Information-driven 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

w-e-nmr

Universiteit Utrecht

[Faculty of Science Chemistry]

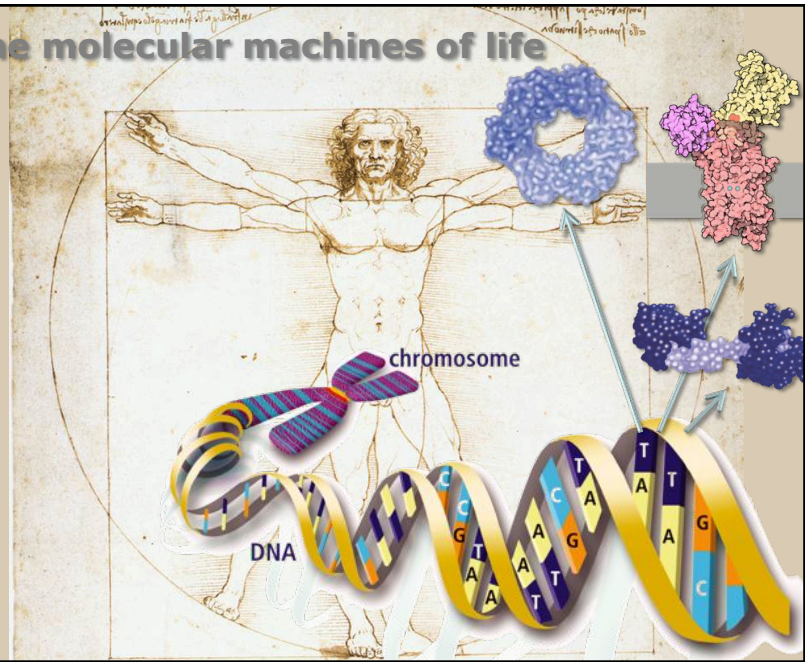
Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- SAXS & docking
- Conclusions & perspectives

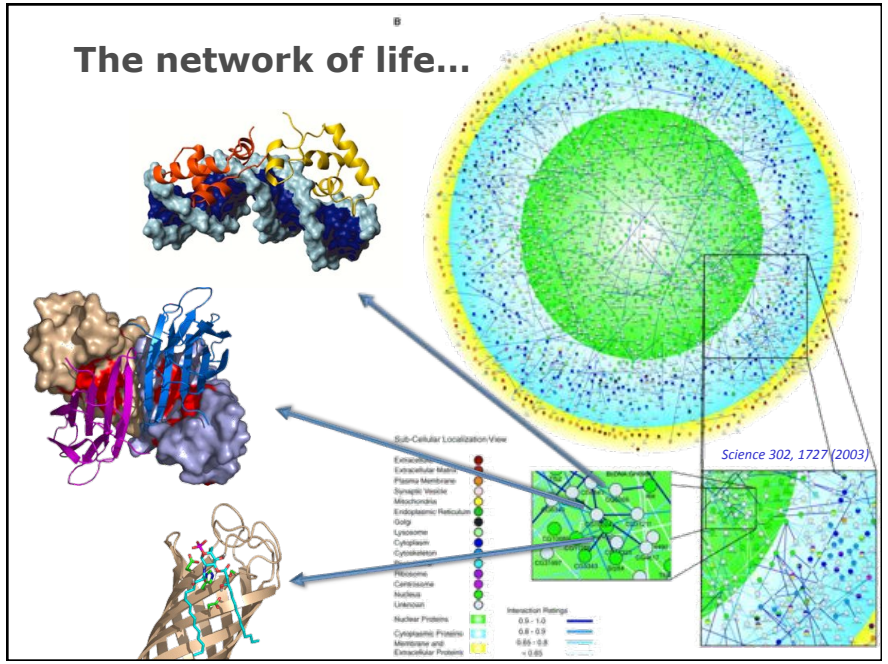
Universiteit Utrecht

[Faculty of Science Chemistry]

The molecular machines of life



The network of life…

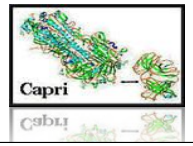


Science 302, 1727 (2003)

Study of biomolecular complexes

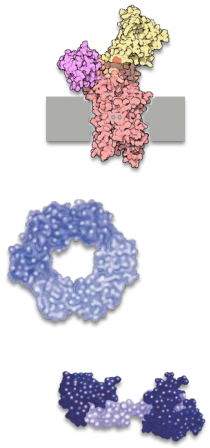
- Classical NMR & X-ray crystallography approaches can be time-consuming
- Problems arise with "bad behaving", weak and/or transient complexes!
- Complementary computational methods are needed!

→ "docking" prediction of the structure of a complex based on the structures of its constituents



"Critical assessment of predicted interactions"
http://capri.ebi.ac.uk

What can we learn from 3D structures (models) of complexes?



- Models provide structural insight into function and mechanism of action
- Models can drive and guide experimental studies
- Models can help understand and rationalize the effect of disease-related mutations
- Models provide a starting point for drug design

Data-driven docking

- There is a wealth of (easily) available experimental data on biomolecular interaction.
- When classical structural studies fail, these are however often not used and the step to modelling (docking) is most of the time not taken.
- These data can be very useful to filter docking solutions or even to drive the docking and thus limit the conformational search problem.

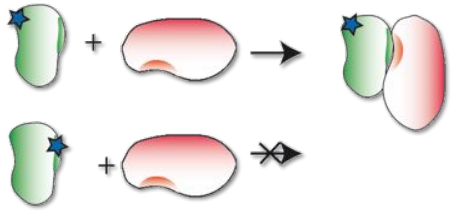
Related reviews

- van Dijk ADJ, Boelens R and Bonvin AMJJ (2005). Data-driven docking for the study of biomolecular complexes. *FEBS Journal* 272 293-312.
- 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.
- de Vries SJ, de Vries M. and Bonvin AMJJ. The prediction of macromolecular complexes by docking. In: *Prediction of Protein Structures, Functions, and Interactions.* Edited by J. Bujnicki Ed., John Wiley & Sons, Ltd, Chichester, UK (2009).
- A.S.J. Melquiond and A.M.J.J. Bonvin. Data-driven docking: using external information to spark the biomolecular rendez-vous. In: Protein-protein complexes: analysis, modelling and drug design. Edited by M. Zacharrias, Imperial College Press, 2010. p 183-209.

Overview

- Introduction
- **Information sources**
- General aspects of docking
- Information-driven docking with HADDOCK
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- SAXS & docking
- Conclusions & perspectives

Experimental sources:
mutagenesis



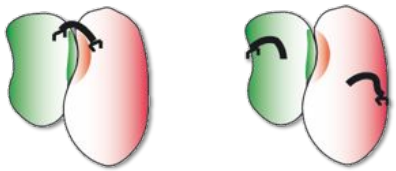
Advantages/disadvantages

+ 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

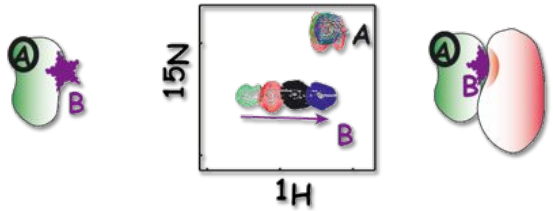
Advantages/disadvantages

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

Detection

- Mass spectrometry
- NMR ^{15}N HSQC

Experimental sources:
NMR chemical shift perturbations



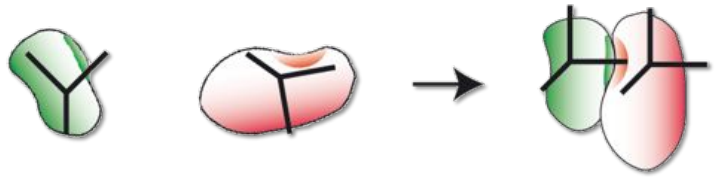
Advantages/disadvantages

+ 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

Experimental sources:
NMR orientational data (RDCs, relaxation)



Advantages/disadvantages

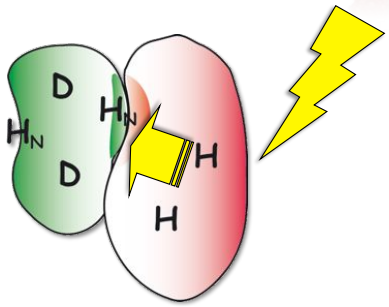
+ Atomic level
- Labeling needed

Detection

- NMR

Experimental sources:
NMR saturation transfer

Amide protons at interface are saturated
==> intensity decrease



Advantages/disadvantages

+ Residue/atomic level
+ No need for assignment if combined with a.a. selective labeling
- Labeling (including deuteration) needed

Other potential experimental sources

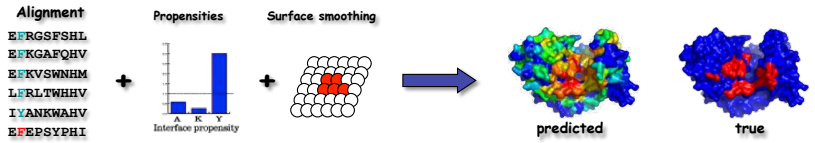
- **Paramagnetic probes** in combination with NMR
- **Cryo-electron microscopy** or tomography and **small angle X-ray scattering** (SAXS) ==> shape information

- 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 **I**nformation does **S**urface **C**onservation **Y**ield?



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

De Vries, van Dijk Bonvin. *Proteins* 2006

What is conservation?

- **Conservation occurs when residues are expected to mutate, but do not mutate, or much more slowly**
- **How to calculate conservation?**
  - Generate a sequence alignment
  - Calculate the expected mutation behavior
  - Calculate deviations from this behavior
  - Is there less change than expected?
- **The residue conservation score is the sum of all deviations from expected behavior**

How to calculate expected conservation?

AFRGTFSHL	AFRGTFSHL
EFRGSFSHL	EFEPSYPHI
Near identical sequences	**Different sequences**
No conservation	**Conservation**

Sequence distance **must be taken into account**

Residue mutation matrix example

- **"Four residue world": Ala, Asp, Glu, Trp**
- **Sequence distance: 1 % mutation**

	Ala	Asp	Glu	Trp
Ala	**99**	0.33	0.33	0.33
Asp	0.33	**99**	0.33	0.33
glu	0.33	0.33	**99**	0.33
Trp	0.33	0.33	0.33	**99**

Residue mutation matrix example

- Some residues mutate however faster than others

	Ala	Asp	Glu	Trp
Ala	98	0.67	0.67	0.67
Asp	0.33	99	0.33	0.33
glu	0.33	0.33	99	0.33
Trp	0.17	0.17	0.17	99.5

Residue mutation matrix example

- Some mutations are more likely than others

	Ala	Asp	Glu	Trp
Ala	98	0.67	0.67	0.67
Asp	0.17	99	0.67	0.17
glu	0.17	0.67	99	0.17
Trp	0.17	0.17	0.17	99.5

Residue mutation matrix example

- You can multiply the matrix by itself to generate distance specific matrices
  - E.g. result of 20 multiplications: 20 % mutation

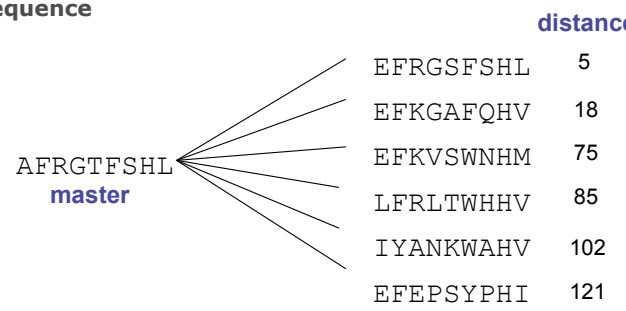
	Ala	Asp	Glu	Trp
Ala	65.96	11.35	11.35	11.35
Asp	2.84	82	11.74	3.42
glu	2.84	11.74	82	3.42
Trp	2.84	3.42	3.42	90.32

Residue mutation matrix

- Several of such matrices exist
- The best known is the Dayhoff (PAM) matrix (Dayhoff *et al.* 1978)
- This matrix is used in Whiscy

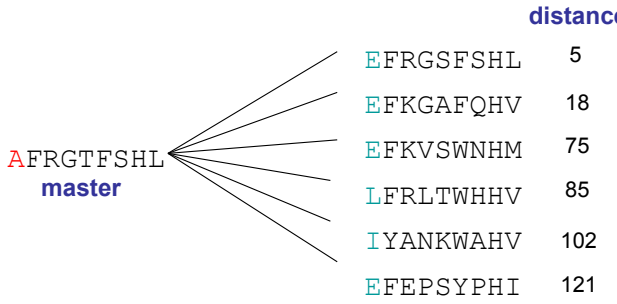
WHISCY calculation

- Take as input a 3D structure and a sequence alignment
- protdist (Felsenstein *et al.*) used to calculate the sequence distances
- WHISCY compares the master sequence to every other sequence

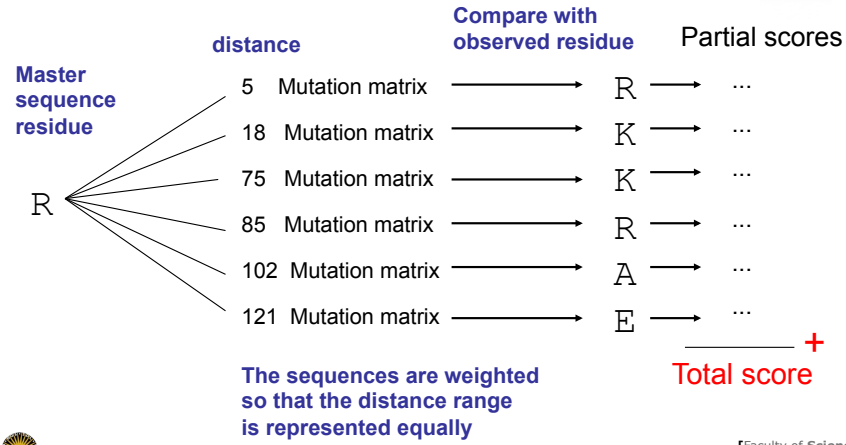


WHISCY calculation

- Each residue is scored independently



WHISCY calculation



Partial score

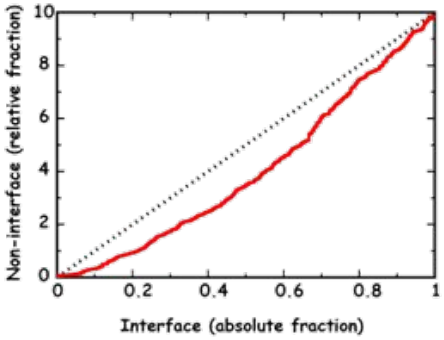
- The partial score is equal to the probability in the distance-dependent mutation matrix
- A correction factor corresponding to the sum of squares of all probabilities is subtracted
- This makes sure that the average score is zero
- WHISCY score > 0 indicates conservation

Testing WHISCY with known complexes

- **Benchmark of 37 protein complexes** (Chen *et al.* 2003)
- **Sequence alignments from the HSSP database** (Sander *et al*. 1991)
  - Some proteins were left out of prediction because of bad sequence alignments
- **Interface definitions by DIMPLOT** (Wallace *et al.* 1995)
  - Residues making contacts across interface (hbond + non-bonded)
- **Surface definition by NACCESS** (Hubbard & Thornton 1993) (15 % accessibility cutoff)

WHISCY raw performance

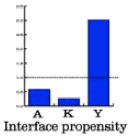
- Fraction of correct versus incorrect predictions for the benchmark



Improving the score using amino acid interface propensities

- Each amino acid has its own interface propensity (from analysis of 3D structures of known complexes):

$$\frac{\text{frequency at the interface}}{\text{frequency at the surface}}$$



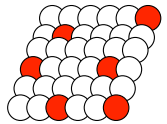
- **WHISCY score converted into a p-value and divided by the a.a. interface propensity**

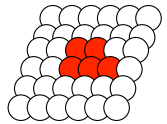
Residue X: score → p = 0.10 → (/ 2.5) → p = 0.04 → higher score

Residue Z: score → p = 0.10 → (/ 0.4) → p = 0.25 → lower score

Improving the score by surface smoothing

- Interface residues are not spread over the surface but form patches


unlikely interface

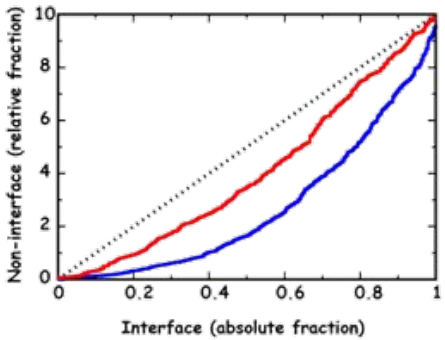

likely interface

- Take the scores of the neighbors into account:
  - Residues with high-scoring neighbors should get a bonus
  - Residues with low-scoring neighbors should get a penalty

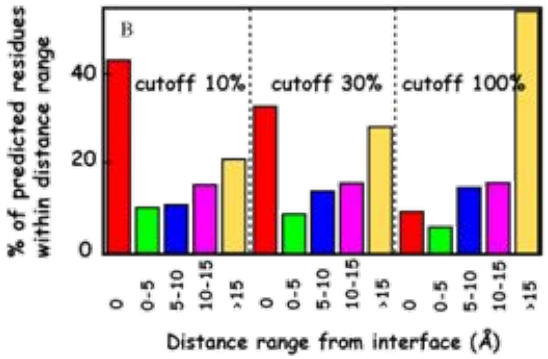
=> Scores are smoothed over a 15Å radius using a Gaussian or optimized step function

WHISCY optimized performance

- **Fraction of correct versus incorrect predictions for the benchmark**



Distribution of predicted interface residues as a function of their distance from the true interface



10% cutoff indicates the WHISCY cutoff resulting in 10% of the true interface predicted

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 **(Porollo & Meller, 2007)**
  - SVM method **(Dong et al., 2007)**
  - ...
  - 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.chem.uu.nl/services/CPORT

CPORT webserver

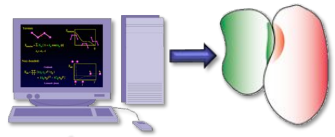
Universiteit Utrecht

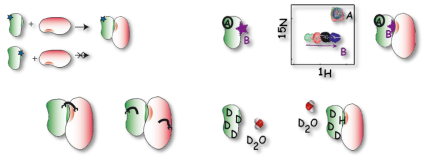
haddock.chem.uu.nl/services/CPORT/

[Faculty of Science Chemistry]

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
- **General aspects of docking**
- Information-driven docking with HADDOCK
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- SAXS & docking
- Conclusions & perspectives

Universiteit Utrecht

[Faculty of Science Chemistry]

A few docking reviews

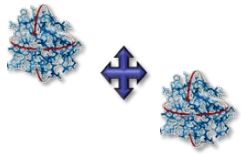
- Halperin *et al.* (2002) "Principles of docking: an overview of search algorithms and a guide to scoring functions". *PROTEINS: Struc. Funct. & Genetics* 47, 409-443.
- Special issues of *PROTEINS:* (2003) (2005) (2007) and (2010) which are dedicated to CAPRI.
- Brooijmans and Kuntz (2003) "Molecular recognition and docking algorithms". *Annu. Rev. Biophys. Biomol. Struct.* 32, 335-373.
- Russell *et al.* (2004) "A structural perspective on protein-protein interactions". *Curr. Opin. Struc. Biol.* 14, 313-324.
- Van Dijk *et al.* (2005) "Data-driven docking for the study of biomolecular complexes." *FEBS J.* 272, 293-312.

Universiteit Utrecht

[Faculty of Science Chemistry]

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?



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

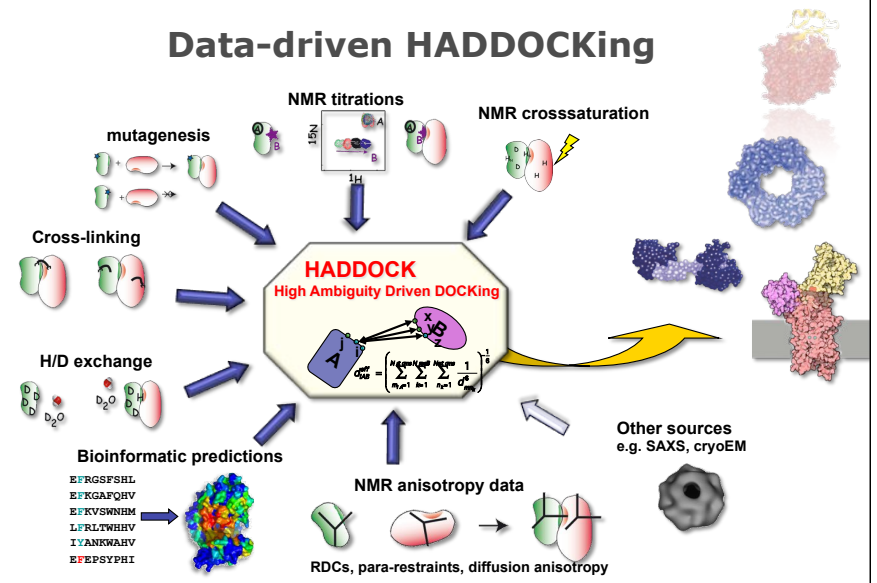
Scoring

- 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,..)

Overview

- Introduction
- Information sources
- General aspects of docking
- **Information-driven docking with HADDOCK**
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- SAXS & docking
- Conclusions & perspectives

Data-driven HADDOCKing

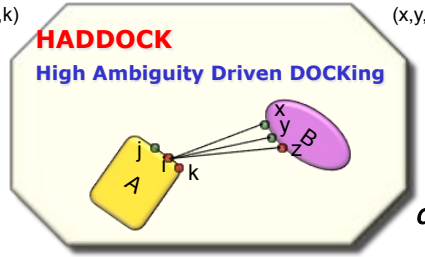


Dominguez, Boelens & Bonvin. *JACS* 125, 173 (2003).

Data-driven docking with HADDOCK

List of interface residues for protein A (i,j,k)

List of interface residues for protein B (x,y,z)



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

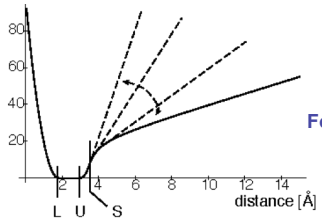
$$d_{iAB}^{eff} = \left(\sum_{m_{i\,A}=1}^{N\,at\,ams} \sum_{k=1}^{N\,resB} \sum_{n_k=1}^{Nat\,ams} \frac{1}{d_{mn_k}^6} \right)^{-\frac{1}{6}}$$

Ambiguous Interaction Restraint: (Nilges & Brunger 1991)
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

Ambiguous Interaction Restraints (AIRs)

- **Soft-square potential (Nilges) used to avoid large forces**

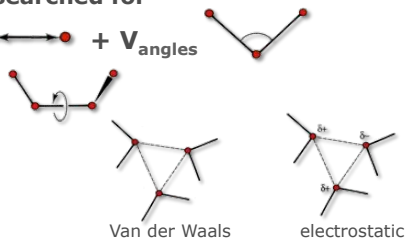


Force becomes constant >2Å violation

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

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}

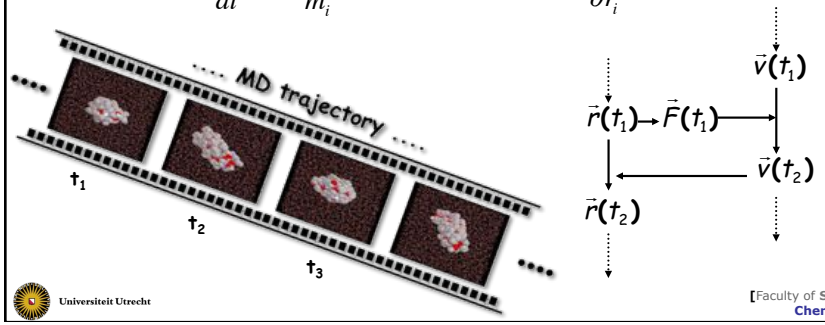


- Search is performed by a combination of gradient driven energy minimization and molecular dynamics simulations

Classical mechanics

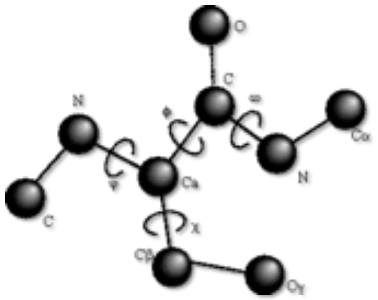
- Molecular dynamics: generates successive configurations of the system by integrating **Newton's second law**

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

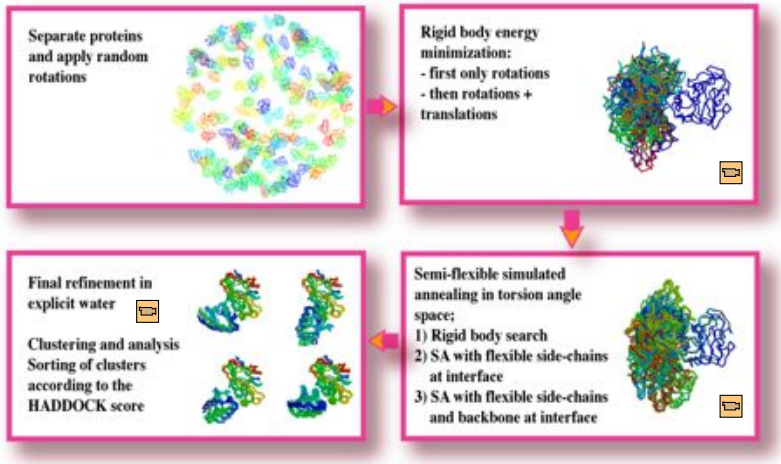


Torsion angle dynamics

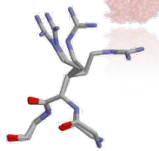
- dynamics time step dictated by bond stretching: waste of CPU time
- important motions are around torsions
- ~ 3 degrees of freedom per AA (vs $3N_{atom}$ for Cartesian dynamics)
- Available in DYANA, X-PLOR, CNS, X-PLOR-NIH



HADDOCK docking protocol

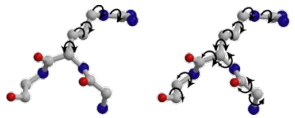


HADDOCK & Flexibility



- **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**
- **Ranking of based on HADDOCK score defined as:**

Rigid: Score = 0.01 E_{air} + 0.01 E_{vdW} + 1.0 E_{elec} + 1.0 E_{desolv} – 0.01 BSA

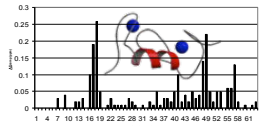
Flexible: Score = 0.1 E_{air} + 1.0 E_{vdW} + 1.0 E_{elec} + 1.0 E_{desolv} – 0.01 BSA

Water: Score = 0.1 E_{air} + 1.0 E_{vdW} + 0.2 E_{elec} + 1.0 E_{desolv}

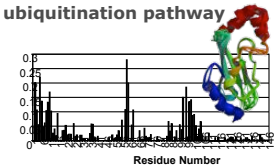
- E_{air}: ambiguous interaction restraint energy
- E_{desolv}: desolvation energy using Atomic Solvation Parameters (Fernandez-Recio et al *JMB* 335, 843 (2004))
- BSA: buried surface area

The Not4 – UbcH5B complex

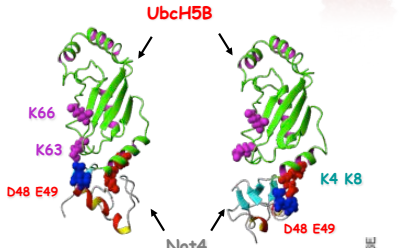
- **Not4: involved in the RNA polymerase II regulation. Contains a N-terminal Ring finger domain** (Hanzawa *et al.*, 2000)



- **UbcH5B: involved in the ubiquitination pathway**

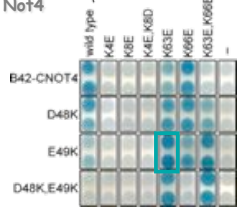


Best Haddock solutions



Haddock directed mutagenesis

==> Altered specificity mutants!

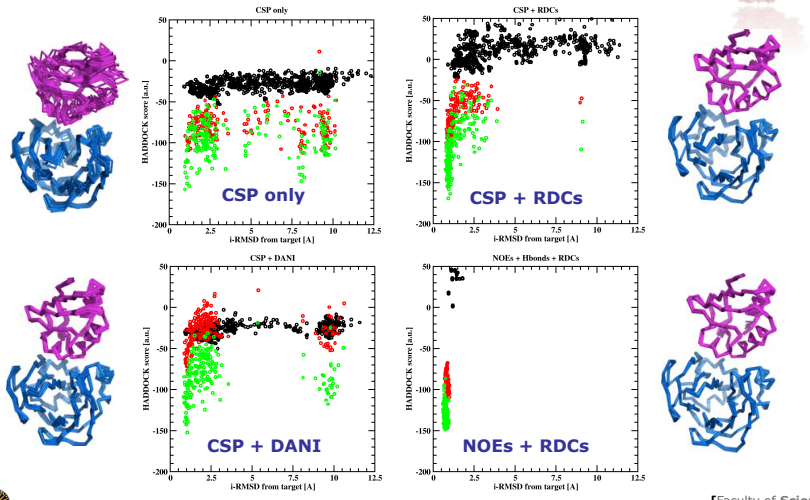


 Dominguez, Bonvin, Winkler, van Schaik, Timmers & Boelens. *Structure* 2004

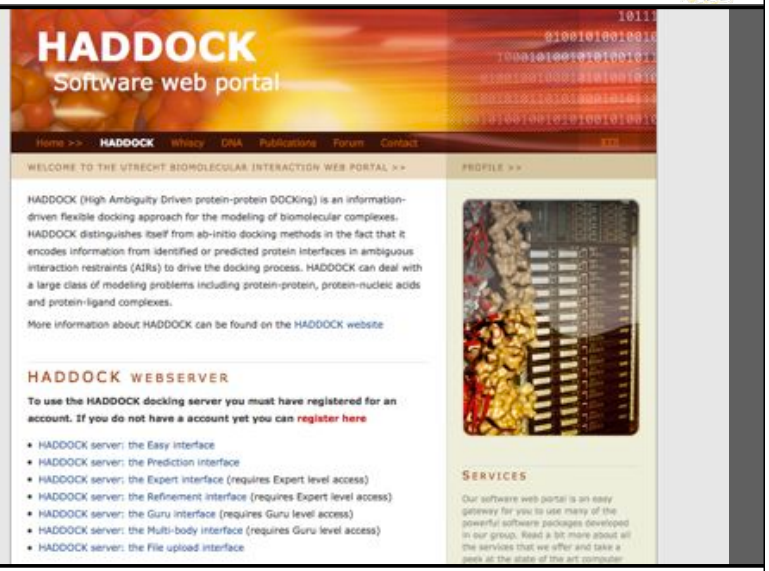
Accuracy <-> Data

When does the model stop and the structure start?

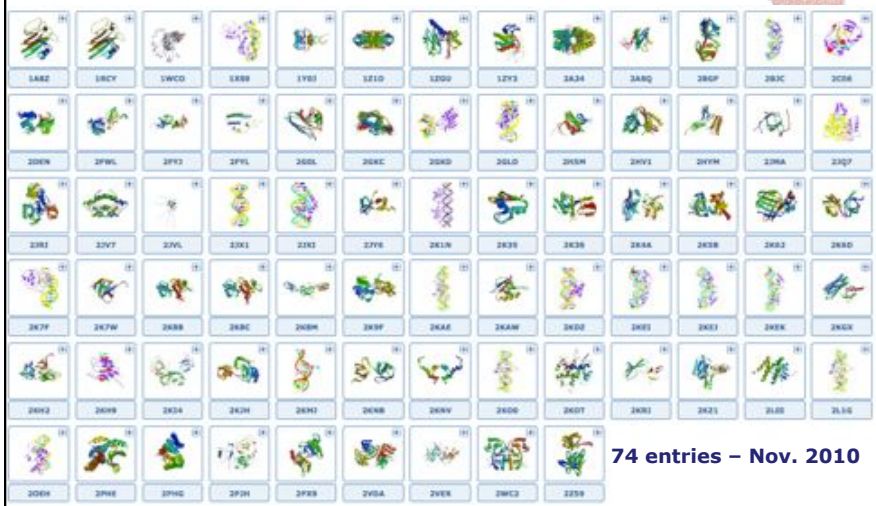
Accuracy <-> Data: E2A-HPR



The HADDOCK web portal

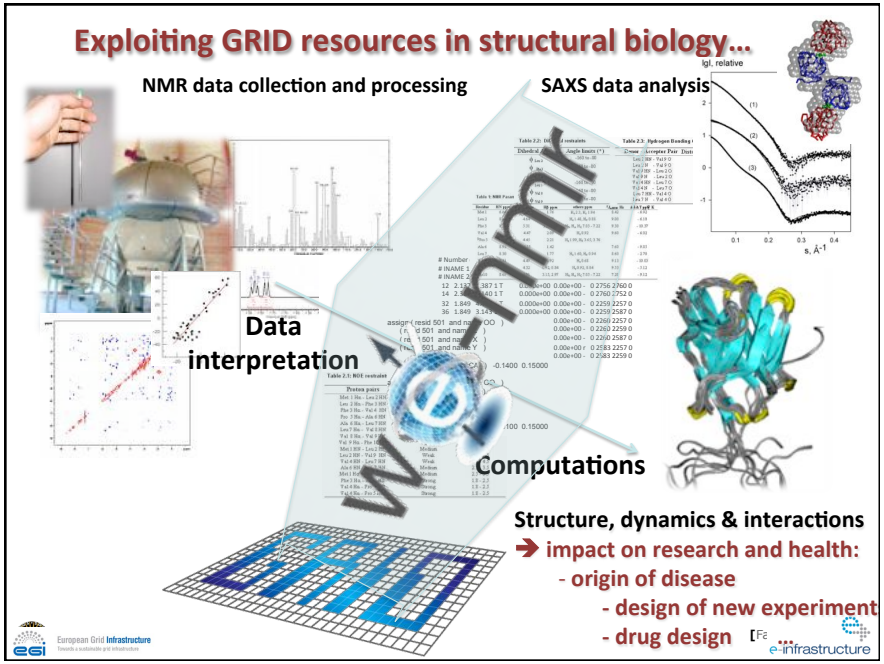


The HADDOCK PDB structure gallery



74 entries – Nov. 2010

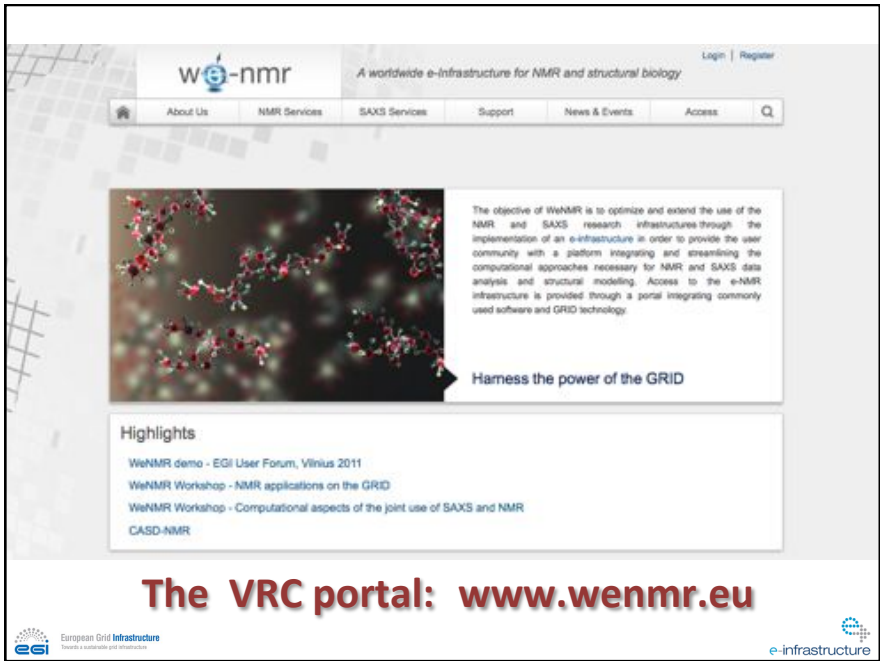
Exploiting GRID resources in structural biology…



WeNMR platform operational and well used!

- Largest global VO in the life sciences
- Over 280 registered users and growing
- >10000 CPUs
- >500 CPU years over the last 12 months
- ~20% of Life Sciences on the Grid
- User-friendly access to e-Infrastructure via web portals

www.wenmr.eu



The VRC portal: www.wenmr.eu

European Grid Infrastructure
e-infrastructure

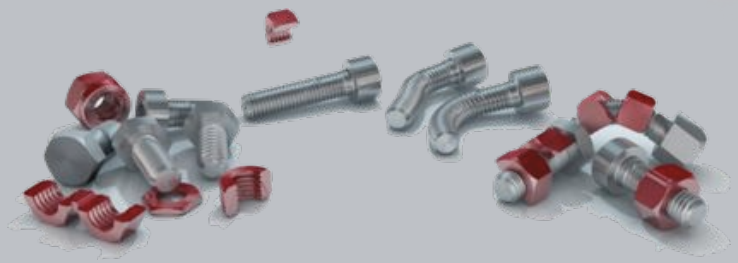
Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- **Protein-DNA HADDOCKing**
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- SAXS & docking
- Conclusions & perspectives

Universiteit Utrecht

[Faculty of Science
Chemistry]

Modeling protein-DNA interactions: Bend and Twist it to make it fit



Modelling of Protein-DNA complexes: a two-stage protocol

Input structures:
- canonical B-DNA
- Protein (ensemble)

1st docking run

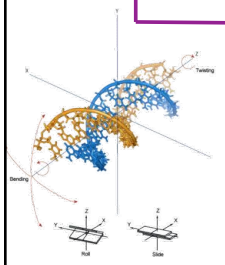
It0 → It1 → Water → Scoring

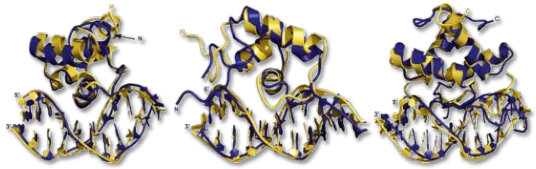
It0: rigid body docking It1: semi-flexible refinement
Water: final refinement in explicit solvent

DNA library generation

2nd docking run

It0 → It1 → Water → Scoring





Cro - O_1R iRMSD = 1.62 Å
Lac - O1 iRMSD = 2.02 Å
Arc - operator iRMSD = 1.90 Å

Van Dijk *et al. Nucl. Acid. Res.* 2006

Generating (custom) nucleic acids structures

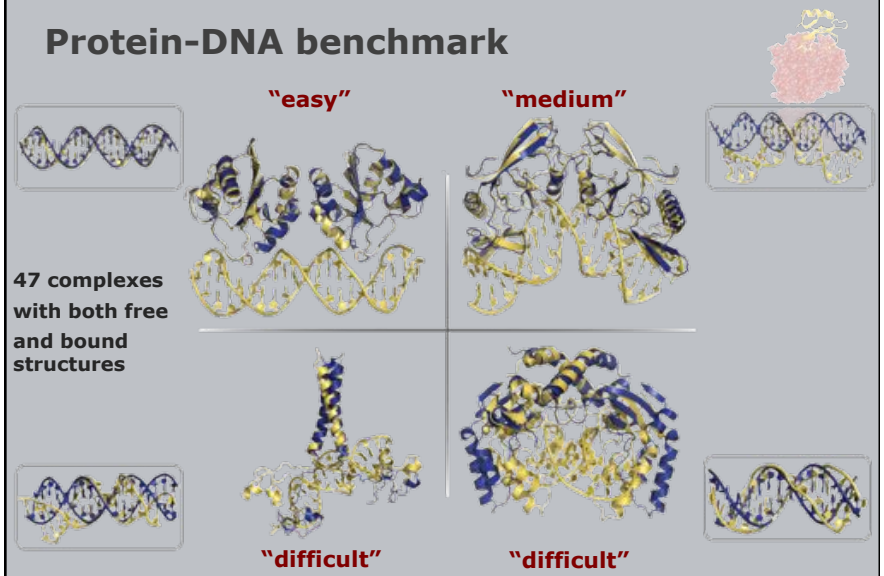
Control over global conformation (bend & twist)
Uses 3DNA (Lu & Olson, NAR 2003)

Generate A-DNA or B-DNA from sequence
Full control over base-pair(step) parameters

Van Dijk & Bonvin
NAR 2009

haddock.chem.uu.nl/dna

Protein-DNA benchmark

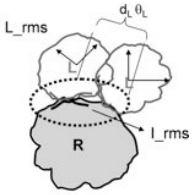


47 complexes with both free and bound structures

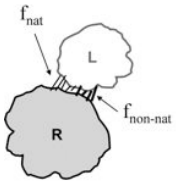
Van Dijk *et al.* NAR 2008

Assessment terminology

	F_{nat}	I-RMSD (Å)	i-RMSD (Å)
High (***)	≥0.5	≤1	≤1
Medium (**)	≥0.3	≤5	≤2
Acceptable (*)	≥0. 1	≤10	≤4
Incorrect	<0. 1	>10	>4

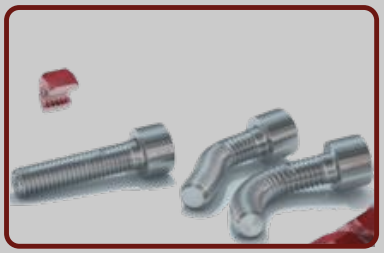


- i-RMSD: Interface RMSD
- l-RMSD: Ligand RMSD
- F_{nat}: Fraction of native contacts



Lensink *et al. Proteins 2007*

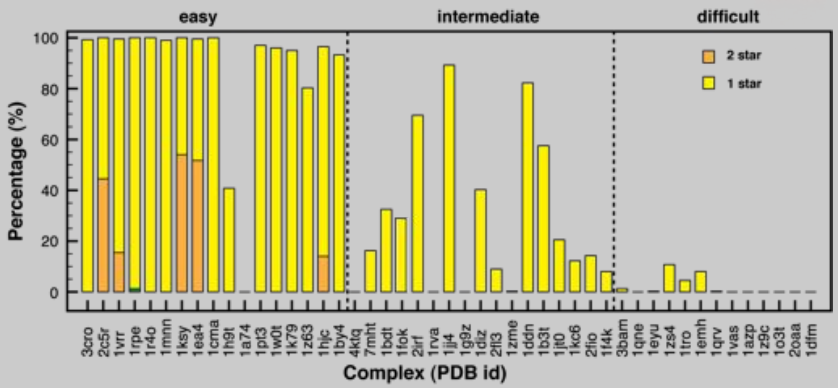
Unbound-Unbound using canonical B-DNA and true interface restraints



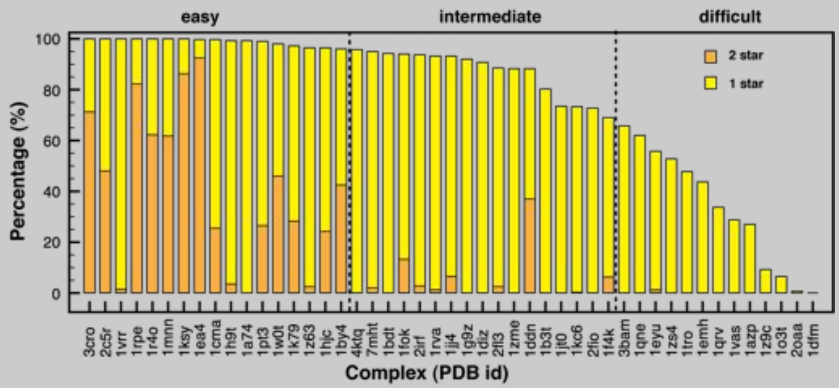
Is the protein-DNA docking procedure able to account for conformation changes, and to what extend?

Van Dijk & Bonvin. NAR 2010

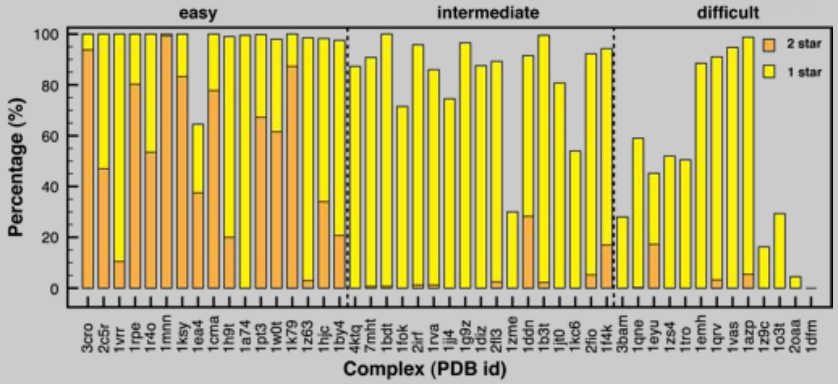
Performance of rigid-body docking only



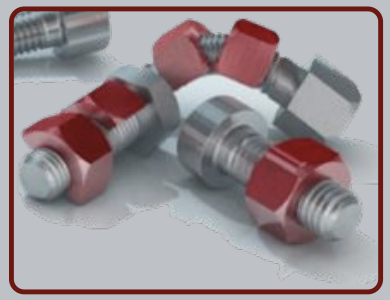
Performance after flexible refinement (1 cycle)



Performance after the 2 steps protocol with custom DNA library



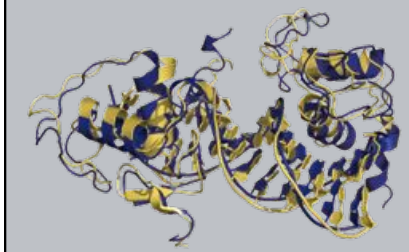
Unbound-Unbound using canonical B-DNA with experimental information



How well does the procedure perform when knowledge-based restraints are used?

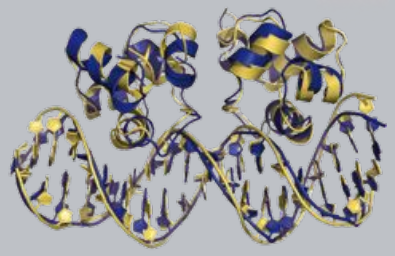
"easy" cases

Retinoic acid receptor



1by4 **
fnat = 0.40
iRMSD = 3.55 Å
dRMSD = 1.50 Å

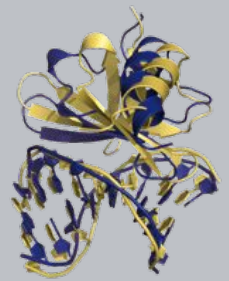
434 Cro protein



3cro **
fnat = 0.50
iRMSD = 2.23 Å
dRMSD = 1.93 Å

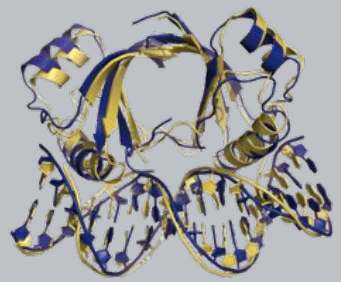
"medium" cases

Hyperthermophile chromosomal protein SAC7D



1azp *
fnat = 0.11
iRMSD = 3.44 Å
dRMSD = 1.58 Å

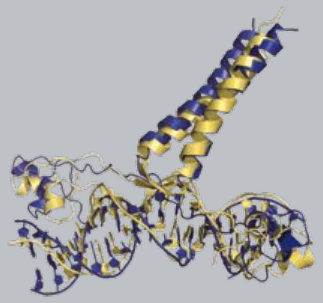
papillomavirus type 18 E2



1jj4 **
fnat = 0.44
iRMSD = 2.63 Å
dRMSD = 2.26 Å

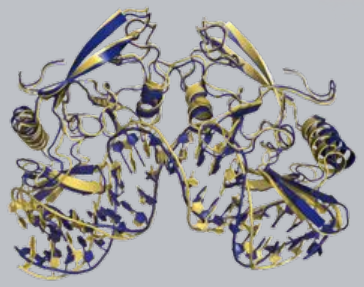
"difficult" cases

PUT3



1zme *
fnat = 0.15
iRMSD = 3.75 Å
dRMSD = 3.23 Å

1-PPOL homing endonuclease



1a74 **
fnat = 0.31
iRMSD = 3.24 Å
dRMSD = 3.70 Å

Overview

HADDOCK's adventures in CAPRI

"Critical assessment of predicted interactions"

http://capri.ebi.ac.uk

- CAPRI is a blind test for protein-protein docking
- Usually 3 weeks for a predictions, 10 models can be submitted
- We participated to rounds 4 to 19 for a total of 27 targets
- For HADDOCK, we derived information to define AIRs from literature and bioinformatic predictions

Van Dijk et al. *Proteins* 2005; de Vries *et al. Proteins* 2007,2010

Performance of the HADDOCK team in CAPRI rounds 13-19

- **29** [1, 1, 2, 1, 1, 1, 0, 0, 0, 0] BU
- 30 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UU
- 32 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UU
- 33 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UH
- **34** [2, 2, 1, 2, 1, 1, 0, 0, 0, 0] UB
- 35 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] HH
- 36 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] BH
- **37** [0, 0, 2, 2, 0, 0, 0, 0, 0, 0] UH (2 *** uploaded)
- 38 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UH
- 39 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UB

} Two-domain protein – crystal structure incompatible with covalently linked domains!!!

- **40** [3, 3, 3, 3, 3, 3, 3, 3, 3, 3] UB
- **41** [1, 1, 2, 2, 1, 1, 1, 1, 1, 1] UH
- **42** [0, 0, 0, 0, 0, 0, 0, 0, 0, 1] HH(H)

1 *, 4 **, 1 *, 12 stars**

Performance of the HADDOCK server in CAPRI rounds 15-19

- 32 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UU
- 33 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UH
- 34 [1, 1, 1, 1, 1, 1, 0, 0, 0, 1] UB
- 35 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] HH
- 36 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] BH
- 37 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UH
- 38 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UH
- 39 [0, 0, 0, 0, 0, 0, 0, 0, 0, 0] UB

} Two-domain protein – crystal structure incompatible with covalently linked domains!!!

- 40 [0, 0, 3, 0, 0, 0, 0, 0, 0, 0] UB
- 41 [1, 1, 2, 1, 0, 0, 0, 0, 0, 0] UH
- 42 [0, 0, 0, 0, 0, 0, 0, 0, 1, 0] HH(H)

1 *, 1 **, 2 *, 7 stars**

HADDOCK's performance in CAPRI

- **Overall performance:**
  - **3***, 9**, 3*** 15 out of 25 (60%)
- **Unbound only performance:**
  - **6**, 2*** 8 out of 13 (62%)
- As good as it gets... (among the top performing methods)
- "wrong" solutions still often have correctly predicted interfaces, but wrong orientations of the components
- ==> still useful to direct the experimental work

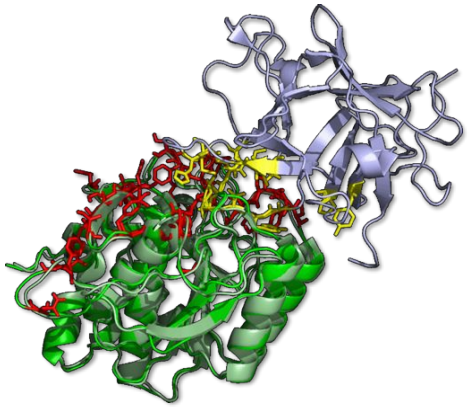
Post-docking interface prediction

Target	Fraction true interface coverage		Fraction overprediciton	
	ligand	receptor	ligand	receptor
T29	0.92	0.88	0.11	0.20
T30	0.84	0.73	0.26	0.39
T32	0.87	0.75	0.25	0.31
T33	0.61	0.42	0.20	0.50
T34	0.61	0.87	0.17	0.10
T37	0.36	0.89	0.66	0.27
T40	0.90	0.96	0.05	0.03
T41	0.89	0.83	0.04	0.15
T42	0.87	0.87	0.14	0.14

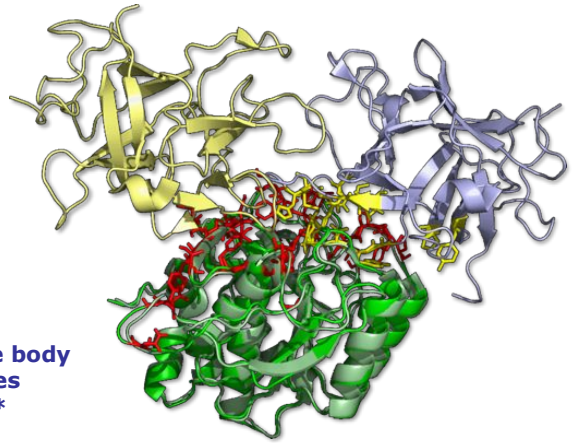
HADDOCK's weakness
(one of them)

Information-driven...

Our T32 failure… (the "easy" one)



Our T32 failure… (the "easy" one)

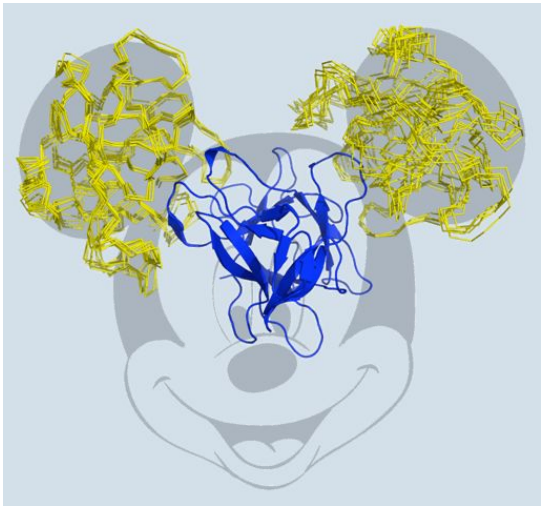


Note: Three body docking does generate ** solutions…

HADDOCK's strength
(one of them)

Information-driven…

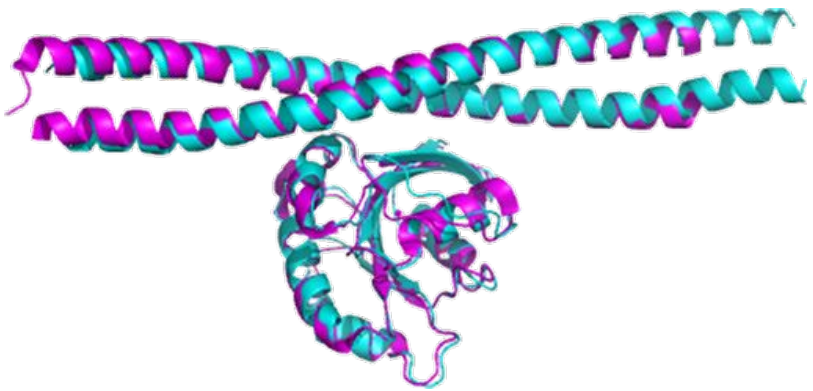
T40



10x ***

T37

** submitted, *** uploaded

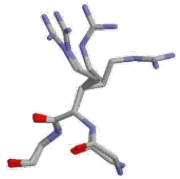


Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- **Small molecule HADDOCKing**
- SAXS & docking
- Conclusions & perspectives

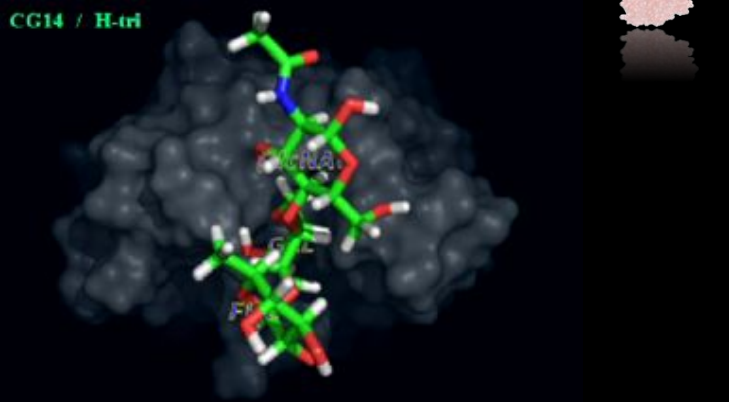
Small molecules docking with HADDOCK

- **Docking protocol issues:**
  - **Pre-sample ligand conformations**
  - **use ensemble for docking**
  - **same for protein**



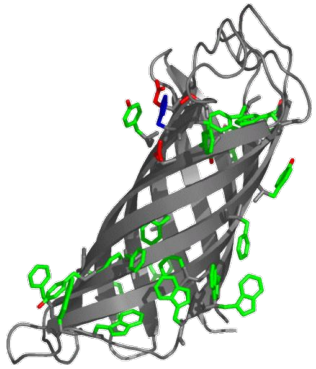
  - **If flexibility is expected to play an important role (e.g. docking of an unstructured peptide onto a protein), perform a fully flexible docking during the simulated annealing phase**

Fully flexible protein-ligand docking



Wu *et al. Glycobiology* 2007

HADDOCK-modelling of substrate binding in PagL, an outer-membrane enzyme involved in LPS-modification

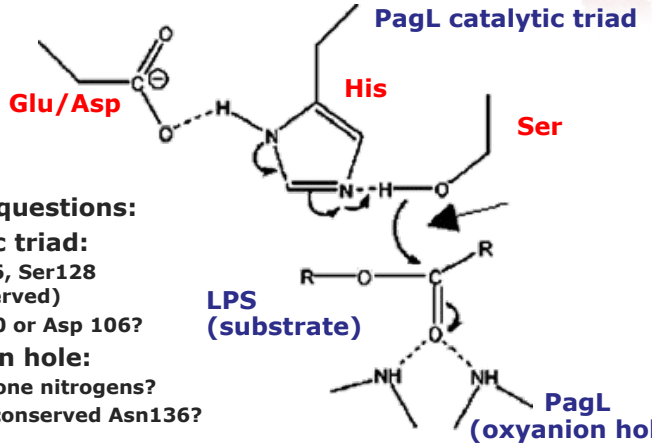


PagL

- Deacetylase (hydrolysis of acylesterbond)
- Activity found in S. typhimurium, B. Bronchiseptica and P. aeruginosa
- PagL homologues found in more than 10 bacterial species
- Crystal structure solved in Utrecht
- Only three residues conserved (Phe104, His126, Ser128)
- Site directed mutagenesis: serine hydrolase

Crystal and Structural Chemistry
- Wietske Lambert
- Lucy Vandeputte-Rutten
- Piet Gros

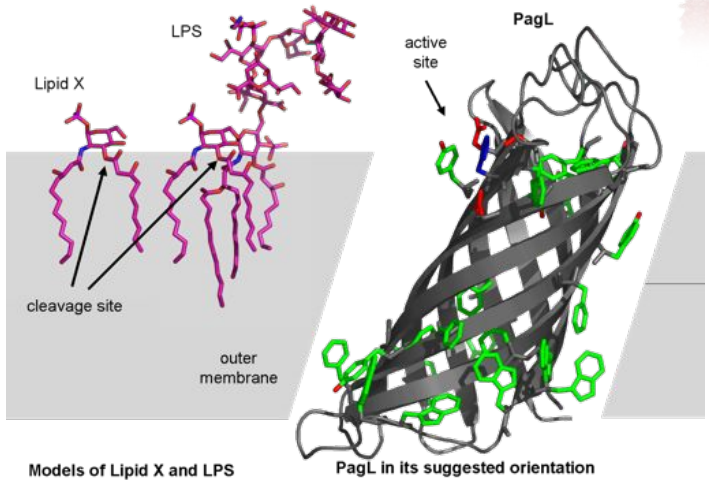
PagL: serine hydrolase mechanism



Still open questions:

- **catalytic triad:**
  - His126, Ser128 (conserved)
  - Glu140 or Asp 106?
- **oxyanion hole:**
  - backbone nitrogens?
  - semi-conserved Asn136?

Substrate recognition by PagL

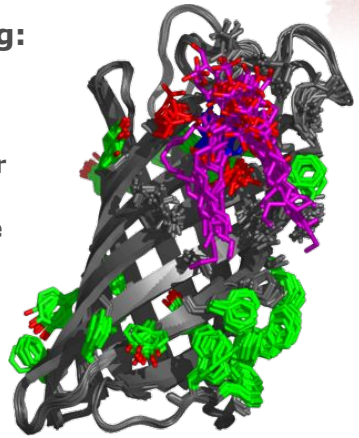


Models of Lipid X and LPS

PagL in its suggested orientation

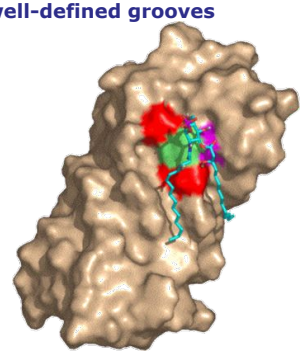
Lipid x docking onto PagL

- **Information for docking:**
  - **reaction mechanism**
    - carbonyl C of lipid x close to active site Ser of PagL
    - ester O of lipid x close to active site His of PagL
  - **hydrophobicity**
    - acyl chains of lipid x should be in the membrane



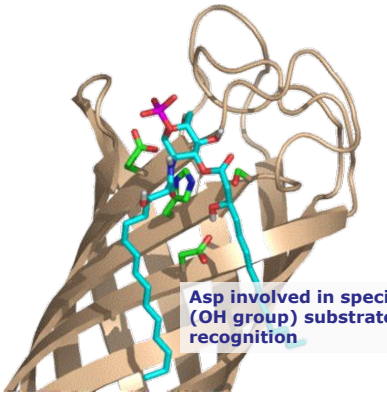
New insights from docking:

Lipid x acyl chains bind in well-defined grooves



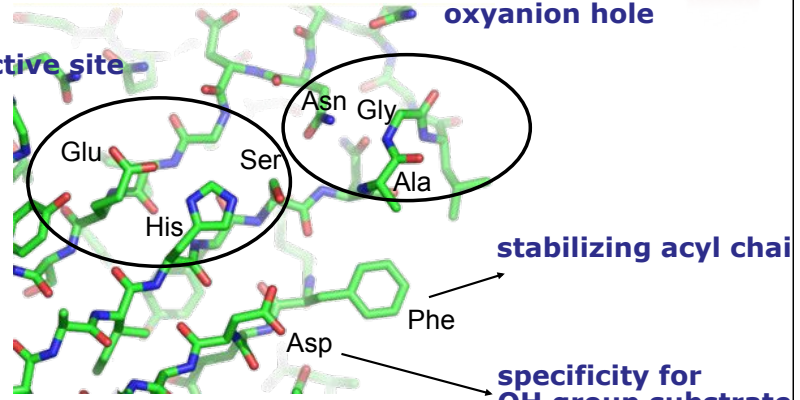
HADDOCK best solution

Catalytic triad: Ser-His-Glu triad



Asp involved in specific (OH group) substrate recognition

PagL active site



Lutten et al. PNAS 2006

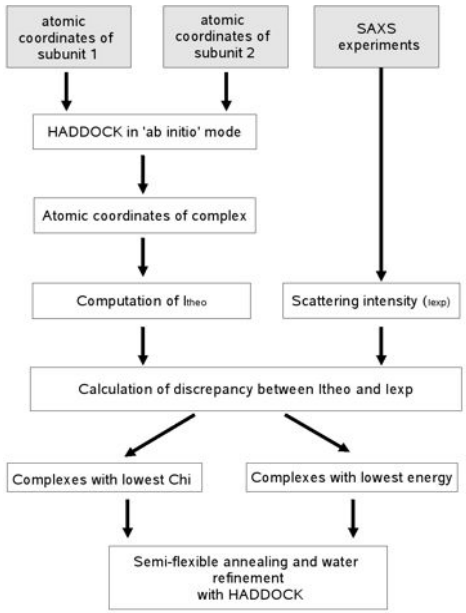
Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- **SAXS & docking**
- Conclusions & perspectives

Combining SAXS and docking

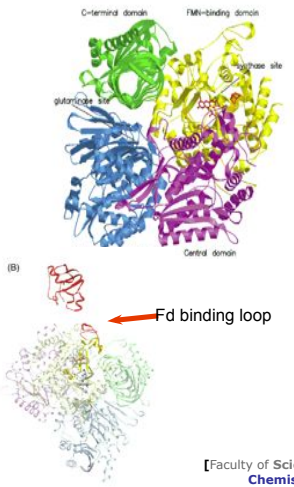
A possible strategy

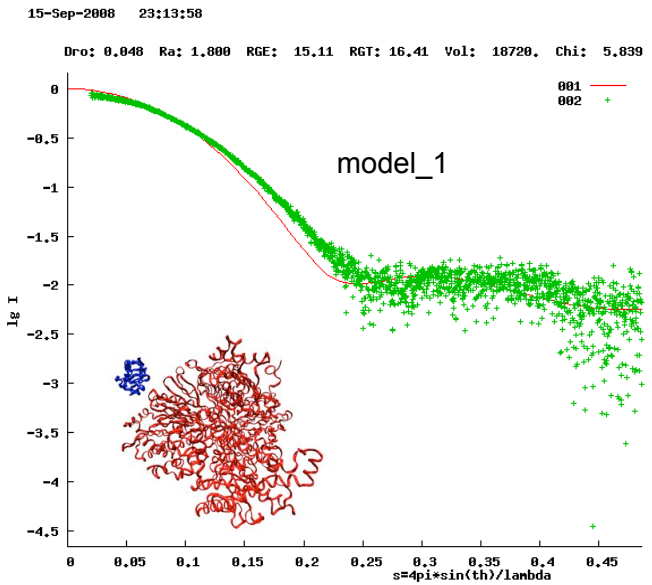
Crysol

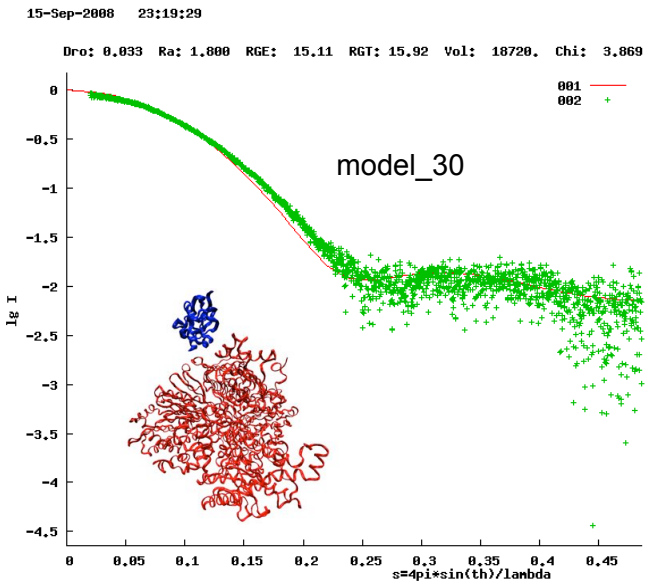


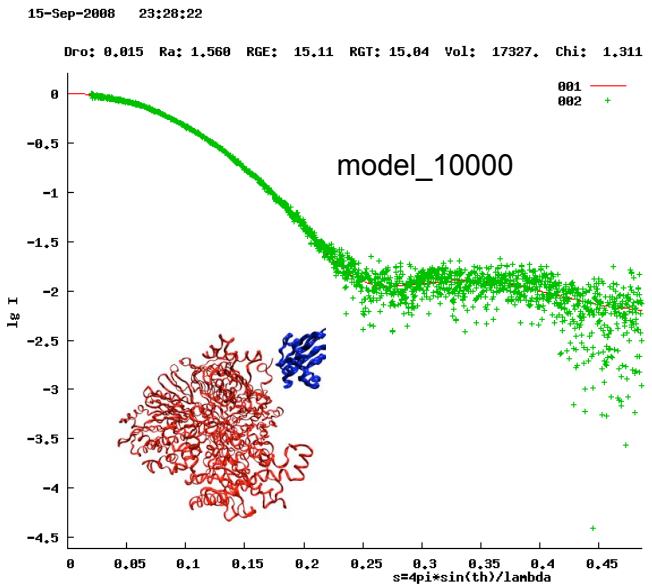
Combining SAXS & docking: one example

- GltS catalyzes the formation of two molecules of L-glutamate from L-glutamine and 2-oxoglutarate
- X-ray structures with substrate and inhibitor have been reported
- SAXS data on GltS and its physiological electron donor ferredoxin (Fd):
  - Suggests an equimolar (1:1) complex.
  - Model based on crystal structure of Fd:Fd-GltS(1:1) fits the SAXS data with χ^2 = 1.3

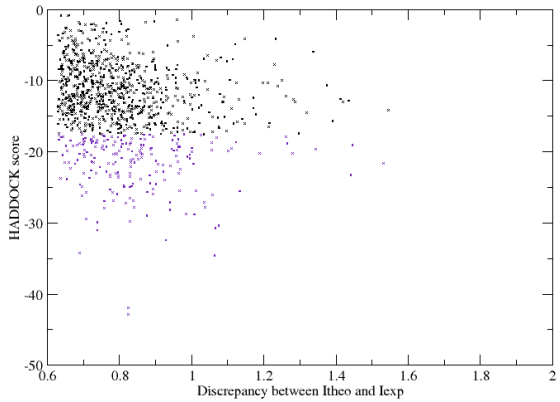




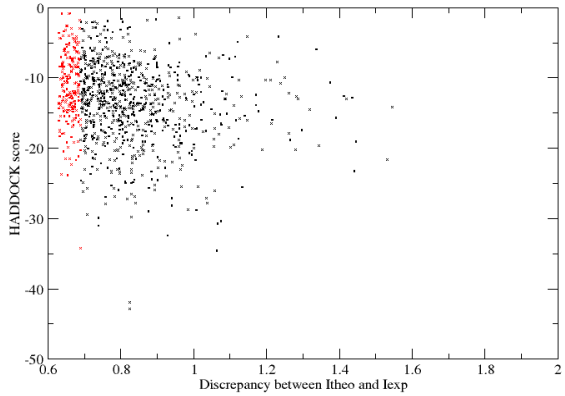




Selection based on HADDOCK energy

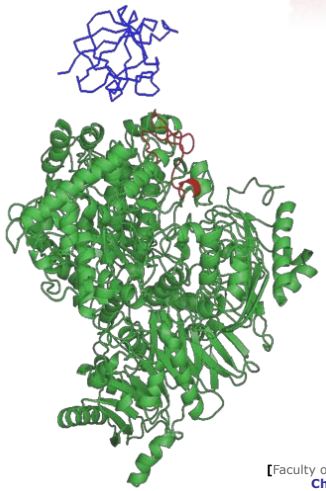


Selection based χ square

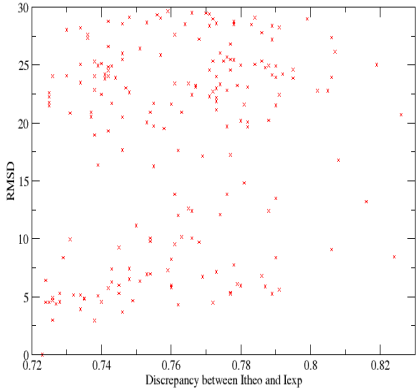


SAXS driven HADDOCK model (one of them …)

- (one of the) HADDOCK model selected based on χ2 has Ferredoxin close to the anticipated Fd-binding loop.
- Fits well to the experimental data (χ2 = 0.8)



χ^2 versus RMSD… a unique, well defined solution???



Overview

- Introduction
- Information sources
- General aspects of docking
- Information-driven docking with HADDOCK
- Protein-DNA HADDOCKing
- HADDOCK's adventures in CAPRI
- Small molecule HADDOCKing
- SAXS & docking
- **Conclusions & perspectives**

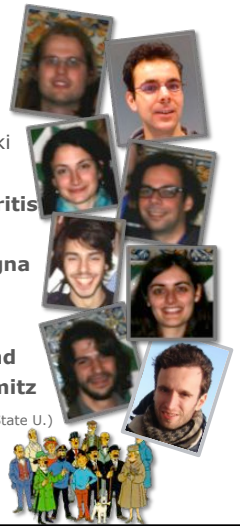
Conclusions & Perspectives

- **Data-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**
- **Data-driven docking is complementary to classical structural methods**
- **Many challenges however remain:**
  - **Scoring**
  - **Predicting and dealing with conformational changes**
  - **Predicting binding affinities**

Acknowledgements

The HADDOCK team

- Cyril Dominguez
- Aalt-Jan van Dijk
- **Sjoerd de Vries**
- **Marc van Dijk**
- Mickaël Krzeminski
- **Ezgi Karaca**
- **Panagiotis Kastritis**
- **Joao Rodrigues**
- **Annalisa Bordogna**
- Aurélien Thureau
- Tsjerk Wassenaar
- **Adrien Melquiond**
- **Christophe Schmitz**
- Victor Hsu (Oregon State U.)
- Rolf Boelens
- Alexandre Bonvin



Babis Kalodimos'lab
Rutger University

Marc Timmers lab
Utrecht Medical Center

Piet Gros lab
Utrecht Science Faculty

€€:

NWO — Visitor grant, VICI, NCF (BigGrid)

SPINE II, Extend-NMR, NDDP, HPC-Europe, BacABs, e-NMR

Barcelona Supercomputing Center — Centro Nacional de Supercomputación