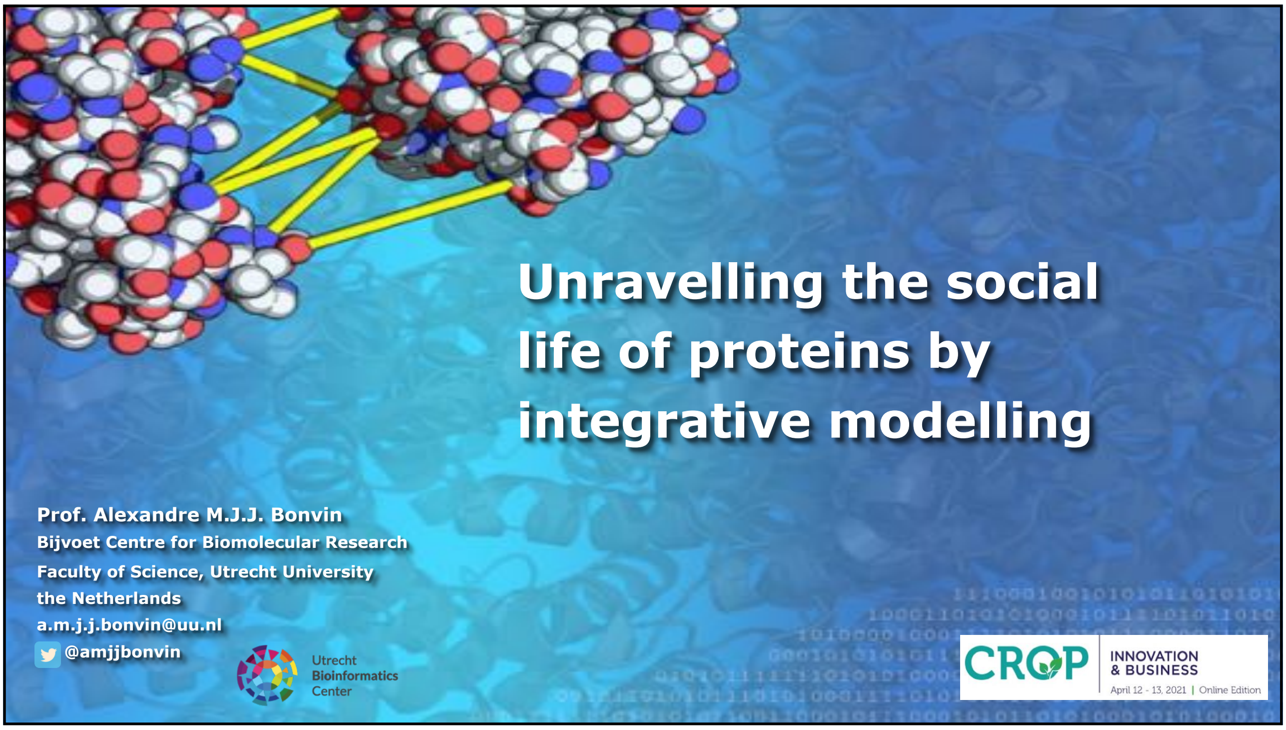
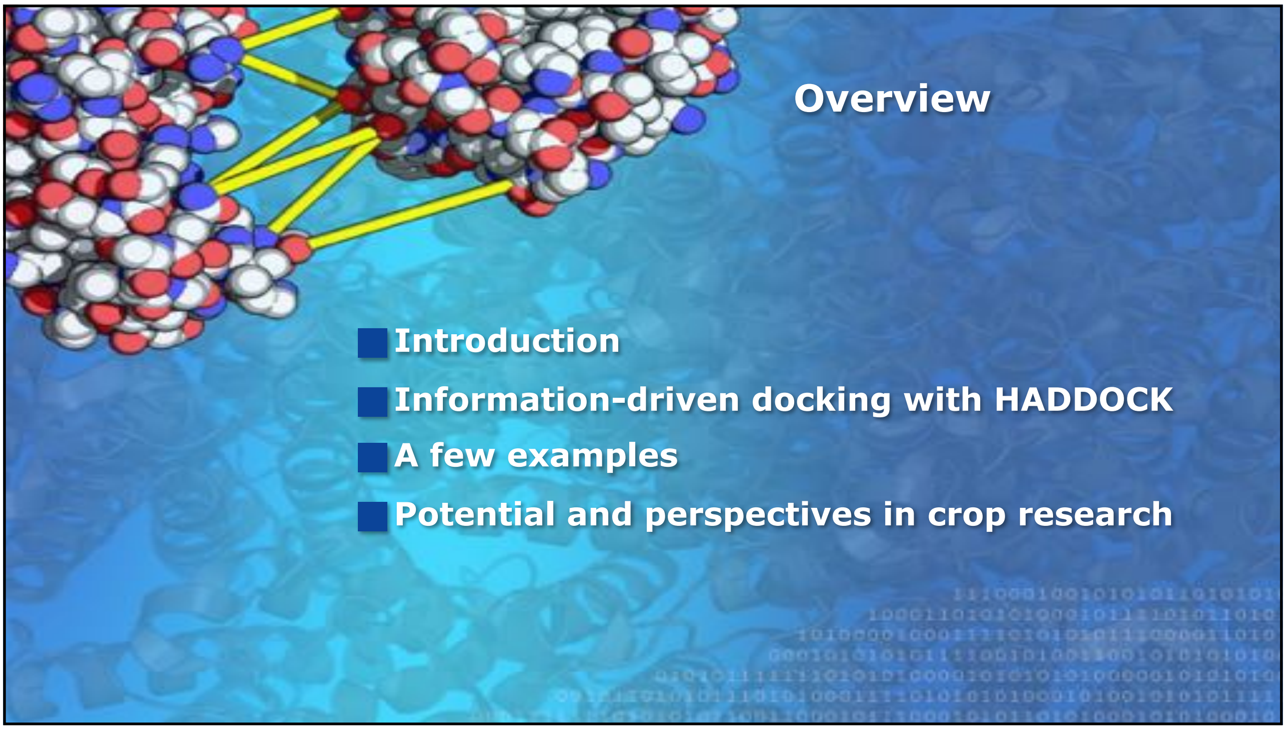


Unravelling the social life of proteins by integrative modelling

Prof. Alexandre M.J.J. Bonvin
Bijvoet Centre for Biomolecular Research
Faculty of Science, Utrecht University
the Netherlands
a.m.j.j.bonvin@uu.nl
@amjjbonvin

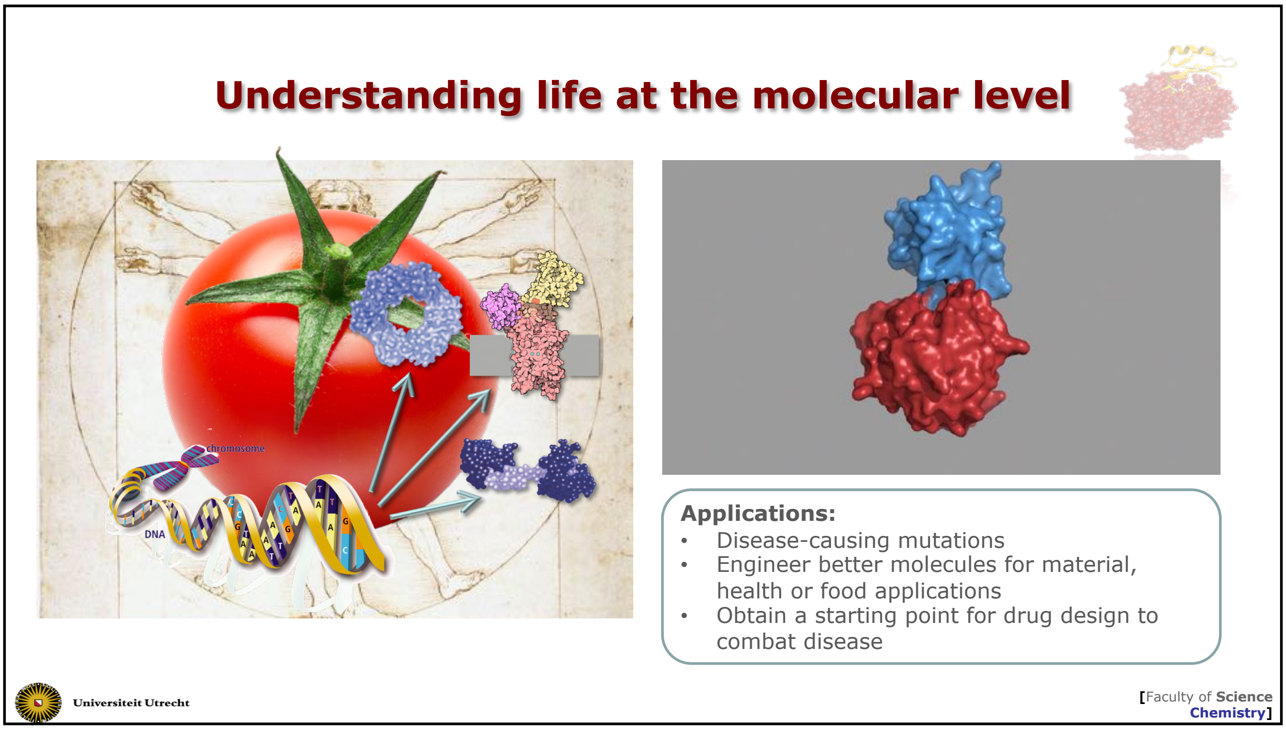
Utrecht
Bioinformatics
Center

CROP
INNOVATION
& BUSINESS
April 12 - 13, 2021 | Online Edition



Overview

- Introduction
- Information-driven docking with HADDOCK
- A few examples
- Potential and perspectives in crop research



Understanding life at the molecular level

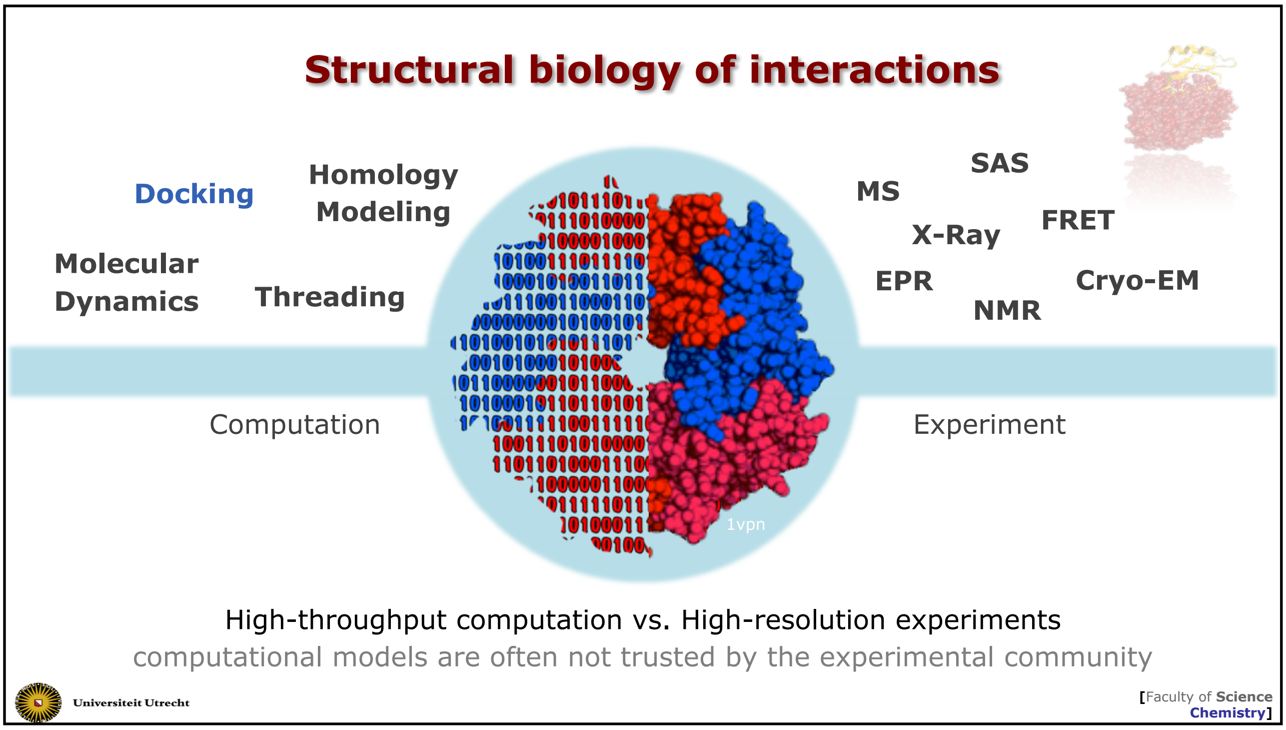
Applications:

- Disease-causing mutations
- Engineer better molecules for material, health or food applications
- Obtain a starting point for drug design to combat disease

Universiteit Utrecht

[Faculty of **Science**
Chemistry]

4



Structural biology of interactions

Docking

Homology Modeling

Molecular Dynamics

Threading

SAS

MS

X-Ray

FRET

EPR

Cryo-EM

NMR

Computation

Experiment

High-throughput computation vs. High-resolution experiments

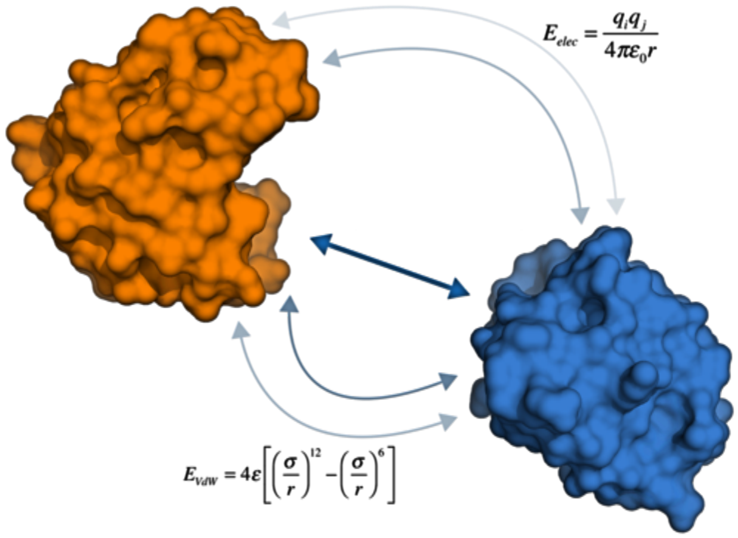
computational models are often not trusted by the experimental community

Universiteit Utrecht

[Faculty of **Science**
Chemistry]

5

Solving 3D puzzles by molecular docking

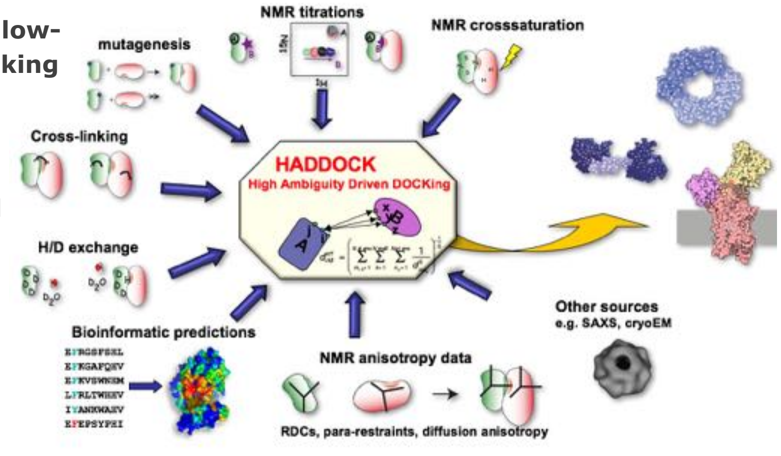


$$E_{elec} = \frac{q_i q_j}{4\pi\varepsilon_0 r}$$

$$E_{VdW} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

HADDOCK:
An integrative modeling platform

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



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

http://www.bonvinlab.org/software

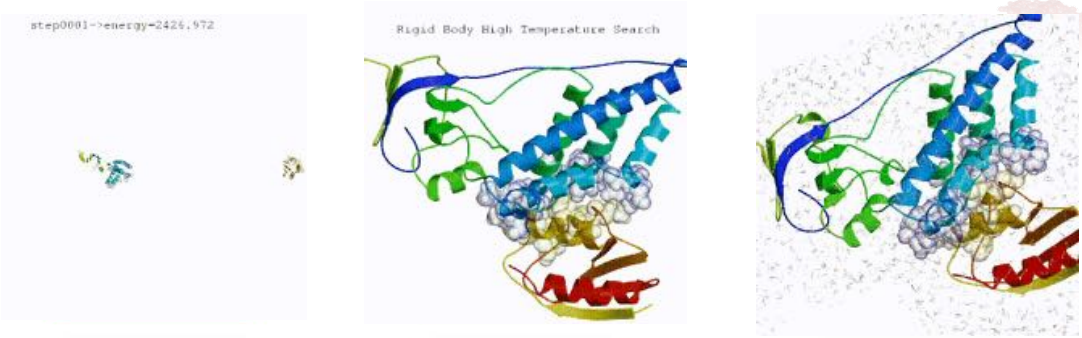
HADDOCK

it0

it1

itw

HADDOCK



it0

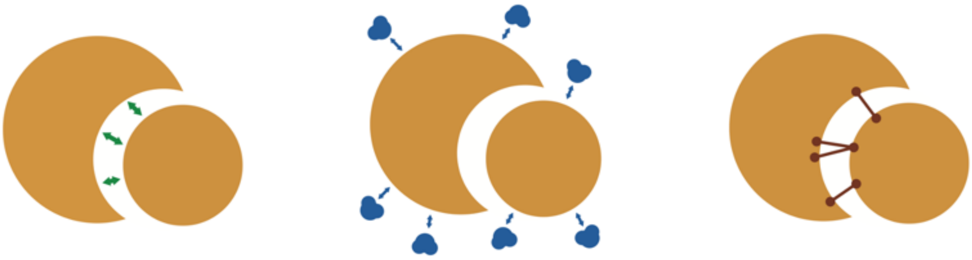
it1

itw

HADDOCK

Final score (in explicit solvent)

$$HS = 0.2\,E_{elec} + 1.0\,E_{vdw} + 1.0\,E_{desolv} + 0.1\,E_{AIR}$$

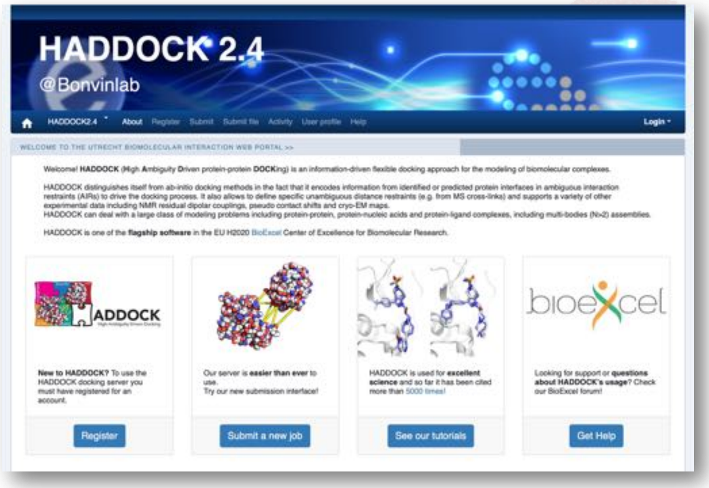


Haddock web portal

- > 21500 registered users
- > 330000 served runs since June 2008
- > 55% on the GRID

De Vries *et al.* Nature Prot. 2010

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



Welcome! **HADDOCK** (**H**igh **A**mbiguity **D**riven protein-protein **DOCK**ing) 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. It also allows to define specific unambiguous distance restraints (e.g. from MS cross-links) and supports a variety of other experimental data including NMR residual dipolar couplings, pseudo contact shifts and cryo-EM maps.
HADDOCK can deal with a large class of modeling problems including protein-protein, protein-nucleic acids and protein-ligand complexes, including multi-bodies (N>2) assemblies.

HADDOCK is one of the **flagship software** in the EU H2020 BioExcel Center of Excellence for Biomolecular Research.

New to HADDOCK? To use the HADDOCK docking server you must have registered for an account.

Our server is **easier than ever** to use.
Try our new submission interface!

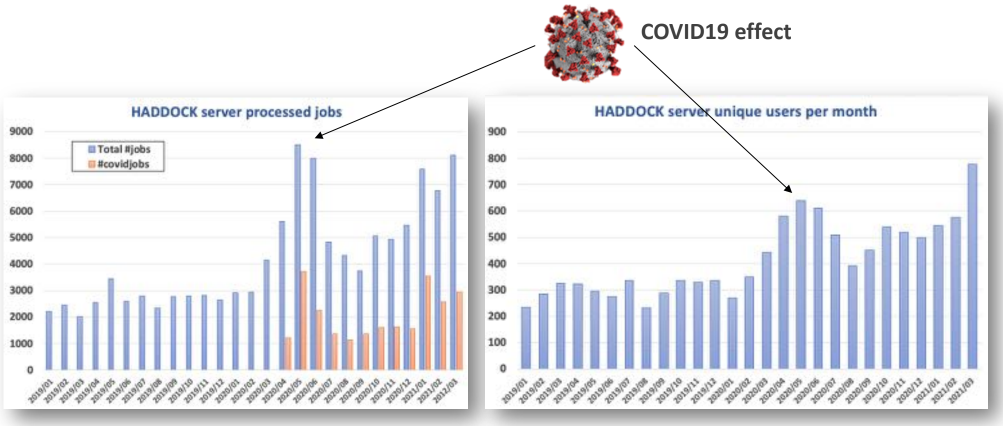
HADDOCK is used for **excellent science** and so far it has been cited more than 5000 times!

Looking for support or **questions about HADDOCK's usage**? Check our BioExcel forum!

https://wenmr.science.uu.nl

HADDOCK: Meeting the increased demand

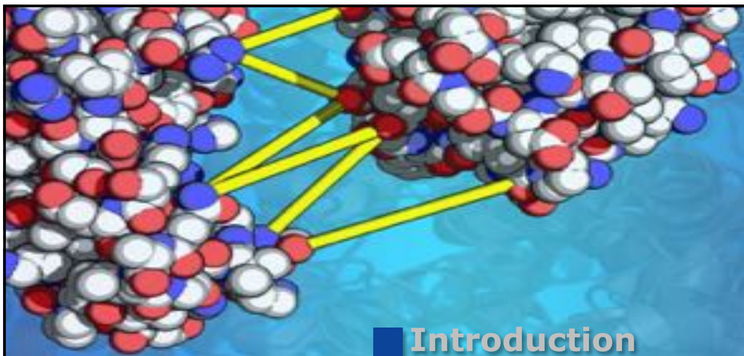
The HADDOCK workflow machinery was modified to improve its efficiency and meet the increased demand (allows to run more processes in // - relevant toward exascale).


COVID19 effect

Universiteit Utrecht

[Faculty of **Science**
Chemistry]

12



Overview

- Introduction
- Information-driven docking with HADDOCK
- **A few examples**
- Potential and perspectives in crop research

13

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

nature COMMUNICATIONS

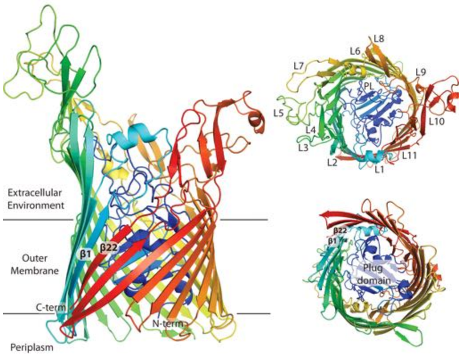
ARTICLE

Received 21 Jan 2016 | Accepted 21 Sep 2016 | Published 31 Oct 2016 DOI: 10.1038/ncomms13308 OPEN

Structure of the bacterial plant-ferredoxin receptor FusA

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

- **Iron import machinery in gram-negative bacteria** (pectobacterium, plant pathogen)
- **First complete crystal structure of such a receptor**

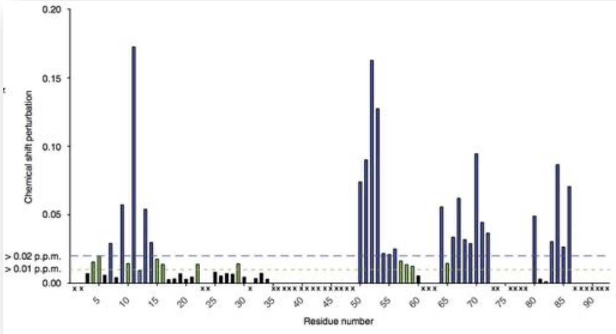


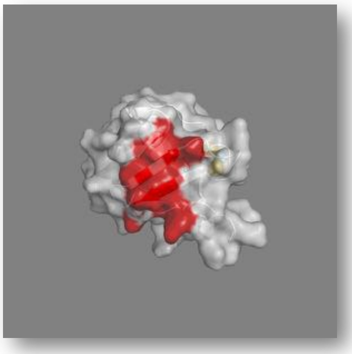
Universiteit Utrecht

[Faculty of **Science**
Chemistry]

Docking strategy

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





Universiteit Utrecht

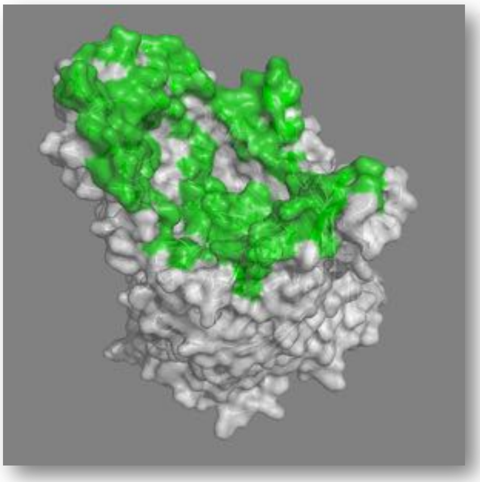
[Faculty of **Science**
Chemistry]

Docking strategy

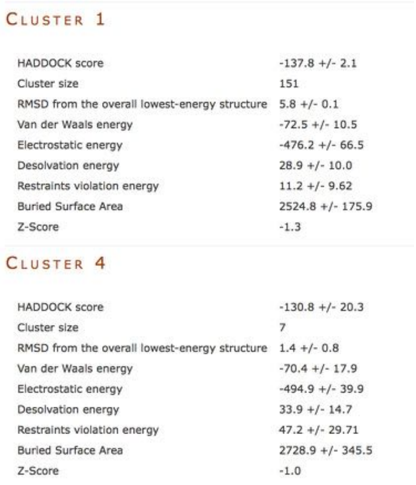
- **No info for FusA (expect that the binding occurs in the extracellular part)**

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

  → Definition of passive refined in a second docking run



Model of the FusA-ferredoxin complex

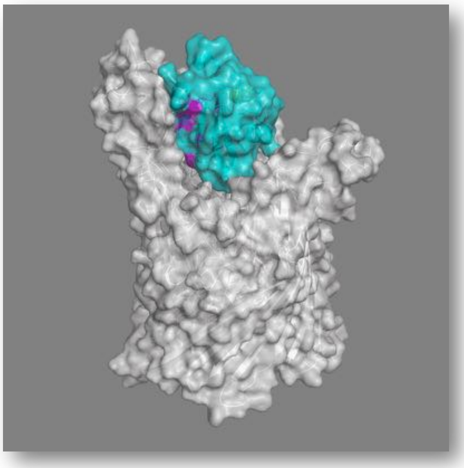


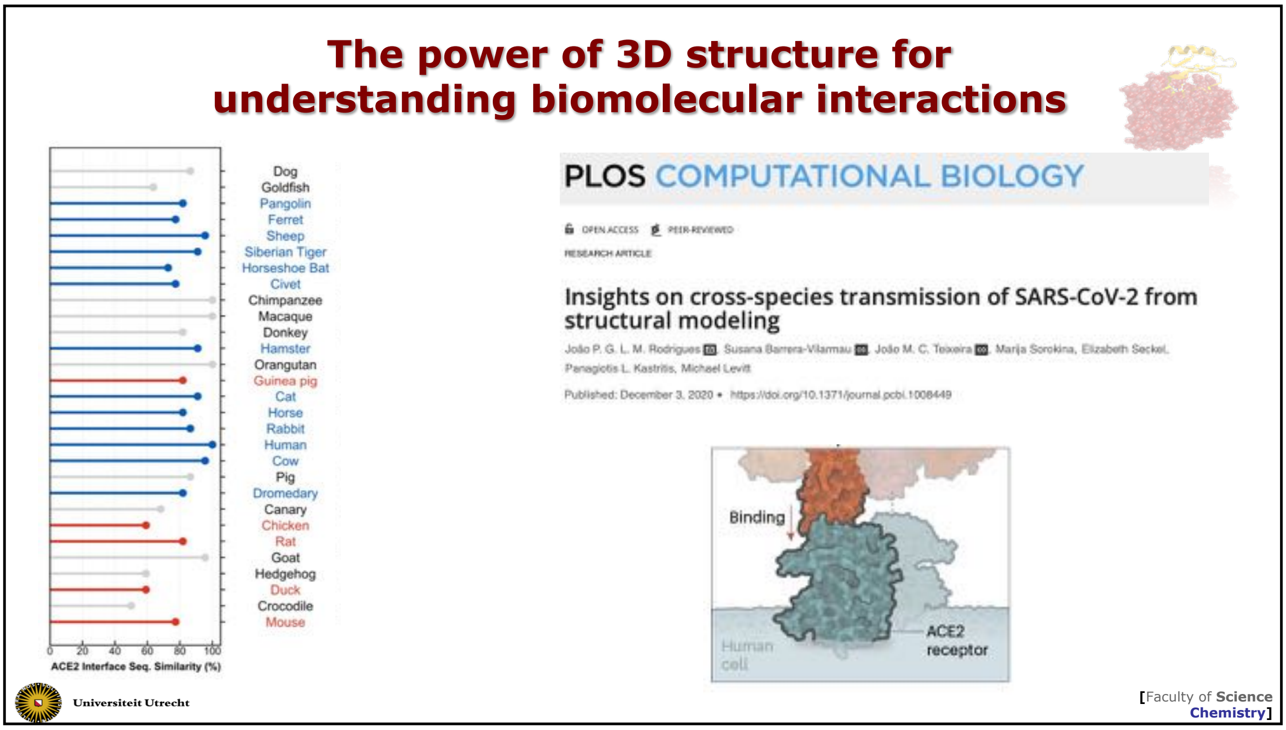
CLUSTER 1

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

CLUSTER 4

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





The power of 3D structure for understanding biomolecular interactions

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

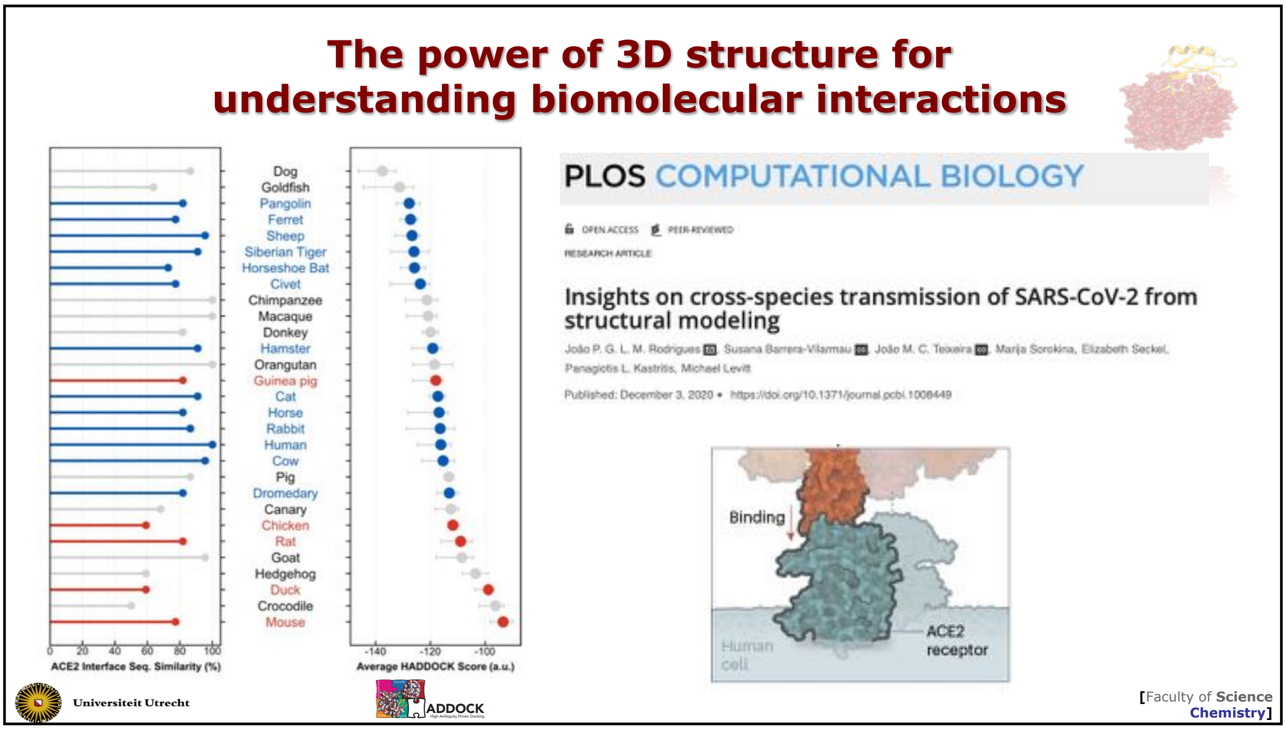
Insights on cross-species transmission of SARS-CoV-2 from structural modeling

João P. G. L. M. Rodrigues, Susana Barrera-Vilarmau, João M. C. Teixeira, Marija Sorokina, Elizabeth Seckel, Panagiotis L. Kastritis, Michael Levitt

Published: December 3, 2020 • https://doi.org/10.1371/journal.pcbi.1008449

Universiteit Utrecht

[Faculty of Science **Chemistry**]



The power of 3D structure for understanding biomolecular interactions

PLOS COMPUTATIONAL BIOLOGY

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Insights on cross-species transmission of SARS-CoV-2 from structural modeling

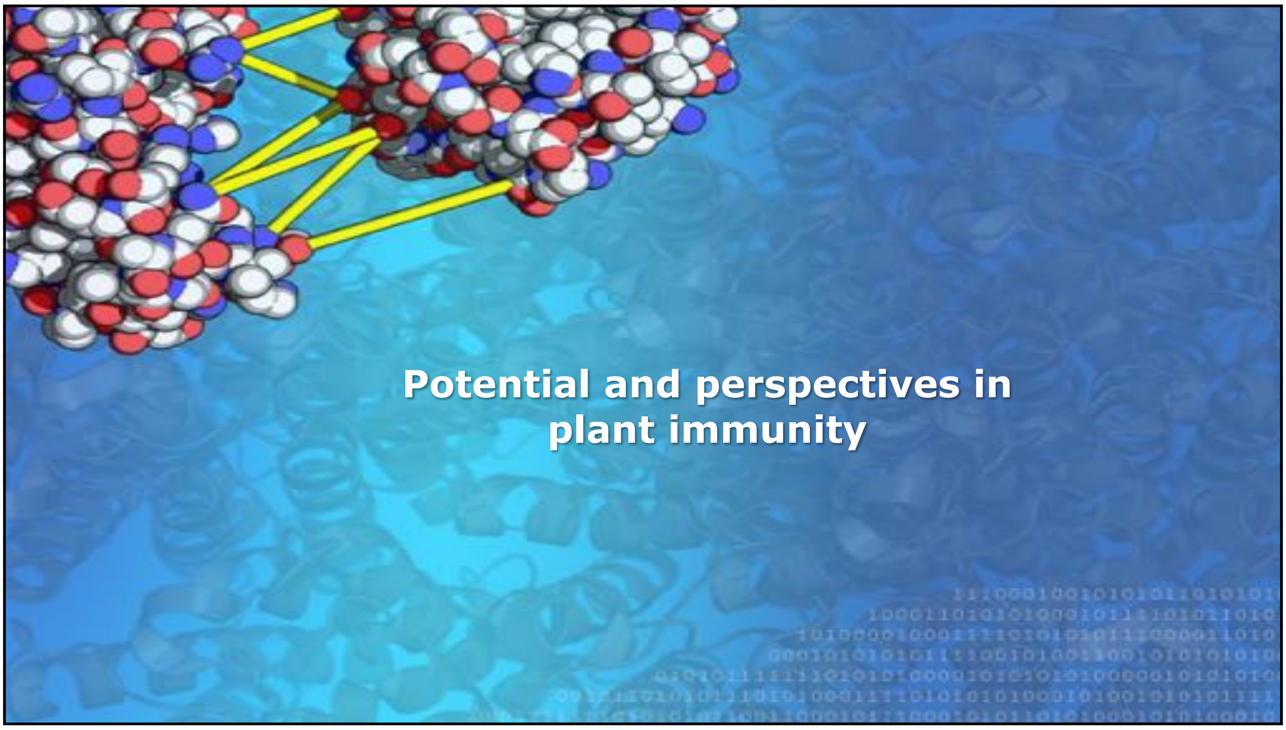
João P. G. L. M. Rodrigues, Susana Barrera-Vilarmau, João M. C. Teixeira, Marija Sorokina, Elizabeth Seckel, Panagiotis L. Kastritis, Michael Levitt

Published: December 3, 2020 • https://doi.org/10.1371/journal.pcbi.1008449

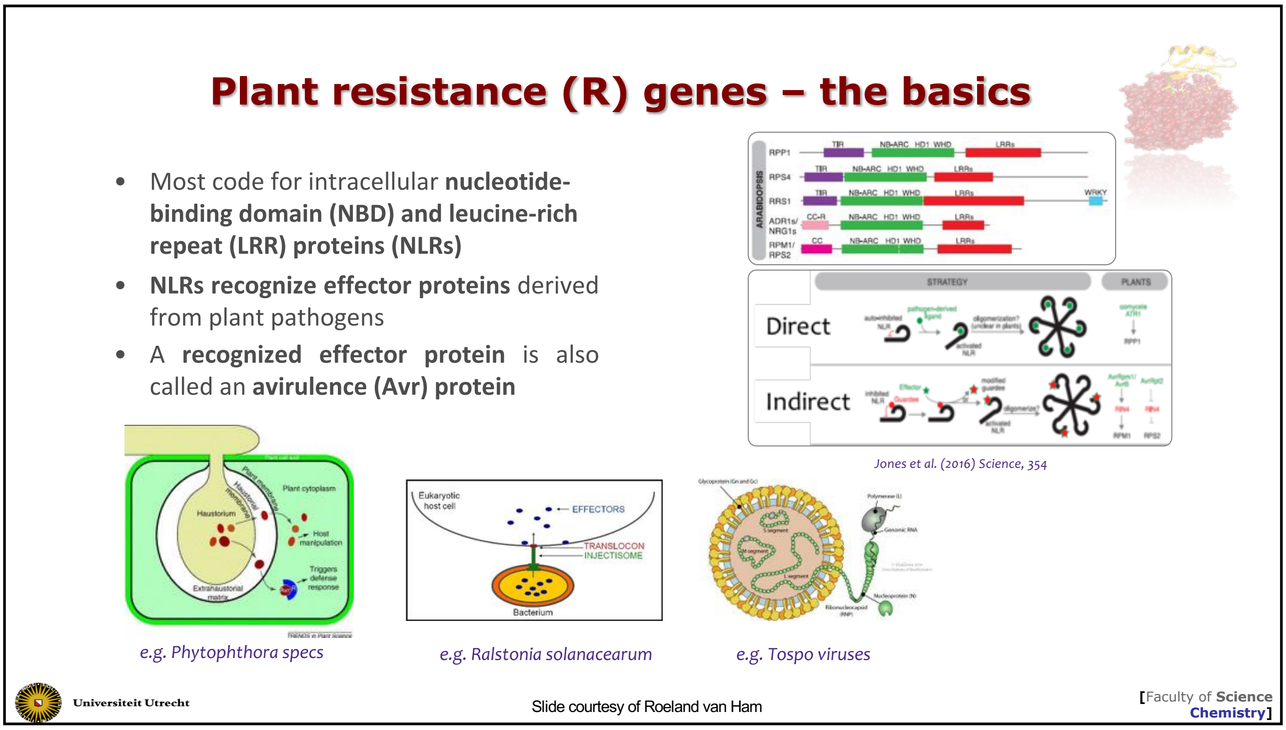
Universiteit Utrecht

HADDOCK

[Faculty of Science **Chemistry**]



Potential and perspectives in plant immunity



Plant resistance (R) genes – the basics

- Most code for intracellular **nucleotide-binding domain (NBD) and leucine-rich repeat (LRR) proteins (NLRs)**
- **NLRs recognize effector proteins** derived from plant pathogens
- A **recognized effector protein** is also called an **avirulence (Avr) protein**

Jones et al. (2016) Science, 354

e.g. Phytophthora specs

e.g. Ralstonia solanacearum

e.g. Tospo viruses

Universiteit Utrecht

Slide courtesy of Roeland van Ham

[Faculty of Science Chemistry]

Potential and perspectives in plant immunity

- Predict complexes between R proteins (e.g. LRR domains) and pathogen components (from sequence and/or structure)

The world this week

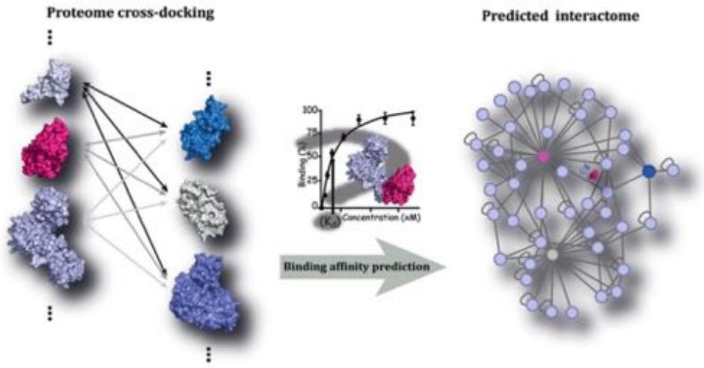
News in focus

A protein's function is determined by its 3D shape.

'IT WILL CHANGE EVERYTHING': AI MAKES GIGANTIC LEAP IN SOLVING PROTEIN STRUCTURES

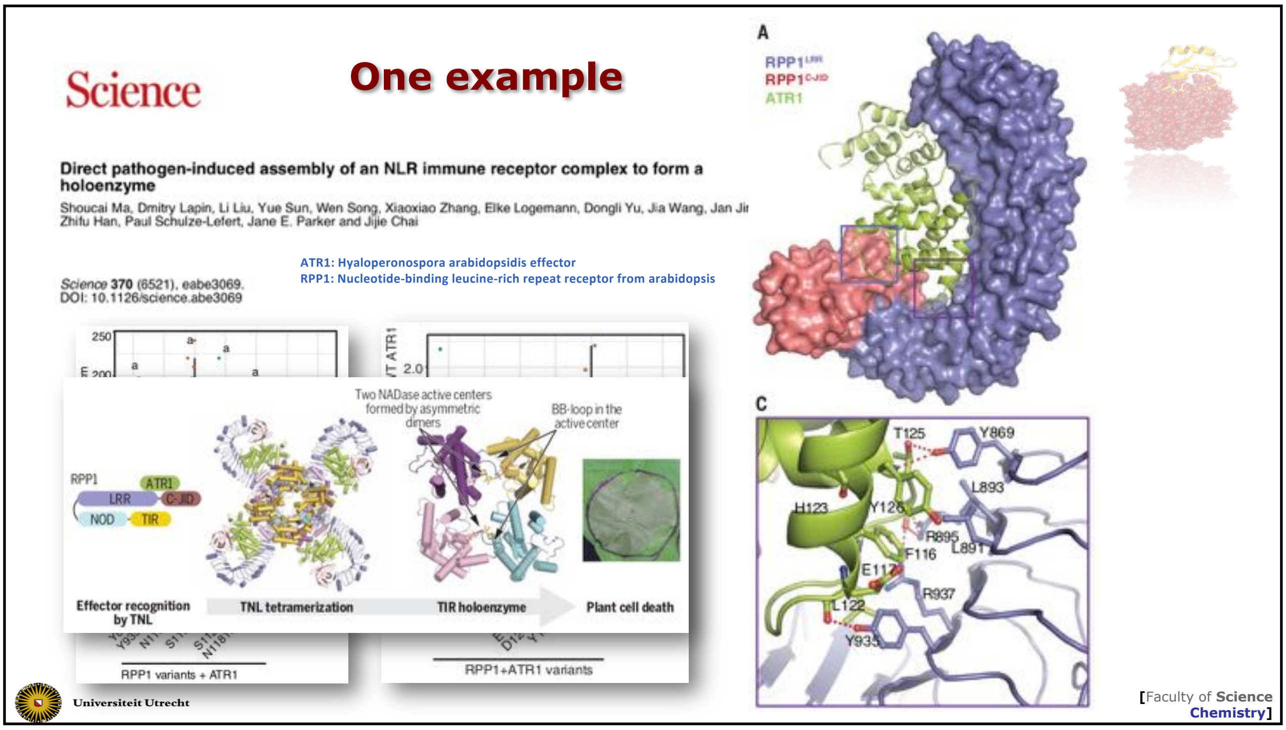
DeepMind's program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Nature | Vol 588 | 10 December 2020 | **203**



Potential and perspectives in plant immunity

- Predict complexes between immunity proteins (e.g. LRR domains) and pathogen components
- Given the knowledge of pairs of protein interacting, identify key residues for the interaction (from sequence and/or structure)



One example

Science

Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme

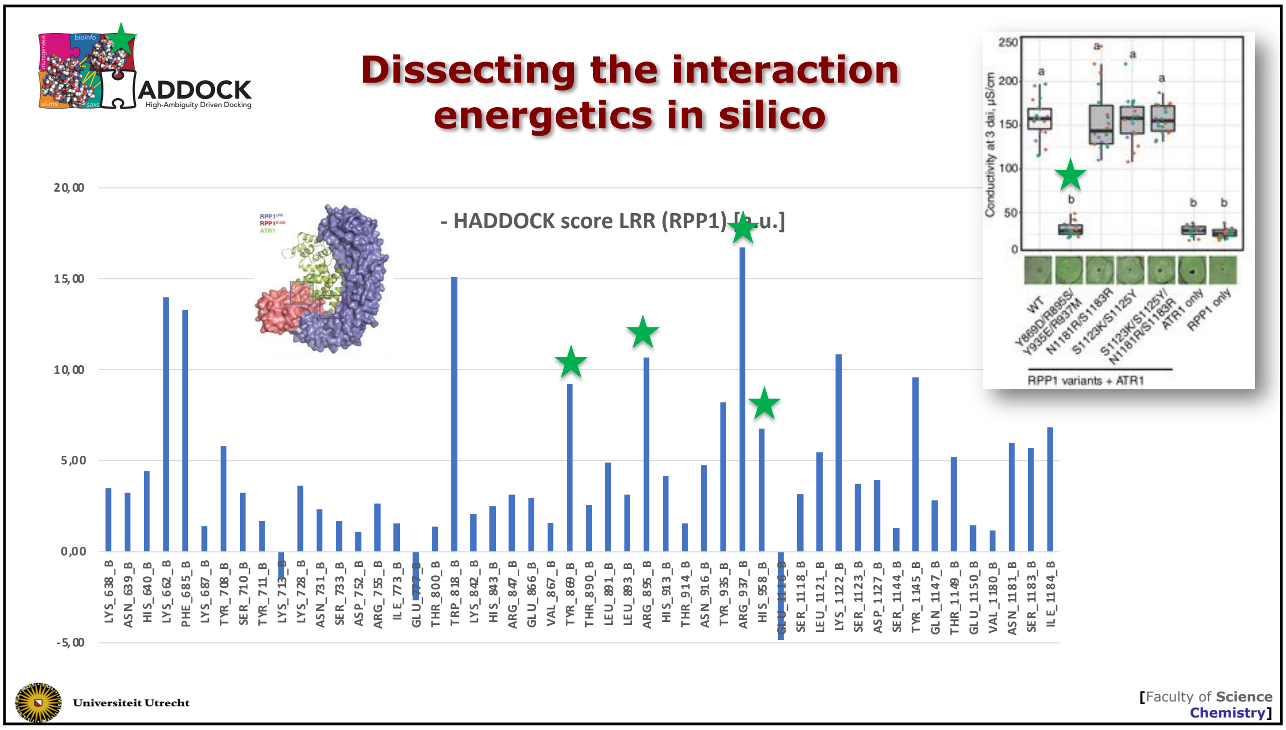
Shoucai Ma, Dmitry Lapin, Li Liu, Yue Sun, Wen Song, Xiaoxiao Zhang, Elke Logemann, Dongli Yu, Jia Wang, Jan Jir Zhifu Han, Paul Schulze-Lefert, Jane E. Parker and Jijie Chai

Science **370** (6521), eabe3069.
DOI: 10.1126/science.abe3069

ATR1: Hyaloperonospora arabidopsidis effector
RPP1: Nucleotide-binding leucine-rich repeat receptor from arabidopsis

Universiteit Utrecht

[Faculty of **Science**
Chemistry]



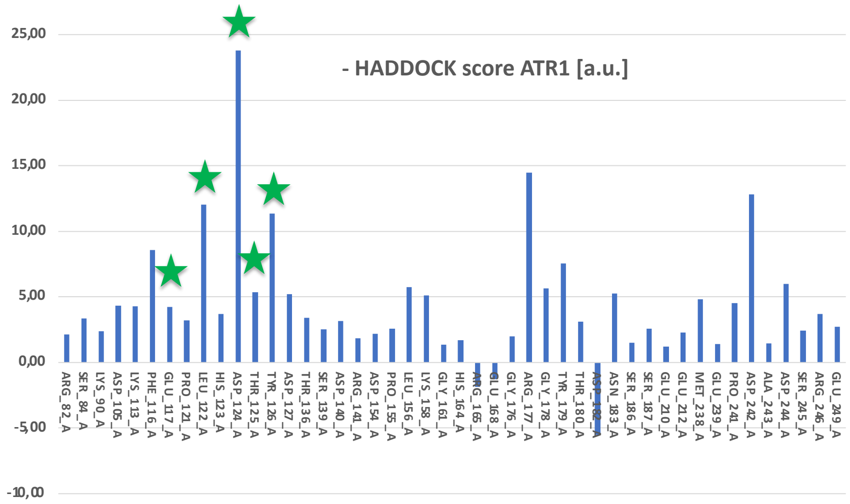
Dissecting the interaction energetics in silico

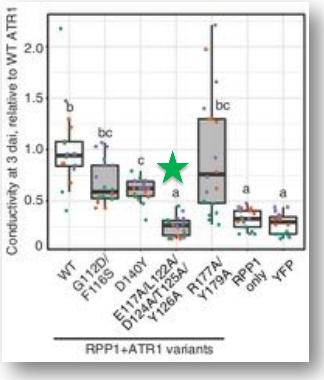
Universiteit Utrecht

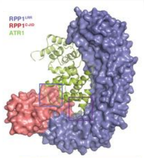
[Faculty of **Science**
Chemistry]

HADDOCK
High-Ambiguity Driven Docking

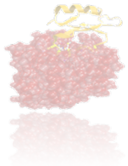
Dissecting the interaction energetics in silico





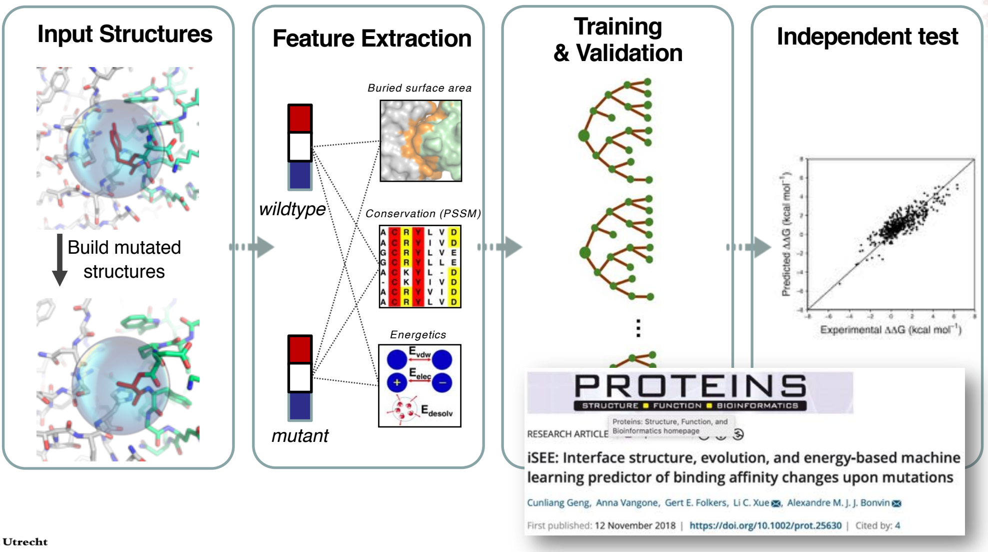


Potential and perspectives in plant immunity



- Predict complexes between immunity proteins (e.g. LRR domains) and pathogen components
- Given the knowledge of pairs of protein interacting, identify key residues for the interaction (from sequence and/or structure)
- Given a mutation in a pathogen protein, can we select/engineer LRR domains that would restore binding (and thus induce immune response)?

From 3D structure to prediction of the impact of mutations



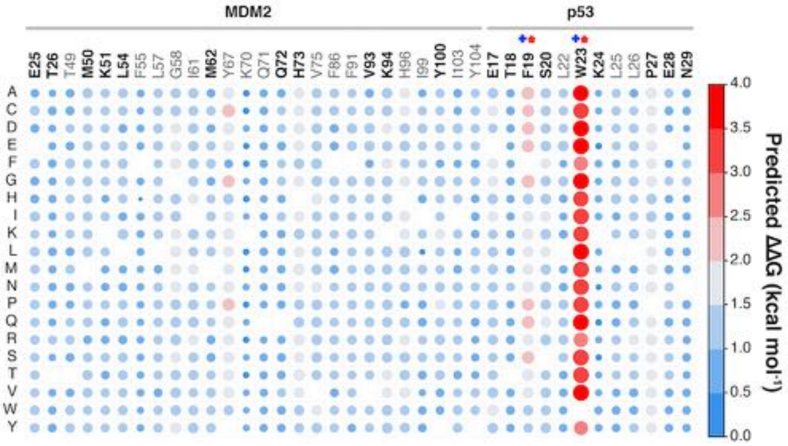
RESEARCH ARTICLE

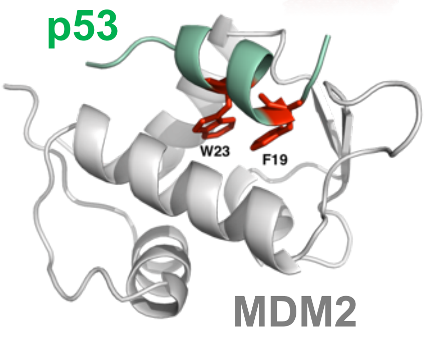
iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations

Cunliang Geng, Anna Vangone, Gert E. Folkers, Li C. Xue, Alexandre M. J. J. Bonvin

First published: 12 November 2018 | https://doi.org/10.1002/prot.25630 | Cited by: 4

Full mutation scanning on MDM2-p53 complex





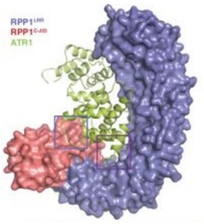
+ Experimental hot-spots

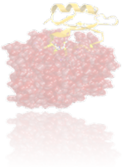
* Predicted hot-spots

iSEE predicted two hot-spots, which were validated by experimental data.

Science

Going back to plants…

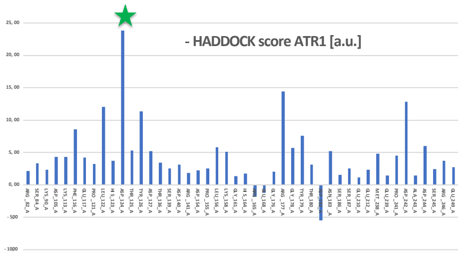




Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme

Shoucai Ma, Dmitry Lapin, Li Liu, Yue Sun, Wen Song, Xiaoxiao Zhang, Elke Logemann, Dongli Yu, Jia Wang, Jan Jirschitzka, Zhifu Han, Paul Schulze-Lefert, Jane E. Parker and Jijie Chai

Science **370** (6521), eabe3069.
DOI: 10.1126/science.abe3069



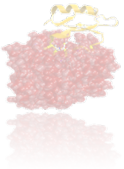
As "pathogen" mutating ASP124

might help me evade recognition…

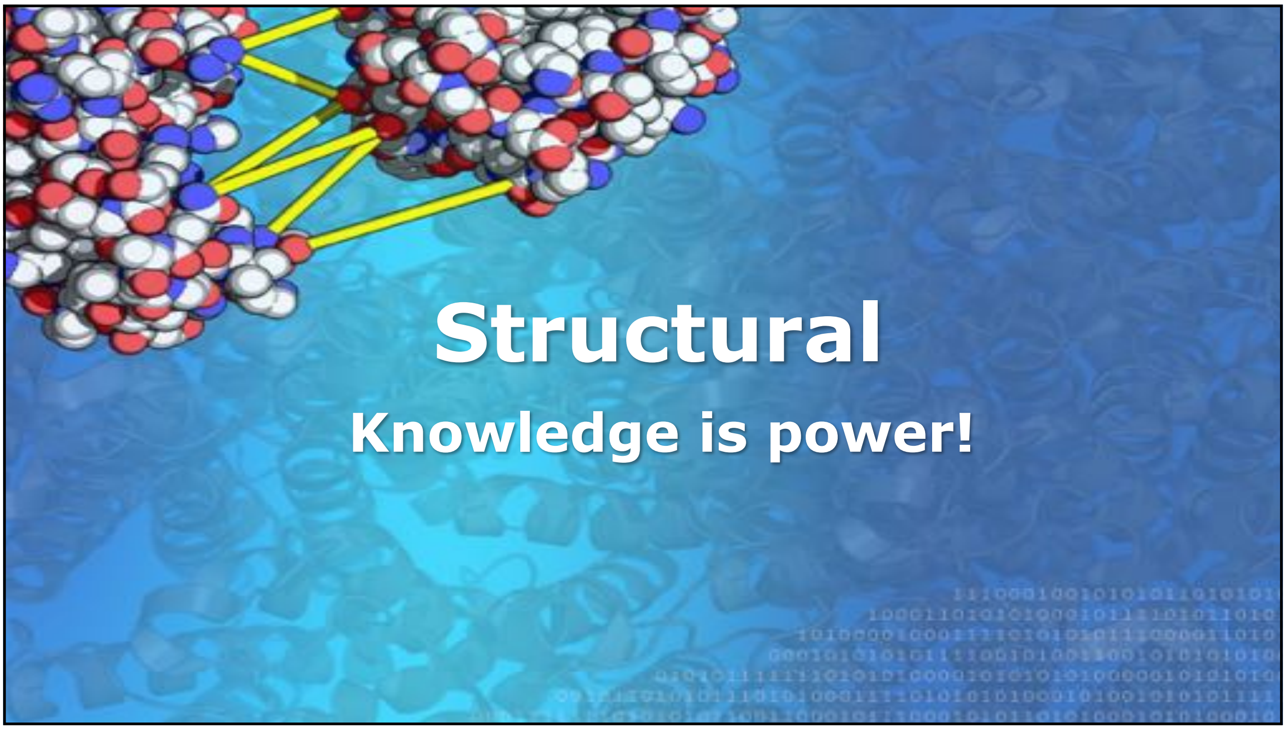
WT HADDOCK score:	**-233.5 [a.u.]**	**ΔHS = 0**
ATR1$_{D124S}$ HADDOCK score:	**-221.6 [a.u.]**	**ΔHS = 11.9**
RPP1$_{R895D}$ ATR1$_{D124S}$ HADDOCK score:	**-228.4 [a.u.]**	**ΔHS = 5.1**

HADDOCK
High-Ambiguity Driven Docking

Potential and perspectives in plant immunity

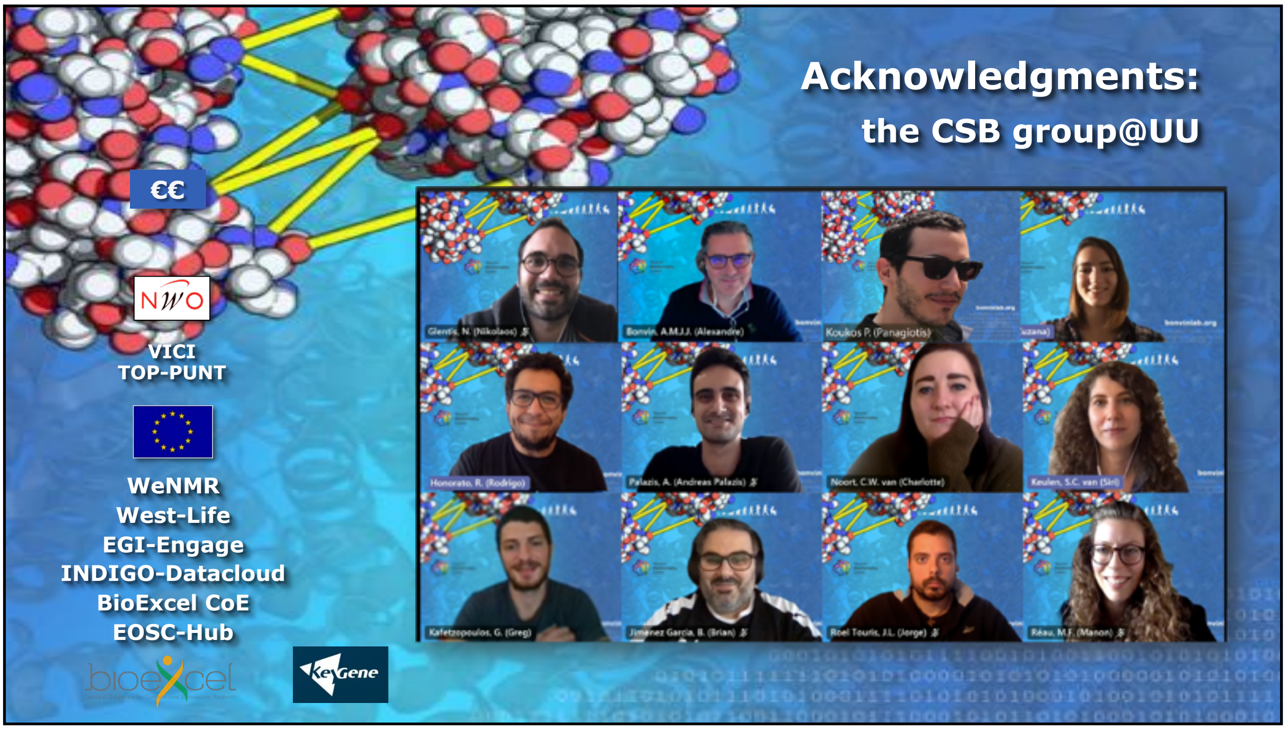


- Predict complexes between immunity proteins (e.g. LRR domains) and pathogen components
- Given the knowledge of pairs of protein interacting, identify key residues for the interaction (from sequence and/or structure)
- Given a mutation in a pathogen protein, can we select/engineer LRR domains that would restore binding?
- Both AI (deep learning) (e.g. Keygene DeepR research programme) and 3D structure can help here!



Structural

Knowledge is power!



Acknowledgments:
the CSB group@UU

€€

NWO
VICI
TOP-PUNT

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

bioexcel
KeyGene