

# Afonso M.S. Duarte

ITQB-UNL, Portugal

Laboratory of Biological Energy Transduction

<http://www.itqb.unl.pt/labs/biological-energy-transduction>

EGI Champion

[http://www.egi.eu/community/egi\\_champions/Afonso\\_Duarte.html](http://www.egi.eu/community/egi_champions/Afonso_Duarte.html)



INSTITUTO  
DE TECNOLOGIA  
QUÍMICA E BIOLÓGICA  
/UNL

Knowledge Creation

FCT

Fundação para a Ciência e a Tecnologia  
MINISTÉRIO DA CIÊNCIA, INOVAÇÃO E DO ENSINO SUPERIOR



# **Structural Biology in the Grid**

*a case study*

**The importance of molecular  
modeling applications**

**The role of the Grid**

**Some examples**

# Proteins are biomolecules that play key functions in living cells

Proteins consist of one or more chains of amino acid residues connected in a linear fashion.

*Chain 1*



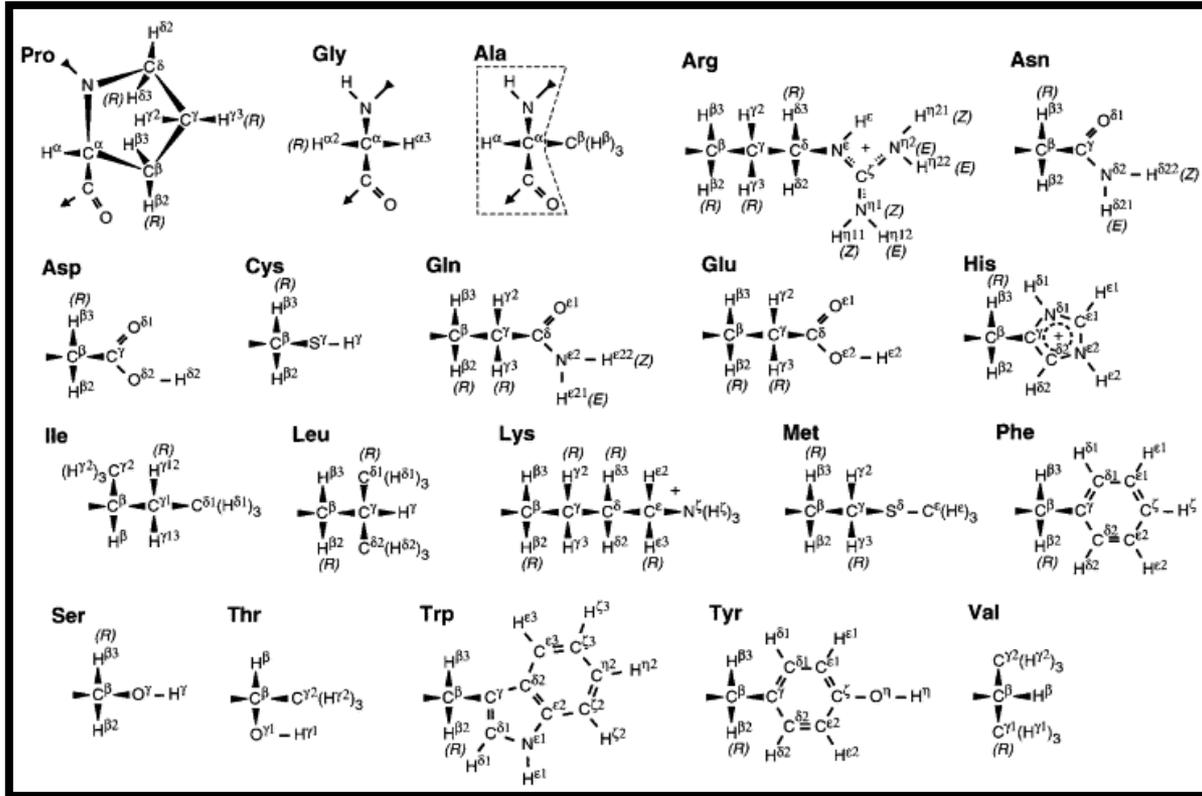
These chains can then fold and form different 3D structures.

Proteins perform an array of functions within living organisms, and are center to almost all processes in life

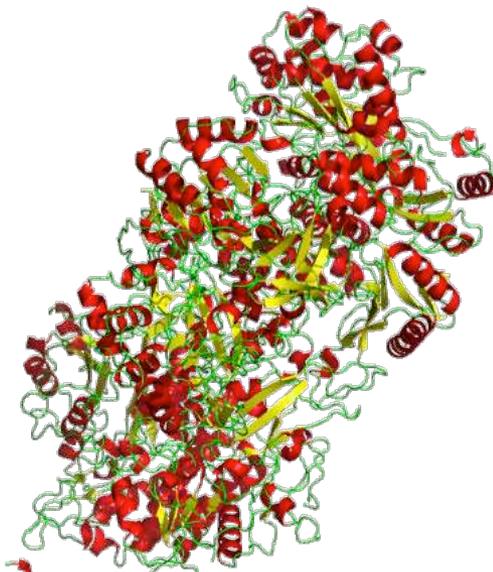
**In Nature there are 20 different amino acids that build-up proteins**  
(<http://www.bmr.b.wisc.edu/referenc/commonaa.php>)

**These have different chemical properties that allow the build up of proteins with different 3D structures and functions in cells**

# 20 amino acids can be combined into proteins with different sizes, shapes and functions



← 20 ≠ amino acids



**Ubiquitin**



**GFP**



**Complex I from the respiratory chain**

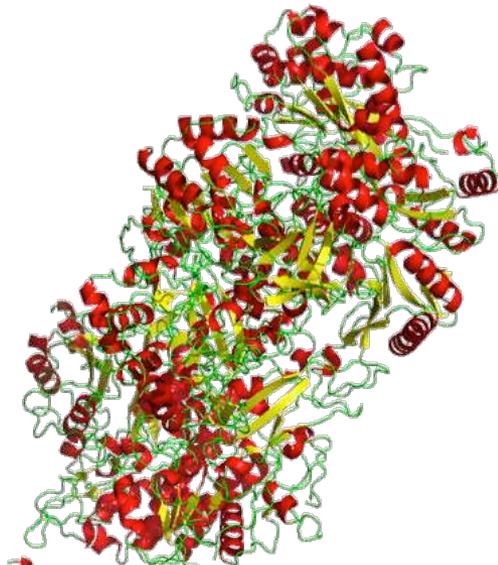
20 Å

# Protein Structure

- Protein structures are determined by techniques that provide atomic resolution (i.e. where atoms are individually detected and a three-dimensional model can be constructed - resolution below 4 Å)

– X-Ray Crystallography

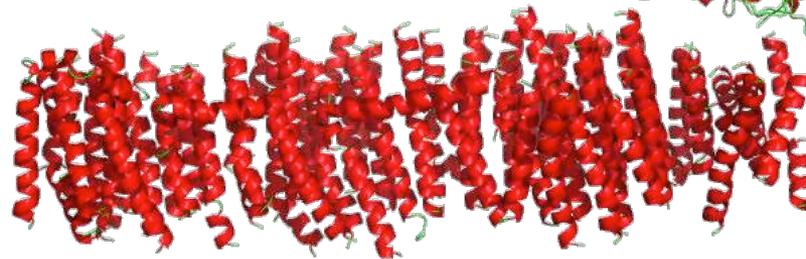
– Nuclear Magnetic Resonance (NMR)  
Spectroscopy



**Ubiquitin**



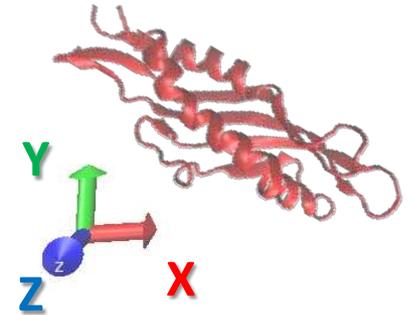
**GFP**



**Complex I from the respiratory chain**

20 Å

# The PDB format



Record type

Number in the pdb file

Type of atom

Amino acid

Chain name

Number in the protein sequence

Other molecular properties

ATOM	1	N	TRP	B	11	17.076	76.256	33.096	1.00	92.88	N
ATOM	2	CA	TRP	B	11	15.670	76.465	33.549	1.00	92.65	C
ATOM	3	C	TRP	B	11	15.616	77.167	33.795	1.00	91.65	C
ATOM	4	O	TRP	B	11	14.610	77.349	35.529	1.00	91.17	O
ATOM	5	CB	TRP	B	11	15.093	75.205	33.755	1.00	93.74	C
ATOM	6	CG	TRP	B	11	13.899	75.145	34.077	1.00	95.26	C
ATOM	7	CD1	TRP	B	11	12.778	74.223	34.753	1.00	95.40	C
ATOM	8	CD2	TRP	B	11	12.608	76.144	33.606	1.00	95.71	C
ATOM	9	CE2	TRP	B	11	11.319	75.758	34.036	1.00	96.10	C
ATOM	10	CE3	TRP	B	11	12.747	77.326	32.862	1.00	95.79	C
ATOM	11	NE1	TRP	B	11	11.451	74.585	34.733	1.00	95.92	N
ATOM	12	CZ2	TRP	B	11	10.175	76.516	33.749	1.00	96.64	C
ATOM	13	CZ3	TRP	B	11	11.610	78.079	32.577	1.00	95.43	C
ATOM	14	CH2	TRP	B	11	10.342	77.669	33.019	1.00	96.33	C
ATOM	15	N	VAL	B	12	16.651	78.158	35.028	1.00	90.37	N

**Coordinates**

X Z Y

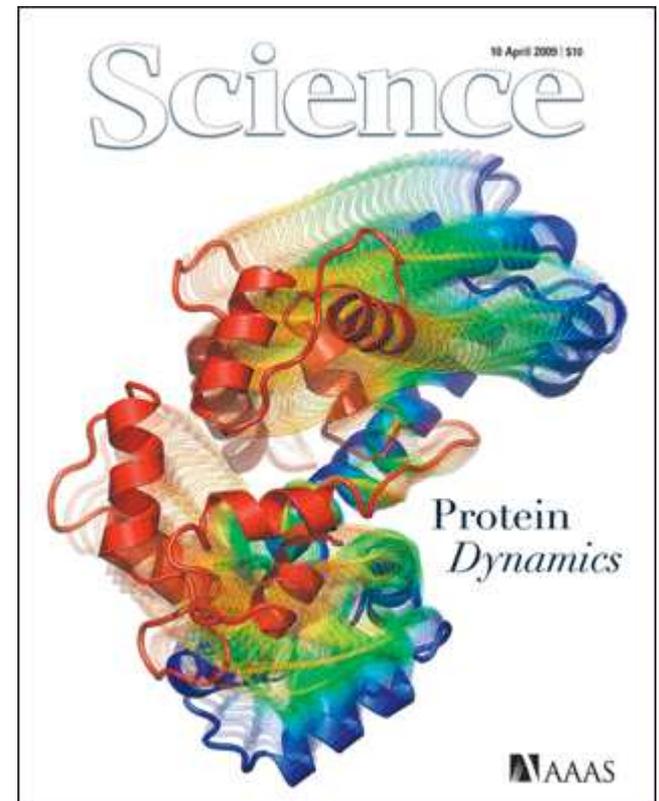
# Protein dynamics

*“what allows proteins to work”*

Dynamics are key for protein function

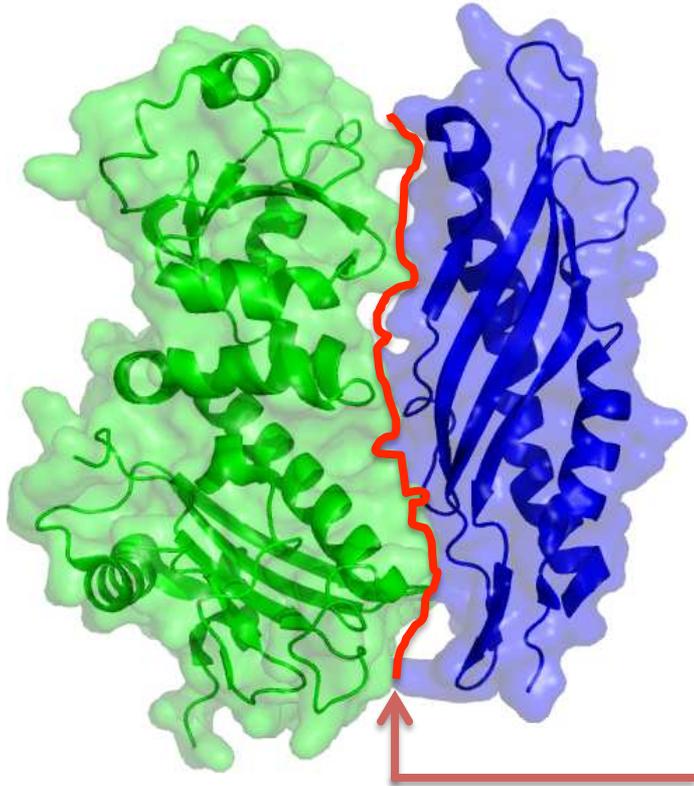
They allow structural changes fundamental for :

- Protein-protein interactions
- Protein-drug interaction
- Activation of enzymatic activity
- ...



# Protein – Protein interactions

*“The way proteins communicate between them”*



Key for protein regulation in cells:

- Activation
- Inhibition
- Intermediary step in cellular processes
- ...

Binding interface (i.e. binding hotspot)

Interaction between **Hsp90** and **Aha1** proteins.

Molecular machine involved in the cell quality control process (**Hsp90**) and its regulator (**Aha1**)

# Experimental techniques

NMR spectroscopy and X-Ray crystallography provide atomic level resolution information on the different steps of protein dynamics and protein-protein interactions

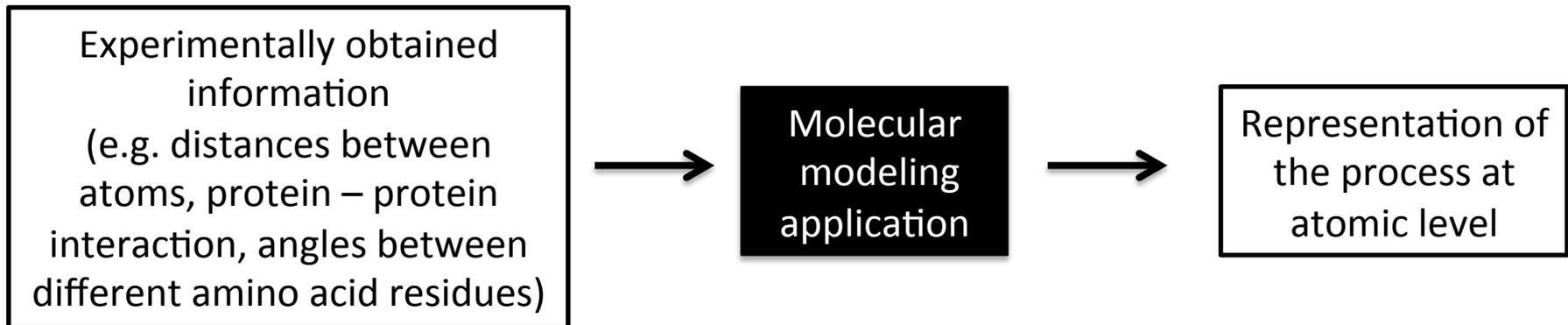
From these one obtains:

***Structural constrains (e.g. distances, angles)***

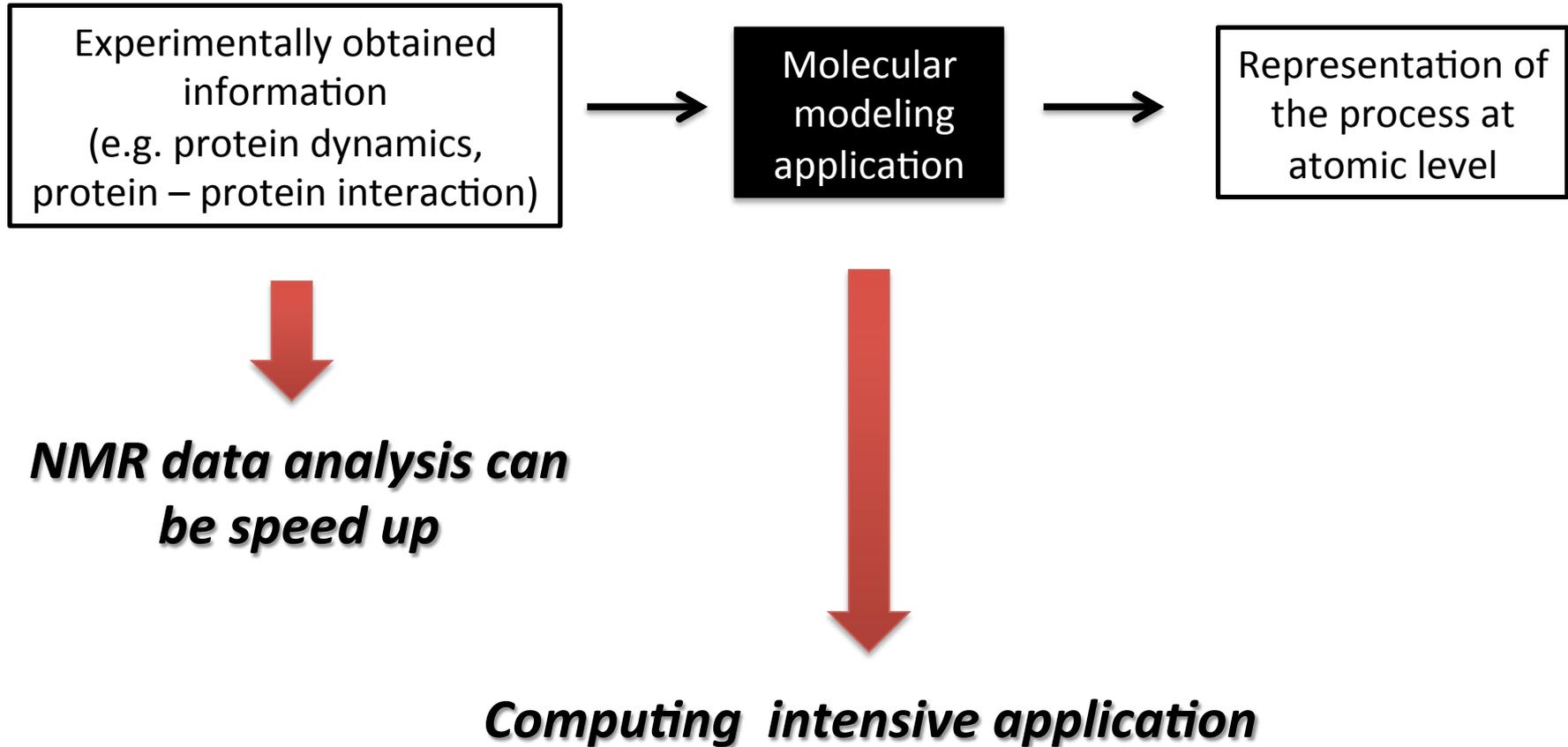
***Snapshots of different steps of the process***

# Computational methods

To obtain a picture of the process (protein dynamics and protein-protein interactions) one uses these constrains as input for the computation of protein models that represent the process under study



# Why Grid Computing?



# The Virtual Organization WeNMR

*“WeNMR is currently the largest Virtual Organization (VO) in life sciences. With its large and worldwide user community, WeNMR has become the first Virtual Research Community officially recognized by the European Grid Infrastructure (EGI).”*

<http://www.wenmr.eu/wenmr/>



**we-nmr**  
A worldwide e-Infrastructure for NMR and structural biology

Login | Register

Home WeNMR Services Market & Resources Support Access

WeNMR is both a three years project funded under the European Commission's 7th Framework Programme (e-Infrastructure RI-261571) and a Virtual Research Community supported by EGI, the largest one within the life science area. WeNMR aims at bringing together complementary research teams in the structural biology and life science area into a virtual research community at a worldwide level and provide them with a platform integrating and streamlining the computational approaches necessary for NMR and SAXS data analysis and structural modelling.

[Get Started >>](#)  
Harness the power of the GRID

# The Virtual Organization WeNMR

“The WeNMR project is a European Union funded international effort to **streamline** and **automate** analysis of Nuclear Magnetic Resonance (NMR) and Small Angle X-Ray scattering (SAXS) imaging data for atomic and near-atomic resolution molecular structures. “

“To facilitate the use of NMR spectroscopy and SAXS in life sciences the WeNMR consortium has **established standard computational workflows and services** through **easy-to-use web interfaces**, while still retaining sufficient flexibility to handle more specific requests. “

# The Virtual Organization WeNMR

**Gaining access to the Grid and to the VO**

<http://www.wenmr.eu/wenmr/>

# HADDOCK

*High Ambiguity Driven biomolecular DOCKing*

based on biochemical and/or biophysical information.

HADDOCK

v2.1

[Main](#)

[Introduction](#)

[Project setup](#)

[Download](#)

[Installation](#)

[Publications](#)

[Software links](#)

[Manual](#)

[PDB files](#)

[AIR restraints](#)

[RDC restraints](#)

[DANI restraints](#)

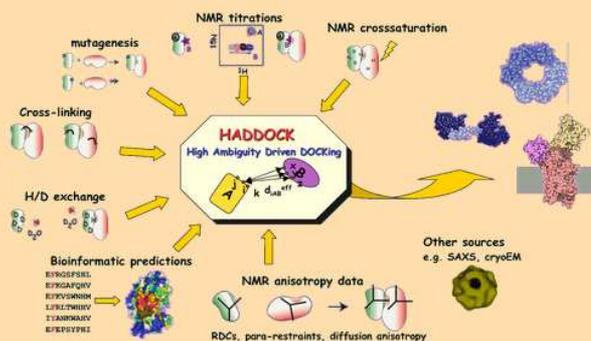
[A new project](#)

[The docking](#)

HADDOCK

*High Ambiguity Driven biomolecular DOCKing*

based on biochemical and/or biophysical information.



Version: 2.1 (February 10, 2010) ([changes](#))

Authors: Alexandre Bonvin, Utrecht University

Contact: Bijvoet Center for Biomolecular Research  
Padualaan 8, 3584 CH Utrecht, the Netherlands  
Email: [a.m.i.bonvin@uu.nl](mailto:a.m.i.bonvin@uu.nl)

home

HADDOCK  
Software web portal

[Home](#) [HADDOCK](#) [Whisky](#) [CPORT](#) [DNA](#) [SQUEEZE](#) [Publications](#) [HADDOCK](#) [Help](#) [Contact](#) [RSS](#)

WELCOME TO THE UTRECHT BIOMOLECULAR INTERACTION WEB PORTAL >>

The Utrecht Biomolecular Interactions software portal provides access to software tools developed in the Computational Structural Biology group / NMR Research Group of Utrecht University with a main focus on the characterization of biomolecular interactions. Please note that this site is in active development.

## HADDOCK WEB DOCKING



HADDOCK (High Ambiguity Driven protein-protein DOCKing) is an information-driven flexible docking approach for the modeling of biomolecular complexes. HADDOCK distinguishes itself from ab-initio docking methods in the fact that it encodes information from identified or predicted protein interfaces in ambiguous interaction restraints (AIRs) to drive the docking process. HADDOCK can deal with a large class of modelling problems including protein-protein, protein-nucleic acids and protein-ligand complexes. | [GO TO SERVICE >>](#)

## FCC CLUSTERING

EMSA based detection of protein-protein and protein-ligand complexes

PROFILE >>



Universiteit Utrecht

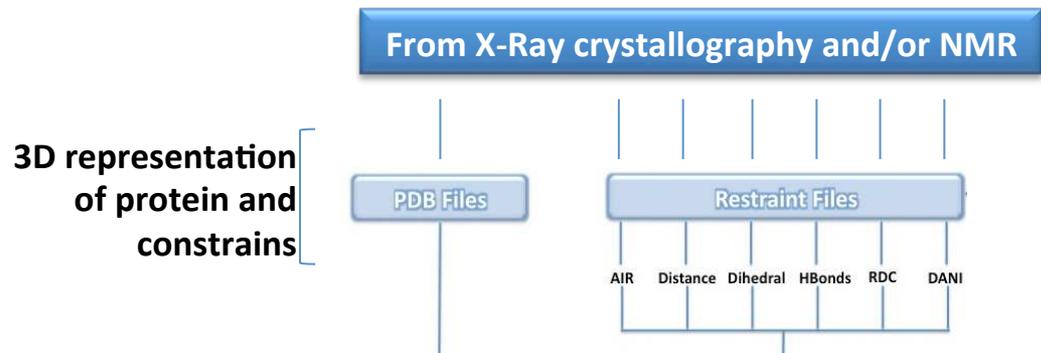


<http://www.nmr.chem.uu.nl/haddock/>

<http://haddock.science.uu.nl/>

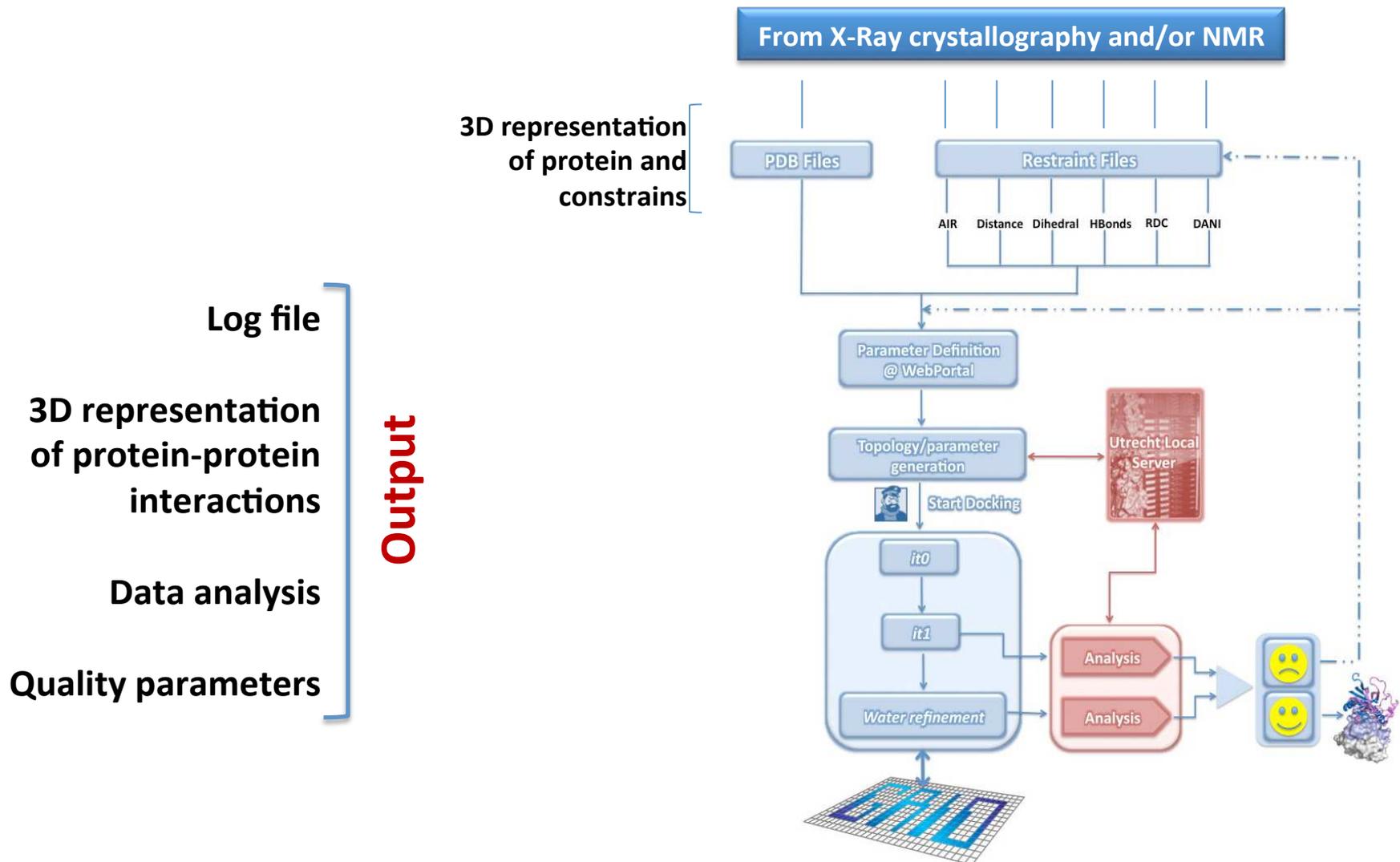
# Protein – Protein interactions

## *HADDOCK in the Grid*



# Protein – Protein interactions

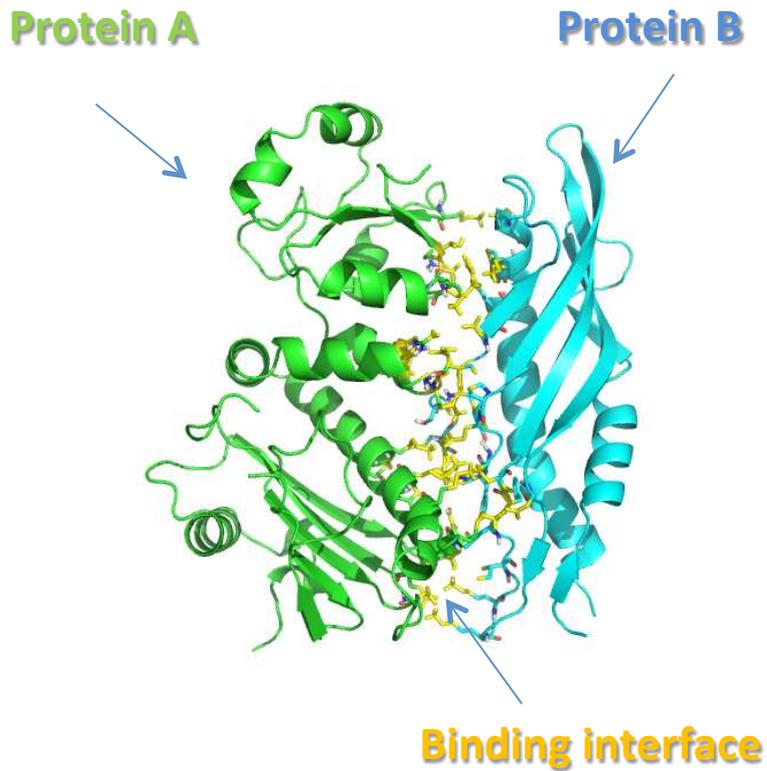
## *HADDOCK in the Grid*



# Demo

# Protein – Protein interactions

## *HADDOCK in the Grid*



**Protein – Protein interactions**

# GROMACS

*“perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles.”*

The screenshot shows the GROMACS website homepage. At the top left, the GROMACS logo is displayed with the tagline "FAST. FLEXIBLE. FREE." Below the logo, there are links for "Log in" and "Register". A search bar is present with a "Search" button. A "Main pages" dropdown menu is also visible. The main navigation menu on the left includes: Gromacs, About Gromacs, Developer Zone, Documentation, Downloads, Gromacs papers, Jobs, Project ideas, and Support. The main content area features a "Gromacs" heading with a key icon, followed by a row of six icons and their corresponding text: "get the source code" (with a folder icon), "search the mailing list" (with a magnifying glass icon), "online manual" (with a book icon), "git repository access" (with a Git logo icon), "Find us on Facebook" (with a Facebook logo icon), and a Google+ logo icon. Below this row, a paragraph of text describes the funding support for GROMACS development, mentioning the European Research Council, the Swedish Research Council, the Swedish Foundation for Strategic Research, the Swedish National Infrastructure for Computing, and the Swedish Foundation for International Cooperation in Research and Higher Education. It also mentions that several other grant agencies provide funding, including NIH and NSF in the US, and the DFG in Germany. A second paragraph states that GROMACS is part of the ScalaLife Competence Center (http://scalalife.eu) and that the center is being developed as a one-stop-shop for users and developers of Life Science software. Various training events on high-performance computing and optimization techniques for improving scalability and performance will be organized by the center. The ScalaLife logo is visible to the right of this paragraph. Below the text, there is a "Sticky" notice: "Can't Post to the mailing list? Read the 'important information' section on the mailing list page. Send a mail if you want to be registered as a contributor and be able to upload files or modify the webpage content." The "News" section follows, with a date of "July 5, 2013" and a headline: "GROMACS 4.6.3 is out! Download the source tarball, and check out the release notes."

<http://www.gromacs.org/>

# Protein dynamics

## ***GROMACS in the Grid***

Structure from X-Ray crystallography and/or NMR



Selection of force field

Solvation of protein structure

Energy minimization of the system

Position restrained equilibration

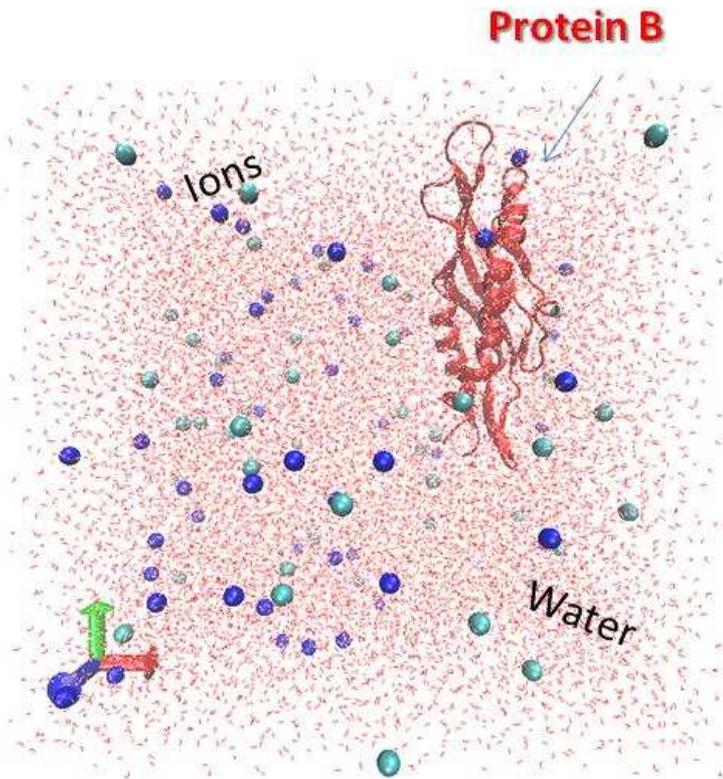
Molecular dynamics calculation

@ WeNMR / GRID

# Protein dynamics

## ***GROMACS in the Grid***

Structure from X-Ray crystallography and/or NMR



Selection of force field

Solvation of protein structure

Energy minimization of the system

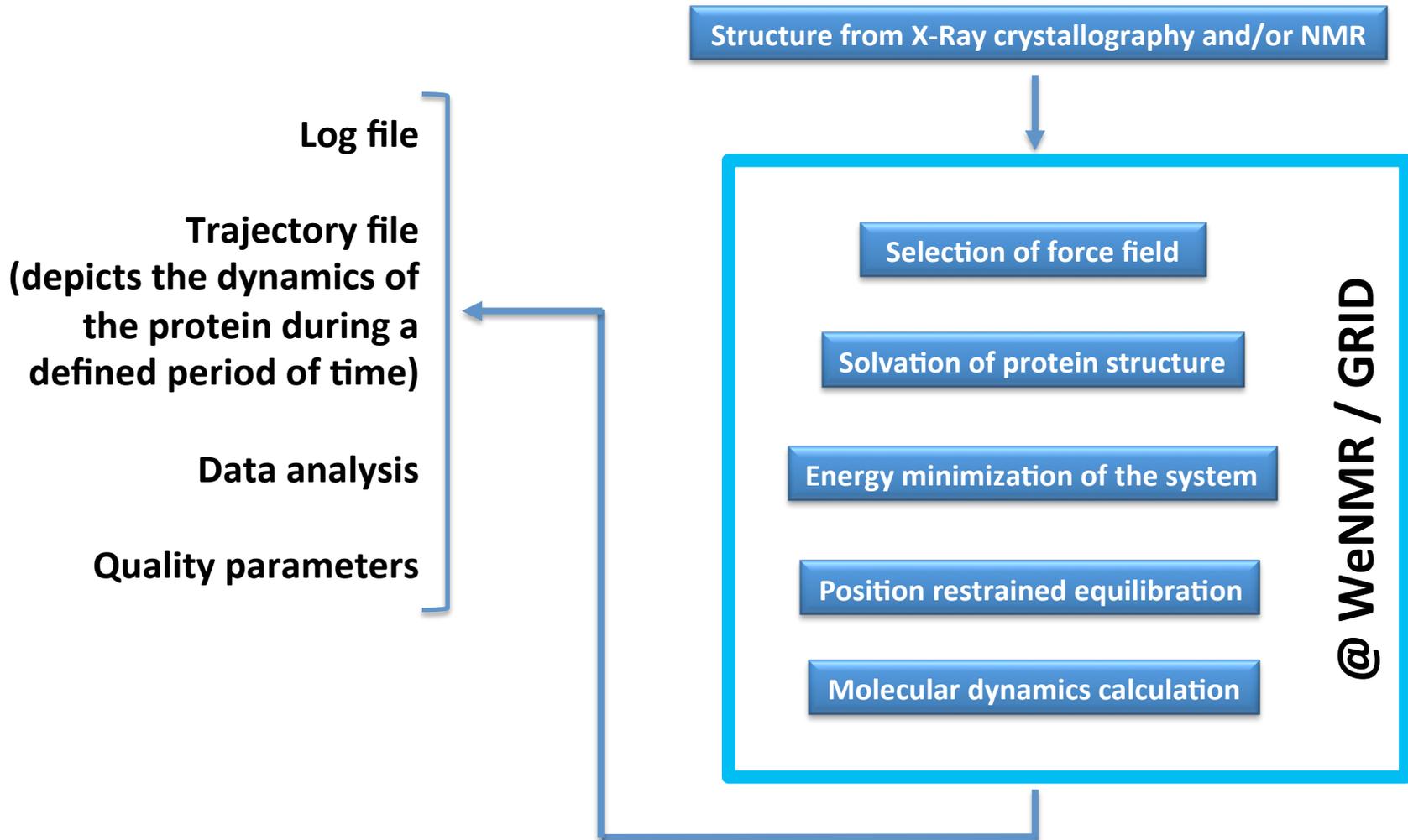
Position restrained equilibration

Molecular dynamics calculation

@ WeNMR / GRID

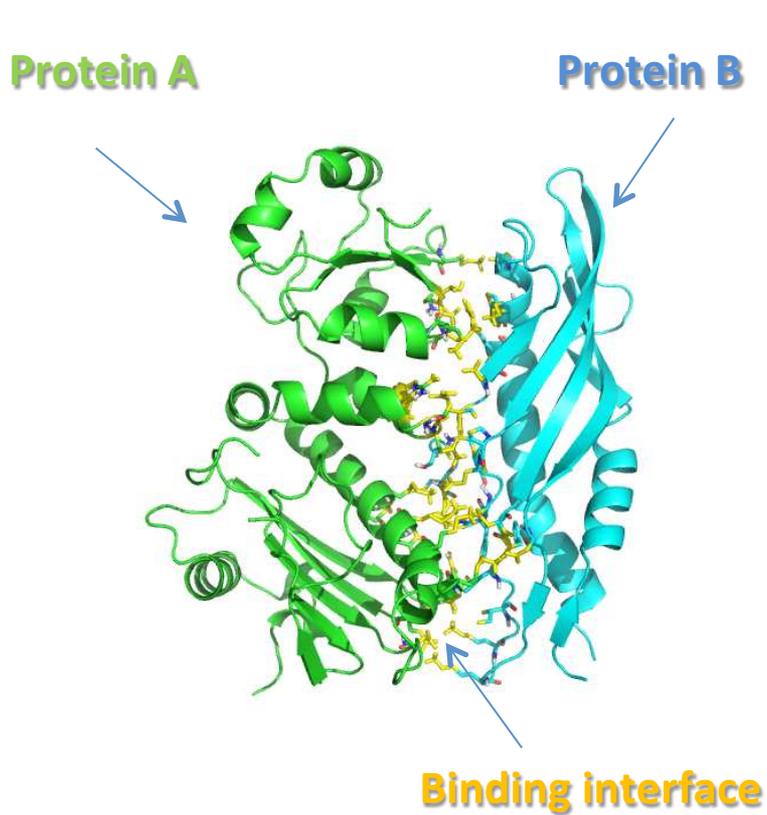
# Protein dynamics

## ***GROMACS in the Grid***

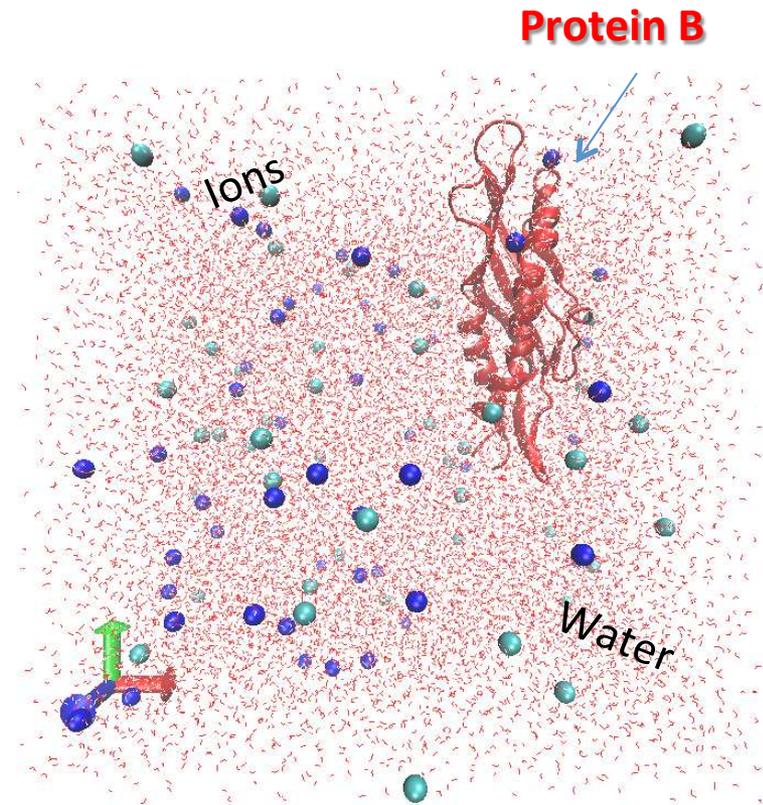


# Demo

# Two examples of the use of Grid application in Structural Biology



**Protein – Protein interactions**



**Protein dynamics**

# More Examples/Application in Structural Biology and other fields

esi Applications Database

Software for research communities

Welcome to the EGI Applications Database

The EGI Applications Database (AppDB) is a central service that stores and provides to the public information about:

- software solutions for scientists and developers to use,
- the programmers and the scientists who developed them, and
- publications derived from the registered solutions.

All software filed in the AppDB is ready to be used on the European Grid Infrastructure. Reusing software products from the AppDB means that scientists and software developers may find a solution than can directly be used on the European Grid Infrastructures without reinventing the wheel. ...read more

Top rated

Application	Category	Rating
SCMS-EMI	Science Gateways	★★★★★ 31
WatG Browser	Tools	★★★★★ 31
ASTRA-ANCIENT	Applications	★★★★★ 31
InSilicoLab	Science Gateways	★★★★★ 31
AutoDock Gateway	Science Gateways	★★★★★ 31

...see more

<http://appdb.egi.eu/>

<https://www.egi.eu/case-studies/index.html>



## The EGI Champions

EGI Champions are enthusiastic scientists using grid computing for their research and keen to go to conferences and spread the word about the benefits of working with EGI.



**Welcoming new users to the grid**

**Bridge users and developers**

<https://wiki.egi.eu/wiki/Champions>

[http://www.egi.eu/community/egi\\_champions/champions\\_network.html](http://www.egi.eu/community/egi_champions/champions_network.html)