



# PUXANO

Artificial Intelligence: A Catalyst for Innovation  
in Experimental Structural Biology

April 30th 2024 – EuroCC Austria

Presented by Wouter Van Putte, CEO and Founder of PUXANO

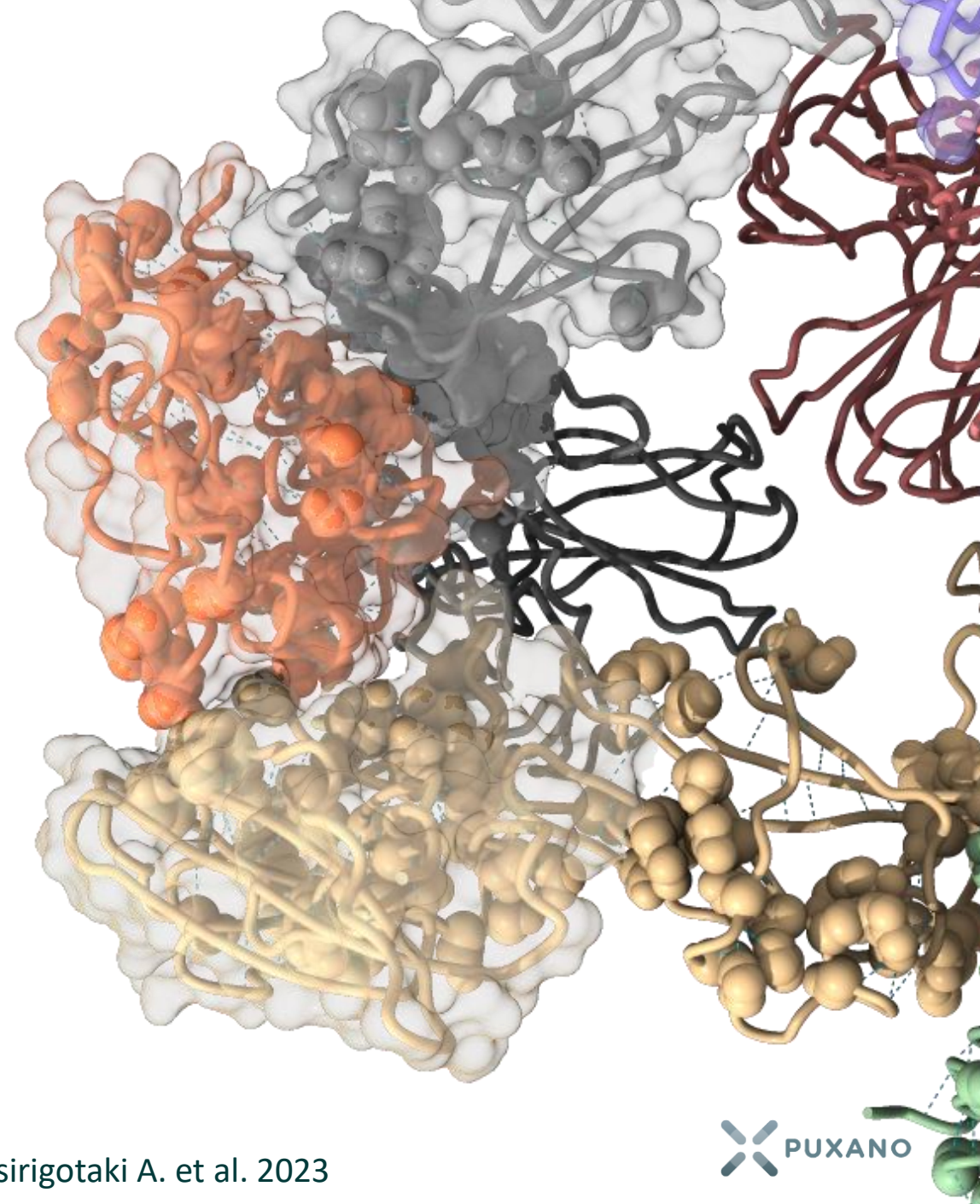
# PUXANO

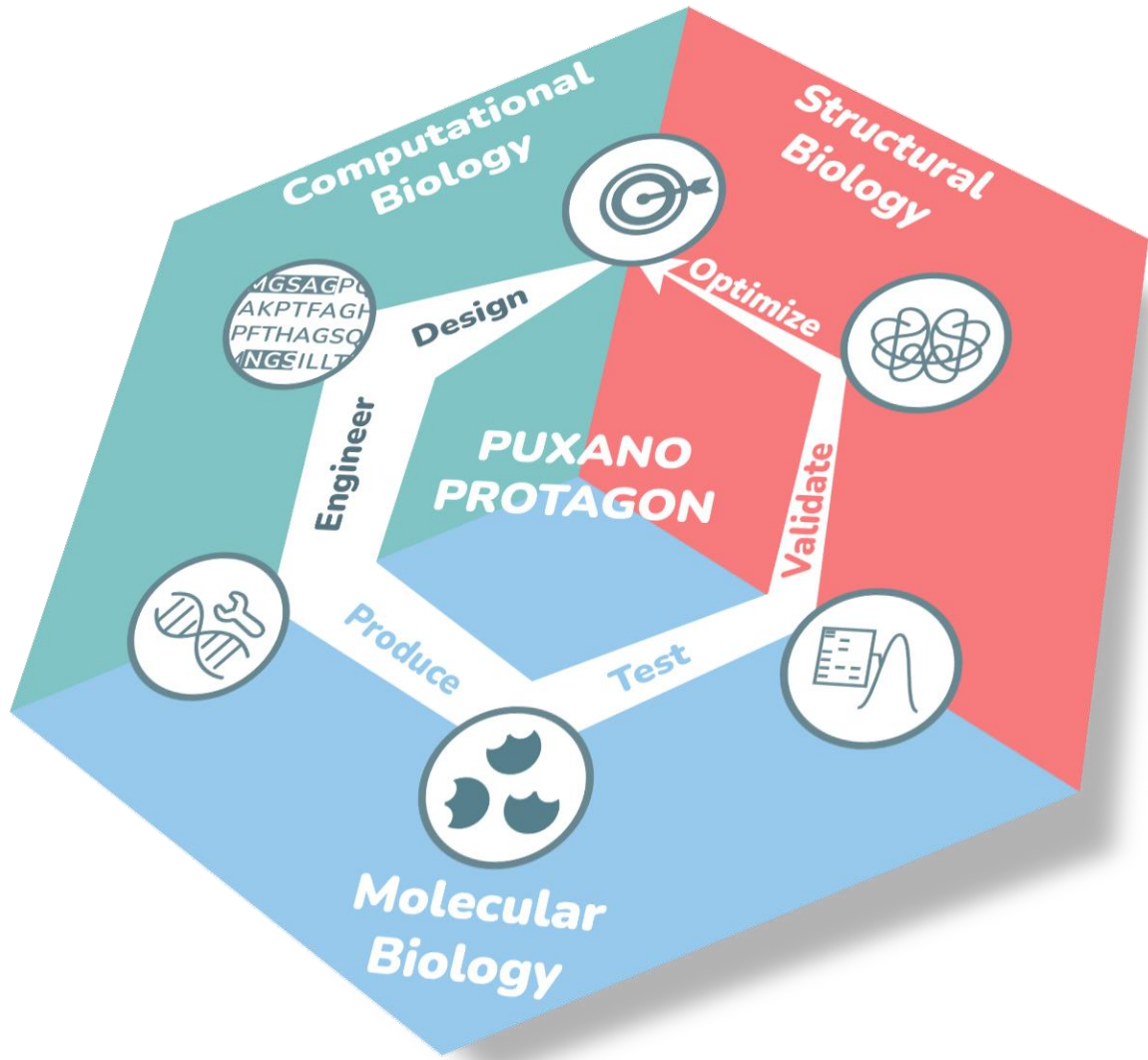
PUXANO functions as a platform-based CRO, focusing on:

- Protein engineering and design
- Structure-based protein research

Dedicated to the development of:

- protein-based therapeutics,
- vaccines
- crop protection agents or biostimulants





# PROTAGON

by PUXANO

- Computational:
  - Structure-guided protein engineering
  - Diffusion-based protein design
- Experimental
  - HT screening for protein expression, purification and biophysical metrics
- Structural:
  - cryoEM-based protein structure determination using the PXN Grid portfolio.

# WHY DO PROTEIN STRUCTURES MATTER?

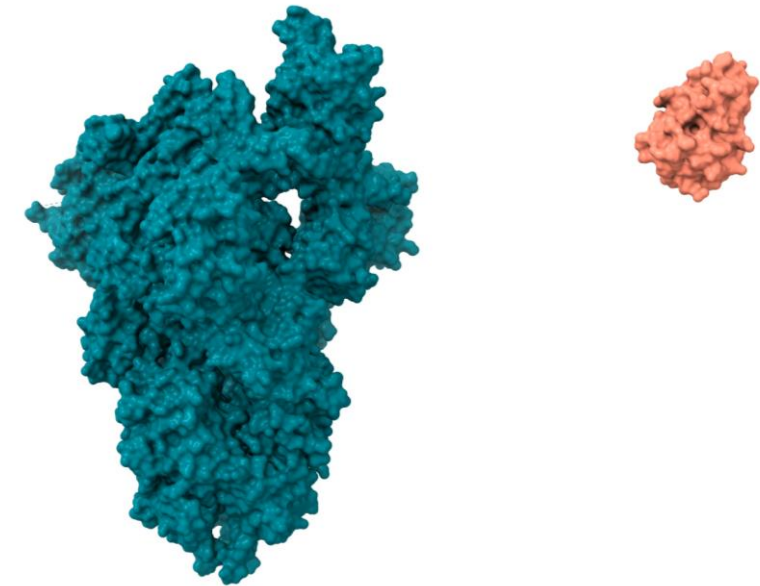


POWDER

[...]  
PLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPF  
GEVFNATRFASVYAWNKRISNCVADYSVLYNSASFSTFKC  
YGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIADY  
NYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRFRKSNLK  
PFERDISTEYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG  
YQPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGL  
TGTGVLTESNKKFLPFQFGRDIADTTDAVRDPQTLEILDITP  
CSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLT  
PTWRVYSTGNSVFQTRAGCLIGAEHVNNSYECDIPIGAGICA  
SYQTQTNsprarsvasqsiiaytmslgaensvaysnnsiaip  
TNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSF  
CTQLNRALTGIAV [...]

QVQLVESGGGLM<sup>Q</sup>AGGSLR<sup>L</sup>LSCAVS<sup>G</sup>RTF<sup>S</sup>TAA<sup>M</sup>GWFRQ  
APGKEREFVAAIRW<sup>S</sup>GGSA<sup>Y</sup>ADSVKGR<sup>F</sup>TISRDKAKNTVYL  
QMNSLKYEDTAVYYCASYQA<sup>T</sup>RSLLSDYATWPYDYWGQGT  
QVTVSSKHHHHHH

AMINO ACID SEQUENCE



STRUCTURE

# EXPERIMENTAL PROTEIN STRUCTURE DETERMINATION

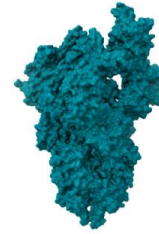
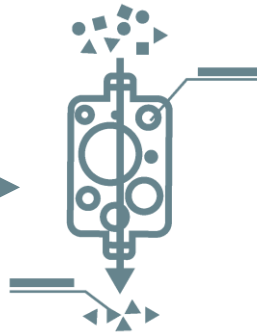
PROTEIN CONSTRUCT  
DESIGN & GENE SYNTHESIS



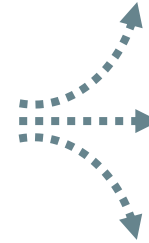
PROTEIN  
EXPRESSION



PROTEIN  
PURIFICATION



X-RAY CRYSTALLOGRAPHY



NMR

SINGLE PARTICLE CRYO-EM

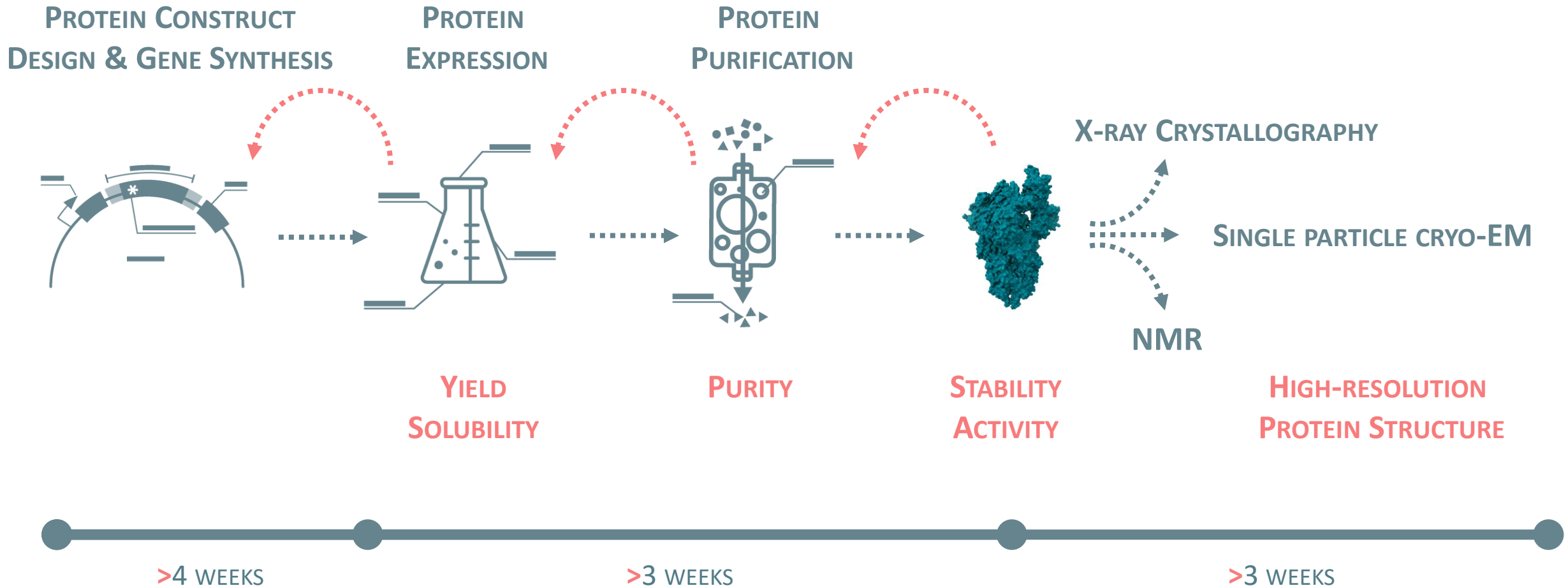


4 WEEKS

3 WEEKS

4 WEEKS

# TRIAL-AND-ERROR



# AI-TOOLS FOR EXPERIMENTAL STRUCTURAL BIOLOGY

## DE NOVO PROTEIN STRUCTURE PREDICTION

[...]  
PLSETKCTLKSFTVEKGIYQTSNFRVQPTESI  
VRFPNITNLCPFGEVFNATRFASVYAWNRK  
RISNCVADYSVLYNSASFSTFKCYGVSPKTL  
NDLCFTNVYADSFVIRGDEVROQIAPGQTGKI  
ADYNYKLPDDFTGCVIAWNSNNLDSKVGG  
NUNYLYRLFRKSNLKPFFERDIST ■ QAGSTP  
CNGVEGFNC ■ SYGFQPTNGVGYQPYPYR  
VVVLSFELLHAPATVCGPKKSTNLVKNKCV  
NFNFNGLTGTGVLTESNKKFLPFQQFGRDIA  
DTTDAVRDPQTLEILDITPCSFGGVSVITPGT  
NTSNQVAVLYQDVNCTEVPVAIHADQLTPT  
WRVYSTGSNVVFQTRAGCLIGAEHVNNSYEC  
DIPIGAGICASYQTQTNsprrarsvasqsIIA  
YTMSLGAENSVAYSNNsIAIPTNFTISVTTEIL  
PVSMTKTSVDCTMYICGDSTECsnlllQYGS  
FCTQLNRALTGIIV [...]

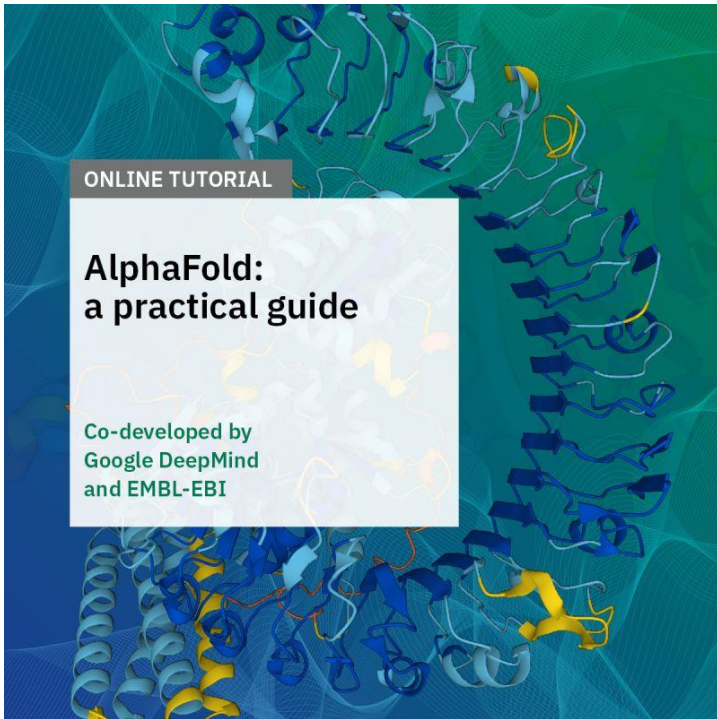
PART 1

PART 2

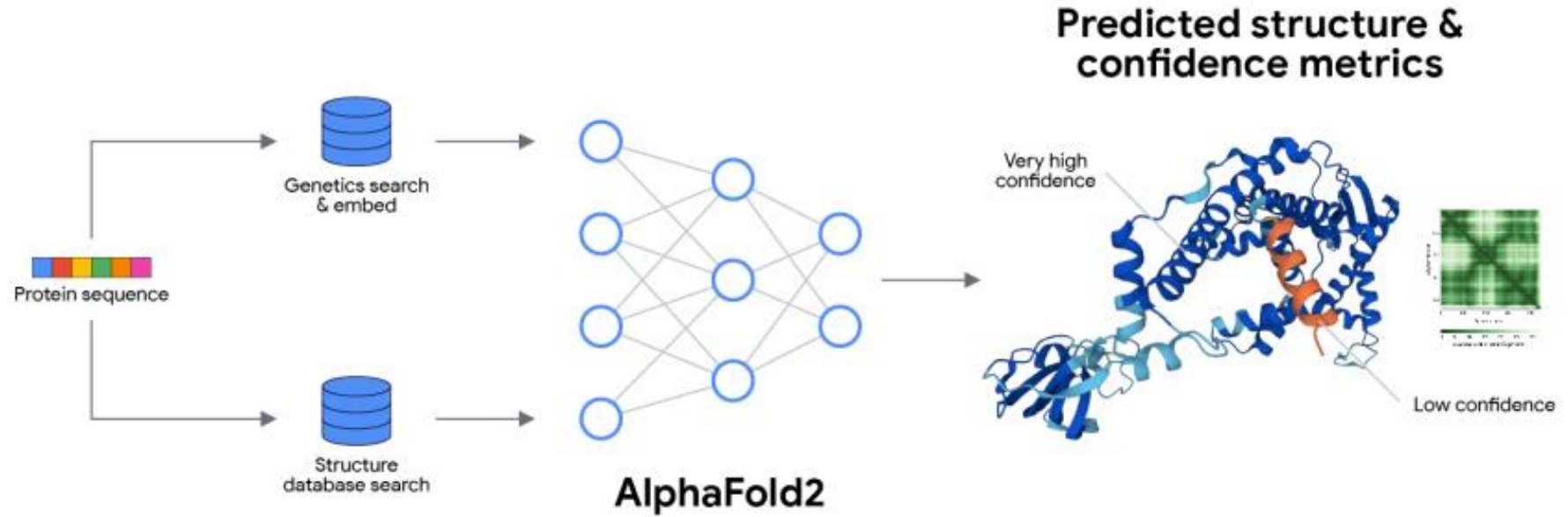


DE NOVO PROTEIN DESIGN

# PART 1: PROTEIN STRUCTURE PREDICTION



<https://www.ebi.ac.uk/training/online/courses/alphafold/>

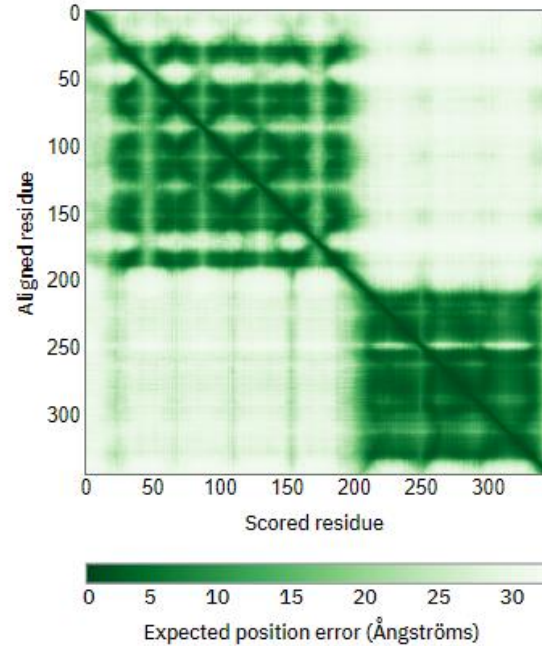
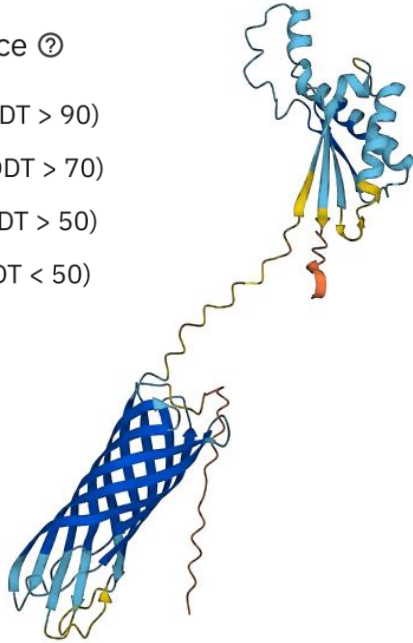




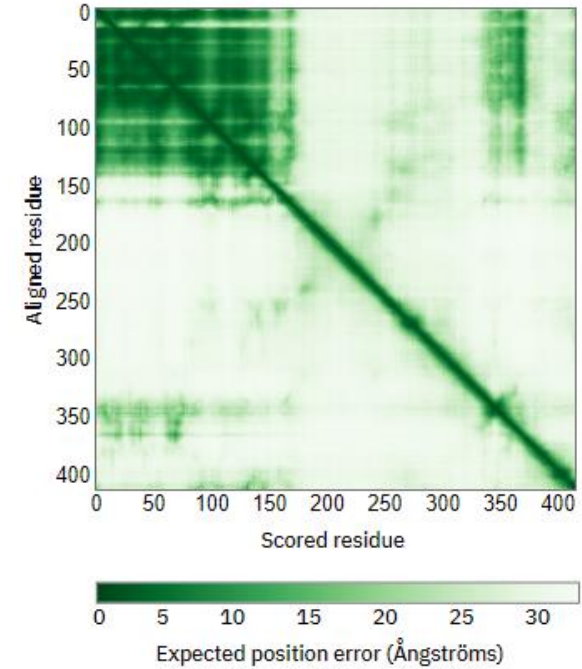
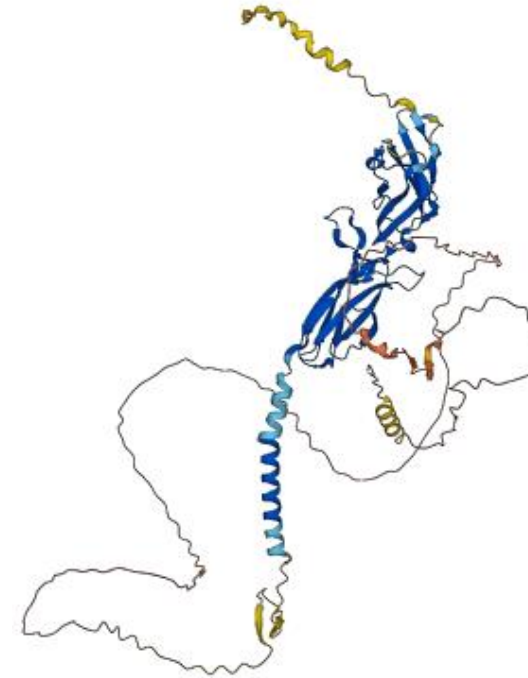
# ALPHAFOLD PREDICTION CASE STUDIES

Model Confidence ☺

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



**Outer membrane protein A**

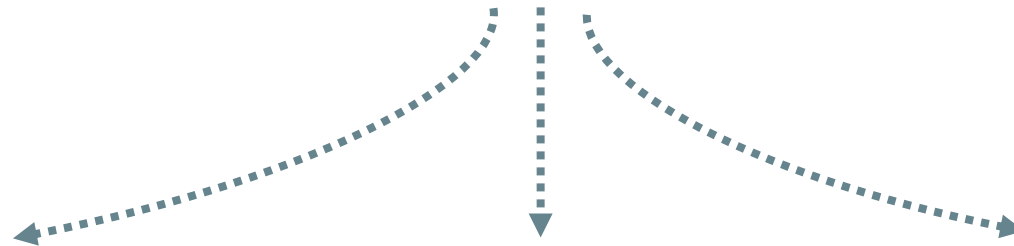


**Interleukin-20 receptor subunit alpha**

Limitations: MSA  $x < 30$  seqs, AA length, mutations, conformation, etc.

# ALPHAFOLD AS A TOOL

## ALPHAFOLD



### TARGET SELECTION

USE PREDICTION QUALITY AS A DATAPOINT TO SELECT TARGETS, NEXT TO DISEASE DATA, METABOLIC PATHWAYS, PATENTS, ETC.

### PROTEIN CONSTRUCT DESIGN

DELINEATE INTERESTING REGIONS, IDENTIFY IDRS, IDENTIFY DOMAINS, ETC.

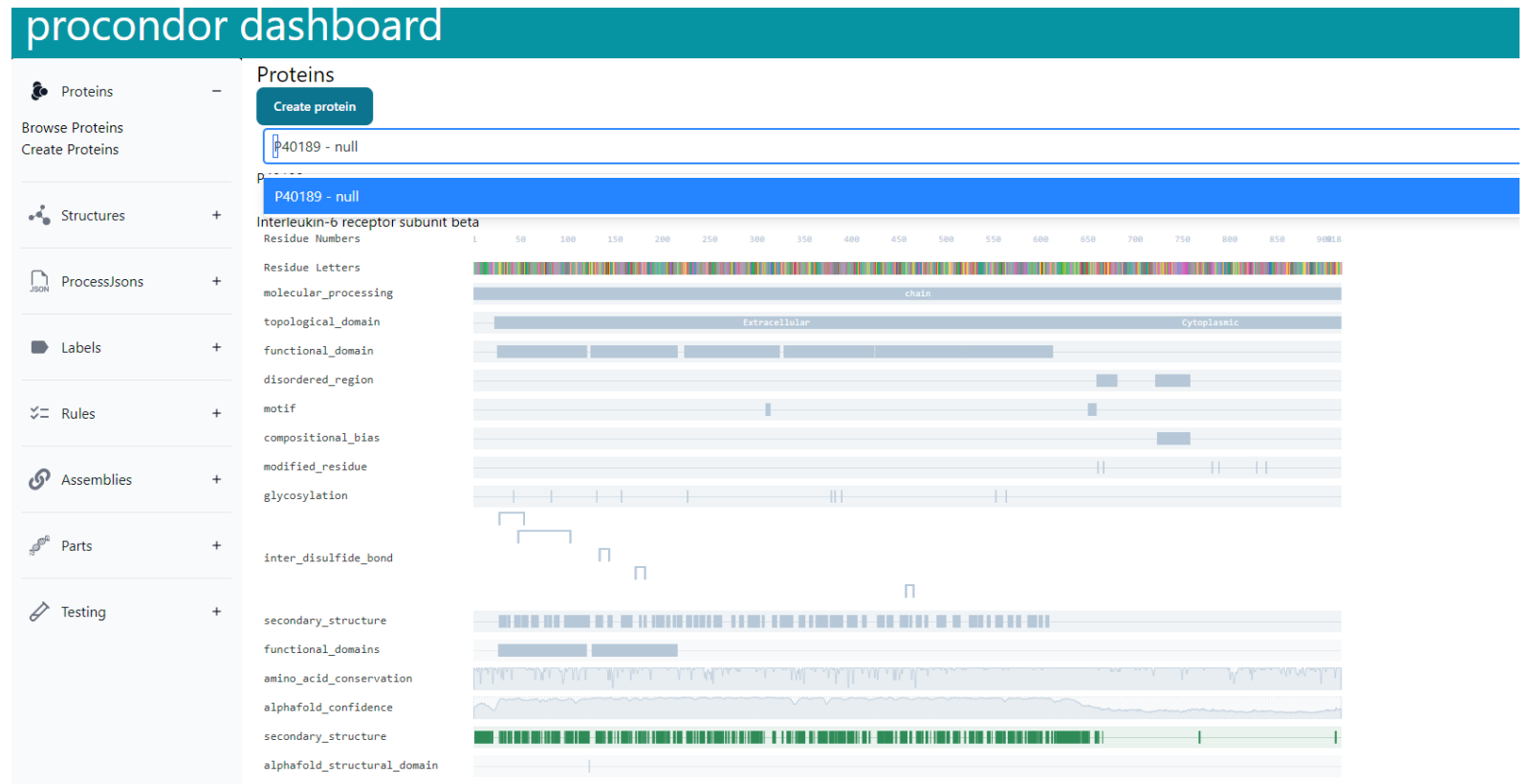
### PROTEIN STRUCTURE DETERMINATION

FIT AF MODEL IN EXPERIMENTAL MAP + OPTIMIZE (PHENIC/ISOLDE), COMPARE MODEL PROJECTIONS WITH EXPERIMENTAL DATA, ETC.

# ALPHAFOLD AS A TOOL FOR CONSTRUCT DESIGN

## PROTEIN CONSTRUCT DESIGN

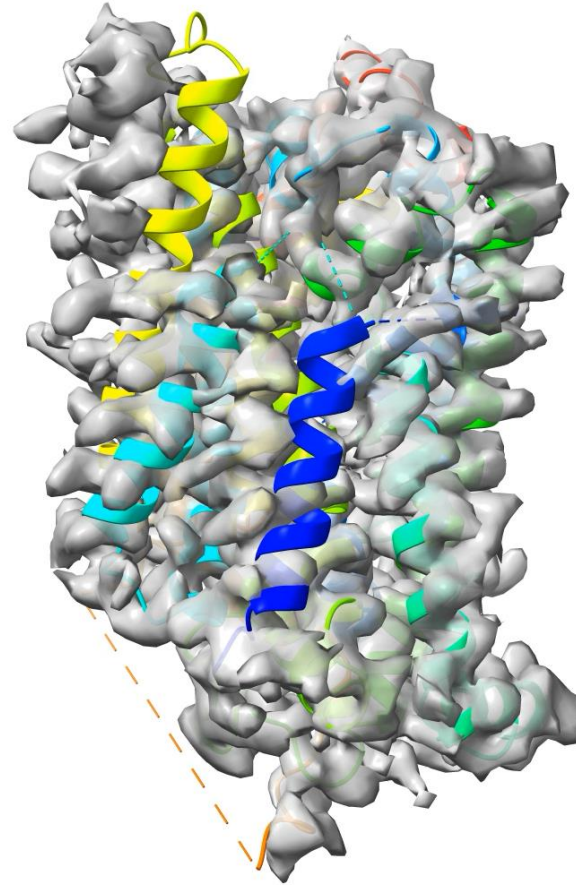
DELINEATE INTERESTING REGIONS,  
IDENTIFY IDRs,  
IDENTIFY DOMAINS,  
ETC.



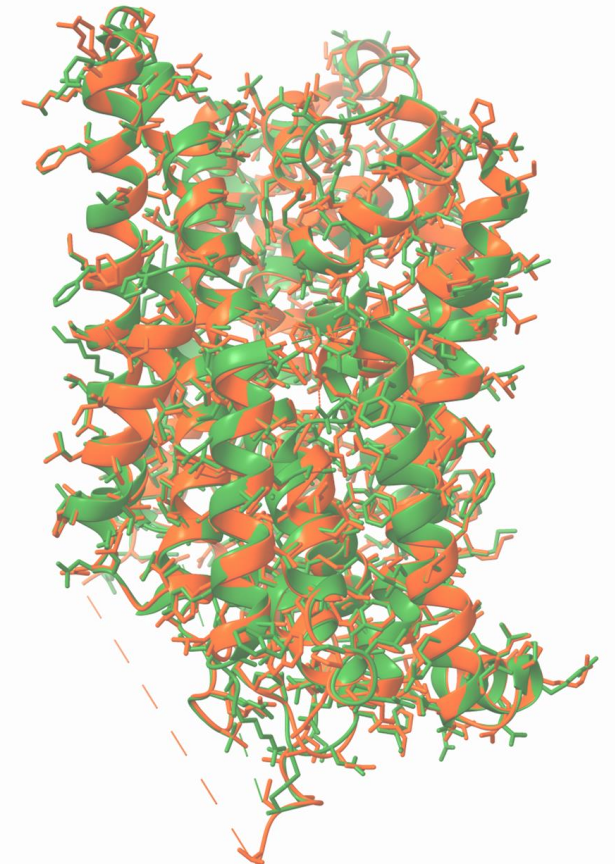
# ALPHAFOLD AS A TOOL FOR STRUCTURE DETERMINATION

## PROTEIN STRUCTURE DETERMINATION

FIT AF MODEL IN EXPERIMENTAL  
MAP + OPTIMIZE  
(PHENIX/ISOLDE), COMPARE  
MODEL PROJECTIONS WITH  
EXPERIMENTAL DATA, ETC.



Predicted structure after Isolde  
Refined PDB structure



Ravera et al. 2022 (Nature) - EMD: 26806, PDB: 7UUY

# OPTIMIZING AF2

- Parameter-based
  - Recycling
  - Seeds
- MSA
  - Db
  - DeepMSA2
- Template
- Validation
  - Other: ColabFold, OpenFold, ...
  - Single sequence: OmegaFold, ESMFold, SoloSeq
  - Whole-genome (DNA): Evo

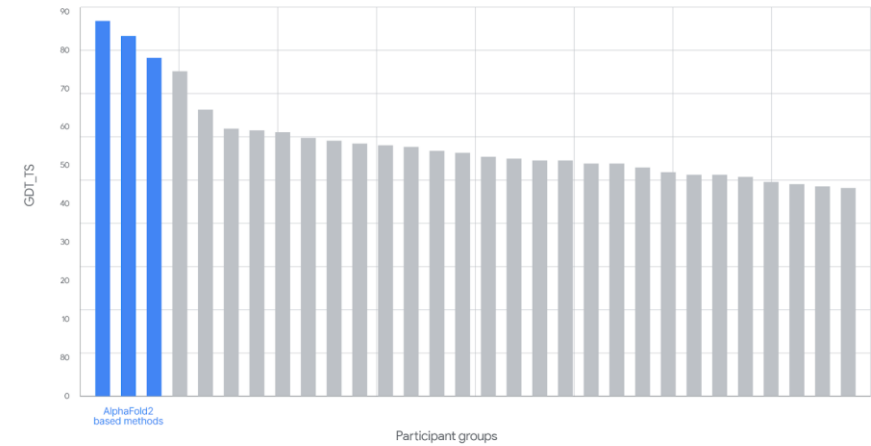
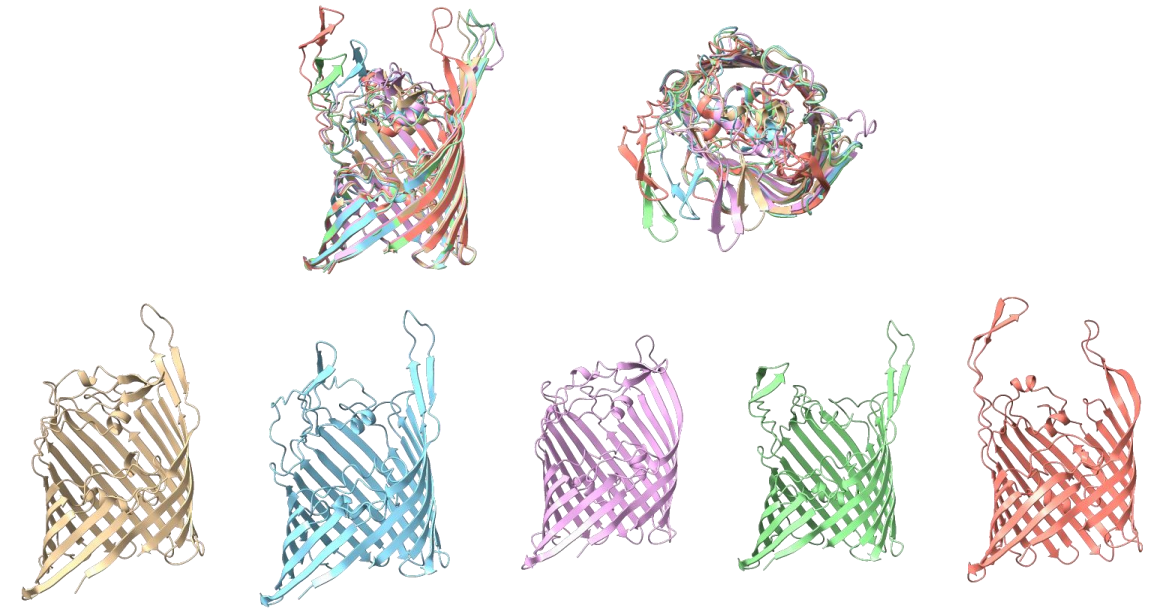
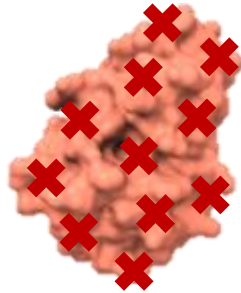


Figure 9. The highest-scoring entries in CASP15 in 2022. All top performers used some version of AlphaFold2 in their predictions.

# PART 2: PROTEIN DESIGN AND ENGINEERING

```
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE
```

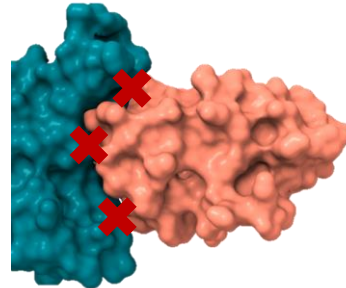
Many irrelevant sequences



Random or Sequence Based  
Mutagenesis

```
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE  
RLSCAVSGRTFSTAAMGWFRQAPGKERE
```

Less irrelevant sequences



Structure-Guided  
Protein Engineering

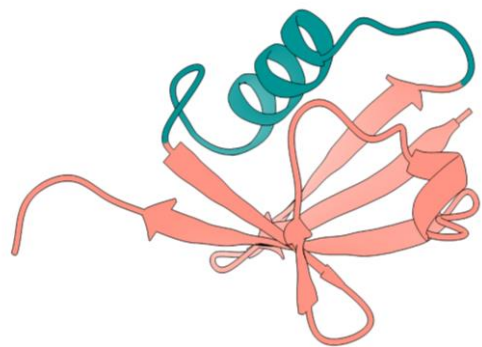
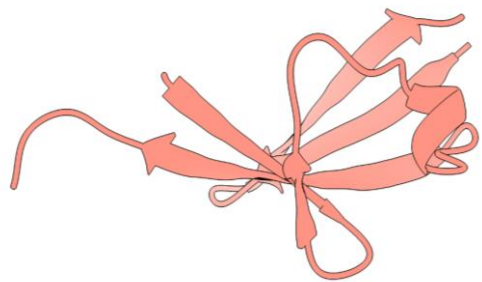
```
RSTTDGHEIKLMNPQRSTVWYCLDEFGED  
ACKKEFGHACDEFGHEIKLMNPQRSTVWY  
LLFRFGHEIKLMNPQRSTVWYADDEFGCL  
YTWCCEFKLMNPQRSTVWYCDDEFHEIKL  
MNASCDEFPPQRSTVWYACDEFGHEIKLAC  
CDEFGHEIKLACDEFGHEIKLPQRSTVWAD  
EFGHEIKLMNPQRSTVWYACDEFGHHALC  
DEFEIKLMNPQRSTVWYACDDEFGHEIKLC  
DEFMNYACDEFGHEIKLMNPQRSTVWALC  
DEFYACDEFGHEIKLMNPQRSTVWYACDYLL
```

More relevant sequences



Diffusion Generative AI Based  
Protein design

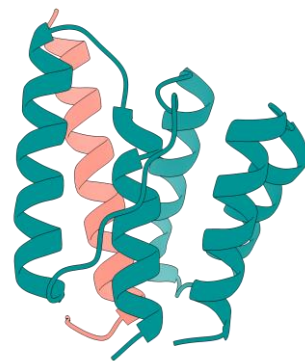
# DIFFUSION DESIGN



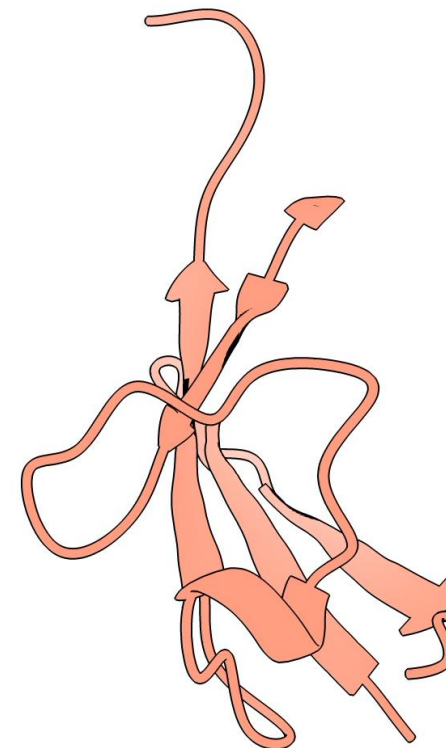
INPAINTING



SCAFFOLDING

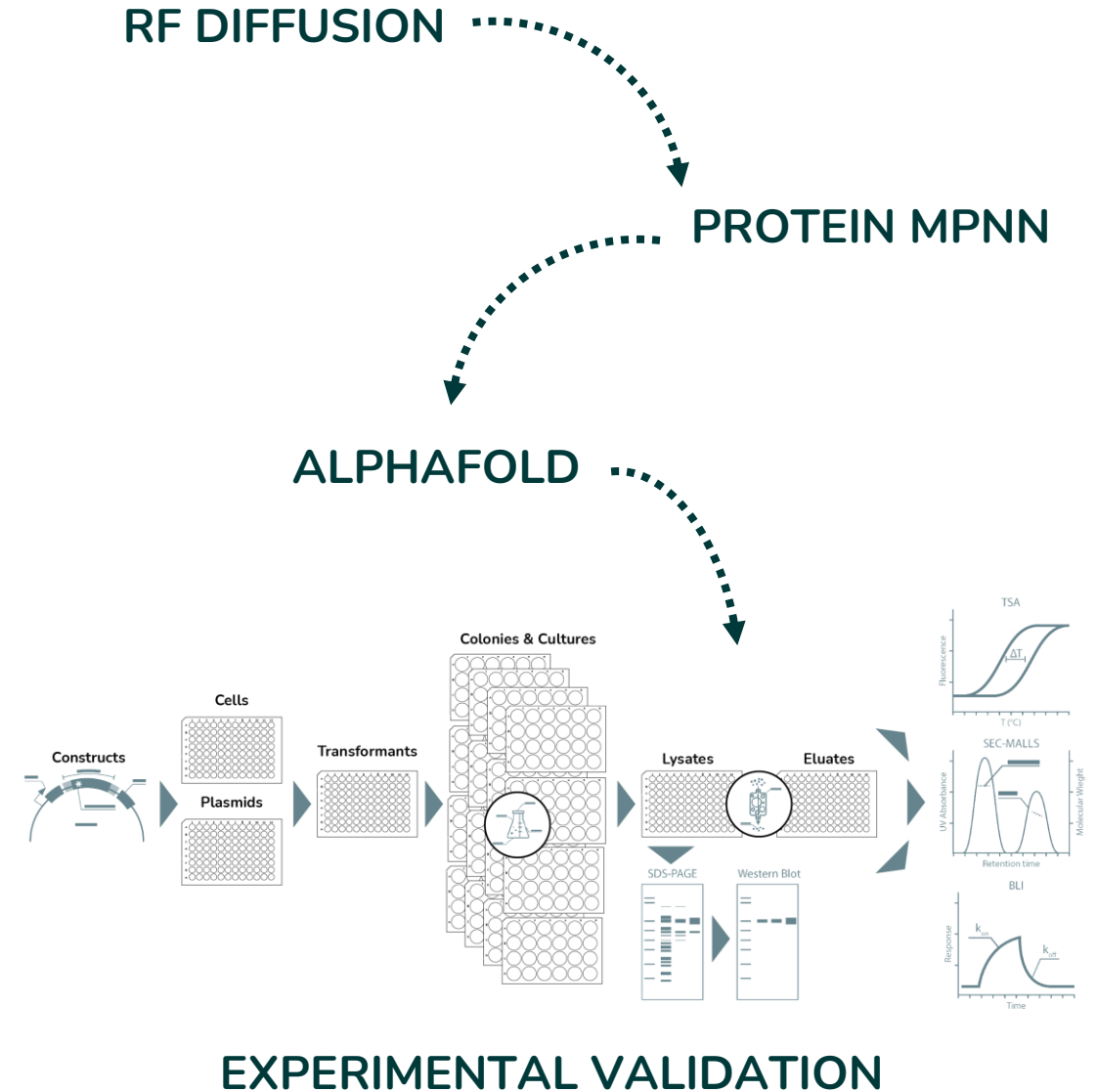


BINDER DESIGN



# RF DIFFUSION PIPELINE

- Diffusion model (backbone)
  - RF Diffusion, Genie, Chroma ...
  - Task specific models
- ProteinMPNN (AA seq)
  - Task specific models
- In silico validation tools
  - Designability (AF2)
  - Diversity
- Experimental validation

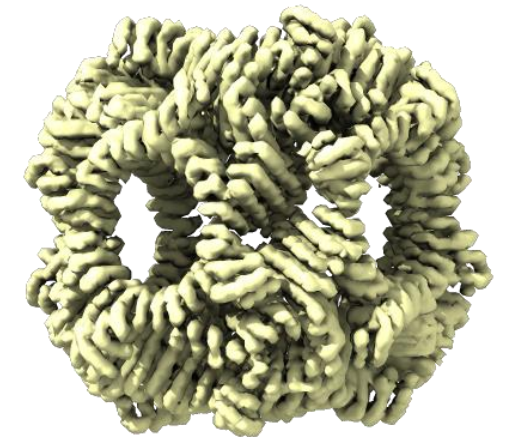
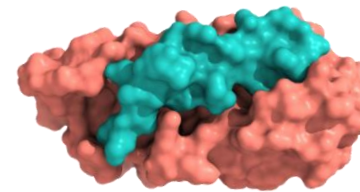
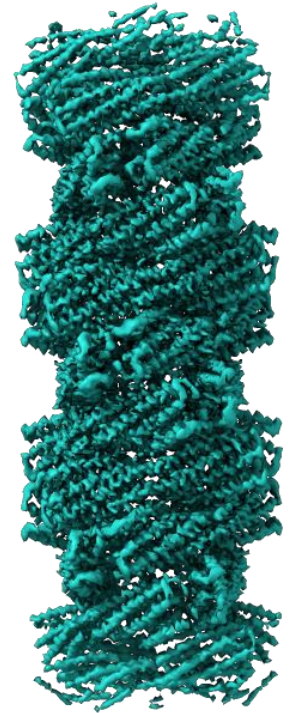
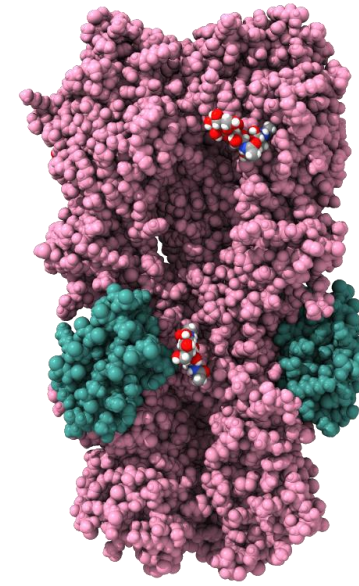


Watson et al. 2023;  
Dauparas et al. 2022



# NEW PROTEIN DESIGN APPLICATIONS

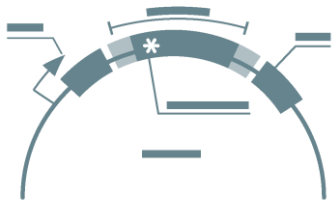
- Binders (VHH, mini, peptide, ...)
- Symmetry
- Enzyme
- Building blocks
- pH-sensors
- Soluble membrane proteins
- ...



Shen et al. 2024; Huddy et al. 2024;  
Torres et al. 2023; Watson et al. 2023

# INTEGRATING AI INTO EXPERIMENTAL STRUCTURAL BIOLOGY

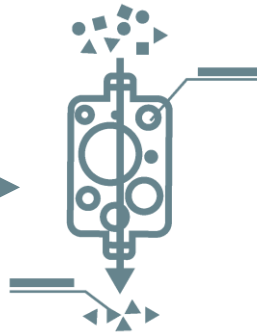
PROTEIN CONSTRUCT  
DESIGN & GENE SYNTHESIS



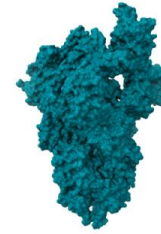
PROTEIN  
EXPRESSION



PROTEIN  
PURIFICATION



X-RAY CRYSTALLOGRAPHY



SINGLE PARTICLE CRYO-EM

NMR

DE NOVO PROTEIN STRUCTURE PREDICTION AND DESIGN

4 WEEKS

3 WEEKS

4 WEEKS

# AI-DRIVEN PROTEIN RESEARCH AND ETHICS

## De Novo Prediction

AlphaFold2	Deepmind/Google	UK/US
ESMFold	Facebook	US
RosettaFold	Univ. Washington – Baker	US
OmegFold	Helixon	CN
OpenFold	Columbia Univ. - AlQuraishi	US

## De Novo Design

RFDiffusion	Univ. Washington – Baker	UK/US
Evodiff	Microsoft	US
DreamFold	Dreamfold	UK
Chroma	Generate Biomedicines	US
Genie	Columbia Univ. - AlQuraishi	US

- Potential misuse (biosafety and biosecurity)
  - Toxins
  - Viruses
- Disruptions to the natural environment
  - Release of altered organisms
- IP
- Access Inequality
  - Architecture dev vs retraining vs user
  - Gradations in open source: usage vs training
  - Active players : US/CN/UK, industry vs academics vs tech companies
  - carbon and water footprint

# REFERENCES

- AlphaFold
  - <https://www.ebi.ac.uk/training/online/courses/alphafold/>
  - <https://alphafold.ebi.ac.uk/>
  - <https://www.nature.com/articles/s41586-023-06415-8>
  - <https://www.nature.com/articles/s41592-023-02087-4>
  - <https://deepmind.google/technologies/alphafold/>
- Diffusion-based protein design
  - <https://www.nature.com/articles/s41586-023-06415-8>
  - <https://link.springer.com/article/10.1038/s41565-024-01641-1>
  - <https://www.nature.com/articles/s41586-024-07188-4>
  - <https://www.nature.com/articles/s41586-023-06953-1>
  - <https://www.science.org/doi/10.1126/science.add2187>
  - <https://www.bakerlab.org/publications/>
- Reviews
  - <https://www.nature.com/nbt/volumes/42/issues/2>
  - [https://www.cell.com/cell/fulltext/S0092-8674\(23\)01402-2](https://www.cell.com/cell/fulltext/S0092-8674(23)01402-2)
- Ethics and protein design
  - <https://www.science.org/doi/10.1126/science.ado1671>
  - Supplementary: <https://www.biorxiv.org/content/10.1101/2024.02.27.582234v2>

# PUXANO

Wouter Van Putte, CEO and Founder of PUXANO

[info@puxano.com](mailto:info@puxano.com)

[www.puxano.com](http://www.puxano.com)

