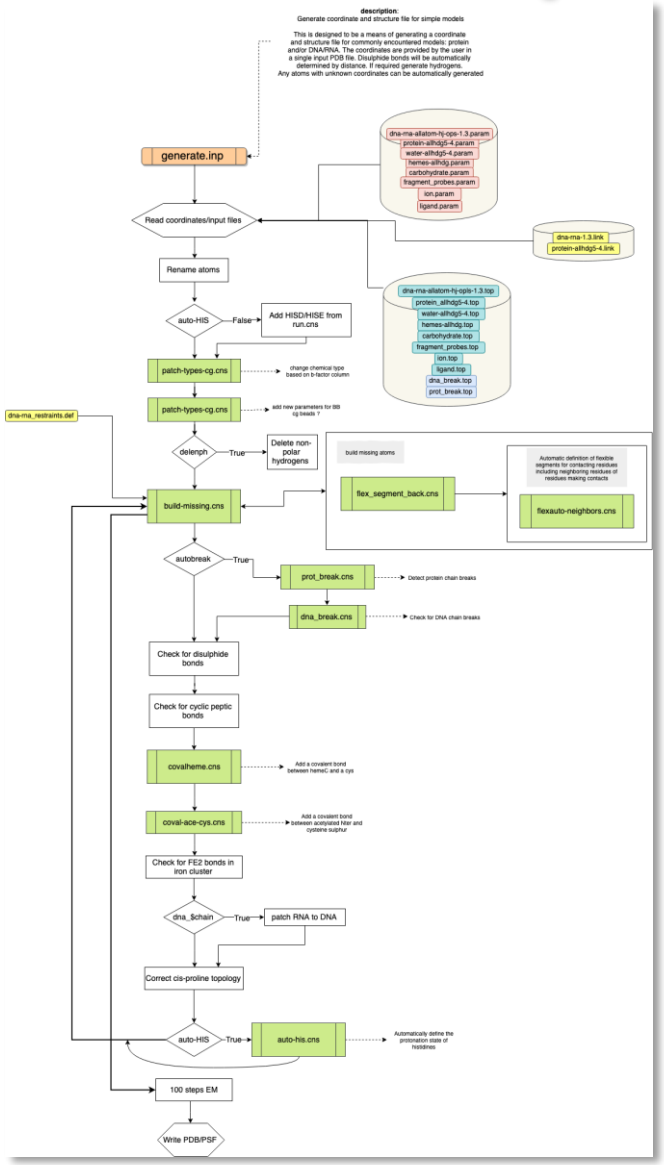


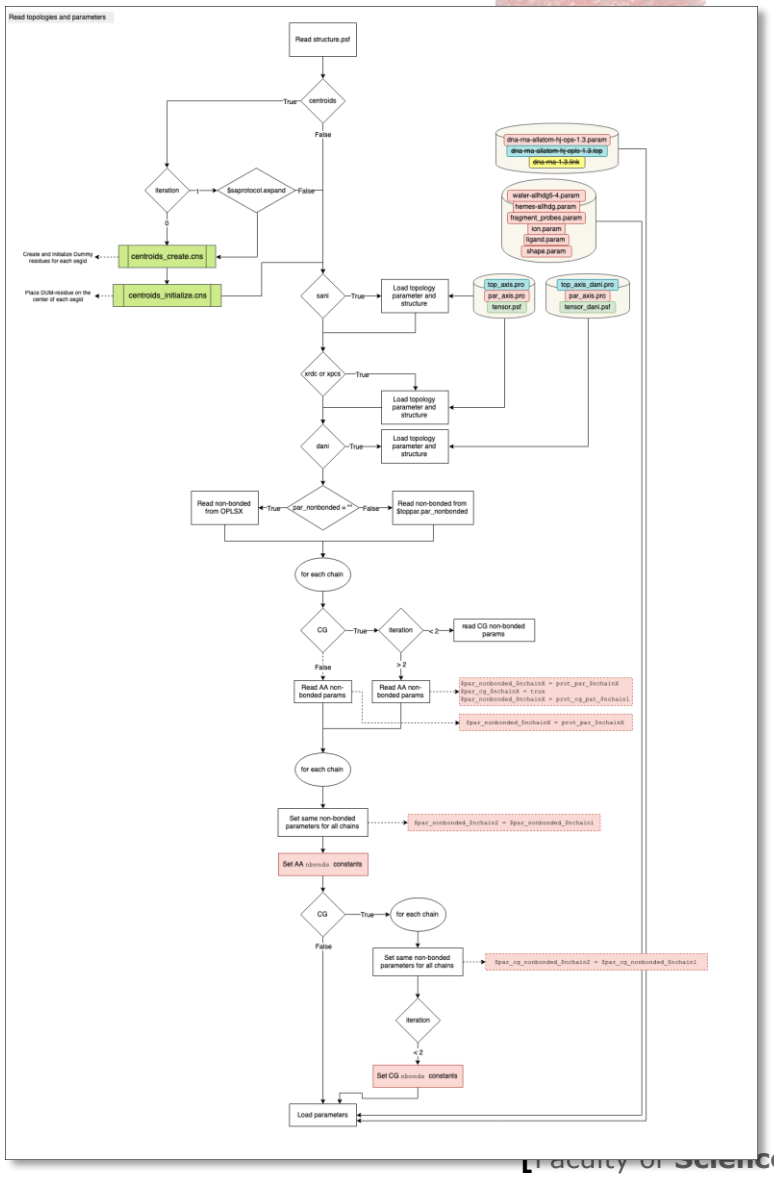
- (Information-driven) docking is useful to generate models of biomolecular complexes, even when little information is available
- **While such models may not be fully accurate, they provide working hypothesis and can still be sufficient to explain and drive the molecular biology behind the system under study**
- … and with a little bit of effort they can be validated!
- Information-driven docking is complementary to classical structural methods

Perspectives – Development

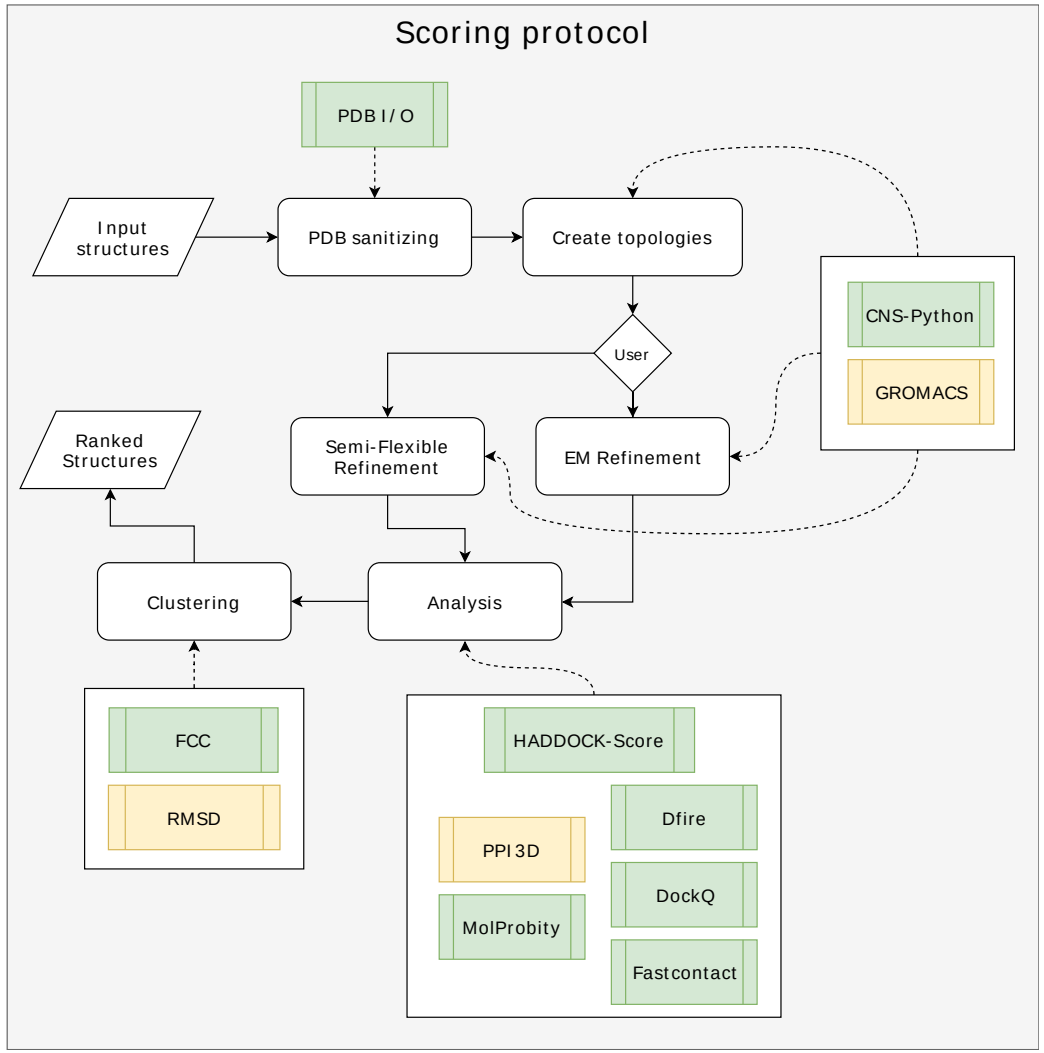
Modularization of HADDOCK

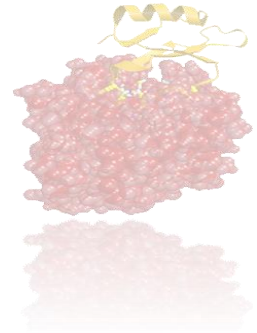
- Creating a map of all internal routines and establishing dependencies
- Thorough code documentation
- Evaluating bottlenecks
- Breaking down CNS routines
- E.g.: pre-processing script for structure check and generation of topology





Modularization of HADDOCK

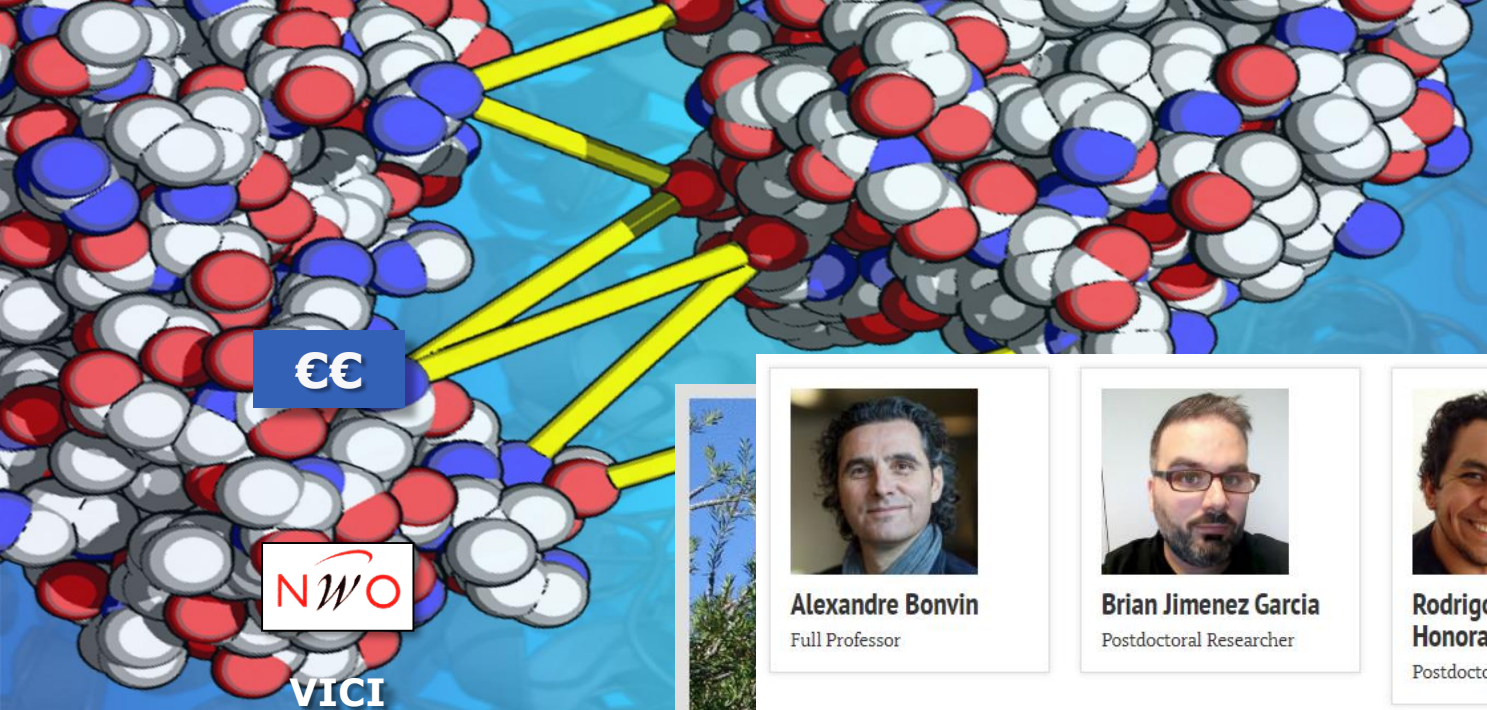




HADDOCK 3
High-Ambiguity Driven Docking

- **Development of a CNS-Python wrapper**
- **Integration of new tools**
- **Testing and optimization of new features**

Acknowledgments:
the CSB group@UU



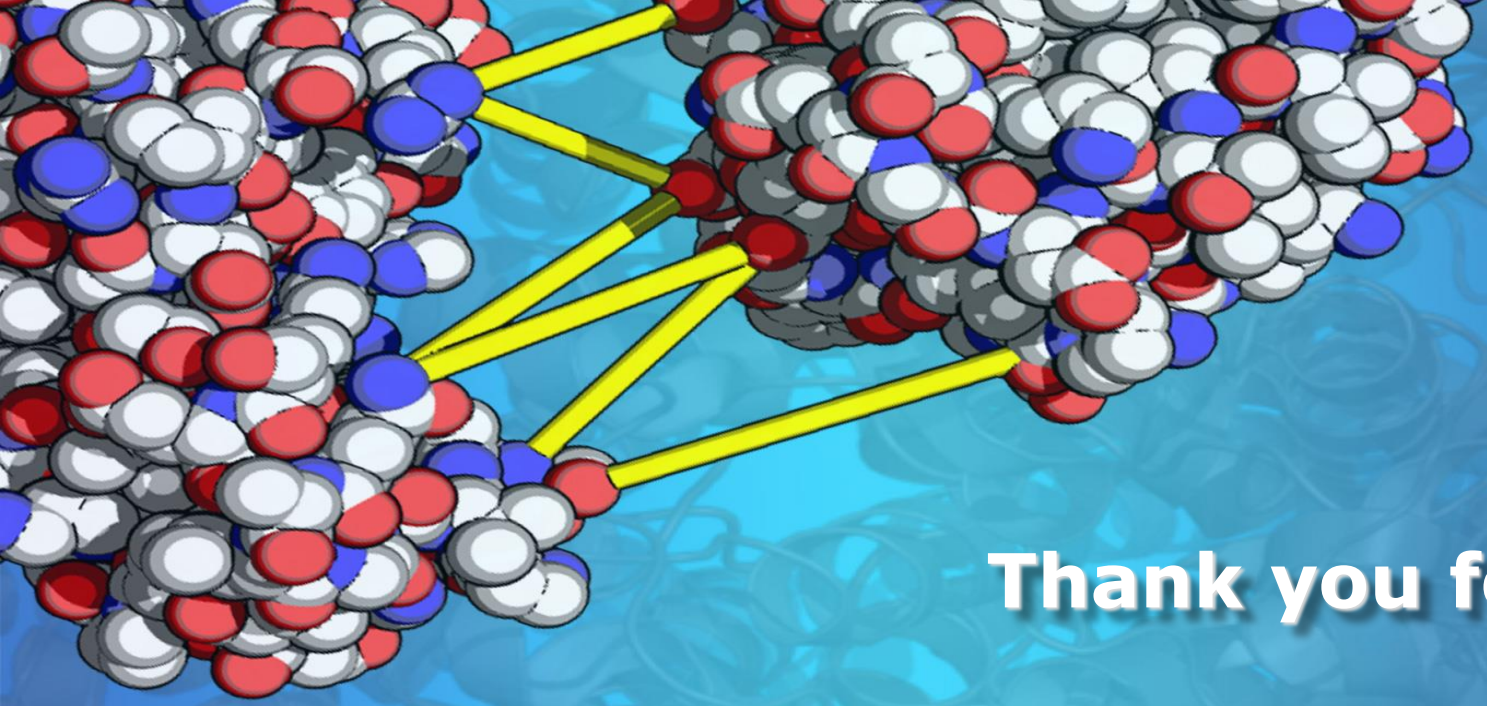
€€

NWO

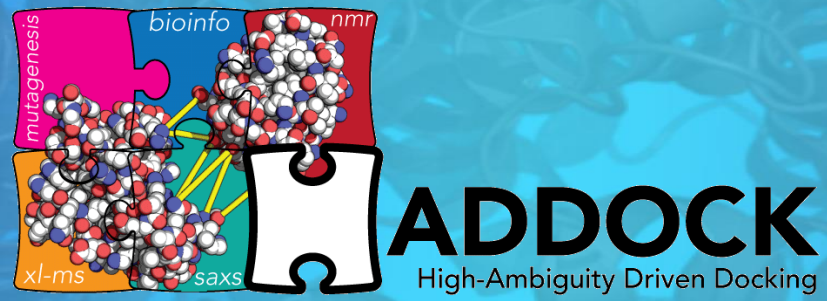
VICI
TOP-PUNT

WeNMR
West-Life
EGI-Engage
INDIGO-Datacloud
BioExcel CoE
EOSC-Hub

bioexcel
Centre of Excellence for Computational Biomolecular Research



Thank you for your attention!



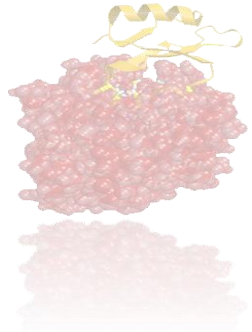
HADDOCK
High-Ambiguity Driven Docking

HADDOCK online:

- **http://haddock.science.uu.nl**
- **http://bonvinlab.org/software**
- **http://ask.bioexcel.eu**

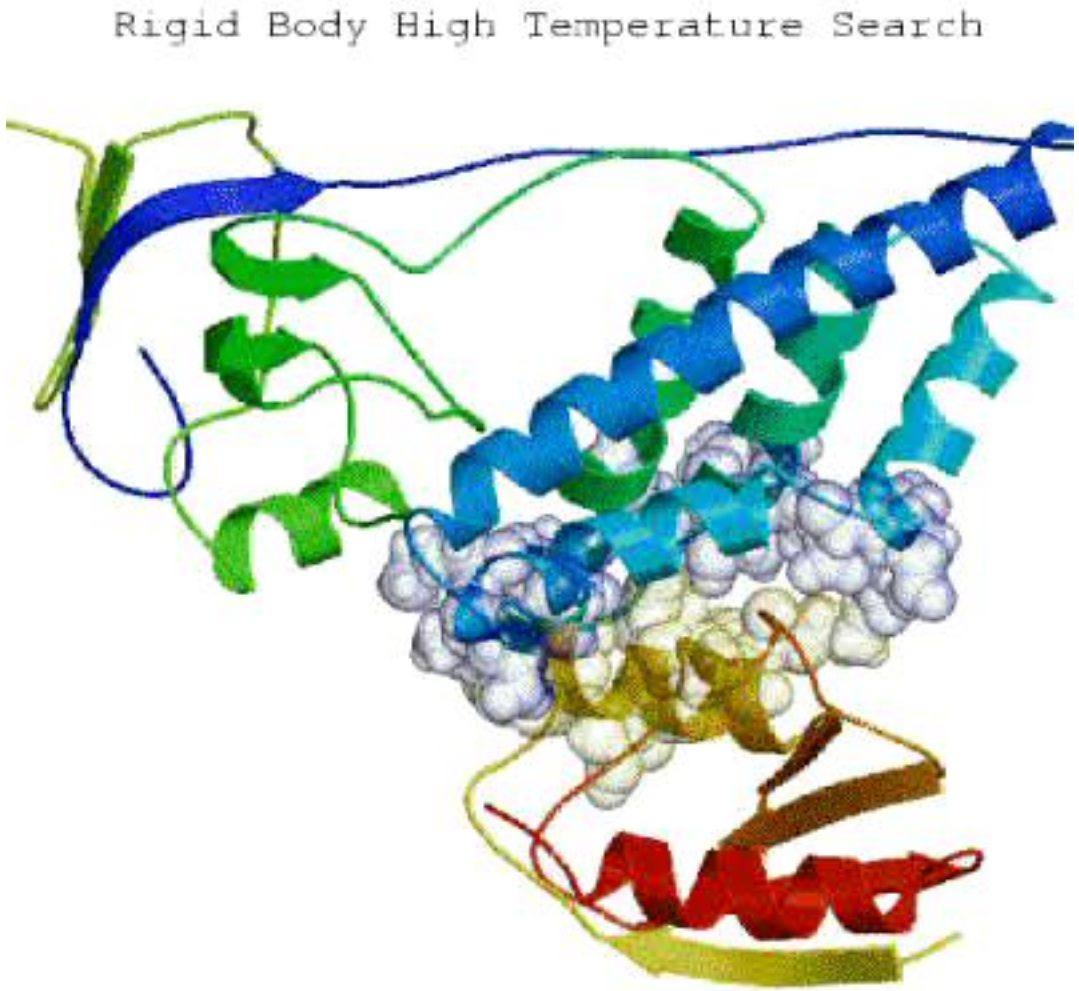
Utrecht Bioinformatics Center

Rigid body energy minimization



step0001->energy=2426.972

Semi-flexible SA refinement in torsion angle space



Refinement in explicit water