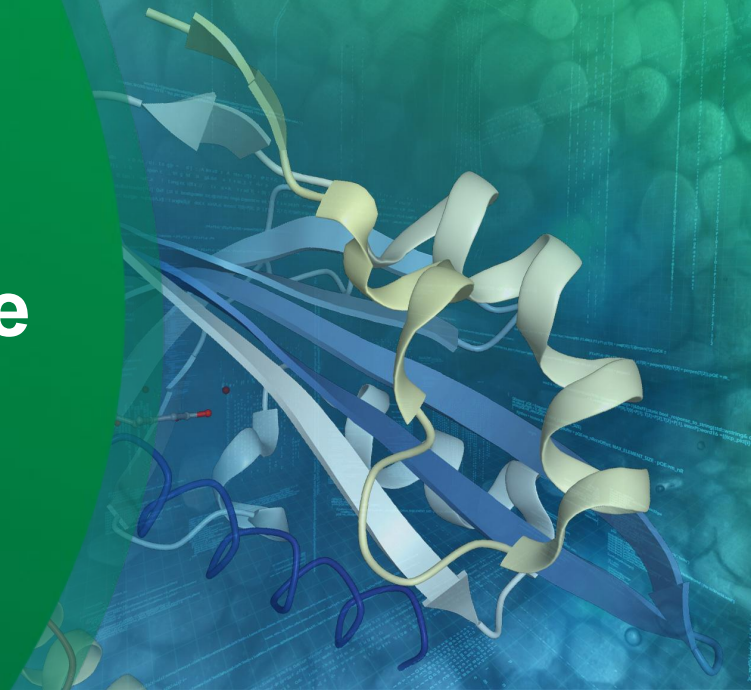
AlphaFold Protein Structure Database & 3D-Beacons

Navigating the flood of predicted protein structures

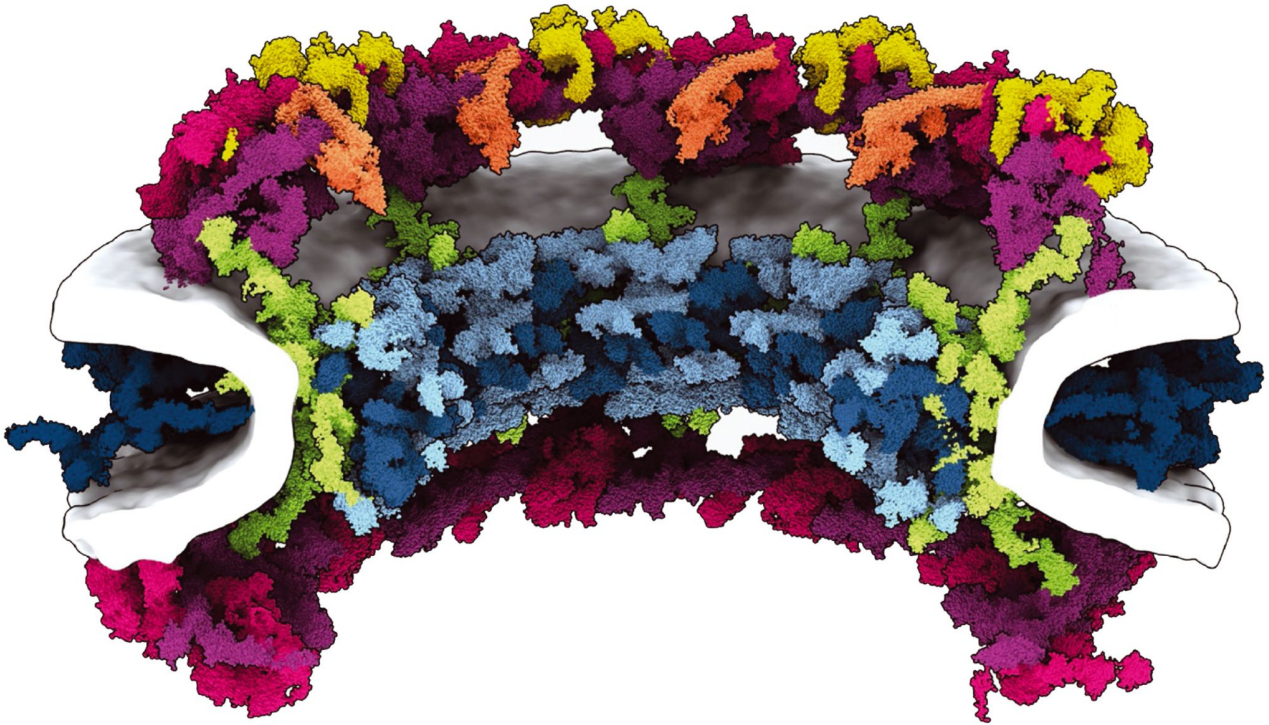


Mihaly Varadi, PhD
PDBe Coordinator

12 October 2023

EMBL-EBI

Predicted structures can help EM-based structure determination



A 70-MDa model of the human nuclear pore complex scaffold architecture.[1]

[1] Mosalaganti,S. et al. (2022) AI-based structure prediction empowers integrative structural analysis of human nuclear pores. Science, 376

Main messages

- The amount of experimentally determined proteins is limited (~200,000 PDB entries)
- There are over 900 million **protein structure predictions** available with their **confidence metrics** from databases like AlphaFold DB and ESM Metagenomic Atlas
  - These models have **limitations**
  - It is critical that researchers **use the confidence metrics** together with the predicted atomic coordinates
- AlphaFold DB is not the only data resource for predicted protein structures
  - The **3D-Beacons Network** allows researchers to retrieve models from other repositories

ToC

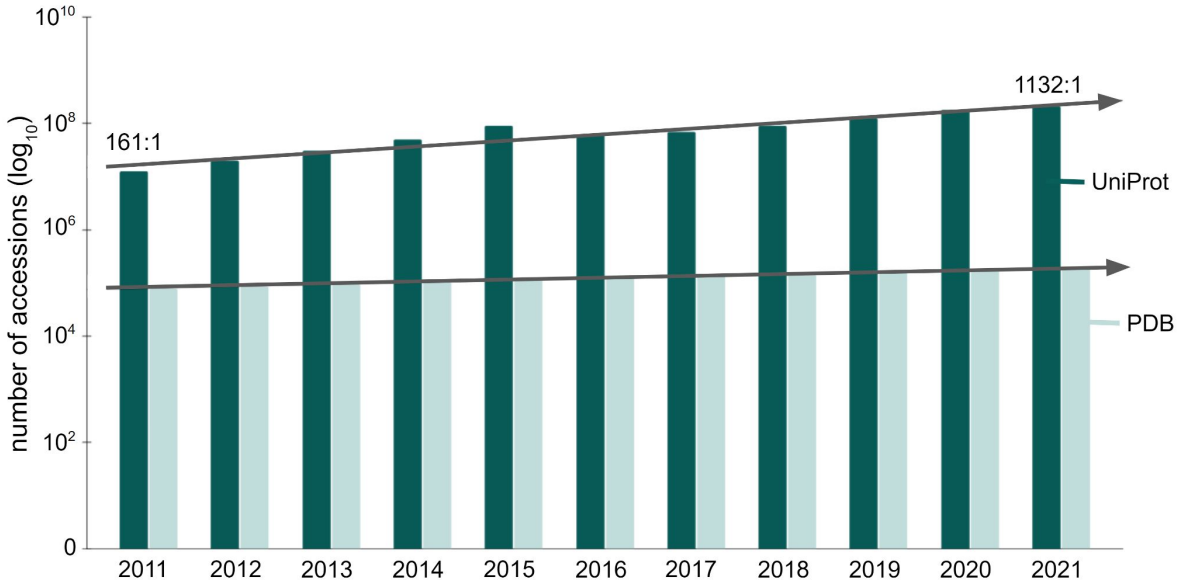
- Why do we need predicted protein structures?
- Briefly introducing AlphaFold
- How to access AlphaFold predictions

EMBL-EBI

Inferring protein function from 3D structures

- The structure-function paradigm states that the 3-dimensional **structure** of a protein **determines its function**
- Structural biology and structural bioinformatics are key components of the toolbox to gain functional insights for proteins
  - Structure determination
  - Bioinformatics analysis studies
  - Scientific software that derives functional annotations from protein structures
- However, the **relatively small number of experimentally determined protein structures** is a major limitation

The gap between sequences and structures



Growth of the UniProt and the PDB databases[1]

The UniProt database contains 252,170,925* protein sequences

The Protein Data Bank contains 210,342* entries, corresponding to 62,790* unique proteins

This gap grew 10x within the last decade

Predicted structures could help close the gap

[1] Varadi, M *et al*: 3D-Beacons: decreasing the gap between protein sequences and structures through a federated network of protein structure data resources; GigaScience, 2022 Nov 30

* Numbers are accurate as of 6th Oct 2023

Predicted protein structures to the rescue!

- Anfinsen's dogma postulated for over 50 years that the structure of (smaller, globular) proteins under standard physiological conditions **can be determined based only on its amino acid sequence**[2]
- For half a century, researchers and software developers created increasingly sophisticated algorithms and computational tools to tackle this grand challenge
- 2020 saw major breakthroughs in applying AI techniques to model protein structures

[2] Anfinsen, C B. Principles that Govern the Folding of Protein Chains. Science. 1973 Jul 20;181(4096):223-30.

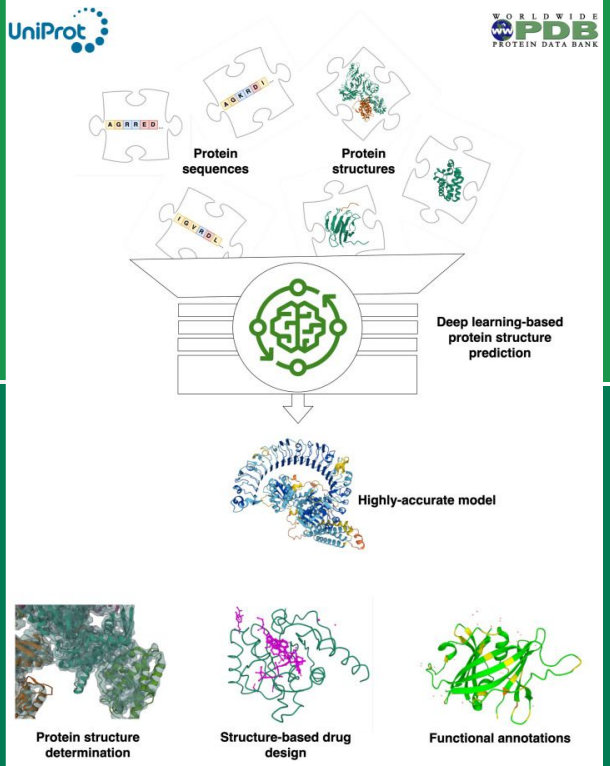
The most frequent structure modeling approaches

- Homology-based models
  - Homology-based techniques require a template structure of a reasonably close homologous sequence
- *Ab initio* models
  - These models do not require existing, similar protein structures
  - Often, these methods use force fields used in molecular dynamics
- Deep learning models
  - These models are a subclass of *ab initio* models, as they do not depend on templates
  - Several groups achieved outstanding performance in applying deep learning techniques for protein structure modeling
    - AlphaFold (https://www.alphafold.ebi.ac.uk/)
    - ESMFold (https://github.com/facebookresearch/esm)
    - RoseTTAFold (https://robetta.bakerlab.org/)
    - OpenFold (https://openfold.io/)

The impact of accurate predicted structures

The new generation of protein structure predictors took advantage of the **massive amounts of existing data**

Open source databases like **UniProt** and **PDB**, and the **decades of work by researchers** made these AI tools possible[3]



Now, these tools **give back to the scientific community**, and drive many breakthroughs

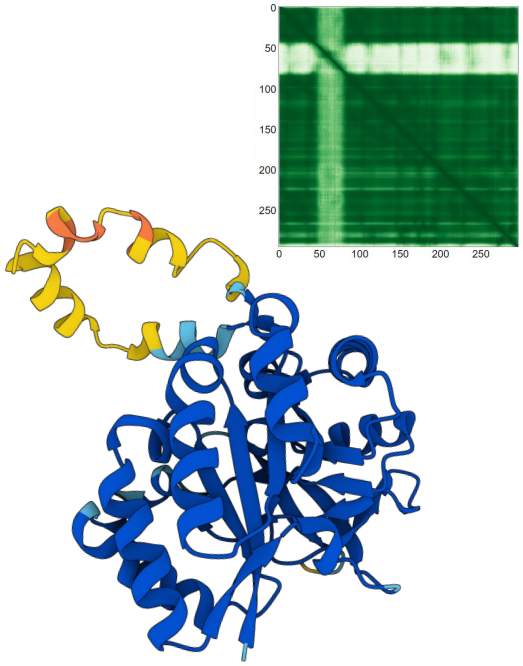
The key impacts are on **structure determination, drug discovery and functional analysis**

[3] Varadi, M *et al*; The opportunities and challenges posed by the new generation of deep learning-based protein structure predictors; Current Opinion in Structural Biology, 2023 Feb 18

Fighting antimicrobial resistance

Bacteria are becoming resistant to many of our antibiotics

Bacteria can evade antibiotics by deploying enzymes which modify the membrane so that drug molecules cannot bind



Probable 4-deoxy-4-formamido-L-arabinose-phosphoundecaprenol deformylase ArnD
https://alphafold.ebi.ac.uk/entry/P76472

AlphaFold and other new generation prediction tools excel at **predicting membrane protein structures**

Researchers can develop drugs against the enzymes' structures

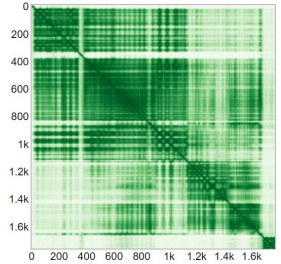
Source: https://unfolded.deepmind.com/stories/accelerating-the-race-against-antibiotic-resistance

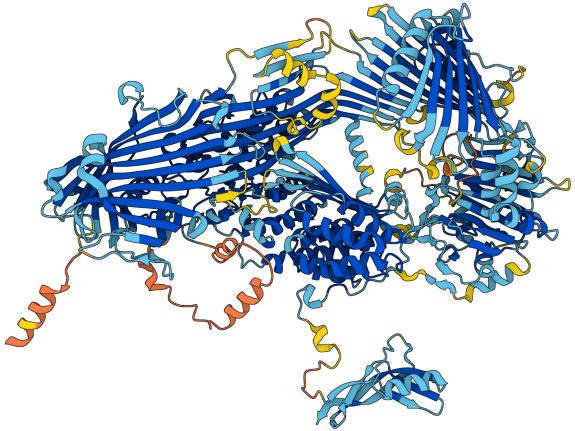
EMBL-EBI

Understanding disease resistance in bees

Honeybee populations are on a very disturbing decline globally

To increase their chances of survival, we need to study the fundamentals of the bee immune system





Vitellogenin
https://alphafold.ebi.ac.uk/entry/Q868N5

Vitellogenin is a key protein which can bind to pathogen proteins like an antibody and contributes to immunity that is passed on

Until AlphaFold become available, there were no reliable structures of bee vitellogenin due to the size of the protein

Source: https://unfolded.deepmind.com/stories/how-the-honeybee-could-help-protect-species-around-the-world

EMBL-EBI

ToC

- Why do we need predicted protein structures?
- Briefly introducing AlphaFold
- How to access AlphaFold predictions

AlphaFold protein structure predictor

- AlphaFold (2.0) is an AI-based tool developed by Google DeepMind
- AlphaFold can predict protein structures **based only on amino acid sequences**
- AlphaFold produces **three independent outputs**:
  - Predicted atomic coordinates
  - Confidence measures per amino acids (pLDDT)
  - Confidence measure per residue pairs (Predicted Aligned Error, PAE)

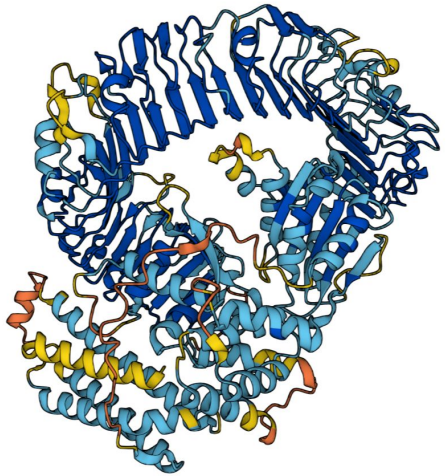
Predicted atomic coordinates

- AlphaFold creates atomic coordinates in two data formats
  - PDB format
    - This is the legacy format of the Protein Data Bank (PDB)
    - It is **obsolete** and **will not be supported for new PDB entries** from the end of 2023
  - mmCIF format
    - This data format is the official format of the PDB archive
- The mmCIF format contains much more information than the PDB format
- Both formats contain the atomic coordinates in a very similar way

```
ATOM 1    N N   . MET A 1 1   ? 5.963   65.398  -32.038 1.0 39.15 ? 1   MET A N   1 Q5VSL9 UNP 1   M
ATOM 2    C CA  . MET A 1 1   ? 6.670   66.225  -31.038 1.0 39.15 ? 1   MET A CA  1 Q5VSL9 UNP 1   M
ATOM 3    C C   . MET A 1 1   ? 6.626   65.441  -29.741 1.0 39.15 ? 1   MET A C   1 Q5VSL9 UNP 1   M
ATOM 4    C CB  . MET A 1 1   ? 6.024   67.613  -30.879 1.0 39.15 ? 1   MET A CB  1 Q5VSL9 UNP 1   M
ATOM 5    O O   . MET A 1 1   ? 5.532   65.058  -29.366 1.0 39.15 ? 1   MET A O   1 Q5VSL9 UNP 1   M
ATOM 6    C CG  . MET A 1 1   ? 6.645   68.653  -31.821 1.0 39.15 ? 1   MET A CG  1 Q5VSL9 UNP 1   M
```

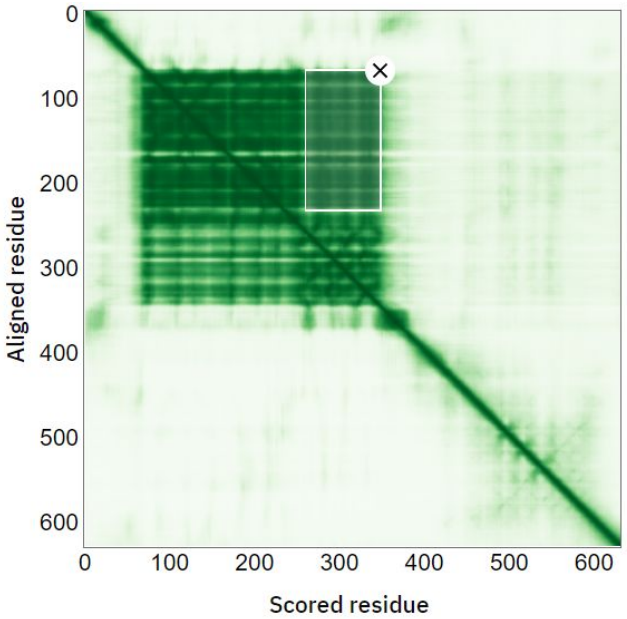
Confidence metrics: pLDDT

- One of the independent outputs of AlphaFold is the amino acid-level confidence measure called pLDDT
- It is a **local accuracy metric**
- We colour-code residues based on their pLDDT scores

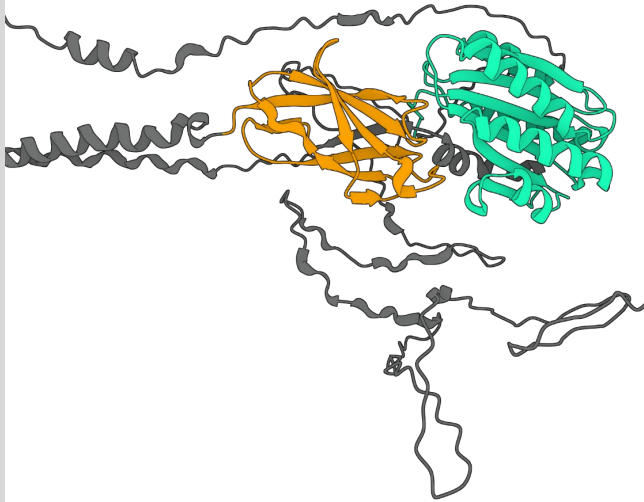


Probable disease resistance protein At1g58602
https://www.alphafold.ebi.ac.uk/entry/Q8W3K0

Confident relative position

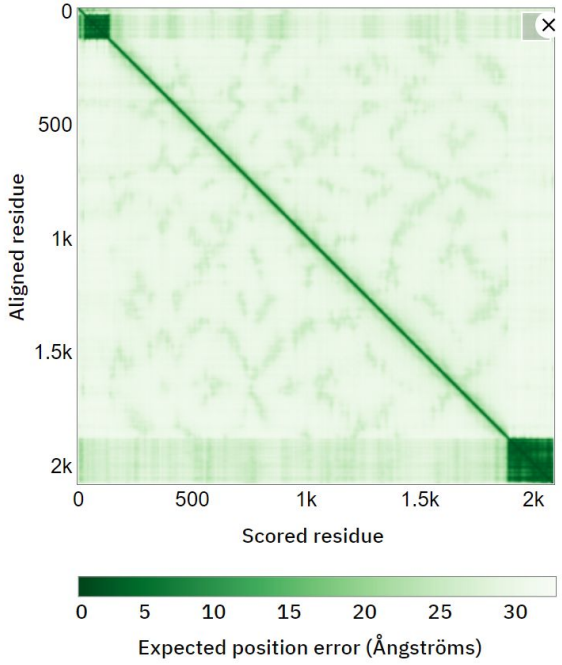


- Predicted Aligned Error, or PAE is an independent output of AlphaFold
- PAE is pairwise, i.e. it has a value for every residue pair
- It measures the **confidence in the relative position of two amino acids**
- We display the PAE scores using interactive heatmaps

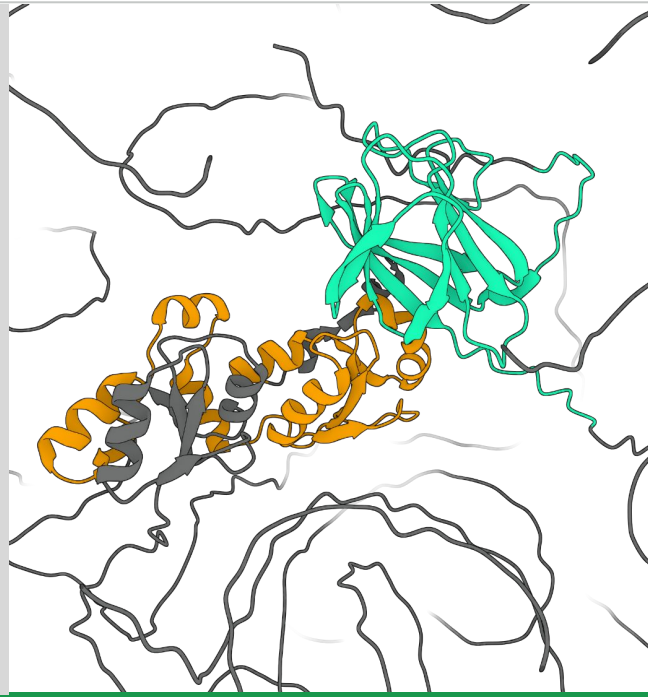


Anthrax toxin receptor-like
https://www.alphafold.ebi.ac.uk/entry/A6NF34

Not confident relative position



- Predicted Aligned Error, or PAE is an independent output of AlphaFold
- PAE is pairwise, i.e. it has a value for every residue pair
- It measures the **confidence in the relative position of two amino acids**
- We display the PAE scores using interactive heatmaps



Mediator of DNA damage checkpoint protein 1
https://www.alphafold.ebi.ac.uk/entry/Q14676

Limitations of AlphaFold

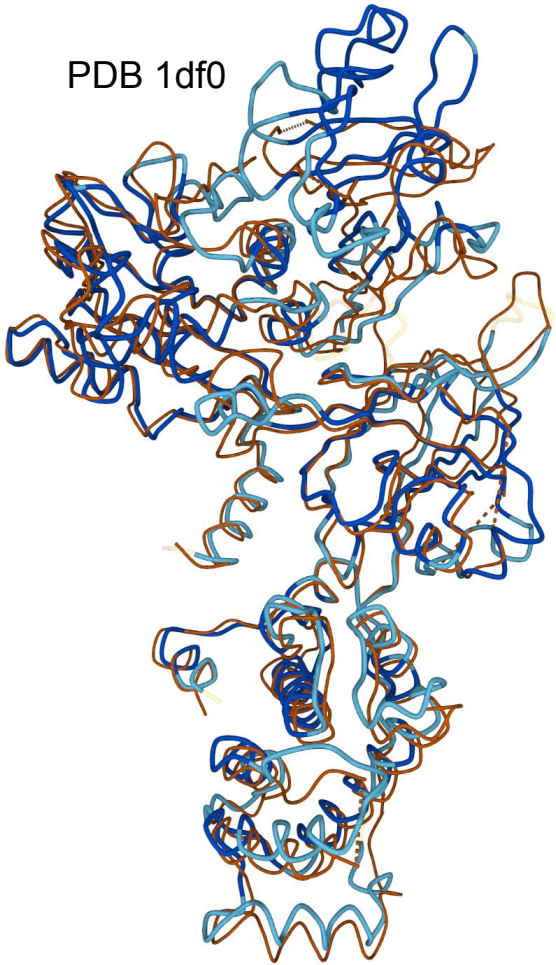
- AlphaFold was not trained to
  - Predict the effects of mutations
    - But Google DeepMind just published AlphaMissense to tackle this problem[4]
  - Predict the binding of ligand molecules
    - But others tools like AlphaFill[5] aim to add small molecules
  - Predict nucleic acid structures

[4] Cheng J. et al; Accurate proteome-wide missense variant effect prediction with AlphaMissense; Science; 2023 Sep 19
[5] Hekkelman M.L. et al; AlphaFill: enriching AlphaFold models with ligands and cofactors; Nature Methods; 2022 Nov 24

Limitations of AlphaFold (cont.)

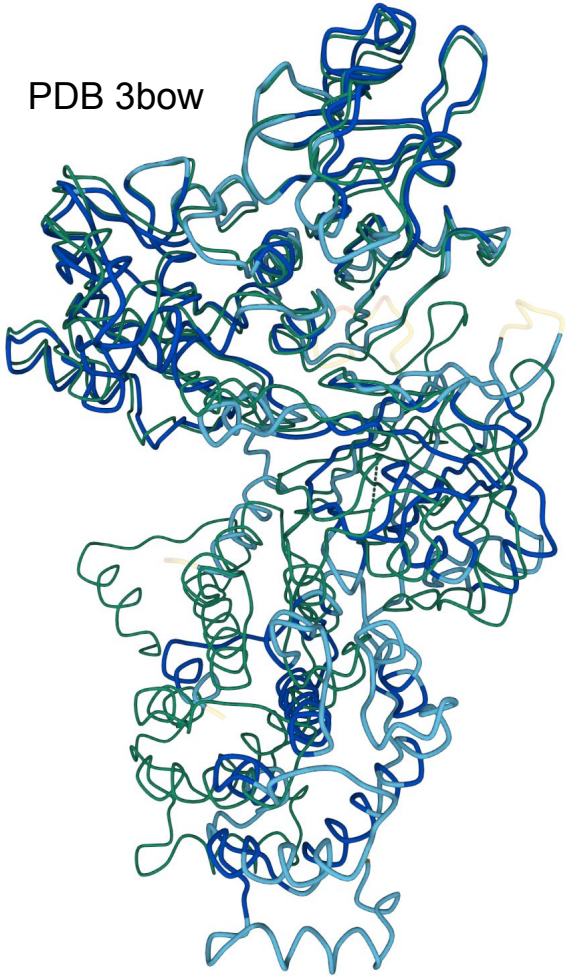
- AlphaFold only accepts the 20 standard amino acids in its input
- AlphaFold (by default) predicts five models per run
  - However, these models are generally very similar
  - In other words, AlphaFold usually cannot predict conformational variability in a protein
  - Researchers found ways to force AlphaFold to come up with alternative conformation by providing shallow, engineered multiple sequence alignments as input
- If a protein has multiple distinct, biologically relevant conformations, **you may not know which conformation AlphaFold predicts**

PDB 1df0

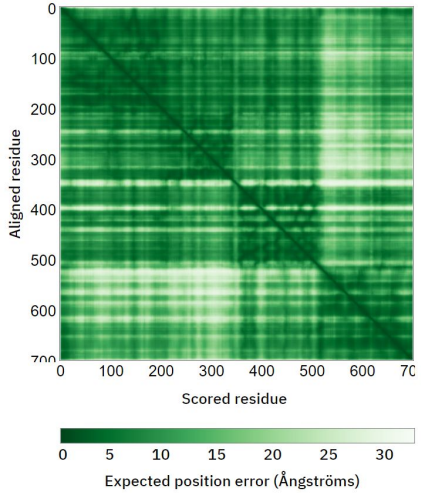


RMSD with inactive form: 2.84Å

PDB 3bow



RMSD with active form: 5.72Å

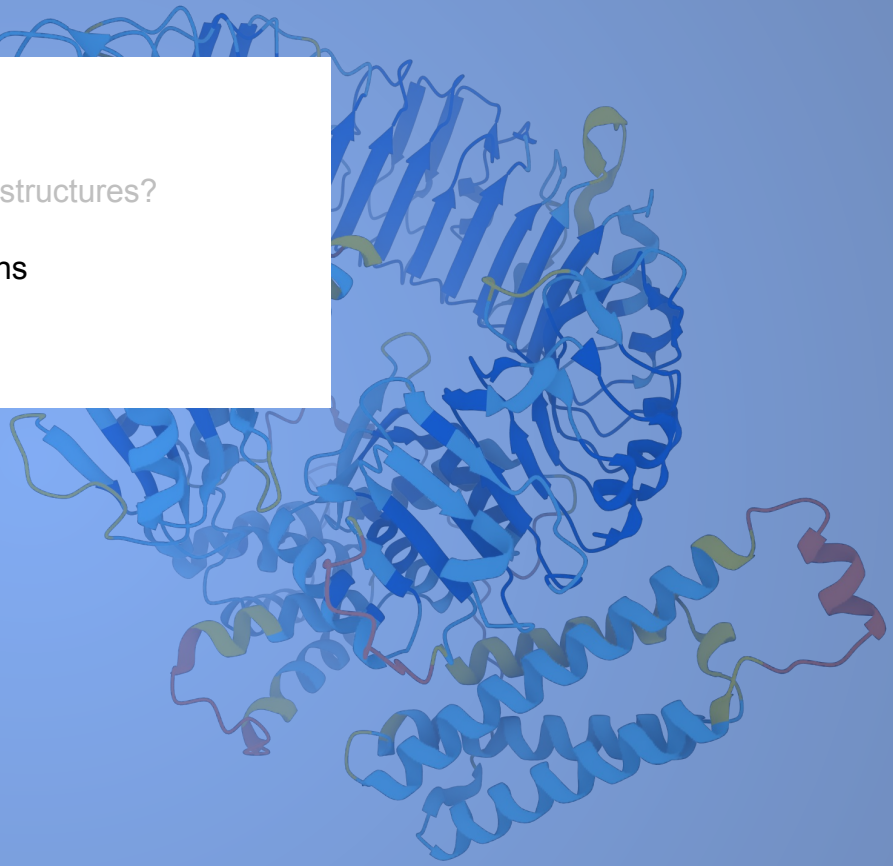


Source: https://pdbe-kb.org/proteins/Q07009
(Calpain)

EMBL-EBI

ToC

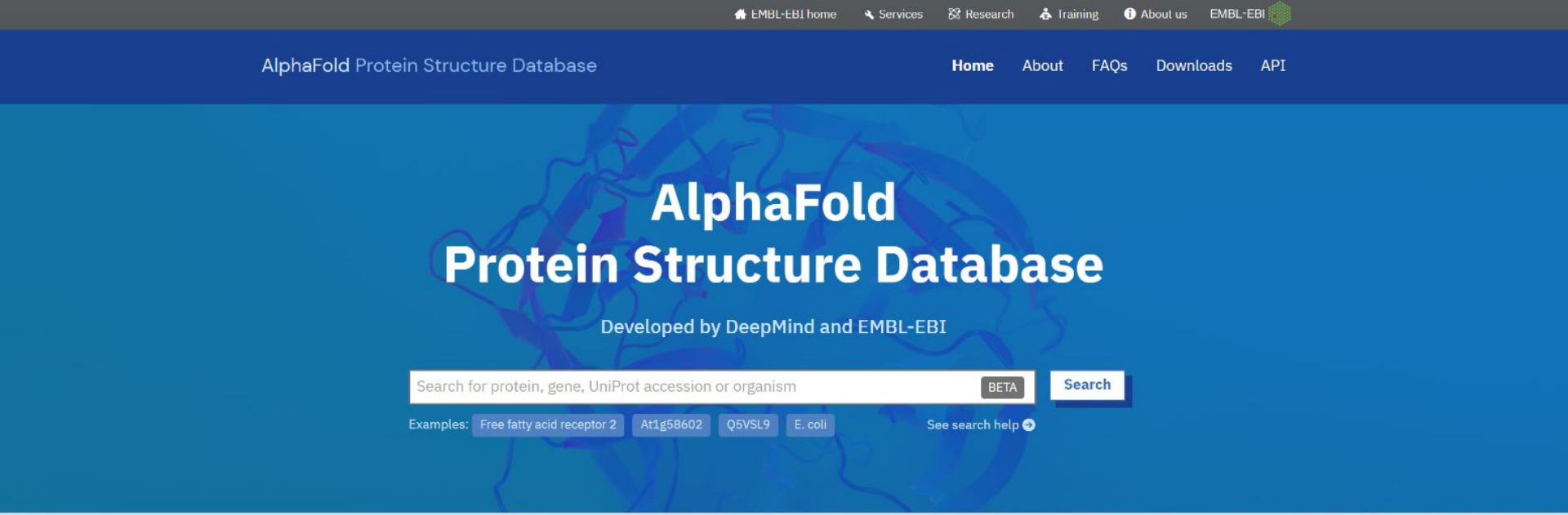
- Why do we need predicted protein structures?
- Briefly introducing AlphaFold
- How to access AlphaFold predictions



EMBL-EBI

How to use AlphaFold

- There are three main ways of accessing predicted protein structures from AlphaFold:
  - The **source code is publicly** accessible at https://github.com/deepmind/alphafold/
    - Using the source code gives total control over the process
    - Not for the faint of heart
  - **Interactive Google Colab** notebook: https://bit.ly/alphafoldcolab
    - More limited in terms of configuration
    - Easier to use, harder to break
    - Researchers can provide the input sequence(s)
  - For the **AlphaFold database** at https://www.alphafold.ebi.ac.uk/
    - These are pre-calculated models and confidence metrics
    - More than 214 million protein structures and 47 proteomes of key model organisms
    - The database does not run predictions based on input sequences



AlphaFold DB provides open access to over 200 million protein structure predictions to accelerate scientific research.

https://alphafold.ebi.ac.uk

EMBL-EBI

What types of data do we store?

- For 214,684,311 proteins:
  - Predicted atomic coordinates of **monomeric structures** (mmCIF and PDB formats)
  - Residue-wise confidence metrics (pLDDT)
  - Residue pairwise predicted aligned errors (PAE)
  - Metadata
    - Cross-references
    - Identifiers
    - Species name
    - Gene name
    - Biological function (from UniProt)

What do we *not yet* have in the database?

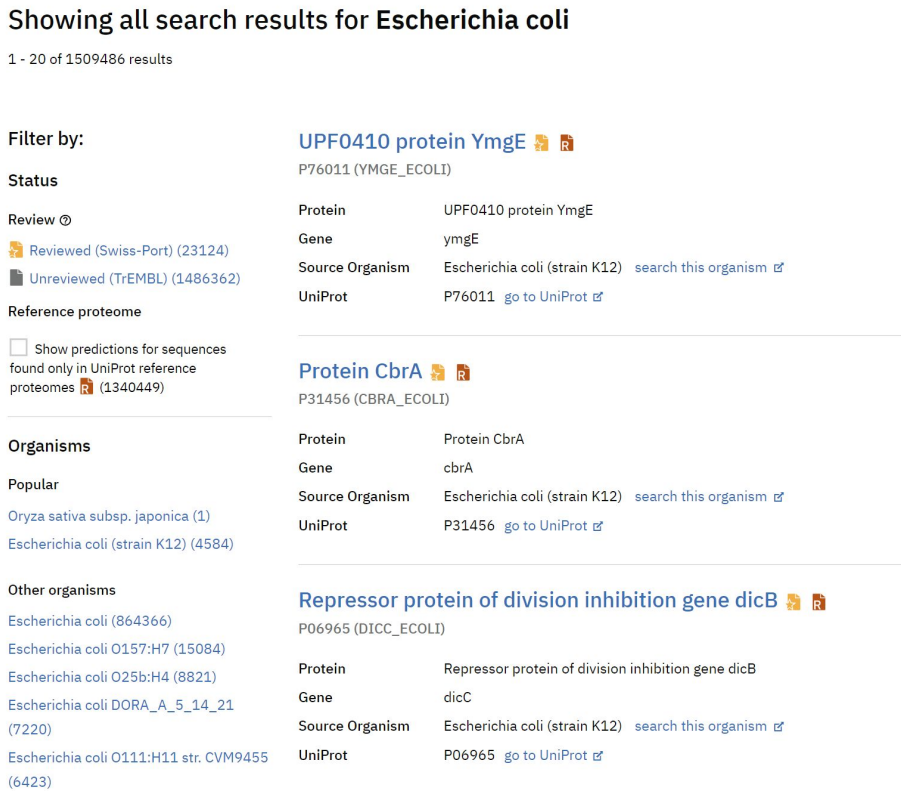
- There are no multiple conformations of protein structures
- There are no viral proteins
- There are no isoforms
- There are no assemblies
- There are no mutant structures
- There are no ligands

Searching in AlphaFold DB

AlphaFold DB can be searched by protein names, UniProt accessions, gene names and species

We have recently added filters for reviewed/unreviewed sequences

We continuously work on the search system to make the data more discoverable

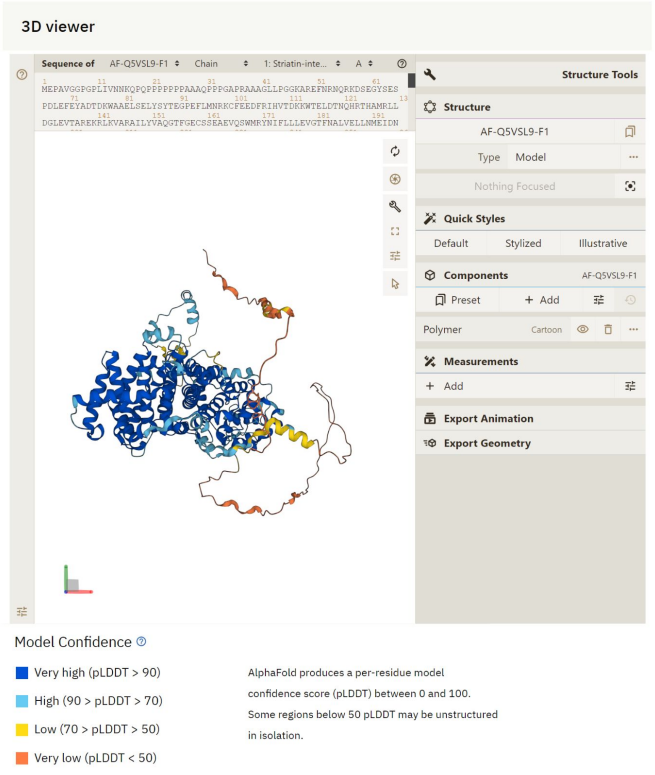


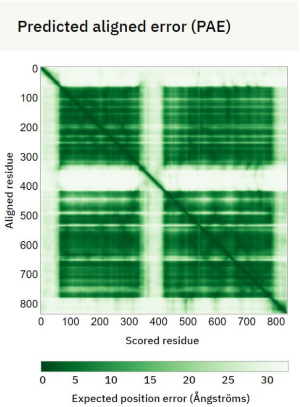
AlphaFold DB prediction pages

AlphaFold DB predictions have dedicated pages

These pages provide download options and 2D/3D data visualizations for structures and confidence metrics

We are expanding these pages to include protein similarity data and functional annotations





Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – go to Help section below for more information.

Data access via FTP

- The **complete proteomes available** through the public FTP (File Transfer Protocol) area of EMBL-EBI
  - http://ftp.ebi.ac.uk/pub/databases/alphafold/
- The FTP provides access to current and previous versions of the predictions
- The **README.txt** file provides detailed information on the types of data available
- The following additional data can be accessed through the FTP:
  - A list of every UniProt accession for which we have predicted structures
  - The amino acid sequences of every predicted structure in FASTA format
- Important limitation: **The PAE data files are not yet available via FTP**

Programmatic access

- The AlphaFold database has a **public API** for programmatic access
  - Documentation at https://www.alphafold.ebi.ac.uk/api-docs
- The API is well suited to retrieve data in bulk
  - For example, for a long list of UniProt accessions
- The API return the data in JSON (JavaScript Object Notation) format
  - For example https://www.alphafold.ebi.ac.uk/api/prediction/P07550

Programmatic access (example JSON)

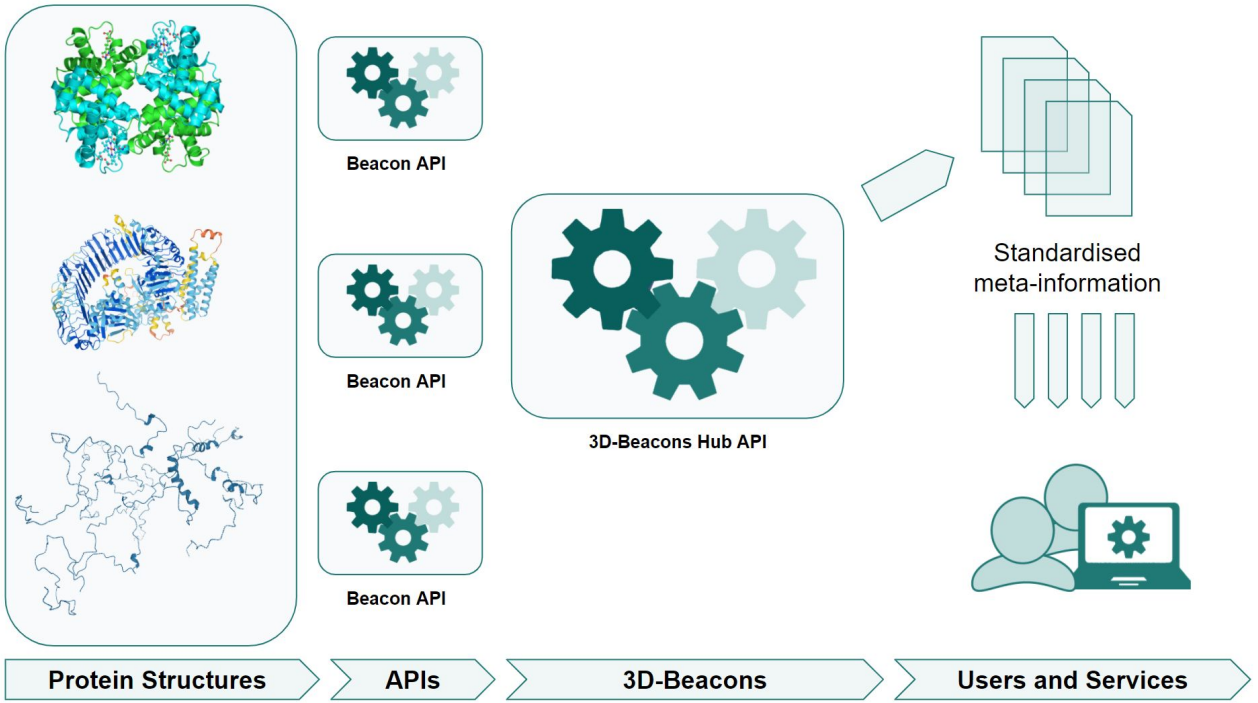

```
[
  {
    "entryId": "AF-P07550-F1",
    "gene": "ADRB2",
    "uniprotAccession": "P07550",
    "uniprotId": "ADRB2_HUMAN",
    "uniprotDescription": "Beta-2 adrenergic receptor",
    "taxId": 9606,
    "organismScientificName": "Homo sapiens",
    "uniprotStart": 1,
    "uniprotEnd": 413,
    "uniprotSequence": "MGQPGNGSAFLLAPNGSHAPDHDVTQERDEVWVVGMGIVMSLIVLAIVFGNVLVITAIAKFERLQTVTNYFITSLACAD
    IYCRSPDFRIAFQELLCLRRSSLKAYGNGYSSNGNTGEQSGYHVEQEKENKLLCEDLPGTEDFVGHQGTVPSDNIDSQGRNCSTNDSLL",
    "modelCreatedDate": "2021-07-01",
    "latestVersion": 2,
    "allVersions": [
      1,
      2
    ],
    "cifUrl": "https://alphafold.ebi.ac.uk/files/AF-P07550-F1-model_v2.cif",
    "bcifUrl": "https://alphafold.ebi.ac.uk/files/AF-P07550-F1-model_v2.bcif",
    "pdbUrl": "https://alphafold.ebi.ac.uk/files/AF-P07550-F1-model_v2.pdb",
    "paeImageUrl": "https://alphafold.ebi.ac.uk/files/AF-P07550-F1-predicted_aligned_error_v2.png",
    "paeDocUrl": "https://alphafold.ebi.ac.uk/files/AF-P07550-F1-predicted_aligned_error_v2.json"
  }
]
```

AlphaFold DB is a member of 3D-Beacons

The 3D-Beacons Network[6] provides standardised access to experimentally determined and predicted protein structures and their confidence metrics

It is an open collaboration between many model providers including AlphaFold DB, PDBe, Model Archive, AlphaFill, Swiss-Model and others

Users can find protein structures from any member provider based on UniProt accession or sequence search



[6] M. Varadi et al., 3D-Beacons: Decreasing the gap between protein sequences and structures through a federated network of protein structure data resources, BioRxiv, 2022 Jul

UniProt accession | Find structures by UniProt accession | Find

UniProt accession examples: P38398 P01308

P38398 (BRCA1_HUMAN) - 64 Structures available

Information

Protein	Breast cancer type 1 susceptibility protein Go to UniProt
Gene	BRCA1
Source organism	*Homo sapiens*
Biological function	E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys-6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA ... Show more

31
Experimentally Determined Structures

0
Conformational Ensembles

3
Template-based models

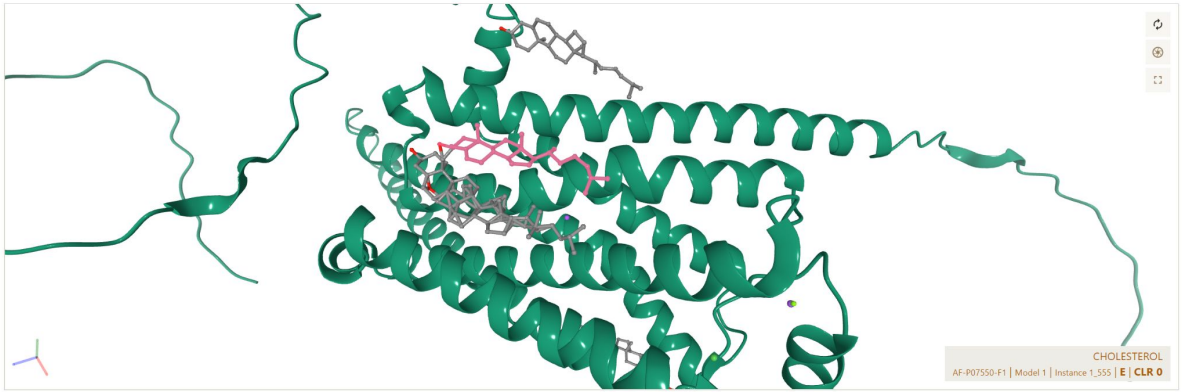
30
Ab-initio Models

AlphaFill, an example data provider

AlphaFill is a newer data resource who joined the 3D-Beacons Network

The main purpose of AlphaFill is to model obligate small molecule that we know should bind to a predicted AlphaFold model

AlphaFold cannot dock small molecules, even though the orientation of the side chains is often correct, as if the ligand was present



Source: https://www.ebi.ac.uk/pdbe/pdbe-kb/3dbeacons/search/P07550

EMBL-EBI

Main messages

- The amount of experimentally determined proteins is limited (~200,000 PDB entries)
- There are over 900 million **protein structure predictions** available along with their **confidence metrics** from databases like AlphaFold DB and ESM Metagenomic Atlas
  - These models have **limitations**
  - It is critical that researchers **use the confidence metrics** together with the predicted atomic coordinates
- AlphaFold DB is not the only data resource for predicted protein structures
  - The **3D-Beacons Network** allows researchers to retrieve models from other repositories

Thank you!


Damian Bertoni


Paulyna Magana

318,955,415
API calls


Ivanna Pidruchna

1,469,354
Total users

224
Countries*


Maxim Tsenkov

4,450,262
Total page views

948,194
File downloads


Urmila Paramval

20,912
Bulk archive downloads

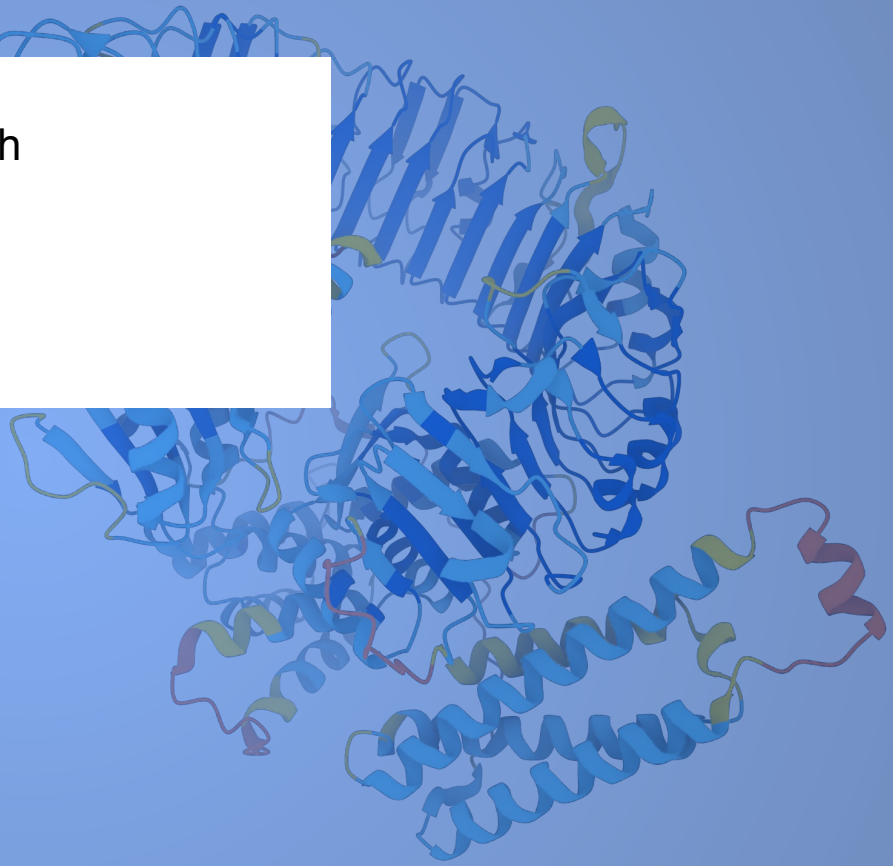

Sreenath Nair

* includes dependant and disputed territories

EMBL-EBI

Coming soon

- Sequence-based search
- Foldseek clusters
- Structure-based search



EMBL-EBI

Showing all search results for **Job ID [job id]**

1 - 20 of 869 results from BLASTP

Filter by:

Status

Review

Reviewed (Swiss-Prot) (18909)

Unreviewed (TrEMBL) (123402112)

Reference proteome

Show predictions for sequences found only in UniProt reference proteomes (79)

Organisms

Popular

Homo sapiens (149)

Mus musculus (73)

Sort by HSP score: highest first

Autism susceptibility gene 2 protein

Q8WXX7 (AUTS2_HUMAN)

Protein	Autism susceptibility gene 2 protein
Gene	AUTS2
Source organism	Homo sapiens go to search
UniProt	Q8WXX7 go to UniProt
Experimental structures	15 PDB structures for Q8WXX7 go to PDBe-KB

Sequence match **HSP score:** 100 **E-value:** 5.482e-231

Identity: 100% **Positives:** 281/300 (100%) **Gaps:** 281/300 (100%)

```
Your query      1  MPTVDDILEQVGESGWFQKQAFLILCLLSAAFAPICVGIVFLGFTPDH-HCQSPGVAELS  59
                   MPTVDDILEQVGESGWFQKQAFLILCLLSAAFAPICVGIVFLGFTPDH-HCQSPGVAELS
Q8WXX7          1  MPSFDEALQRVGEFGRFQRRVFLLLCLTGVTFAFLFVGVVFLGTQPDHYWCRGPSVAALA  60
```

Show full alignment

Structure similarity cluster

Predicted structures in the AlphaFold Protein Structure Database clustered using MMseq2 and Foldseek. This data is provided by the AFDB Clusters.

AFDB50/MMseqs2 (3221) | AFBD/Foldseek (45)

AlphaFold database protein sequences clustered by the MMseqs2 algorithm (Steinegger M. and Soeding J., Nat. Commun. 9, 2018). Each cluster is comprised of sequences that fulfil two criteria: maintaining a maximum sequence identity of 50% and achieving a 90% bi-directional sequence overlap with the longest sequence of the cluster representative.

1 selected: Download as

- PDB file
- mmCIF file
- Binary mmCIF file
- PAE (JSON) file
- FASTA file

Taxonomic filter

	A	Description	Species	Sequence length	Average pLDDT
☑	A	Organic Protein patched homolog 2 (inferred by orthology to a human protein) resistance protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	A	Resistance-nodulation-cell division superfamily	*Jatropha curcas*	174	77.00
☐	AF-Q3U5Q7-F1	Protein patched homolog 2	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	Protein patched homolog 2	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	SSD domain-containing protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	SSD domain-containing protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	SSD domain-containing protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	SSD domain-containing protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	SSD domain-containing protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00
☐	AF-Q3U5Q7-F1	SSD domain-containing protein	*Pyrobaculum calidifontis JCM 11548*	174	77.00

Results per page: 10 1-10 of 45 Previous page Next page

Similar structures

Structure search is powered by Foldseek and PDBe.

1985 hits in AFBD and PDB

1 selected: Download as ⌄

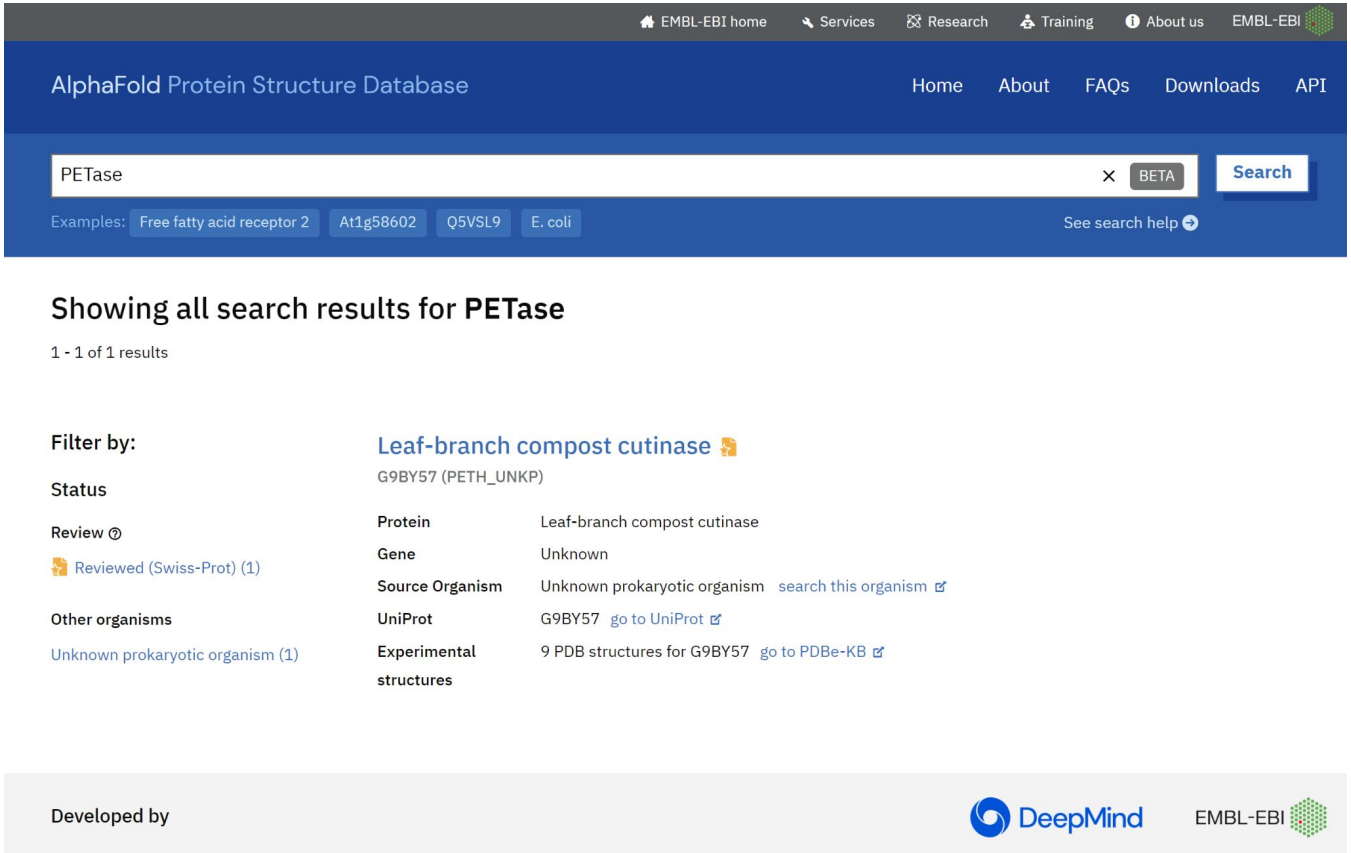
- PDB file
- mmCIF file
- Binary mmCIF file
- PAE (JSON) file
- FASTA file

● Show all entries ○ Show only PDB entries ○ Show only AFDB entries Taxonomic filter

	fier	Description	Species	Sequence identity	Sequence length	E-value	Average pLDDT
☑		Protein ecdysoneless homolog	*Homo sapiens*	100%	174	5.42e-30	77.00
☐		Organic hydroperoxide resistance protein	*Mus musculus*	100%	174	5.42e-30	77.00
☐		Organic hydroperoxide resistance protein	*Staphylococcus xylosus*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00
☐	AF-Q3U5Q7-F1	Organic hydroperoxide resistance protein	*Pyrobaculum calidifontis JC...*	100%	174	5.42e-30	77.00

Results per page: 10 ⌄ 1-10 of 1985 ‹ Previous page Next page ›

Why would you use structure clusters?



EMBL-EBI

Why would you use structure clusters?

Leaf-branch compost cutinase

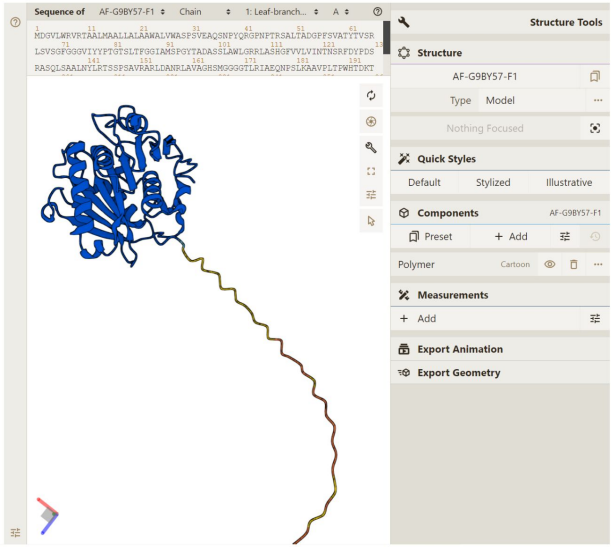
AlphaFold structure prediction

Download **PDB file** **mmCIF file** **Predicted aligned error**

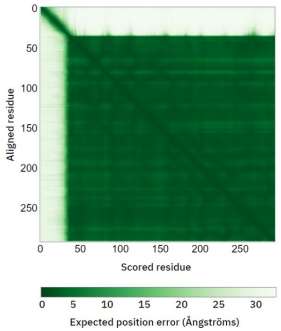
Share your feedback on structure with DeepMind **Looks great** **Could be improved**

Information

3D viewer



Predicted aligned error (PAE)



Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – go to Help section below for more information.

Source: https://alphafold.ebi.ac.uk/entry/G9BY57

EMBL-EBI

Why would you use structure clusters?

AFDB50/MMseqs2 (2341) | **AFBD/Foldseek (178)**

Structural clustering of the protein structure with the highest pLDDT for each AFDB50 cluster using Foldseek Cluster (Barrio-Hernandez & Yeo et al., Nature, 2023). Each cluster is comprised of structures that fulfil two criteria: maintaining an E-value threshold below 0.01 and ensuring a 90% bi-directional structure overlap to the largest structure of a cluster representative.

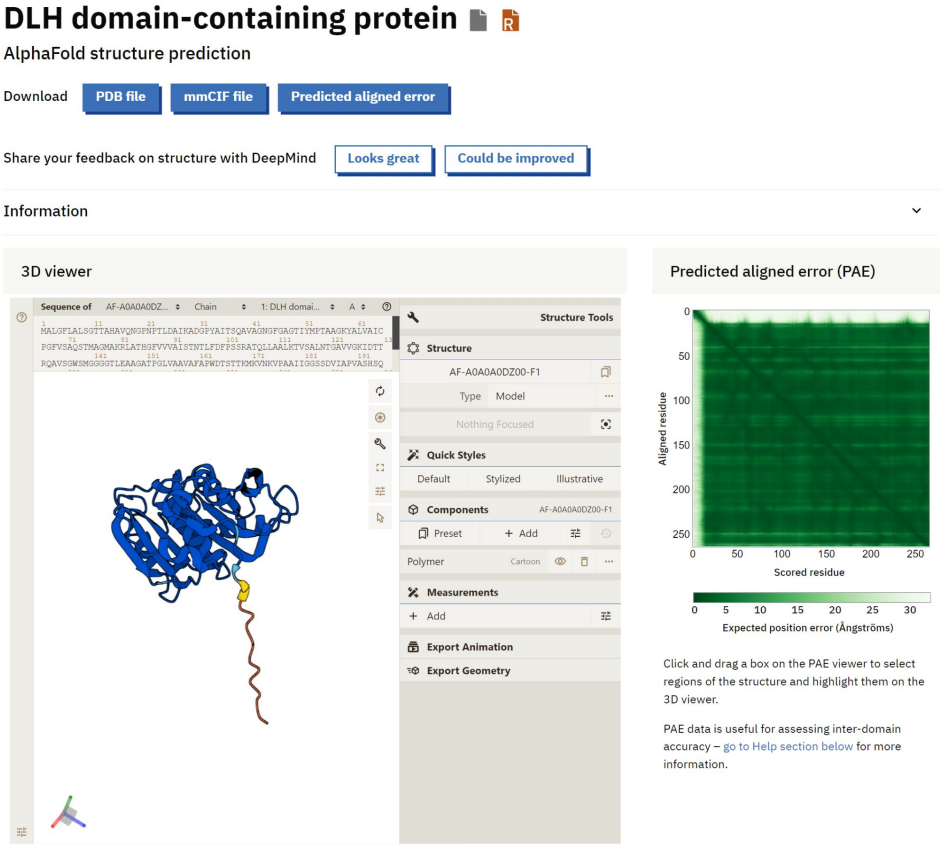
Taxonomic filter

AFDB accession	Description	Species	Sequence length	Average pLDDT
AF-A0A1C5EHN8-F1	Platelet-activating factor acetylhydrolase, isoform II	*Streptomyces sp. MnatMP-M17*	273	95.75
AF-A0A349SVV4-F1	Alpha/beta hydrolase	*Gammaproteobacteria bacterium*	273	95.75
AF-A0A2N8SPS3-F1	Alpha/beta hydrolase	*Pseudomonas stutzeri*	281	95.56
AF-A0A6G3T934-F1	Alpha/beta hydrolase	*Streptomyces rubrogriseus*	274	95.5
AF-A0A7Y0JP13-F1	Alpha/beta hydrolase	*Actinoplanes sp. TBRC 11911*	271	95.44
AF-A0A1H0Q4E1-F1	Dienelactone hydrolase	*Actinopolyspora xinjiangensis*	272	95.31
AF-A0A8B0ELD5-F1	Alpha/beta hydrolase	*Streptomyces sp. LBUM 1480*	271	95.25
AF-A0A399VEE9-F1	Alpha/beta hydrolase	*Nakamurella silvestris*	263	95.12
AF-A0A4R4N608-F1	Alpha/beta hydrolase	*Actinomadura bangladeshensis*	271	94.88
AF-A0A0A0DZ00-F1	DLH domain-containing protein	*Aquabacterium sp. NJ1*	265	94.81

Items per page: 10 1 – 10 of 178

EMBL-EBI

Why would you use structure clusters?



Source: https://alphafold.ebi.ac.uk/entry/A0A0A0DZ00

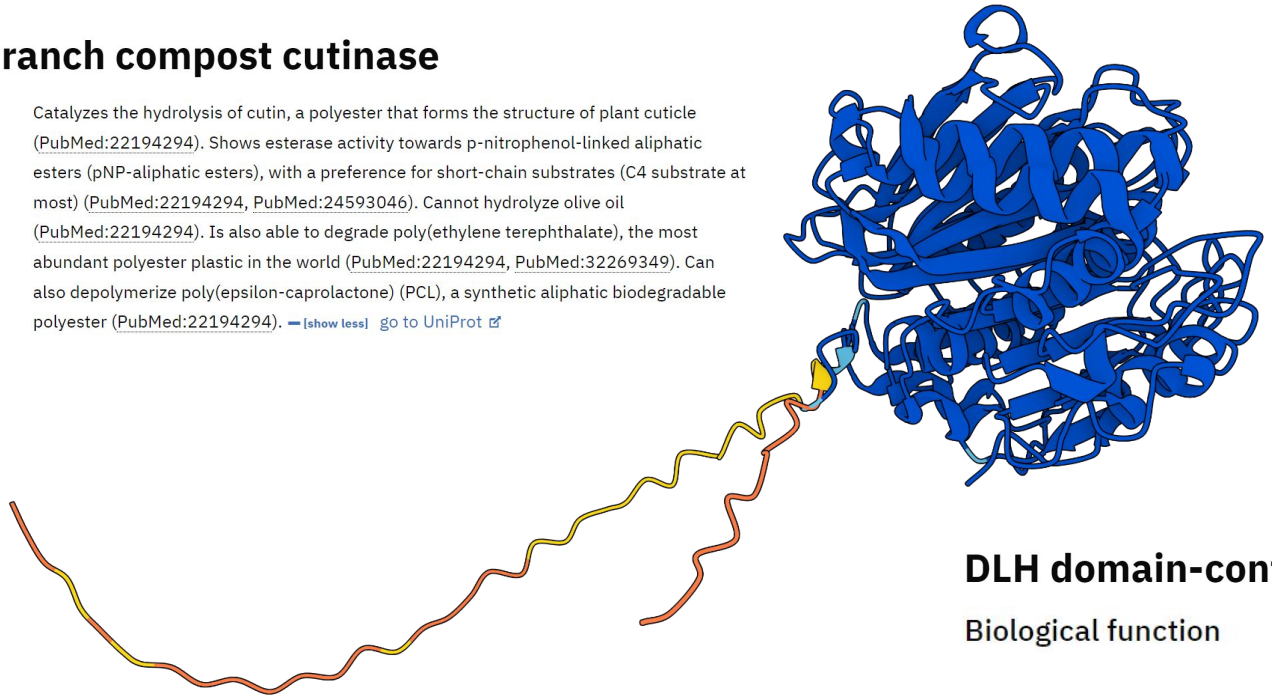
EMBL-EBI

Why would you use structure clusters?

Leaf-branch compost cutinase

Biological function

Catalyzes the hydrolysis of cutin, a polyester that forms the structure of plant cuticle (PubMed:22194294). Shows esterase activity towards p-nitrophenol-linked aliphatic esters (pNP-aliphatic esters), with a preference for short-chain substrates (C4 substrate at most) (PubMed:22194294, PubMed:24593046). Cannot hydrolyze olive oil (PubMed:22194294). Is also able to degrade poly(ethylene terephthalate), the most abundant polyester plastic in the world (PubMed:22194294, PubMed:32269349). Can also depolymerize poly(epsilon-caprolactone) (PCL), a synthetic aliphatic biodegradable polyester (PubMed:22194294). — [show less] go to UniProt



DLH domain-containing protein

Biological function Not available.

EMBL-EBI