HADDOCK: High Ambiguity Driven DOCKing

A protein-protein docking approach based on biochemical and/or biophysical data

Frontiers of NMR VIII
Taos 2003

Alexandre Bonvin
Utrecht University

Determination of Protein-Protein complexes

- In PDB: >15000 protein structures but <900 structures of complexes (<100 by NMR) !
- NMR: typically based on intermolecular NOEs and residual dipolar couplings (e.g. Clore *PNAS* (2001) 97: 9021)
  - Collection of restraints difficult and lengthy process
  - Requires complete assignments (side-chains) at the interface
- Ab-initio docking: typically based on theoretical knowledge and not on experimental data
- NMR titration data that are easily available at the stage of backbone assignment are typically not used!

Determination of Protein-Protein complexes

- Our aim:

  Combine biochemical or biophysical data and docking methods to solve protein–protein complexes

- HADDOCK package using CNS and python scripts derived from ARIA (Linge & Nilges)

Requisites

- 3D structures of the two partners must be known
- No large conformational change must take place upon complex formation
- Information available to map the interaction interface of both partners, e.g.
  - Chemical shift perturbation
  - Mutagenesis data
  - Any other type of data

Outline

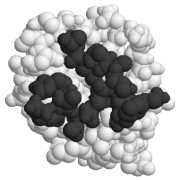
- Introduction
- **Methodology**
- Validation
- Application
- Conclusions & Perspectives

Ambiguous Interaction Restraints (AIRs)

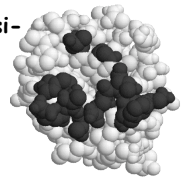
- Two kinds of residues are considered:
  - The active residues
    - Known to be involved in the interaction (NMR data, mutagenesis)
    - and high solvent accessibility (>50%)
  - The passive residues
    - All surface neighbor residues
    - and high solvent accessibility (>50%)

AIRs definition

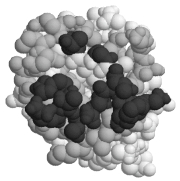
1. Residues with large CSP or from mutagenesis data



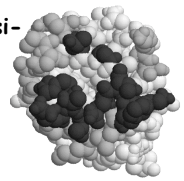
2. Filtering using solvent accessible surface area



3. All surface neighbors

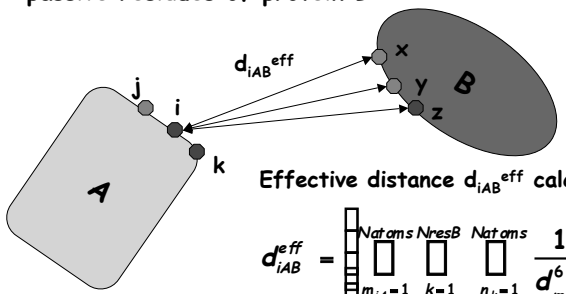


4. Filtering using solvent accessible surface area



Ambiguous Interaction Restraints (AIRs)

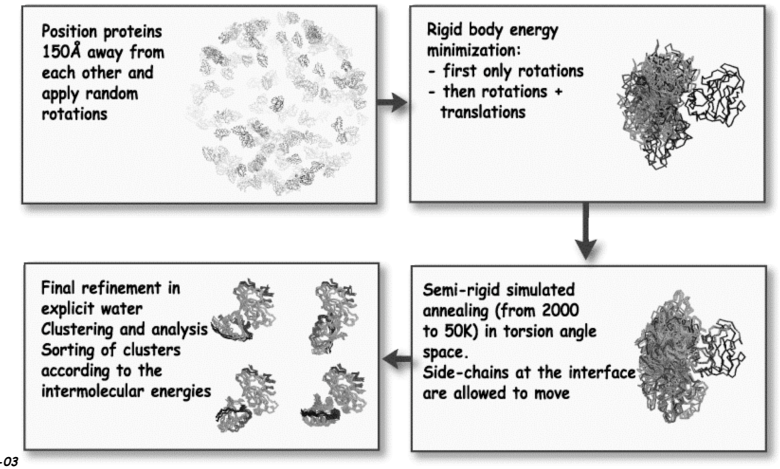
- AIR defined as an ambiguous distance restraint with a maximum value of 2-3Å between any atom of an active residue i of protein A and any atom of all active and passive residues of protein B



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

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

HADDOCK docking protocol

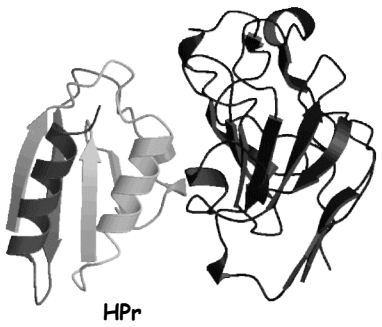


Outline

- **Introduction**
- **Methodology**
- **Validation**
  - **E2A-HPr** *using chemical shift perturbation data*
- **Application**
- **Conclusions & Perspectives**

E2A-HPr

NMR complex (1GGR)
(Wang *et al. EMBO 2000*)



Free proteins:

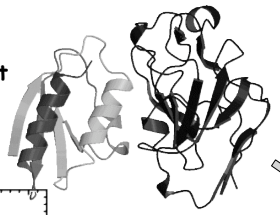
	Rmsd from cplx
E2A (1F3G)	0.36 Å
HPr (1POH)	0.17 Å

 # AIR restraints

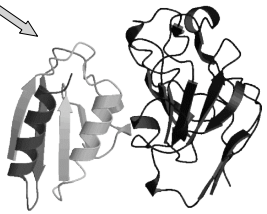
From E2A: 11 active and 4 passive
From HPr: 9 active and 1 passive

E2A-HPr docking results

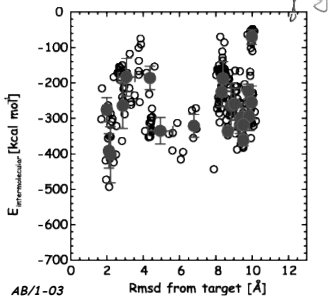
Best of best cluster (lowest energy)



Backbone 2.3Å
Interface 2.1Å

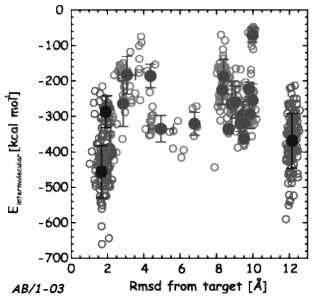


Target
(PDB: 1GGR)



E2A-HPr new docking results

1GGR



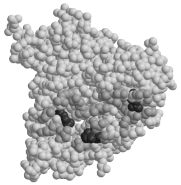
Best cluster	"single" Haddock	"ensemble" HADDOCK
Rmsd from 1GGR [Å]	2.5 ± 0.2	1.4 ± 0.2
iRmsd from 1GGR [Å]	2.1 ± 0.1	1.7 ± 0.2
E_{inter} [kcal mol^{-1}]	-392 ± 49	-443 ± 92
BSA [Å^2]	1453 ± 114	1676 ± 150

- 30 start structures for HPr (1HDN)
- 1 start structure for E2A
- N_{trial} = 10
- 1500 conformations at it0 (15000 trials)
- non-bonded scaling

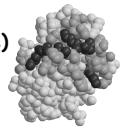
AB/1-03

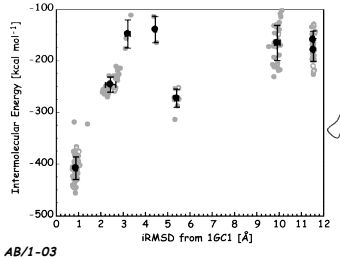
gp120-CD4 docking from mutagenesis data

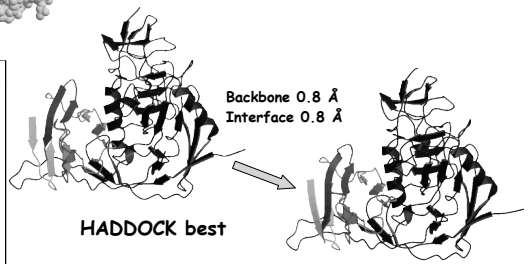
gp120:
4 AIR restraints (active)
19 additional residues
at interface (passive)



CD4:
7 AIR restraints (active)
15 additional residues
at interface (passive)







HADDOCK best

Target (1GC1)
(Kwong *et al. Nature* 1998)

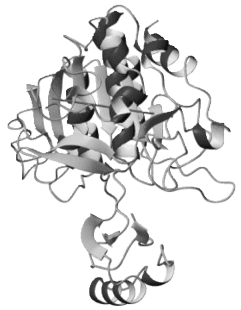
AB/1-03

Outline

- Introduction
- Methodology
- Validation
- Application
  - Subtilisin-CI2
- Conclusions & Perspectives

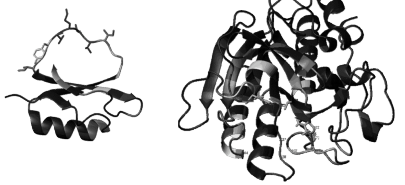
AB/1-03

Subtilisin (savinase)-CI2 complex

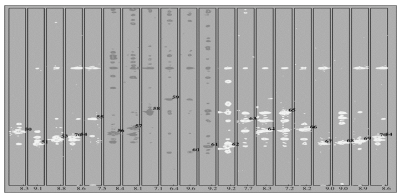


1SNI: Subtilisin Novo-CI2 complex
(C.A. McPhalen & M. N. G. James (1998), *Biochemistry* **27**, 6582)

□□(^{15}N)
0ppm >2ppm



^{15}N-NOESY-HSQC on U-[^{15}N,^{13}C, ^{2}H] CI2-subtilisin

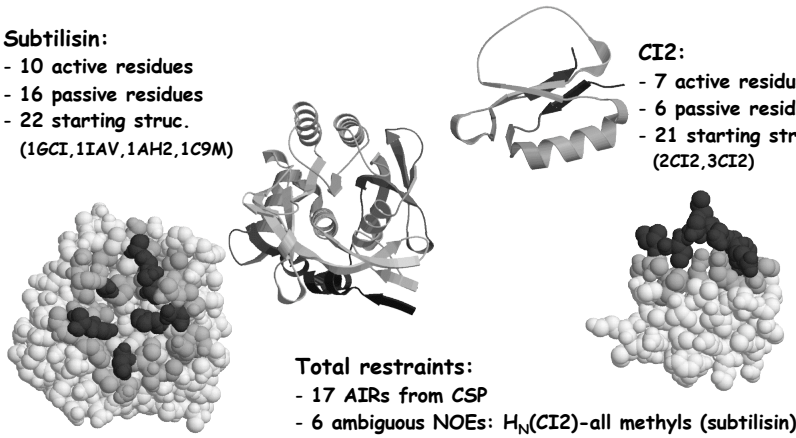


AB/1-03

Subtilisin-CI2: AIRs definition

Subtilisin:
- 10 active residues
- 16 passive residues
- 22 starting struc. (1GCI,1IAV,1AH2,1C9M)

CI2:
- 7 active residues
- 6 passive residues
- 21 starting struc. (2CI2,3CI2)

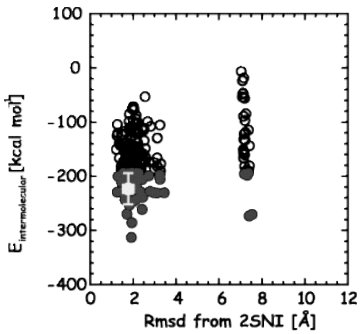


Total restraints:
- 17 AIRs from CSP
- 6 ambiguous NOEs: H_N(CI2)-all methyls (subtilisin)

AB/1-03

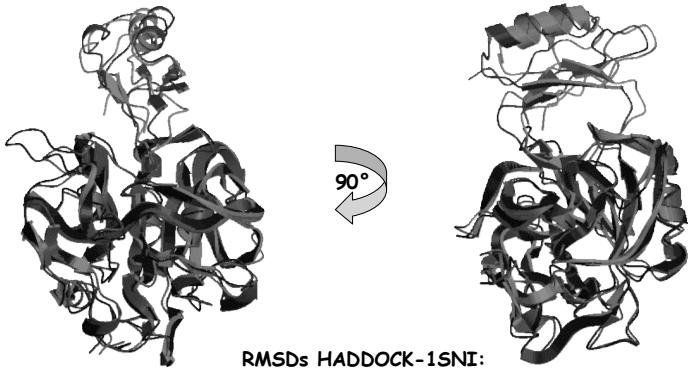
Subtilisin-CI2: HADDOCK results

- Analysis of the 50 lowest energy structures gives only one cluster containing 27 structures



AB/1-03

Subtilisin-CI2: HADDOCK vs homologuous crystal structure 1SNI



RMSDs HADDOCK-1SNI:
- backbone both: 1.5 Å
- interface only: 1.7 Å

AB/1-03

Outline

- Introduction
- Methodology
- Validation
- Application
- Conclusions & perspectives

AB/1-03

Conclusions & perspectives

- High ambiguity restraints (AIRs) are able to drive the docking process
  - Limit the number of solutions compared to ab-initio docking
  - Allow the inclusion of any type of information (NMR, mutagenesis, cross-linking, …)
- Flexible docking
- Allow the study of weak and transient complexes
- Genome -> proteome -> interactosome !!!

Acknowledgements

HADDOCK

- Cyril Dominguez
- Patrick Fuchs
- Rolf Boelens

Subtilisin-CI2

- Carine van Heijenoort
- Leo Koharudin
- Klaartje Houben
- Rick Bott (GCI)
- Grant Granshaw (GCI)
- Rolf Boelens

€€ NWO

AB/1-03

The End.

Thank you for your attention!

HADDOCK online: www.nmr.chem.uu.nl/haddock
Poster #403 (Saturday)

AB/1-03