Protein structure prediction

“In science, as in life,
it is always bad
to fall in love with models”.

“All models are wrong,
but some are useful”.

George Box

Prediction strategies

- Homology modelling – comparison to homologous proteins

Prediction strategies

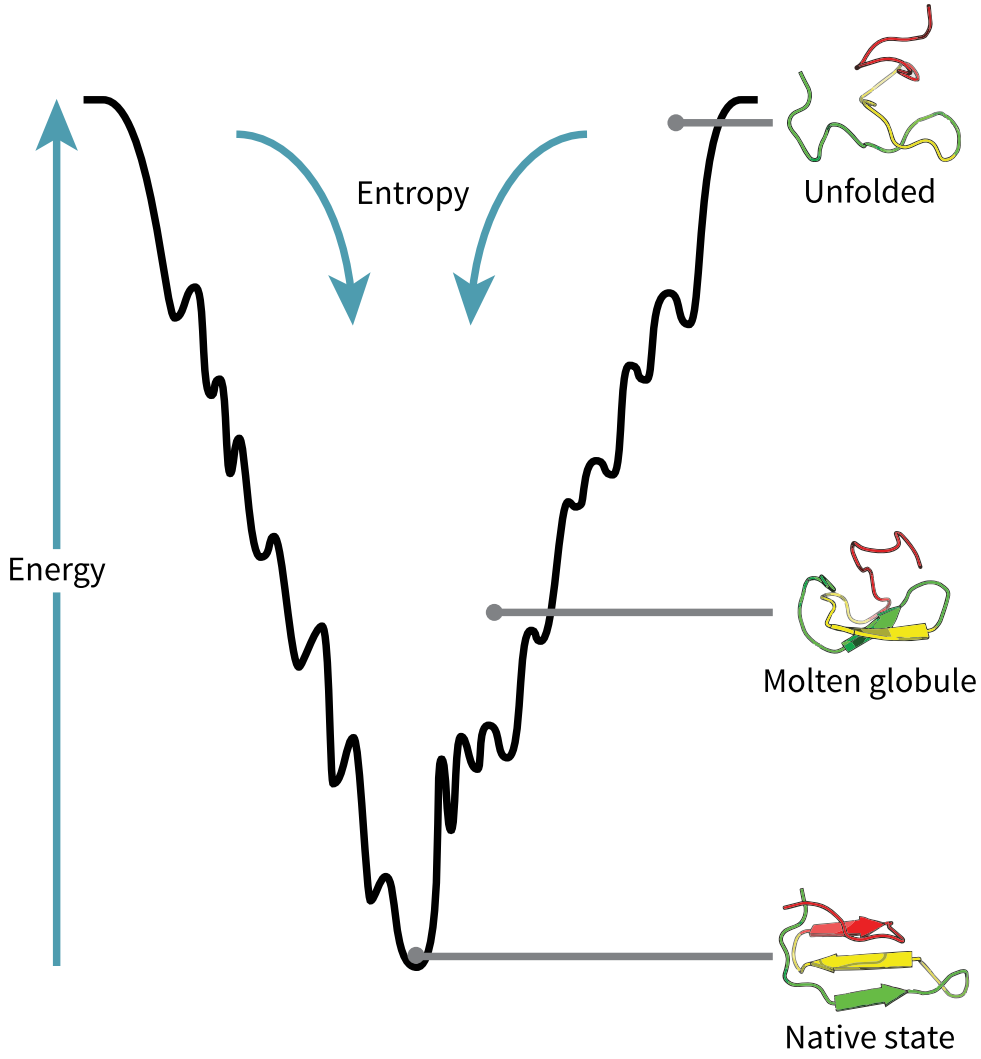
- Homology modelling – comparison to homologous proteins
- *De novo* or *ab initio* – prediction from primary structure

Anfinsen‘s dogma

„*The native structure is determined only by the protein‘s amino acid sequence, at least for small globular proteins.*“

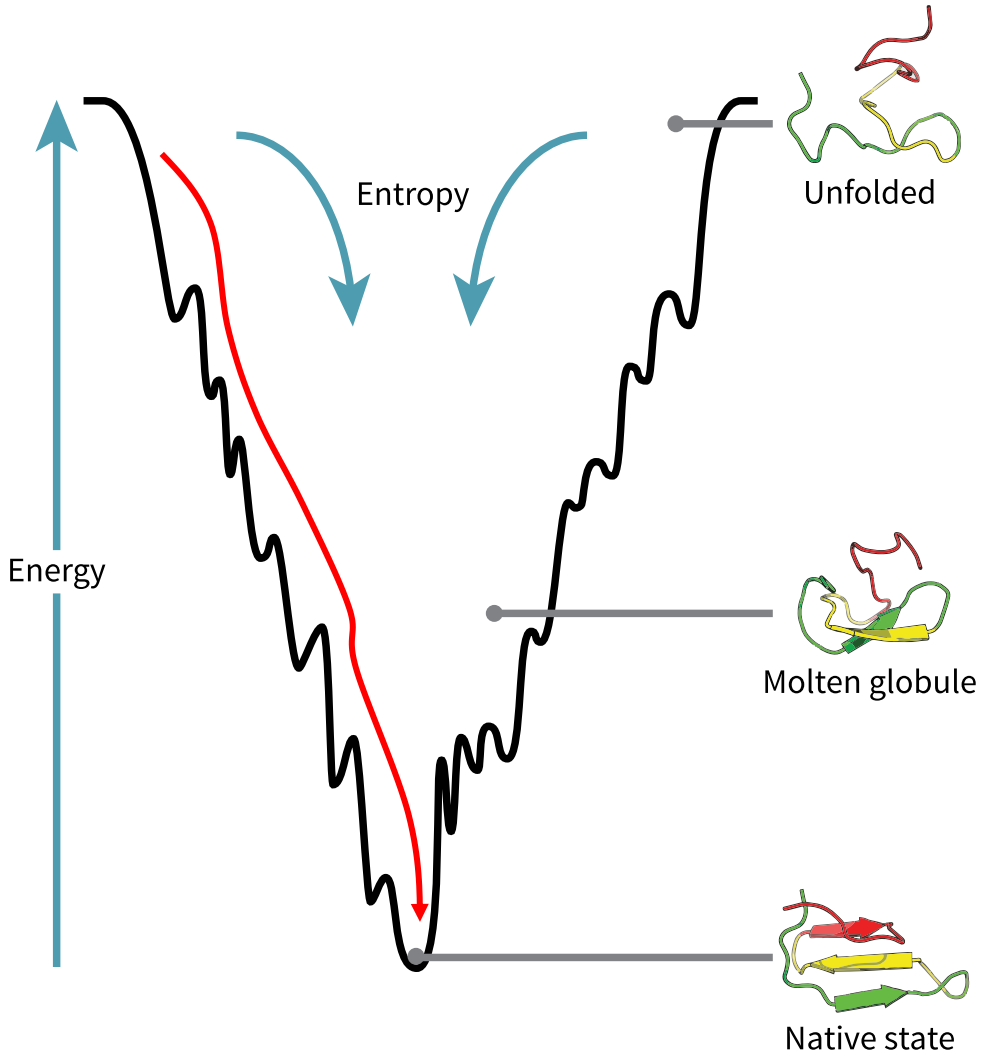
Christian B. Anfinsen

What folds a protein?



- Intrinsic forces

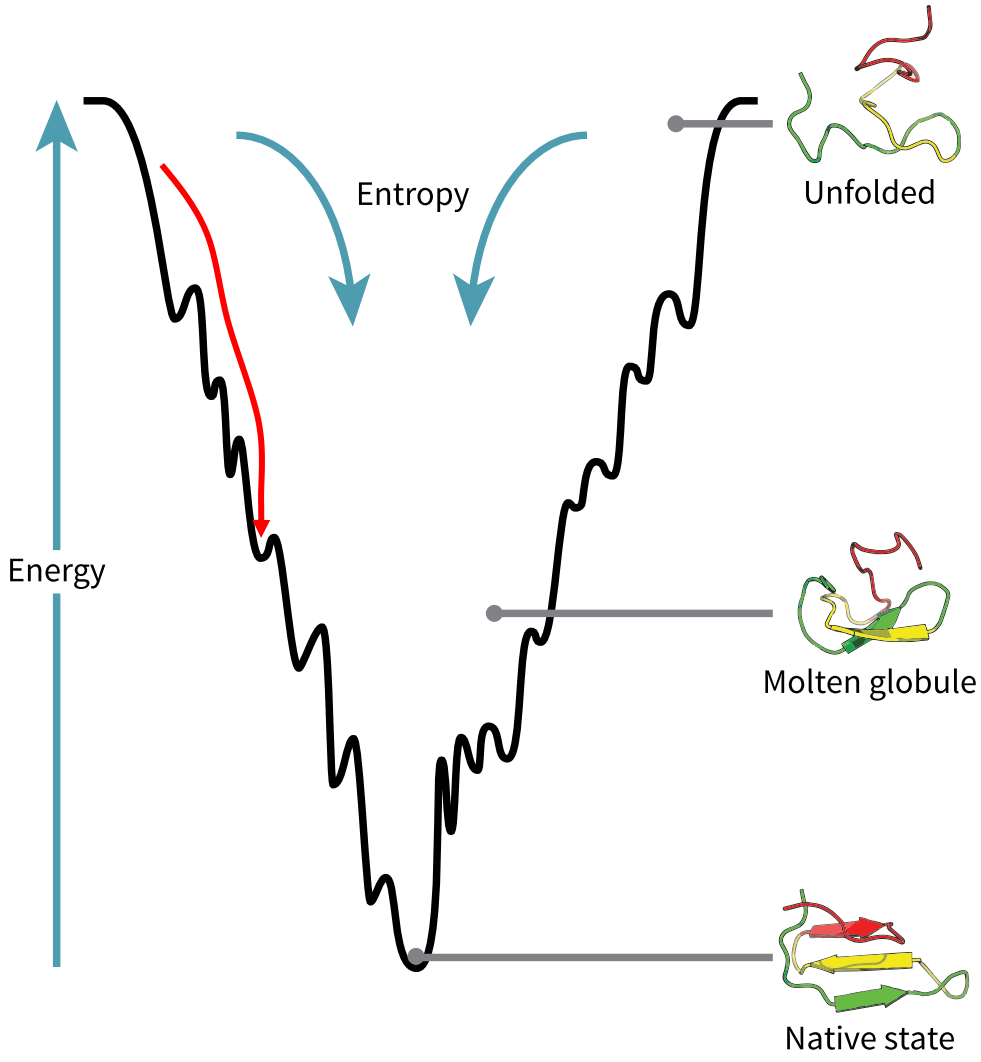
Minimization



Tries to take the protein to a conformation in the global minimum.

But it ends up trapped at local minima.

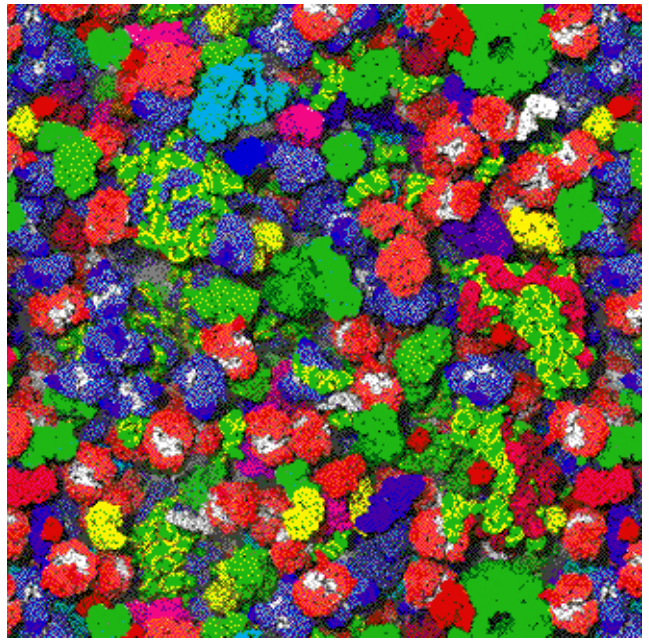
Minimization



Tries to take the protein to a conformation in the global minimum.

But it ends up trapped at local minima.

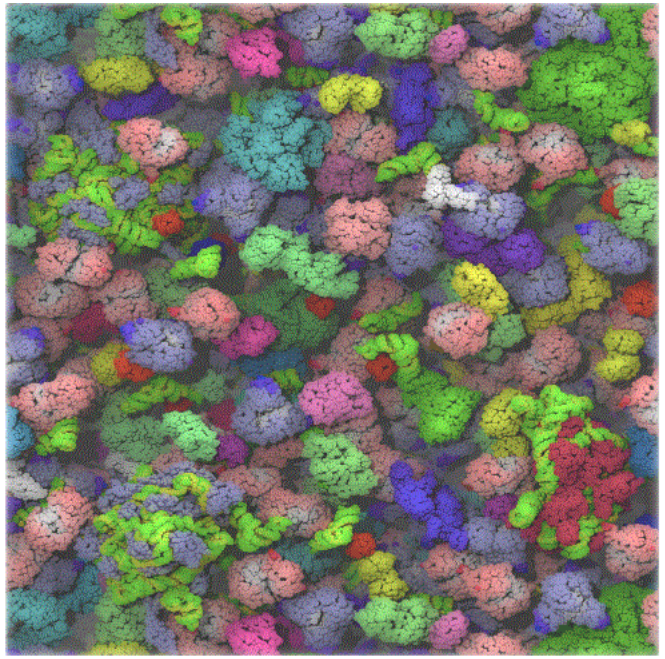
What folds a protein?



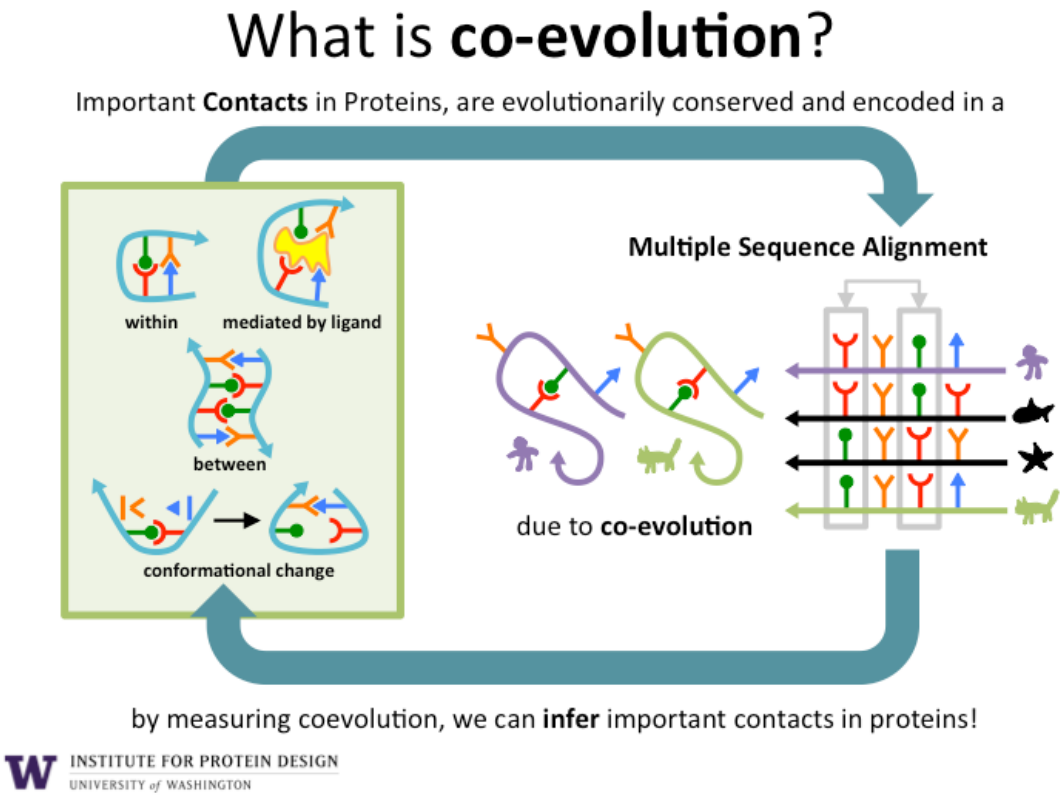
McGuffee and Elcock, 2010

15 μs simulation of cytoplasm

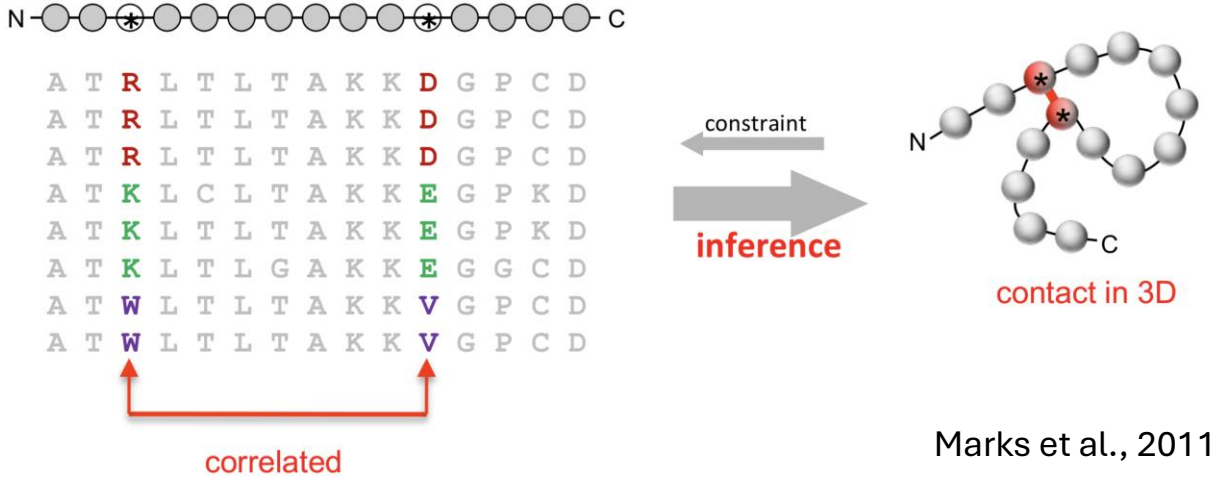
- Environmental conditions



What folds a protein?



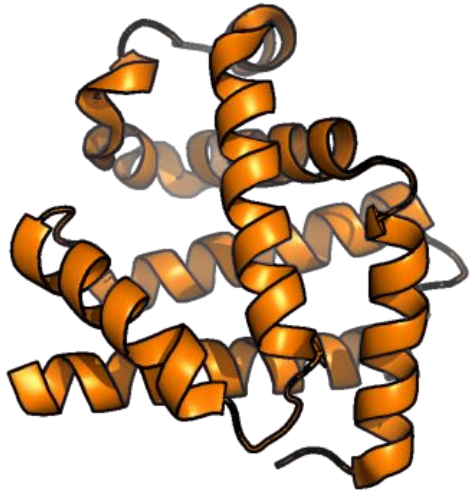
- Evolutionary constraints



Marks et al., 2011

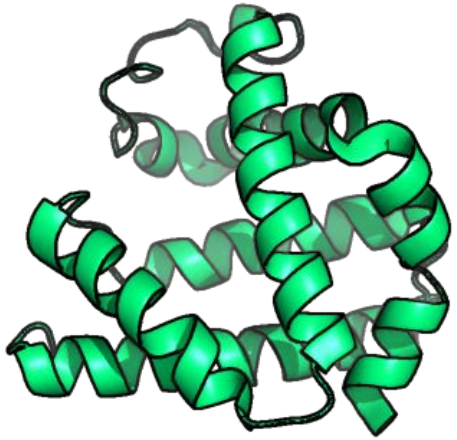
Low sequence similarity but high structural homology

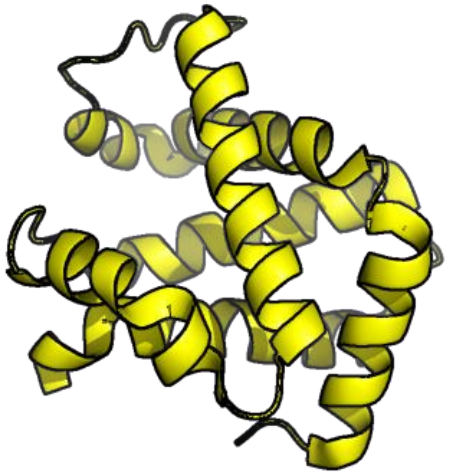
Human myoglobin



African elephant myoglobin
80% sequence identity

Pigeon myoglobin
25% sequence identity



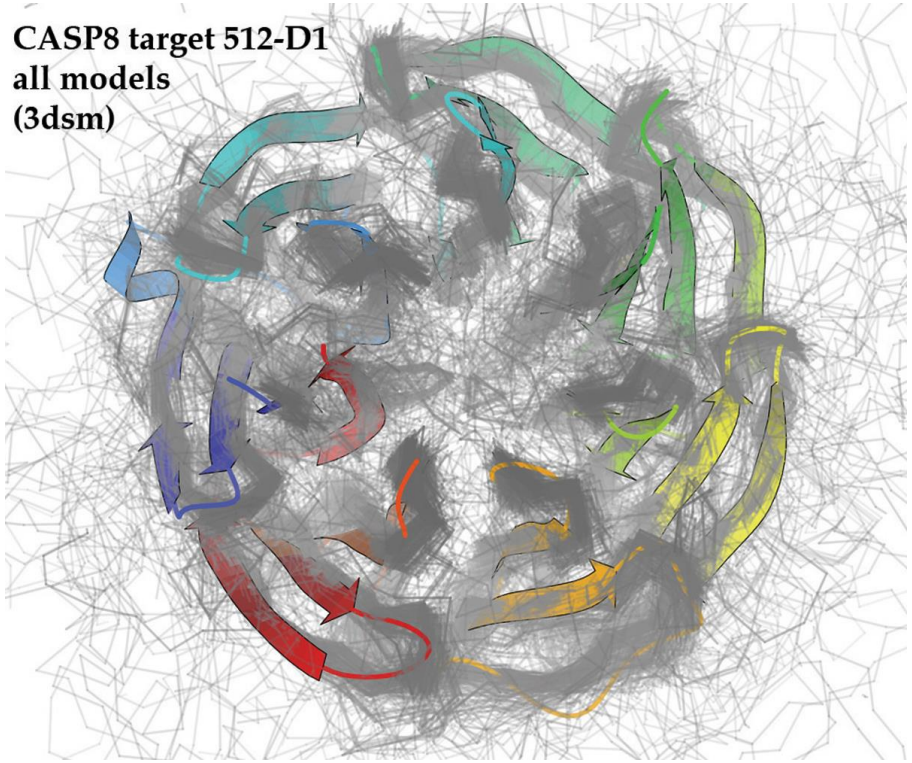


Black-fin tuna myoglobin
45% sequence identity

Rubiera, https://www.blopig.com/blog/

CASP

- **Critical Assessment of Structure Prediction**



- Every 2 years since 1994
- An experiment to objectively evaluate the predictions of a community
- Help advance prediction methods
- „world championship“
- Target structures were very recently solved and not yet published

Alphafold will change everything

nature

Explore content | About the journal | Publish with us | Subscribe

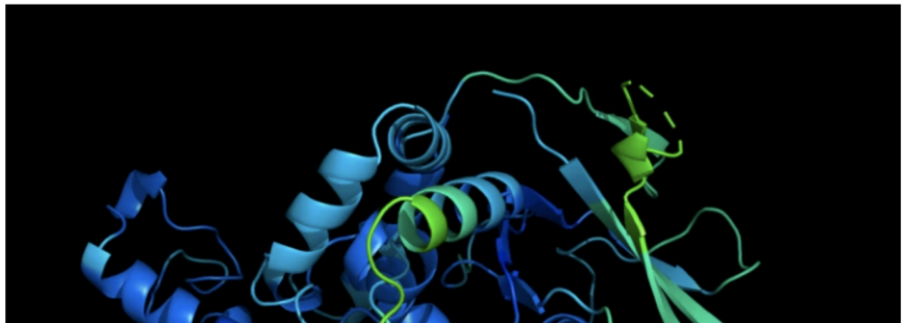
nature > news > article

NEWS | 30 November 2020

'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures

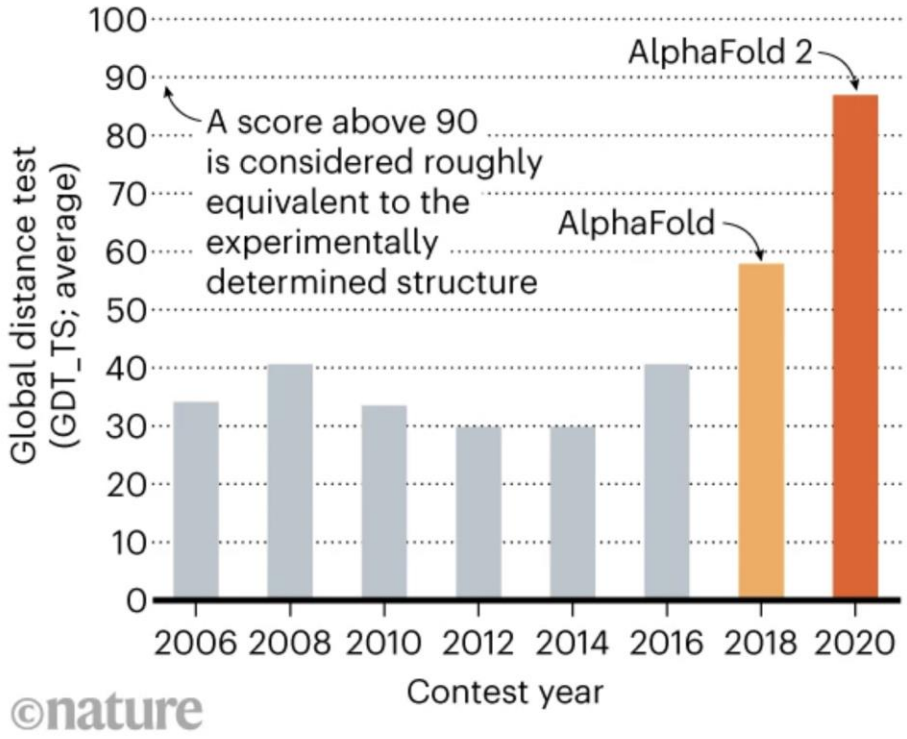
Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway

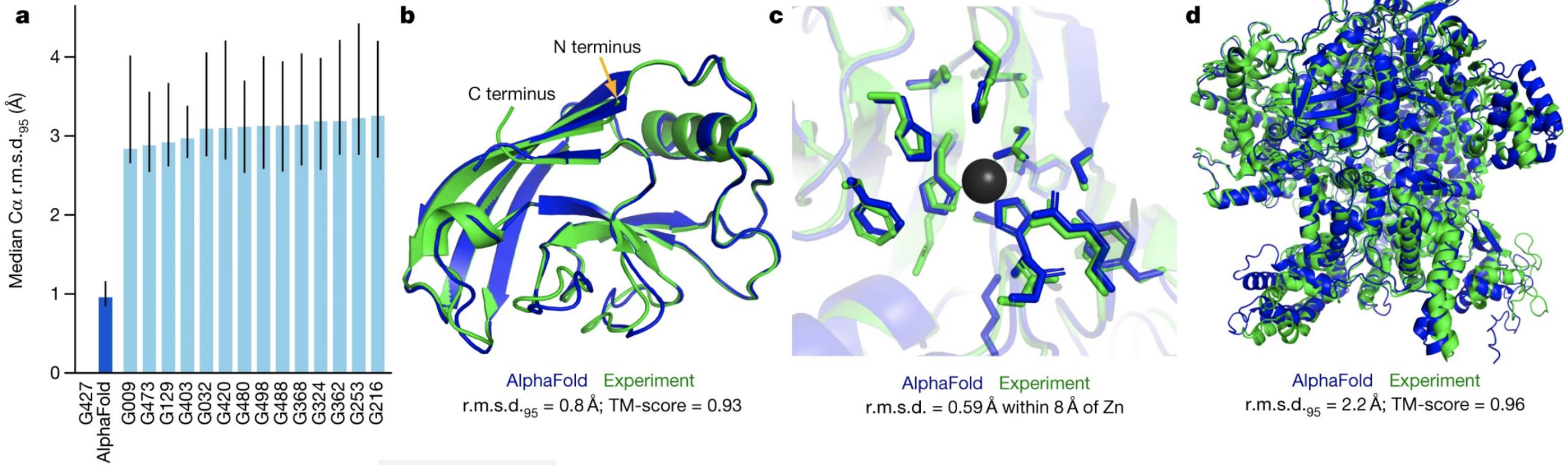


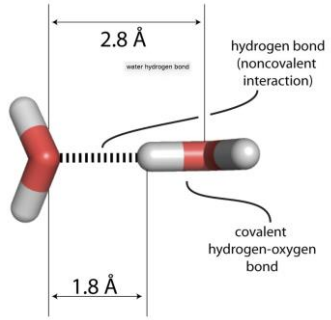
STRUCTURE SOLVER

DeepMind's AlphaFold 2 algorithm significantly outperformed other teams at the CASP14 protein-folding contest — and its previous version's performance at the last CASP.



AlphaFold accuracy

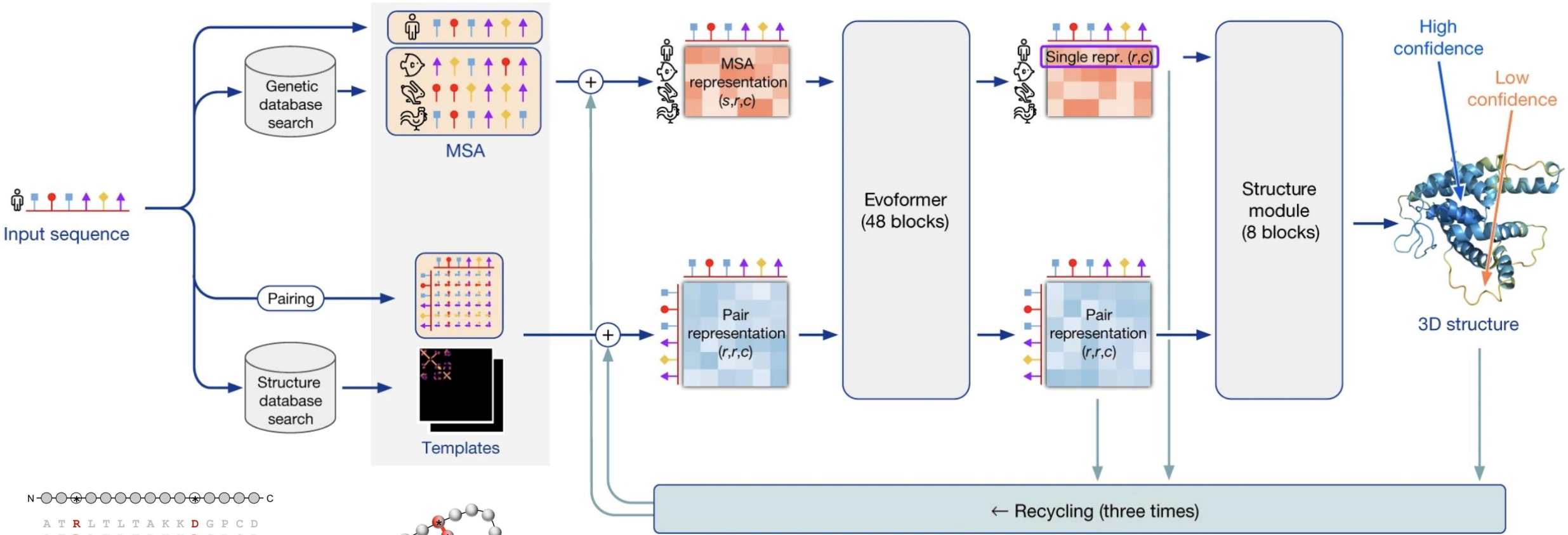


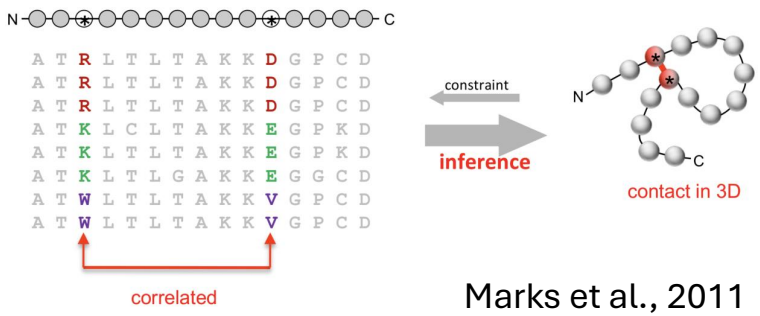


- Compared by RMSD and TM-score

Jumper et al., 2021

AlphaFold pipeline

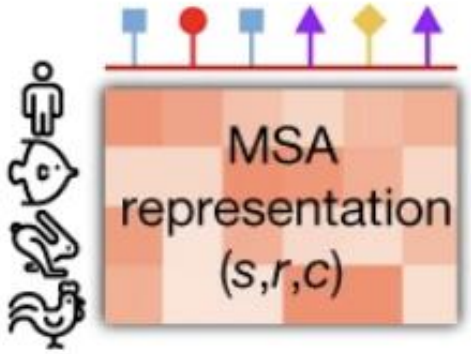


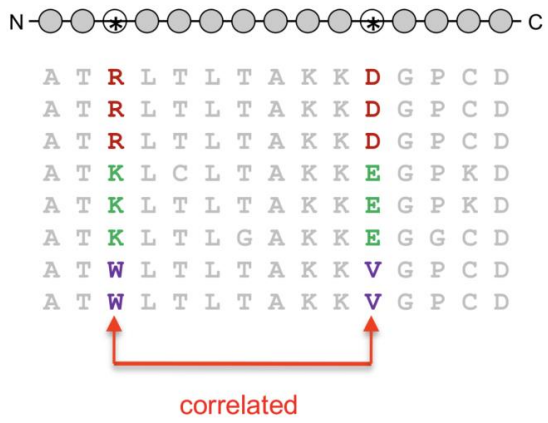


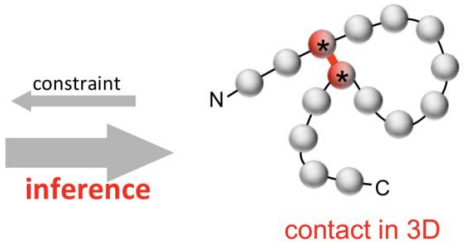
Marks et al., 2011

Jumper et al., 2021

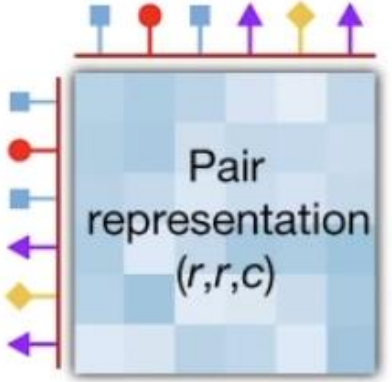
AlphaFold pipeline

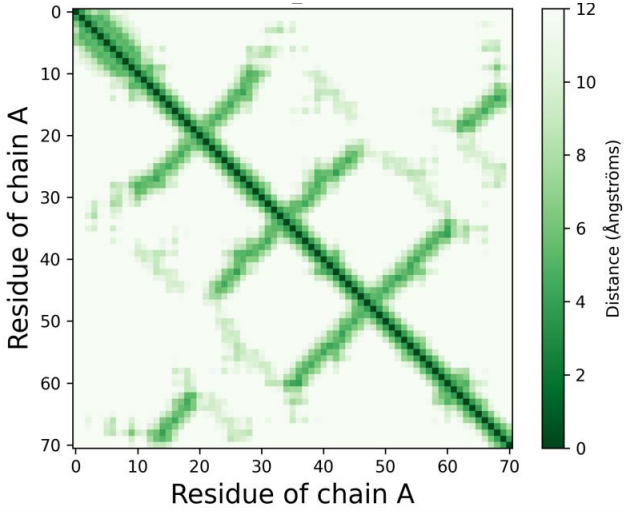






Marks et al., 2011





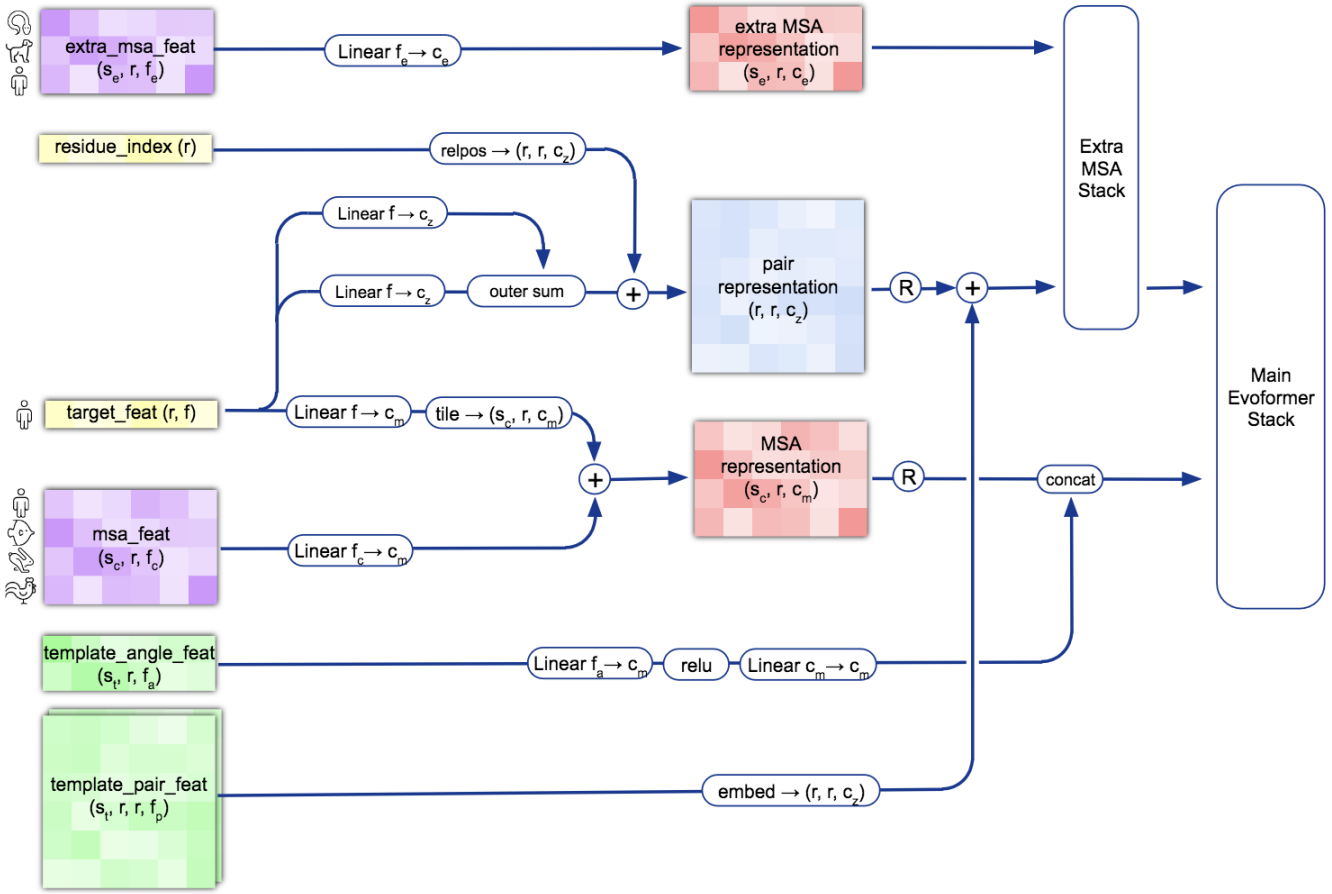
Like a contact map, but with more information

Keys to success of AlphaFold2?

- A great database
    - **Big Fantastic Database -** 65,983,866 families , 2,204,359,010 protein sequences from reference databases, metagenomes and metatranscriptomes
    - **PDB**
    - **PDB70**
    - **Uniref90**
    - **Uniclust30**
    - **Uniprot**
    - **MGnify**

Keys to success of AlphaFold2?

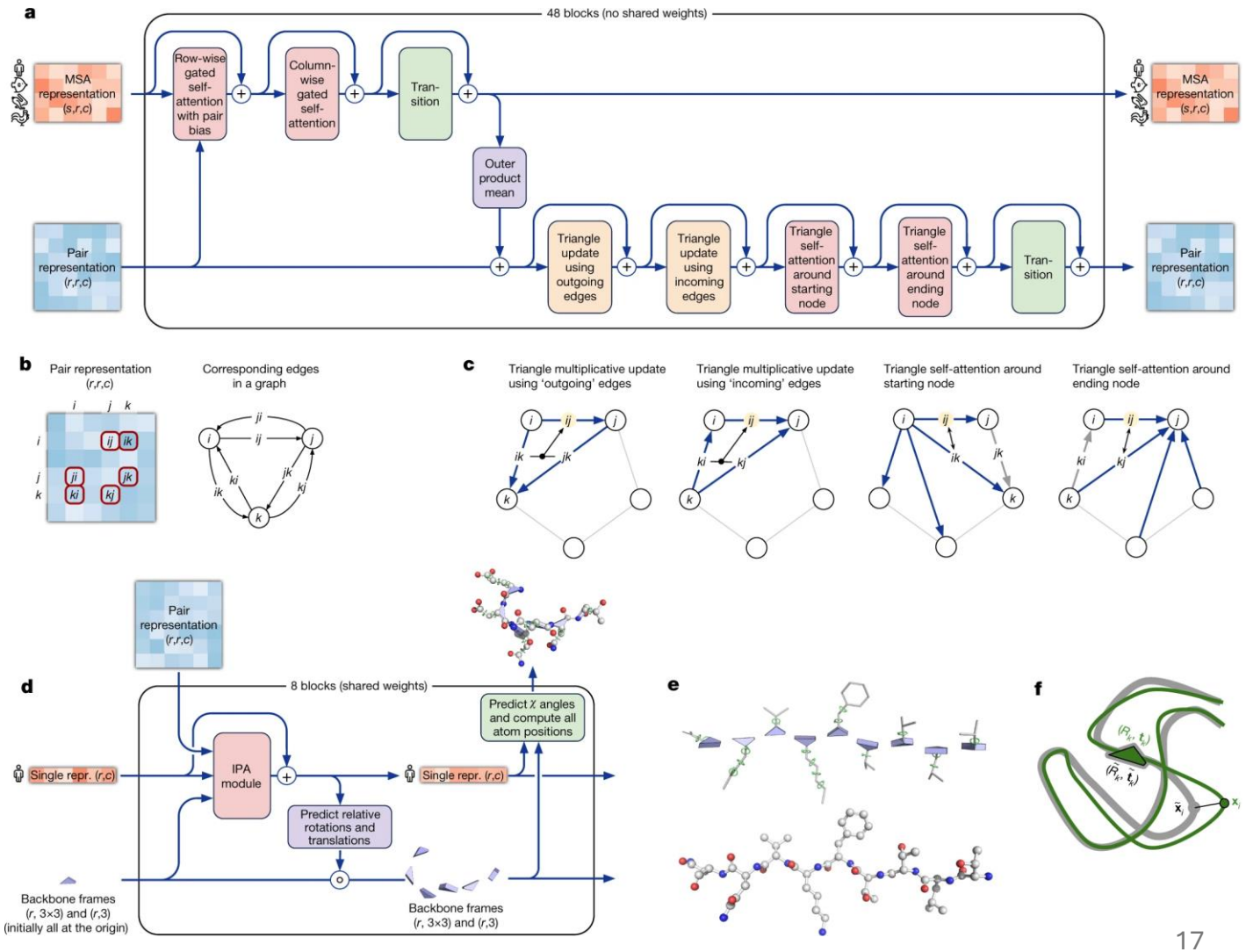
- Abbility to handle complex (a great team of experts and a great infrastructure)



Jumper et al., 2021

Keys to success of AlphaFold2?

- A neural network that alternates between structural and geometrical data and evolutionary data (Evoformer)
- And other clever architectural tricks and assumptions like a „gas of residues“

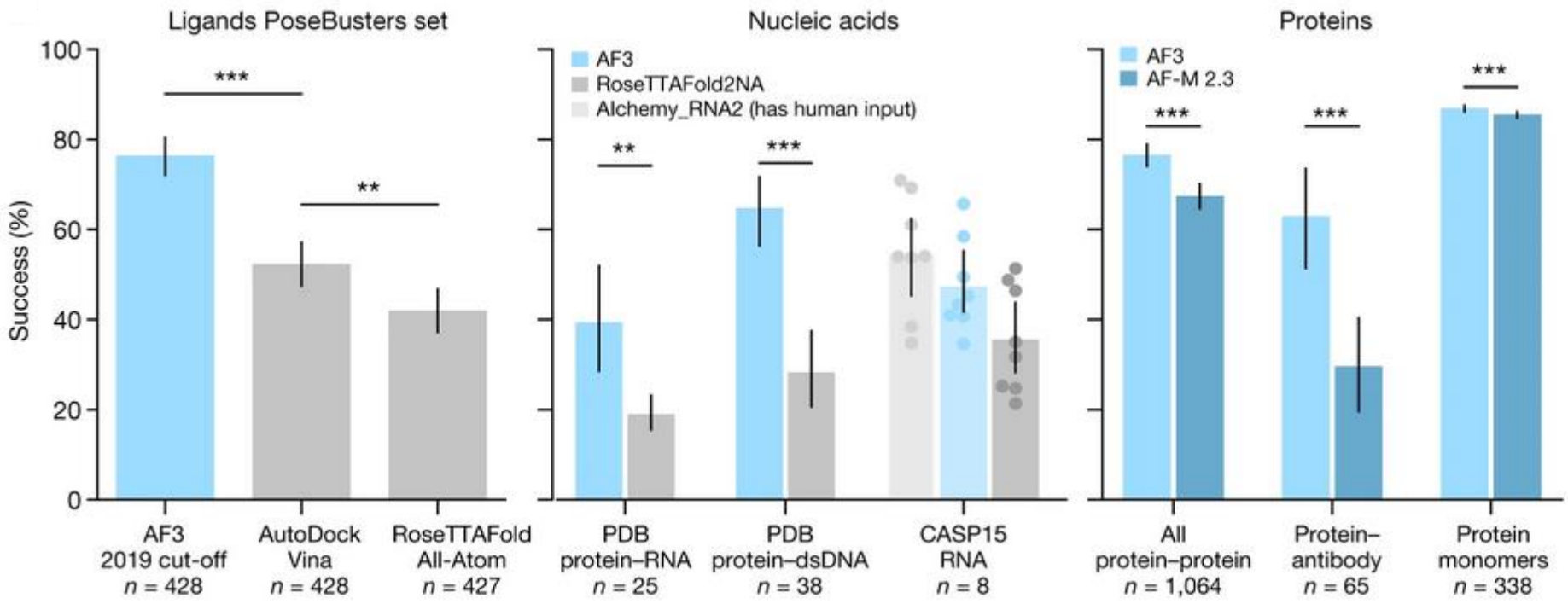


Jumper et al., 2021

More on how Alphafold works

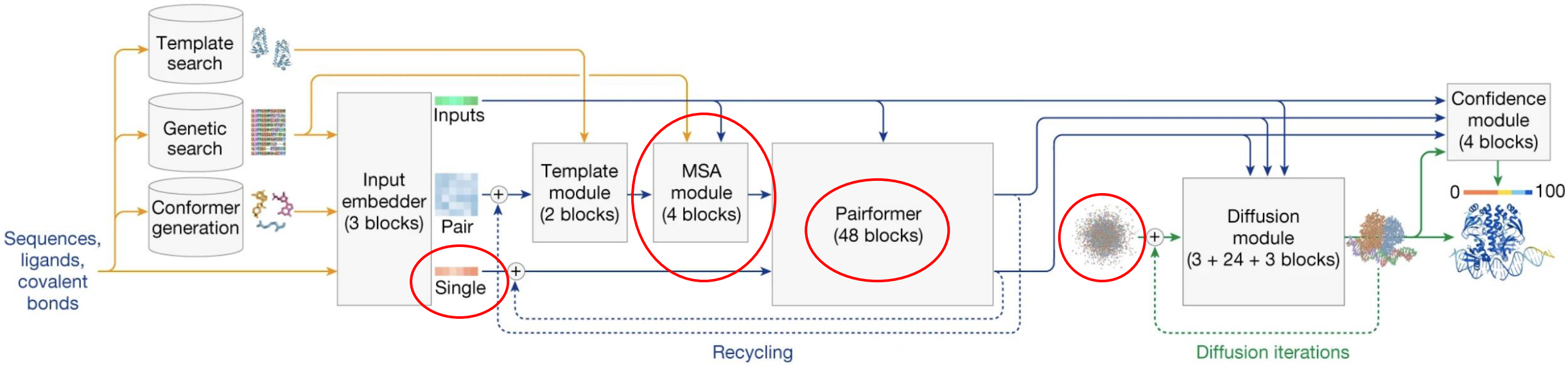
- **Highly Accurate Protein Structure Prediction with AlphaFold | SimonKohl**
  - **https://www.youtube.com/watch?v=tTN0MM2CQLU**
- Oxford Protein Informatics Group - **AlphaFold 2 is here: what's behind the structure prediction miracle**
  - **https://www.blopig.com/blog/2021/07/alphafold-2-is-here-whats-behind-the-structure-prediction-miracle/**

Novelties of AlphaFold3



Abramson et al., 2024

Novelties of AlphaFold3



Abramson et al., 2024

Interpretation of predictions

Usual metrics

- RMSD – Root Mean Square Deviation
- TM-score – Template Modelling score
- LDDT – Local Distance Difference Test
- GDT – Global Distance Test (used by CASP)

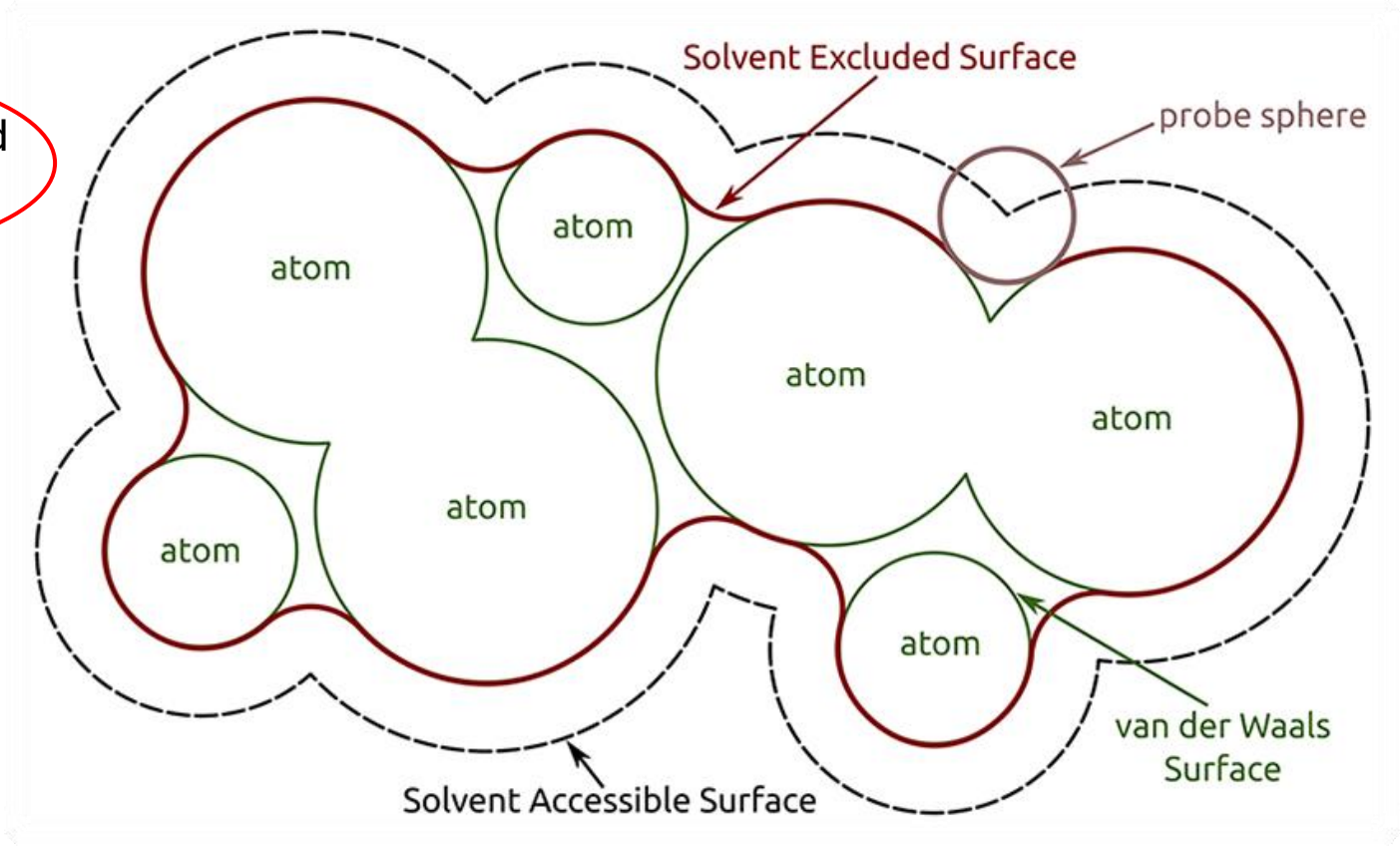
AlphaFold metrics

- pLDDT – predicted LDDT
- PAE – Predicted Aligned Errors

RASA with AlphaFold3

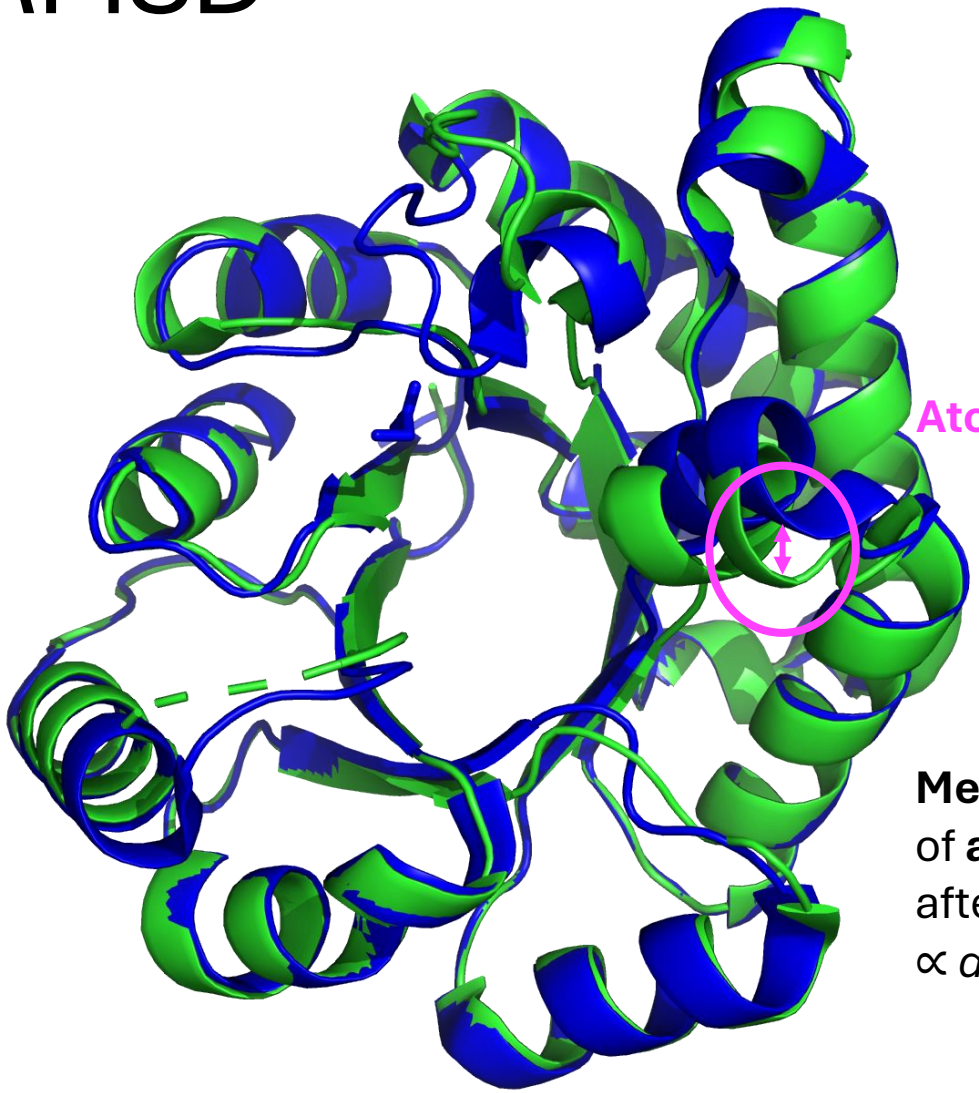
- Relative solvent Accessible Surface Area

Compared to extended conformation



https://www.cgl.ucsf.edu/chimera/docs/UsersGuide/representation.html

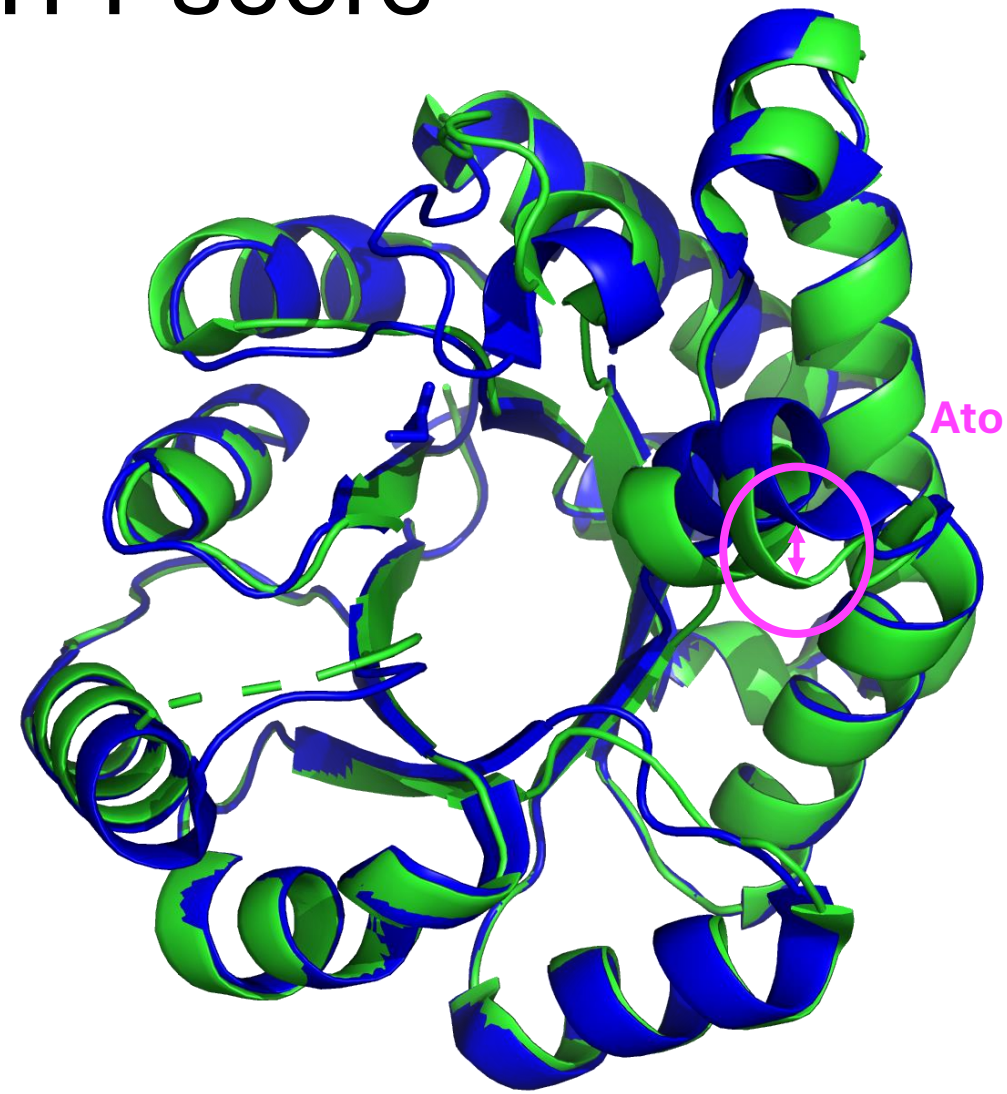
RMSD



Atomic deviation d_i

Mean deviation
of **all** aligned atoms
after **superposition**
$\propto d_i$

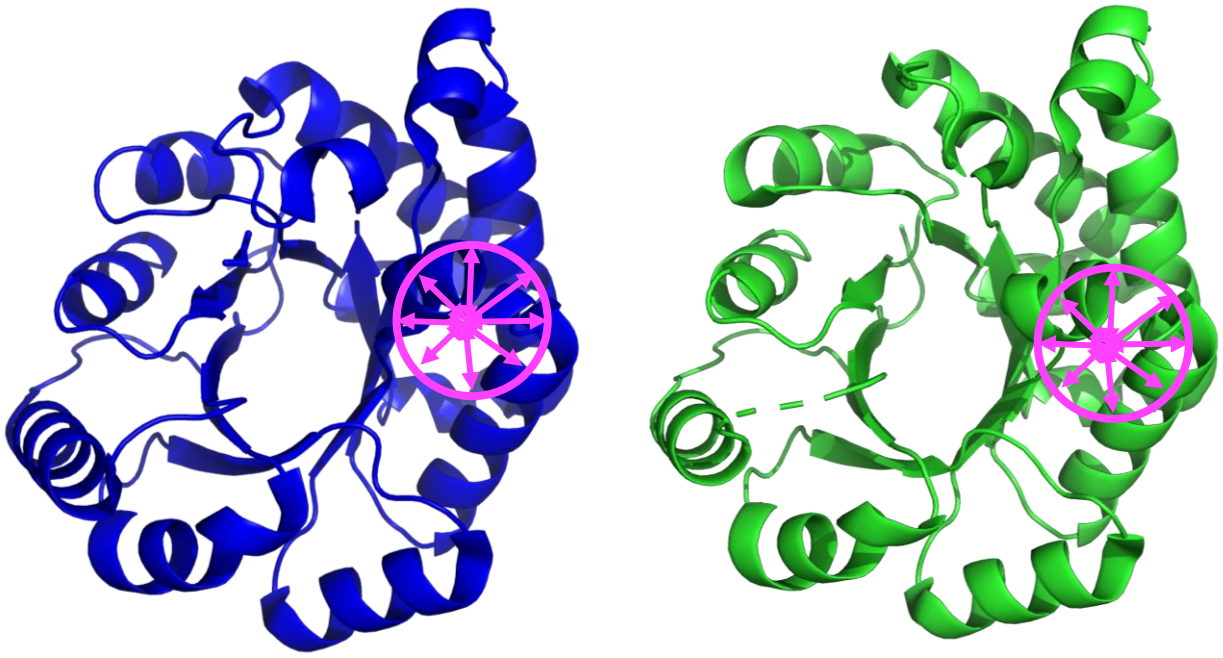
TM-score



A **score** between (0,1] considering **all** aligned atoms after **superposition**
$\propto 1/d_i$

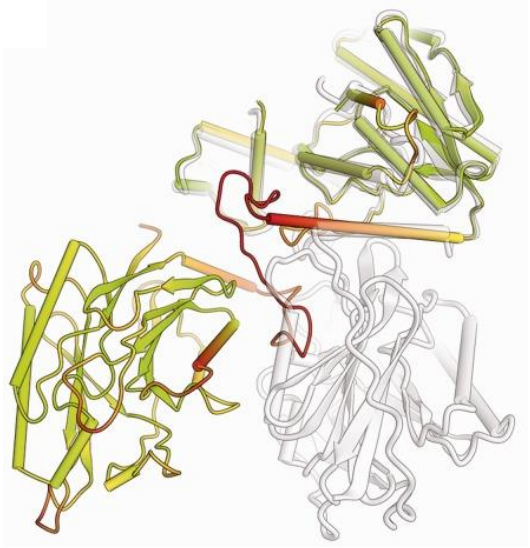
Intended to be more accurate because smaller deviations weigh more than larger deviations.

LDDT



List of d_i *within a threshold*

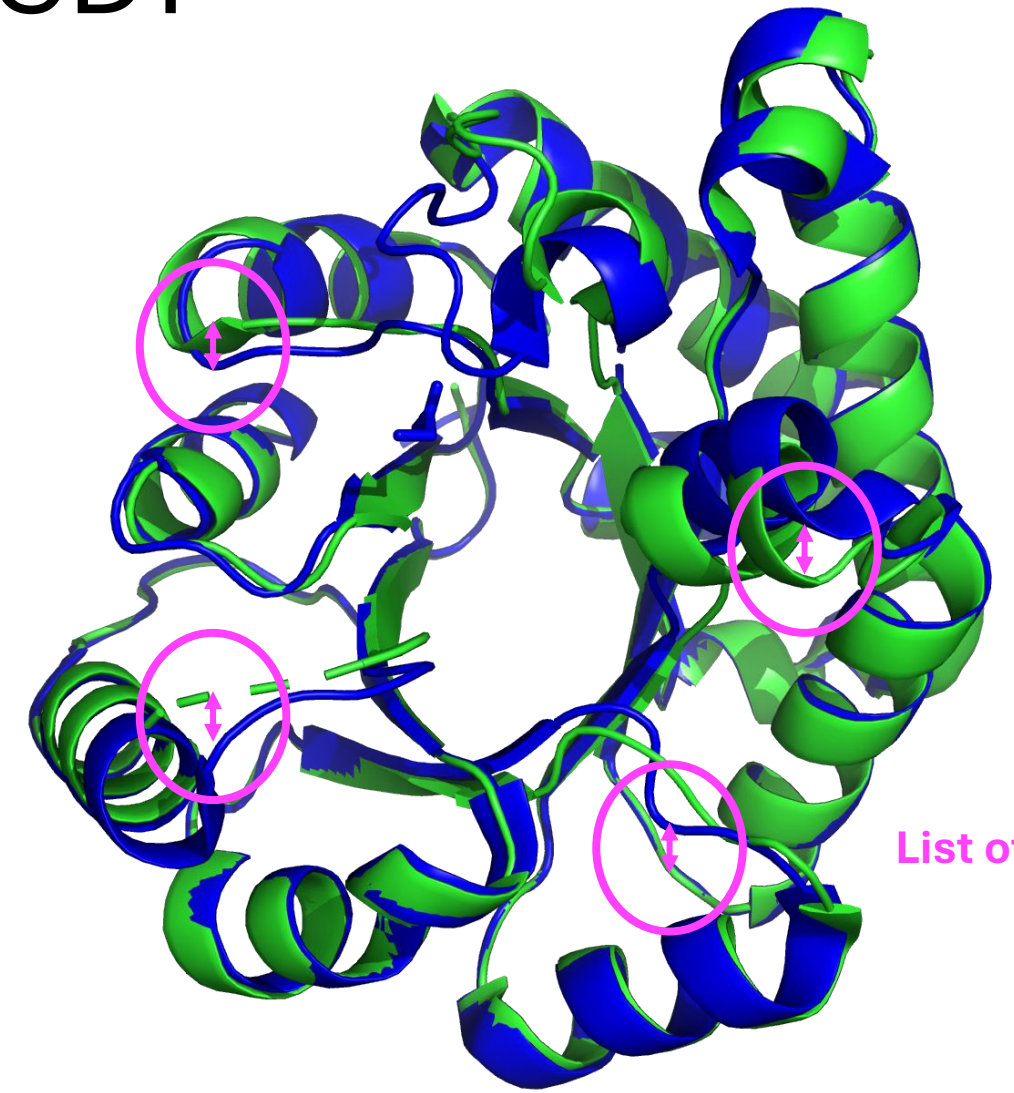
Superposition free,
percentage of conserved
local inter-atomic distances



Mariani et al., 2013

Good for comparison of structures with separate domains

GDT

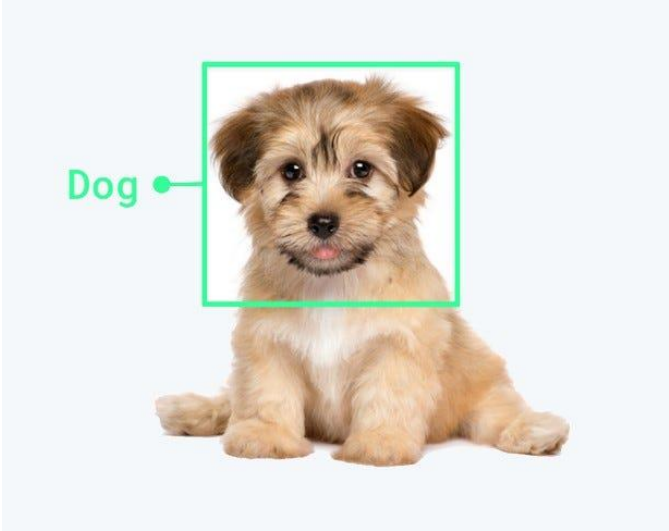


A **percentage** of residues that deviate within a **set of cut-offs** after iterative **superpositions**

List of deviating residues

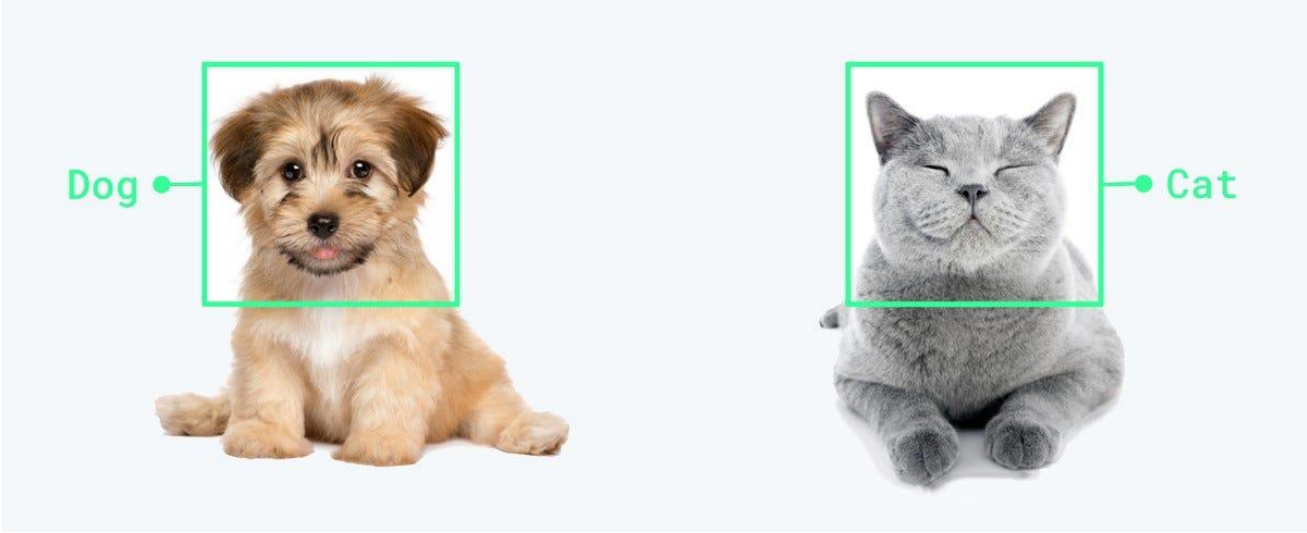
Correct in relation to others

Ground truth



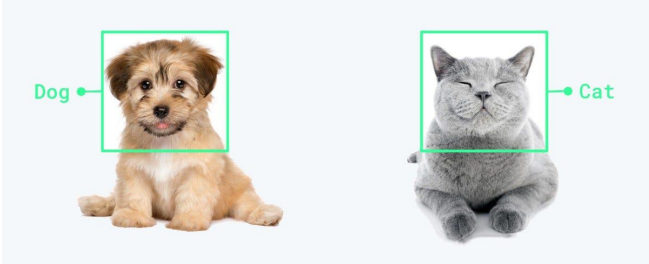
Correct in relation to others

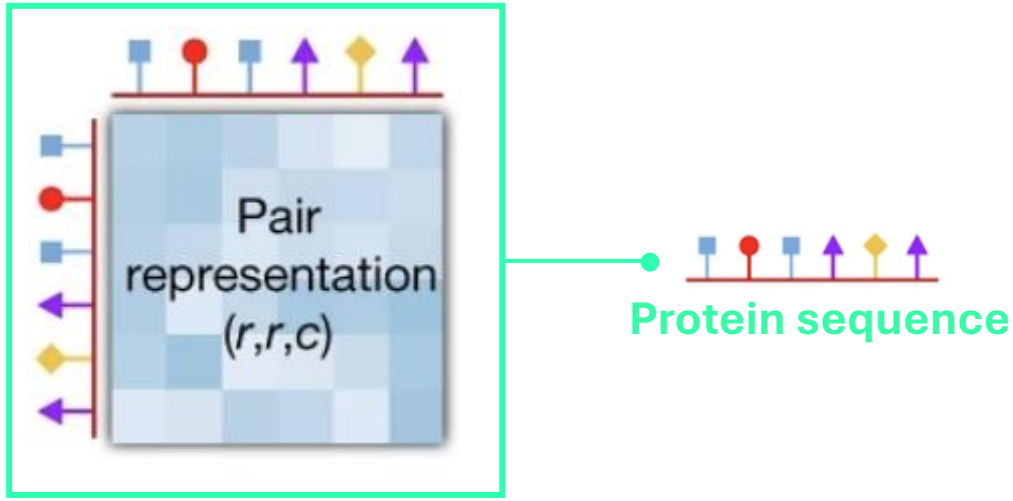
Ground truth



Correct in relation to others

Ground truth

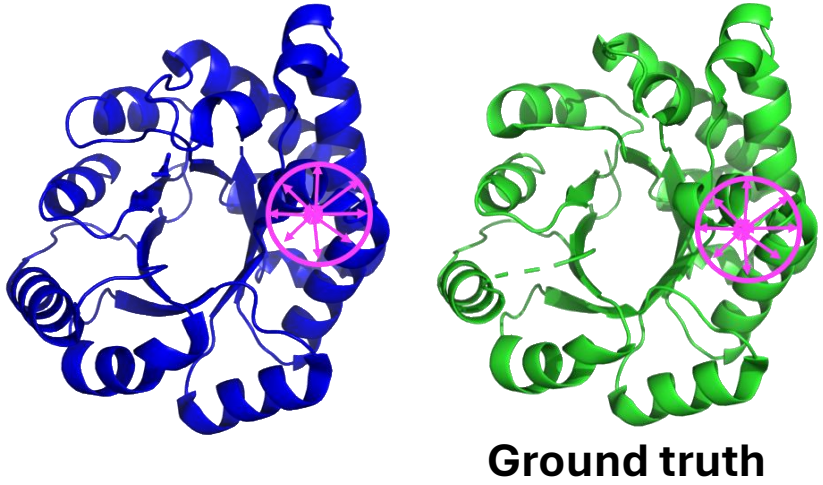




PDB

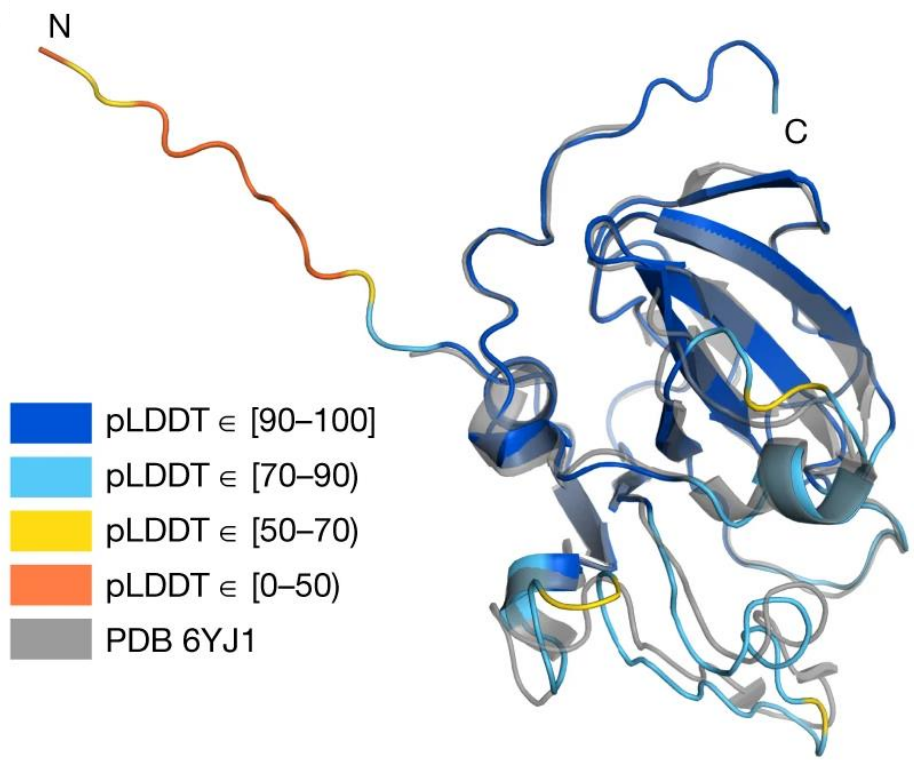
Contact map-like
(with more information)

pLDDT



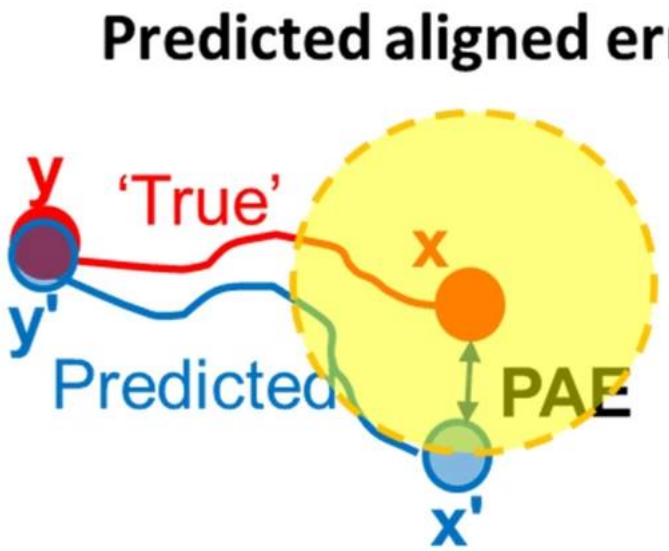
Ground truth

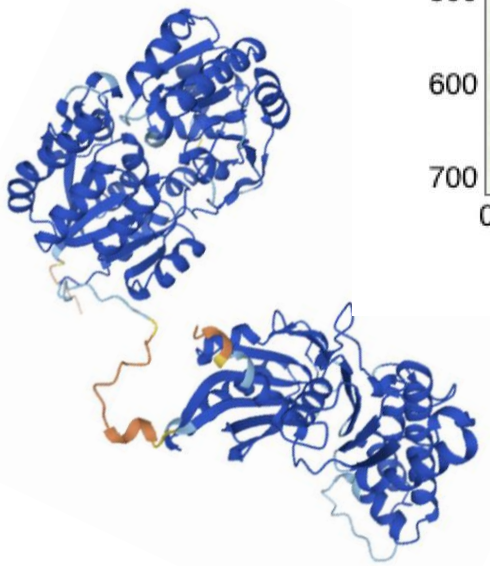
Per residue score

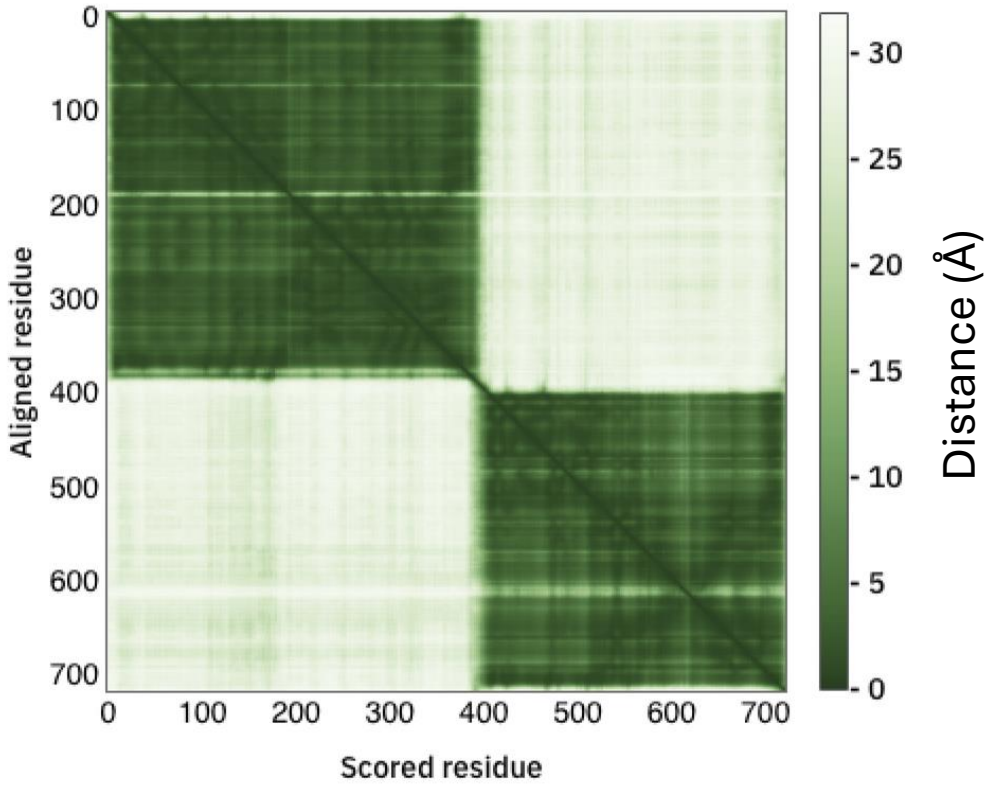


pLDDT > 90 – high confidence
pLDDT > 70 – confident / generally correct backbone
pLDDT < 70 – low confidence
pLDDT < 50 - not confident / disordered

PAE


Predicted aligned error (PAE)
y
'True'
x
y'
Predicted
PAE
x'

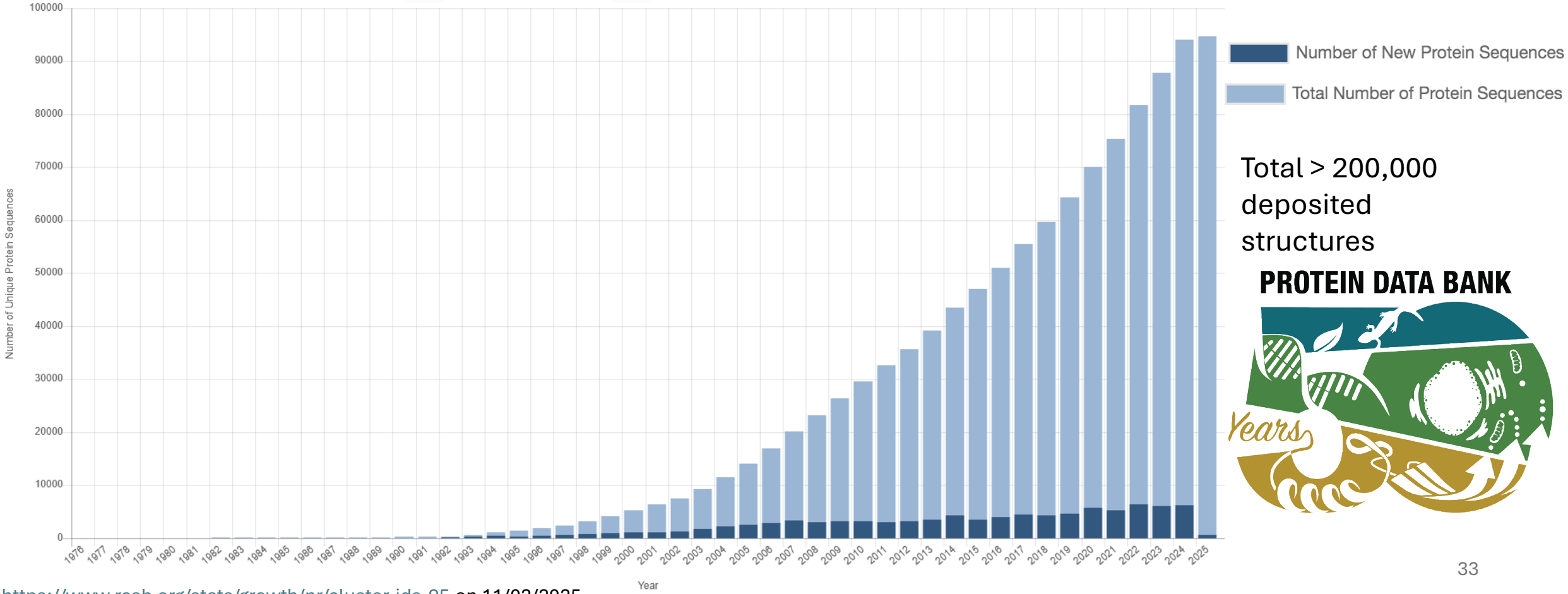



Aligned residue
Scored residue
0
100
200
300
400
500
600
700
30
25
20
15
10
5
0
Distance (Å)

Let‘s do some predictions!

Protein structure as a digital resource

- 1971 was stablished the Protein Data Bank with only 7 structures



Total > 200,000 deposited structures

PROTEIN DATA BANK

Years

https://www.rcsb.org/stats/growth/nr/cluster-ids-95 on 11/02/2025

Protein structure as a digital resource

- 2003 was announced the worldwide Protein Data Bank - wwPDB

https://www.rcsb.org/pages/about-us/index

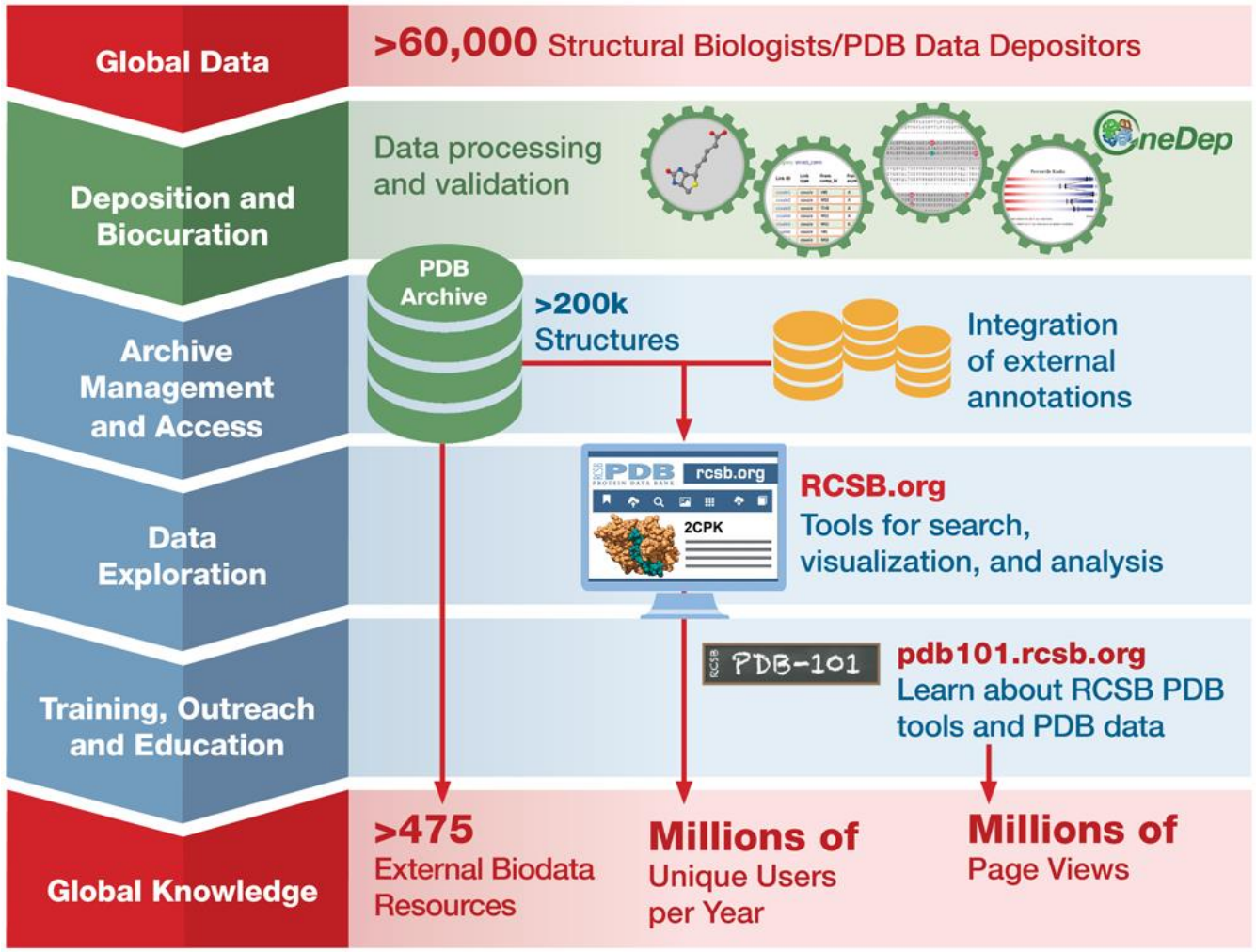
The PDB website

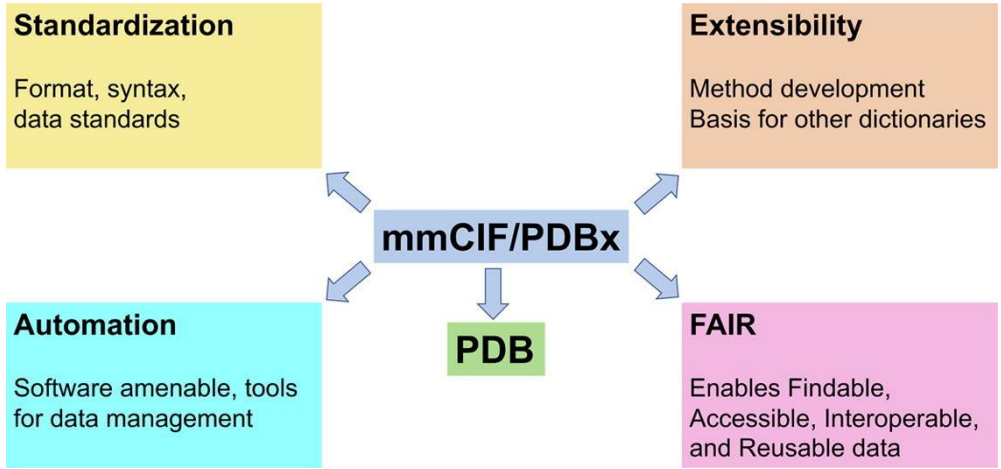
https://www.rcsb.org/

The coordinates files

- .pdb and .cif (*Crystallographic Information File*) are old formats adopted since 1977 and 1991, respectively, to document molecular structure – proteins, nucleic acids and ligands
- The .pdb format has not been updated since 2012, and the current standard PDB archive distribution format is the PDBx/mmCIF (Protein Data Bank Exchange/ macromolecular Cryistallographic Information File)
- .cif format allows bigger macromolecules than .pdb format (limited to 62 chains and 99999 atom records)
- AlphaFold outputs .pdb format
- Both files are machine and human readable formats

Global outreach of the Protein Data Bank





https://www.rcsb.org/pages/about-us/index

Westbrook et al., 2022

The headers

- Contain much of experimental and documentation details

.pdb

```
HEADER    LYASE                                   04-MAY-12   4F0H
TITLE     UNACTIVATED RUBISCO WITH OXYGEN BOUND
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN;
COMPND   3 CHAIN: A;
COMPND   4 SYNONYM: RUBISCO LARGE SUBUNIT;
COMPND   5 EC: 4.1.1.39;
COMPND   6 MOL_ID: 2;
COMPND   7 MOLECULE: RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN;
COMPND   8 CHAIN: B;
COMPND   9 SYNONYM: RUBISCO SMALL SUBUNIT;
COMPND  10 EC: 4.1.1.39
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: GALDIERIA SULPHURARIA;
SOURCE   3 ORGANISM_COMMON: RED ALGA;
SOURCE   4 ORGANISM_TAXID: 130081;
SOURCE   5 MOL_ID: 2;
SOURCE   6 ORGANISM_SCIENTIFIC: GALDIERIA SULPHURARIA;
SOURCE   7 ORGANISM_COMMON: RED ALGA;
SOURCE   8 ORGANISM_TAXID: 130081
KEYWDS    ALPHA BETA DOMAIN, CATALYTIC DOMAIN TIM BARREL,
KEYWDS   2 CARBOXYLASE/OXYGENASE, NITROSYLATION, CHLOROPLAST, LYASE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    B.STEC
REVDAT   2   12-DEC-12 4F0H    1       JRNL
REVDAT   1   14-NOV-12 4F0H    0
JRNL        AUTH   B.STEC
JRNL        TITL   STRUCTURAL MECHANISM OF RUBISCO ACTIVATION BY CARBAMYLATION
JRNL        TITL 2 OF THE ACTIVE SITE LYSINE.
JRNL        REF    PROC.NATL.ACAD.SCI.USA        V. 109 18785 2012
JRNL        REFN                   ISSN 0027-8424
JRNL        PMID   23112176
JRNL        DOI    10.1073/PNAS.1210754109
REMARK   2
REMARK   2 RESOLUTION.    1.96 ANGSTROMS.
REMARK   3
```

.cif

```
data_4F0H
#
_entry.id   4F0H
#
_audit_conform.dict_name       mmcif_pdbx.dic
_audit_conform.dict_version    5.281
_audit_conform.dict_location   http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
loop_
_database_2.database_id
_database_2.database_code
PDB   4F0H
RCSB  RCSB072299
WWPDB D_1000072299
#
loop_
_pdbx_database_related.db_name
_pdbx_database_related.db_id
_pdbx_database_related.details
_pdbx_database_related.content_type
PDB 4F0K . unspecified
PDB 4F0M . unspecified
#
_pdbx_database_status.status_code                     REL
_pdbx_database_status.entry_id                        4F0H
_pdbx_database_status.recvd_initial_deposition_date   2012-05-04
_pdbx_database_status.deposit_site                    RCSB
_pdbx_database_status.process_site                    RCSB
_pdbx_database_status.status_code_sf                  REL
_pdbx_database_status.status_code_mr                  ?
_pdbx_database_status.SG_entry                        ?
_pdbx_database_status.status_code_cs                  ?
_pdbx_database_status.methods_development_category    ?
_pdbx_database_status.pdb_format_compatible           Y
#
_audit_author.name           'Stec, B.'
_audit_author.pdbx_ordinal   1
#
_citation.id                        primary
_citation.title                     'Structural mechanism of RuBisCO activation by carbamylation of the active site lysine.'
_citation.journal_abbrev            Proc.Natl.Acad.Sci.USA
```

The header

- Contains much of experimental and documentation details
- Is only partially needed depending on the program used to read it
- For pymol:

.pdb

.cif

```
data_MyPolymerCIF
#
loop_
_atom_site.group_PDB
_atom_site.id
_atom_site.type_symbol
_atom_site.label_atom_id
_atom_site.label_alt_id
_atom_site.label_comp_id
_atom_site.label_asym_id
_atom_site.label_entity_id
_atom_site.label_seq_id
_atom_site.pdbx_PDB_ins_code
_atom_site.Cartn_x
_atom_site.Cartn_y
_atom_site.Cartn_z
_atom_site.occupancy
_atom_site.B_iso_or_equiv
_atom_site.pdbx_formal_charge
_atom_site.auth_seq_id
_atom_site.auth_comp_id
_atom_site.auth_asym_id
_atom_site.auth_atom_id
_atom_site.pdbx_PDB_model_num
```

The coordinates entry lines

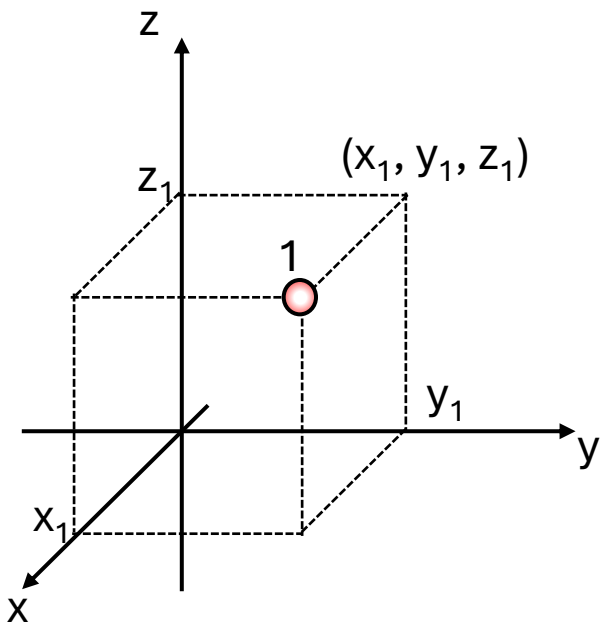
.pdb

```
ATOM      1  N    PRO A  27      18.254  54.186 -22.797  1.00 53.35           N
ATOM      2  CA   PRO A  27      17.690  52.844 -22.682  1.00 53.37           C
ATOM      3  C    PRO A  27      18.564  51.908 -21.783  1.00 53.24           C
ATOM      4  O    PRO A  27      19.115  50.895 -22.257  1.00 50.82           O
ATOM      5  CB   PRO A  27      17.659  52.369 -24.147  1.00 51.91           C
ATOM      6  CG   PRO A  27      18.673  53.271 -24.894  1.00 52.61           C
ATOM      7  CD   PRO A  27      19.231  54.261 -23.898  1.00 53.07           C
ATOM      8  N    TYR A  28      18.669  52.246 -20.493  1.00 51.56           N
ATOM      9  CA   TYR A  28      19.634  51.581 -19.613  1.00 49.67           C
ATOM     10  C    TYR A  28      19.310  50.171 -19.219  1.00 49.67           C
ATOM     11  O    TYR A  28      20.210  49.351 -19.134  1.00 49.06           O
ATOM     12  CB   TYR A  28      19.819  52.335 -18.325  1.00 48.01           C
ATOM     13  CG   TYR A  28      20.340  53.693 -18.509  1.00 48.13           C
ATOM     14  CD1  TYR A  28      21.705  53.916 -18.629  1.00 46.64           C
ATOM     15  CD2  TYR A  28      19.460  54.772 -18.573  1.00 48.85           C
ATOM     16  CE1  TYR A  28      22.204  55.188 -18.792  1.00 48.05           C
ATOM     17  CE2  TYR A  28      19.939  56.058 -18.741  1.00 50.76           C
ATOM     18  CZ   TYR A  28      21.319  56.261 -18.850  1.00 50.78           C
ATOM     19  OH   TYR A  28      21.781  57.551 -19.020  1.00 52.62           O
```

.cif

```
ATOM   1    N N   . PRO A 1 27  ? 18.254 54.186 -22.797 1.00 53.35 ? 27   PRO A N   1
ATOM   2    C CA  . PRO A 1 27  ? 17.690 52.844 -22.682 1.00 53.37 ? 27   PRO A CA  1
ATOM   3    C C   . PRO A 1 27  ? 18.564 51.908 -21.783 1.00 53.24 ? 27   PRO A C   1
ATOM   4    O O   . PRO A 1 27  ? 19.115 50.895 -22.257 1.00 50.82 ? 27   PRO A O   1
ATOM   5    C CB  . PRO A 1 27  ? 17.659 52.369 -24.147 1.00 51.91 ? 27   PRO A CB  1
ATOM   6    C CG  . PRO A 1 27  ? 18.673 53.271 -24.894 1.00 52.61 ? 27   PRO A CG  1
ATOM   7    C CD  . PRO A 1 27  ? 19.231 54.261 -23.898 1.00 53.07 ? 27   PRO A CD  1
ATOM   8    N N   . TYR A 1 28  ? 18.669 52.246 -20.493 1.00 51.56 ? 28   TYR A N   1
ATOM   9    C CA  . TYR A 1 28  ? 19.634 51.581 -19.613 1.00 49.67 ? 28   TYR A CA  1
ATOM   10   C C   . TYR A 1 28  ? 19.310 50.171 -19.219 1.00 49.67 ? 28   TYR A C   1
ATOM   11   O O   . TYR A 1 28  ? 20.210 49.351 -19.134 1.00 49.06 ? 28   TYR A O   1
ATOM   12   C CB  . TYR A 1 28  ? 19.819 52.335 -18.325 1.00 48.01 ? 28   TYR A CB  1
ATOM   13   C CG  . TYR A 1 28  ? 20.340 53.693 -18.509 1.00 48.13 ? 28   TYR A CG  1
ATOM   14   C CD1 . TYR A 1 28  ? 21.705 53.916 -18.629 1.00 46.64 ? 28   TYR A CD1 1
ATOM   15   C CD2 . TYR A 1 28  ? 19.460 54.772 -18.573 1.00 48.85 ? 28   TYR A CD2 1
ATOM   16   C CE1 . TYR A 1 28  ? 22.204 55.188 -18.792 1.00 48.05 ? 28   TYR A CE1 1
ATOM   17   C CE2 . TYR A 1 28  ? 19.939 56.058 -18.741 1.00 50.76 ? 28   TYR A CE2 1
ATOM   18   C CZ  . TYR A 1 28  ? 21.319 56.261 -18.850 1.00 50.78 ? 28   TYR A CZ  1
ATOM   19   O OH  . TYR A 1 28  ? 21.781 57.551 -19.020 1.00 52.62 ? 28   TYR A OH  1
```

Cartesian coordinates



ATOM 1 C CA 1 A x_1 y_1 z_1

The coordinates entry lines

.pdb

1. Entry type
2. Index
3. Atom type
4. amino acid residue type
5. Chain name
6. Amino acid residue number
7. X coordinate
8. Y coordinate
9. Z coordinate
10. Occupancy
11. Temperature factor
12. General atom type

.cif

1. Entry type
2. Index
3. General atom type
4. Atom type
5. Alternate conformation
6. amino acid residue type
7. Chain name
8. Chain number
9. Amino acid residue number
10. PDB insertion code
11. X coordinate
12. Y coordinate
13. Z coordinate
14. Occupancy
15. Temperature factor
16. Net integer charge
17. Author‘s amino acid residue number
18. Author‘s chain name
19. Author‘s chain number
20. Author‘s atom type
21. Model number

https://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Index/index.html

Indicators of structural flexibility

- Occupancy - percentage among alternate conformations (sum up 1)
- Temperature Factors (B-value) - amount of smearing of the electron density of the atom – AlaphaFold‘s confidence (plDDT)
- Model number – for solution NMR structures
- Missing atoms - .cif lists them with ?, .pdb skips them

Most common other entry lines

.pdb

```
HETATM 4685  O1  OXY A 501      32.806  71.080  19.591  1.00 36.49           O
HETATM 4686  O2  OXY A 501      32.767  71.346  18.525  1.00 30.70           O
HETATM 4687  P   PO4 A 502      28.002  69.877  10.767  0.60 36.70           P
HETATM 4688  O1  PO4 A 502      27.636  69.337  12.136  0.60 34.88           O
HETATM 4689  O2  PO4 A 502      26.823  70.597  10.165  0.60 36.45           O
HETATM 4690  O3  PO4 A 502      29.132  70.873  10.871  0.60 37.20           O
HETATM 4691  O4  PO4 A 502      28.481  68.732   9.914  0.60 35.05           O
HETATM 4692  P   PO4 A 503      28.843  73.673  23.642  0.60 26.50           P
HETATM 4693  O1  PO4 A 503      27.707  73.193  22.759  0.60 28.85           O
HETATM 4694  O2  PO4 A 503      28.624  75.119  24.044  0.60 28.55           O
HETATM 4695  O3  PO4 A 503      28.823  72.842  24.897  0.60 27.15           O
HETATM 4696  O4  PO4 A 503      30.139  73.552  22.905  0.60 26.19           O
HETATM 4697  O   HOH A1001      42.027  54.564  33.635  1.00 11.57           O
HETATM 4698  O   HOH A1002      41.224  48.003  -7.265  1.00  5.70           O
HETATM 4699  O   HOH A1003      50.970  54.849  33.382  1.00  4.39           O
HETATM 4700  O   HOH A1004      41.029  57.318  33.353  1.00  5.41           O
```

.cif

```
HETATM 4683 O O1  . OXY C 3 .   ? 32.806 71.080 19.591  1.00 36.49 ? 501  OXY A O1  1
HETATM 4684 O O2  . OXY C 3 .   ? 32.767 71.346 18.525  1.00 30.70 ? 501  OXY A O2  1
HETATM 4685 P P   . PO4 D 4 .   ? 28.002 69.877 10.767  0.60 36.70 ? 502  PO4 A P   1
HETATM 4686 O O1  . PO4 D 4 .   ? 27.636 69.337 12.136  0.60 34.88 ? 502  PO4 A O1  1
HETATM 4687 O O2  . PO4 D 4 .   ? 26.823 70.597 10.165  0.60 36.45 ? 502  PO4 A O2  1
HETATM 4688 O O3  . PO4 D 4 .   ? 29.132 70.873 10.871  0.60 37.20 ? 502  PO4 A O3  1
HETATM 4689 O O4  . PO4 D 4 .   ? 28.481 68.732 9.914   0.60 35.05 ? 502  PO4 A O4  1
HETATM 4690 P P   . PO4 E 4 .   ? 28.843 73.673 23.642  0.60 26.50 ? 503  PO4 A P   1
HETATM 4691 O O1  . PO4 E 4 .   ? 27.707 73.193 22.759  0.60 28.85 ? 503  PO4 A O1  1
HETATM 4692 O O2  . PO4 E 4 .   ? 28.624 75.119 24.044  0.60 28.55 ? 503  PO4 A O2  1
HETATM 4693 O O3  . PO4 E 4 .   ? 28.823 72.842 24.897  0.60 27.15 ? 503  PO4 A O3  1
HETATM 4694 O O4  . PO4 E 4 .   ? 30.139 73.552 22.905  0.60 26.19 ? 503  PO4 A O4  1
HETATM 4695 O O   . HOH F 5 .   ? 42.027 54.564 33.635  1.00 11.57 ? 1001 HOH A O   1
HETATM 4696 O O   . HOH F 5 .   ? 41.224 48.003 -7.265  1.00 5.70  ? 1002 HOH A O   1
HETATM 4697 O O   . HOH F 5 .   ? 50.970 54.849 33.382  1.00 4.39  ? 1003 HOH A O   1
HETATM 4698 O O   . HOH F 5 .   ? 41.029 57.318 33.353  1.00 5.41  ? 1004 HOH A O   1
HETATM 4699 O O   . HOH F 5 .   ? 47.279 88.807 28.394  1.00 6.48  ? 1005 HOH A O   1
HETATM 4700 O O   . HOH F 5 .   ? 31.581 46.787 -12.605 1.00 8.32  ? 1006 HOH A O   1
```

Heteroatoms

Atoms from molecules different than amino acids or nucleic acids

Most common other entry lines

.pdb

```
ATOM      1  N   ALA A   1      45.202  42.596  18.258  1.00 25.18           N
ANISOU    1  N   ALA A   1    3206   3173   3186     12     -9     -8       N
ATOM      2  CA  ALA A   1      45.411  43.522  19.407  1.00 24.86           C
ANISOU    2  CA  ALA A   1    3148   3149   3146      2      3      0       C
```

.cif

```
#
loop_
_atom_site_anisotrop.id
_atom_site_anisotrop.type_symbol
_atom_site_anisotrop.pdbx_label_atom_id
_atom_site_anisotrop.pdbx_label_alt_id
_atom_site_anisotrop.pdbx_label_comp_id
_atom_site_anisotrop.pdbx_label_asym_id
_atom_site_anisotrop.pdbx_label_seq_id
_atom_site_anisotrop.pdbx_PDB_ins_code
_atom_site_anisotrop.U[1][1]
_atom_site_anisotrop.U[2][2]
_atom_site_anisotrop.U[3][3]
_atom_site_anisotrop.U[1][2]
_atom_site_anisotrop.U[1][3]
_atom_site_anisotrop.U[2][3]
_atom_site_anisotrop.pdbx_auth_seq_id
_atom_site_anisotrop.pdbx_auth_comp_id
_atom_site_anisotrop.pdbx_auth_asym_id
_atom_site_anisotrop.pdbx_auth_atom_id
1    N N   . ALA A 2   ? 0.3206 0.3173 0.3186 0.0012  -0.0009 -0.0008 1   ALA A N
2    C CA  . ALA A 2   ? 0.3148 0.3149 0.3146 0.0002  0.0003  0.0000  1   ALA A CA
```

Anisotropic Atomic Displacement Parameters

- High resolution structures may give more details on flexibility
- 6 values to indicate atomic displacements in 3D

Let‘s see these coordinates in PyMOL

What can we expect from AlphaFold?

- Regarding nucleic acids, membrane proteins or ligands?
- Regarding complexes (multimeric proteins)? And big complexes?
- Regarding flexibility?
- Regarding single mutations?
- Did AlphaFold solve the folding mechanism?

Alphafold outputs many models

	initial training	first fine-tuning		second fine-tuning				
Model	1	1.1	1.2	**1.1.1**	**1.1.2**	**1.2.1**	**1.2.2**	**1.2.3**
Parameters initialized from	Random	Model 1	···	Model 1.1	···	Model 1.2	···	···
Number of templates N_{templ}	4	4	0	4	···	0	···	···
Sequence crop size N_{res}	256	···	···	384	···	···	···	···
Number of sequences N_{seq}	128	512	···	···	···	···	···	···
Number of extra sequences $N_{\text{extra\_seq}}$	1024	···	···	5120	1024	5120	···	1024
Initial learning rate	10^{-3}	$5 \cdot 10^{-4}$	···	···	···	···	···	···
Learning rate linear warm-up samples	128000	0	···	···	···	···	···	···
Structural violation loss weight	0.0	1.0	···	···	···	···	···	···
"Experimentally resolved" loss weight	0.0	0.01	···	···	···	···	···	···
Training samples ($\cdot 10^6$)	9.2	1.1	1.7	0.3	0.6	1.4	1.1	2.4
Training time	6d 6h	1d 10h	2d 3h	20h	1d 13h	4d 1h	3d	5d 12h

5 models

AlphaFold Multimer runs each of the 5 models with 5 different random seeds for MSA sampling, resulting in 25 different models.

equations

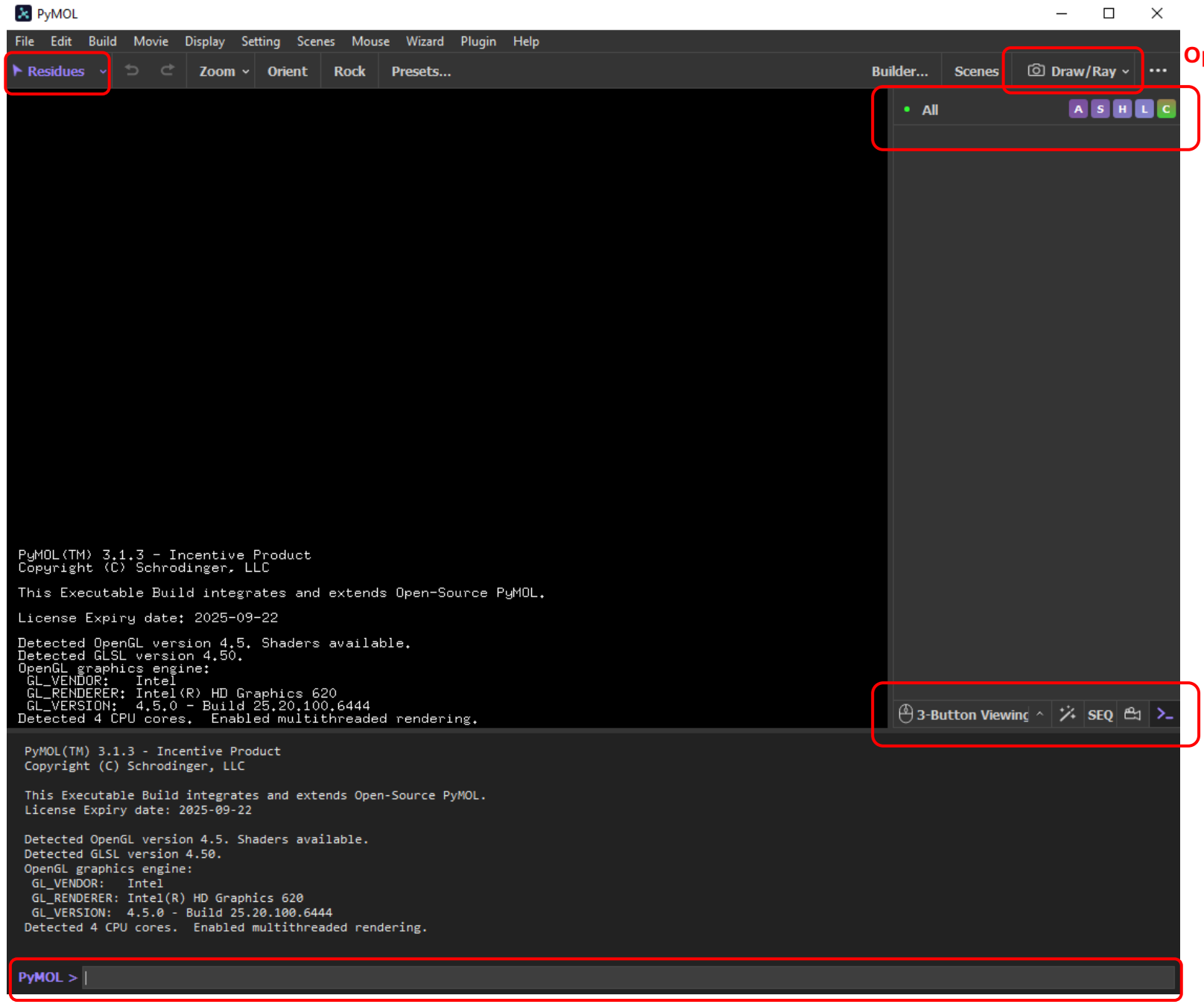
- Rmsd

$$\mathrm{RMSD}(\mathbf{v}, \mathbf{w}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|v_i - w_i\|^2}$$
$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n} ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}$$

- Tm score

$$\text{TM-score} = \max\left[\frac{1}{L_{\text{target}}}\sum_{i}^{L_{\text{common}}}\frac{1}{1+\left(\frac{d_i}{d_0(L_{\text{target}})}\right)^2}\right]$$

- https://en.wikipedia.org/wiki/Template_modeling_score
- iptm https://www.biorxiv.org/content/10.1101/2021.10.04.463034v1.full.pdf



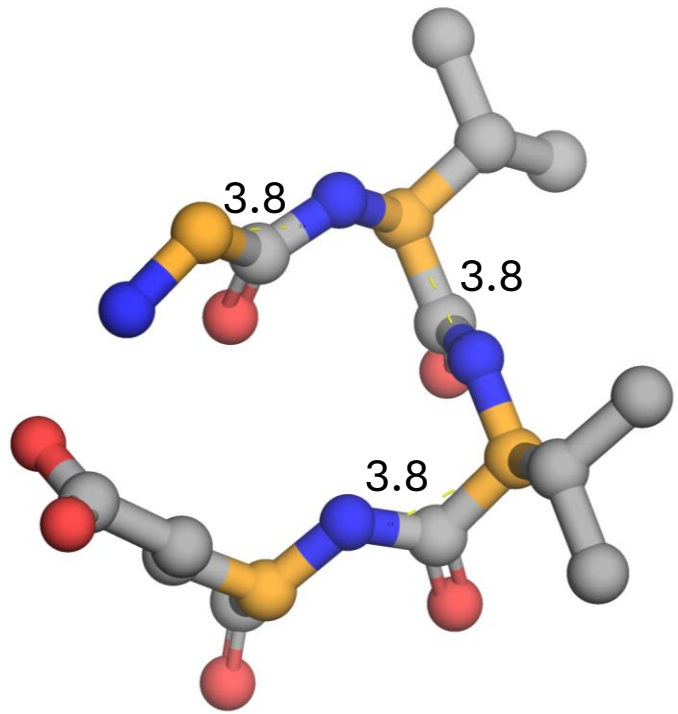
More view options

Options for image rendering

Action
Show
Hide
Label
Colour
for each object

Mouse and keys options,
in view or edit mode
Wizard shortcuts
Show or hide sequence
Movie editing options

Command line


3.8
3.8
3.8