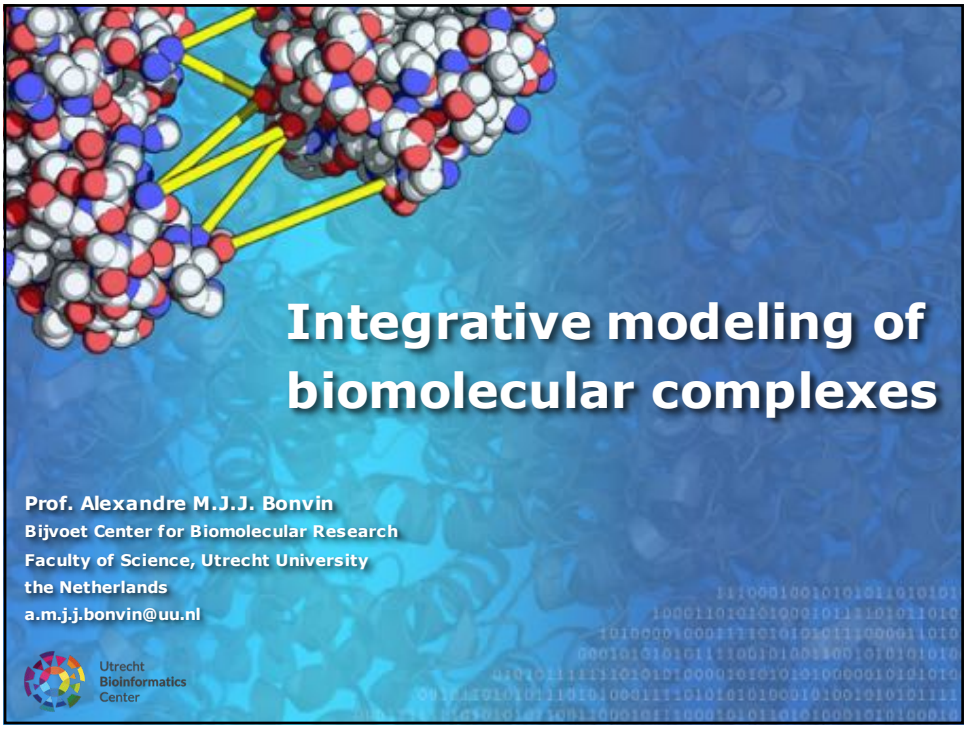
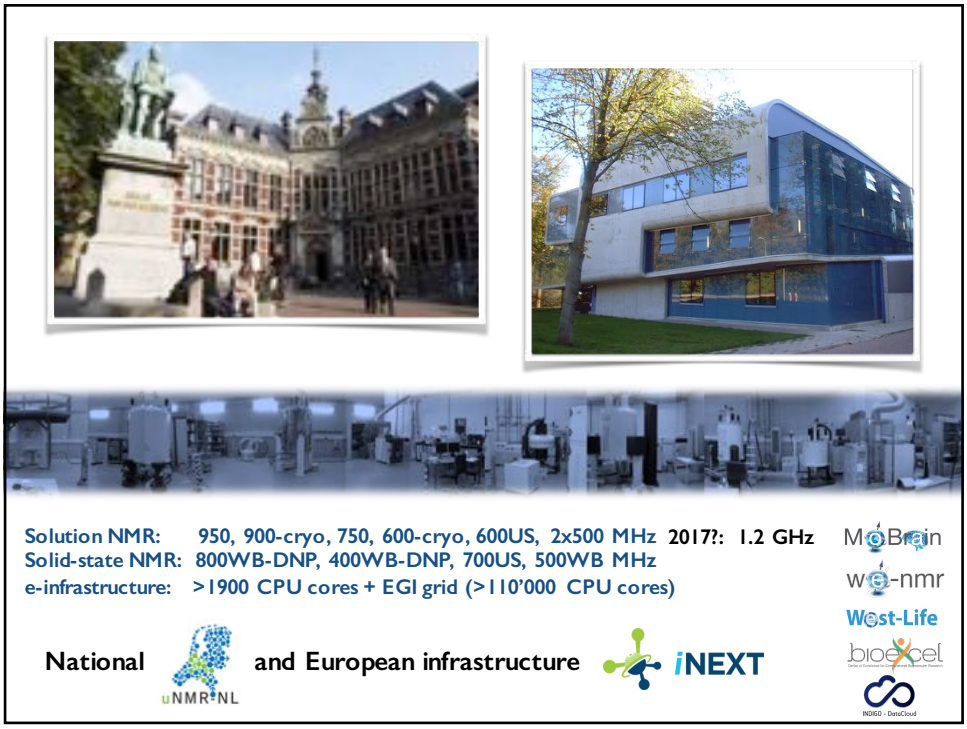
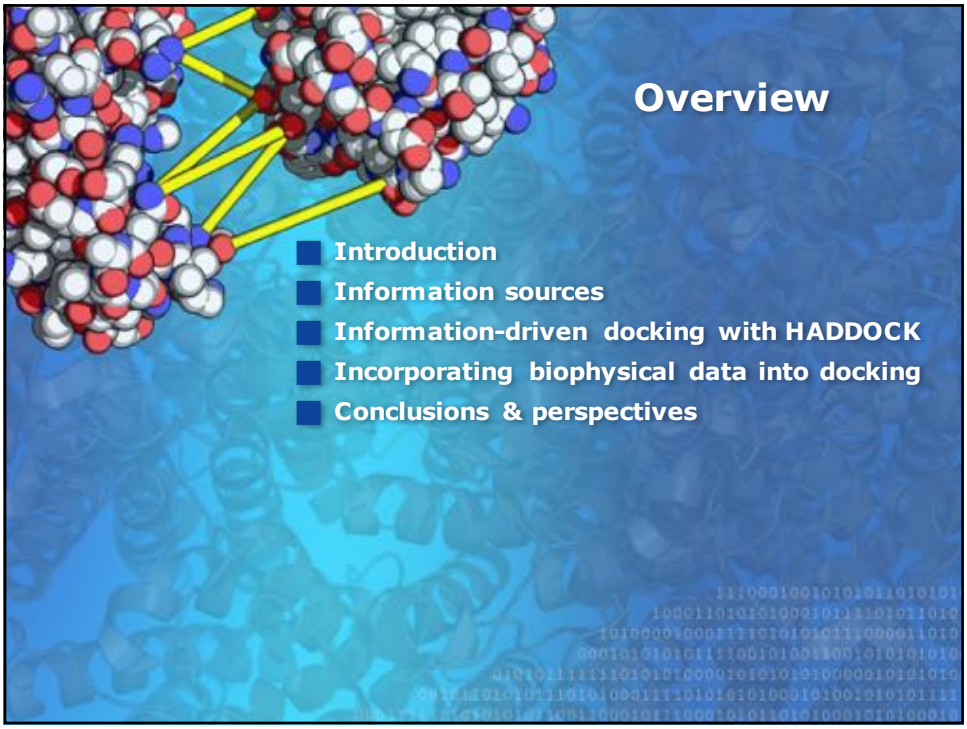
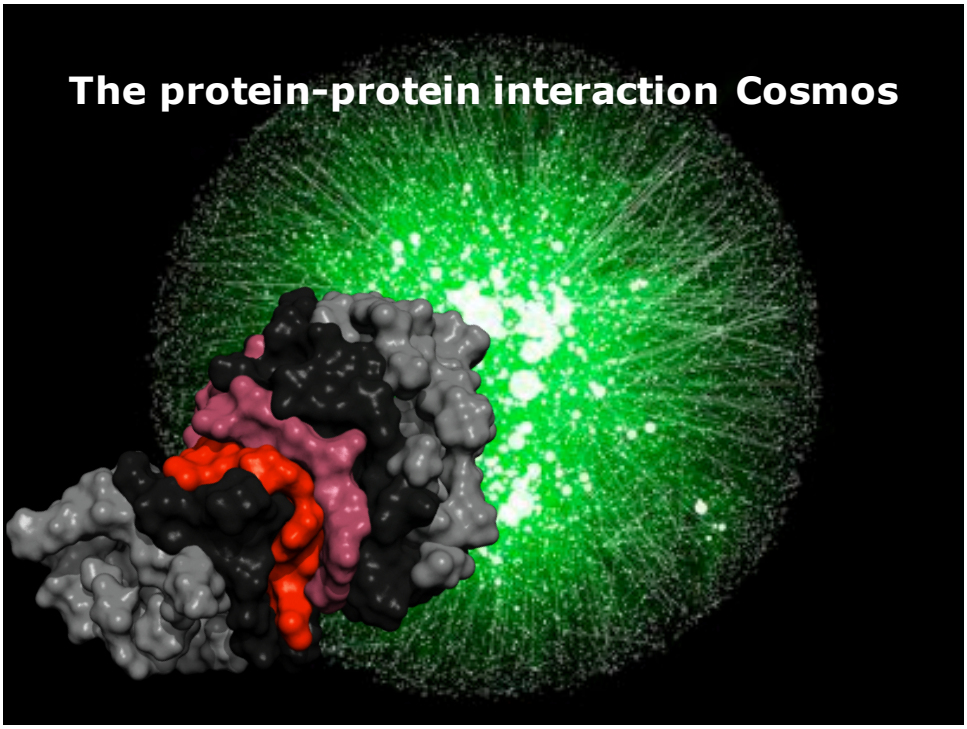

Integrative modeling of biomolecular complexes
Prof. Alexandre M.J.J. Bonvin
Bijvoet Center for Biomolecular Research
Faculty of Science, Utrecht University
the Netherlands
a.m.j.j.bonvin@uu.nl
Utrecht Bioinformatics Center

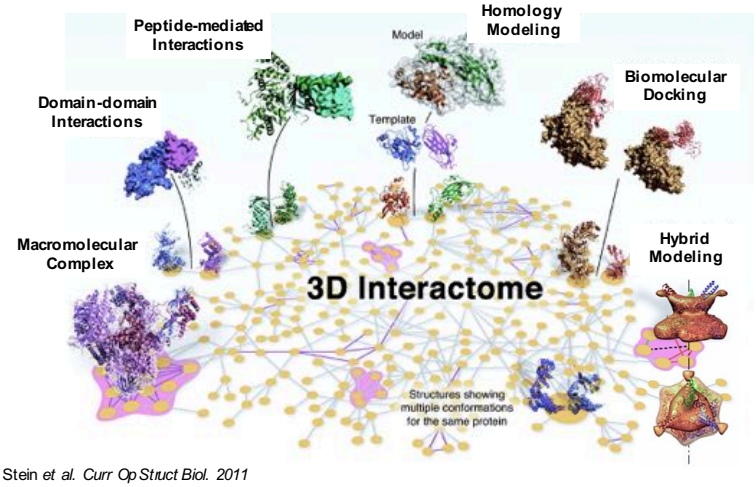

Solution NMR: 950, 900-cryo, 750, 600-cryo, 600US, 2x500 MHz 2017?: 1.2 GHz
Solid-state NMR: 800WB-DNP, 400WB-DNP, 700US, 500WB MHz
e-infrastructure: >1900 CPU cores + EGI grid (>110'000 CPU cores)
National
uNMR-NL
and European infrastructure
iNEXT
MoBrain
w-nmr
West-Life
bioexcel
INDIGO - DataCloud


Overview
Introduction
Information sources
Information-driven docking with HADDOCK
Incorporating biophysical data into docking
Conclusions & perspectives


The protein-protein interaction Cosmos

Adding the 3rd dimension

Experimental Structures | Computational Models

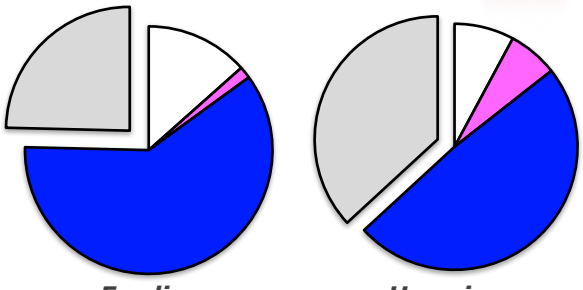


Stein *et al. Curr Op Struct Biol. 2011*

Structural coverage of interactomes

Unique interactions in interactomes

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

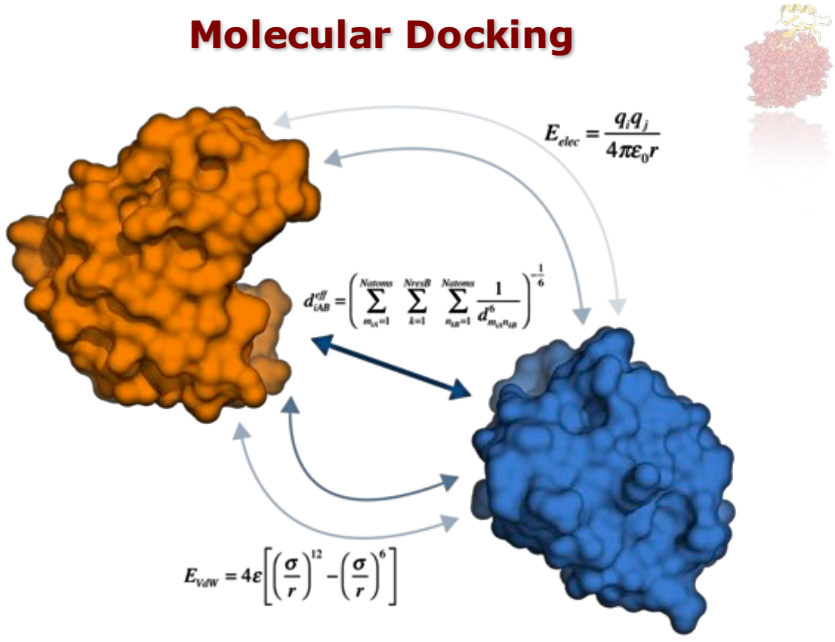


Statistics from Interactome3D (2013-01)

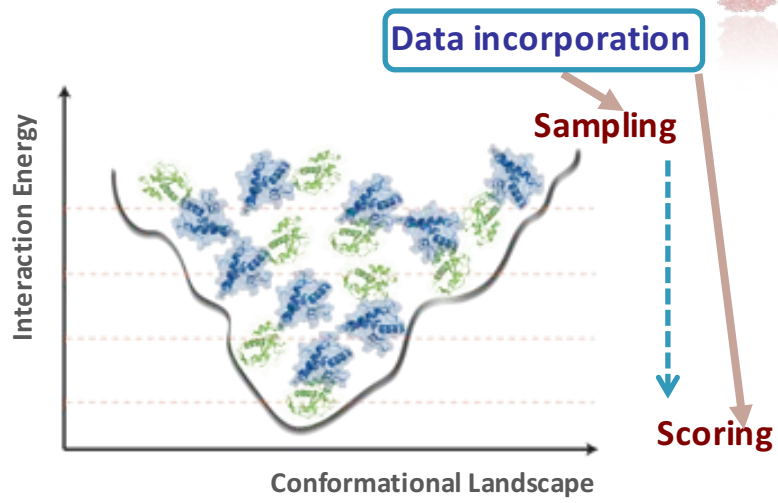
Mosca *et al.* Nature Methods 2013

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

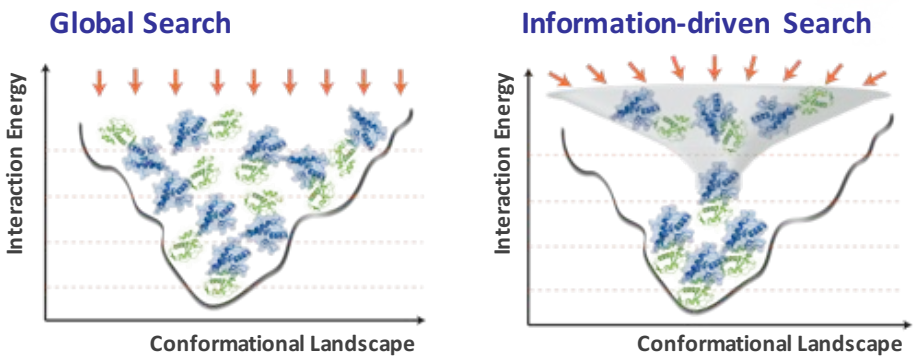
Molecular Docking



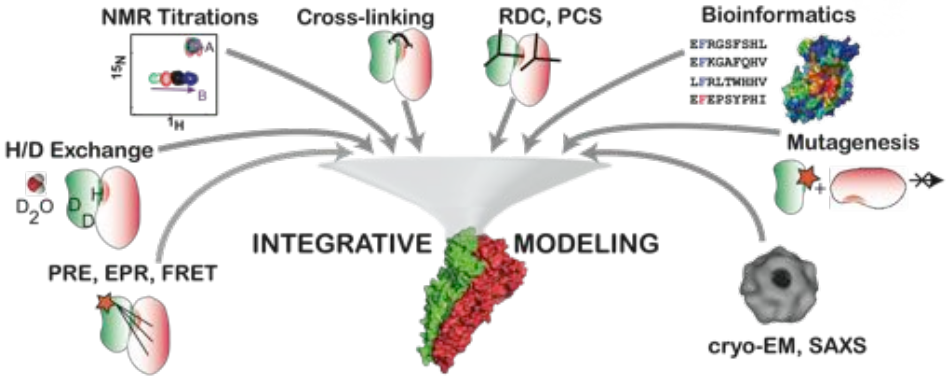
Methodology



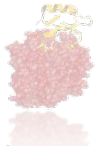
Data Integration during Sampling



What is Integrative Modeling?



Why integrative modelling?



For Experimentalists

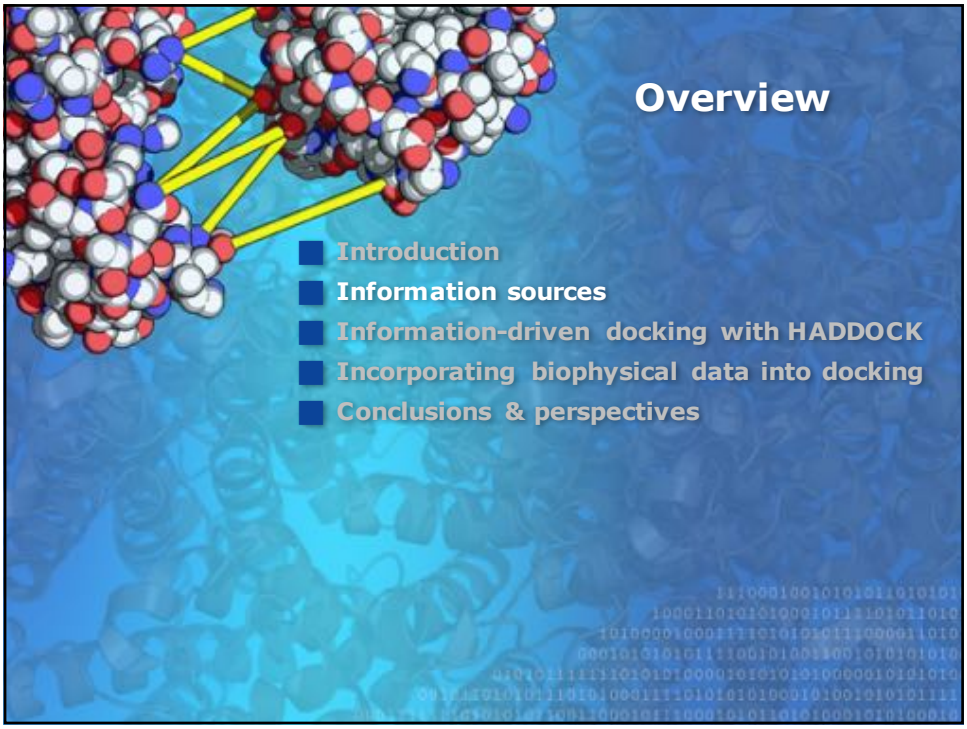
- ✓ New hypothesis to drive experiments
- ✓ Speed up structure determination
- ✓ Increase our understanding of function

For Modelers

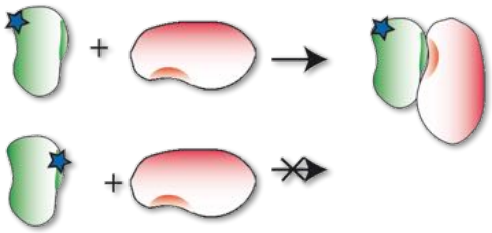
- ✓ Decrease high false positive rate
- ✓ Ease accuracy assessment

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) and (2013), 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).



Experimental sources: mutagenesis



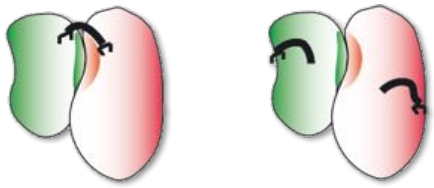
Advantages/disadvantages

+ 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

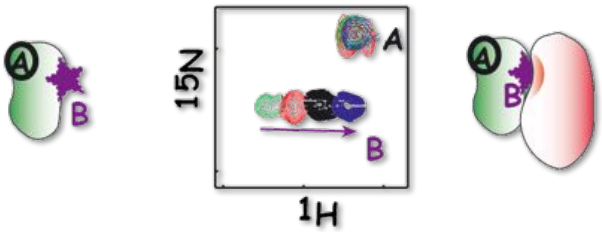
Advantages/disadvantages

+ 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

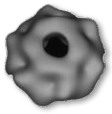
+ Residue/atomic level
+ No need for assignment if combined with a.a. selective labeling
- Direct vs indirect effects
- Labeling needed

Detection

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

Other potential experimental sources

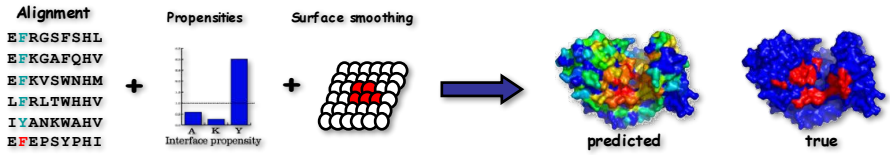
- **Paramagnetic probes** in combination with NMR
- **Cryo-electron microscopy** or tomography and **small angle X-ray scattering** (SAXS) ==> shape information



- 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

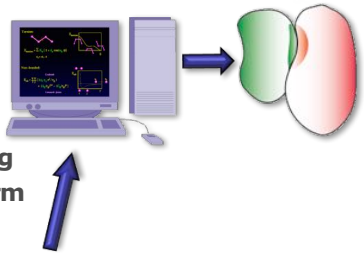
- PPISP (Zhou & Shan,2001; Chen & Zhou, 2005)
  http://pipe.scs.fsu.edu/ppisp.html
- ProMate (Neuvirth et al., 2004)
  http://bioportal.weizmann.ac.il/promate
- WHISCY (De Vries et al., 2005)
  http://www.nmr.chem.uu.nl/whiscy
- PINUP (Liang et al., 2006)
  http://sparks.informatics.iupui.edu/PINUP
-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

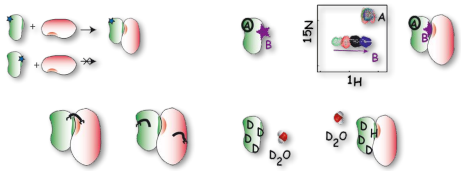
Consensus interface prediction (CPORT)
haddock.science.uu.nl/services/CPORT

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**





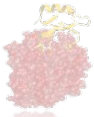
Universiteit Utrecht

[Faculty of **Science** **Chemistry**]

Overview

- Introduction
- Information sources
- **Information-driven docking with HADDOCK**
- Incorporating biophysical data into docking
- Conclusions & perspectives

HADDOCK:
An integrative modeling platform



Incorporates ambiguous and low-resolution data to aid the docking

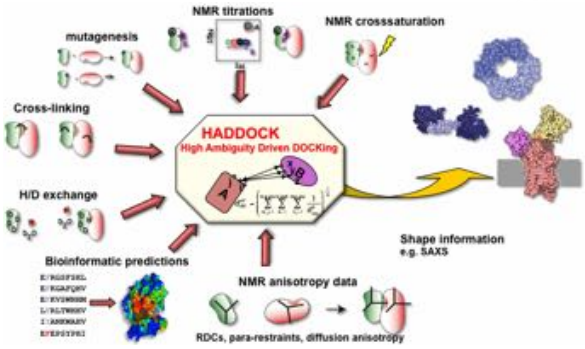
Capable of docking up to 6 molecules

Symmetries can be leveraged

Powerful algorithms to handle flexibility at the interface

Final flexible refinement in explicit solvent

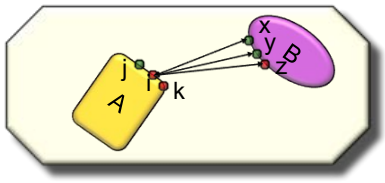
One of the best performing software in CAPRI



Data-driven docking with HADDOCK

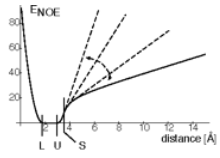
List of interface residues for protein A

List of interface residues for protein B



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

$$d_{iAB}^{eff} = \left(\sum_{m_A=1}^{Nstrms} \sum_{k=1}^{NresB} \sum_{n_k=1}^{Nstrms} \frac{1}{d_{mn_k}^6} \right)^{\frac{1}{6}}$$



Ambiguous Interaction Restraint:
a residue must make contact with any residue from the other list

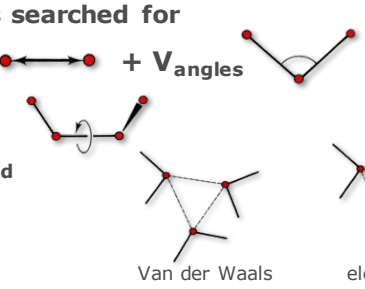
Different fraction of restraints (typically 50%) randomly deleted for each docking trial to deal with inaccuracies and errors in the information used

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

(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}

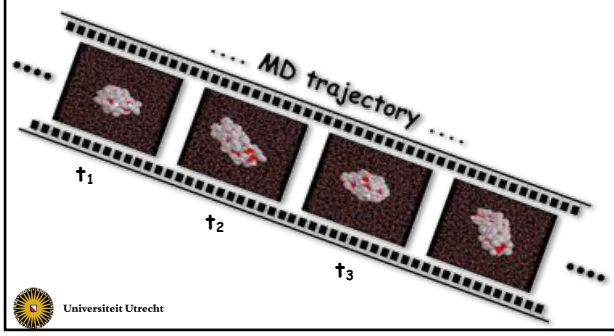


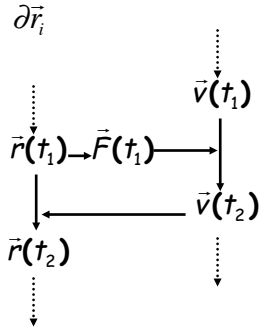
- Search is performed by a combination of gradient driven energy minimization and molecular dynamics simulations

Classical mechanics

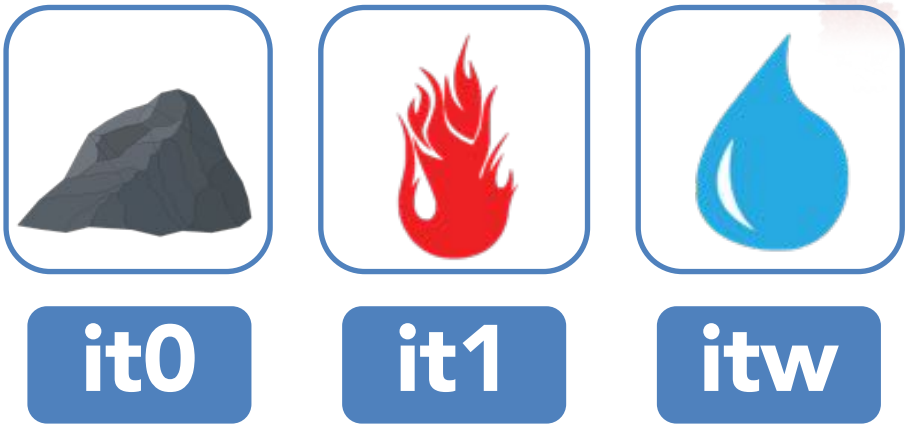
- Molecular dynamics: generates successive configurations of the system by integrating **Newton's second law**

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





HADDOCK docking protocol



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)

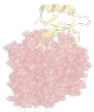
Typically, 10.000 conformations are sampled but only the best 1.000 are written to disk.

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.

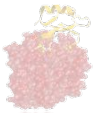
Flexibility reproduces conformation changes up to 2Å, typical of small induced fit.

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



itw

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.

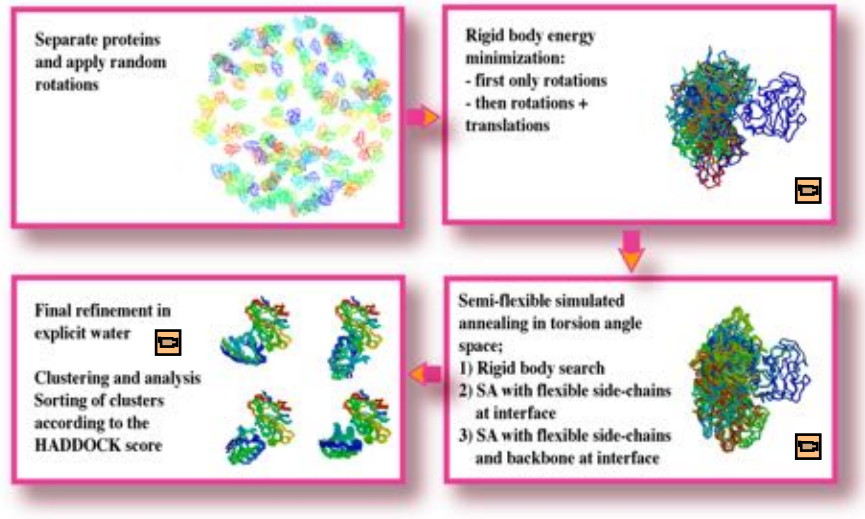
Explicit solvent models include TIP3P water and DMSO (membrane mimic).

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

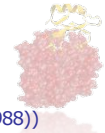
HADDOCK docking protocol



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



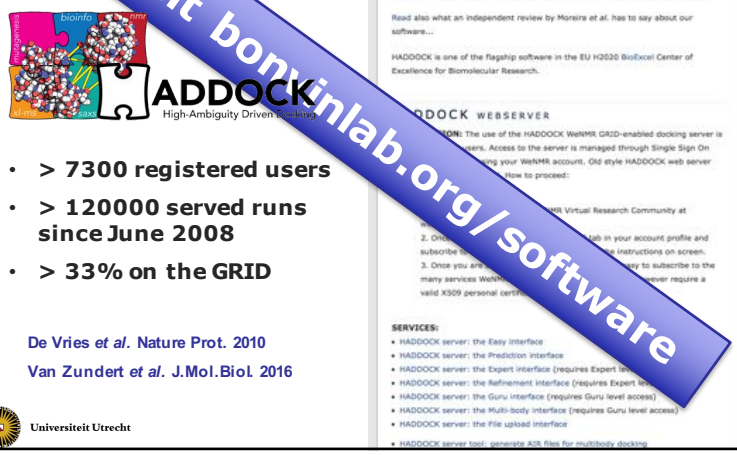
- OPLS non-bonded parameters (Jorgensen, *JACS* 110, 1657 (1988))
- 8.5Å non-bonded cutoff, switching function, ε=10
- Clustering of solutions

- Ranking of 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

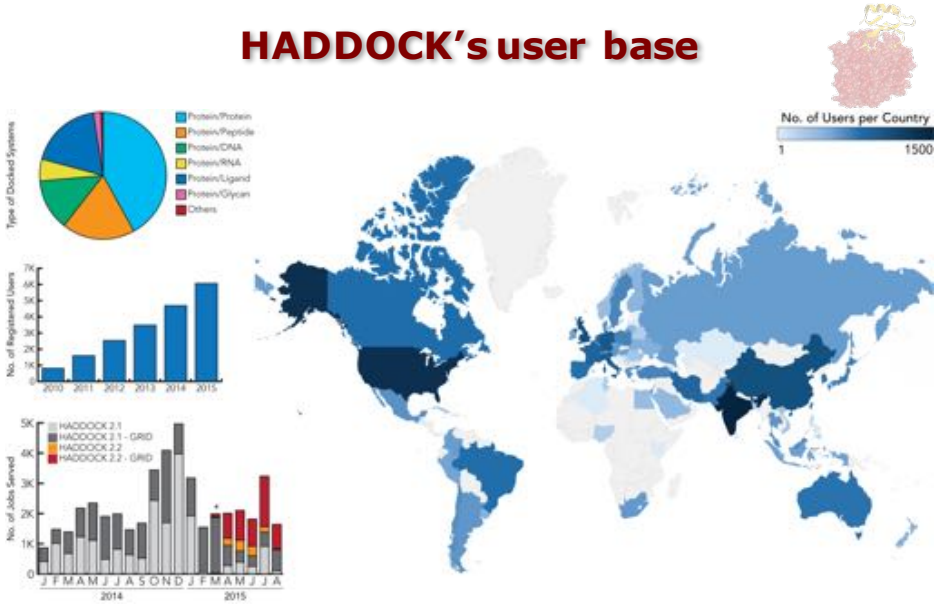


- > 7300 registered users
- > 120000 served runs since June 2008
- > 33% on the GRID

De Vries *et al.* Nature Prot. 2010

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

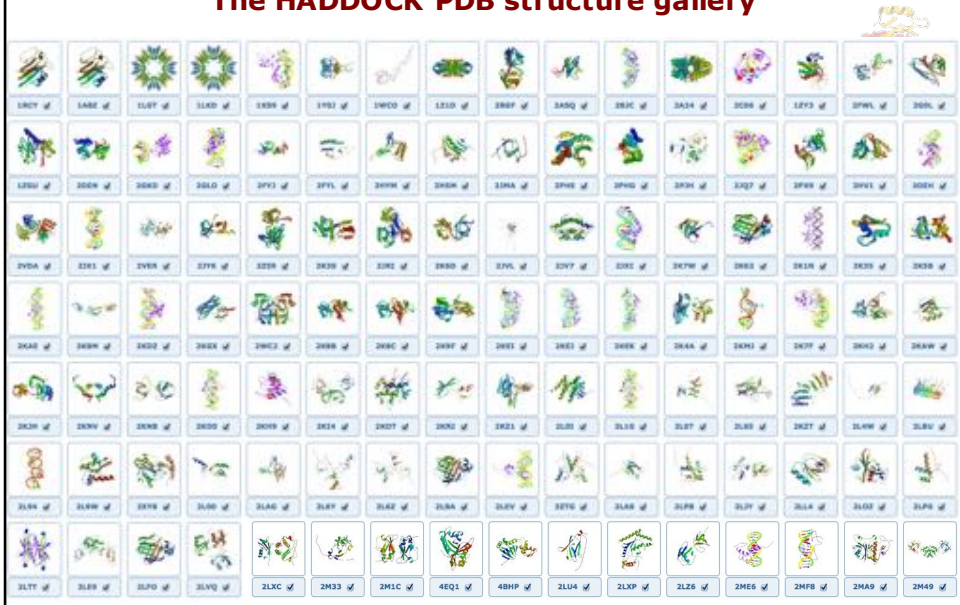
HADDOCK's user base



Universiteit Utrecht

[Faculty of **Science** **Chemistry**]

The HADDOCK PDB structure gallery



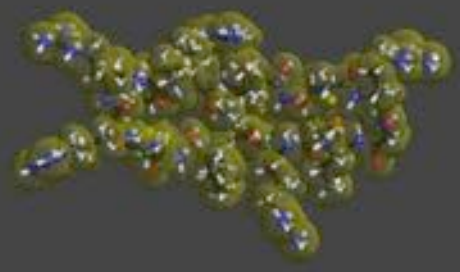
>120 entries – Jan. 2015

Image collage from http://www.pdb.org

Universiteit Utrecht

[Faculty of **Science** **Chemistry**]

Fully flexible small molecule docking

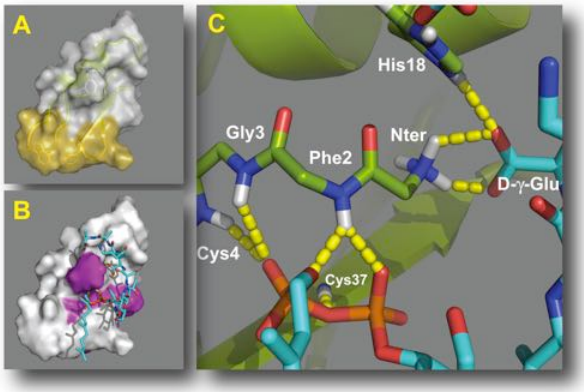


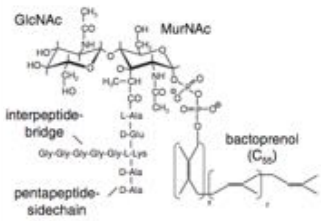
Retrocyclin-lactose docking with HADDOCK, ABz2004

Can deal with complex molecules

Science
AAAS

Plectasin, a Fungal Defensin, Targets the Bacterial Cell Wall Precursor Lipid II
Tanja Schneider, *et al.*
Science **328**, 1168 (2010);
DOI: 10.1126/science.1185723





Universiteit Utrecht

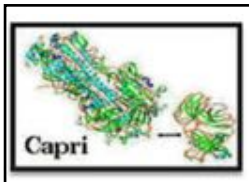
[Faculty of **Science** **Chemistry**]

HADDOCK's adventures in CASP-CAPRI

"Critical assessment of predicted interactions"

http://capri.ebi.ac.uk

- We only participated in the CASP-CAPRI round as server and scorers
  - Number of targets
  - Lack of information
  - Questionable biological relevance
- Two strategies:
  - Ab-initio docking with symmetry restraints
  - Template-based modelling



CAPRI Round 30
CASP-11 (2014)

Assessment of predicted protein complexes

Marc F. Lensink
CNRS, University Lille North of France, France, CAPRI-MC
Sameer Velankar
PDBe European Bioinformatics Institute (EBI) UK, CAPRI-MC
Andriy Kryshtafovych
UC Davis, US, CASP
Shoshana J. Wodak
VIB-VUB, Brussels Belgium & Hospital for Sick Childre..,
University of Toronto, Canada, CAPRI-MC
Shoshana.wodak@gmail.com

CASP-CAPRI Session, CASP-11 meeting 2014

CAPRI Predictor Ranking	
Guerois	17
Huang	16
Seok	15
Zou	14
Weng, Vajda/ Kozakov, Shen	13
Vakser, Grudinin, Fernandez-Recio	11
Lee	10
Tomii, Bates	8
Kihara	7
Sali	6
Negi	5
Zhou	4
Tovchigrechko, Eisenstein	3
Ritchie	2
Xiao, Gray, Fernandez-Fuentes	1

CAPRI Server Ranking	
HADDOCK	15
CLUSPRO	14
SWARMDOCK	11
GRAMM-X	6
LZERD	3
DOCK/PIERR	1

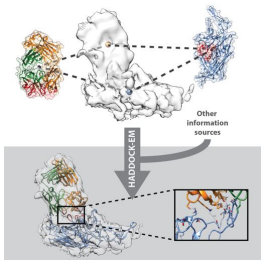
CAPRI Scorer Ranking	
Bonvin	19
Seok, Huang, Bates	17
Zou, Weng	16
Kihara	15
Fernandez-Recio	14
Oliva, Grudinin	13
Gray	10
LZERD	6
Lee	3

CASP Predictor Ranking	
Seok	15
Umeyama	13
Tomii, Dunbrack	8
Nakamura	7
Luethy	5
Baker	3
Wallner, Skwark	1

CASP Server Ranking	
ROSETTASERVER	9
SEOK_SERVER	7
RAPTOX-X_Wang, NNS_Lee	1

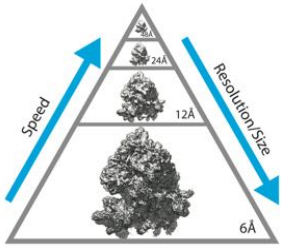
*Ranking by # of **INTERFACES** for which at least one acceptable solution was obtained. In total 42 interfaces were assessed for 25 targets.*

Recent developments



HADDOCK is now capable of using **cryo-EM** data in combination with all other supported sources of information

Van Zundert & Bonvin, Structure 2015

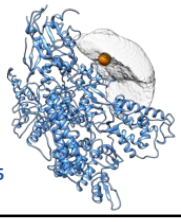


PowerFit: fast and sensitive rigid body fitting in lower-resolution densities

Van Zundert & Bonvin,J. Proteome. Res. 2015

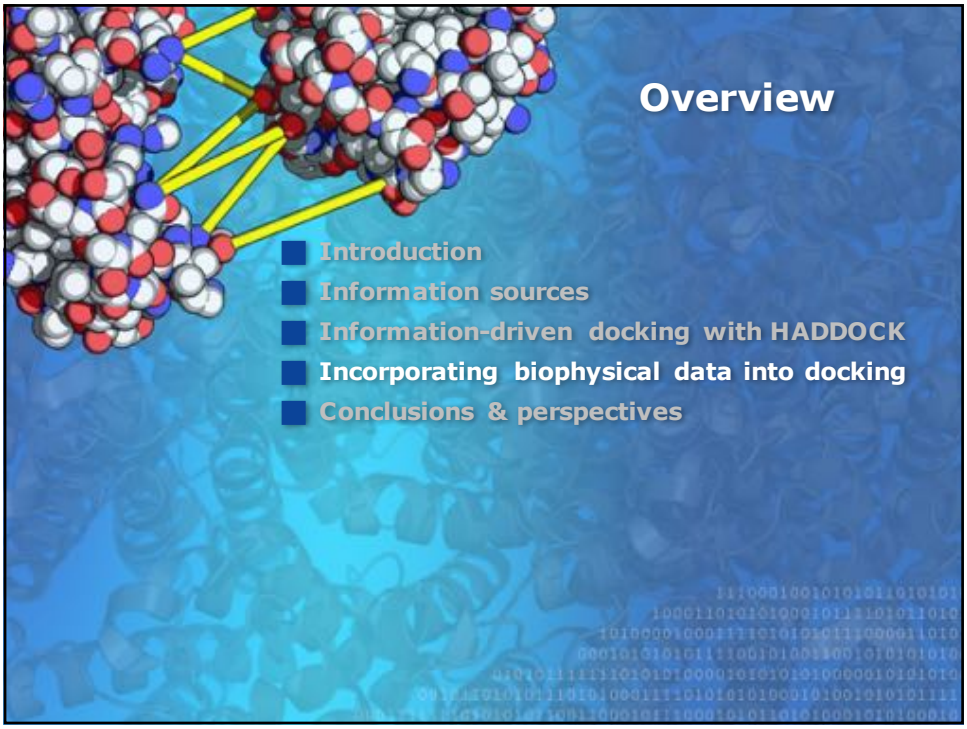
DisVis: quantification and visualization of the information content of distance restraints

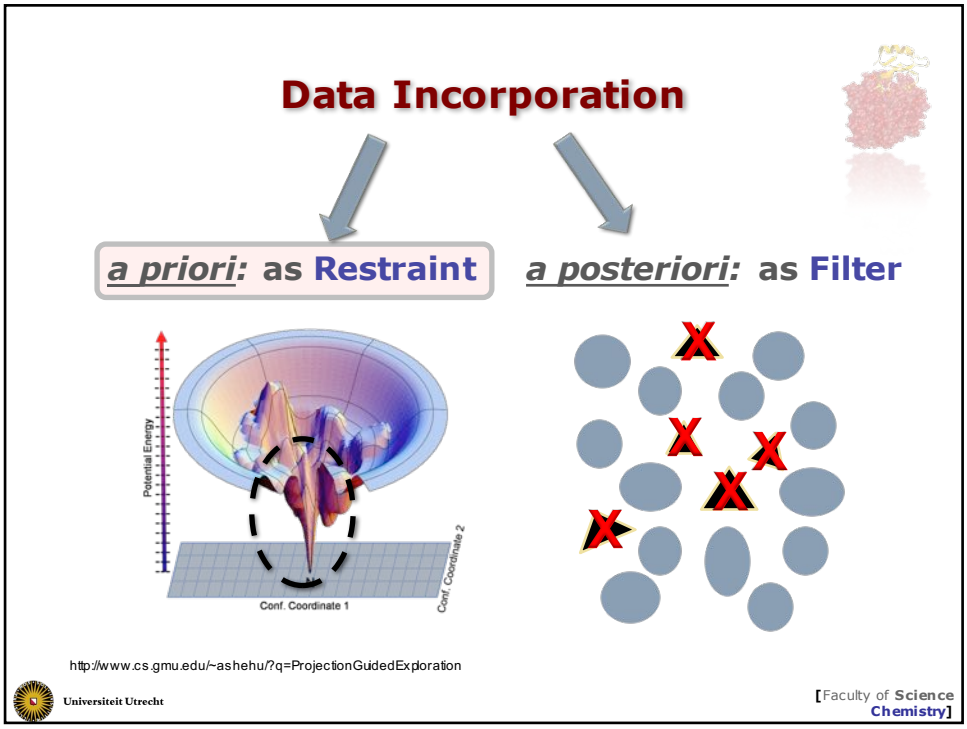
Van Zundert & Bonvin,Bioinformatics 2015



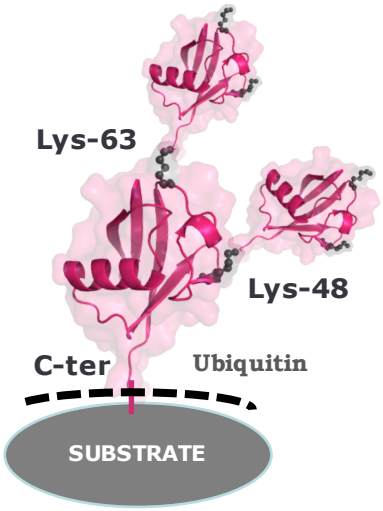
Universiteit Utrecht

[Faculty of **Science** **Chemistry**]


Overview
Introduction
Information sources
Information-driven docking with HADDOCK
Incorporating biophysical data into docking
Conclusions & perspectives


Data Incorporation
a priori: as Restraint
a posteriori: as Filter
Potential Energy
Conf. Coordinate 1
Conf. Coordinate 2
http://www.cs.gmu.edu/~ashehu/?q=ProjectionGuidedExploration
Universiteit Utrecht
[Faculty of Science
Chemistry]

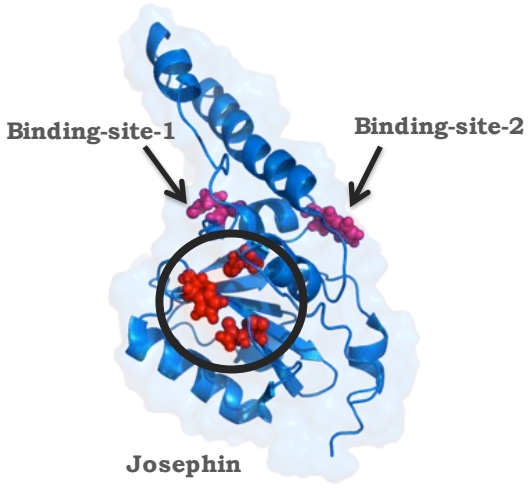
NMR Example: CSP-driven docking



- Ub-cleaving enzyme
  - Josephin
- Which di-Ub linkage type is cleaved, K48 and/or K63 linkage?
- Collaboration with Annalisa Pastore (London, MRC)

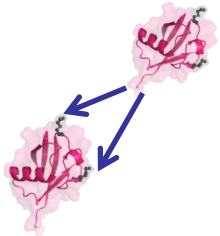
Nicastro et al., *Plos One*, 2010

NMR Example: CSP-driven docking



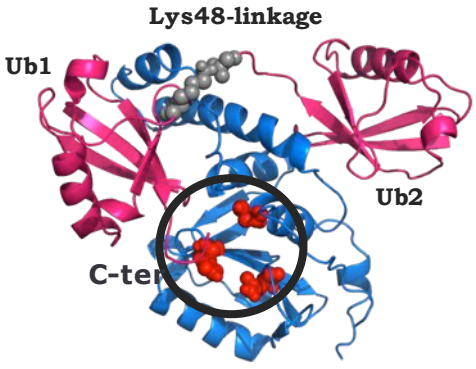
Input for docking:

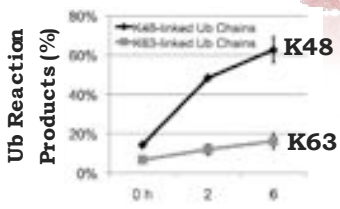
- Catalytic Triad
- 2 Binding-sites
  - CSP + Mutation
- FMD Protocol

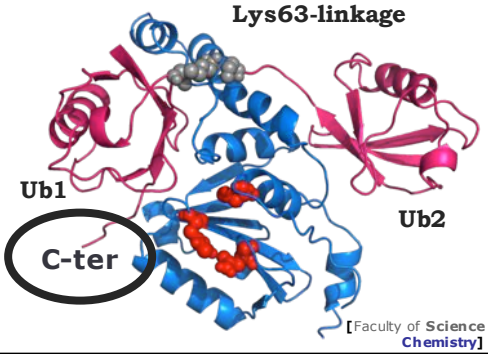


Nicastro et al., *Plos One*, 2010

NMR Example: CSP-driven docking







Nicastro et al., *Plos One*, 2010

Universiteit Utrecht

[Faculty of Science Chemistry]

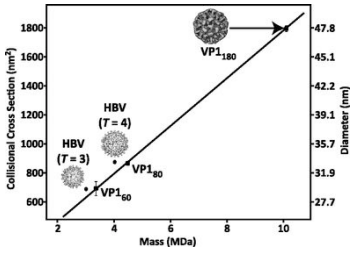
Overview

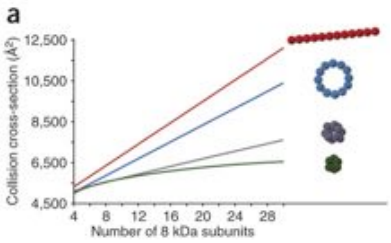
- Introduction
- Information sources
- Information-driven docking with HADDOCK
- Incorporating biophysical data into docking
  - MS CCS as filters in docking
- Conclusions & perspectives

Integration of shape information

Ion Mobility Mass Spectrometry

- **Collision Cross Section (CCS):** rotationally averaged shape adopted by a given molecular ion under particular gas phase conditions





Ruotolo et al.,
Nature Protocols, 2008

Insight into cyanobacterial circadian timing: the KaiB-KaiC interaction

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

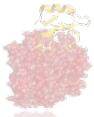
Interactions define the phosphorylation status of KaiC and control the phase of the cycle

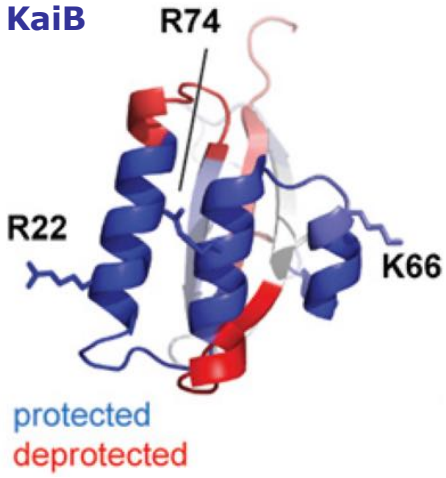
Information from MS:

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

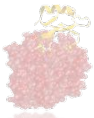
Snijder et al. PNAS 111, 1379 (2014)

The KaiB-KaiC interaction: HDX

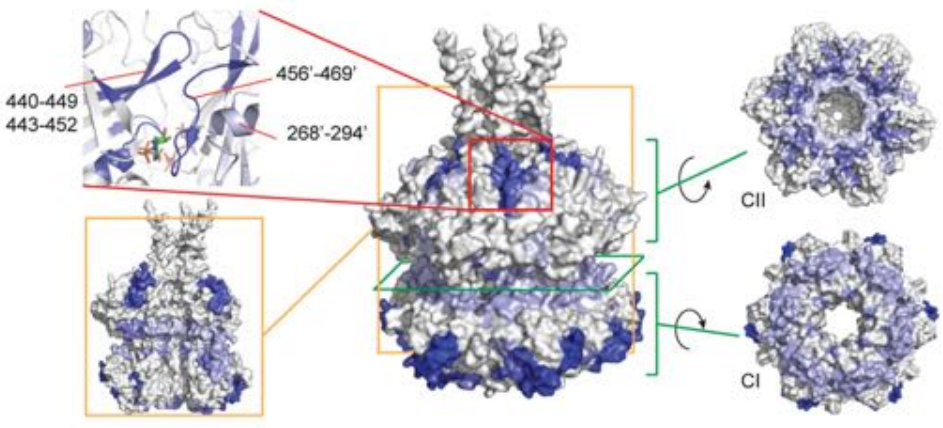




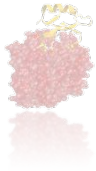
The KaiB-KaiC interaction: HDX



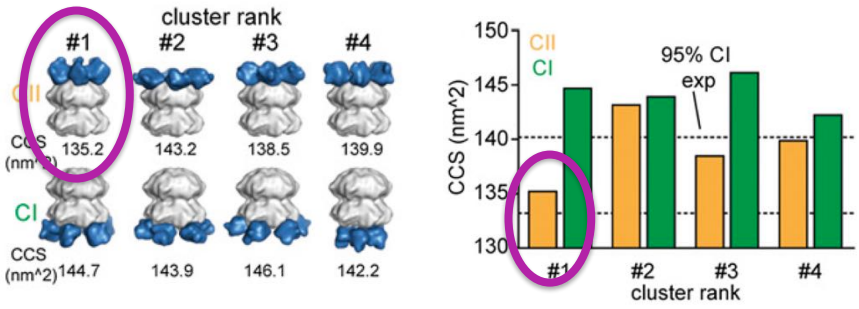
KaiC



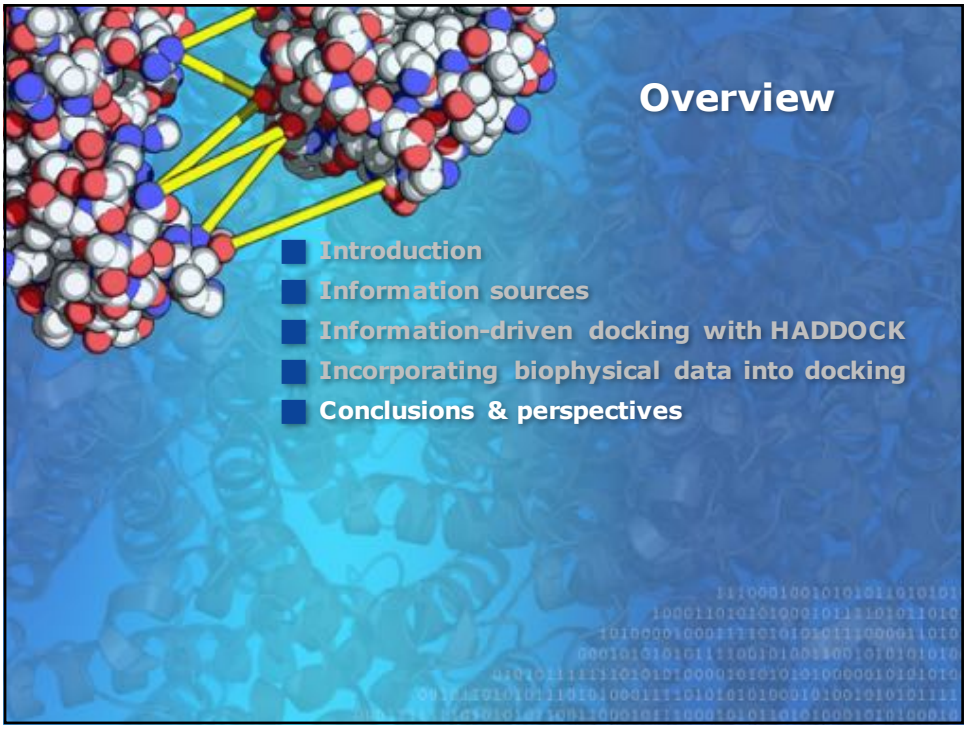
The KaiB-KaiC interaction: CCS



Collision cross section from MS allows to filter the HADDOCKing solutions



HADDOCK best scoring/most populated solution of CII

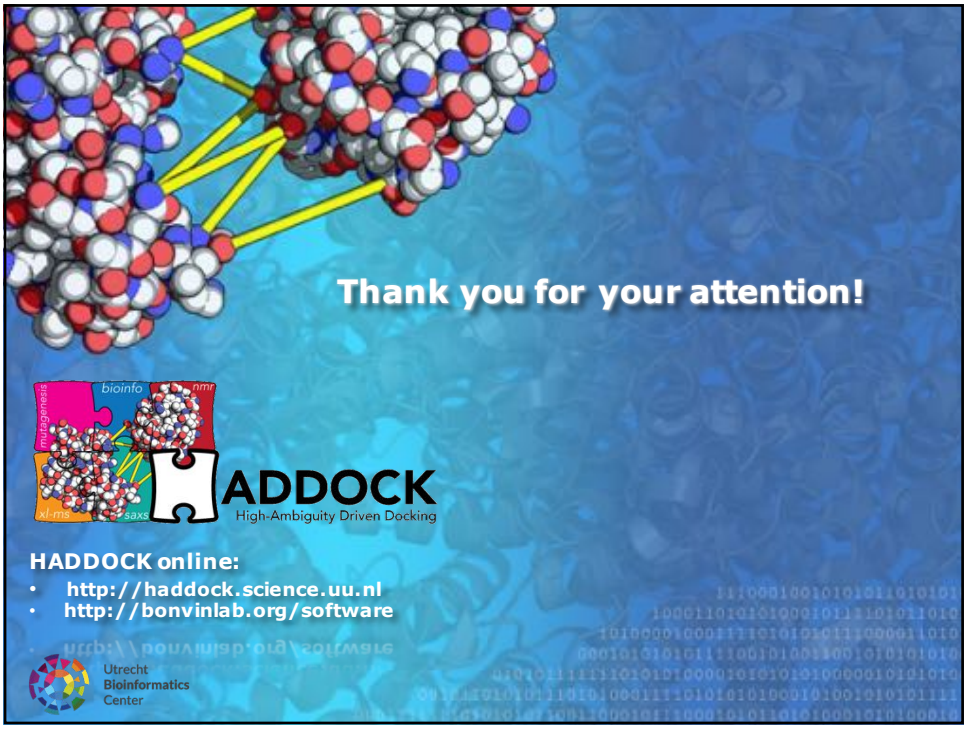


Conclusions

- (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

Universiteit Utrecht

[Faculty of Science
Chemistry]


Thank you for your attention!
bioinfo
nmr
mutagenesis
xl-ms
saxs
HADDOCK
High-Ambiguity Driven Docking
HADDOCK online:
• http://haddock.science.uu.nl
• http://bonvinlab.org/software
Utrecht
Bioinformatics
Center