

37C3
UNLOCKED

AlphaFold – how machine learning changed structural biology forever (or not?)

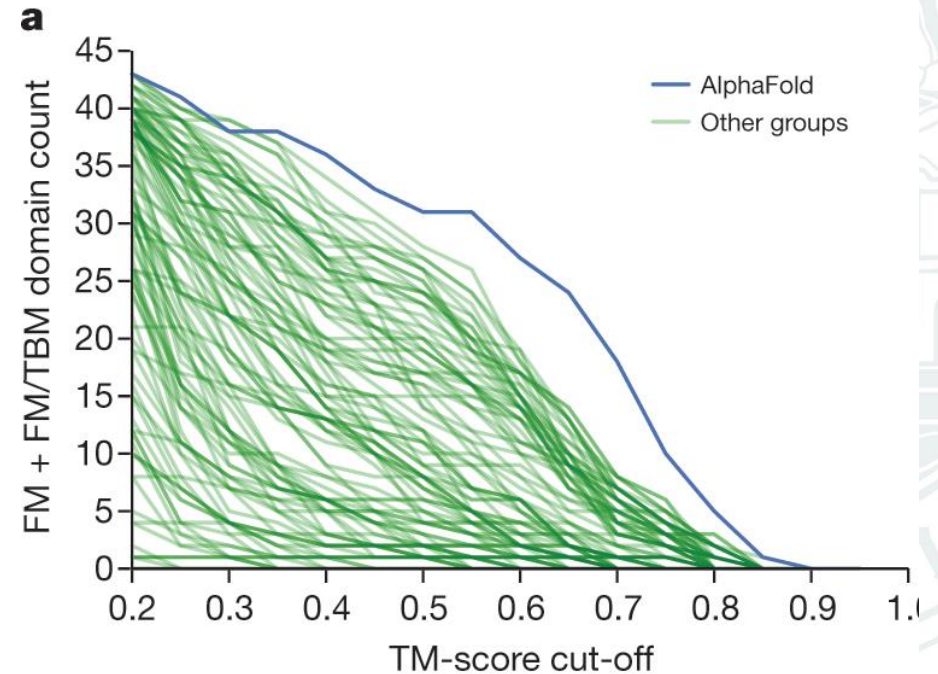
Jan Gebauer

@JanGebauer@fediscience.org

And the winner of the 13th CASP Awards is...



Critical Assessment of protein Structure Prediction



And the winner of the 13th CASP Awards is...

nature

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [articles](#) > article

Article | [Published: 15 January 2020](#)

Improved protein structure prediction using potentials from deep learning

[Andrew W. Senior](#) , [Richard Evans](#), [John Jumper](#), [James Kirkpatrick](#), [Laurent Sifre](#), [Tim Green](#), [Chongli Qin](#), [Augustin Židek](#), [Alexander W. R. Nelson](#), [Alex Bridgland](#), [Hugo Penedones](#), [Stig Petersen](#), [Karen Simonyan](#), [Steve Crossan](#), [Pushmeet Kohli](#), [David T. Jones](#), [David Silver](#), [Koray Kavukcuoglu](#) & [Demis Hassabis](#)

Nature **577**, 706–710 (2020) | [Cite this article](#)



Jan Gebauer @jan_gebauer · 3. Dez. 2018

Interestingly, Google seems to have beaten all the established researchers on structure prediction. However, haven't found any scientific pub yet. Curious if it's better method or "just" more resources. Google's DeepMind predicts 3D shapes of proteins theguardian.com/science/2018/d...



1



3



Jan Gebauer
@jan_gebauer

Found at least some stuff on [#GoogleSheets](#) homepage - however, there is no peer-reviewed publication plannend...?

[#AlphaFold: Using AI for scientific discovery](#)
deepmind.com/blog/alphafold/

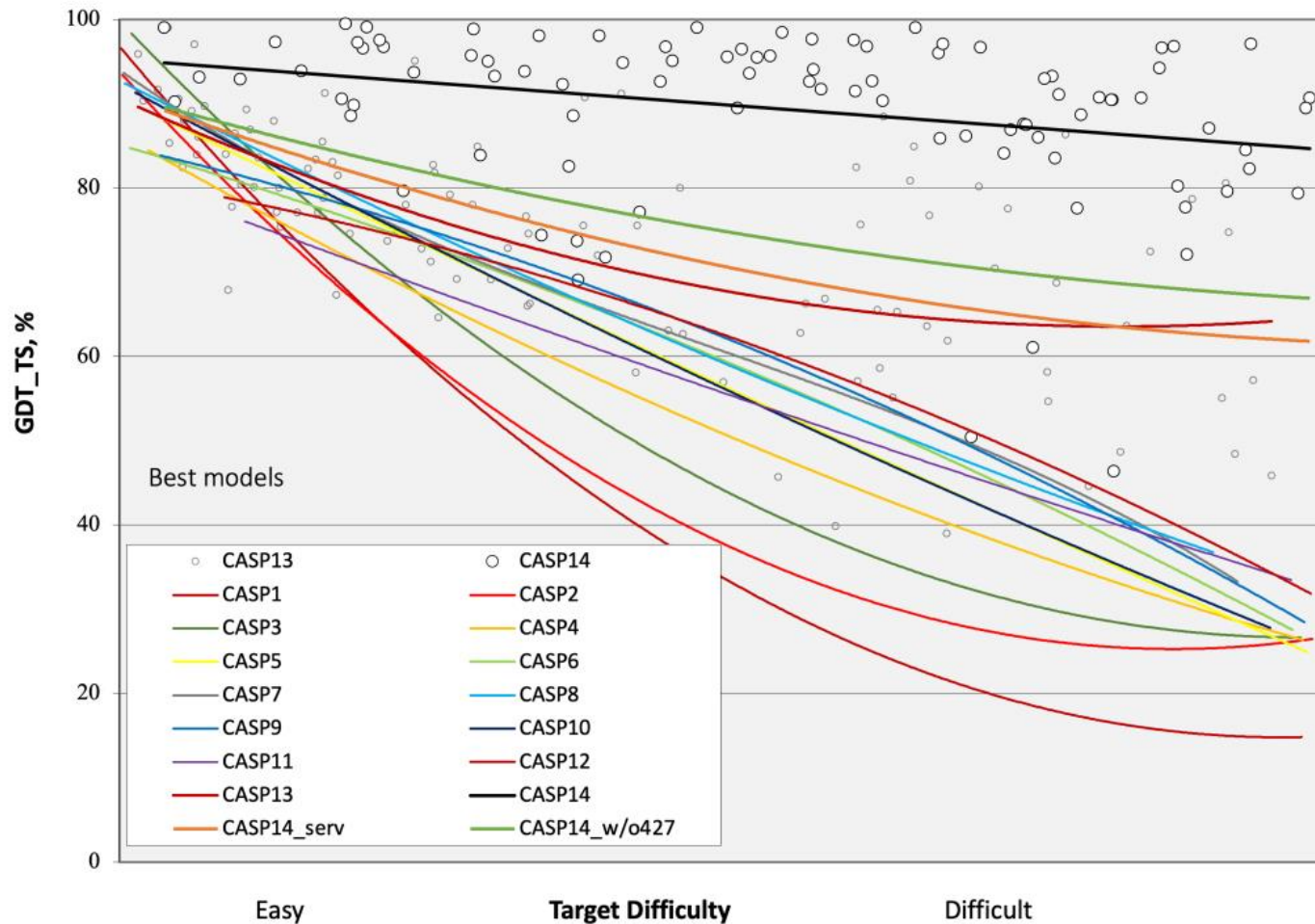
Post übersetzen

10:56 nachm. · 3. Dez. 2018

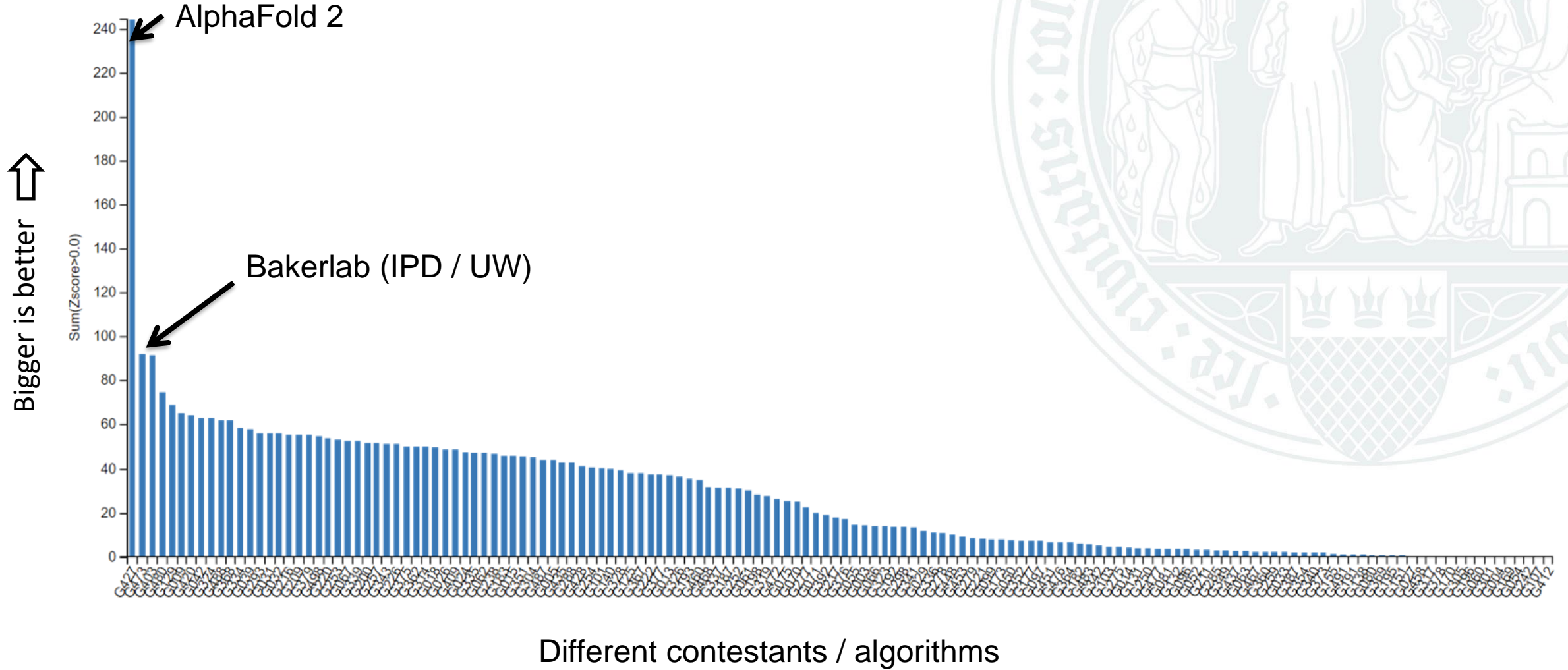
And the winner of the 14th CASP Awards is...



Critical Assessment of protein Structure Prediction



And the winner of the 14th CASP Awards is...



News coverage

nature

“This i
Thorn
Bioinf

Explore

nature

NEWS

‘It
De
ins

The New York Times

London A.I. Lab Claims That Could Accelerate Dr

Researchers at DeepMind say they have solved the “protein folding problem,” a task that has bedeviled scientists for more than 50 years.

<https://www.nytimes.com/2020/11/30/technology/>

Google’s deep-learning program for determining the 3D shapes of proteins could transform biology, say scientists.

By [Ewen Callaway](#)



Illustration: Studio Pong / DER SPIEGEL

KI-System AlphaFold

Should this machine win the Nobel Prize?

The most important discovery of the past year in the field of chemistry was made by software. There should be a Nobel Prize for that. The only question is: to the artificial intelligence – or to its programmers?

Von **Johann Grolle**

October 4th, 2022, 12:15 a.m • aus **DER SPIEGEL 40/2022**

News coverage

□ SEARCH

Spektrum.de

MAGAZINE | ARCHIVE | ABO/SHOP | SERVICE | LOGIN

ASTRONOMY | BIOLOGY | CHEMISTRY | EARTH/ENVIRONMENT | IT/TECH | CULTURAL | MATHEMATICS | MEDICINE | PHYSICS | PSYCHOLOGY/BRAIN RESEARCH

[Home page](#) » [IT/Tech](#) » AlphaFold 2: Everyone can now fold proteins

OPEN-SOURCE-SOFTWARE

Anyone can now fold proteins

Last year, AlphaFold 2 made a breakthrough on one of the most important problems in biology. Now anyone who wants can use the AI software. And it's free of charge.

von [Ewen Callaway](#)

<https://www.spektrum.de/news/alphafold-2-jeder-kann-jetzt-proteine-falten/1896928>

News
19.07.2021
Reading time
approx. 4 minutes
[Press](#)
[Split](#)

The End Is Nigh ...

Google's po

The year 202
corona pande
election. It wi
machines beg

And the year
already based
entirely new f



Are we witnessing the dawn of post-theory science?

Does the advent of machine learning mean the classic methodology of hypothesise, predict and test has had its day?

by [Laura Spinney](#)

What are proteins?



What are proteins?



Coffee
Nutrition Facts

Serving size	100 g	DV
Calories	2 kcal	2%
Total Carbohydrate	0.3 g	0%
Total Fat	0 g	0%
Protein	0.1 g	0%
Niacin	0.2 mg	1%
Potassium	30 mg	1%
Sodium	4 mg	0%
Selenium	0.1 µg	0%

% Daily values (DV) are based on a 2000 calories Diet. DV may be higher or lower depending on your calorie needs.

1



Nutrition Facts
Serving Size 8 fl. oz. (240 mL)
Servings Per Container 2

Amount Per Serving	
Calories	80
	% Daily Value*
Total Fat	0g 0%
Sodium	75mg 3%
Total Carbohydrate	16g 5%
Sugars	16g
Protein	0g
Calcium	2%
Iron	0%

Not a significant source of calories from fat, saturated fat, trans fat, cholesterol, dietary fiber, vitamin A, and vitamin C.
*Percent daily values are based on a 2,000 calorie diet.



Nutrition Facts
2 servings per container
Serving size **1/2 pizza (162g)**

	Per serving	Per container
Calories	340	690
	% DV*	% DV*
Total Fat	21g 27%	42g 54%
Saturated Fat	8g 38%	15g 76%
Trans Fat	0g	0.5g
Cholesterol	50mg 17%	100mg 33%
Sodium	950mg 41%	1900mg 83%
Total Carb.	25g 9%	50g 18%
Dietary Fiber	19g 68%	38g 136%
Total Sugars	4g	7g
Protein	28g 56%	56g 112%
Vitamin D	0.1mcg 0%	0.2mcg 2%
Calcium	440mg 35%	870mg 70%
Iron	1.2mg 6%	2.4mg 15%
Potassium	350mg 8%	700mg 15%

* The % Daily Value (DV) tells you how much a nutrient in a serving of food contributes to a daily diet. 2,000 calories a day is used for general nutrition advice.

What are proteins?

Proteins serve as both the toolkit and the building blocks of your body



1

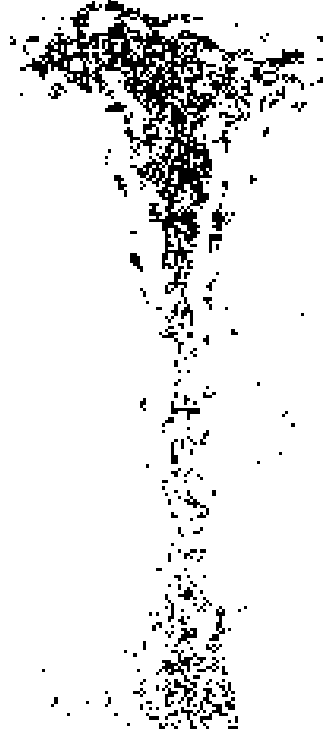
What are proteins?

Proteins are the tool set (and the building blocks) of your body...



How can we understand protein function?

Three “low-res” protein structures ($\sim 8\text{\AA}$):



Motor-domain
(ATPase?)

linker



Effector-domain
(Protease?)

Function: Mixing / Blending

How can we understand protein function?

Three “high-res” protein structures ($\sim 2\text{\AA}$):



Motor-domain
(ATPase?)

linker



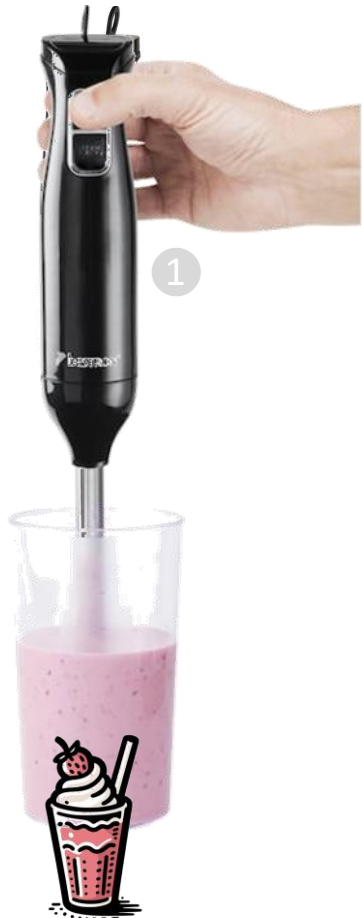
Effector-domain
(Protease?)

$2\text{\AA} = 0.2 \text{ nm} = 0.2 \times 10^{-9} \text{ m}$ (1/1'000'000'000) or 1\AA roughly the distance of two atoms in a molecule

Structures equals functions...



Structures equals functions...



Smoothies

2



bestron hand blender
»ASM 250 Z«, 250 W,
hand blender with 2
speed levels
Item no.: 8836134790

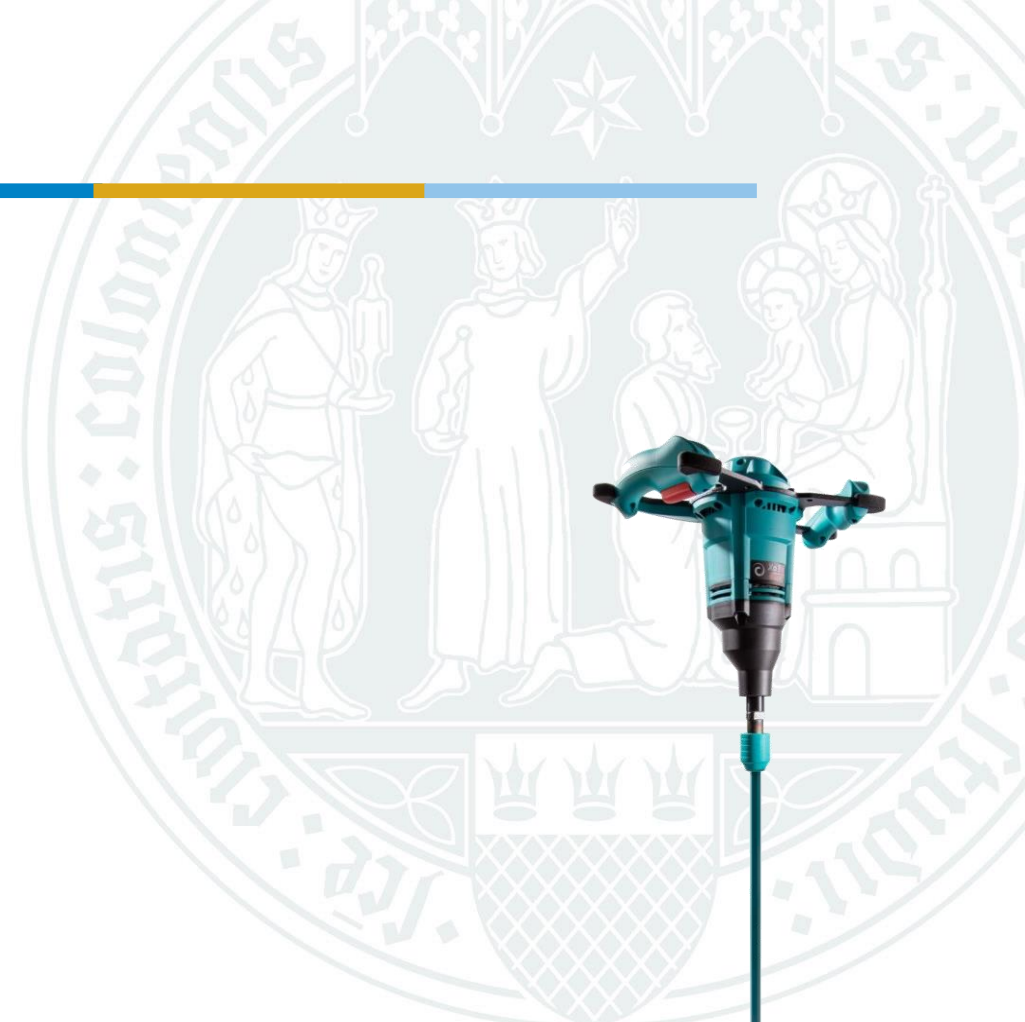
- Hand blender ideal for making soups, shakes, sauces or other dishes



Structures equals functions...



Smoothies



Structures equals functions...



Smoothies



CBD oil



Call us today! 203-267-4600

Shop Products

Homogenizer Applications

Home > Rotor-Stator Homogenizers

Deluxe Plus CBD Homogenizing Package

- Deluxe+ CBD cannabis homogenizing package perfect for all varieties of CBD, cannabinoid and cannabis homogenizing in medium-large beakers & gallon containers



Structures equals functions...



Milkshakes



CBD oil



Mortar



Collomix North America ²

<https://www.collomix.com> > produkte > xo1r-ruehrwerk ⋮

Compact agitator for mortar, glue, spatula | Xo 1 R

The **Xo 1 R** is a particularly light 1-speed agitator for stirrers up to 120 mm and smaller mixing quantities. Compact and yet powerful, this is how the Xo supports...

Structures equals functions...



Milkshakes



CBD oil



Mortar

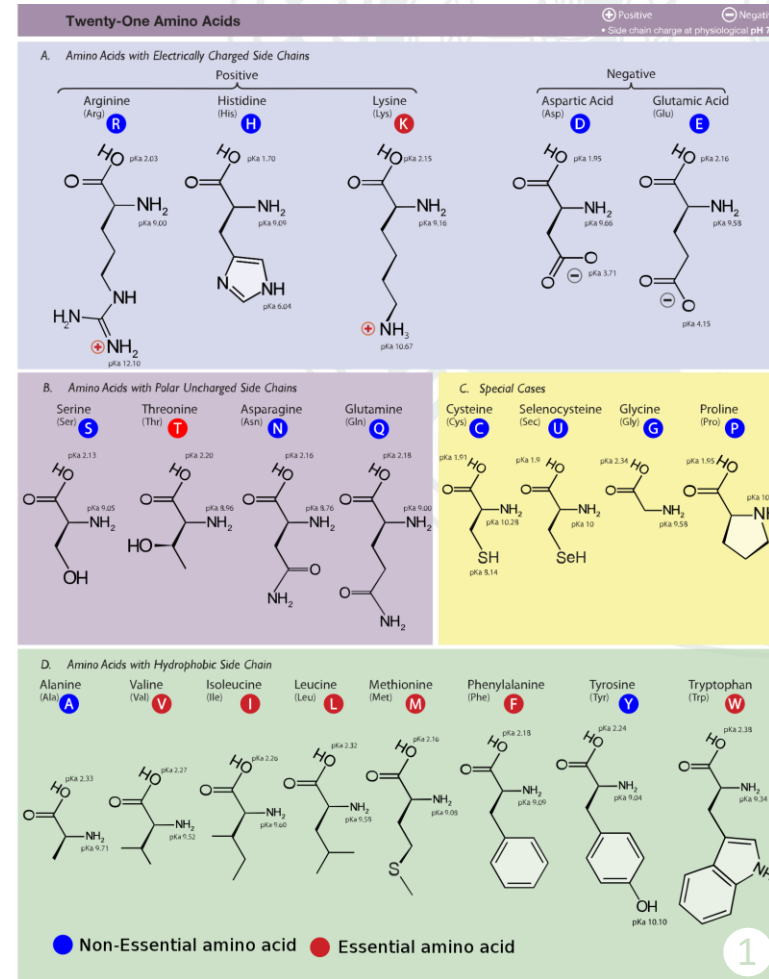
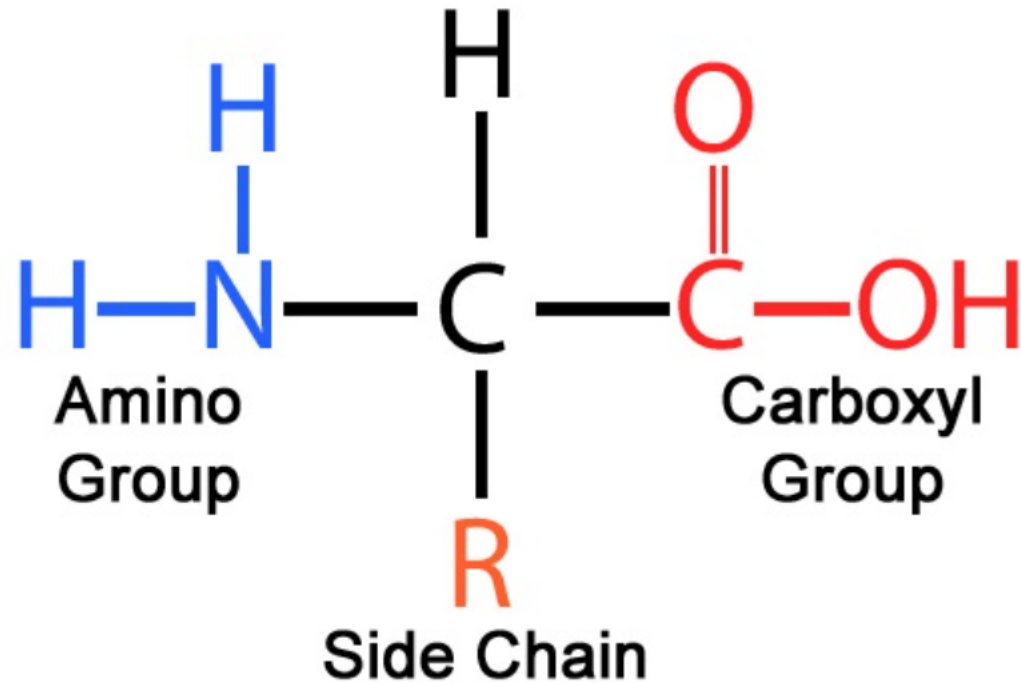
If we know the structure,
we can (often) infer the function!

Structure

Function

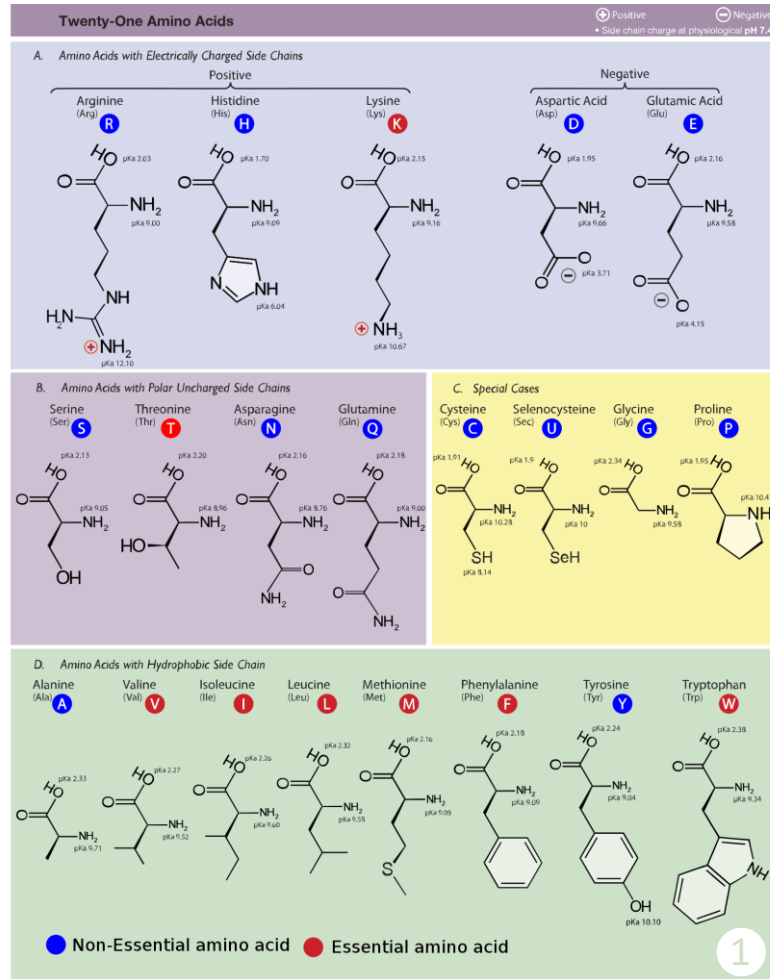
How are proteins build?

Proteins are long chains build from only 21 different building blocks (amino acids)



How are proteins build?

Proteins are long chains build from only 21 different building blocks (amino acids)



Let's simplify , down to six...



Different combinations give different structures.

Exemplary protein, e.g. a cleaving protein from the gut.

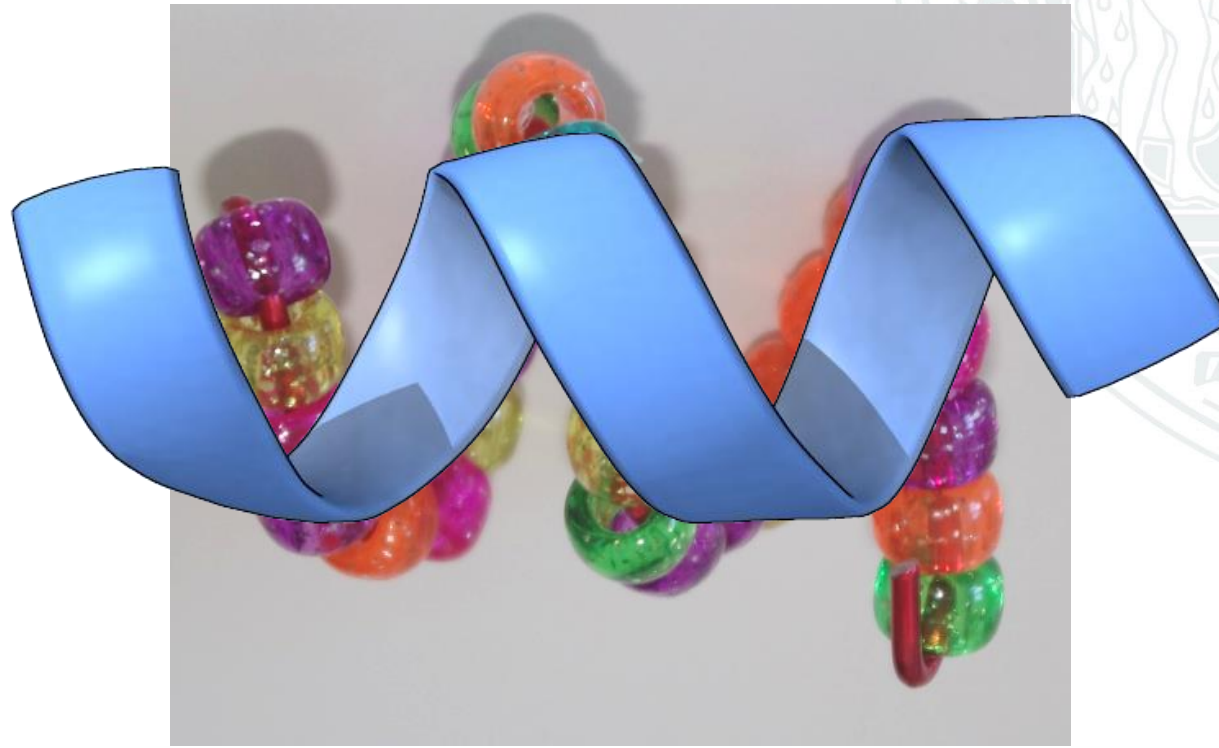


Other protein, e.g. electric conductor from the brain.



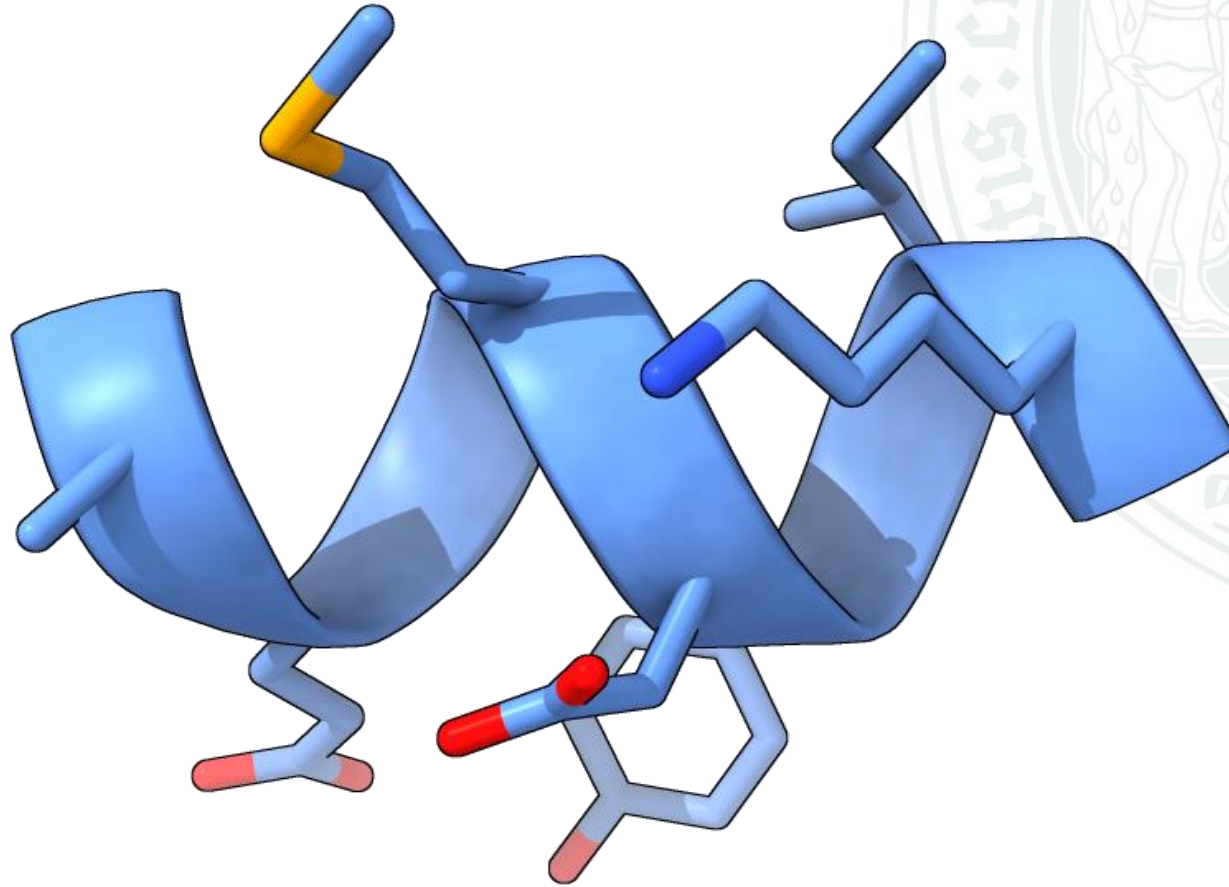
... they form complicated specific (!) “knots”

How do they “really” look?

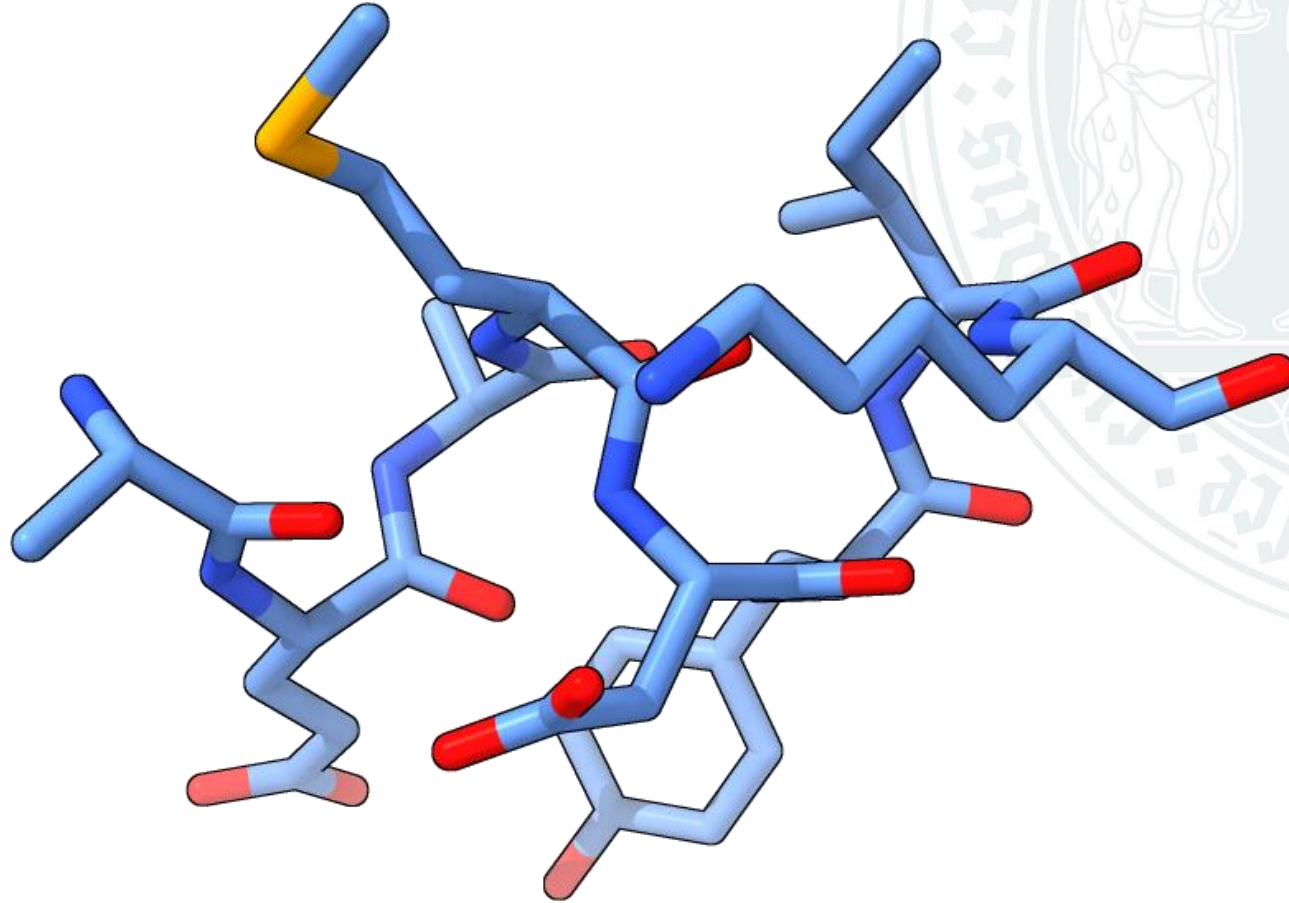


“Cartoon”

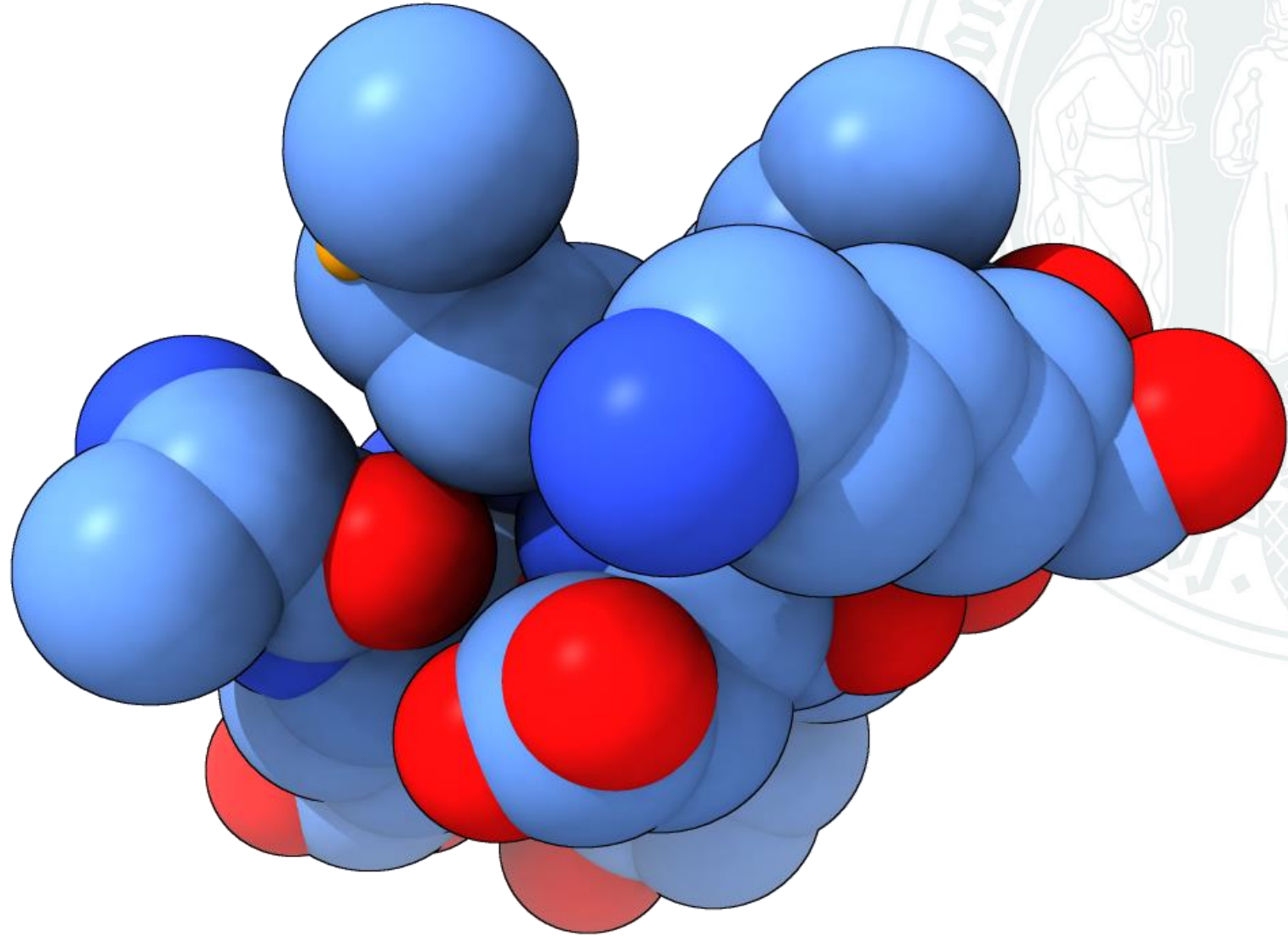
Different representation for different questions...



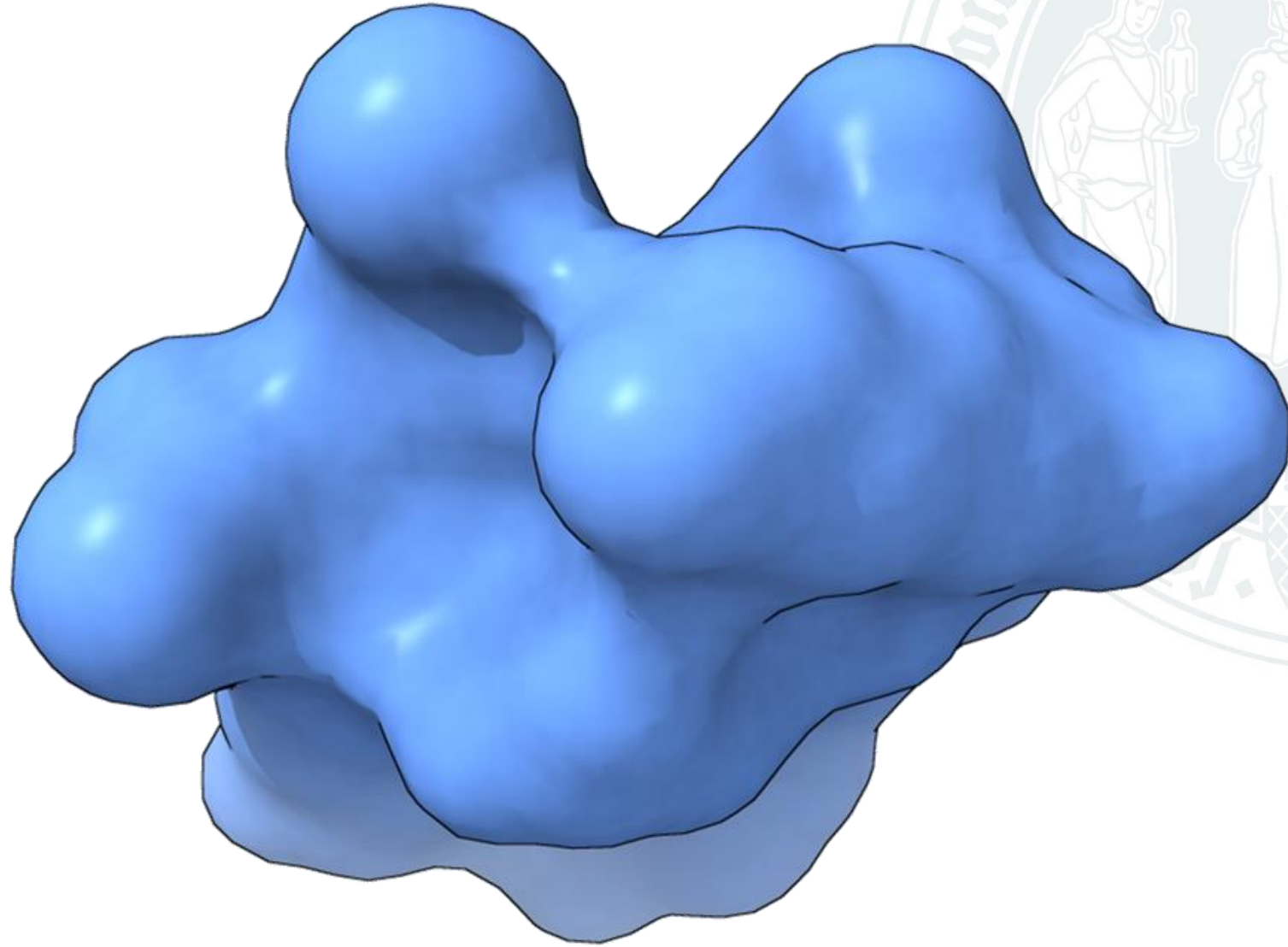
Different representation for different questions...



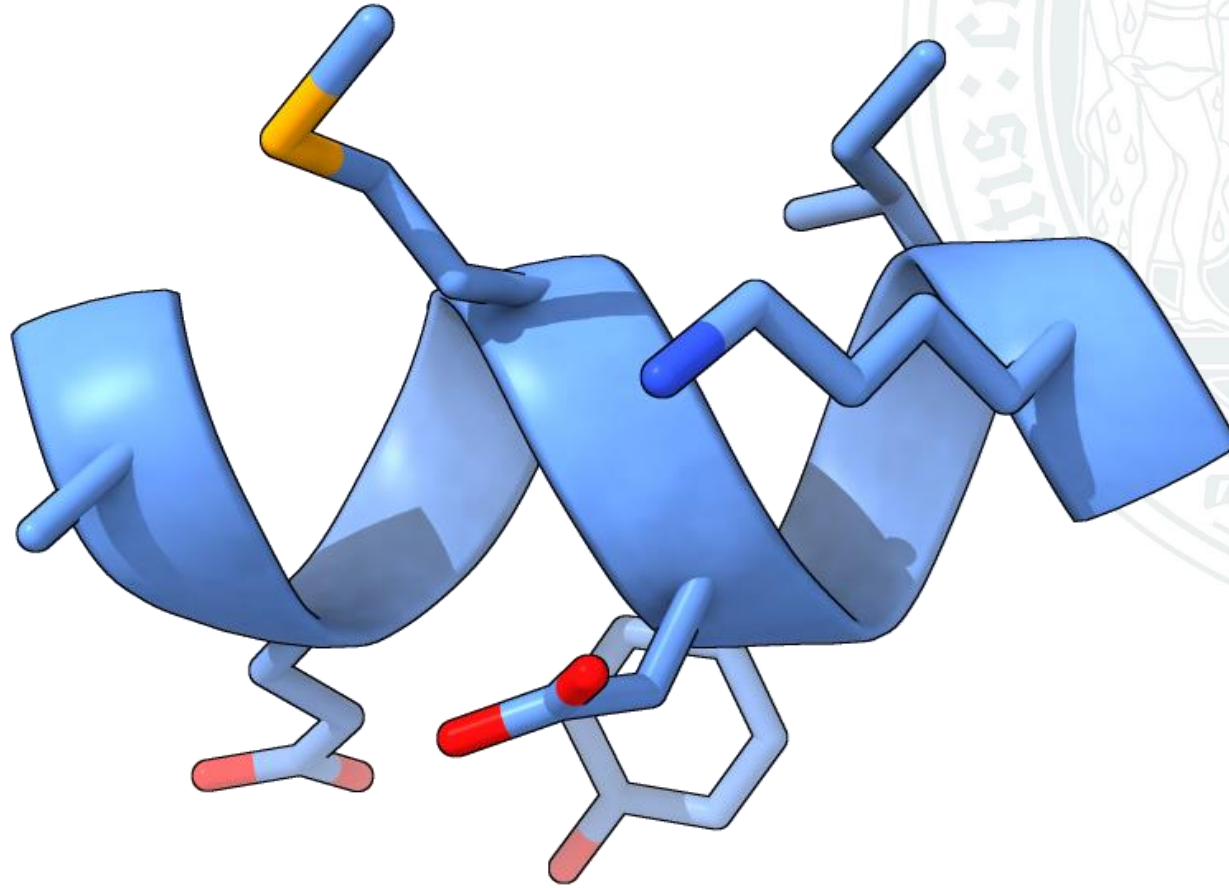
Different representation for different questions...



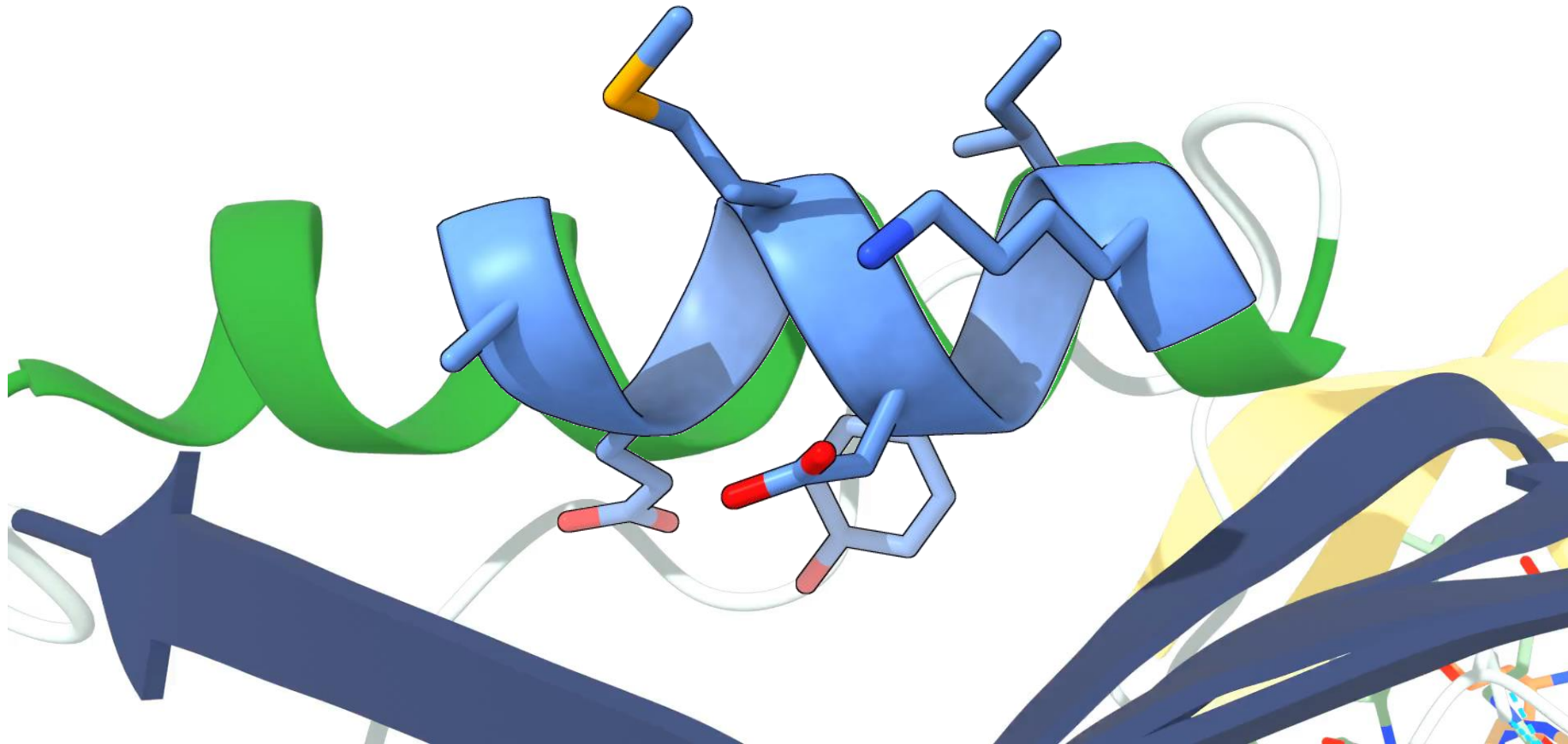
Different representation for different questions...



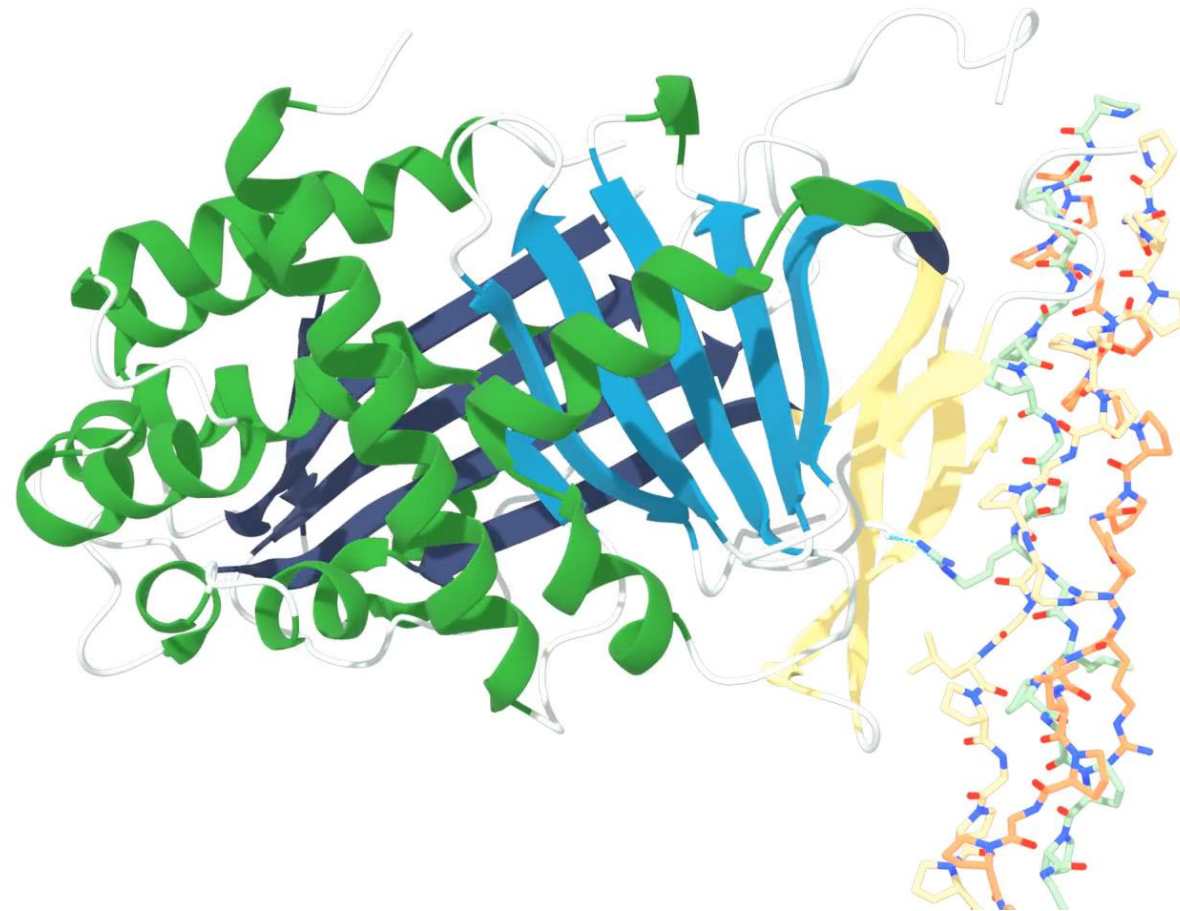
Different representation for different questions...



Obviously they are much bigger...



Obviously they are much bigger...



Amino acid “interactions” define the structure

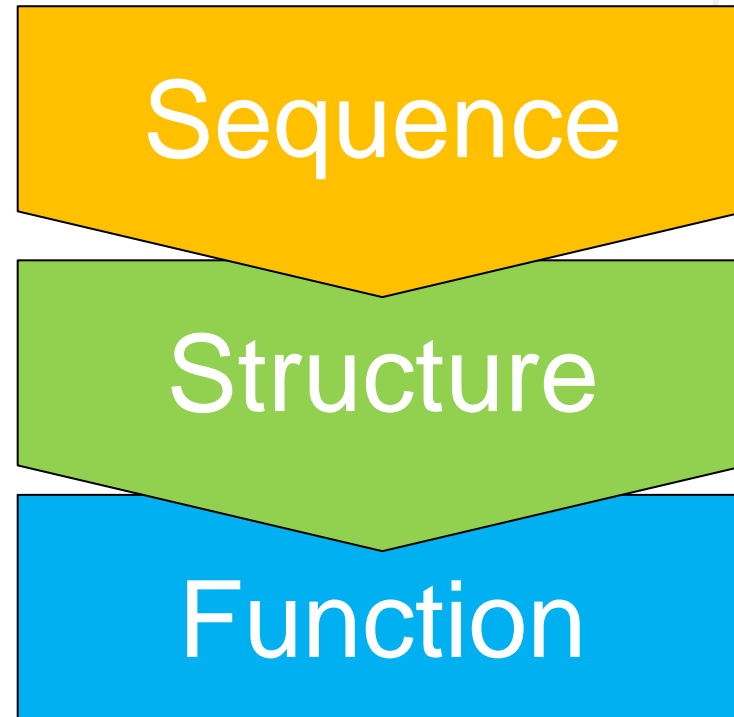
Sequence

Structure

Structure

Function

Amino acid “interactions” define the structure

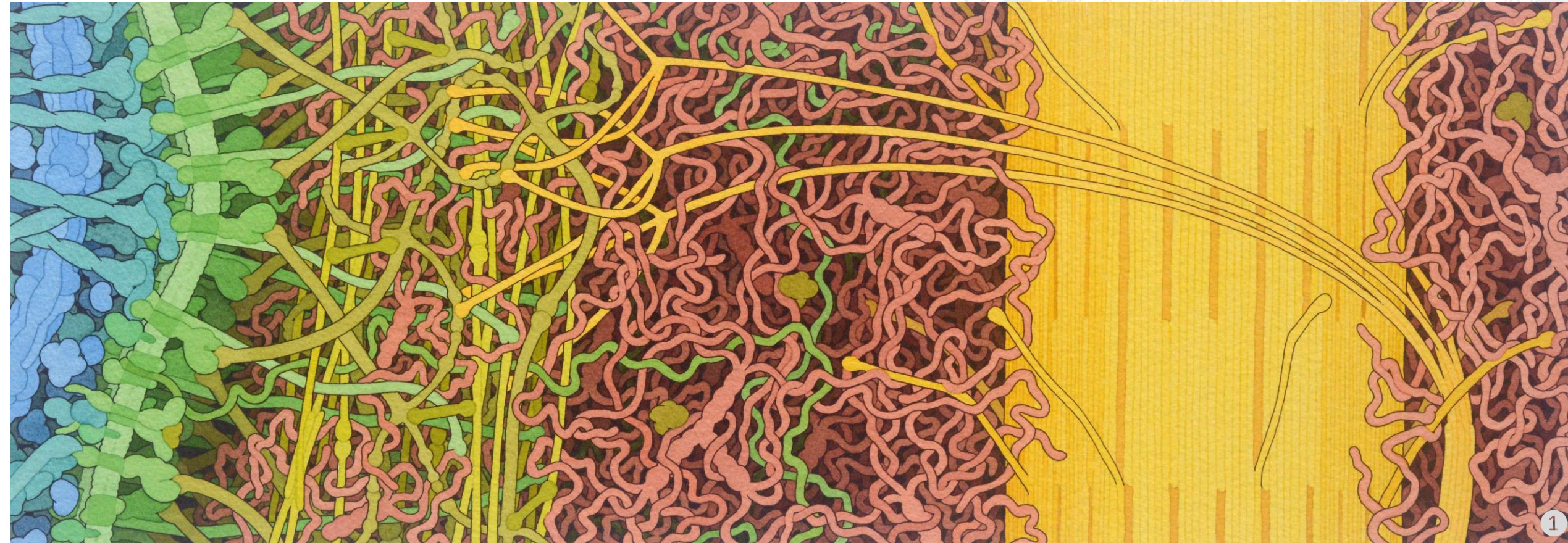


sequence-structure-function

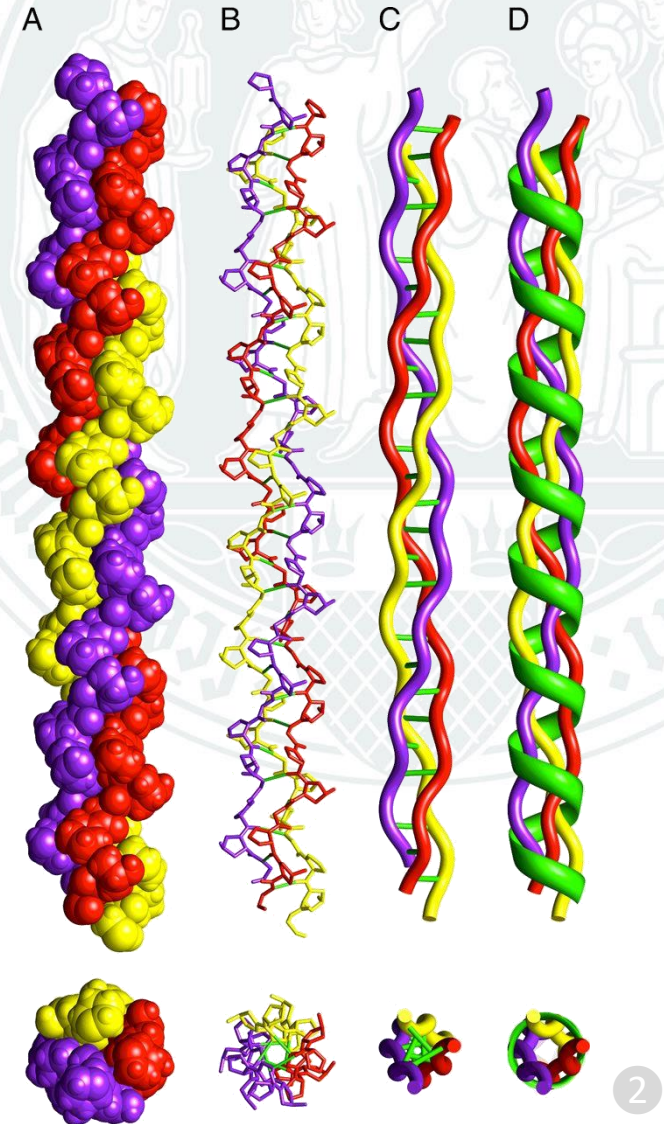
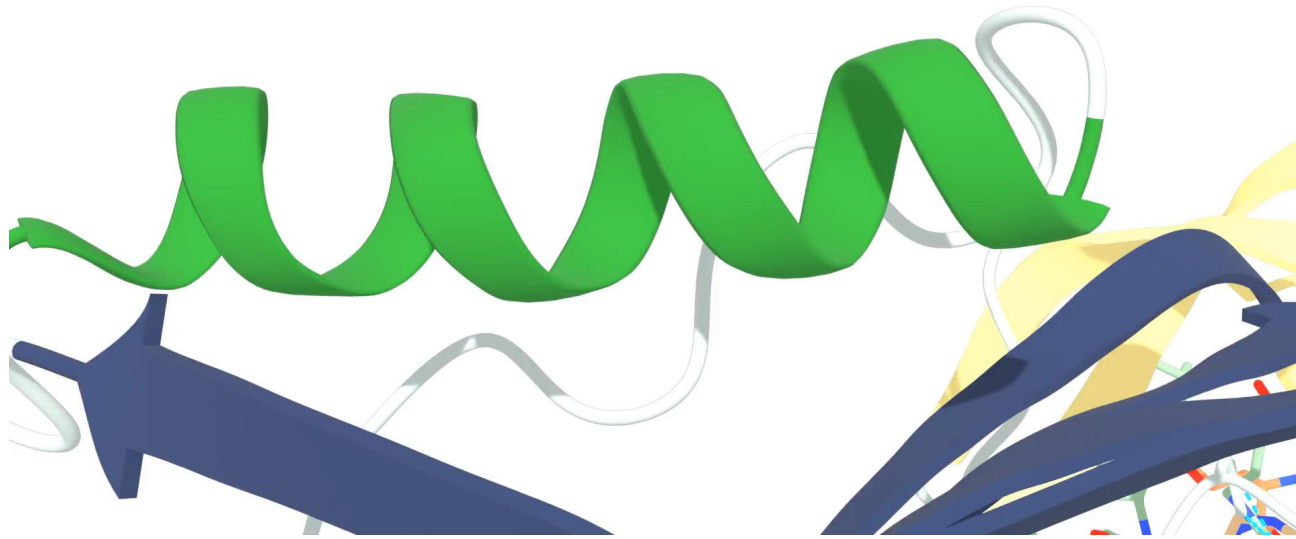
How useful is structural biology?



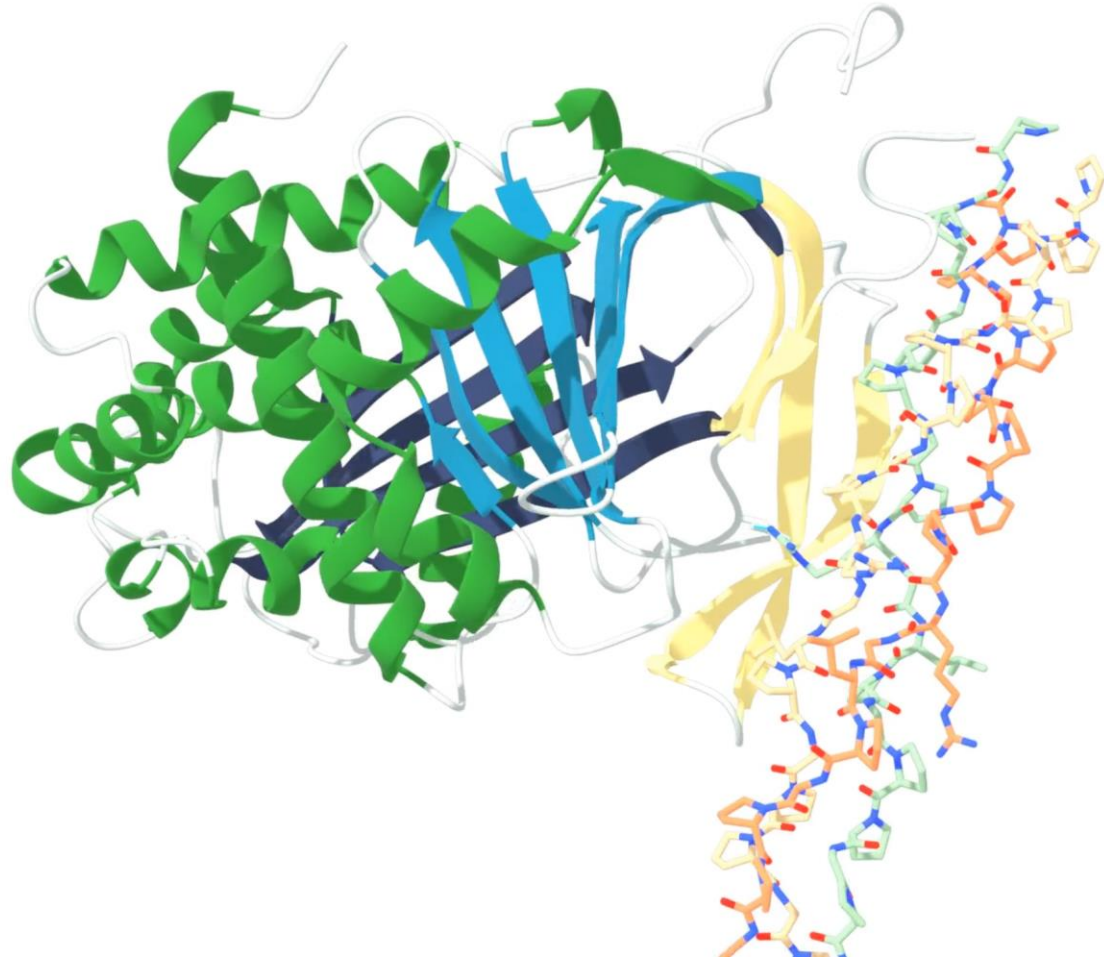
Basic research



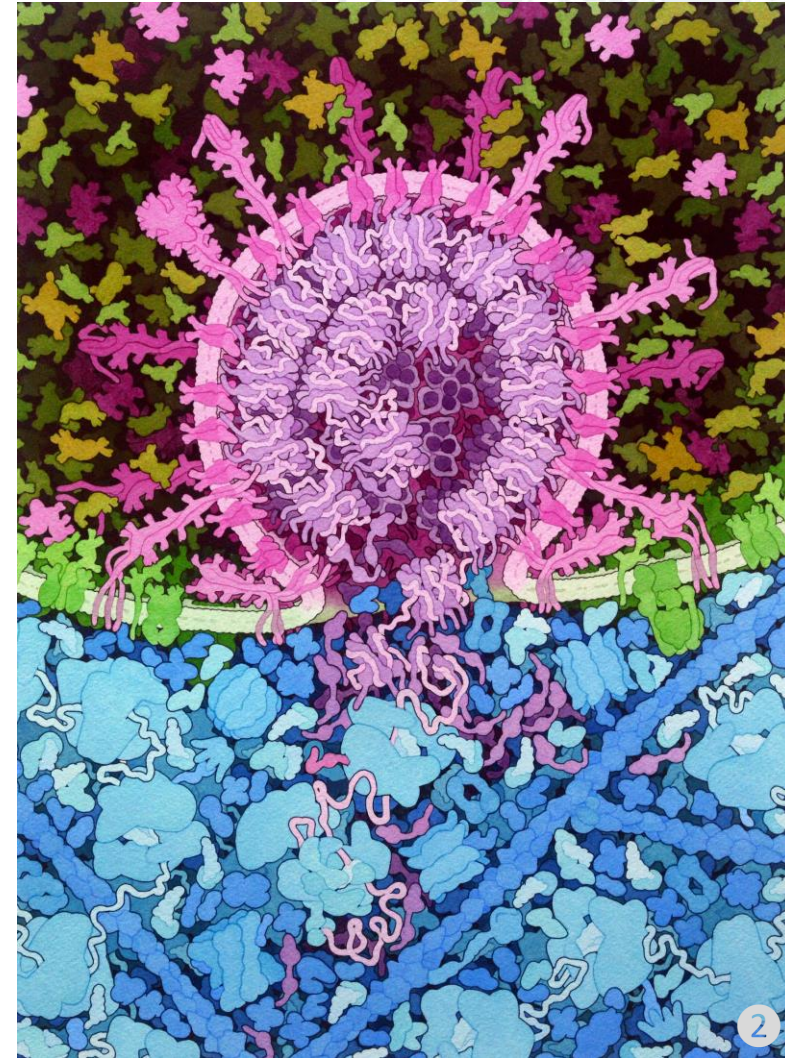
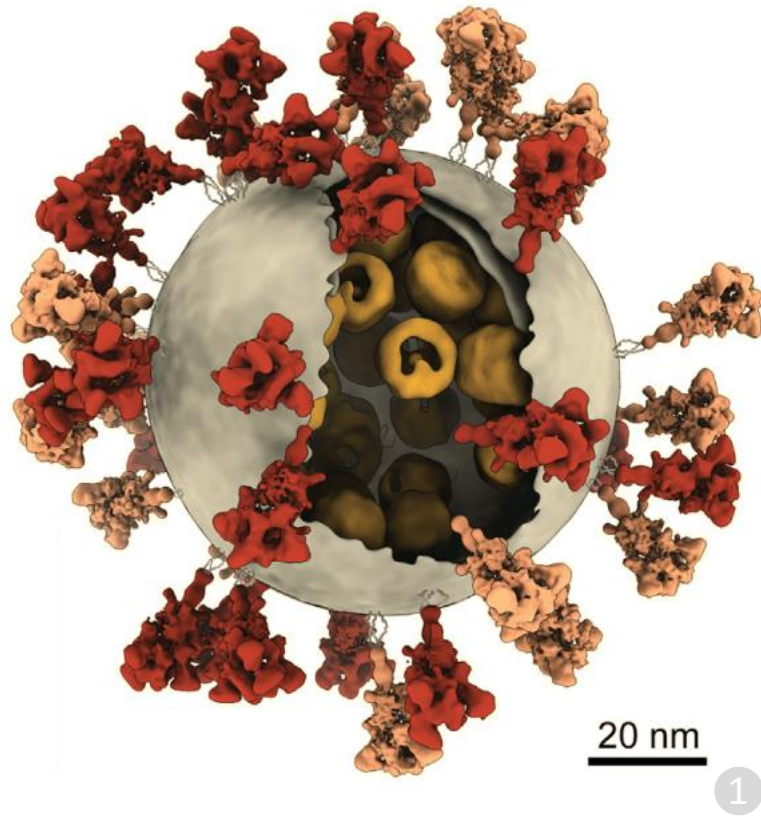
Understanding our basic physiology



Understanding our basic physiology

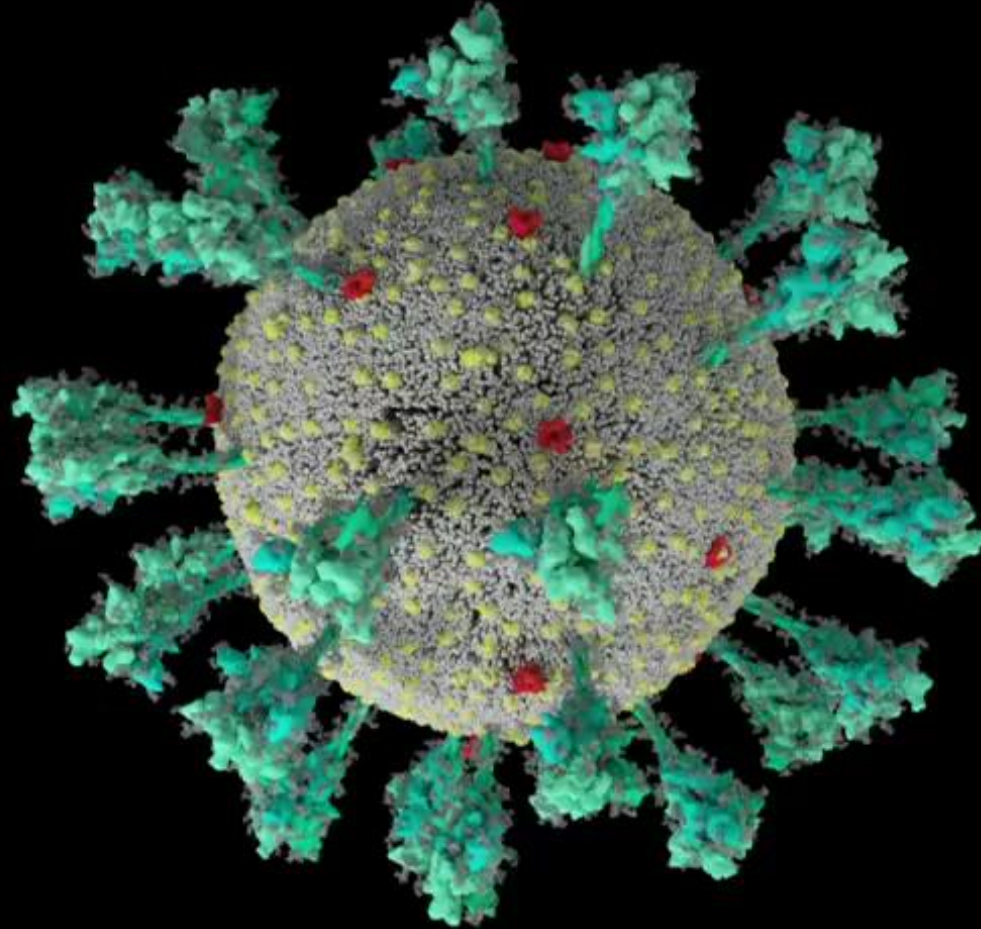


Understanding diseases (and how to treat them)

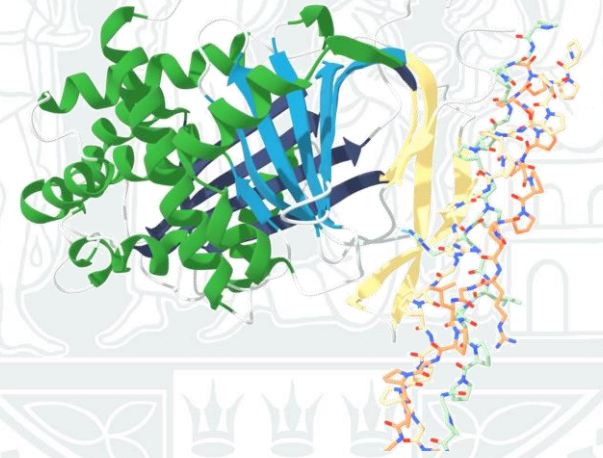
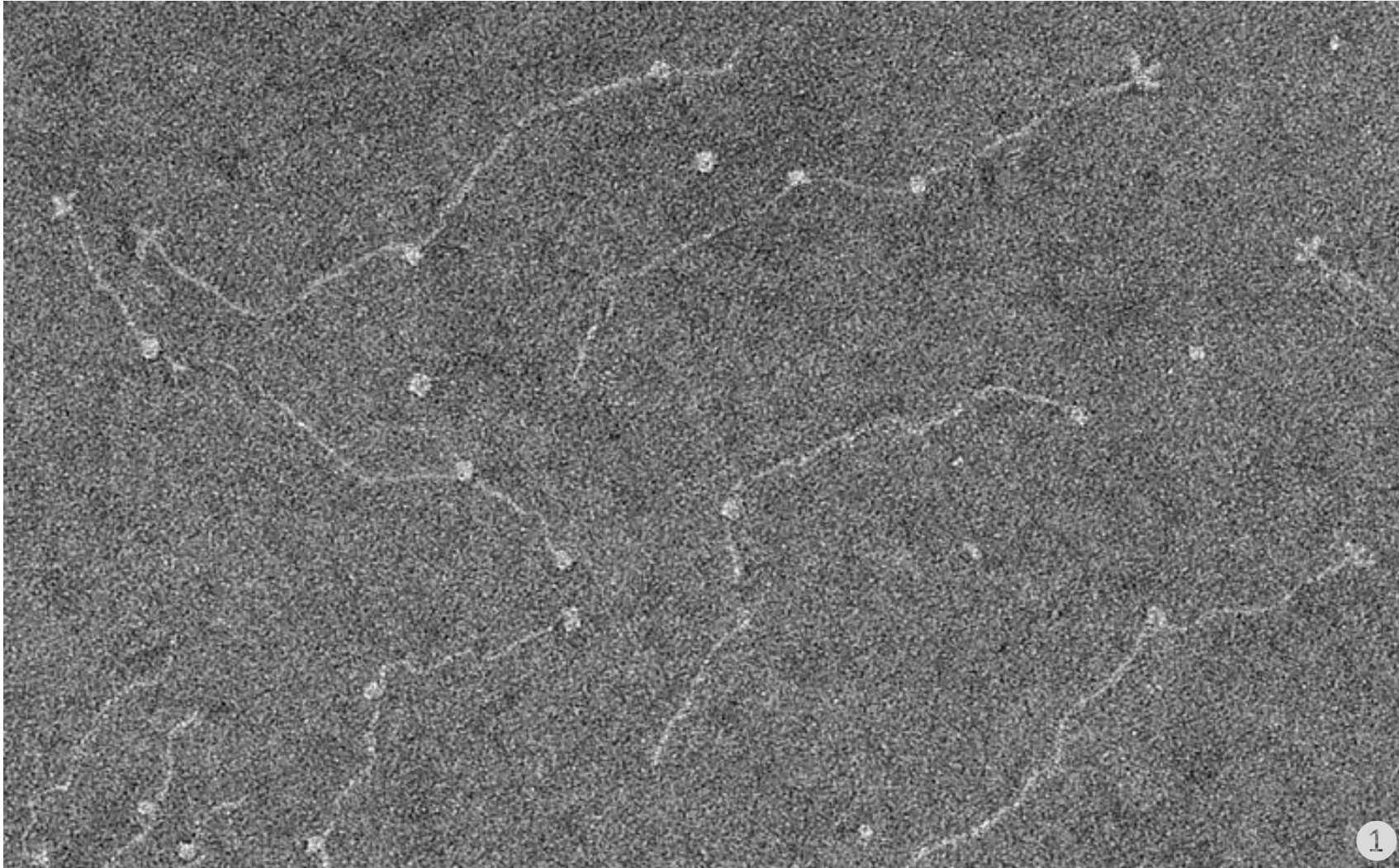


Understanding diseases (and how to treat them)

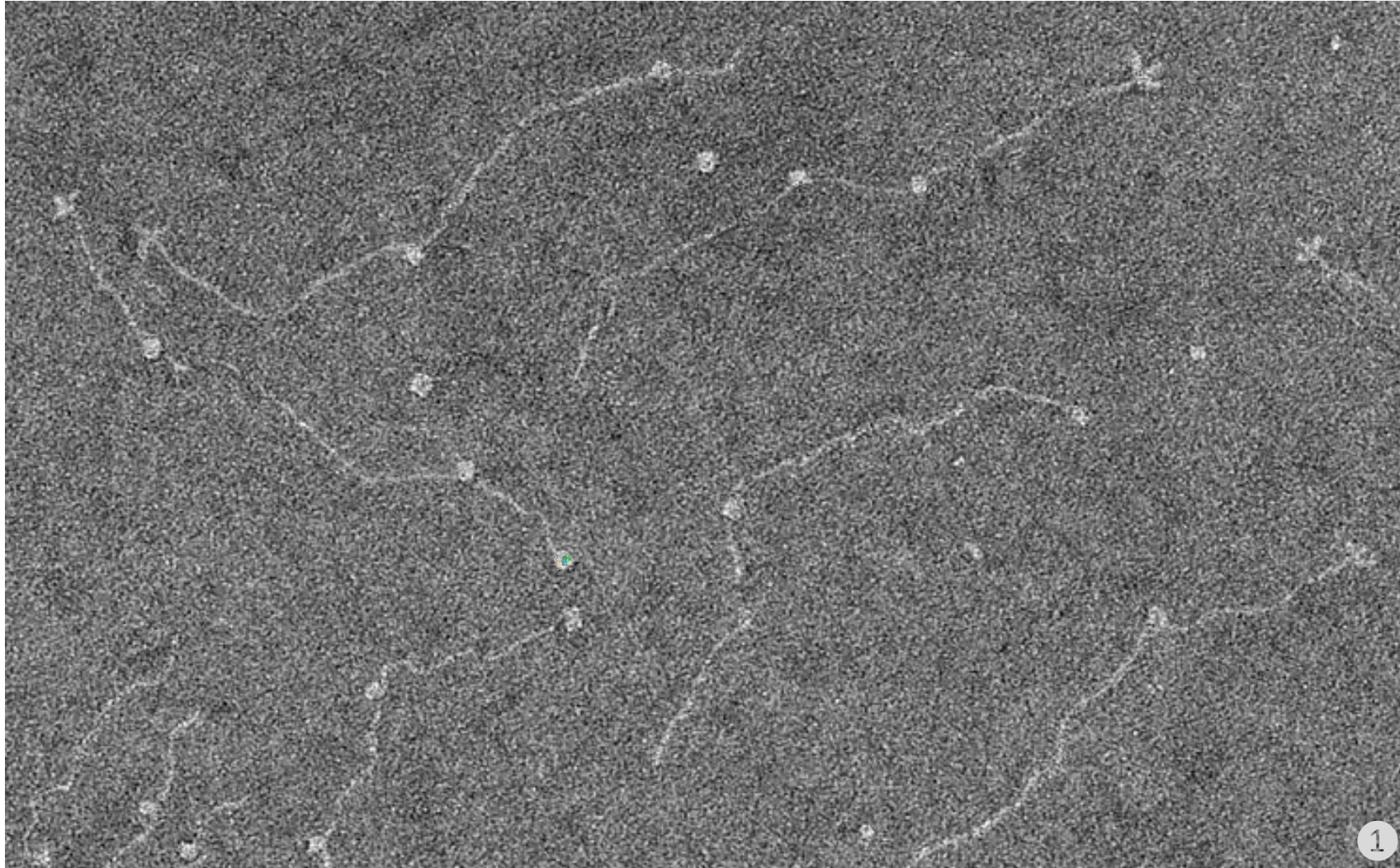
SARS-CoV-2



How do we determine these structures?

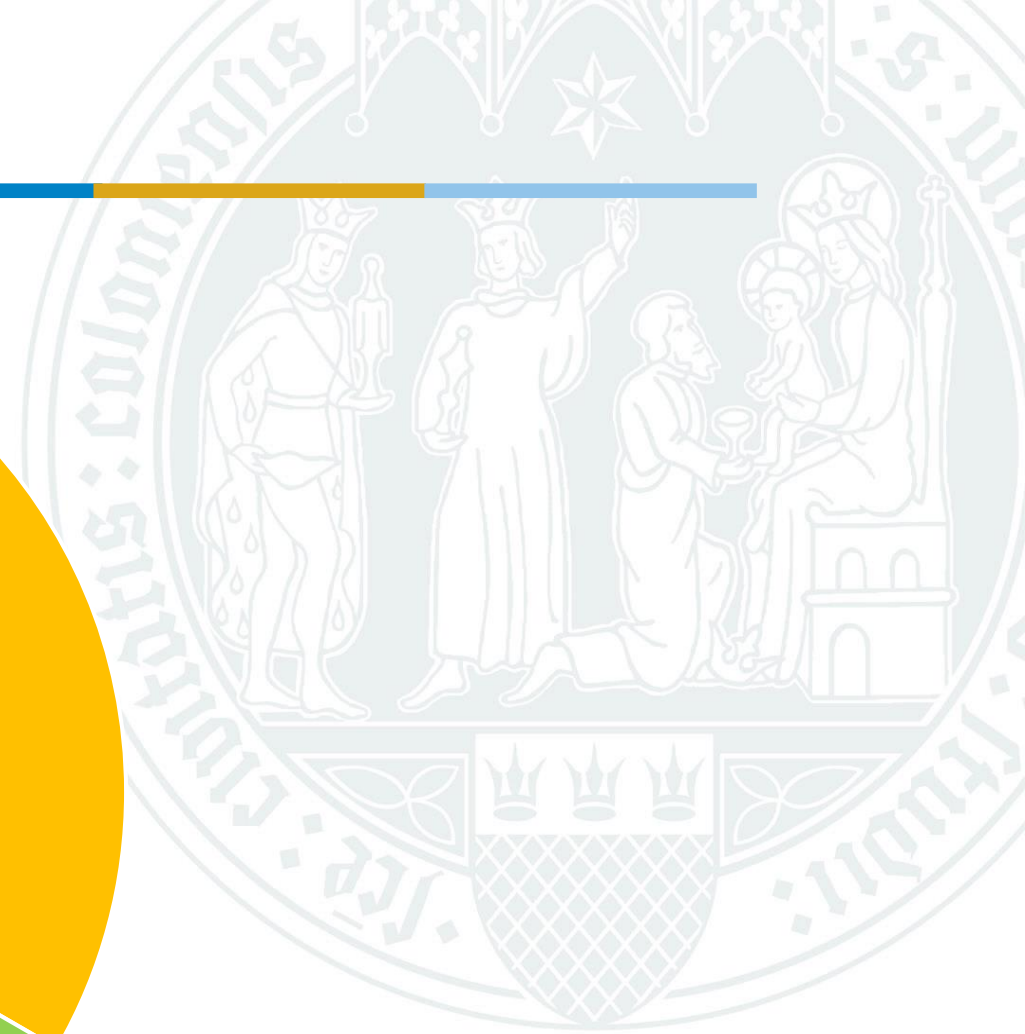
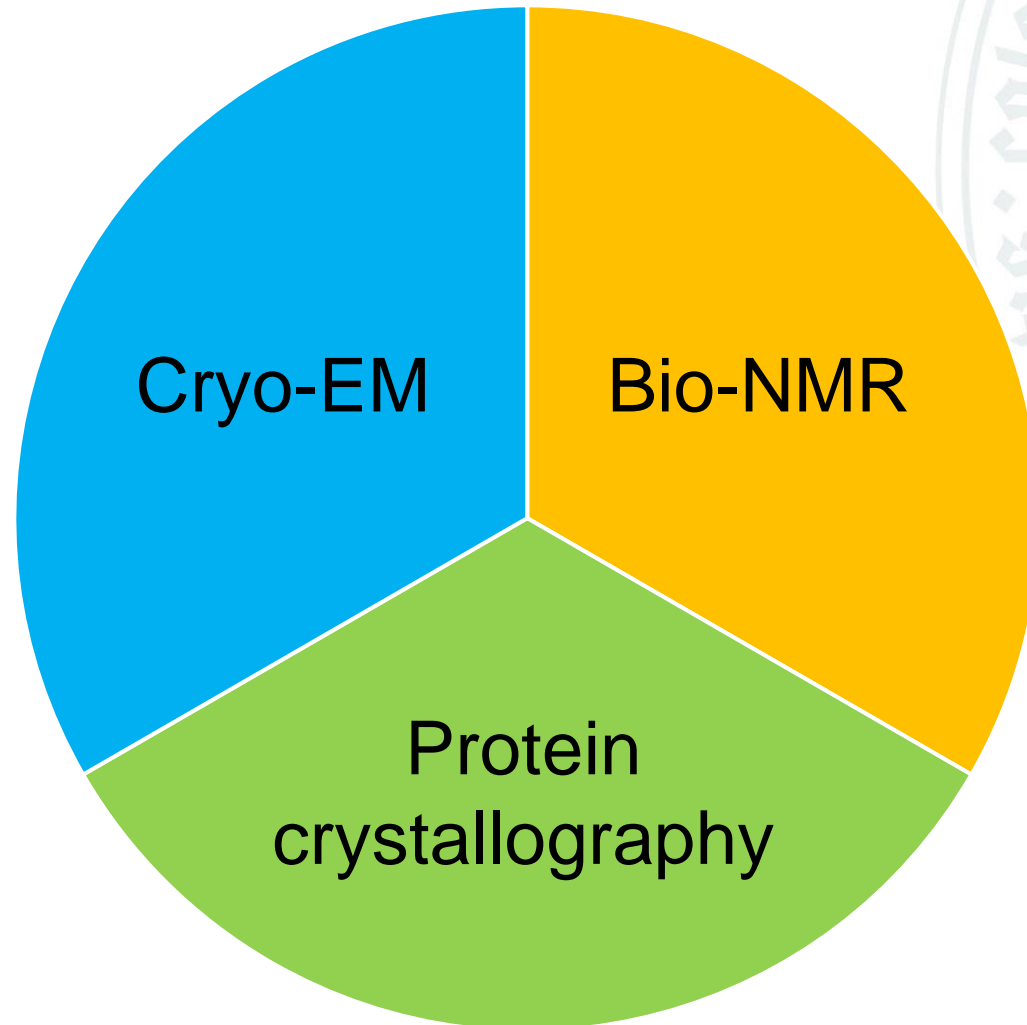


How do we determine these structures?

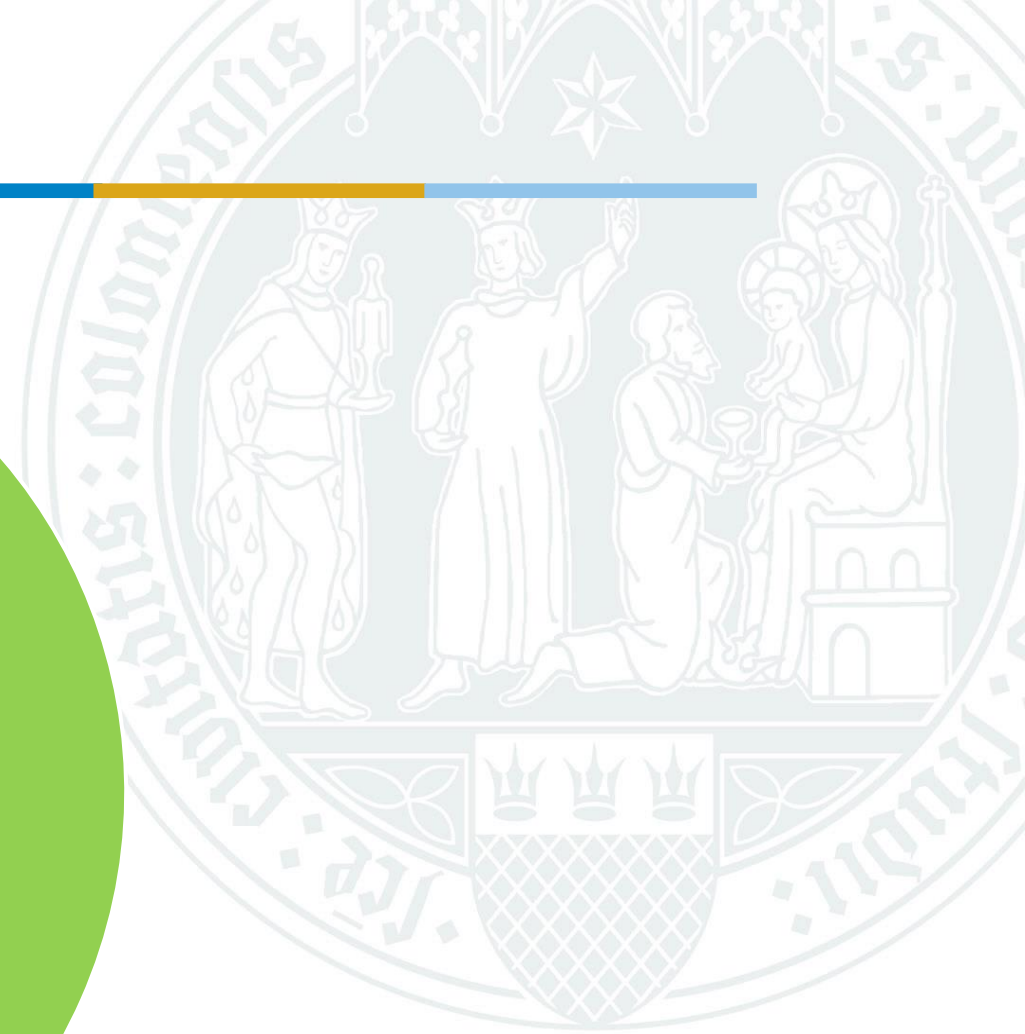
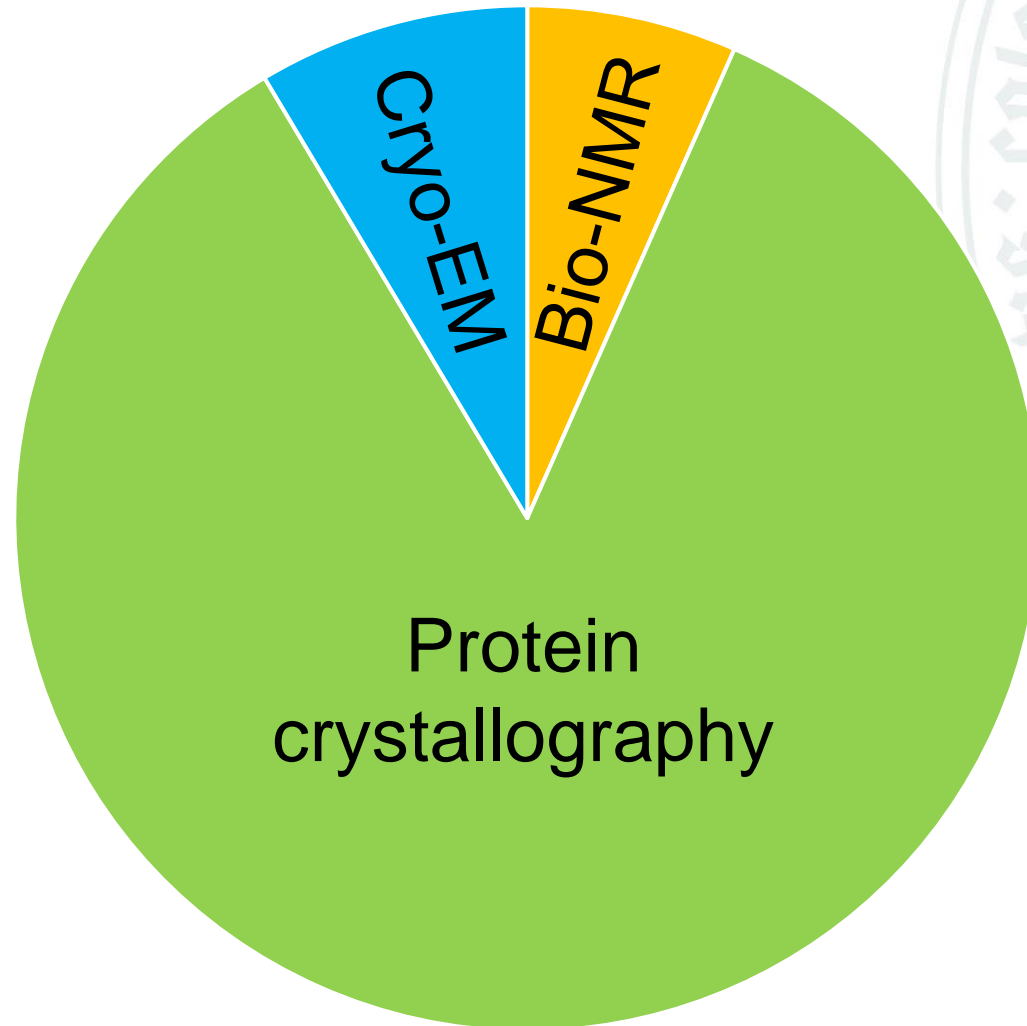


1

Experimental methods

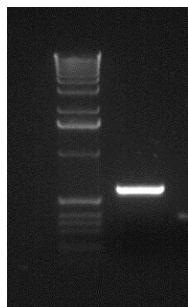


Experimental methods



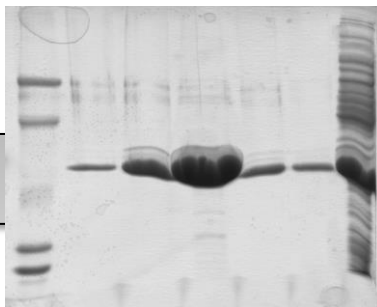
Protein crystallography

Target



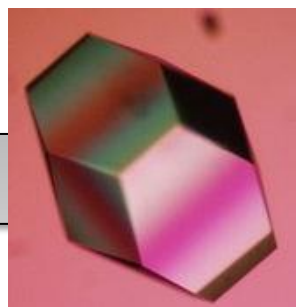
days

Purification



weeks

Crystallisation



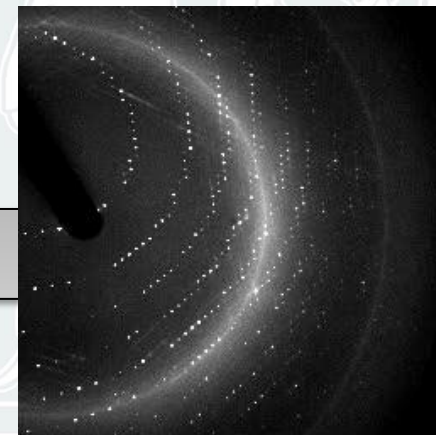
weeks-month

X-Rays

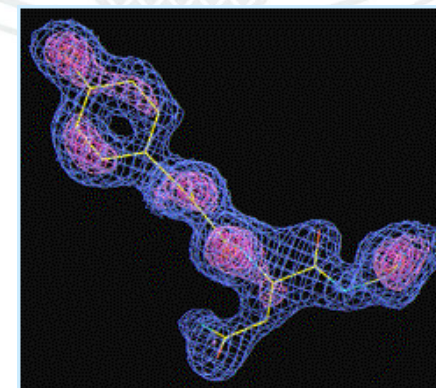


hours

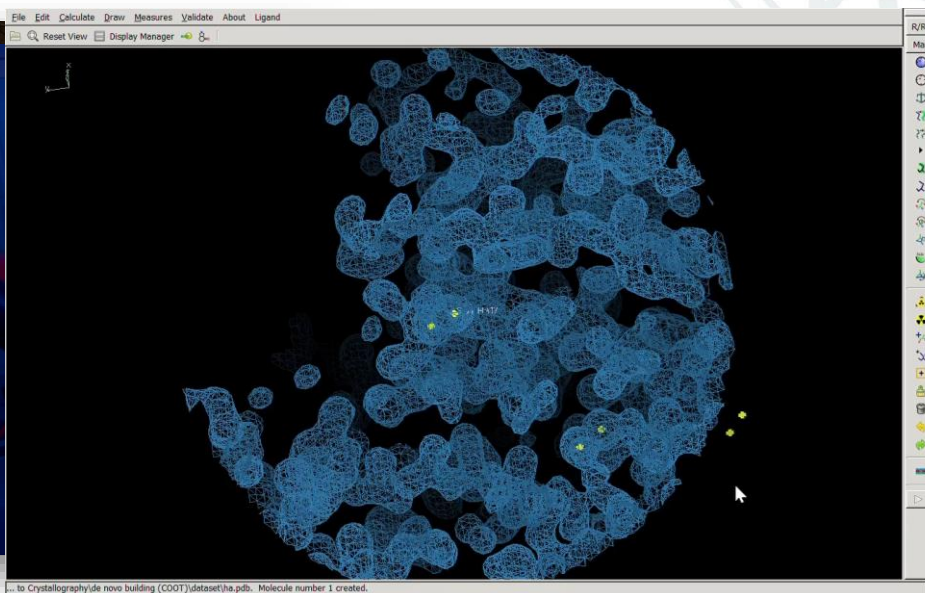
Data



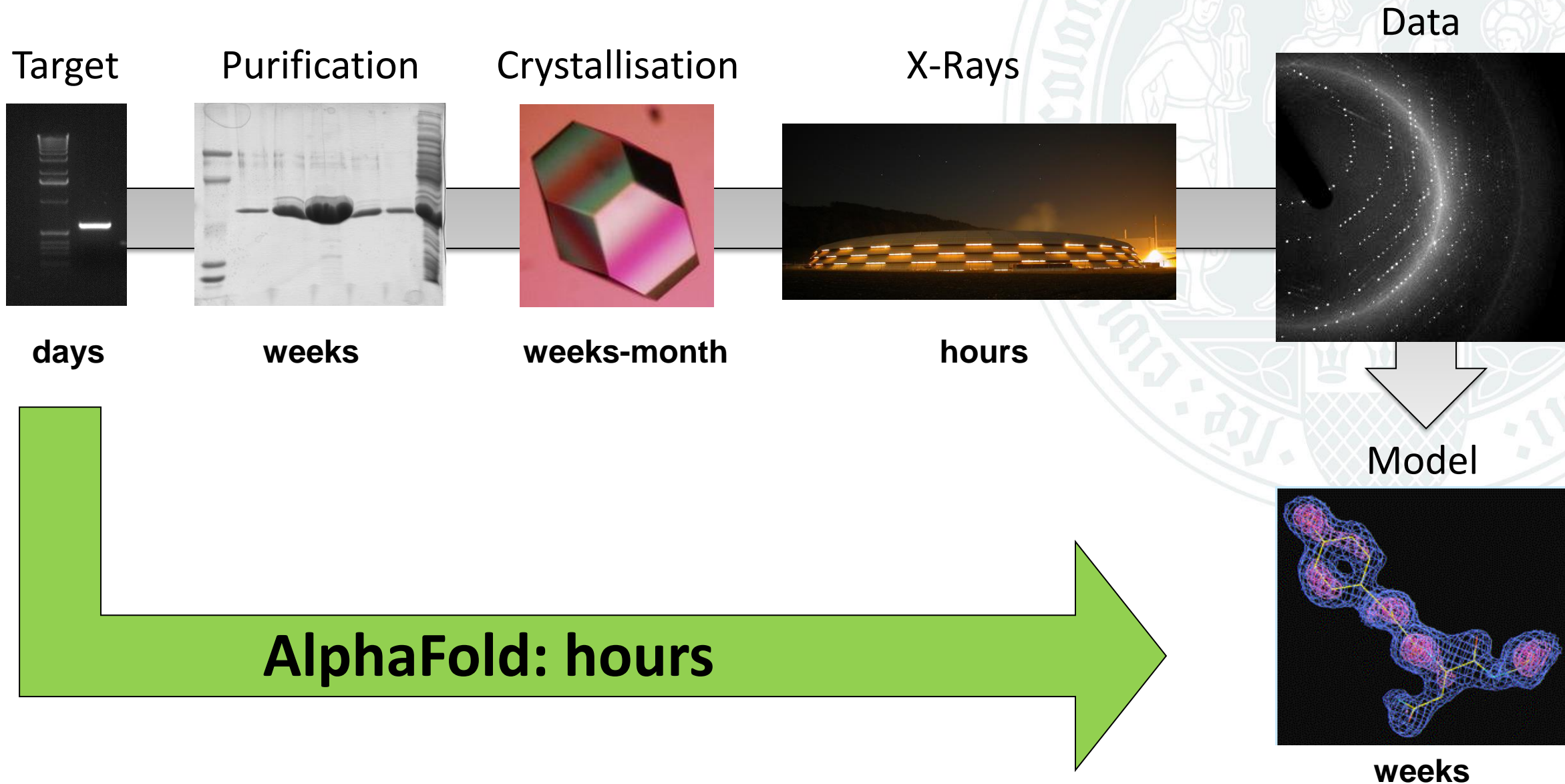
Model



weeks



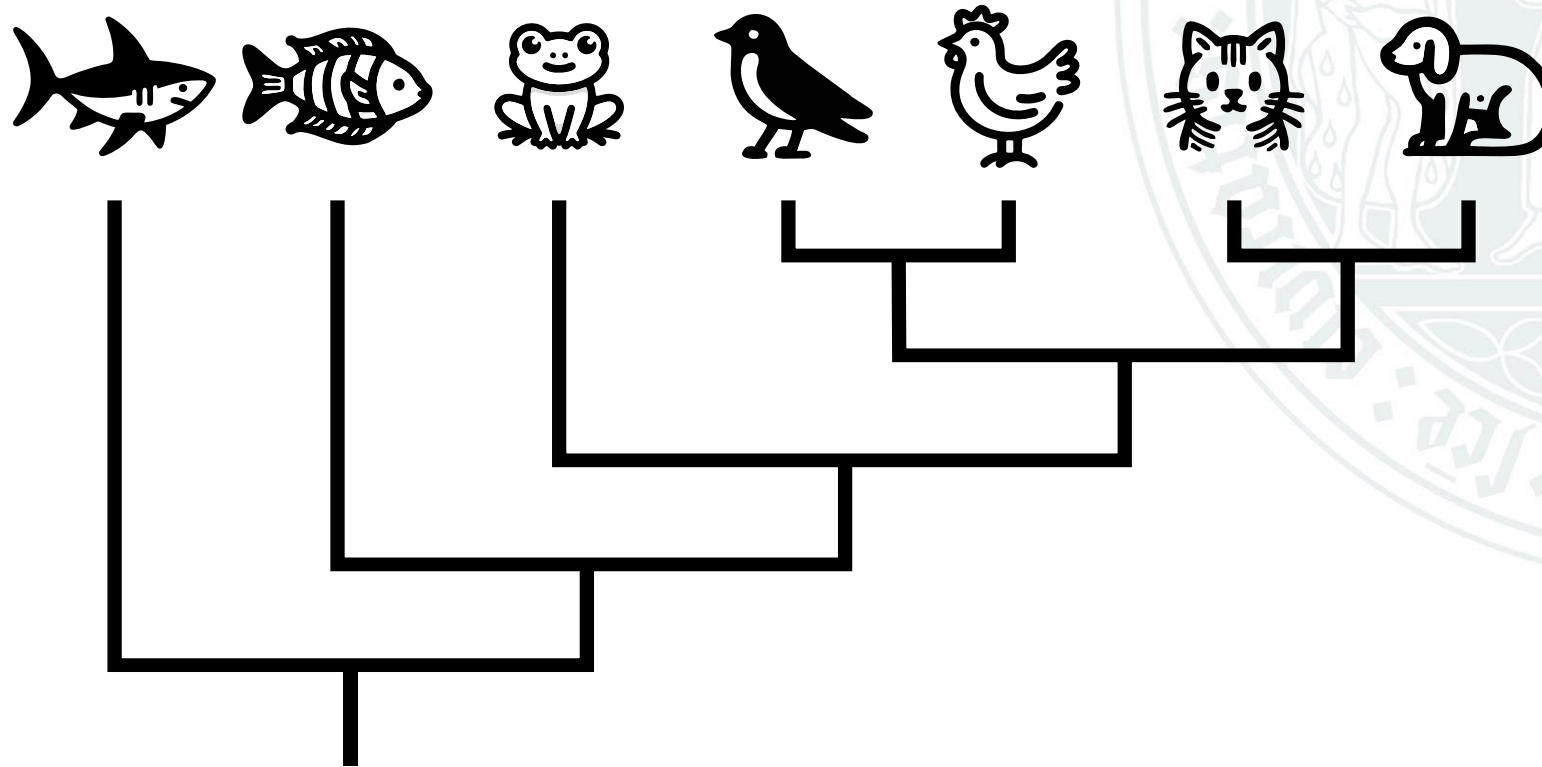
Protein crystallography



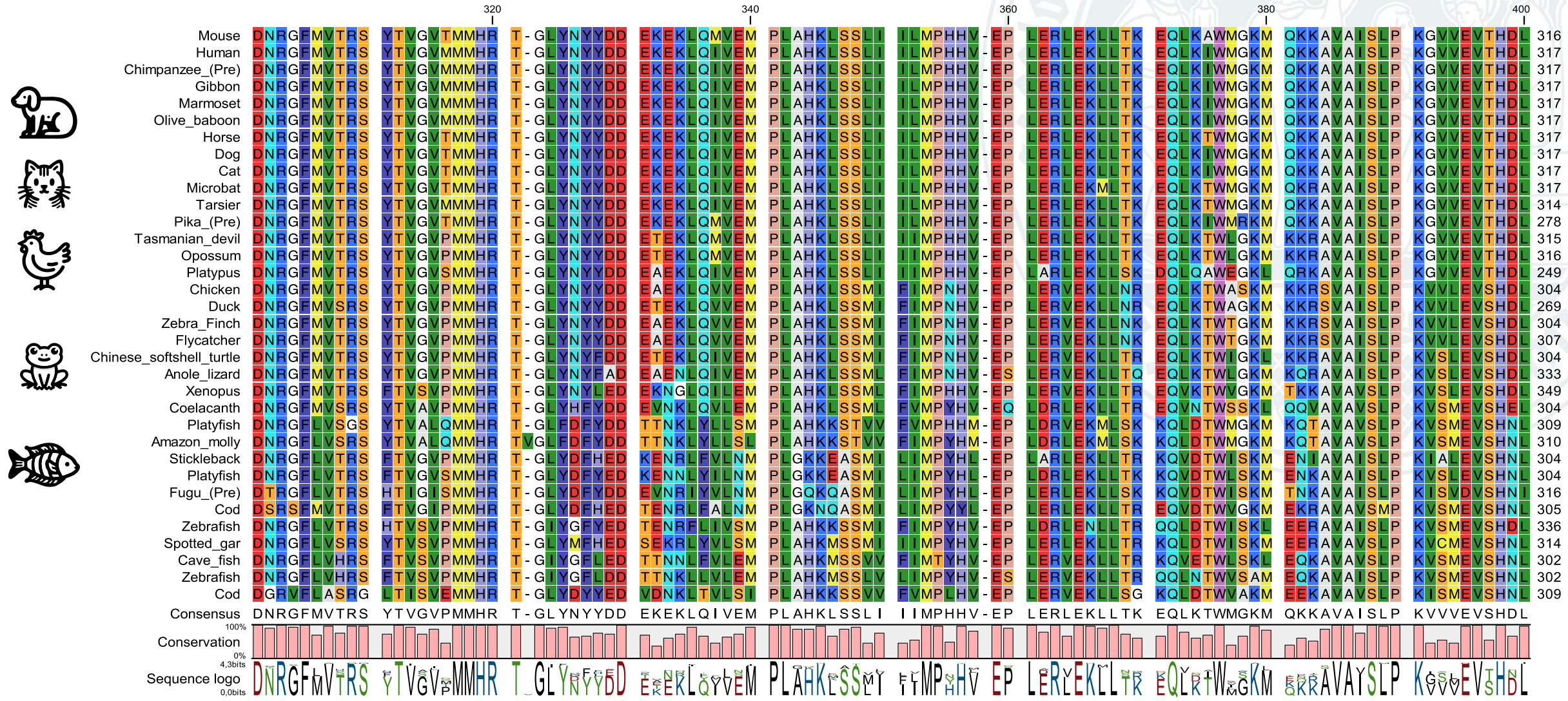
How does AlphaFold work?



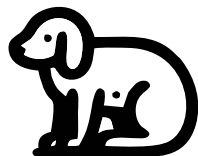
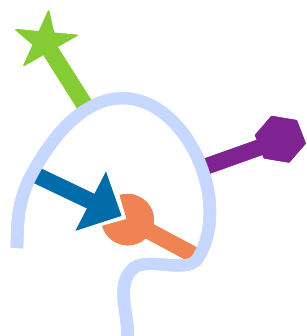
Protein sequences changed over evolution



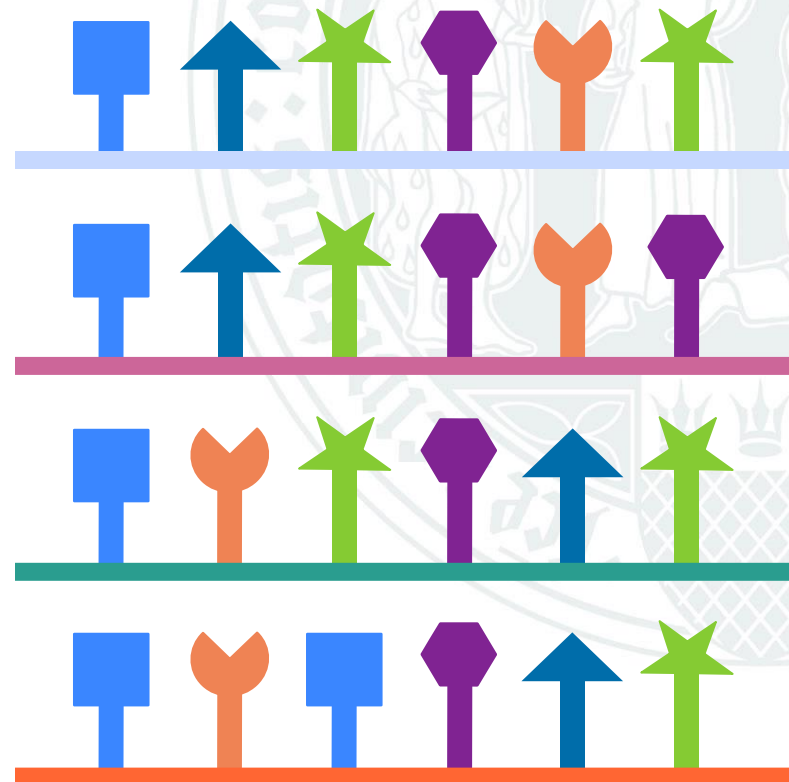
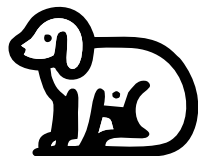
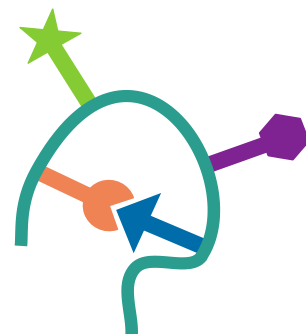
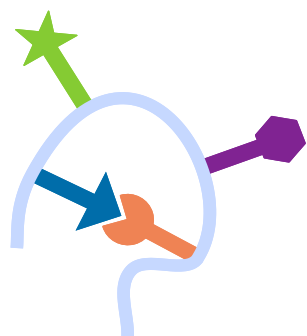
A typical „well-conserved“ protein



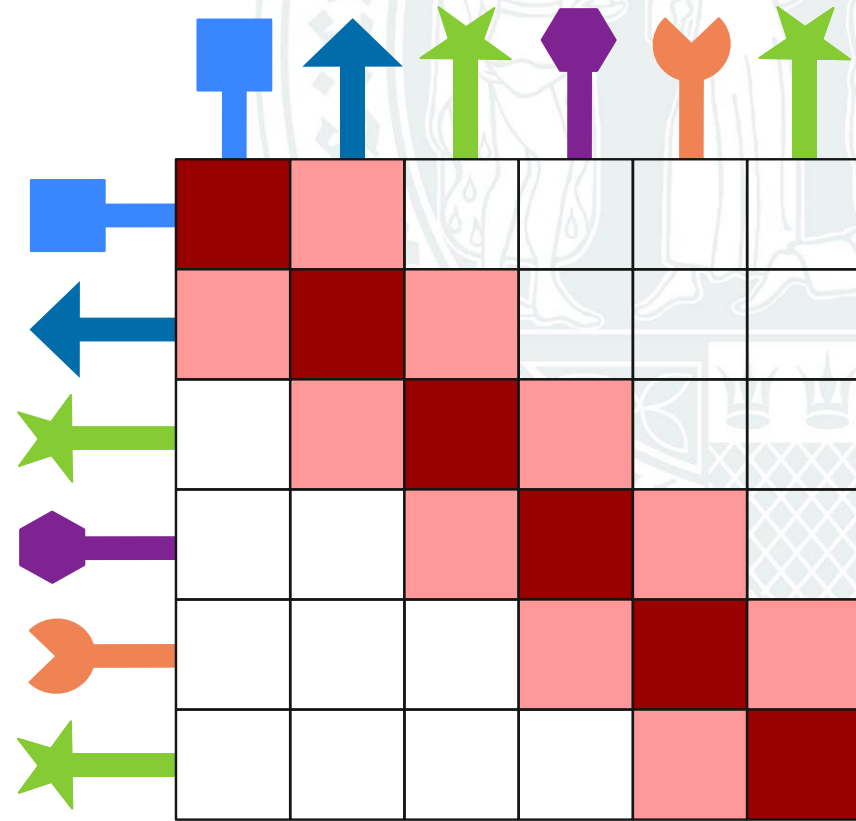
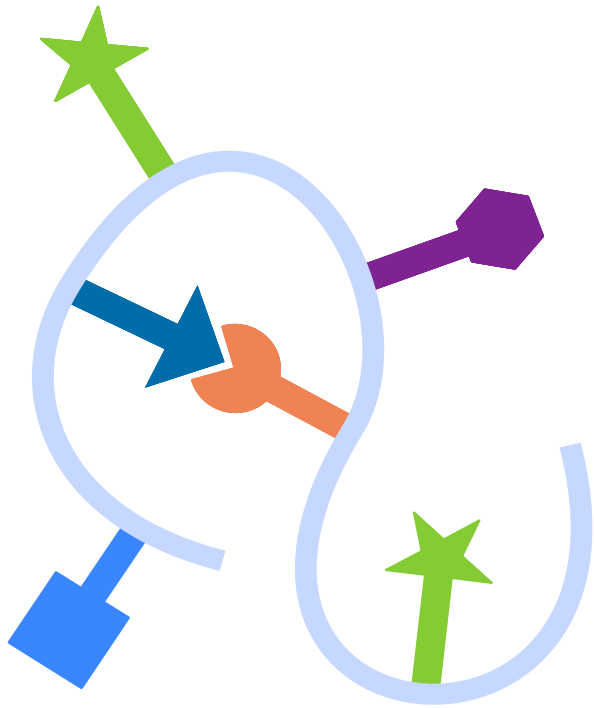
Important residue pairs should co-evolve



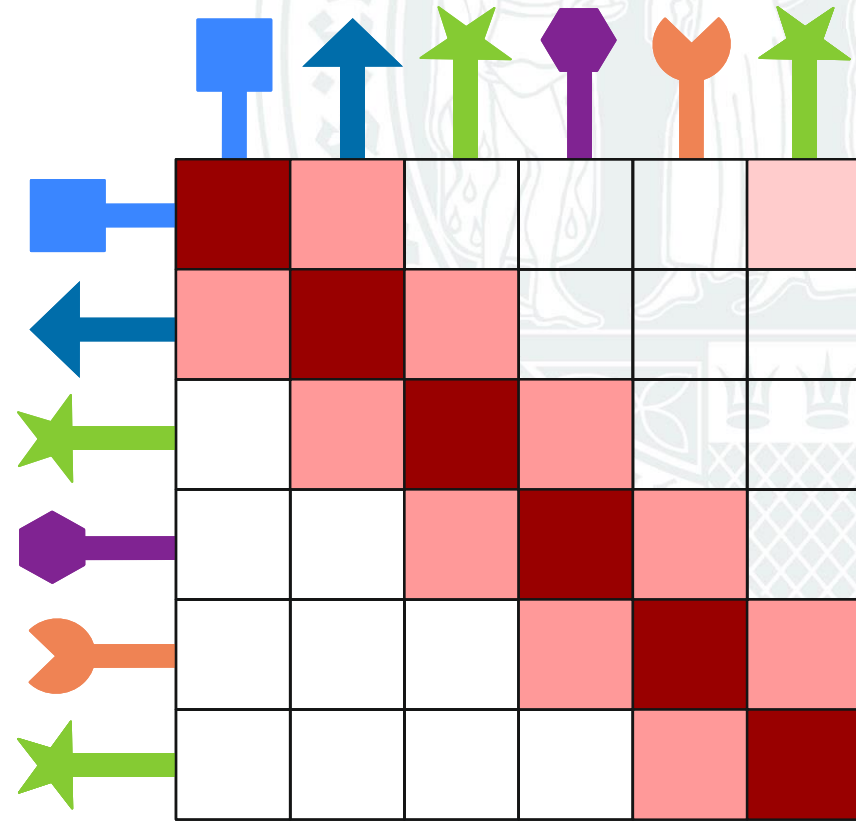
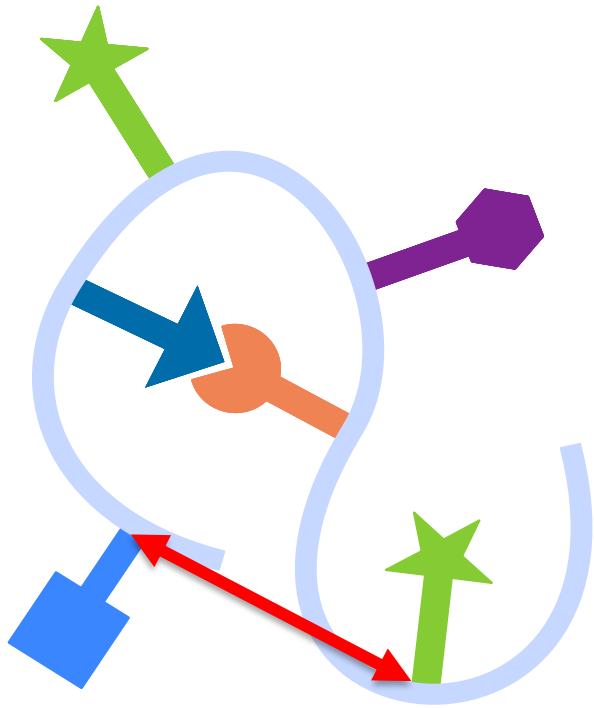
Important residue pairs should co-evolve



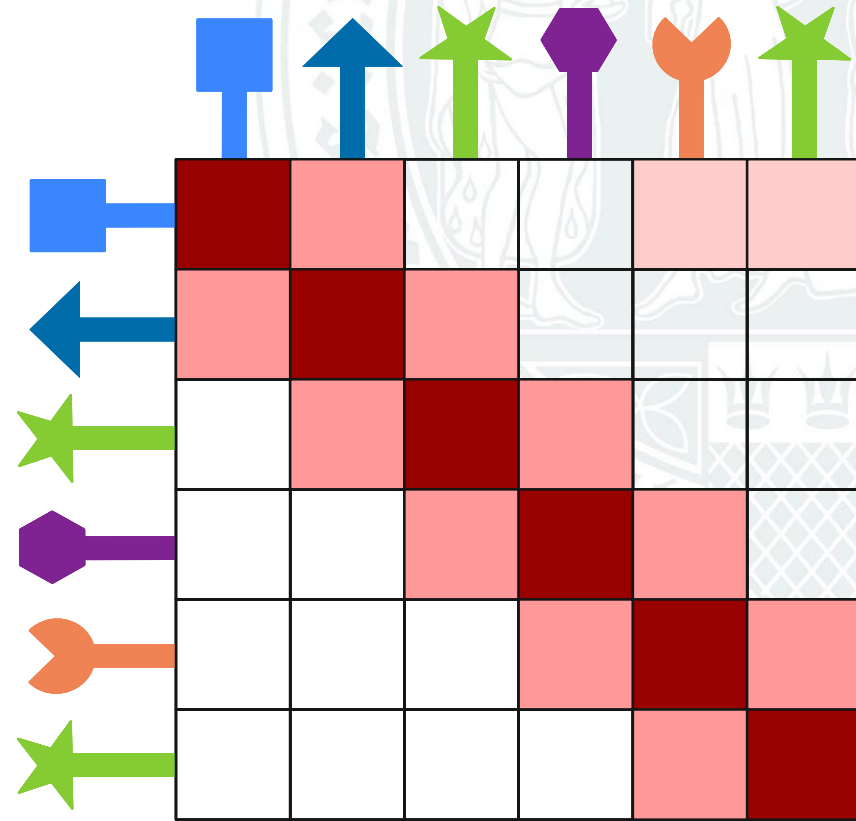
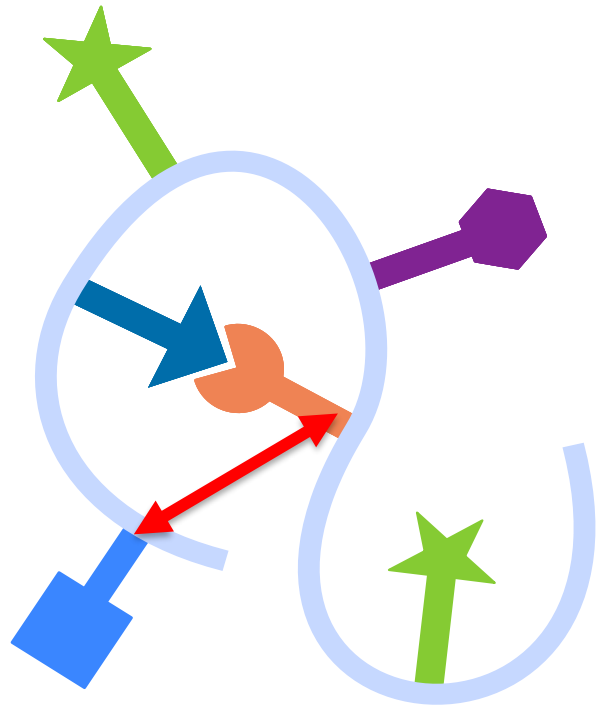
A 2D representation of a 3D world



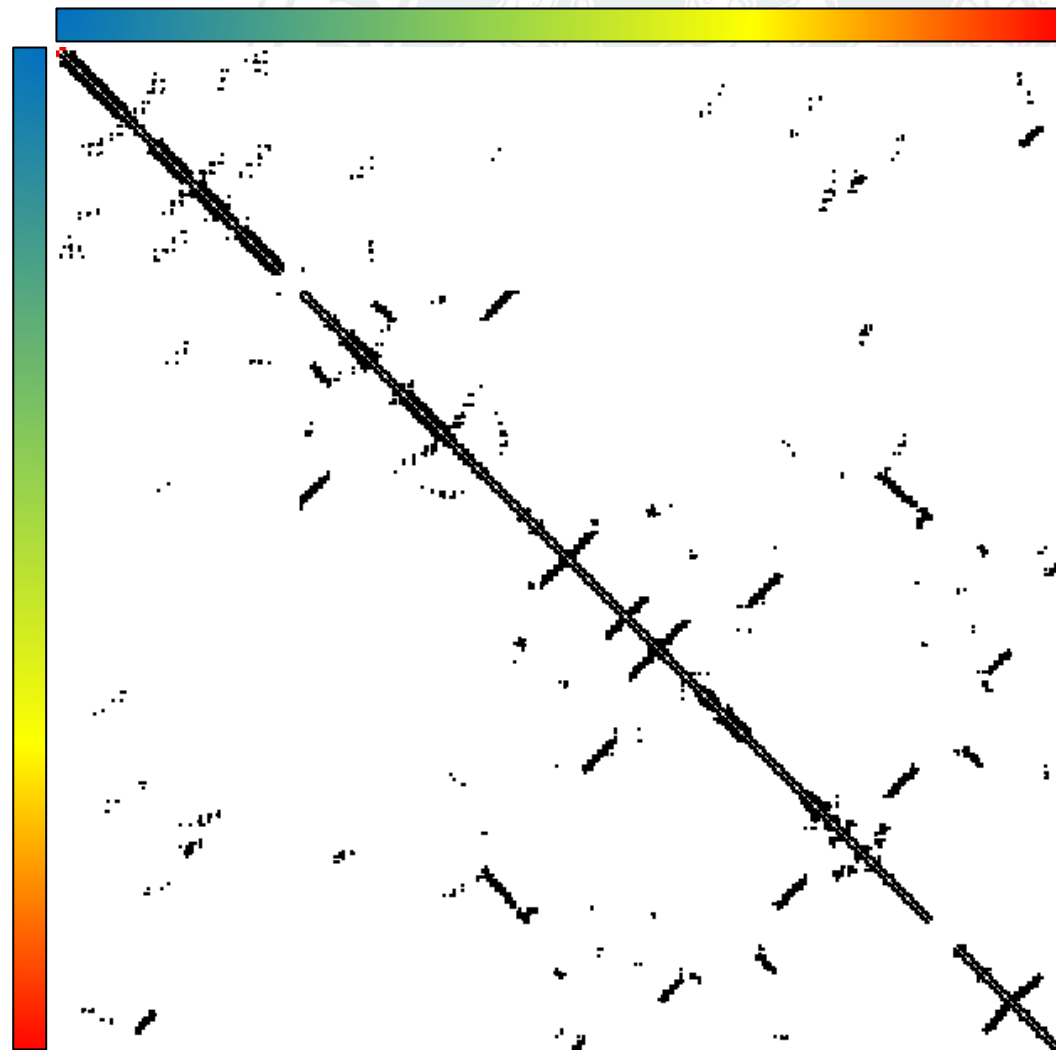
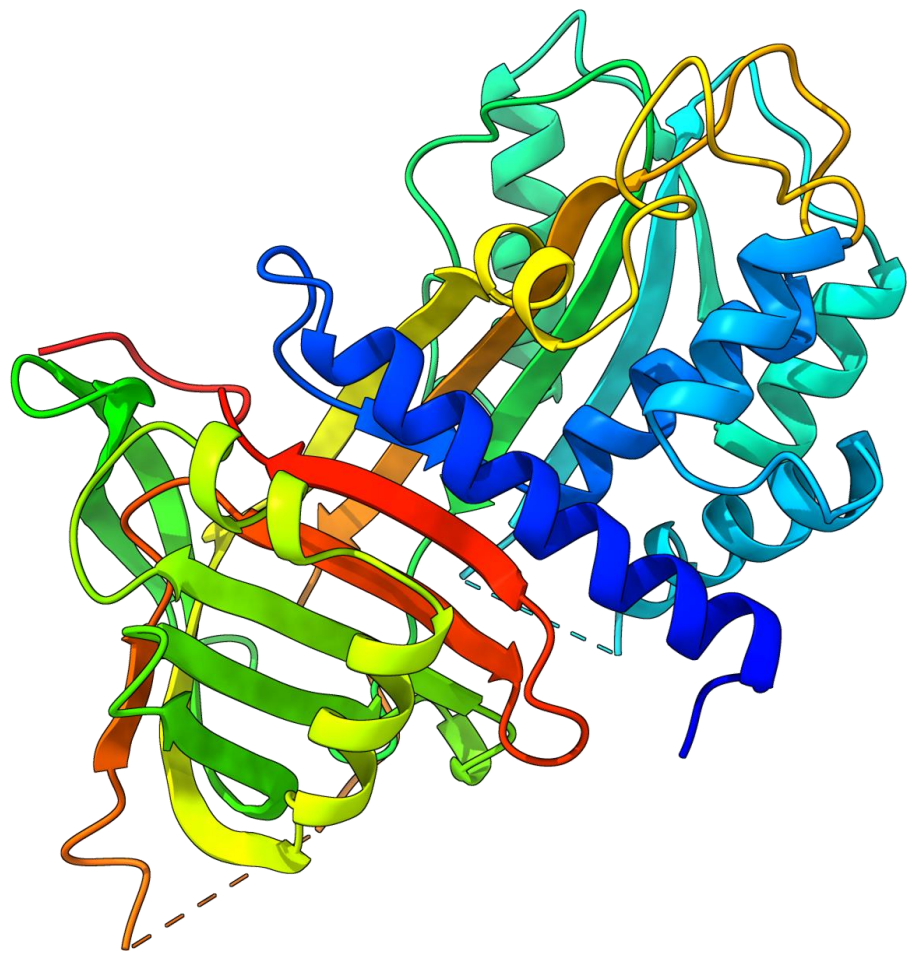
A 2D representation of a 3D world



A 2D representation of a 3D world



A real map



The data source

RCSB PDB PROTEIN DATA BANK

213,818 Structures from the PDB

1,068,577 Computed Structure Models (CSM)

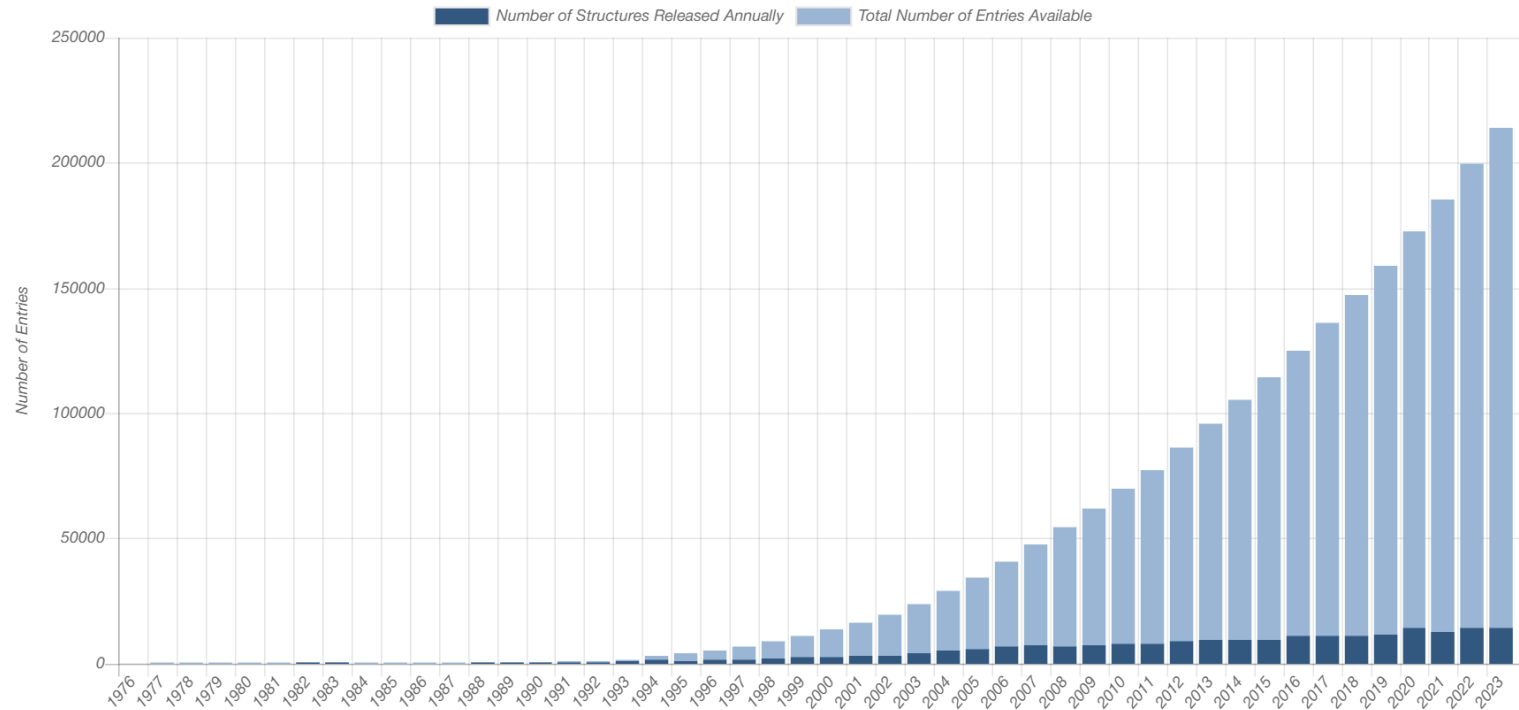
3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM

Advanced Search | Browse Annotations Help

PDB-101 PDB EMDataResource NAKB wwPDB Foundation PDB-Dev

f t y g

PDB Statistics: Overall Growth of Released Structures Per Year






PROTEIN DATA BANK



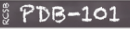






Approx.
215'000
experimentally
determined
structures

One entry ...


RCSB **PDB** PROTEIN DATA BANK  213,818 Structures from the PDB
 1,068,577 Computed Structure Models (CSM)

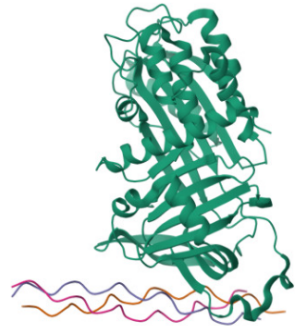
3D Structures Include CSM 




[Advanced Search](#) | [Browse Annotations](#) [Help](#)

[Structure Summary](#) [Structure](#) [Annotations](#) [Experiment](#) [Sequence](#) [Genome](#) [Versions](#)

Biological Assembly 3 



7BDU


Crystal structure of a Hsp47-collagen peptide complex

PDB DOI: <https://doi.org/10.2210/pdb7BDU/pdb>

Classification: **CHAPERONE**

Organism(s): *Canis lupus familiaris*, *Homo sapiens*

Expression System: *Escherichia coli*

Mutation(s): No 

Deposited: 2020-12-22 Released: 2021-11-10

Deposition Author(s): Abraham, E.T., Gebauer, J.M., Baumann, U.

Funding Organization(s): German Research Foundation (DFG)

Experimental Data Snapshot


Method: X-RAY DIFFRACTION






Resolution: 2.49 Å

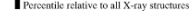

R-Value Free: 0.267


R-Value Work: 0.217


R-Value Observed: 0.220

wwPDB Validation  [3D Report](#) [Full Report](#)

Metric	Percentile Ranks	Value
Rfree		0.267
Clashscore		4
Ramachandran outliers		0.1%
Sidechain outliers		2.6%
RSRZ outliers		3.3%

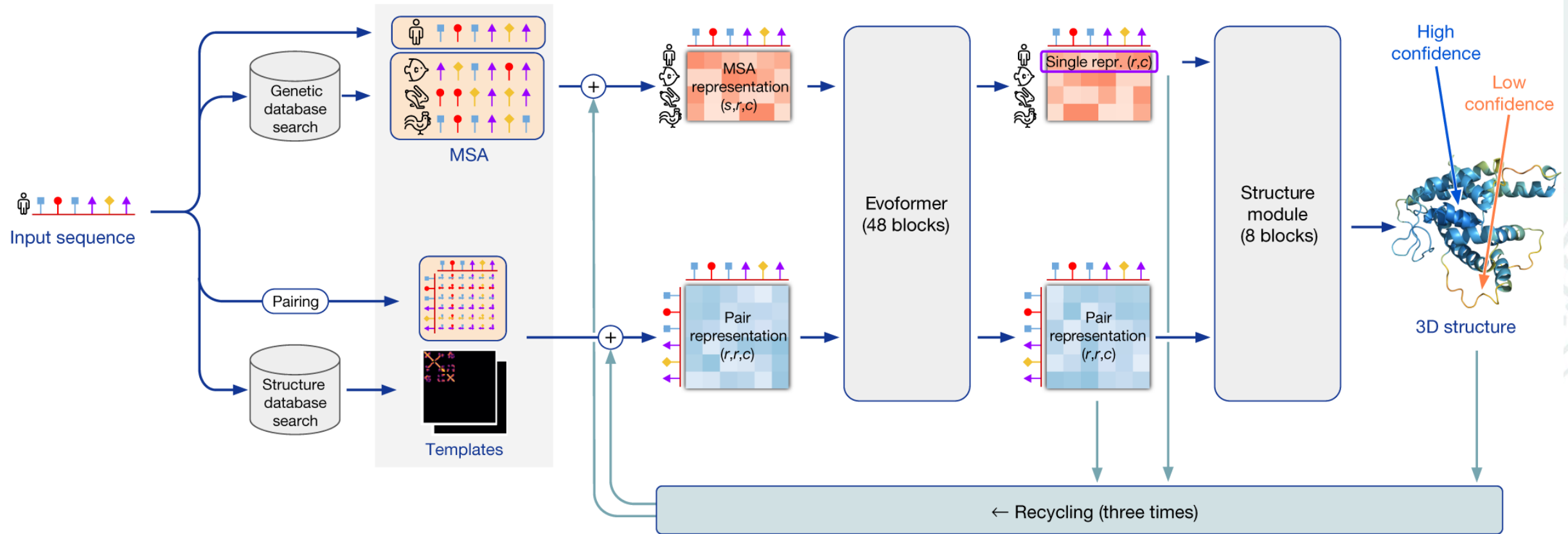
Worse   Better

Local Symmetry: Helical - H  (Explore in 3D)

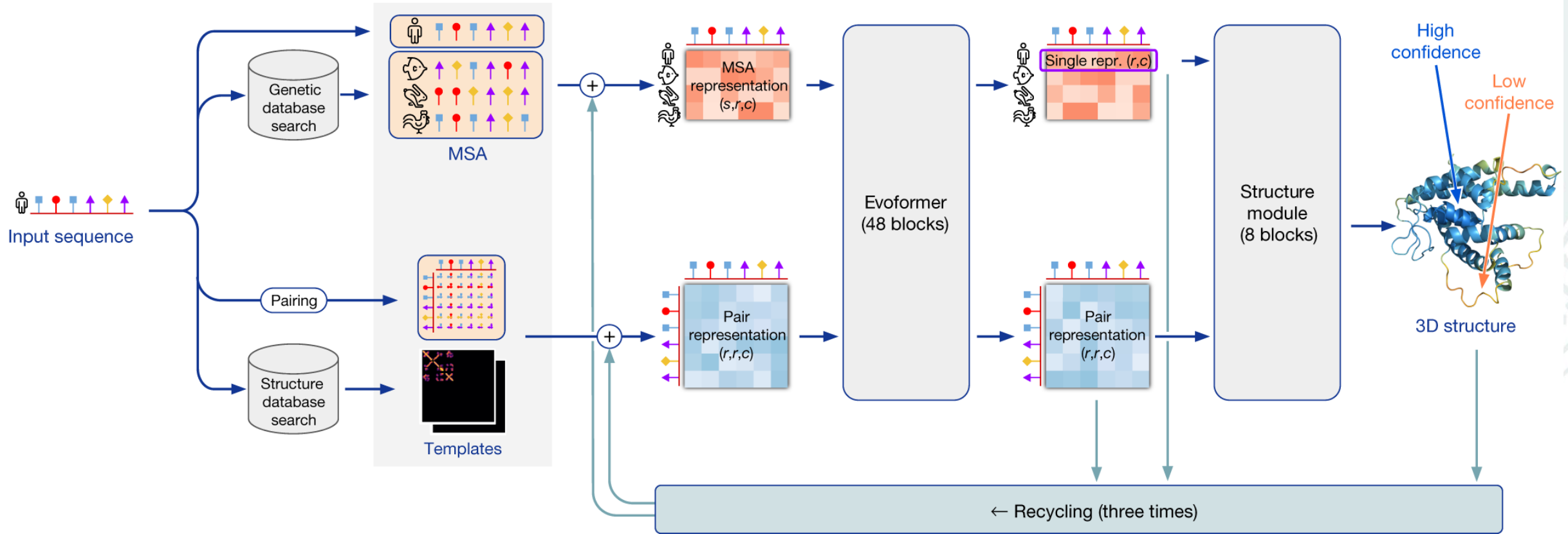
Local Stoichiometry: Homo 3-mer - A3 

This is version 1.1 of the entry. See complete [history](#).

The neuronal network

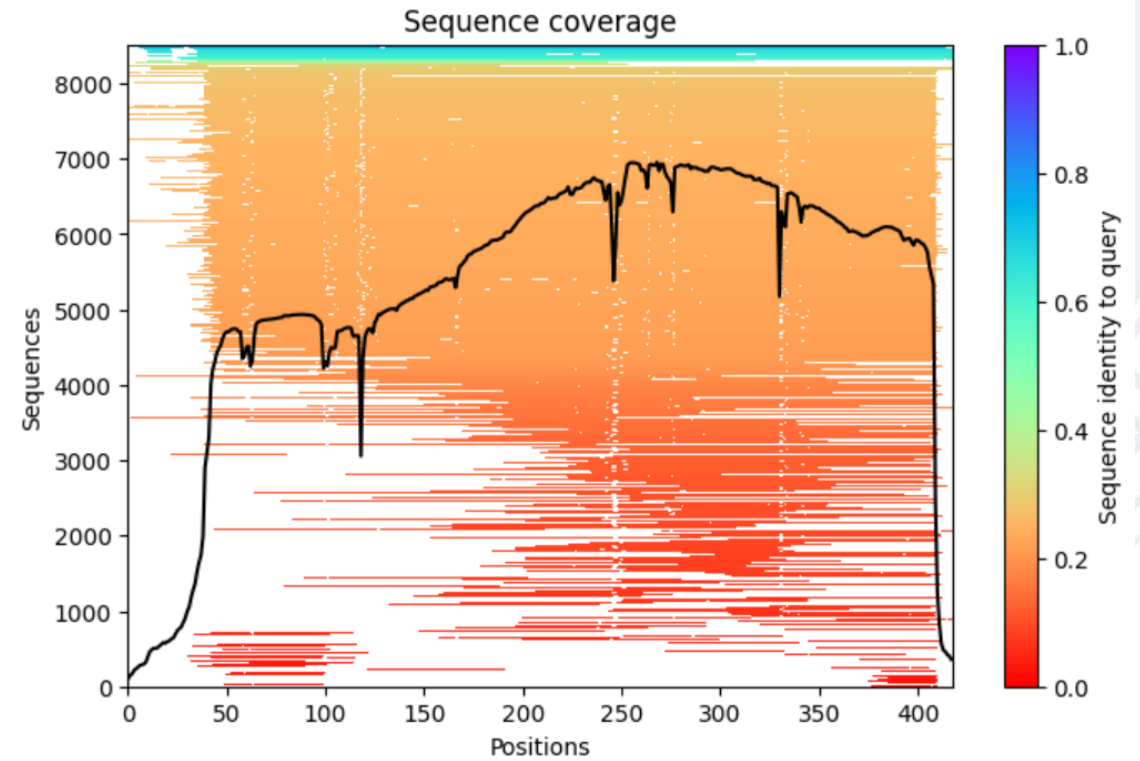
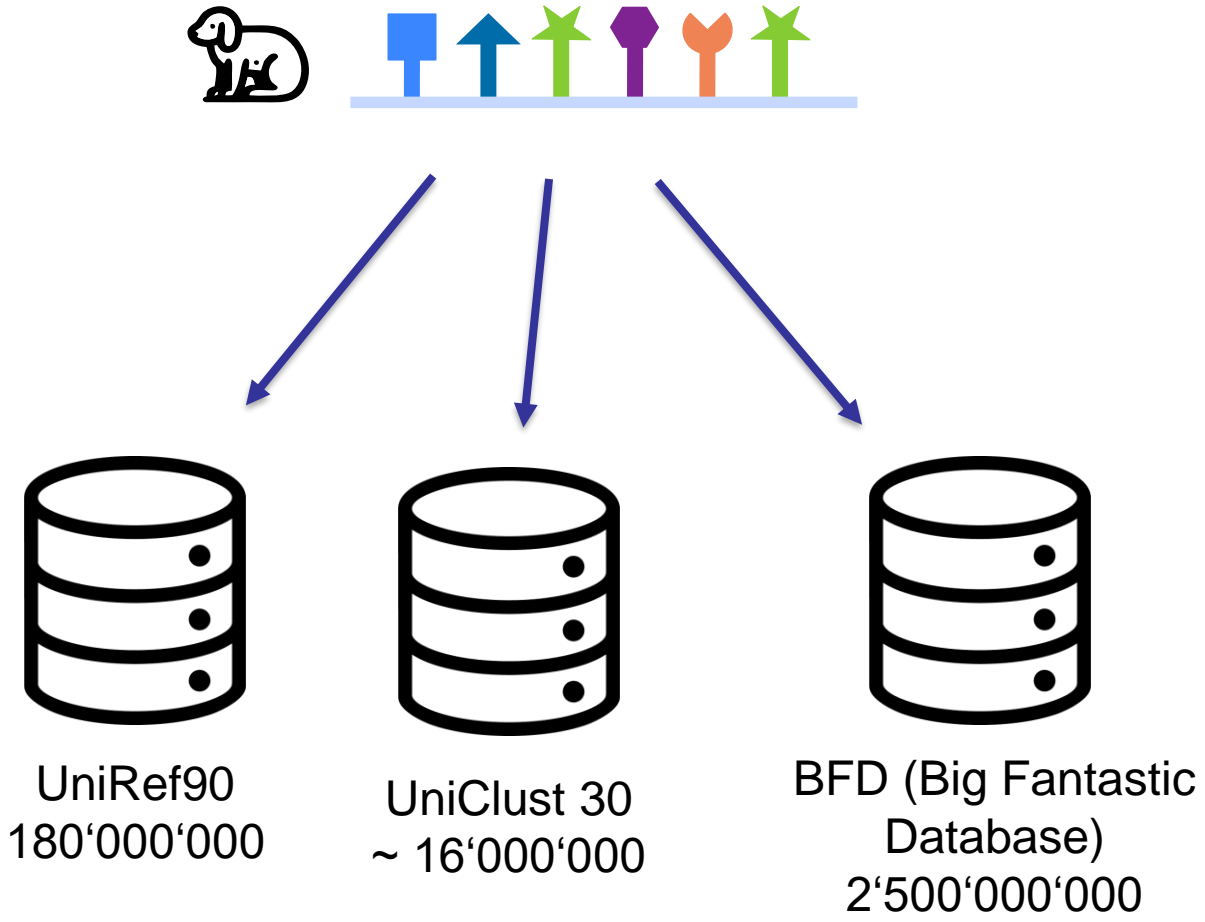


The neuronal network



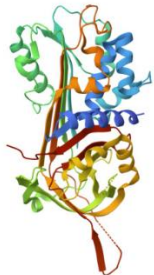
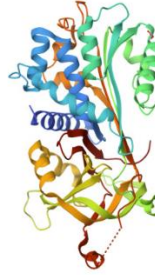




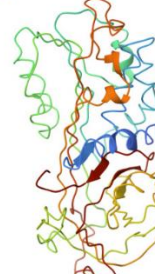


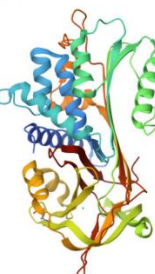
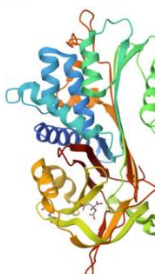

Data gathering

Gathering the data...

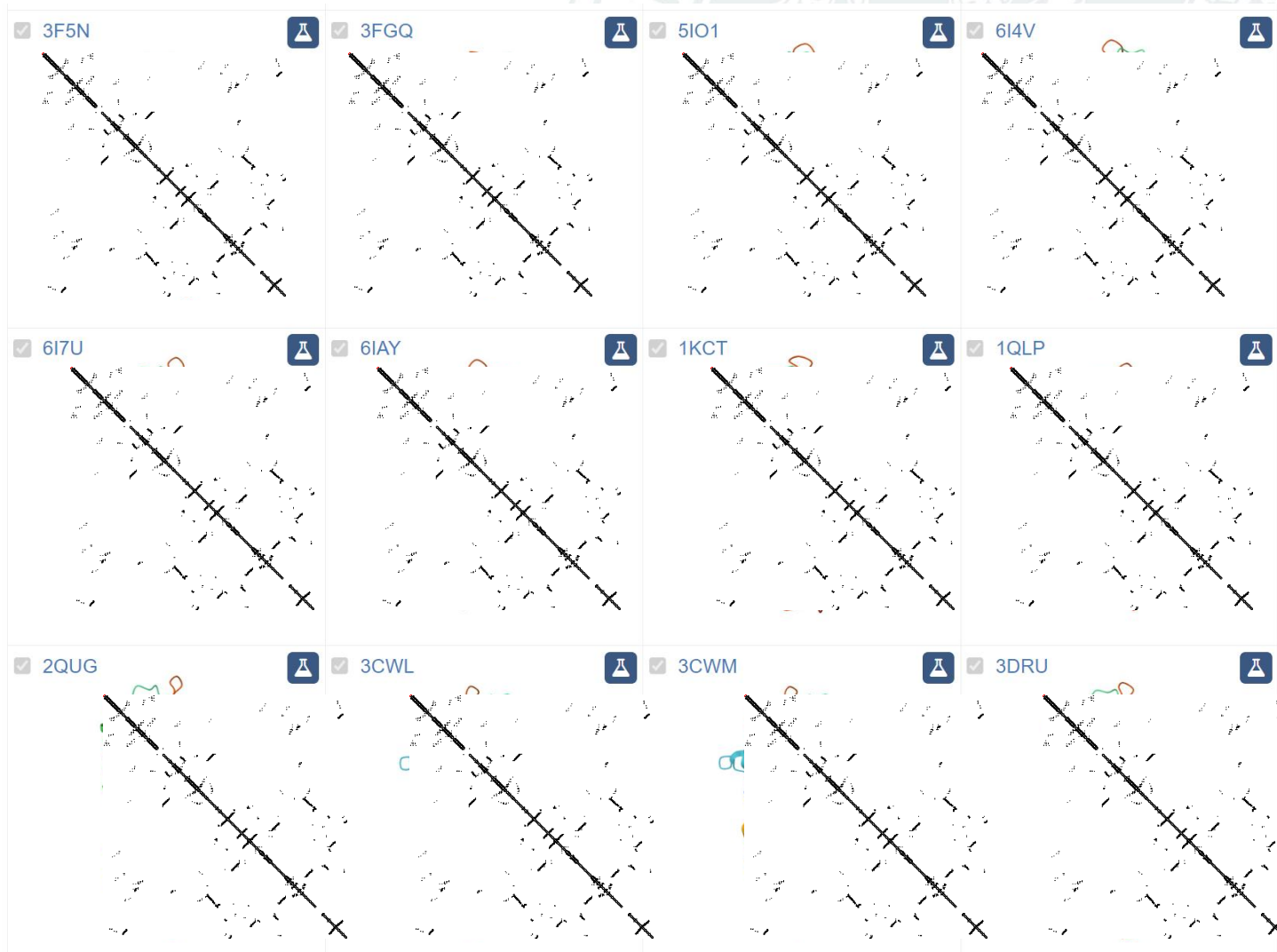


Searching for templates

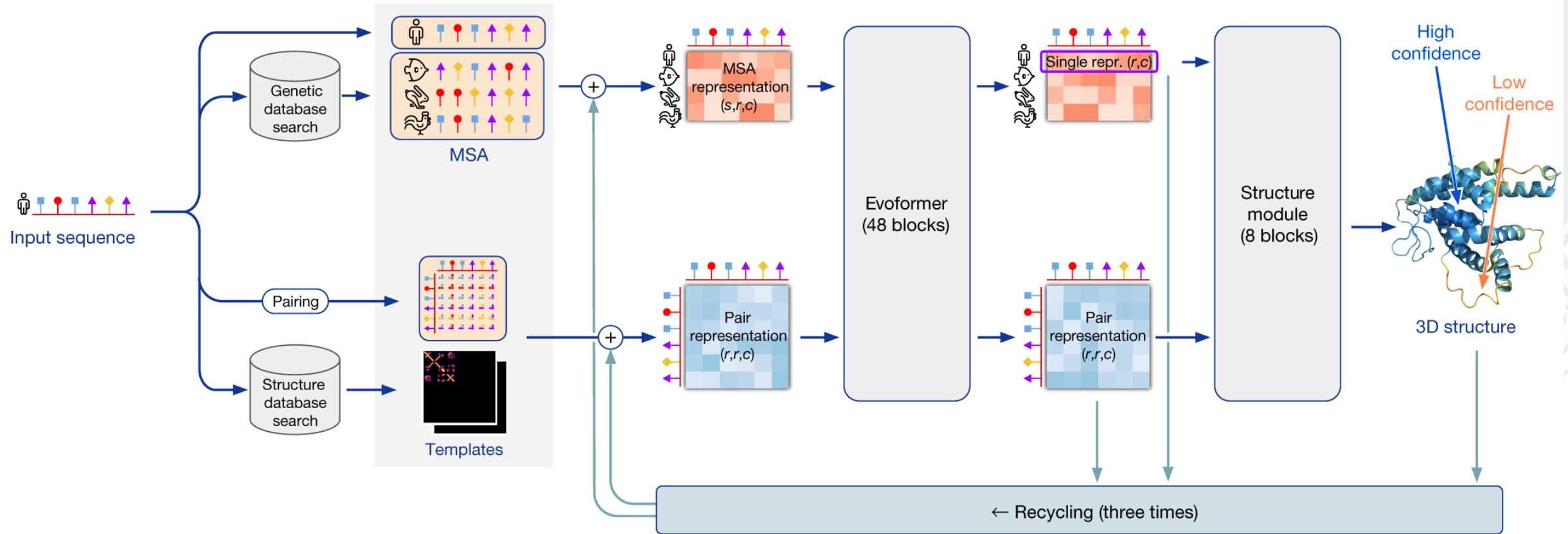


<input checked="" type="checkbox"/> 3F5N 	<input checked="" type="checkbox"/> 3FGQ 	<input checked="" type="checkbox"/> 5IO1 	<input checked="" type="checkbox"/> 6I4V 
<input checked="" type="checkbox"/> 6I7U 	<input checked="" type="checkbox"/> 6IAY 	<input checked="" type="checkbox"/> 1KCT 	<input checked="" type="checkbox"/> 1QLP 
<input checked="" type="checkbox"/> 2QUG 	<input checked="" type="checkbox"/> 3CWL 	<input checked="" type="checkbox"/> 3CWM 	<input checked="" type="checkbox"/> 3DRU 

Searching for templates



The neuronal network

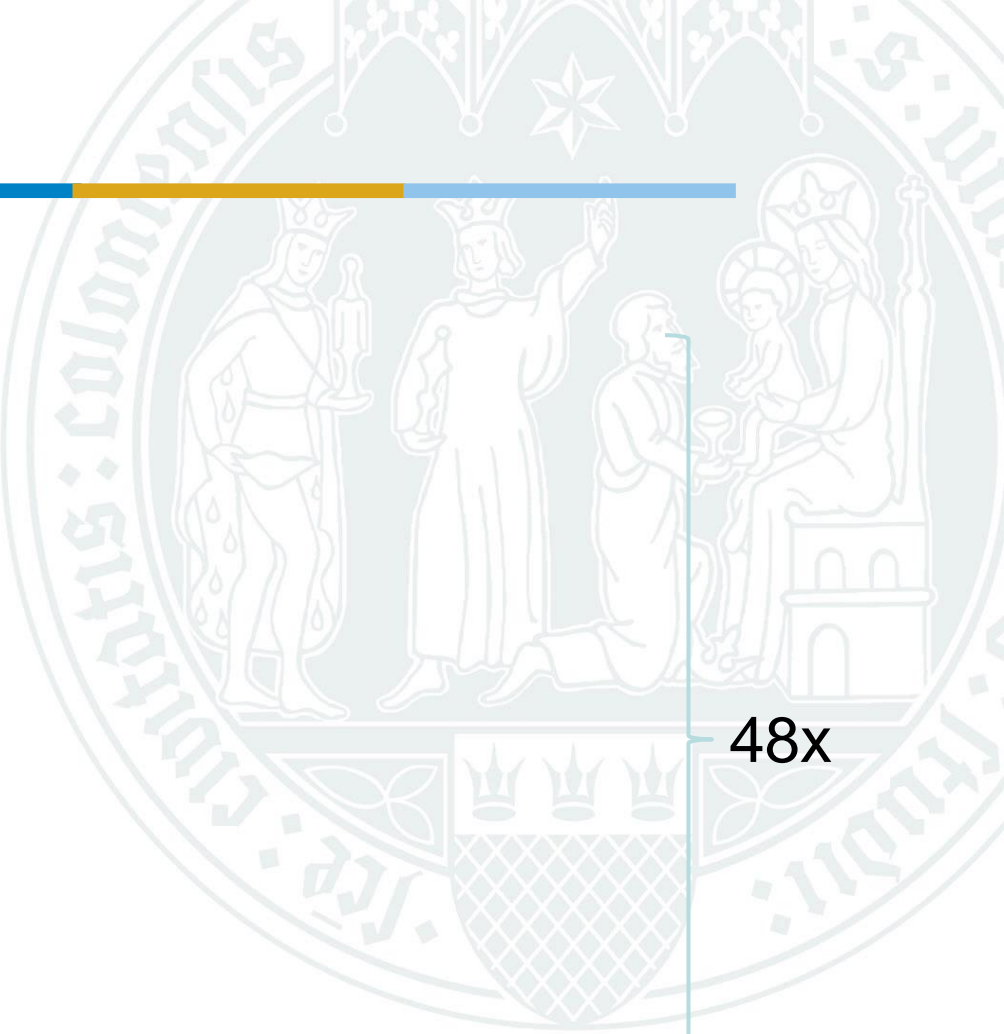
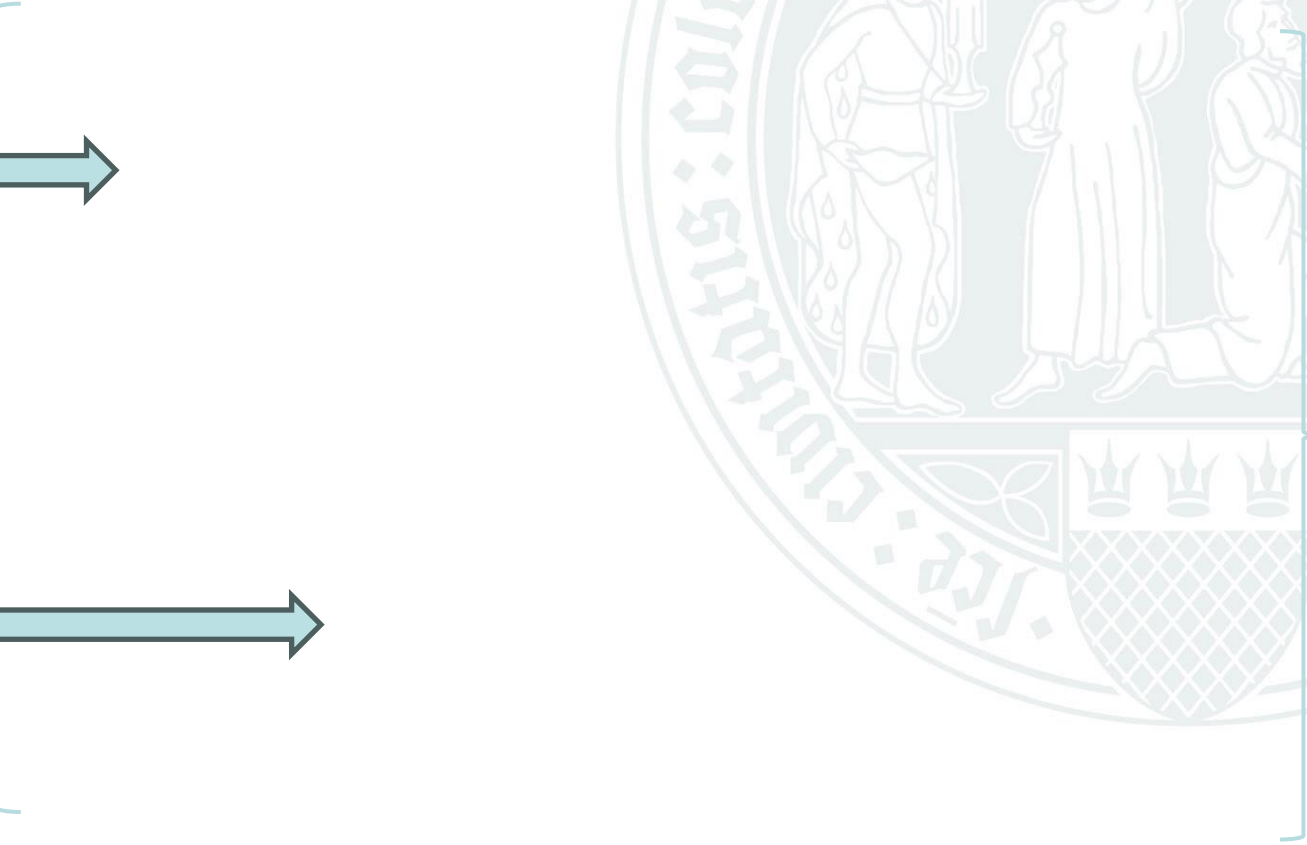
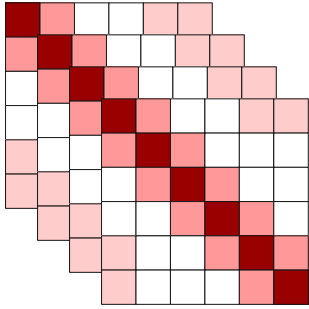
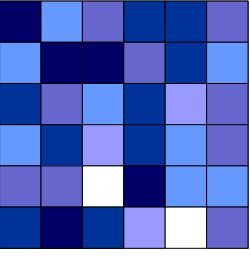
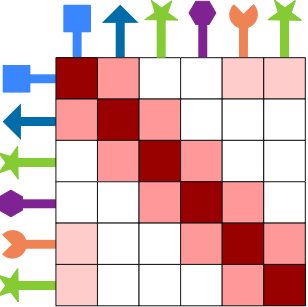
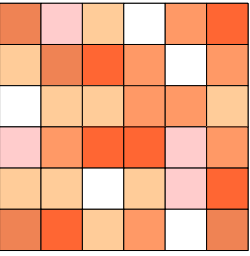
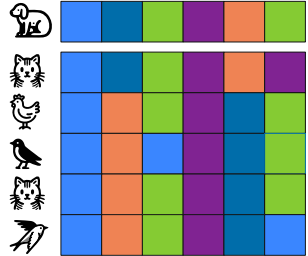


Data gathering

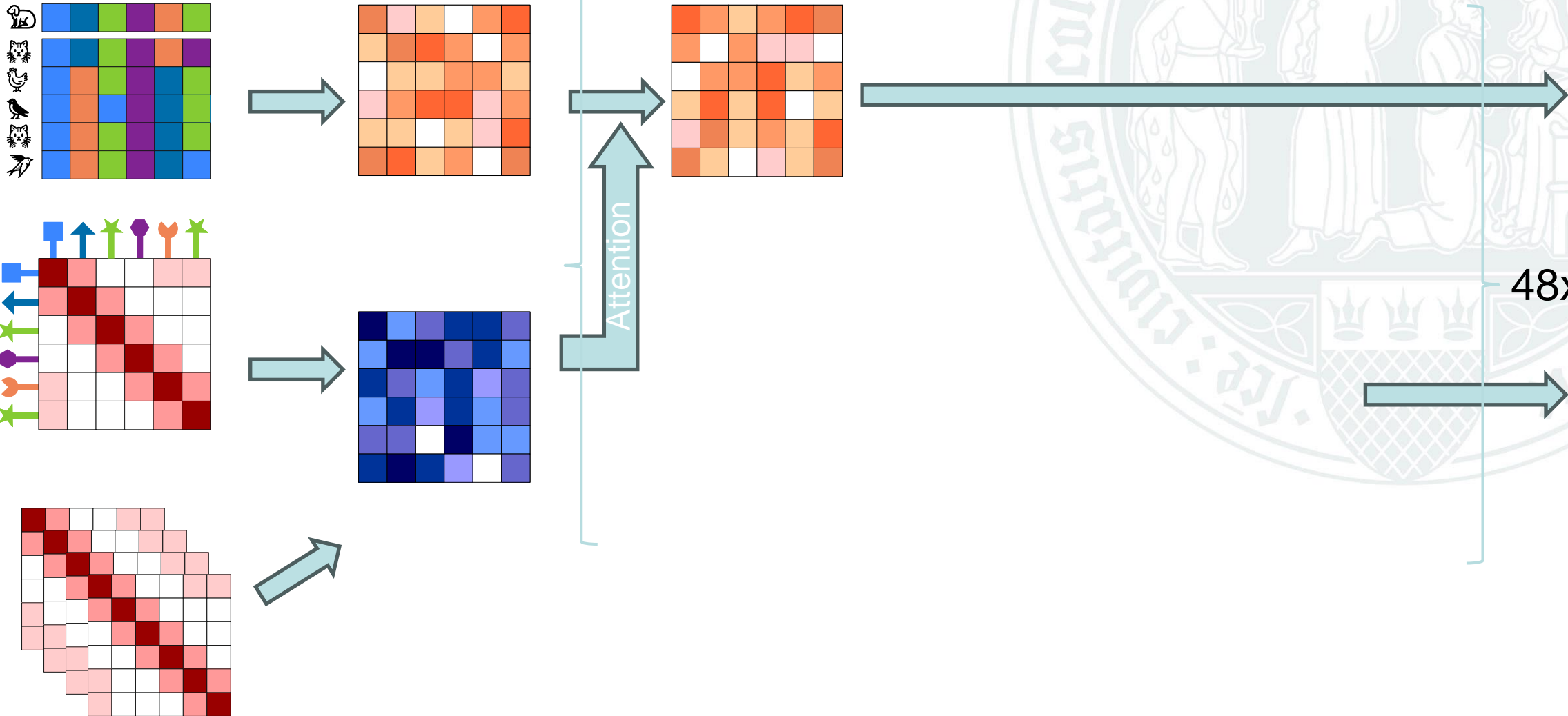
Distance estimation

Folding

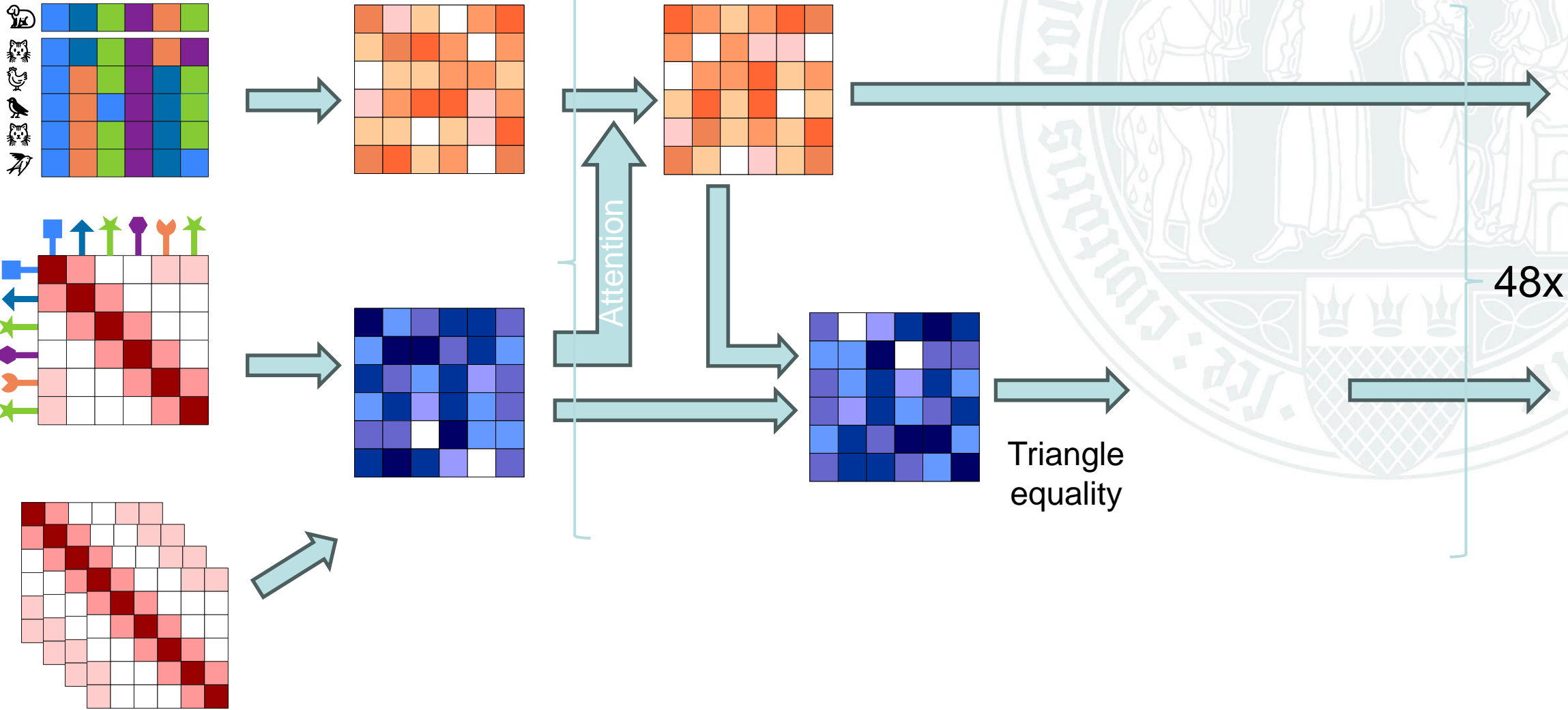
Evoformer



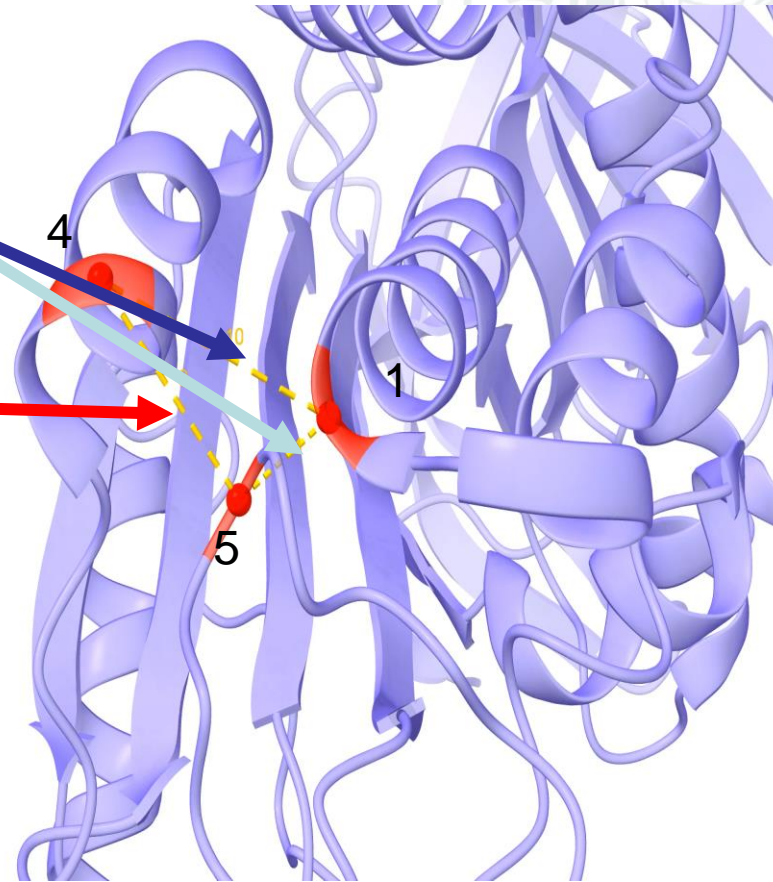
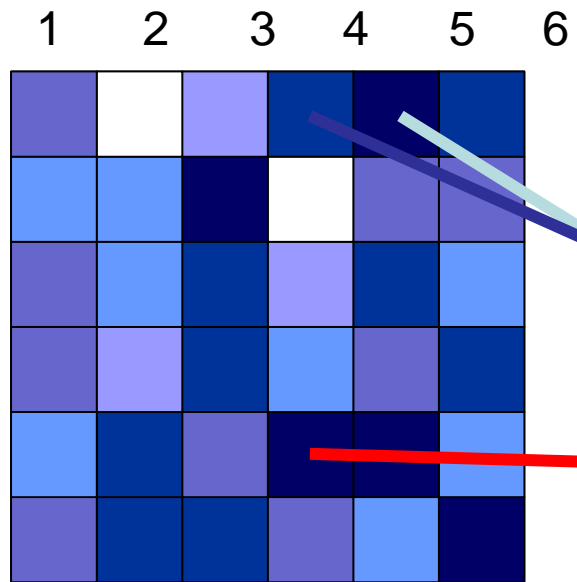
Evoformer



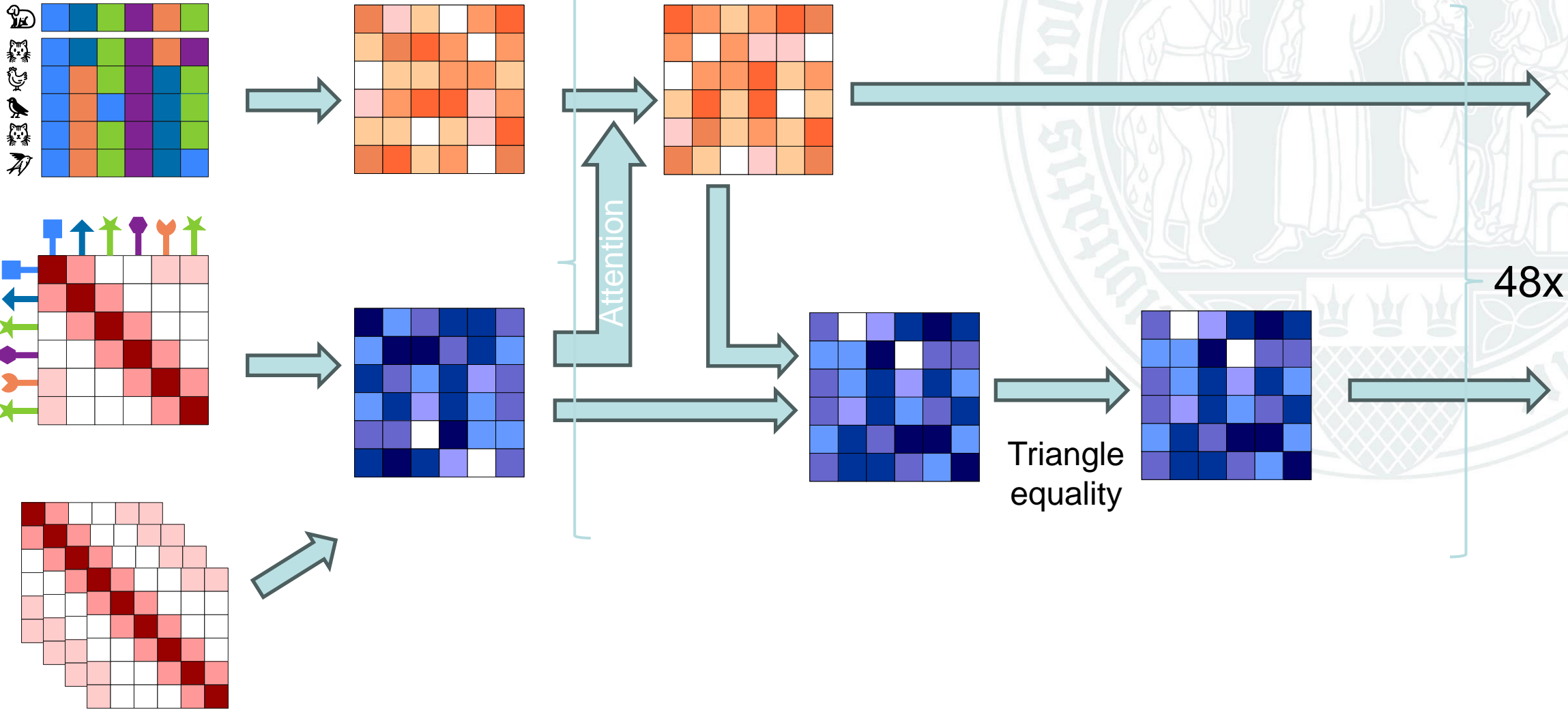
Evoformer



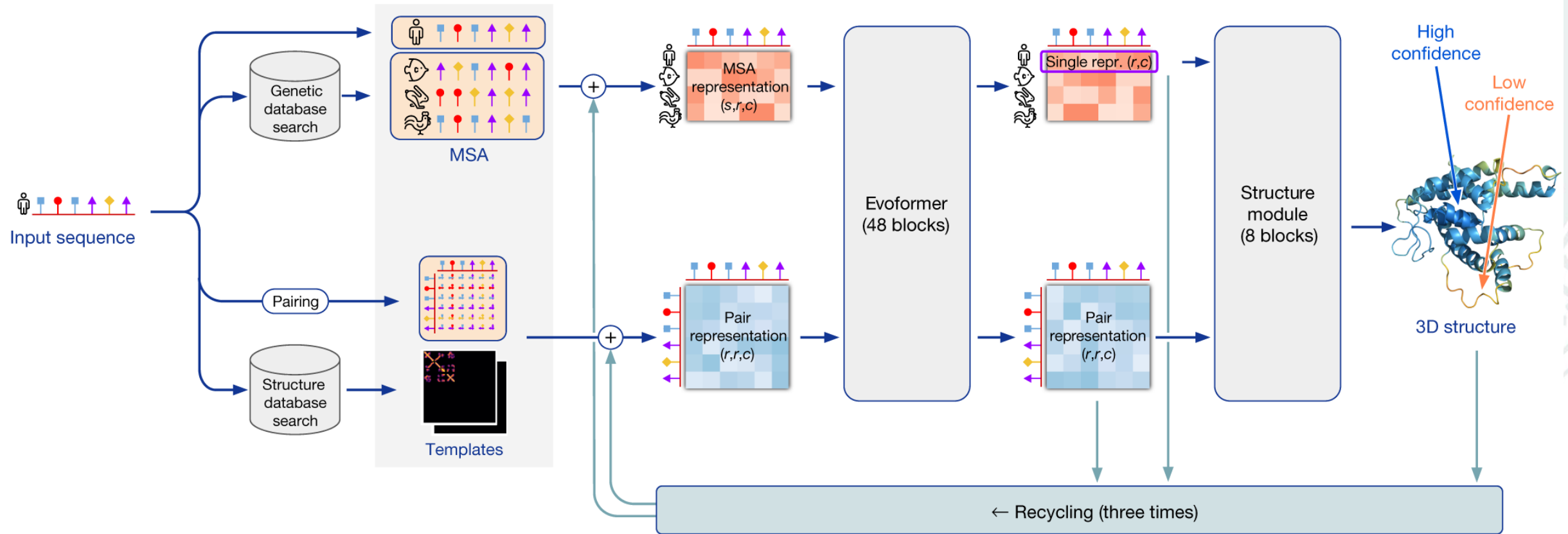
Triangle equality



Evoformer



The neuronal network



Data gathering

Distance estimation

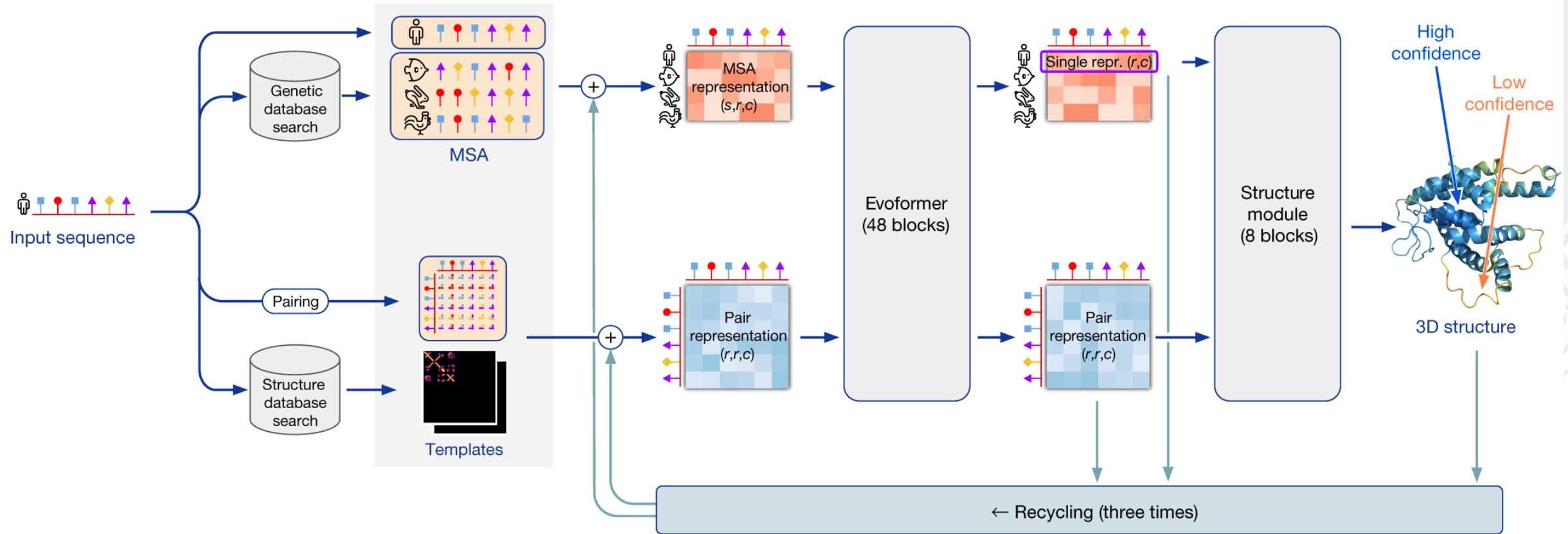
Folding

Modelling the restrains as a „protein gas“



Recycling iteration 0, block 01
Secondary structure assigned from the final prediction

The neuronal network

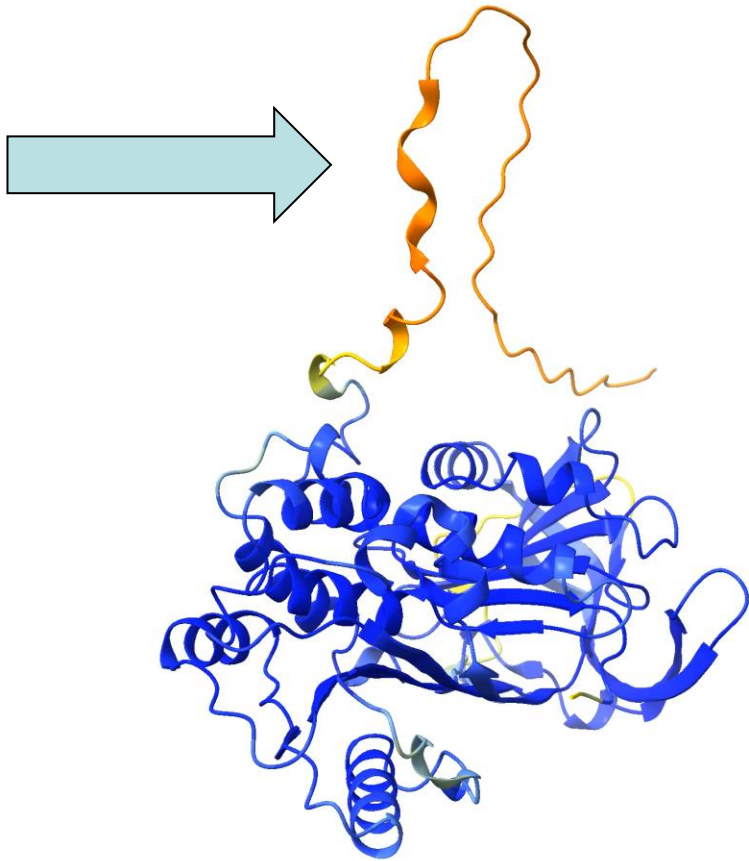


Data gathering

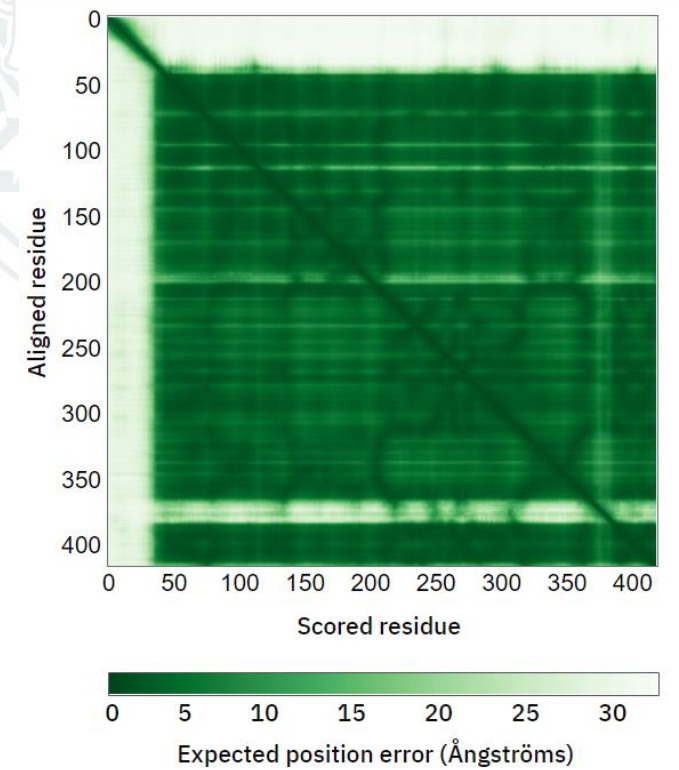
Distance estimation

Folding

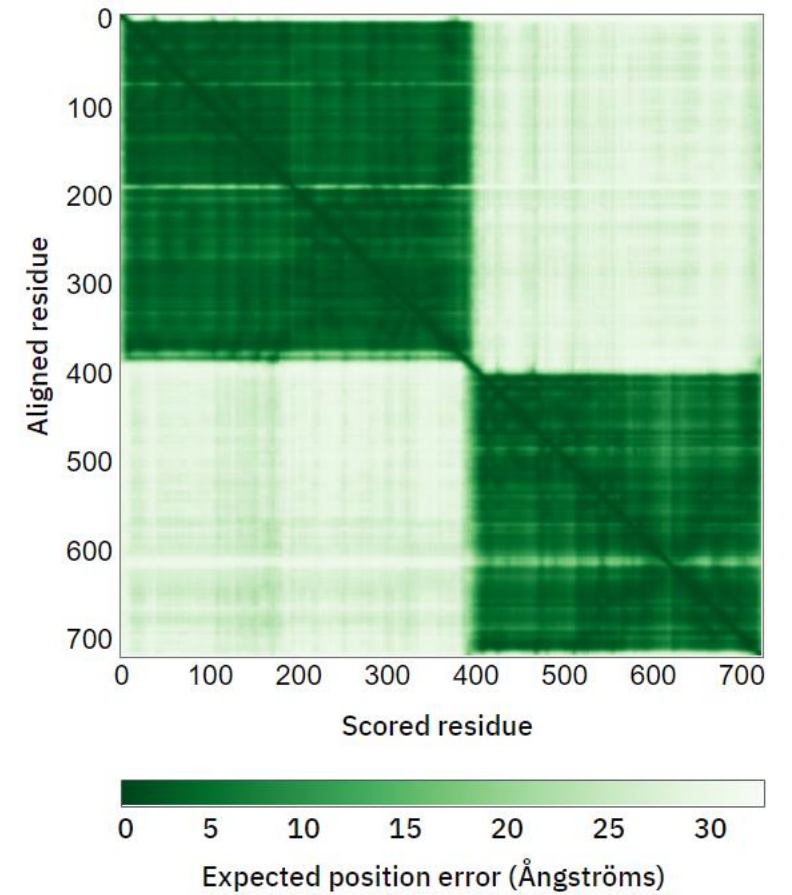
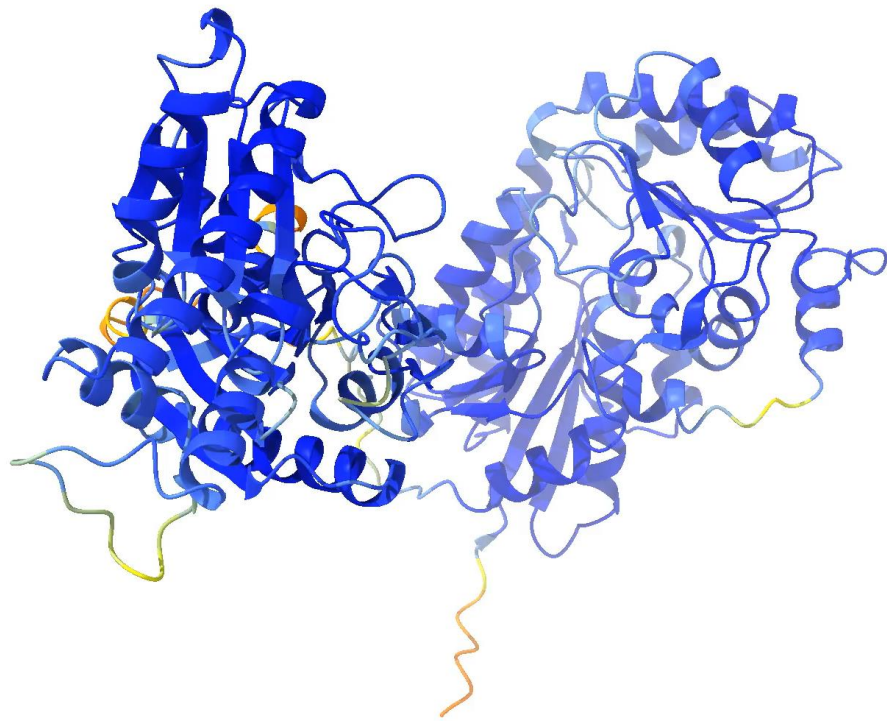
A simple case



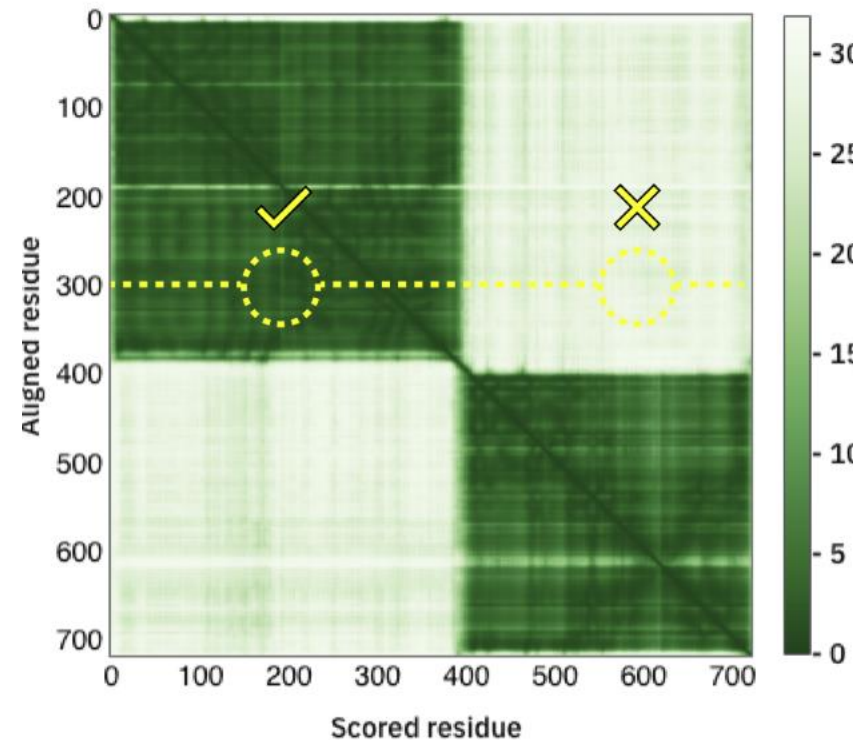
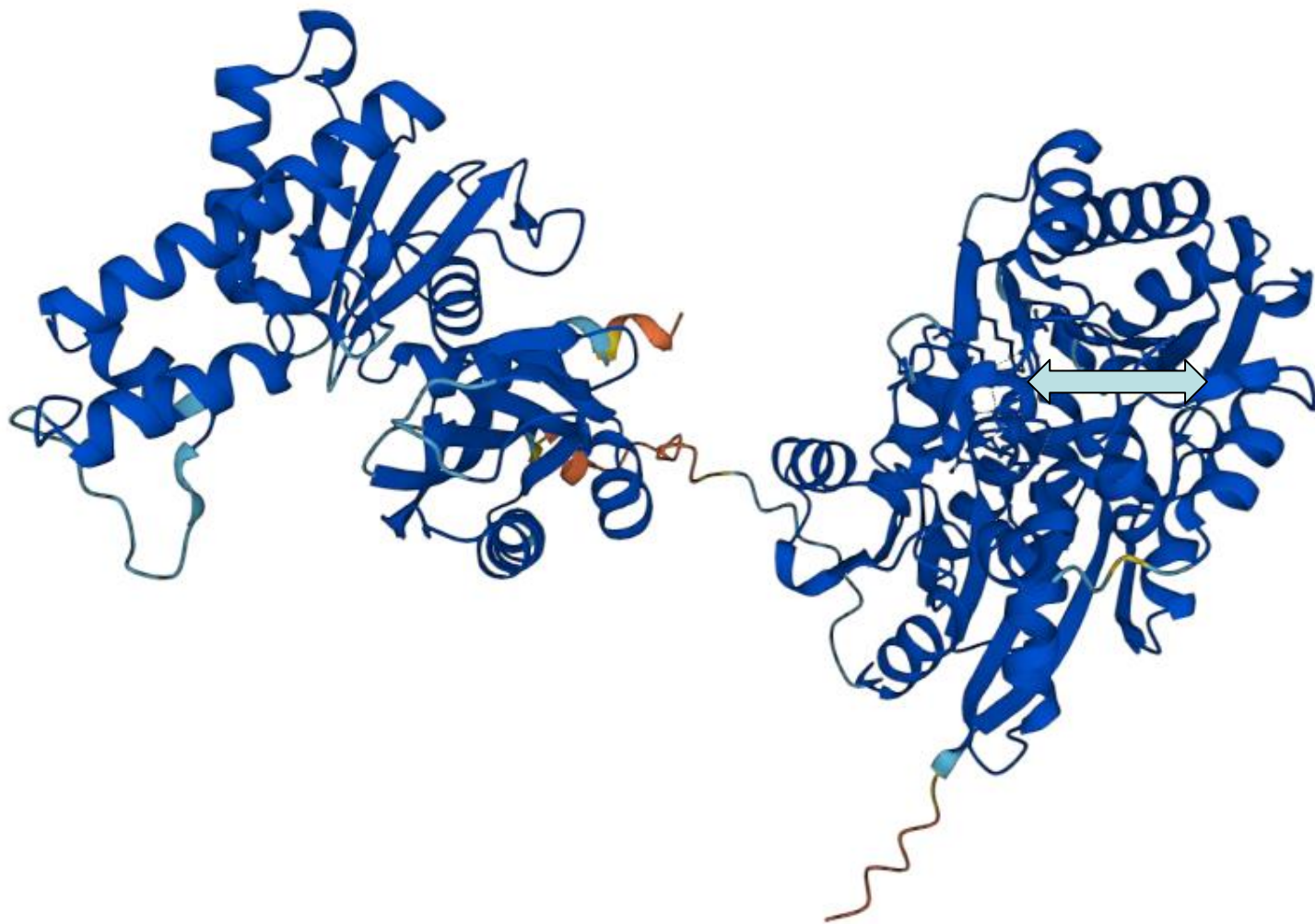
- Very high (pLDDT > 90)
- Confident (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)



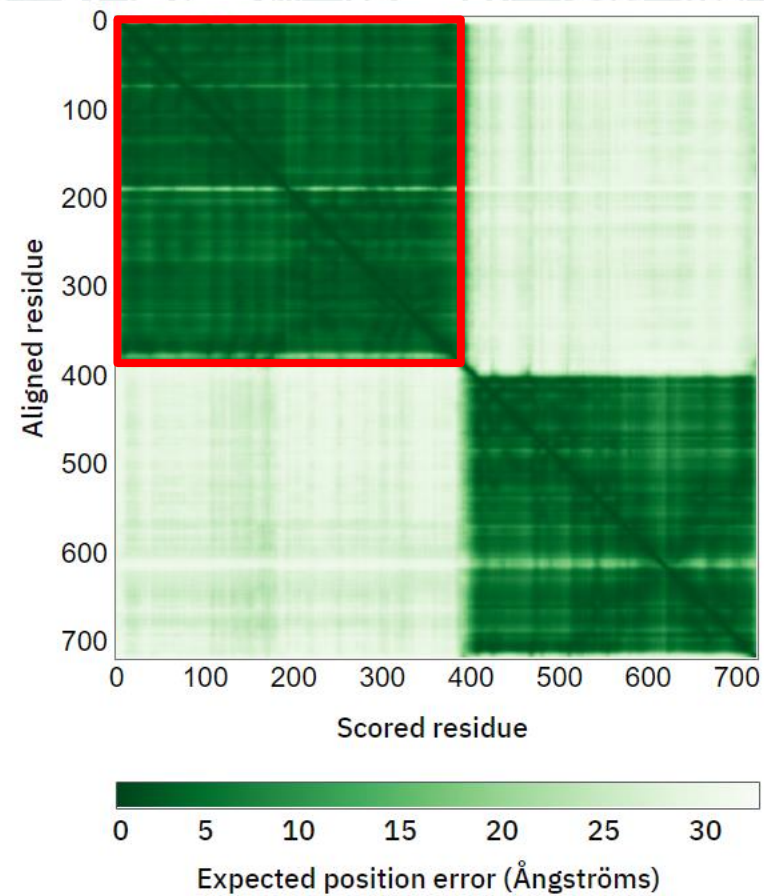
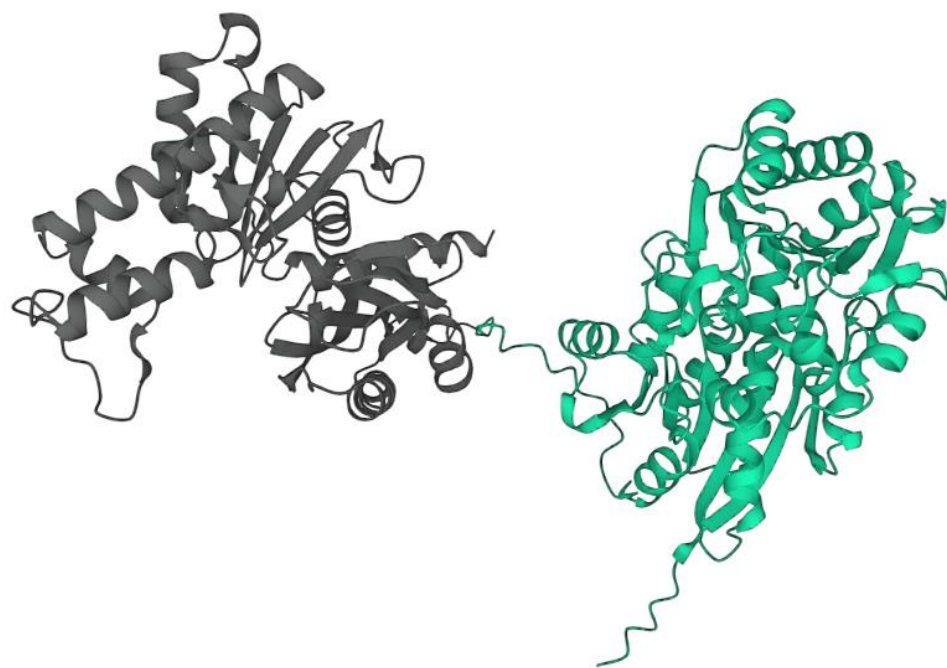
A little more complicated



A little more complicated



Let's have a closer look



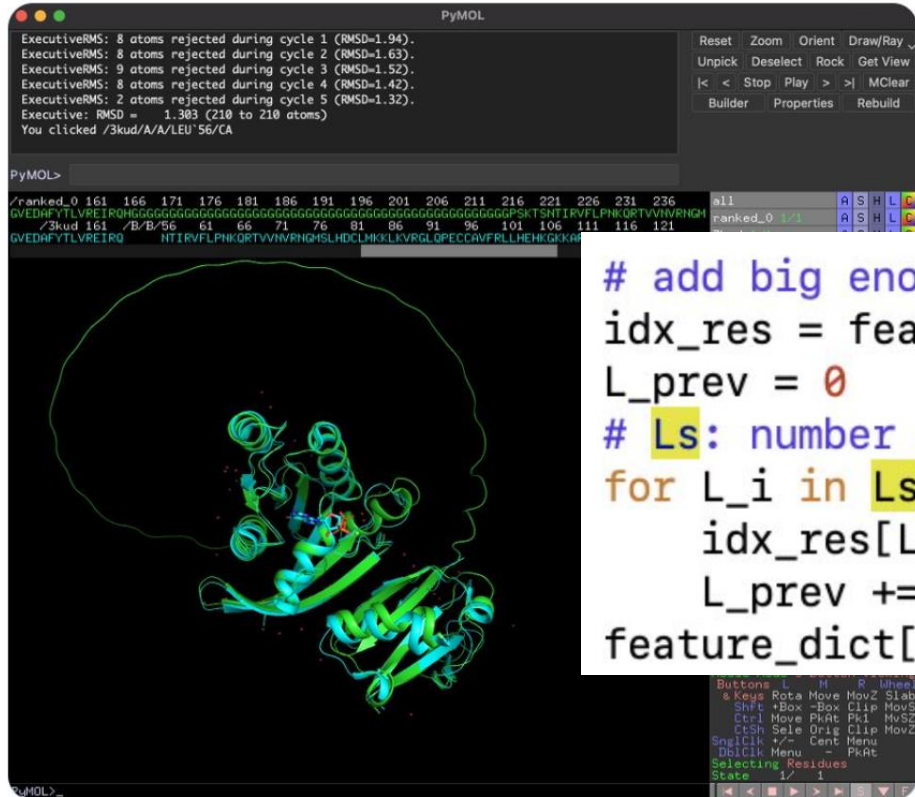
Hacking the network



Yoshitaka Moriwaki
@Ag_smith

AlphaFold2 can also predict heterocomplexes. All you have to do is input the two sequences you want to predict and connect them with a long linker.

[Post übersetzen](#)

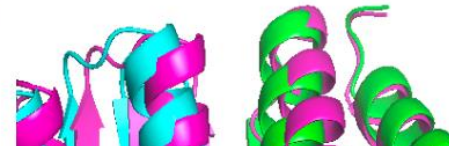


Minkyung Baek
@minkbaek

Adding a big enough number for "residue_index" feature is enough to model hetero-complex using AlphaFold (green&cyan: crystal structure / magenta: predicted model w/ residue_index modification).

[#AlphaFold](#) [#alphafold2](#)

```
to residue index  
'residue_index']
```



```
# add big enough number to residue index to indicate chain breaks  
idx_res = feature_dict['residue_index']  
L_prev = 0  
# Ls: number of residues in each chain  
for L_i in Ls[::-1]:  
    idx_res[L_prev+L_i:] += 200  
    L_prev += L_i  
feature_dict['residue_index'] = idx_res
```

All at your fingertip(s) ...



Sergey Ovchinnikov

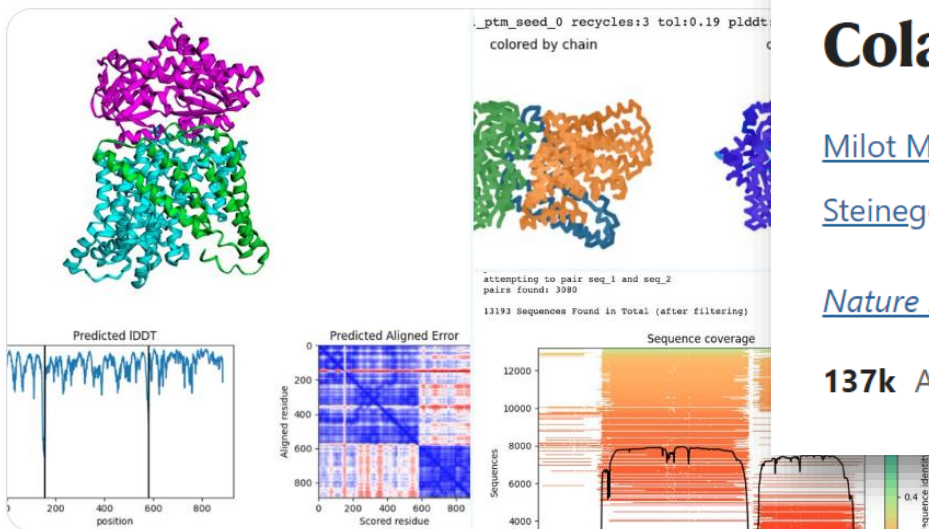
@sokrypton

ColabFold AlphaFold2_advanced, now supports higher-order homo/hetero-complexes (or however number you can fit into)



WARNING: #alphafold was only trained and validated on single (monomers). Modeling complexes is an unintended use case and is experimental!

[Post übersetzen](#)



10:18 nachm. · 8. Aug. 2021

nature methods

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature methods](#) > [brief communications](#) > article

Brief Communication | [Open access](#) | [Published: 30 May 2022](#)

ColabFold: making protein folding accessible to all

[Milot Mirdita](#) ✉, [Konstantin Schütze](#), [Yoshitaka Moriwaki](#), [Lim Heo](#), [Sergey Ovchinnikov](#) ✉ & [Martin Steinegger](#) ✉

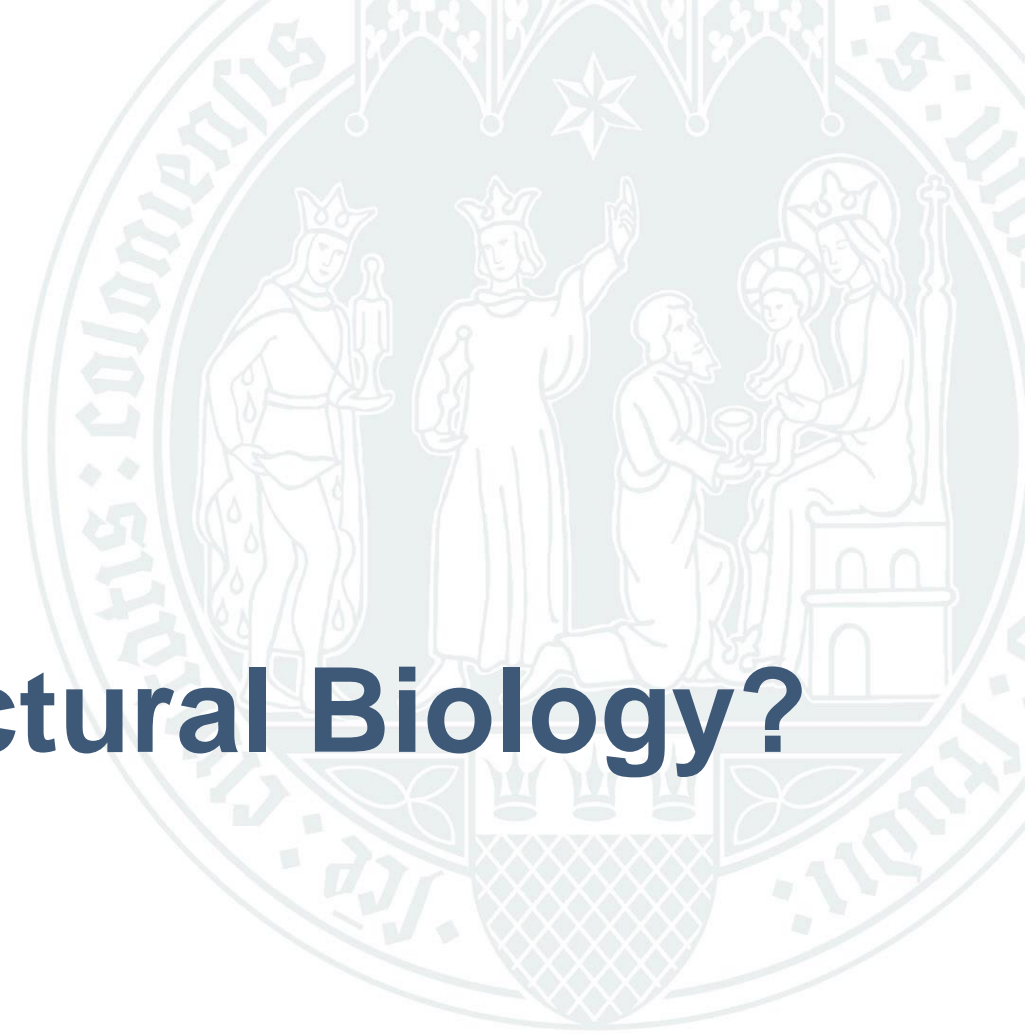
[Nature Methods](#) **19**, 679–682 (2022) | [Cite this article](#)

137k Accesses | **1732** Citations | **415** Altmetric | [Metrics](#)

<https://github.com/sokrypton/ColabFold>



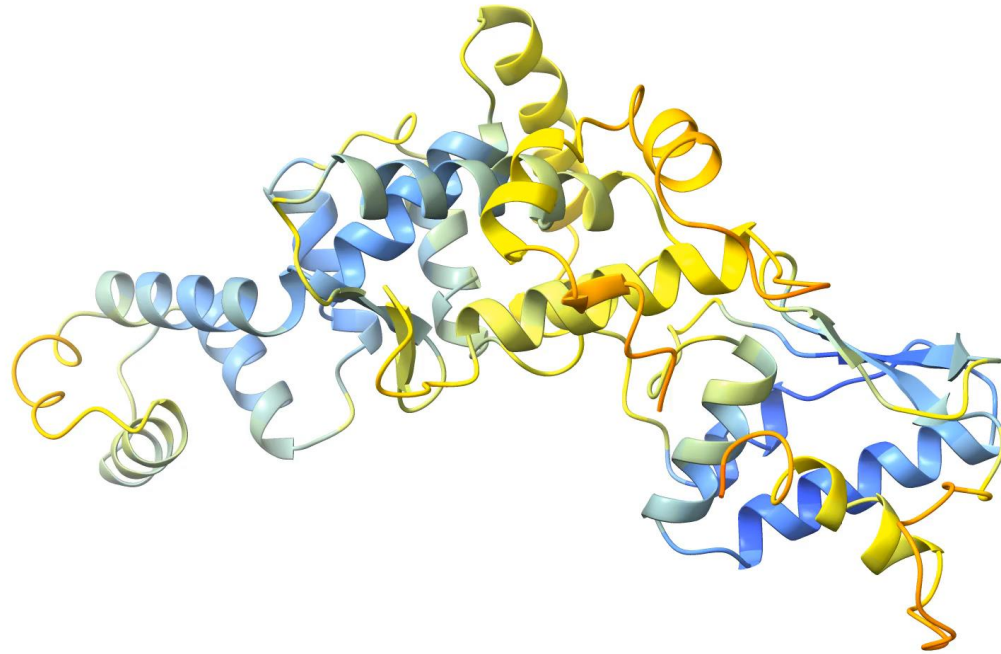
How did it change Structural Biology?



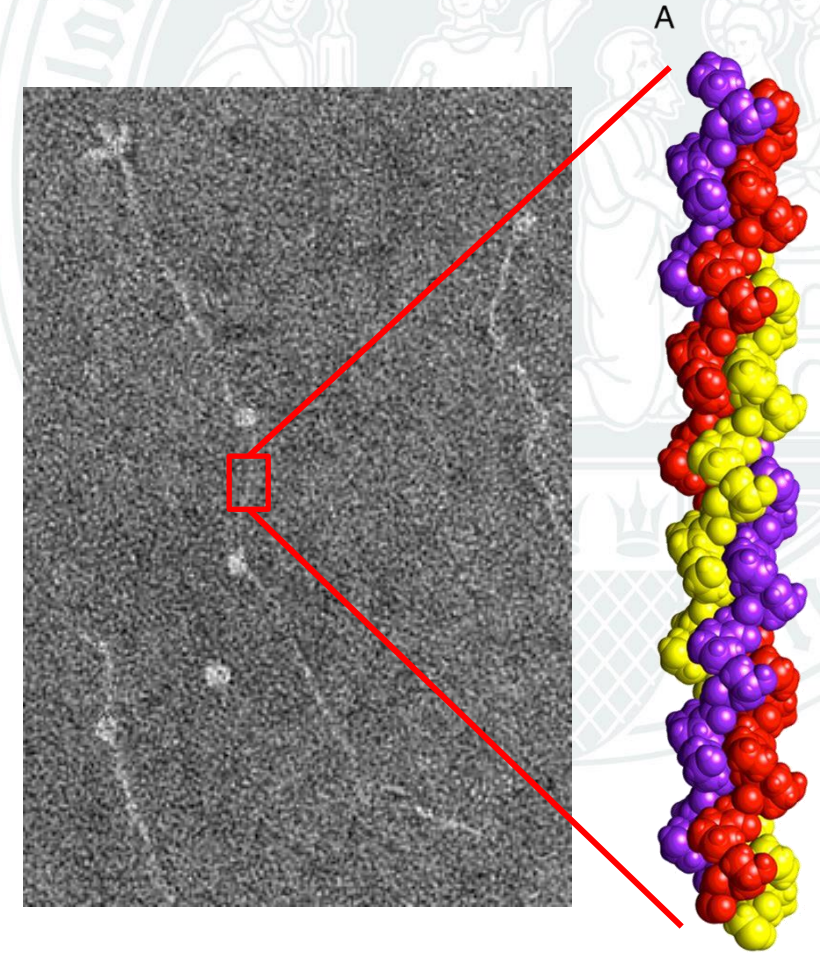
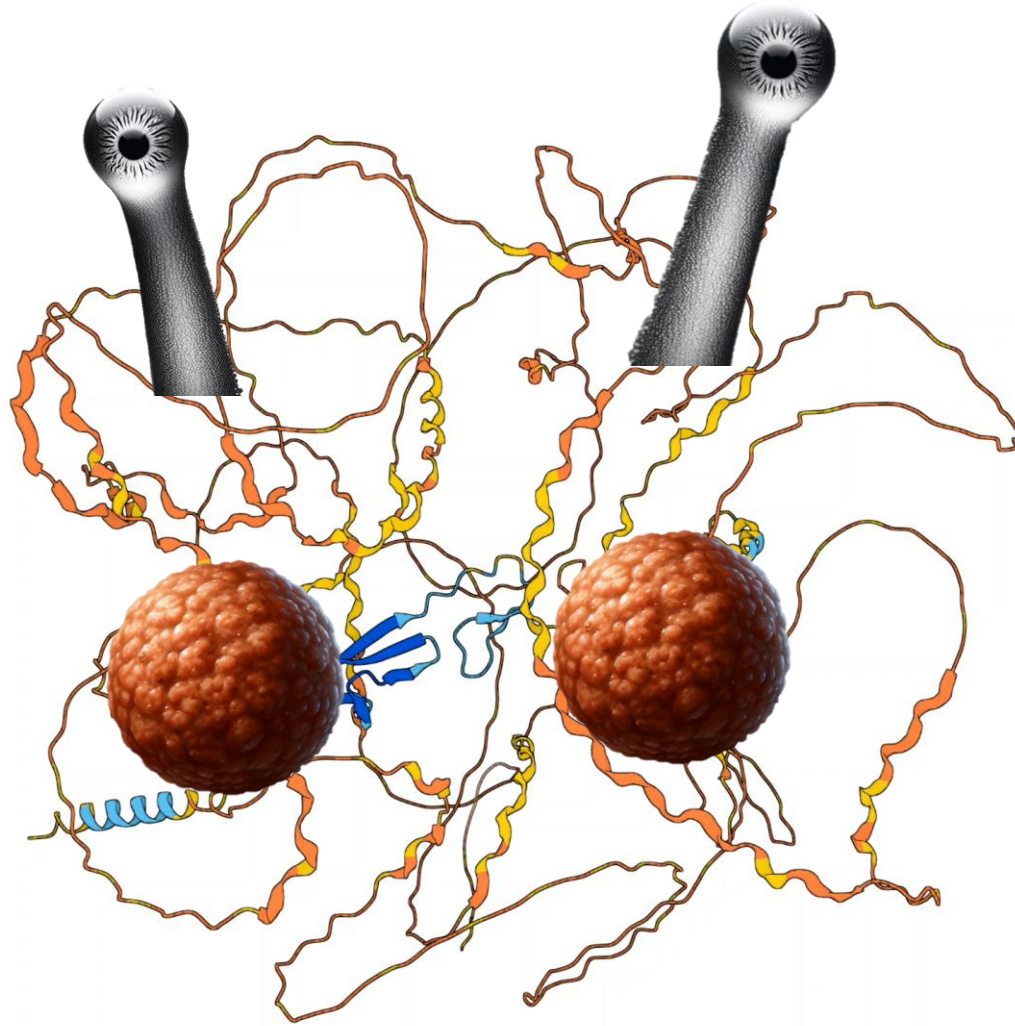
Mostly it's very good



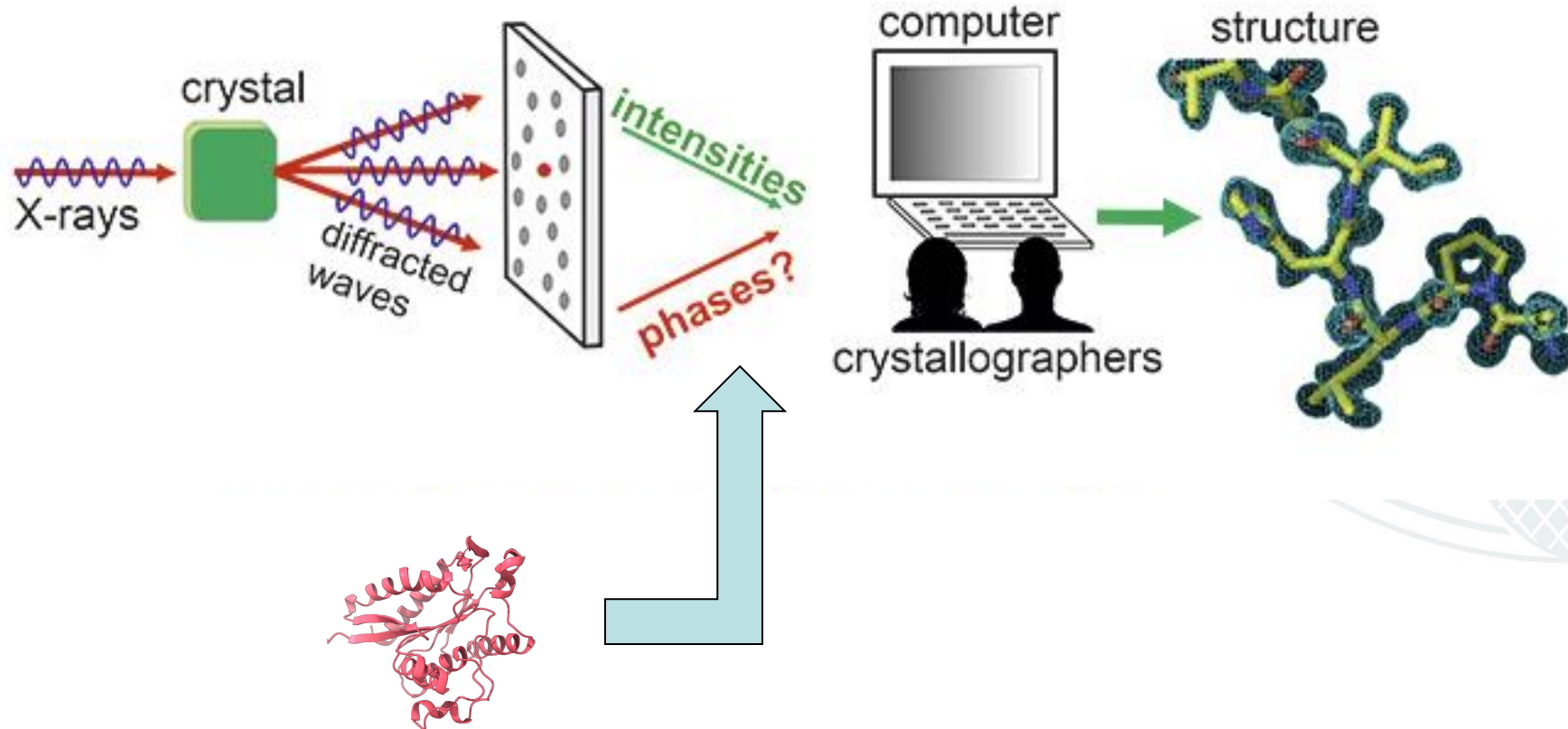
Sometimes it's wrong



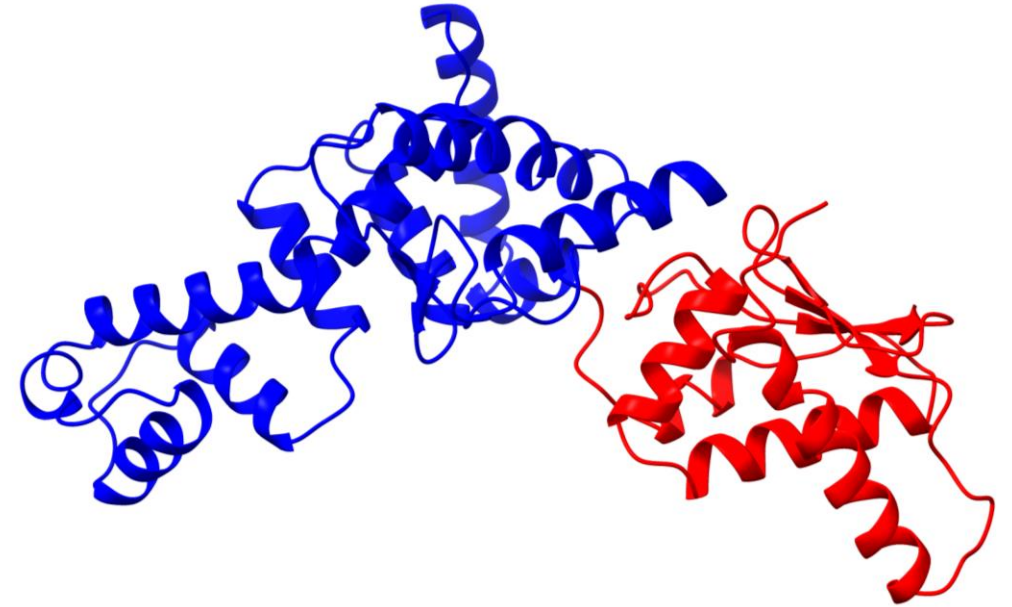
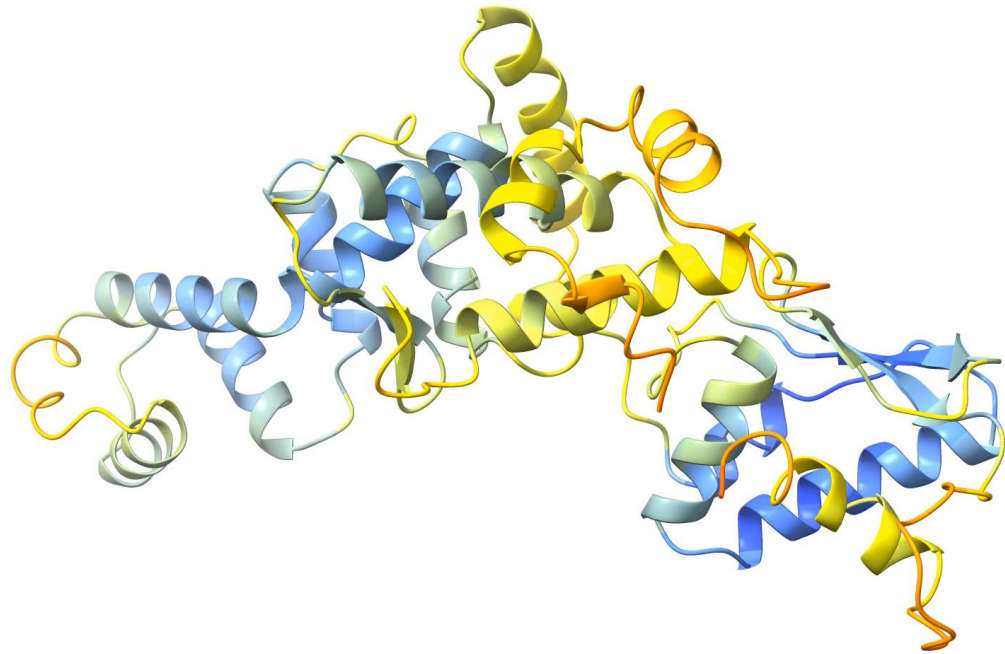
Sometimes it's very wrong...



Helpful for Experimentalists



Helpful for Experimentalists



Use the prediction as initial solution
and correct the errors

Integrative modelling building

Science Current Issue First release papers Archive About Submit manus

HOME > SCIENCE > VOL. 376, NO. 6598 > STRUCTURE OF CYTOPLASMIC RING OF NUCLEAR PORE COMPLEX BY INTEGRATIVE CRYO-EM AND ALPHAFOLD

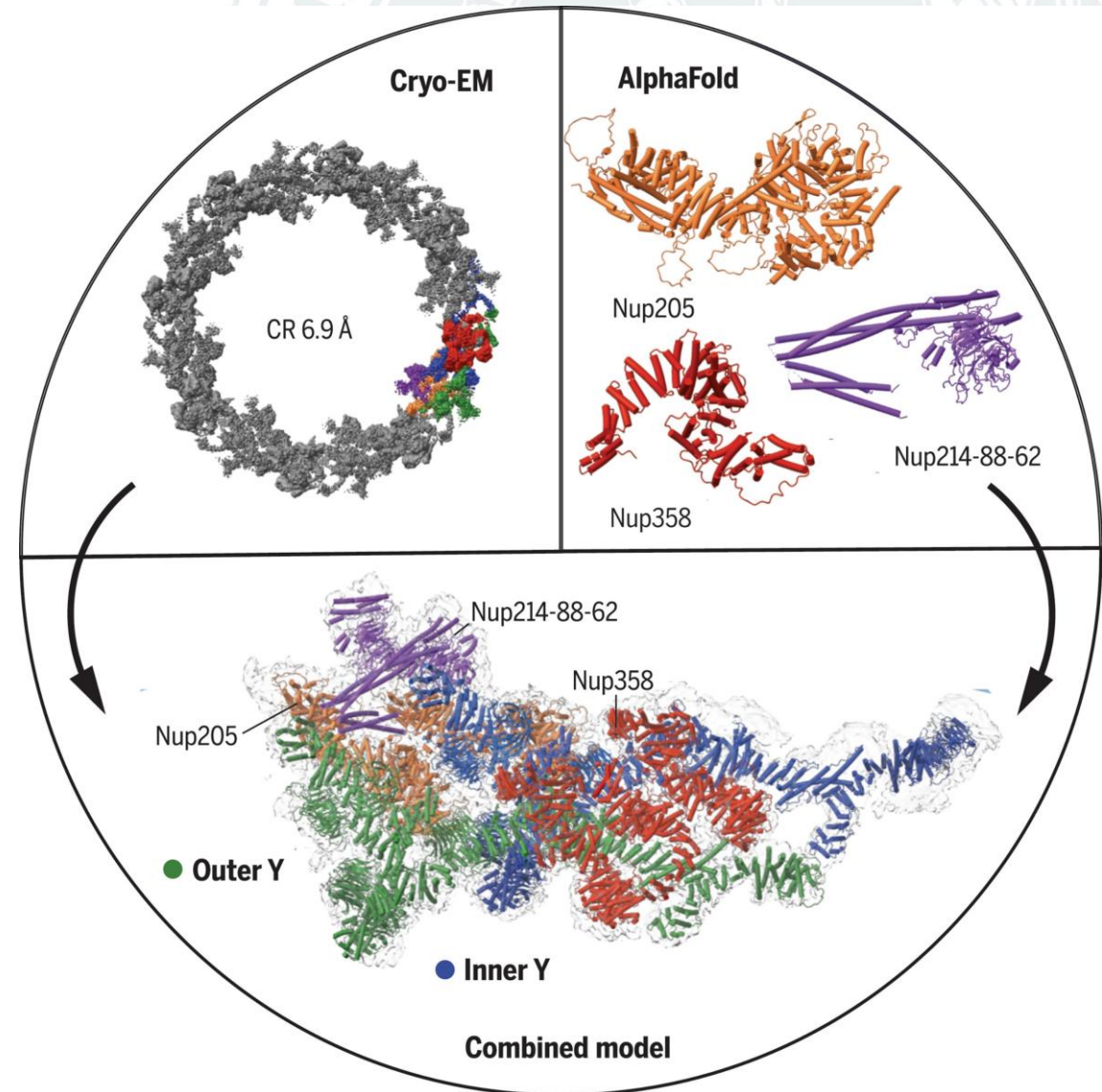
SPECIAL ISSUE RESEARCH ARTICLE | NUCLEAR PORE COMPLEX

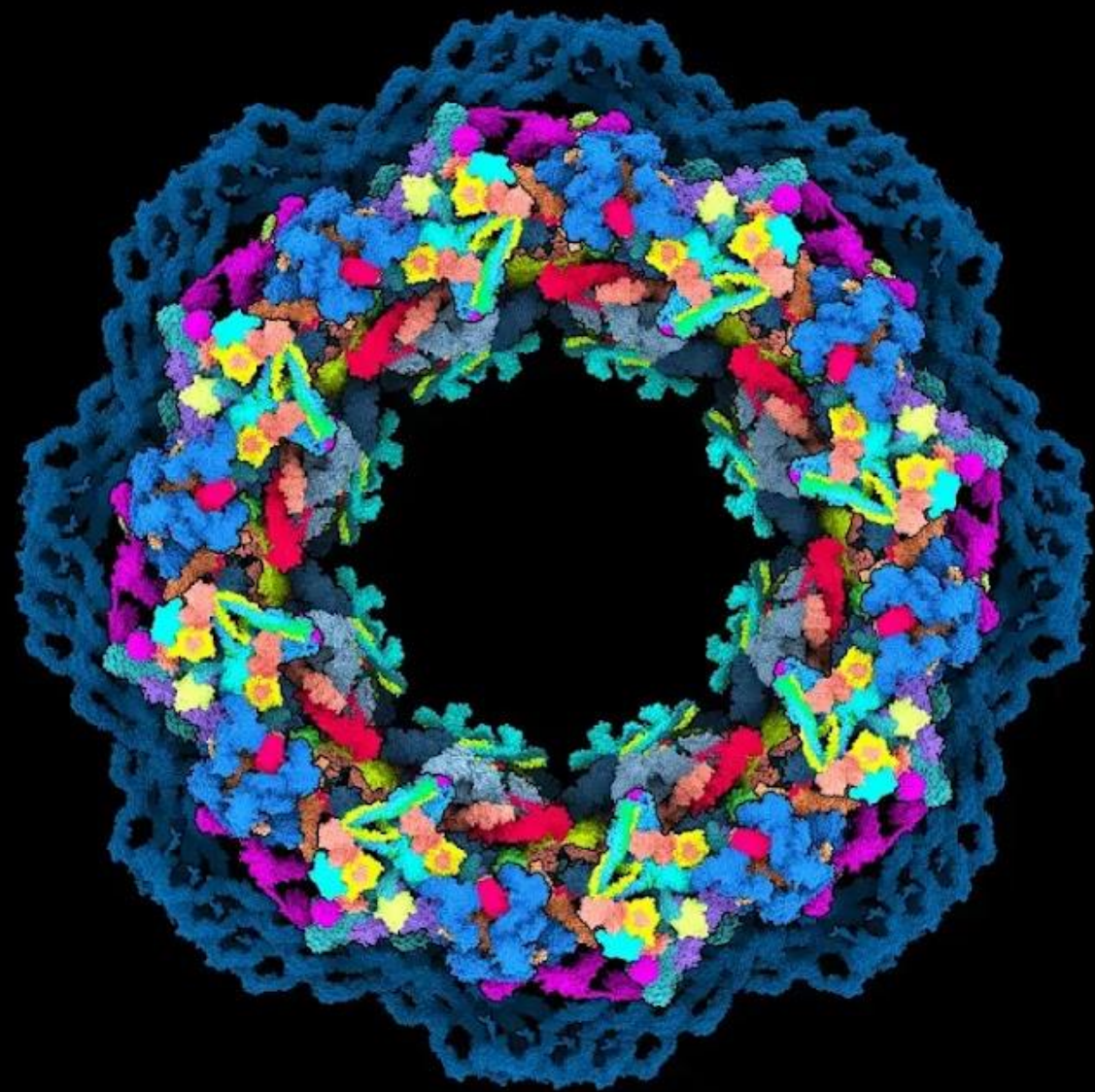
Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold

PIETRO FONTANA, YING DONG, XIONG PI, ALEXANDER B. TONG, COREY W. HECKSEL, LONGFEI WANG, TIAN-MIN FU, CARLOS BUSTAMANTE, AND HAO WU [Authors Info & Affiliations](#)

SCIENCE • 10 Jun 2022 • Vol 376, Issue 6598 • DOI: 10.1126/science.abm9326

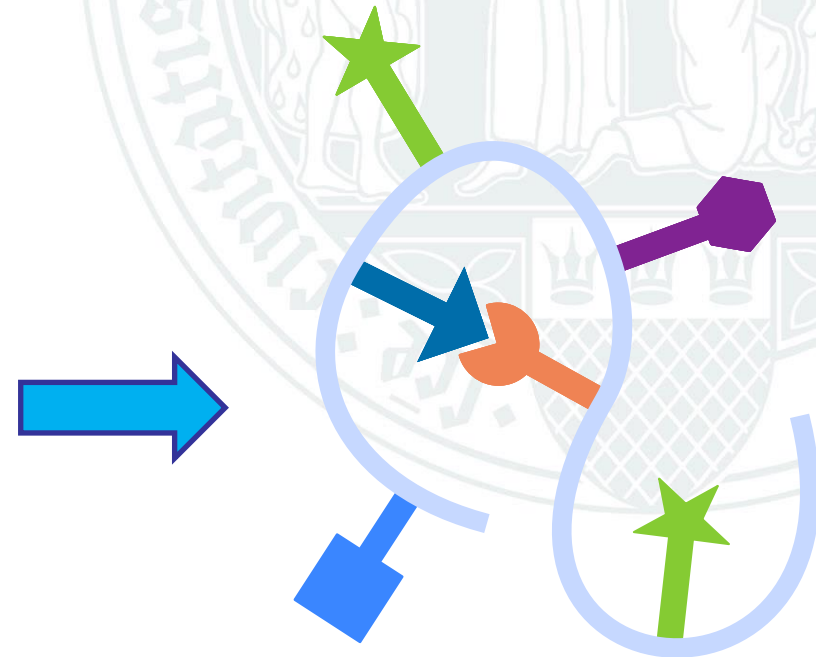
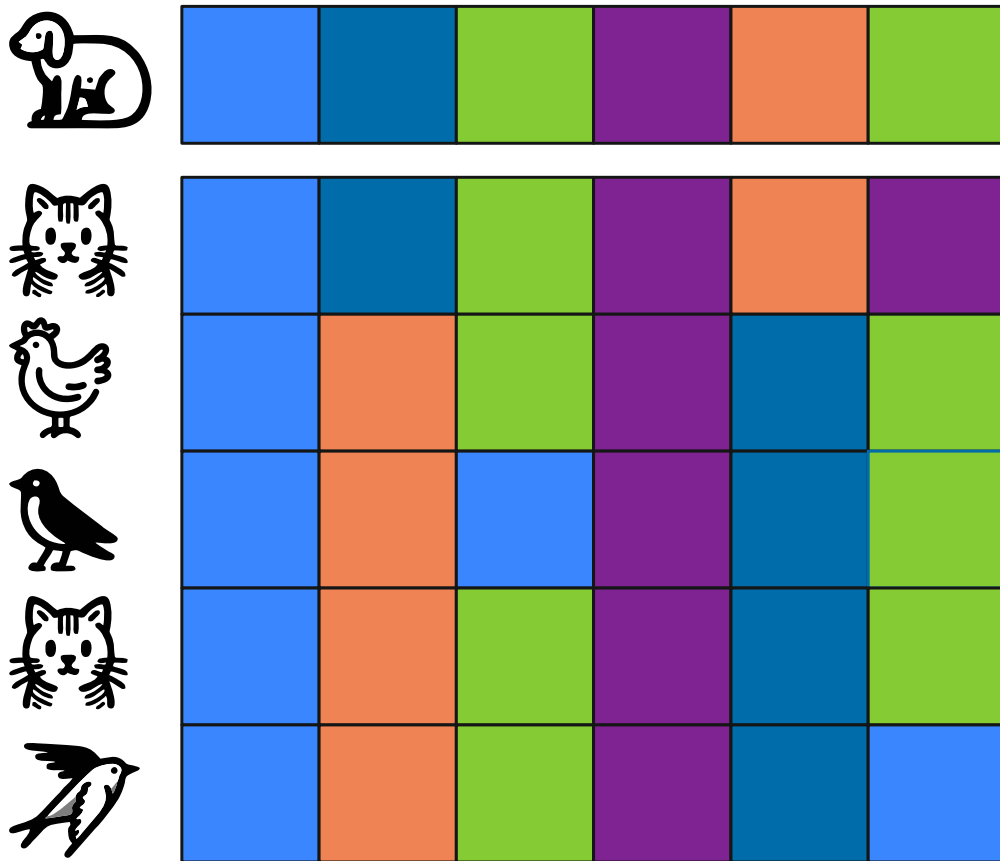
12,161 32





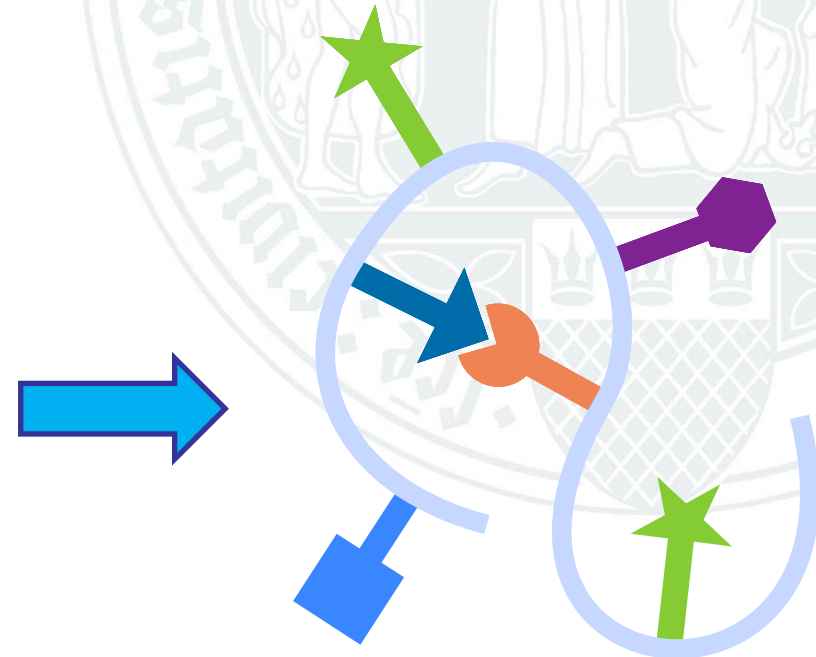
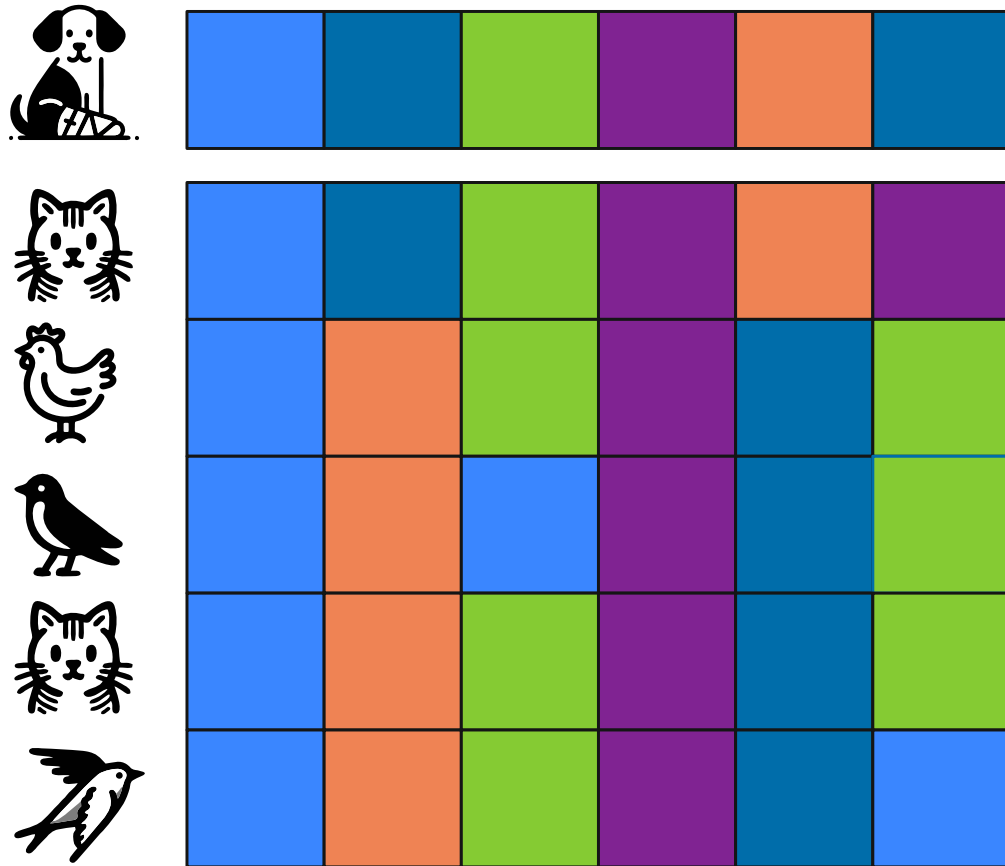
What it cannot do

Can't predict effects of single mutations (yet?)



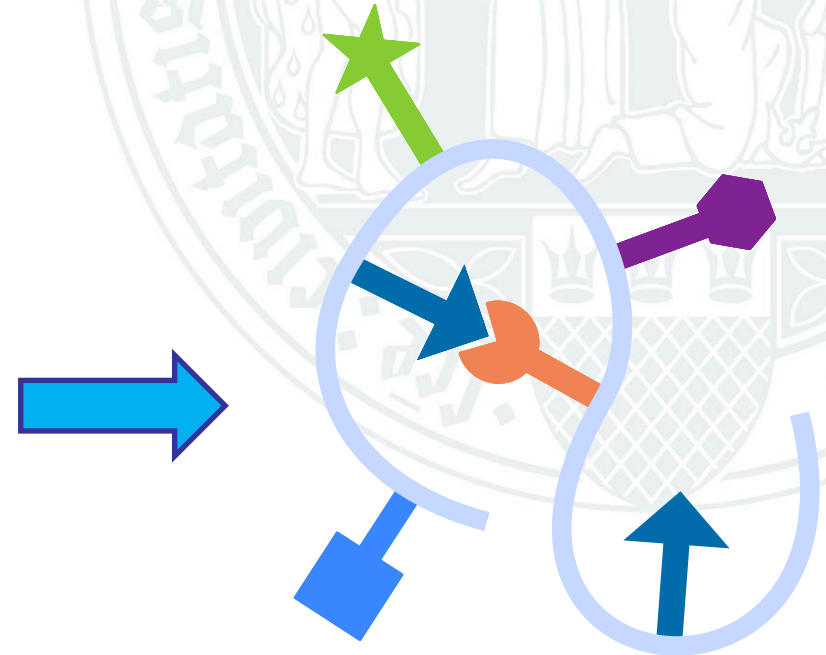
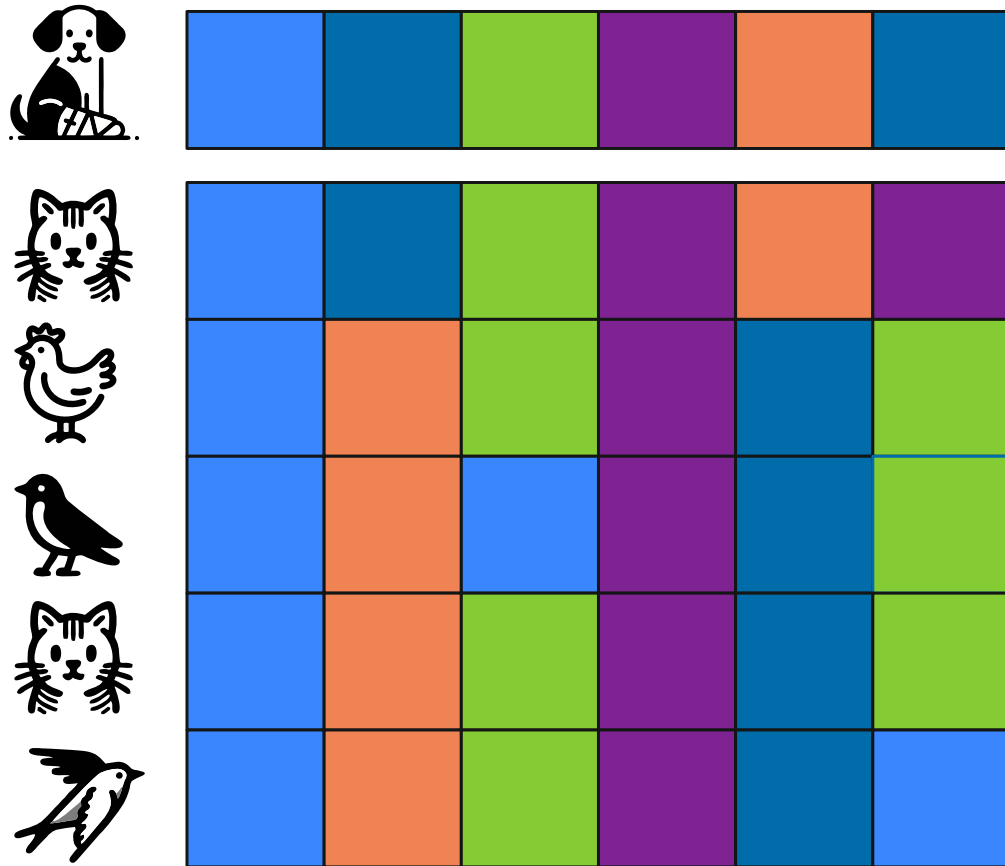
What it cannot do

Can't predict effects of single mutations (yet?)



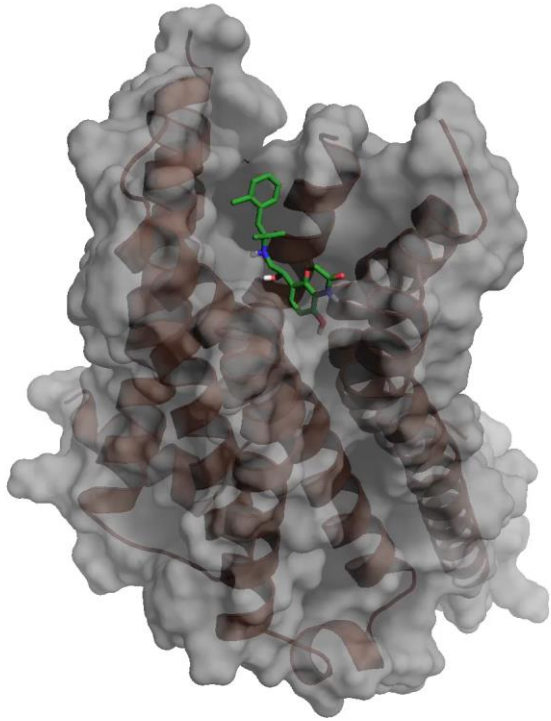
What it cannot do

Can't predict effects of single mutations (yet?)



What it cannot do

Can't predict effects of single mutations (yet?)
Often not good enough for docking (yet)



What it cannot do



Can't predict effects of single mutations (yet?)
Often not good enough for docking (yet)
You can never be sure (as it's only a prediction!)

nature methods



Article

<https://doi.org/10.1038/s41592-023-02087-4>

AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination

Received: 30 January 2023

Accepted: 11 October 2023

Published online: 30 November 2023

Thomas C. Terwilliger^{1,2}✉, Dorothee Liebschner³, Tristan I. Croll⁴,
Christopher J. Williams⁵, Airlie J. McCoy⁴, Billy K. Poon³, Pavel V. Afonine³,
Robert D. Oeffner⁴, Jane S. Richardson⁵, Randy J. Read⁴ &
Paul D. Adams^{3,6}

AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination

[Thomas C. Terwilliger](#) 

[K. Poon](#), [Pavel V. Afonine](#)

[Nature Methods](#) (2023)

[...] We suggest considering AlphaFold predictions as exceptionally useful hypotheses. We further suggest that it is important to consider the confidence in prediction when interpreting AlphaFold predictions and to carry out experimental structure determination to verify structural details, particularly those that involve interactions not included in the prediction.

Exciting times ahead of us

Has AlphaFold changed structural biology forever?

So structural biology has been greatly advanced by these new tools. But it has not been outmoded, replaced, or rendered irrelevant. It's more relevant than ever, and now we can get down to even bigger questions with it.



Derek Lowe

Acknowledgments



Google Deepmind (for developing and sharing).

The Computational Structural Biology Community, for all the previous attempts and ideas.

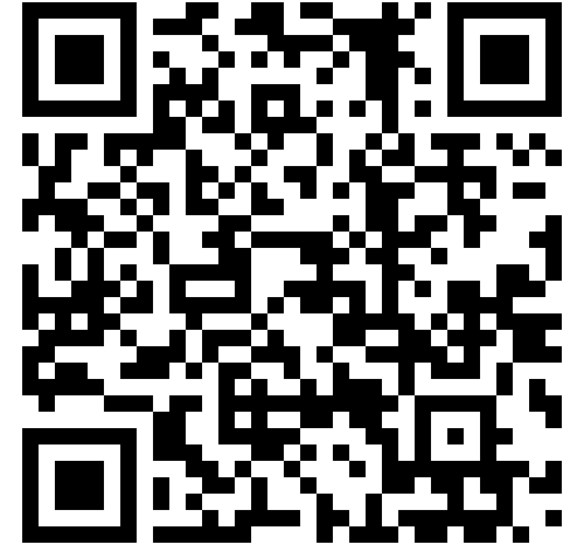
All structural biologist working on this problem for > 50 years, for providing this beautifully annotated datasets.

Fun & Games

Corona Virus Animation



ColabFold – Fold your own proteins



FoldIt – Play being a structural biologist



What's about ColabFold?