The impact of AlphaFold

Marc F. Lensink

UGSF

Computational Biology

What is the purpose of this meeting?

Who are we?

What do we do?

What can we do together?

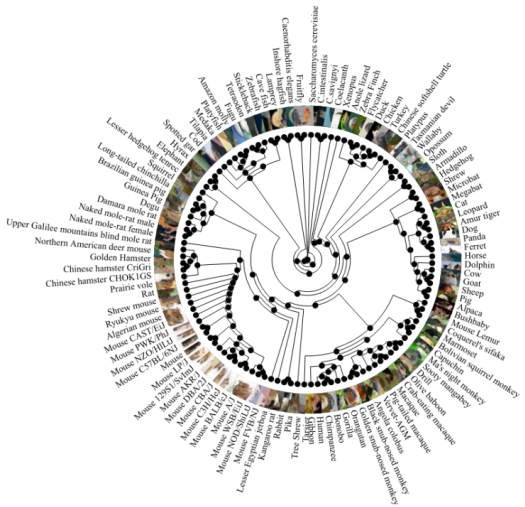
Ralf Blossey	DR1	Theoretical physics; Complex kinetic modeling
Julie Bouckaert	DR2	Crystallography; Biophysics
Guillaume Brysbaert	IR1	Functional genomics; Bioinformatics; Networks
Goedele Roos	CRCN	Density functional calculations; Redox reactions
Jérôme de Ruyck	MCF	Molecular modeling; Crystallography
Marc Lensink	DR2	Protein-protein interaction; Molecular dynamics; Structural bioinformatics; Networks

The impact of protein-protein interaction on the molecular flow of information in the cell

Ralf Blossey	DR1	Theoretical physics; Complex kinetic modeling
Julie Bouckaert	DR2	Crystallography; Biophysics
Guillaume Brysbaert	IR1	Functional genomics; Bioinformatics; Networks
Goedele Roos	CRCN	Density functional calculations; Redox reactions
Jerome de Ruyck	MCF	Molecular modeling; Crystallography
Marc Lensink	DR2	Protein-protein interaction; Molecular dynamics; Structural bioinformatics; Networks

Protein structure and dynamics

Molecular interaction



Sequence and structure space are slowly being filled

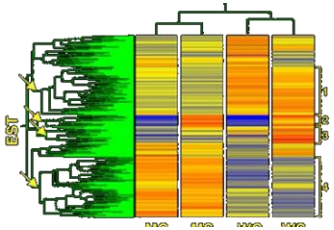
Structure is more conserved than sequence

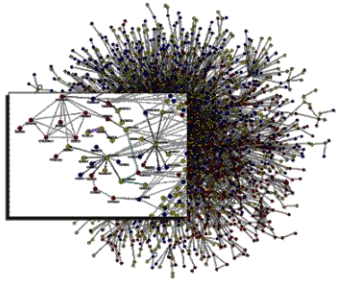
Sampling of the interactome remains sparse

There is a need to clarify how proteins interact

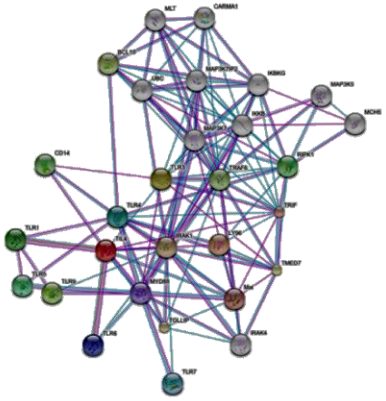
***Genome**: collection of hereditary information*

***Transcriptome**: collection of genes expressed under environmental conditions of the cell*



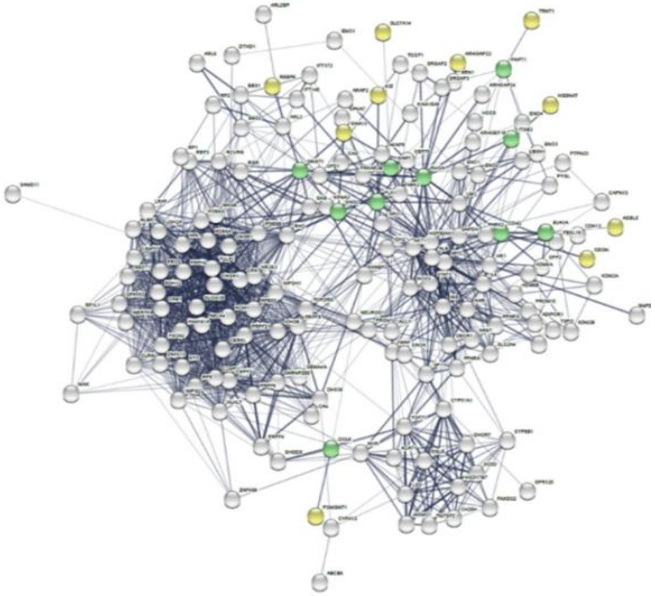


***Proteome**: collection of expressed proteins*



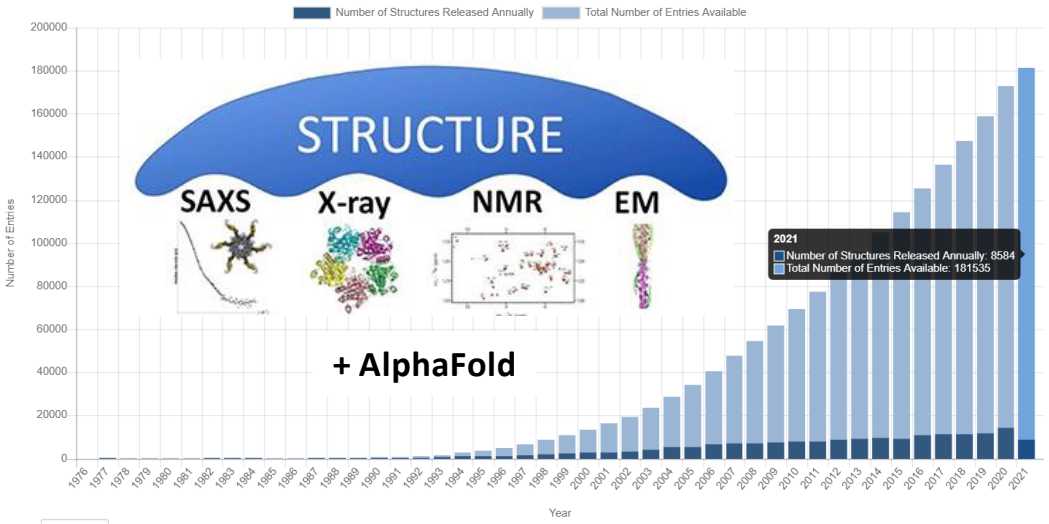
***Interactome**: whole set of molecular interactions*

Protein-protein interaction



UniProt	214M	sequences
WORLDWIDE PDB PROTEIN DATA BANK	173K	structures
STRING	24M	Proteins
	>2000M	Interactions
	>5000	Species

Cellular signaling
Expression regulation
Metabolic substrate channeling



4.

JBC ARTICLE

Assembly of B4GALT1/ST6GAL1 heteromers in the Golgi membranes involves lateral interactions via highly charged surface domains

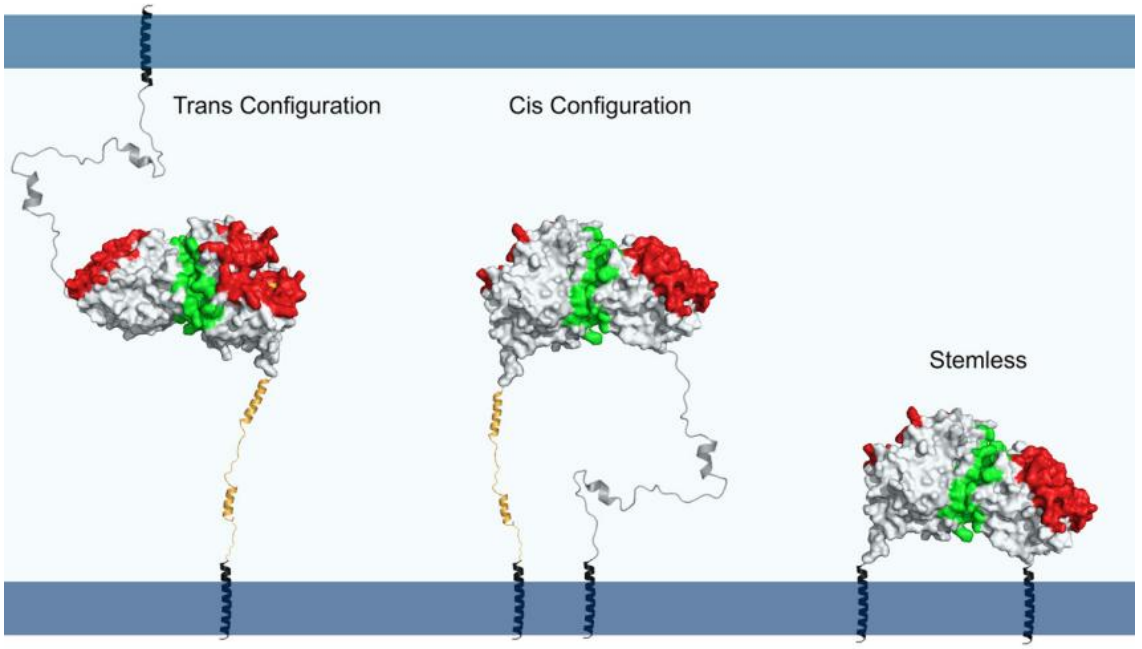
Received for publication, May 27, 2019, and in revised form, August 1, 2019 Published, Papers in Press, August 8, 2019, DOI 10.1074/jbc.RA119.009539

Fawzi Khoder-Agha[‡], Deborah Harrus[‡], Guillaume Brysbaert[§], Marc F. Lensink[§], Anne Harduin-Lepers[§], Tuomo Glumoff[‡], and Sakari Kellokumpu[‡1]

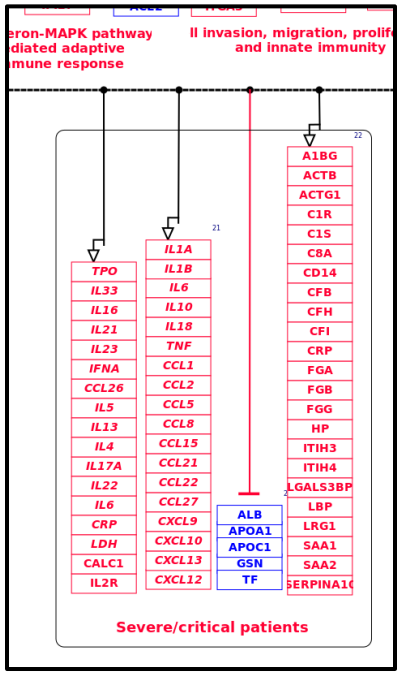
Two of the conclusions from this paper:

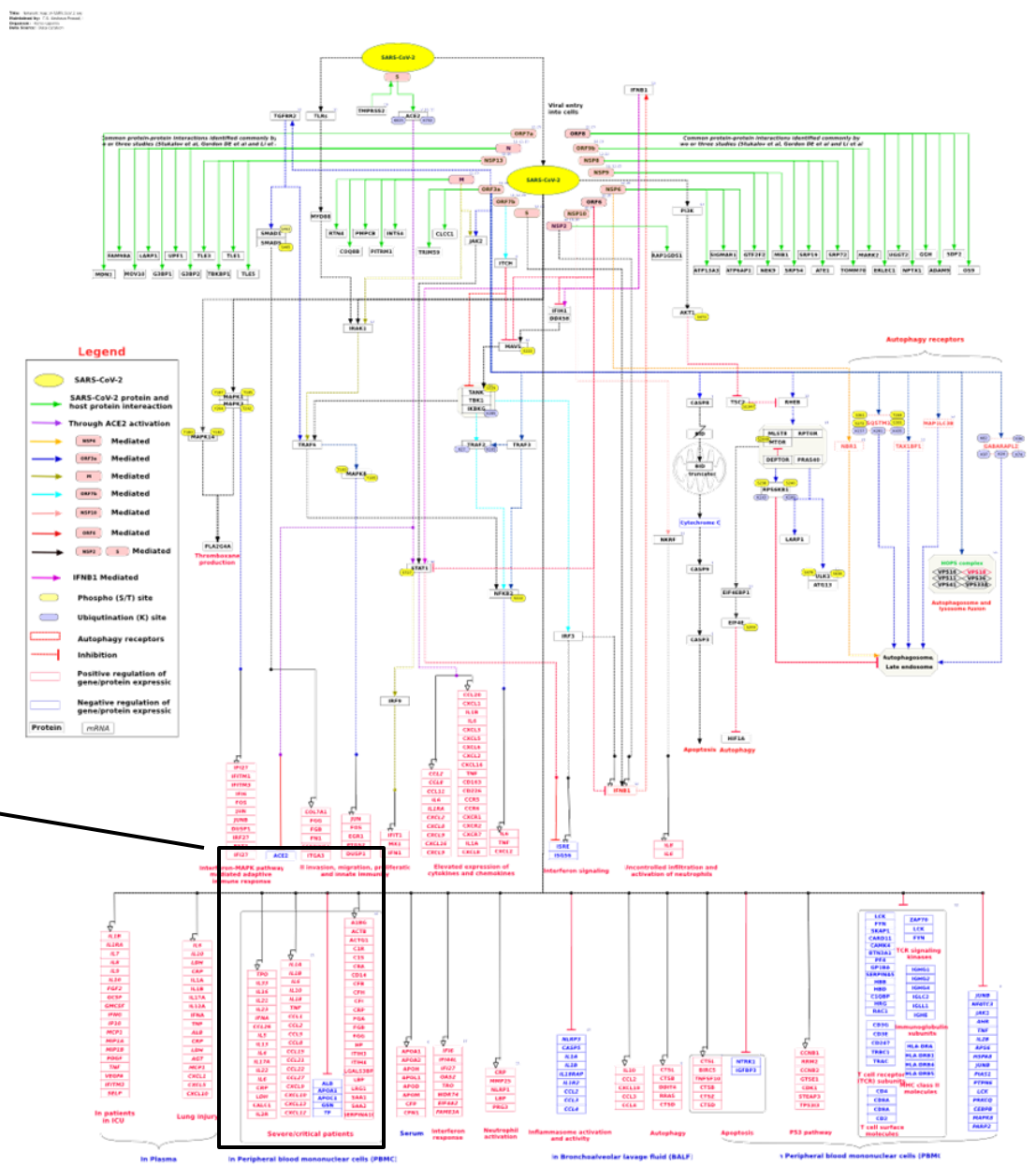
- *"B4GALT1 uses its noncatalytic surface for interaction with ST6GAL1"*

- *"Heteromeric assembly regulates the catalytic activity of B4GALT1 and ST6GAL1"*



Signaling landscape

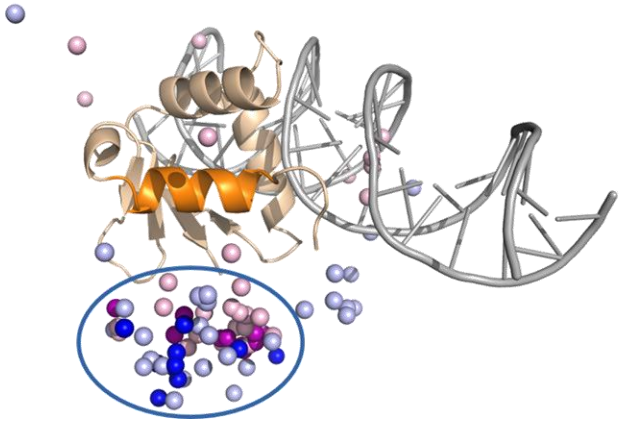


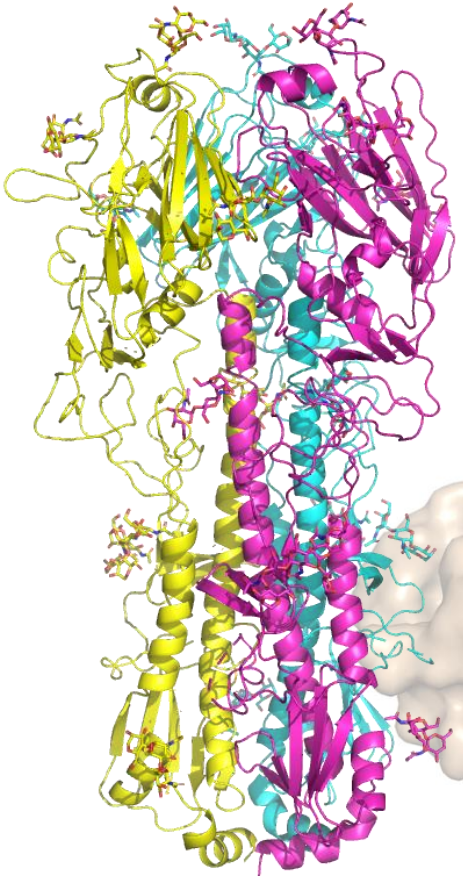


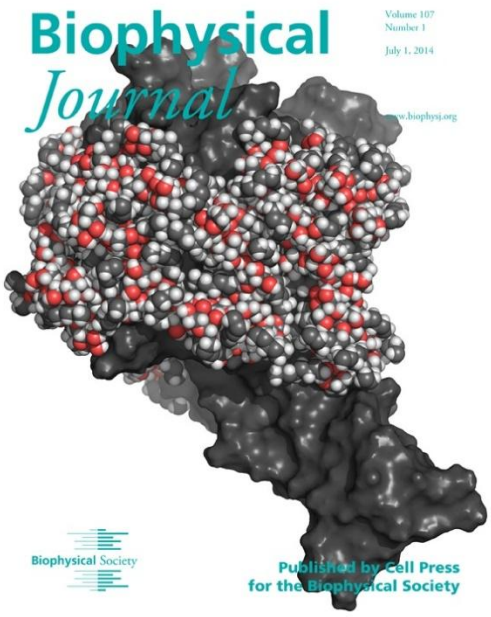
Rex *et al*,
SARS-CoV-2 signaling pathway map: a functional landscape of molecular mechanisms in COVID-19,
J Cell Commun Signal 2021

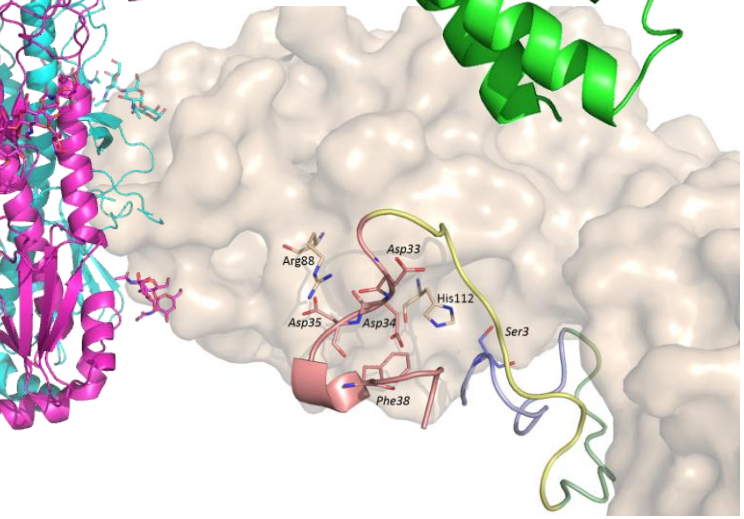
Publicly available through WikiPathways

1. Modeling





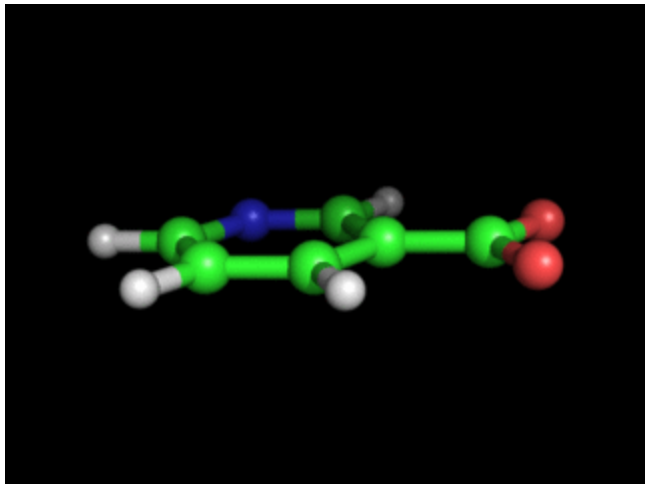

Biophysical
Journal
Volume 107
Number 1
July 1, 2014
www.biophysj.org
Biophysical Society
Published by Cell Press
for the Biophysical Society


Arg88
Asp33
His112
Asp35
Asp34
Ser3
Phe38

Molecular dynamics

$$F = -\nabla V(\vec{r}) \qquad F = m \cdot a$$

$$V = \frac{1}{2} \sum_{i,j \neq i} V_{i,j}$$



Give atoms initial r_0 and v_0

↓

Calculate forces F from $V(r)$

↓

Apply F with small time step Δt to get new positions r_i and velocities v_i

↓

Apply boundary conditions
Apply temperature and pressure control

↓

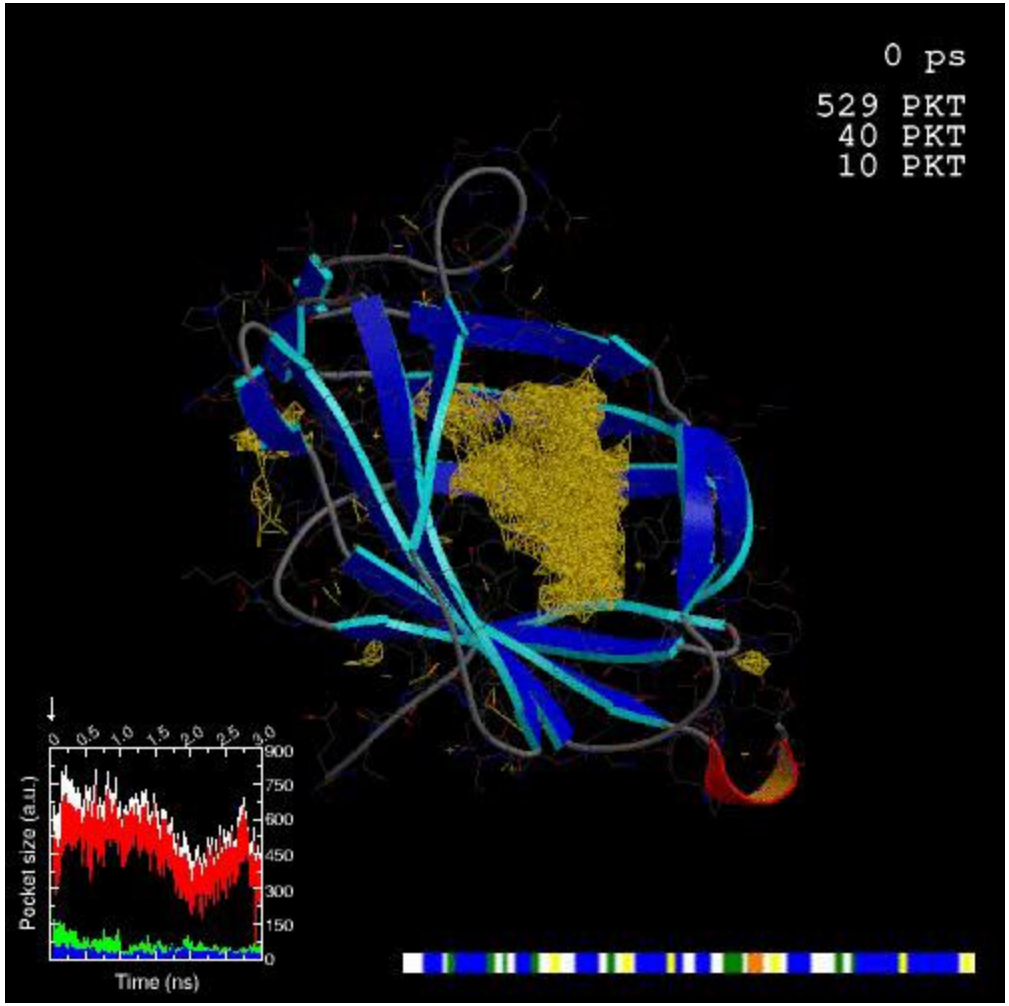
Calculate and output physical quantities of interest

↓

Iterate to next step: $t = t+\Delta t$

Protein dynamics

- Provides a time evolution or trajectory of atomic positions
- Time step of femtoseconds
- Hundreds of nanoseconds required for biological significance



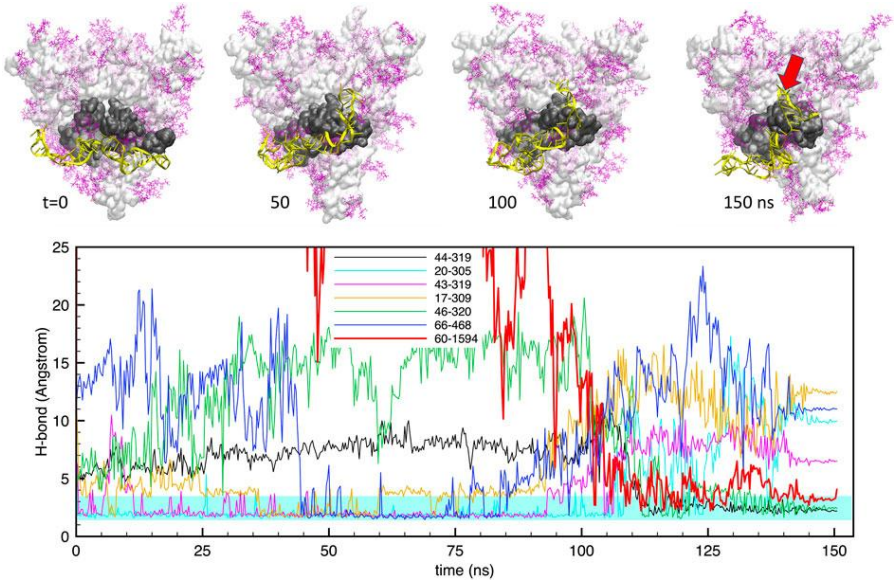
COVID

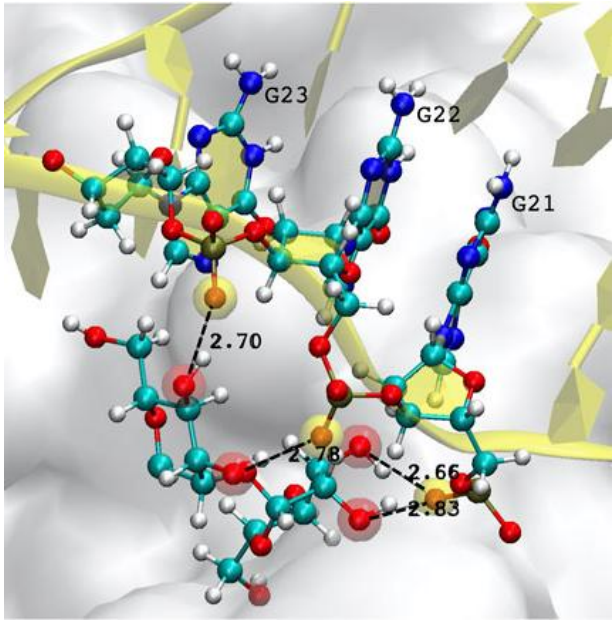
frontiers in Molecular Biosciences

ORIGINAL RESEARCH
published: 12 August 2021
doi: 10.3389/fmolb.2021.713003

DNA Aptamers Block the Receptor Binding Domain at the Spike Protein of SARS-CoV-2

Fabrizio Cleri[1,2]*, *Marc F. Lensink*[3] and *Ralf Blossey*[3]



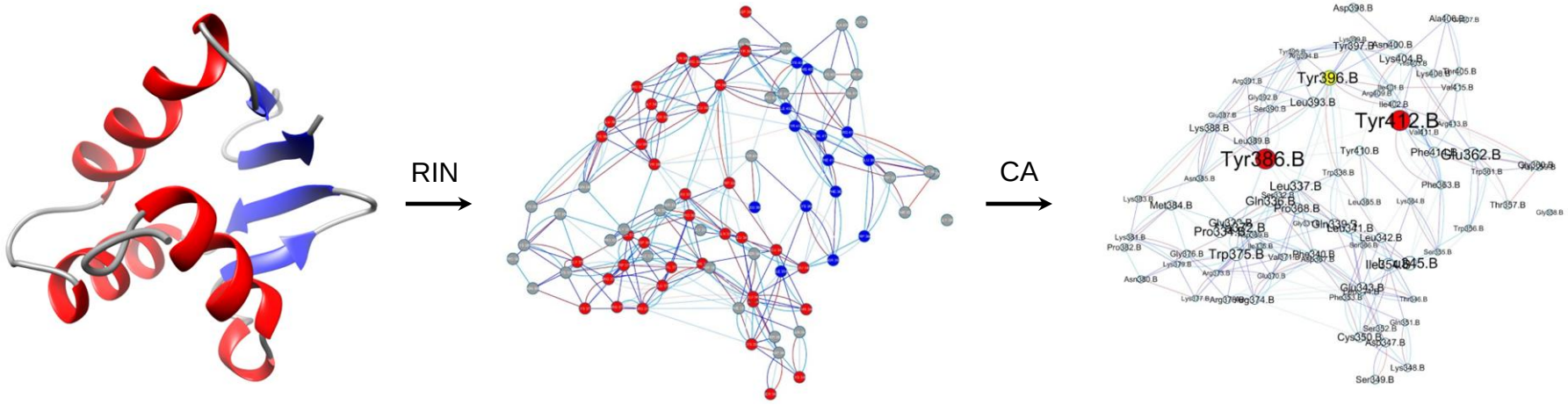


2.

RIN – Residue Interaction Networks

Centrality Analysis

- Centrality analysis identifies important nodes in a network
- Many centrality measures exist
- Computing a Z-score highlights the major ones
- Central residues correlate strongly with functional or structural relevance



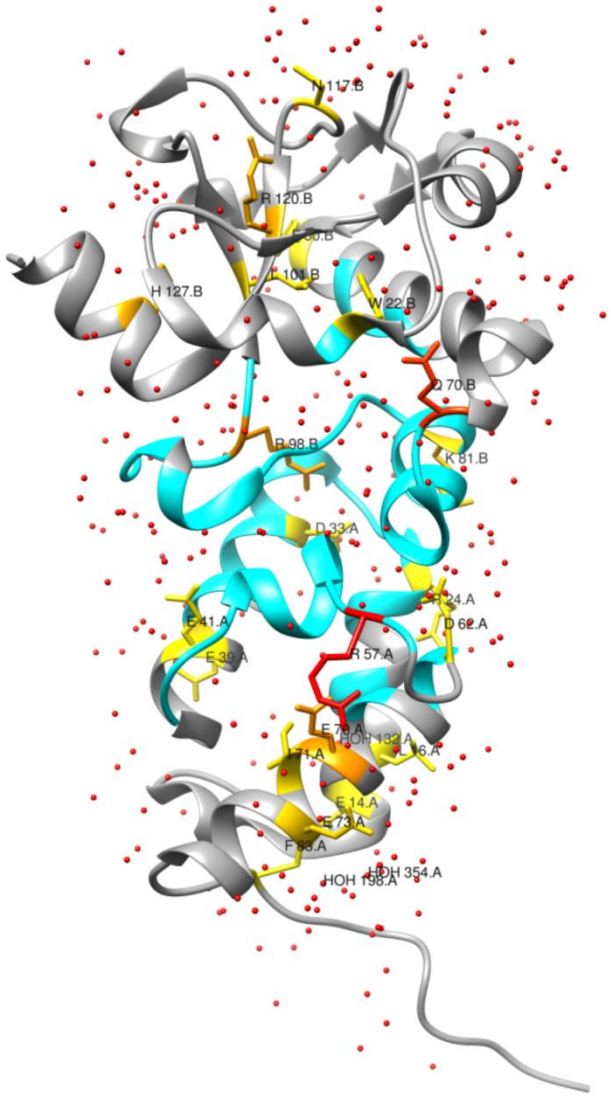
Residue Interaction Networks

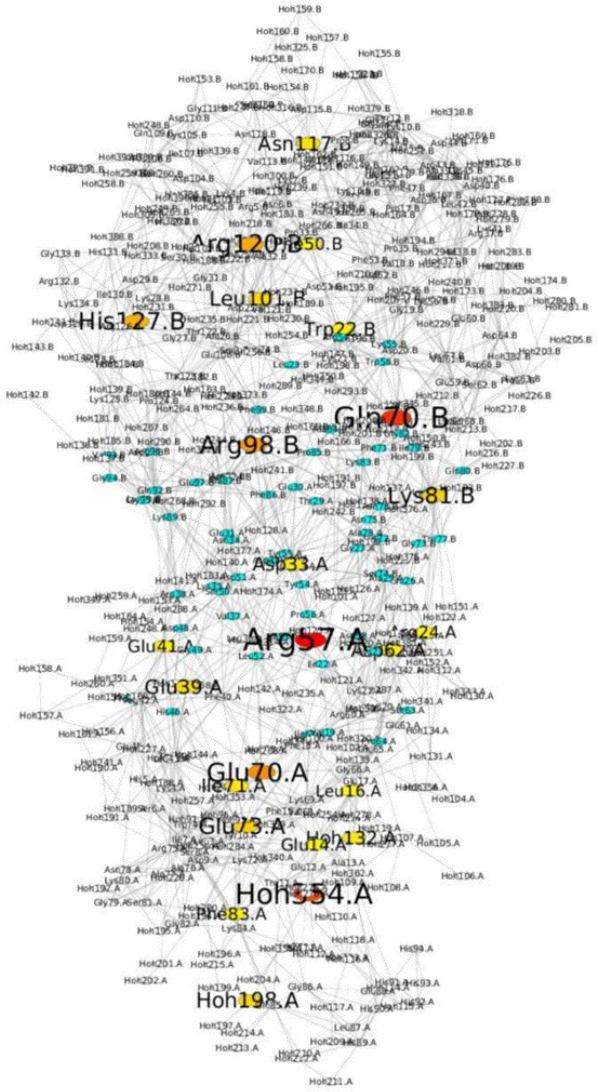
Include water molecules in the analysis

This can also be extended to lipids

… and to sugars

Add dimension:
- Time
- Evolution

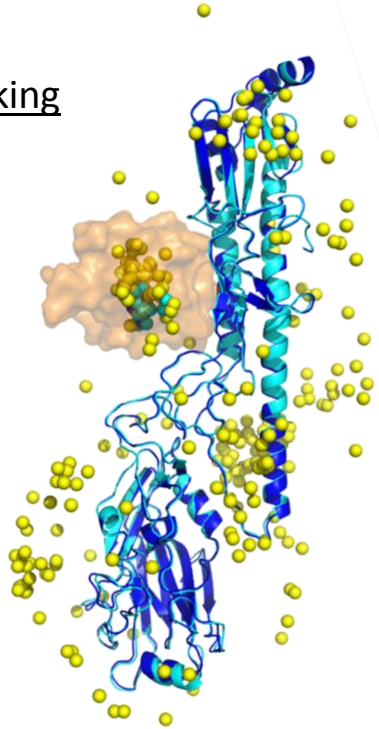




3. Critical Assessment of PRedicted Interactions (CAPRI)

CAPRI organizes blind docking prediction trials

Protein/protein
-Peptides
-Nucleic acids
-Polysaccharides
Oligomer assembly
Multi-domain assembly
Interfacial water positions
Binding affinity



Proteins *publishes a special CAPRI-dedicated issue with every Evaluation Meeting*

Most algorithms developed in CAPRI context

CAPRI has been a proven catalyst
CAPRI *has made it possible to* ***reliably predict*** *what* ***the quality of your docking model*** *will be****!***

Management Committee	
Alexandre Bonvin	NL
Marc Lensink	FR
Michael Sternberg	UK
Sandor Vajda	USA
Ilya Vakser	USA
Sameer Velankar	UK
Zhiping Weng	USA
Scientific Advisory Board	
Rommie Amaro	USA
Stephen Burley	USA
Wah Chiu	USA
Kristina Djinovic Carugo	FR

assessment, organization, website, operations, infrastructure

Marc 

Alexandre 

Shoshana 

Guillaume 

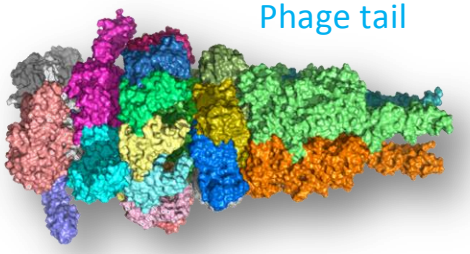
Sameer 

Nurul 

Target diversity

T170

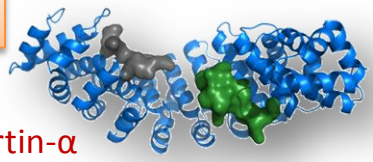
*



Phage tail

T60-64

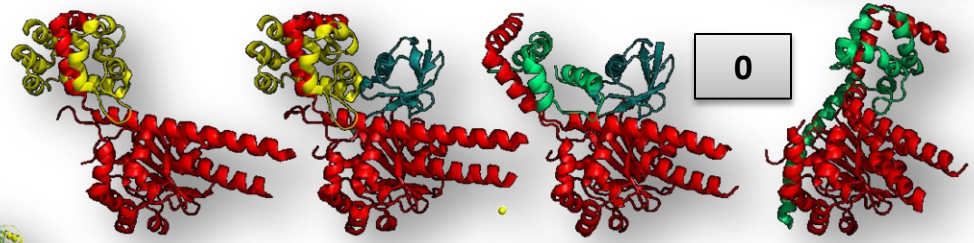
***



NLS
Importin-α

T98-101

Complexes of Ubiquitin carboxyl-terminal hydrolase L5 with subunits of protein degradation or chromatin remodeling complexes

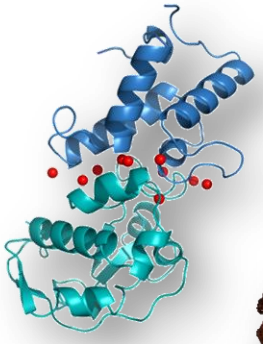


0

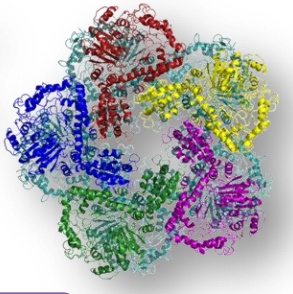
***
+++

T105

Water prediction

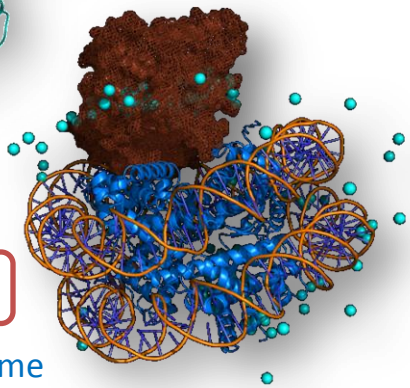


**



*

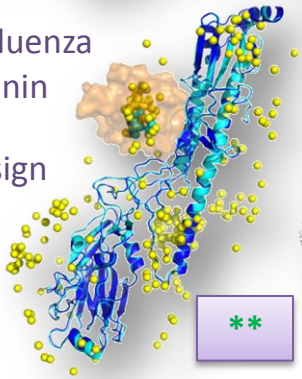
T95



Nucleosome
PRC1 ubiquitylation module

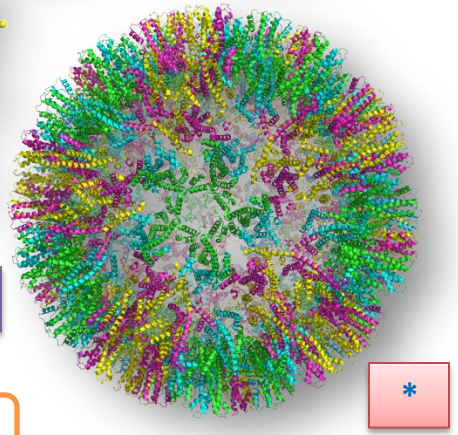
Spanish influenza hemagglutinin

Protein design



T50

**



*

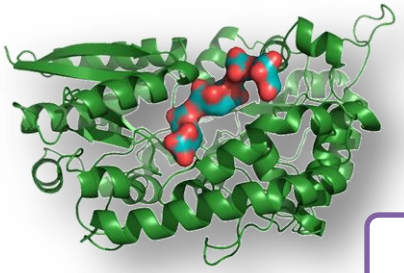
Viral capsid

T180

T136

Decarboxylase

Arabinose metabolism

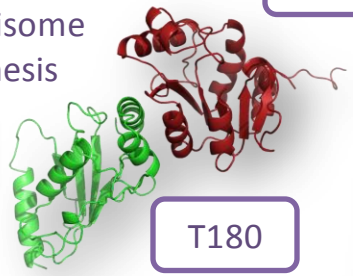


**

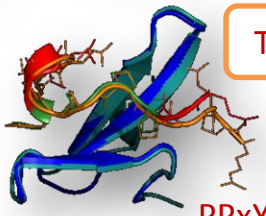
T126-T129

Perioxisome biogenesis

**



T180



T67

***

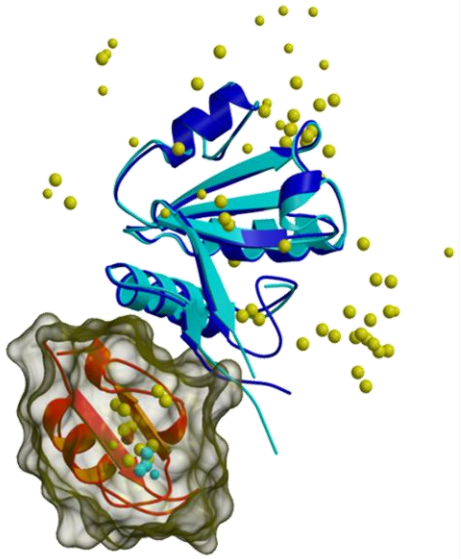
PPxY motif

The CAPRI Community (2019)

Note the date: 8th CAPRI Evaluation Meeting: Grenoble February 12 – 16, 2024

Lensink & Méndez, *Curr Pharm Biotechnol* **2008**;9:77

"Recognition-induced conformational changes in protein-protein docking"



Protein A Coordinates

Protein B Coordinates

Flexibility Analysis

MD-PCA / ED
NMA / GNM / ANM
Graph Theory

Biological Information

Protein A & B Conformers

Rigid body Search

FFT-methods
GA / GH, Monte Carlo
Neural networks

Raw score Filtering

Surface complementarity,
Electrostatics

Energy-based Re-ranking

Knowledge-based potentials,
Force-field based energy functions

MM or MD, implicit or explicit solvent

Refinement

Complex AB Coordinates

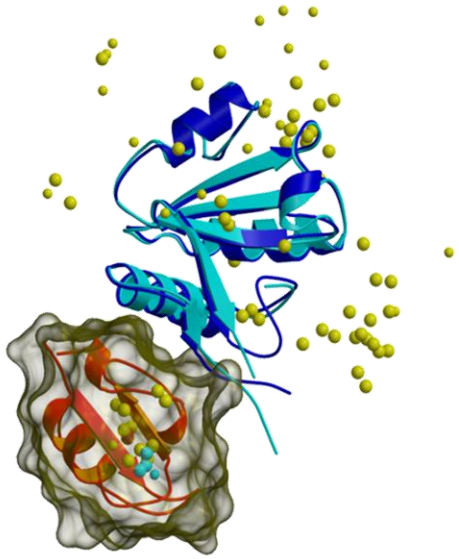
Lensink *et al*, *Proteins* **2023**, Epub

"Impact of AlphaFold on structure prediction of protein complexes"

Protein A Sequence

Protein B Sequence

Biological Databases



AlphaFold?

Complex AB Coordinates

Protein docking…

is more than simply putting two proteins together

1. Exploring
2. Scoring

CAPRI is actively working to push development in all these aspects

- Identifying problem areas
- Searching for targets
- Organizing blind prediction trials

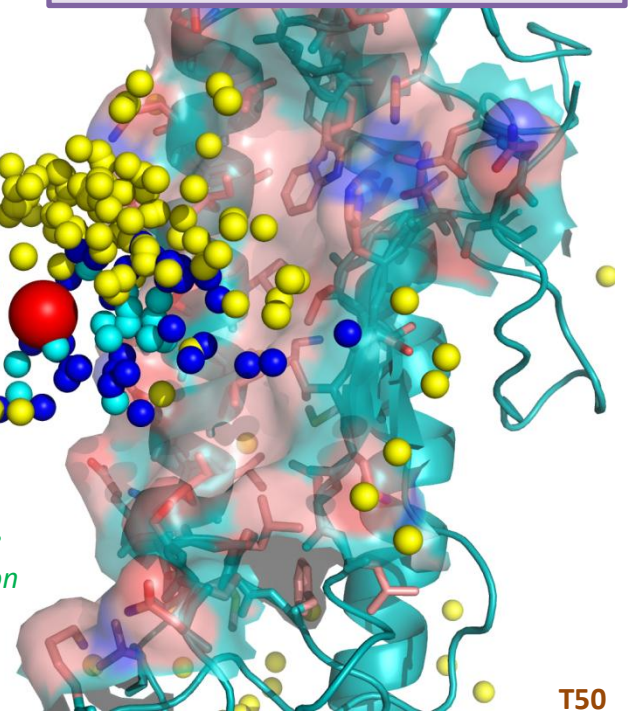
- *k*-space docking
- guided docking
- co-evolution docking
- contact prediction
- machine learning and deep-learning

Incorrect but near-acceptable

Many of the incorrect models do show good interface prediction values.

Target

Every colored sphere corresponds to the geometric center of the predicted ligand position



T50

Designed protein-protein complex
Spanish influenza hemagglutinin

Sarel Fleishman, David Baker
University of Washington, USA

Science 2011;332:816

CAPRI SCORESET v2022

Introduction | Browse | Download | Help

Download:

- T050.1.pdb
- T050.1.json
- T050.1.xml

Model download:

Target ID	Pubmed ID	PDB	Description		Stoichiometry	#Interfaces
T050	21566186	3R2X	*HB36.3 designed protein/flu hemagglutinin*		ABC	1
	Species and kingdom:		Influenza A virus / designed	Viruses		
	Interaction type:		antigen/antibody	unbound		
(Easy)	Interface	1	Number of decoys: 1948	AB:C 327/171:82	ABC	1500 Å²
		Receptor	HEMA_I18A0	7-328/1-173		
		Ligand	*designed*	1-93		

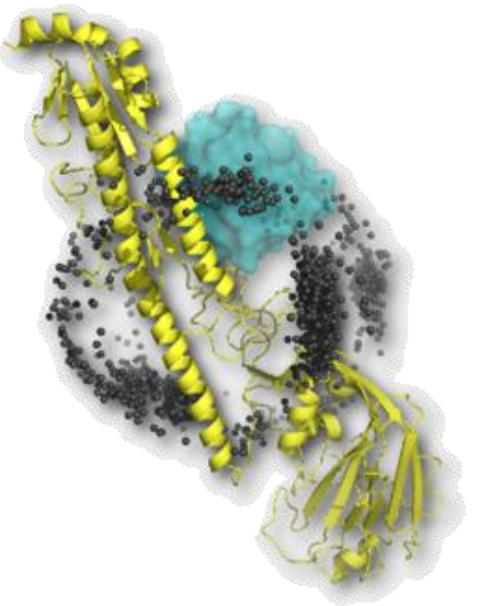
Previous

Next

The CAPRI scoreset v2022 includes **all published CAPRI targets up to CAPRI Round 50**, including joint CAPRI/CASP Rounds.

All models are uniform!
- all atoms are reordered following PDB standard
- all decoys are superimposed on the target receptor entity
- all chain labeling and residue numbering matches the target
- all computed assessment quantities are included in the dataset

And finally:
Target images are aligned such that they match the flag of Ukraine



The CAPRI scoreset v2022 is developed and maintained by:

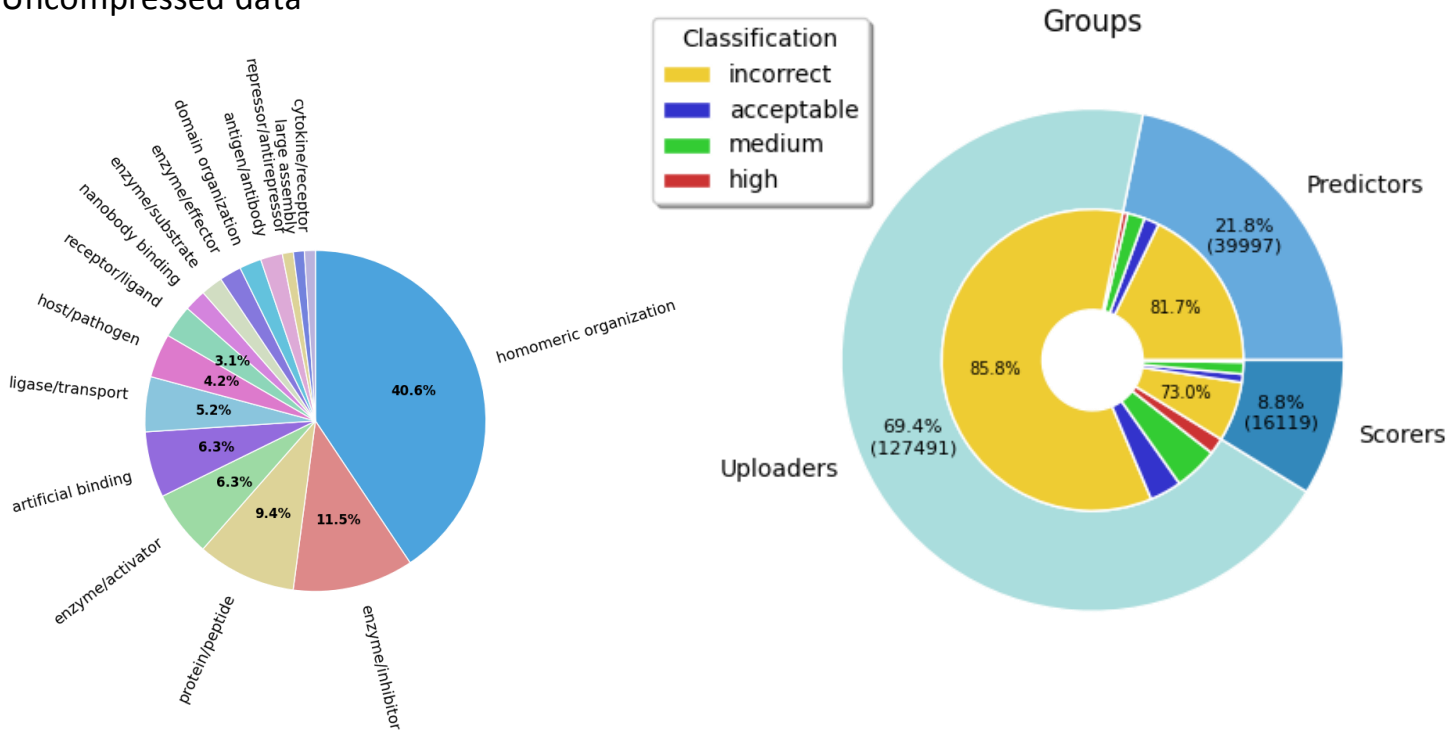
- Marc F. Lensink, CNRS & University of Lille, France
- Theo Mauri, CNRS & University of Lille, France
- Guillaume Brysbaert, CNRS & University of Lille, France
- Shoshana J. Wodak, VUB-VIB, Belgium

cnrs Université de Lille VIB-VUB CENTER FOR STRUCTURAL BIOLOGY

scoreset.org

Scoreset is…

96	Targets	127k	Uploader models	All annotated with all calculated CAPRI metrics
148	Interfaces	40k	Predictor models	
170k	Models	16k	Scorer models	
57.5G	Uncompressed data			



CAPRI / CASP

CAPRI

Since 2001

Critical **A**ssessment of
PRedicted **I**nteractions

Joint prediction rounds since 2014:

25 Targets	Round 30
10 Targets	Round 37
21 Targets	Round 46
12 Targets	Round 50
37 Targets	Round 54

Prediction rounds on a "rolling" basis

Fits with publication schedule

3 to 4 weeks per prediction round

CASP

Since 1994

Structure **P**redictions

CASP11	2014
CASP12	2016
CASP13	2018
CASP14	2020
CASP15	2022

Prediction season

Intense 2 to 3 months

Difference in targets

Mostly hetero-dimers or –trimers
Peptides, sugars, water positions

Incites method development

Mostly obligate, many homo-oligomers
Very large assemblies

Large-scale testing of methodologies

CAPRI / CASP

CAPRI

Since 2001

Critical **A**ssessment of

PRedicted **I**nteraction

Joint prediction rounds since 20

25 Targets	Round		2014
10 Targets	Round 37		2016
21 Targets	Round 46	CASP13	2018
12 Targets	Round 50	CASP14	2020
37 Targets	**Round 54**	**CASP15**	**2022**

Prediction rounds on a "rolling" basis

Fits with publication schedule

3 to 4 weeks per prediction round

Difference i

Mostly hetero-dimers or –trimers

Peptides, sugars, water positions

Incites method development

CASP

Since 1994

ture **P**redictions

Prediction season

Mostly obligate, many homo-oligomers

Very large assemblies

Large-scale testing of methodologies

Prediction over the summer of 2022

Results presented at CASP15 conference in Antalya, December '22

"AlphaFold can accurately predict 3D models of protein structures and has the potential to accelerate research in every field of biology."

"AlphaFold: a solution to a 50-year old grand challenge in biology."

www.nature.com › news

'It will change everything': DeepMind's AI makes gigantic leap ...

Nov 30, 2020 — DeepMind's **AlphaFold 2** algorithm outperformed other teams at the CASP14 protein. Source: DeepMind. DeepMind's 2018 performance at ...

deepmind.com › blog › article › alphafold-a-solution-t...

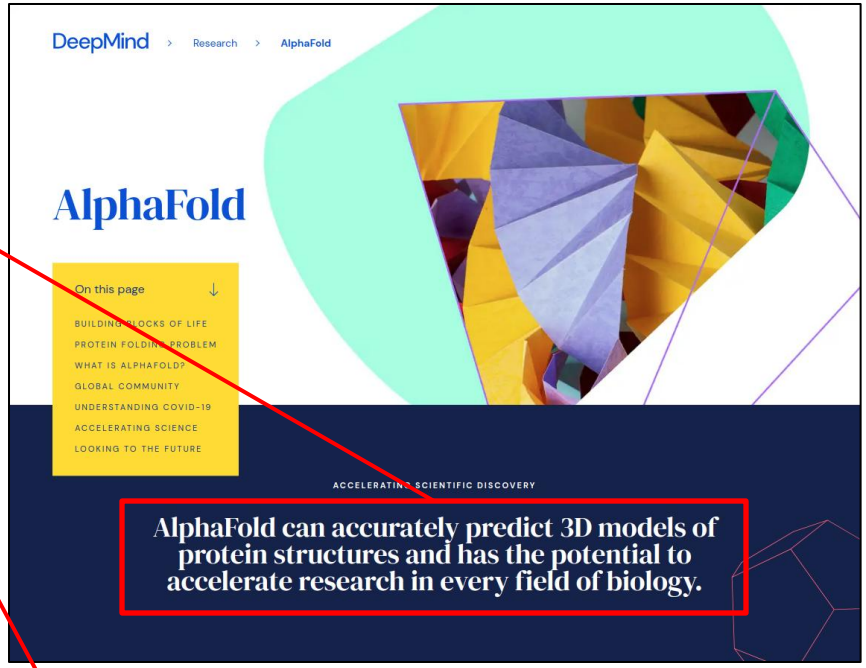
AlphaFold: a solution to a 50-year-old grand challenge in ...

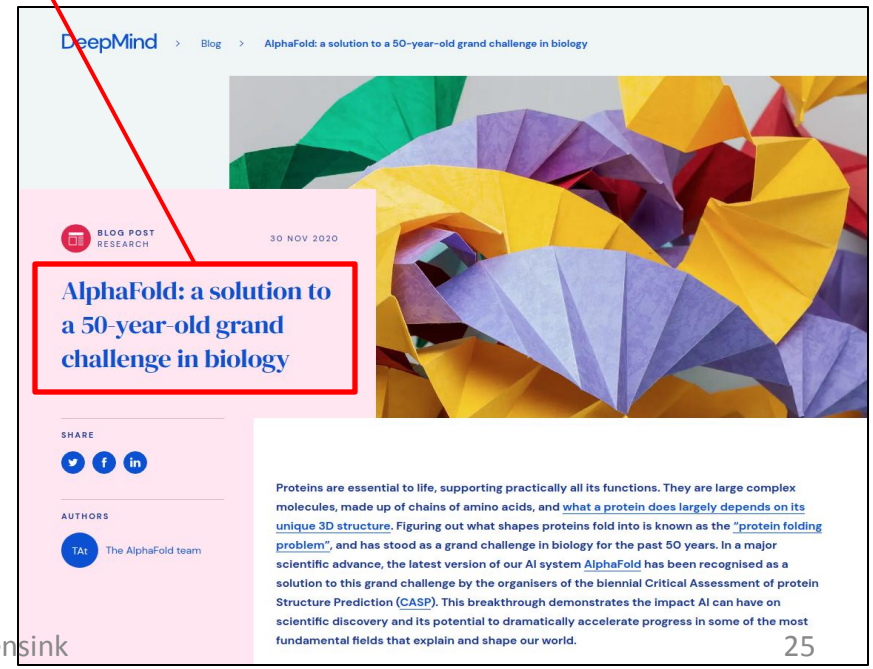
Nov 30, 2020 — **AlphaFold**: The making of a scientific breakthrough · Improvements in the median accuracy of predictions in the free modelling category for the ...

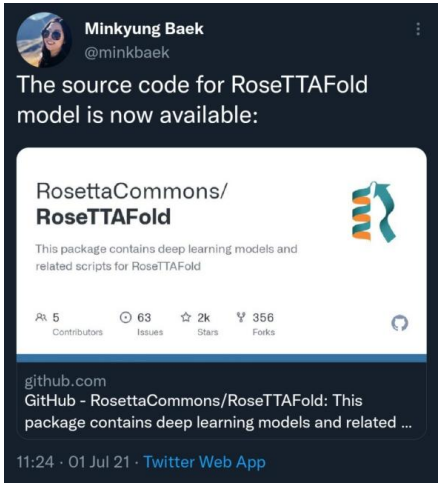
AlphaFold: Using AI for ... · AlphaFold · Computational predictions of ...

'So, either this group is close to solving the folding problem or they cheated somehow.'

Nick Grishin

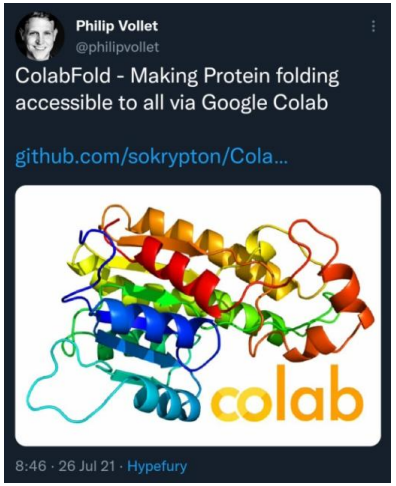




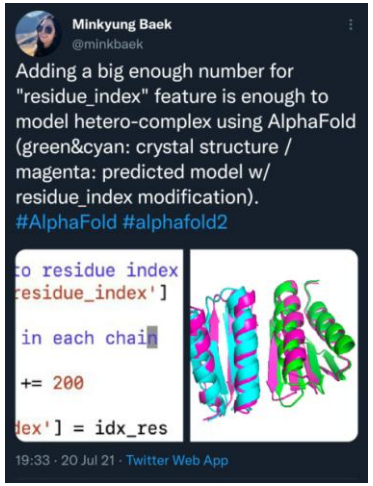


1 Jul 2021

15 Jul 2021



26 Jul 2021



20 Jul 2021

A significant day for the life sciences

Dear all,

Today, EMBL and DeepMind have announced a partnership to make the most complete and accurate database yet of predicted 3D structures for the human proteome freely and openly available to the scientific community. The methods and code for AlphaFold, the AI system that powers the structure predictions in the database, were published last week in Nature and are able to be modified by the scientific community. The AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk) will make an extensive set of pre-calculated AlphaFold computational structures - including all human proteins – accessible to all scientists, consistent with EMBL's open data mission.

An initial two-year formal collaboration between EMBL and DeepMind has been agreed, in order to provide access to this remarkable resource for the world. Millions of AlphaFold predictions will be made available through this new resource hosted by EMBL-EBI. The initial release contains approximately 365,000 structures, and will eventually increase to an estimated 130 million 3D models. In context, that would be 700-times more than in the current contents of the Protein Data Bank (PDB), the global resource of experimentally derived 3D structure data.

I am happy to answer any specific questions you may have.

Kind regards,
Sameer

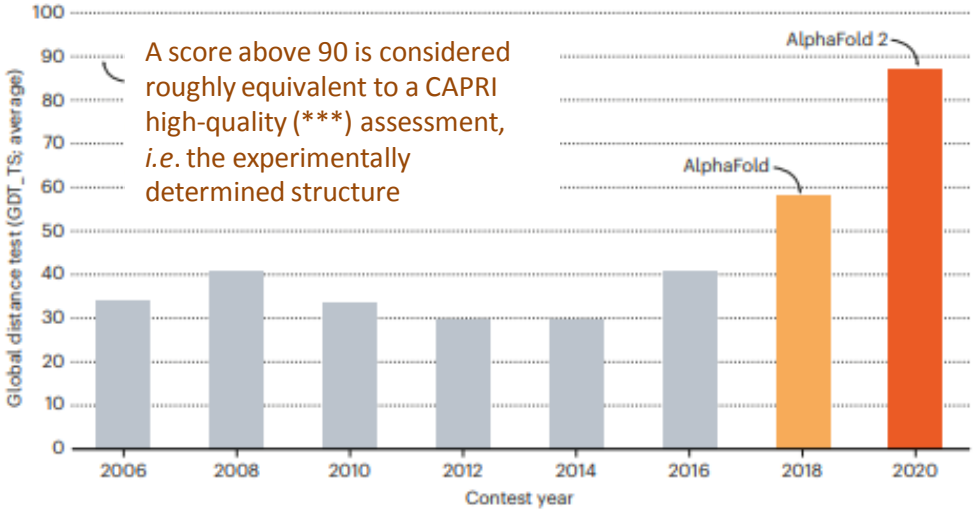
22 Jul 2021

A brief history

In **2020**, **AlphaFold 2** outperforms everybody else in the CASP14 blind structure prediction experiment
In **2021**, the **AlphaFold Protein Structure Database** is launched by DeepMind and EMBL-EBI;
This increases the size of the available protein structure universe **700-fold**

In **2022**, ***deep-learning-based methods were*** expected to be ***applied to assembly prediction*** in CASP15
So where is the state-of-the-art now?

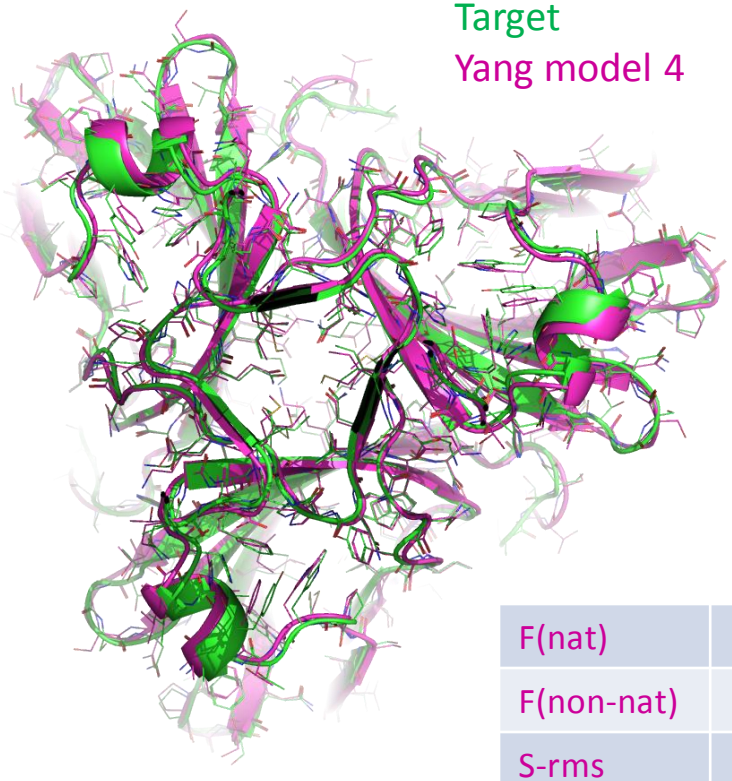
AlphaFold 2 significantly outperformed other teams at the CASP14 protein-folding contest – and its previous version's performance at CASP13



T222	T1173	no pdb	X-ray	2.40	A3

Bacteria
Bdellovibrio bacteriovorus
Cell wall surface anchor

Very good side-chain placement

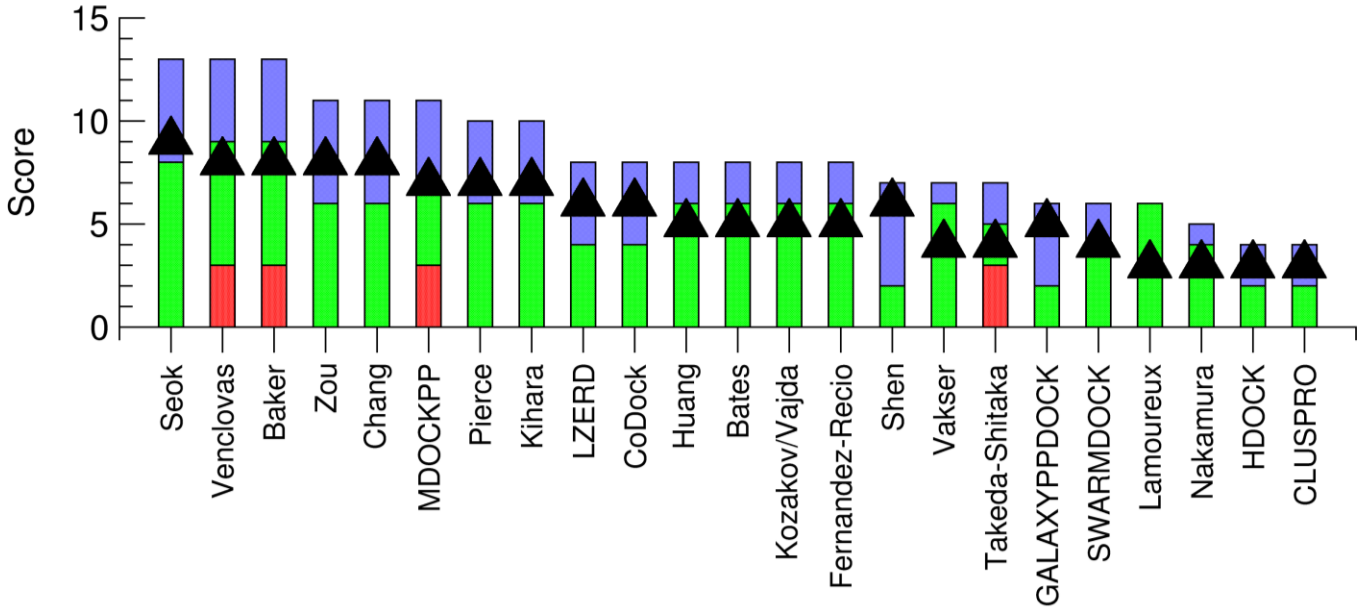


F(nat)	0.857
F(non-nat)	0.153
S-rms	1.85 Å

Group	Performance
Jianyi_Yang	5***
YANG-SERVER, YANG-MULTIMER	5/4***/1**
Wei_Zheng, J_Cheng	4/3***/1**
MULTICOM	4/2***/1**
Many others…	
AF2-Multimer	2/1**

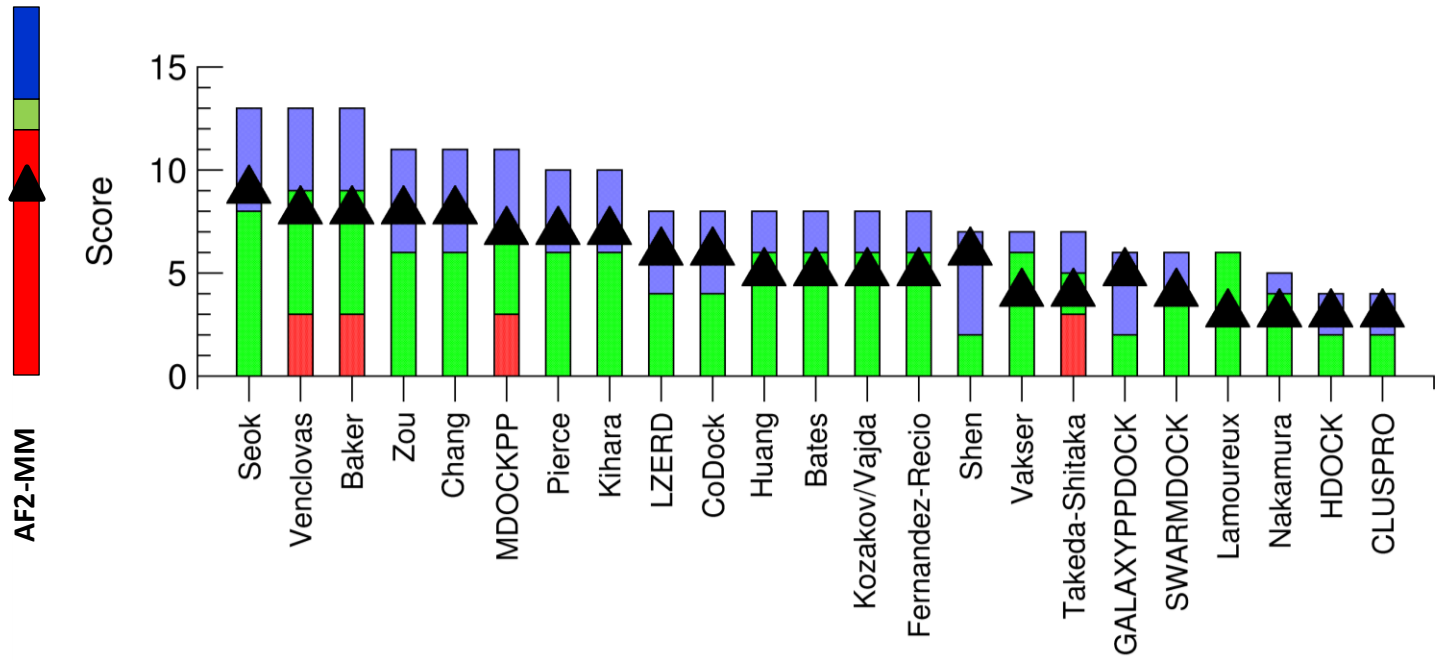
Scorer	Performance
Takeda-Shitaka	10/9***/1**
MULTICOM	10/6***/3**
LZERD, Kihara	9/2***/6**
S_Chang	10/1***/7**
S_Huang, HDOCK	4/1***/2**
Venclovas	10/8**
Zou, MDOCKP	6/5**

CAPRI Round 50 / CASP14 Ranking



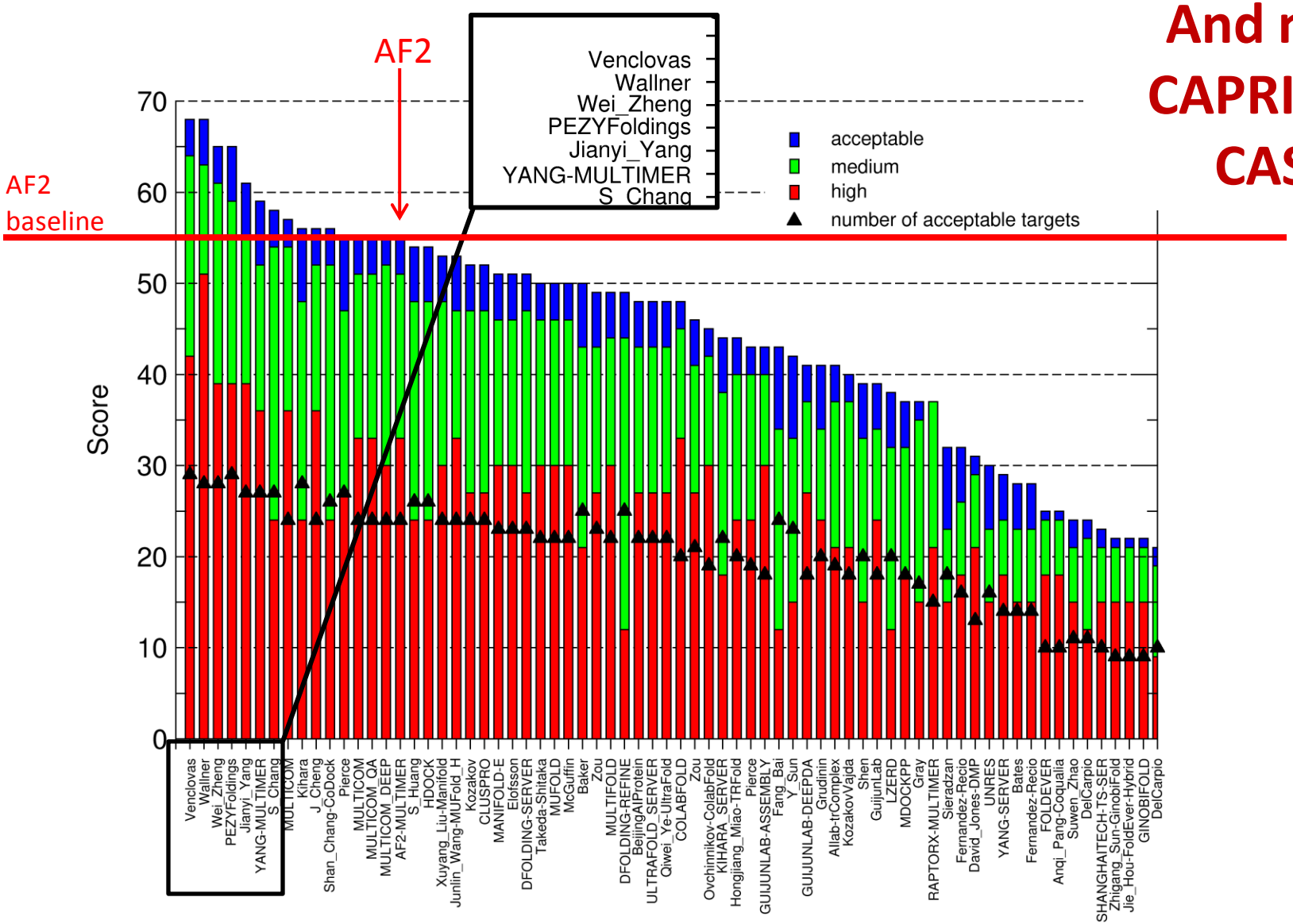
Only AlphaFold2 monomer existed

CAPRI Round 50 / CASP14 Ranking

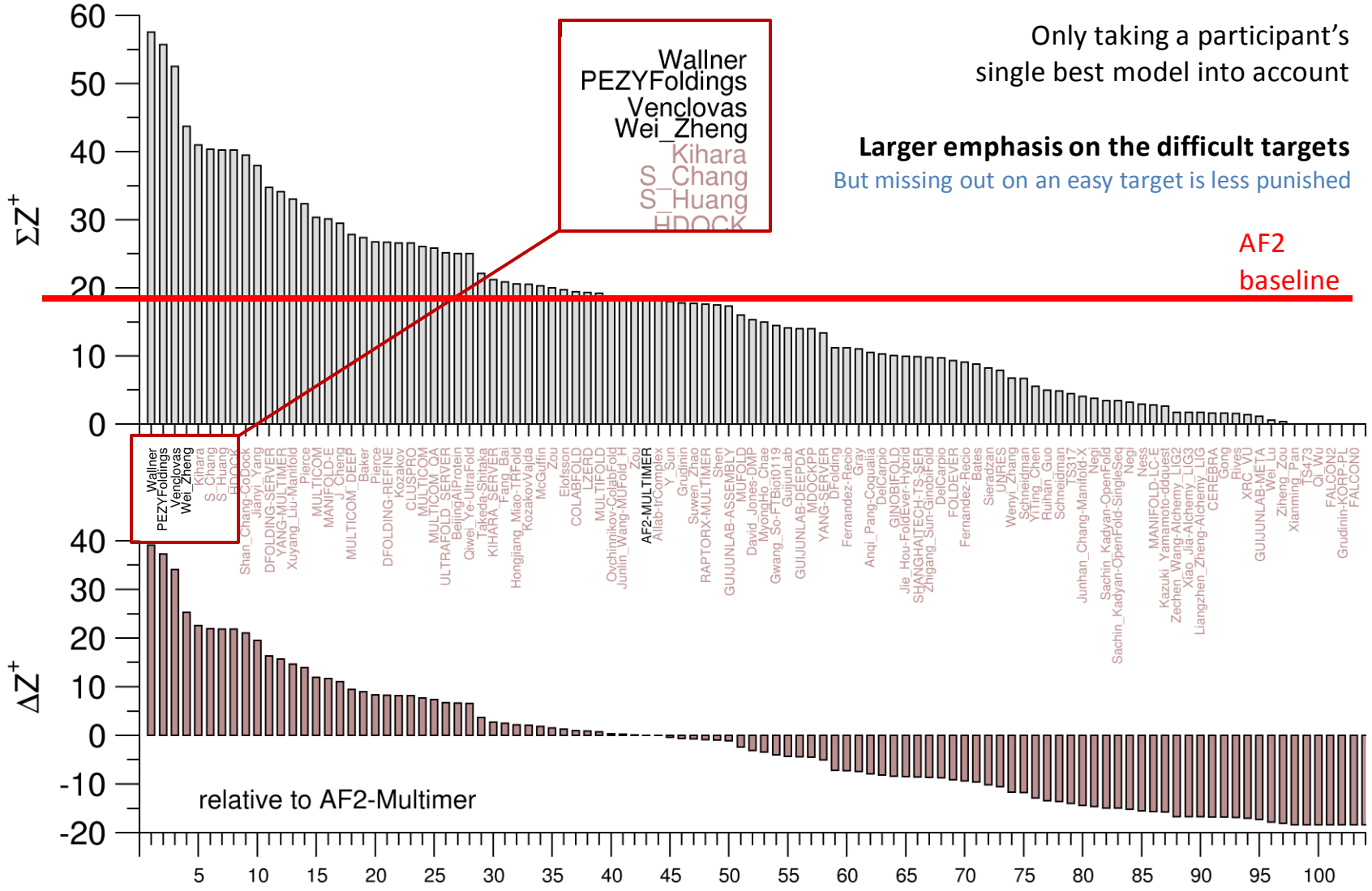


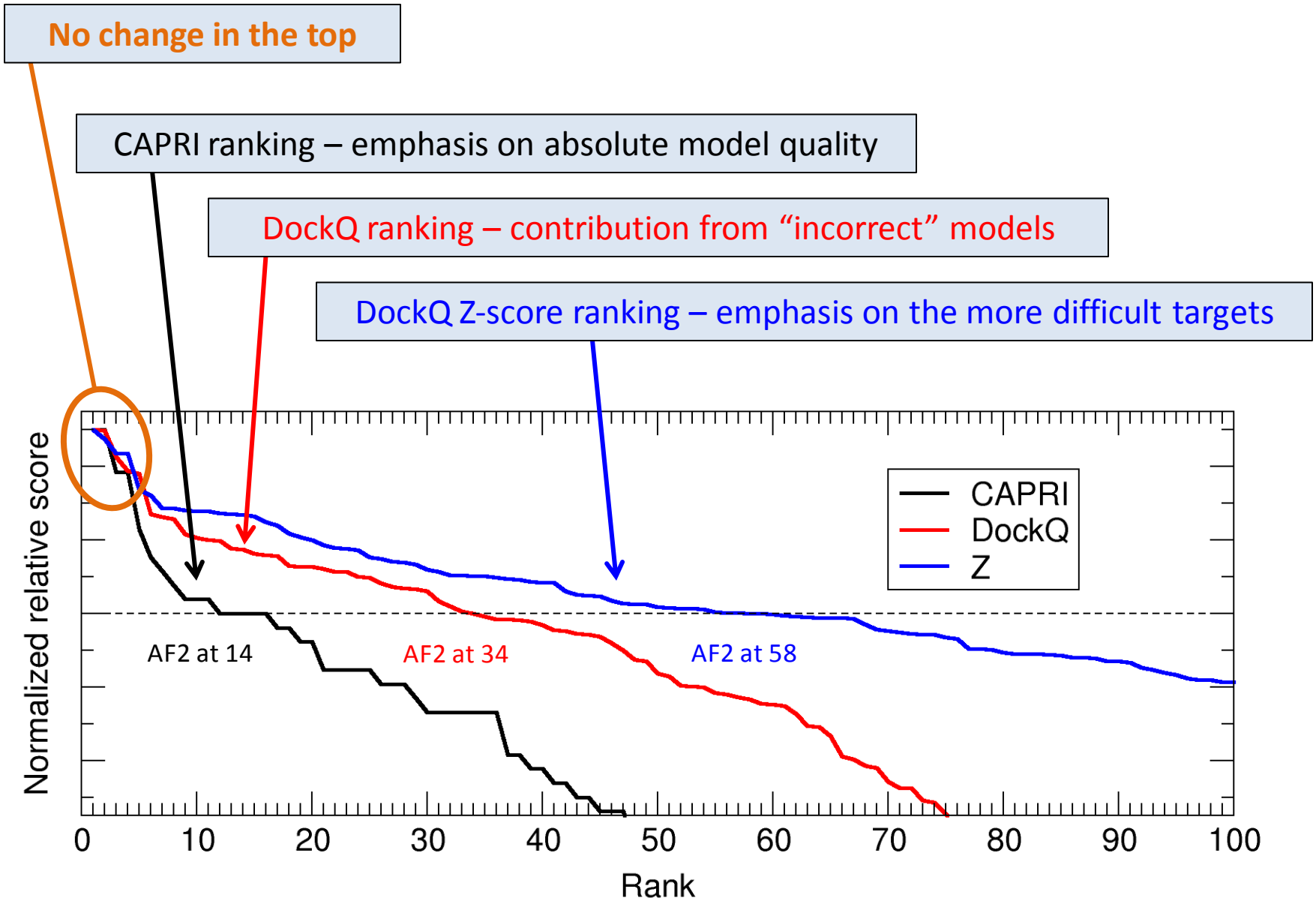
"*Off-the-shelf*" AF2-Multimer: same number of targets, but increased quality

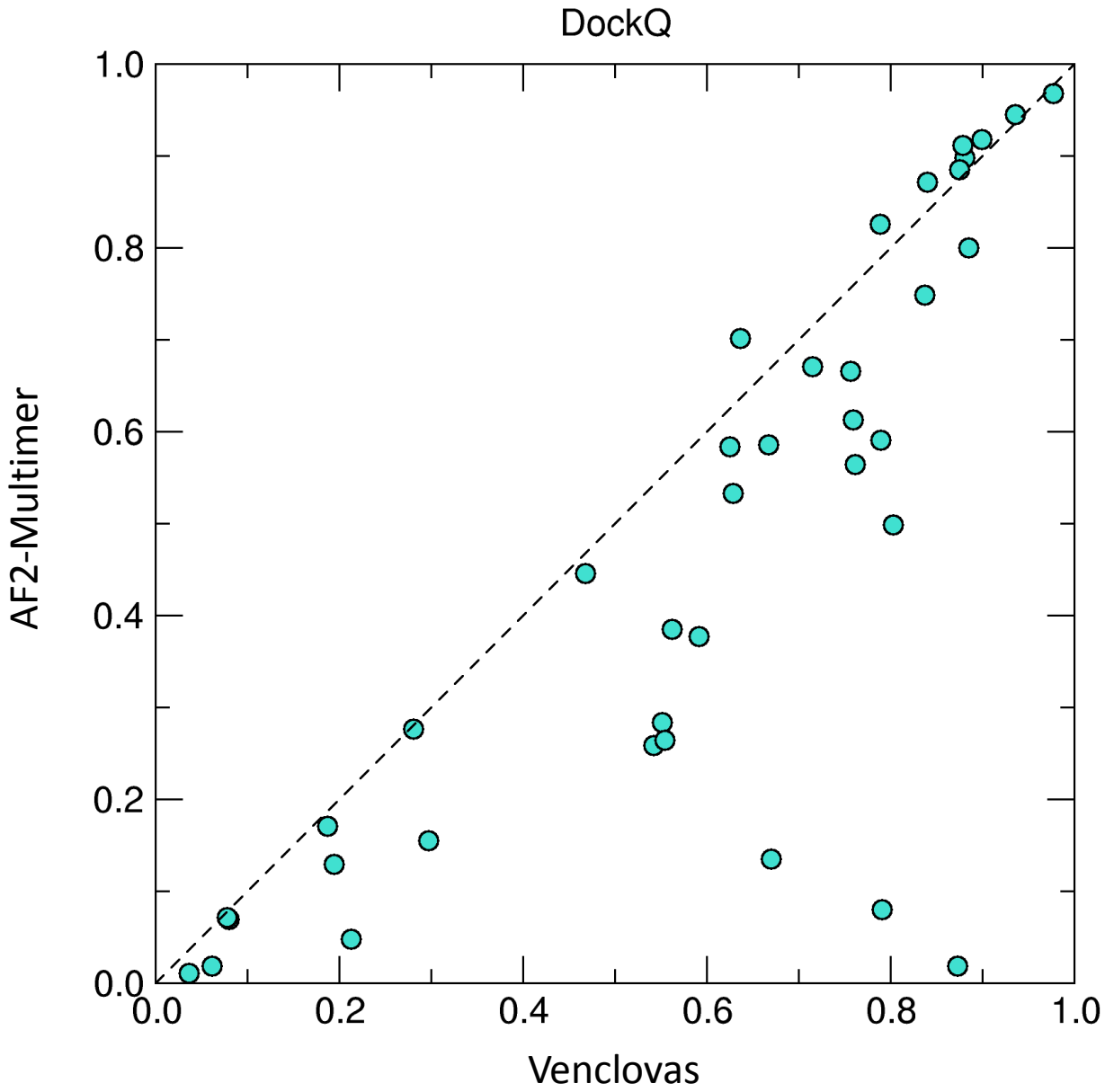
And now: CAPRI R54 CASP15

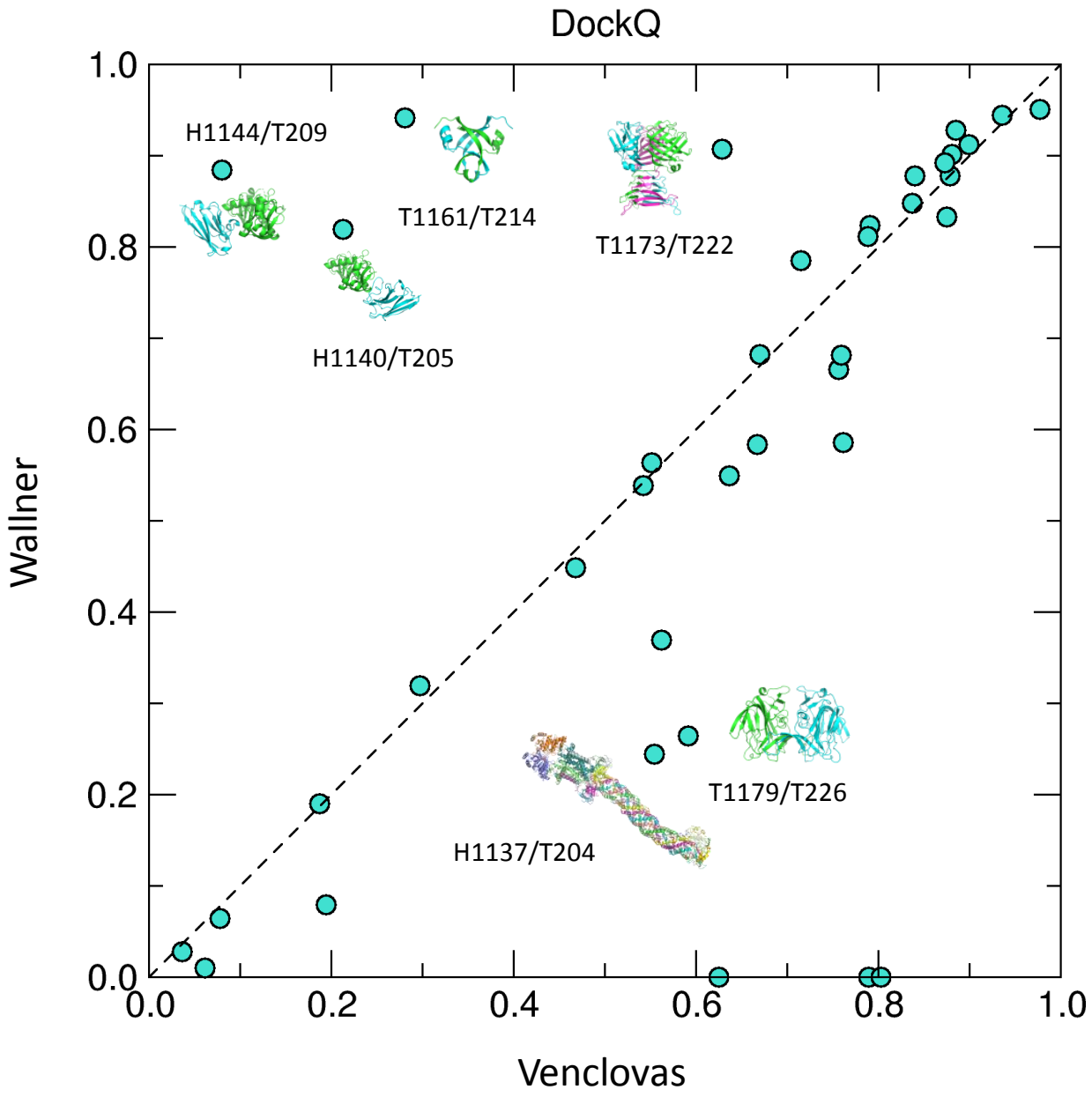


Calculate μ and σ for top-5 models P and TS set, removing exact duplicates
Express DockQ in σ (Z-score), retain only positives
Sum




No change in the top
CAPRI ranking – emphasis on absolute model quality
DockQ ranking – contribution from “incorrect” models
DockQ Z-score ranking – emphasis on the more difficult targets
Normalized relative score
CAPRI
DockQ
Z
AF2 at 14
AF2 at 34
AF2 at 58
0
10
20
30
40
50
60
70
80
90
100
Rank


DockQ
AF2-Multimer
Venclovas
0.0
0.2
0.4
0.6
0.8
1.0


DockQ
Wallner
Venclovas
0.0
0.2
0.4
0.6
0.8
1.0
H1144/T209
T1161/T214
T1173/T222
H1140/T205
T1179/T226
H1137/T204

Received: 7 July 2023 | Revised: 22 September 2023 | Accepted: 28 September 2023

DOI: 10.1002/prot.26609

RESEARCH ARTICLE

PROTEINS WILEY

Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment

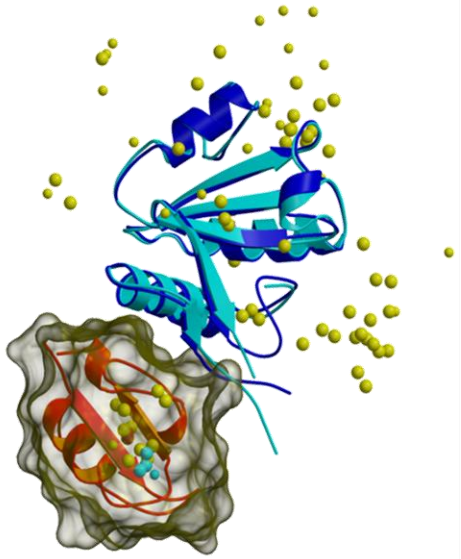
113 authors

Marc F. Lensink[1] | Guillaume Brysbaert[1] | Nessim Raouraoua[1] | Paul A. Bates[2] | Marco Giulini[3] | Rodrigo V. Honorato[3] | Charlotte van Noort[3] | Joao M. C. Teixeira[3] | Alexandre M. J. J. Bonvin[3] | Ren Kong[4] | Hang Shi[4] | Yufeng Liu[4] | Shan Chang[4] | Jian Liu[5] | Zhiye Guo[5] | Xiao Chen[5] |

Proteins special issue on CASP15

Lensink & Méndez, *Curr Pharm Biotechnol* **2008**;9:77

"Recognition-induced conformational changes in protein-protein docking"



Protein A Coordinates

Protein B Coordinates

Flexibility Analysis

MD-PCA / ED
NMA / GNM / ANM
Graph Theory

Biological Information

Protein A & B Conformers

Rigid body Search

FFT-methods
GA / GH, Monte Carlo
Neural networks

Raw score Filtering

Surface complementarity,
Electrostatics

Energy-based Re-ranking

Knowledge-based potentials,
Force-field based energy functions

MM or MD, implicit or explicit solvent

Refinement

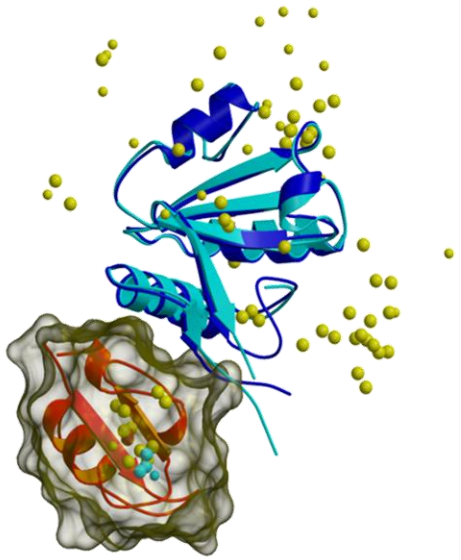
Complex AB Coordinates

Lensink *et al*, *Proteins* **2023**, in press

"Impact of AlphaFold on structure prediction of protein complexes"

Protein A Sequence

Protein B Sequence



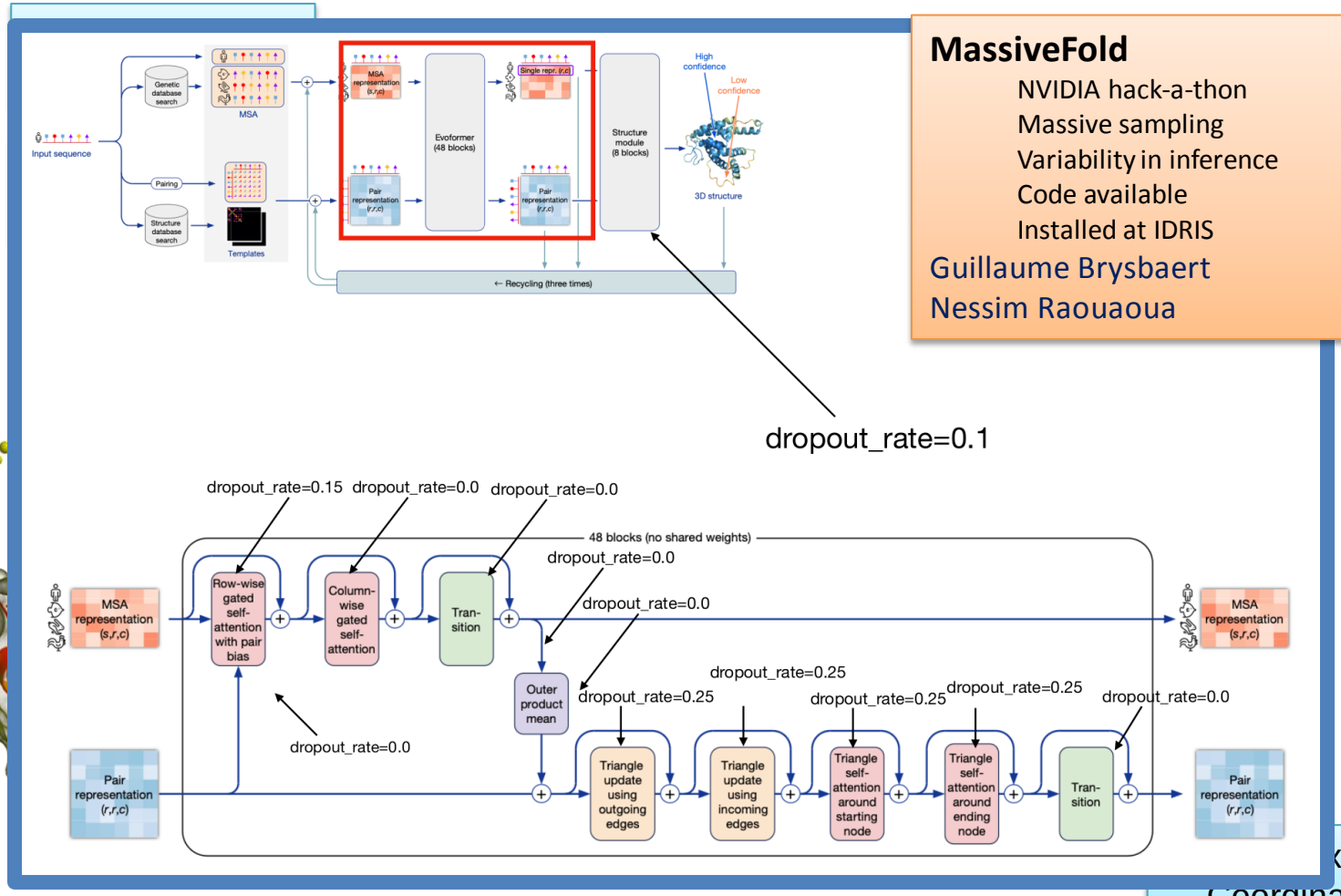
Biological Databases

AlphaFold?

Complex AB Coordinates

Lensink *et al*, *Proteins* **2023**, Epub

"Impact of AlphaFold on structure prediction of protein complexes"



MassiveFold

- NVIDIA hack-a-thon
- Massive sampling
- Variability in inference
- Code available
- Installed at IDRIS

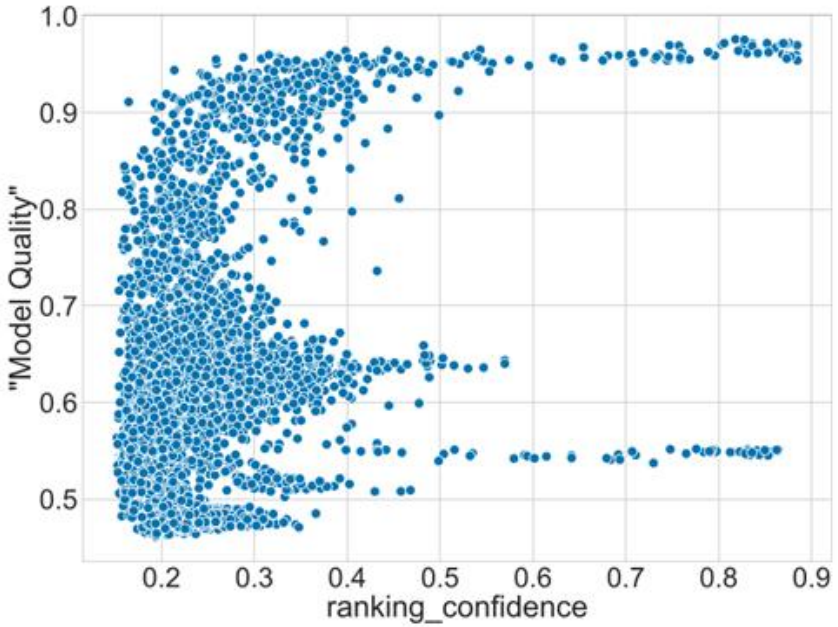
Guillaume Brysbaert

Nessim Raouaoua

How did they get better results than basic AlphaFold runs ?

Björn Wallner:
=> agressive sampling: produced up to 30000 predicted structures per target

- Use AlphaFold to generate models (full_dbs)
- Turn on dropout at inference
  1. All dropout_rates at default
  2. No dropout in the structural module
- Generate many structures with a different random seed
- Use both:
  - *multimer_v1*
  - *multimer_v2*
- Run with and without templates
- Run with an increased number of recycles
- Select based on the ranking_score

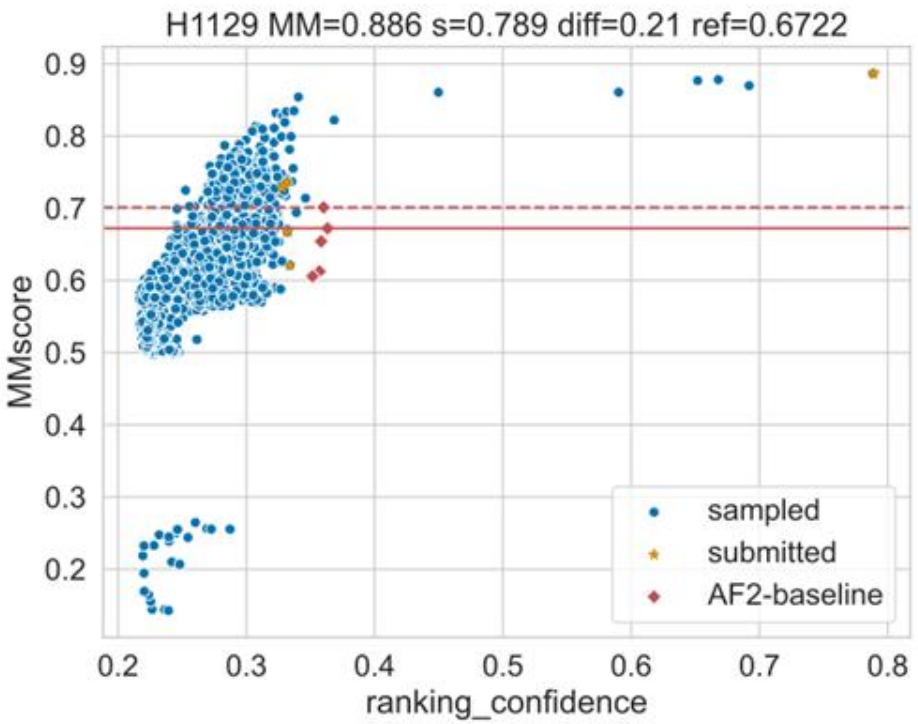


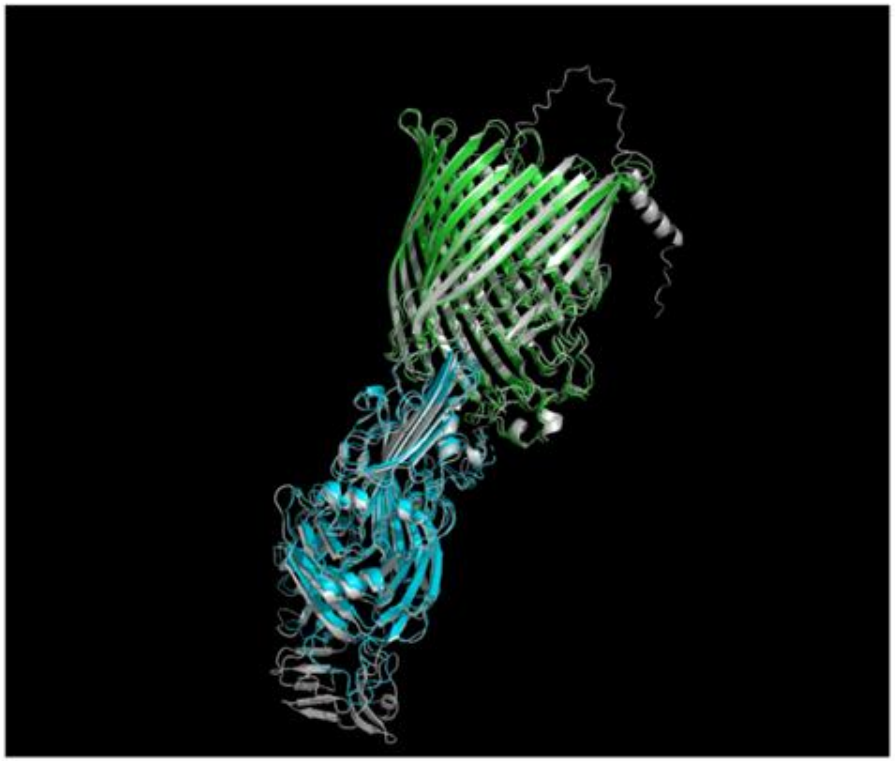
Björn Wallner, CASP15, 2022

How did they get better results than basic AlphaFold runs ?

Björn Wallner:
=> example

H1129





Björn Wallner, CASP15, 2022

Team developments

- MassiveFold – bio.tools/massivefold
  - Enhanced AlphaFold
    - Can be run in parallel
    - Highly customizable (dropout, recycle, number of models)
- RINspector – apps.cytoscape.org/apps/rinspector
  - Centrality and flexibility analysis of protein structures
    - Exists as app for cytoscape
    - Command-line version in development
- Scoreset – scoreset.org
  - Protein-protein interaction benchmark dataset
    - 176k annotated models
    - Includes correct and incorrect models

Computational Biology:
Biomolecular Interactions and Dynamics

Institute for Structural and Functional Glycobiology

- **Protein interaction and regulatory networks**
- **Statistical physics of biomolecular interactions**
- **Structural biology of protein-carbohydrate interactions**
- **Computational modeling and dynamics of biomolecular systems**