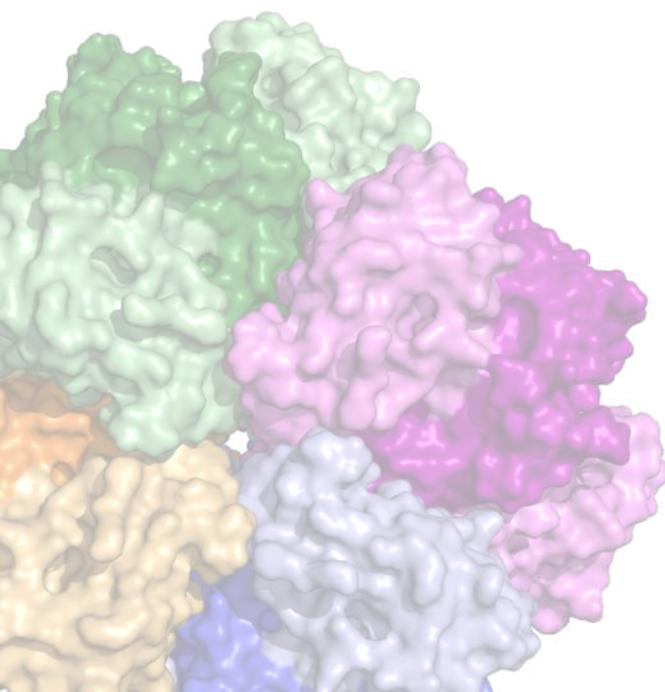


Protein Structure Workshop

In Silico Analysis

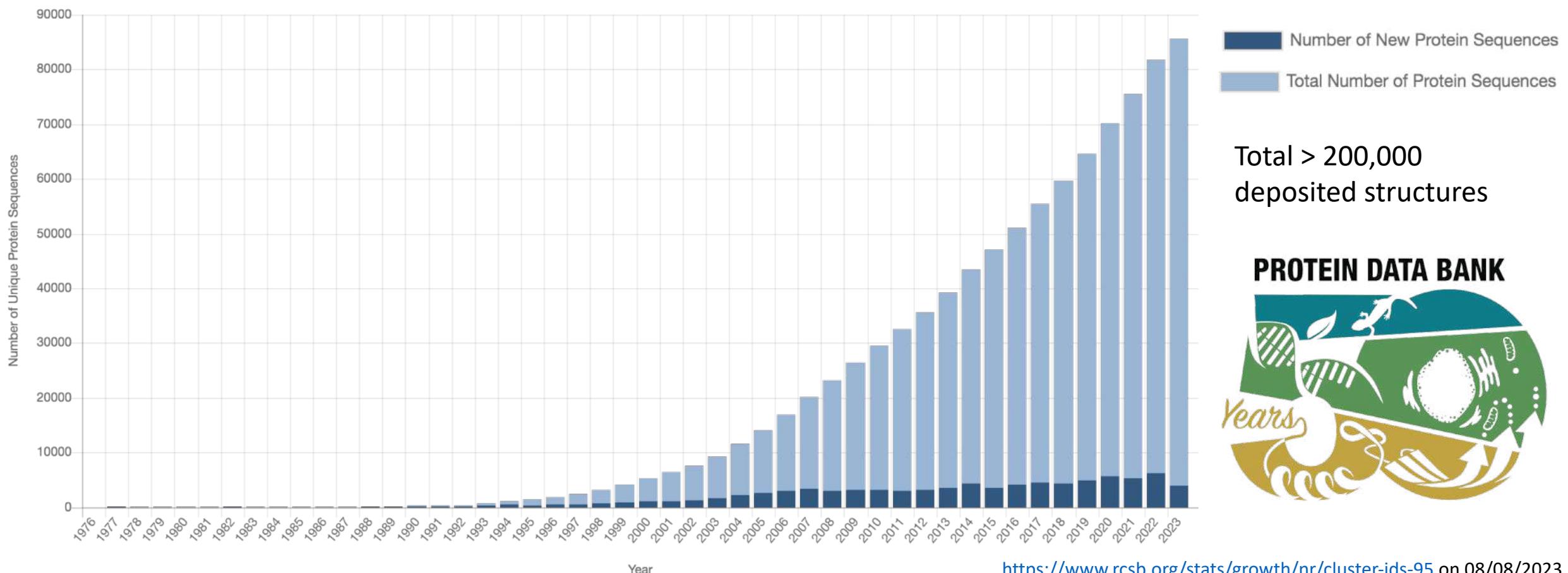
Amanda Souza Câmara



From 18th to 22th September

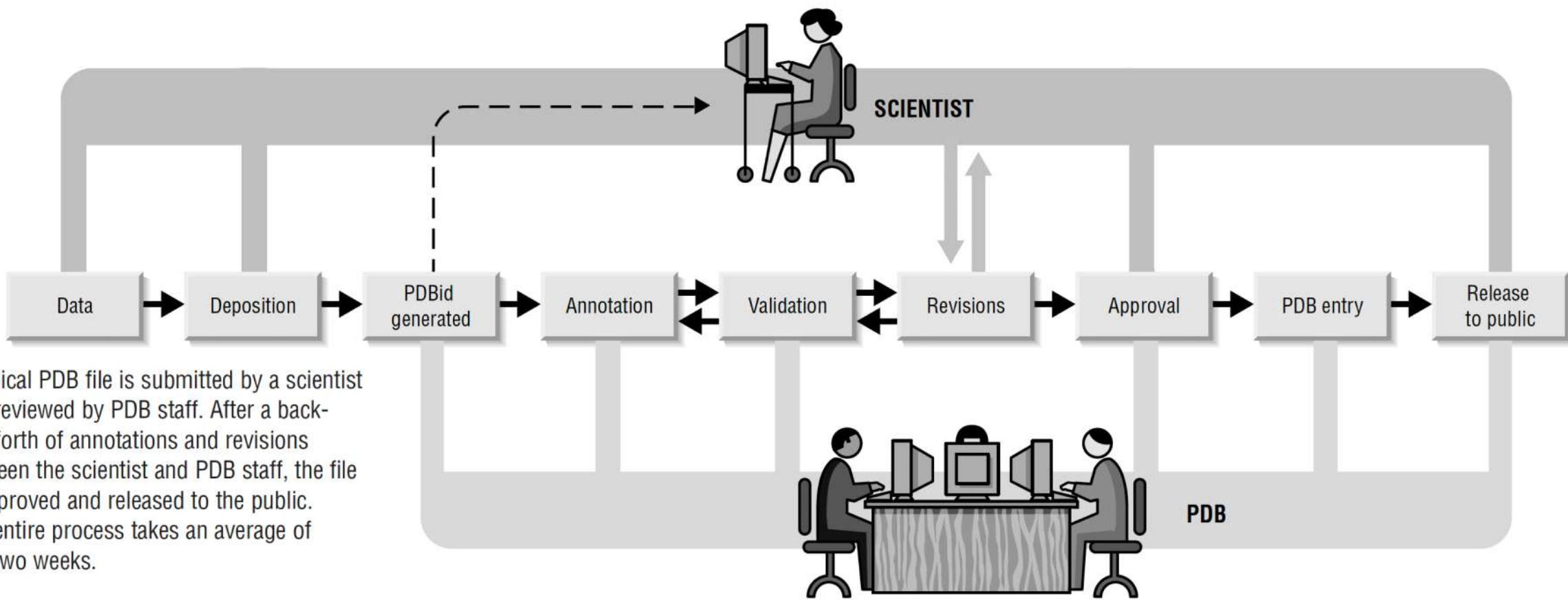
Protein structure as a digital resource

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



The PDB entry

Steps in the Flow of Data Through the PDB



The PDB browser



NCSA Mosaic: Document View

File Options Navigate Annotate Documents

Document Title: The WWW Version of the PDB Browser

Document URL: <http://www.pdb.bnl.gov/cgi-bin/browse>

PDBBrowse

This is the PDB Browser, providing fast access to the entire structure data base. [Help](#) is available at the click of a mouse. (Please note - some Macintosh browsers have trouble with this page. You may have to try Mosaic, MacWeb, or Netscape.) If you are having problems with a search involving the characters *, ?, or +, please see this [description of regular expressions](#).

Display Options

[List Compound](#) [Sort by ID](#) [And Constraints](#)
[Apply Constraints](#) [Search Full PDB](#)

Search Criteria [Send Request](#) (See next section for results.)

ID: Class: Compound:
Author: HAREL Source: Experiment:
Journal: Cryst1: Het:
Formula:
Resolution Worse Than Resolution Better Than

Please see the display of matching ID-Codes below

List of ID Codes [Fetch the Entry with ID Code!](#)

1ace ACETYLCHOLINESTERASE (E.C.3.1.1.7)
1acj ACETYLCHOLINESTERASE (E.C.3.1.1.7) COMPLEXED WITH TACrine
1ack ACETYLCHOLINESTERASE (E.C.3.1.1.7) COMPLEXED WITH EDROPHONIUM
1aci ACETYLCHOLINESTERASE (E.C.3.1.1.7) COMPLEXED WITH DECAMETHONIUM
1gcd GAMMA CHYMOTRYPSIN (E.C.3.4.21.1) COMPLEXED WITH DIETHYL PHOSPHORYL ...

Dave Stampf/PDB Sr. Project Mgr. -- drs@bnl.gov

Back Forward Home Reload Open Save As... Clone New Window Close Window

Protein structure as a digital resource

- 2003 was announced the worldwide Protein Data Bank - wwPDB



The PDB website

RCSB PDB Deposit Search Visualize Analyze Download Learn About Documentation Careers COVID-19 MyPDB Contact us

PDB PROTEIN DATA BANK 208,347 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 CPDB EMDResource NAKB wwPDB Foundation PDB-Dev

New: More Computed Structure Models (CSM) available Learn more

Welcome Deposit Search Visualize Analyze Download Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:
Experimentally-determined 3D structures from the Protein Data Bank (PDB) archive
Computed Structure Models (CSM) from AlphaFold DB and ModelArchive

These data can be explored in context of external annotations providing a structural view of biology.

Explore NEW Features PDB-101 Training Resources

August Molecule of the Month

ATM and ATR Kinases

Latest Entries As of Tue Aug 08 2023

8G4E Green Fluorescence Protein imaged on a cryo-EM imaging scaffold

Features & Highlights

Updated Annotation and Standardization of Peptide Residues In October 2023, wwPDB will roll out updated CCD data files with additional annotation and standardized atom naming of peptide residues.

PDB NextGen Archive Now Provides Intra-molecular Connectivity With this release, intra-molecular connectivity for each residue present in an entry has been provided to help users transitioning from legacy PDB format to PDBx/mmCIF

DNS DNS name changes for PDB archive downloads from RCSB PDB starting September 2023 Programmatic users (ftp, rsync or https) should update scripts as soon as

News

Bragg Your Pattern at IUCr Bragg Your Pattern has something for everyone: school-aged children and their teachers/parents; IUCr attendees; and all structural biology enthusiasts + 08/08/2023

New Poster Available for Download Shiga Toxin 2 in Complex with Ribosomal P-stalk + 08/01/2023

Summer Newsletter Published In this issue: Explore Bioenergy; Upload Structure Files to Search; Preparing PDB Depositions; more. In the Education Corner, learn about Empowering Educators with Research-Grade

PDB at a Glance 63,919 Structures of Human Sequences 16,507 Nucleic Acid Containing Structures More Statistics

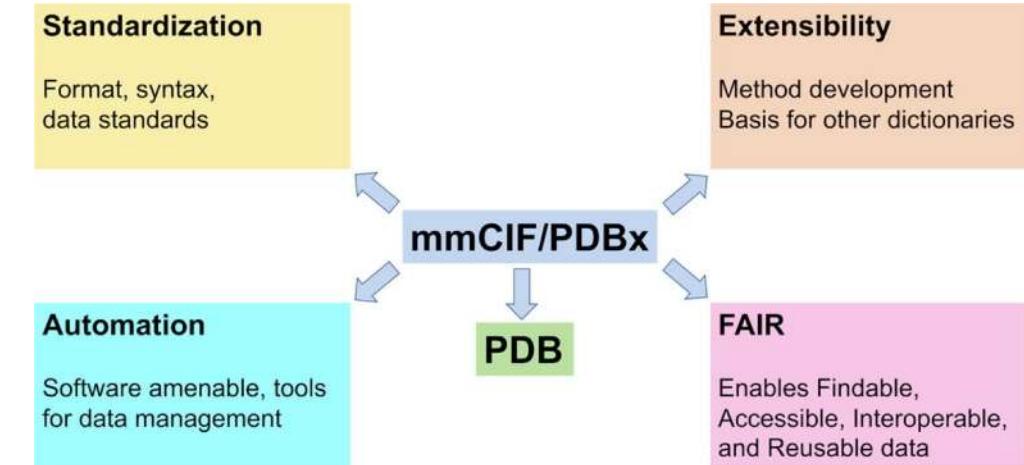
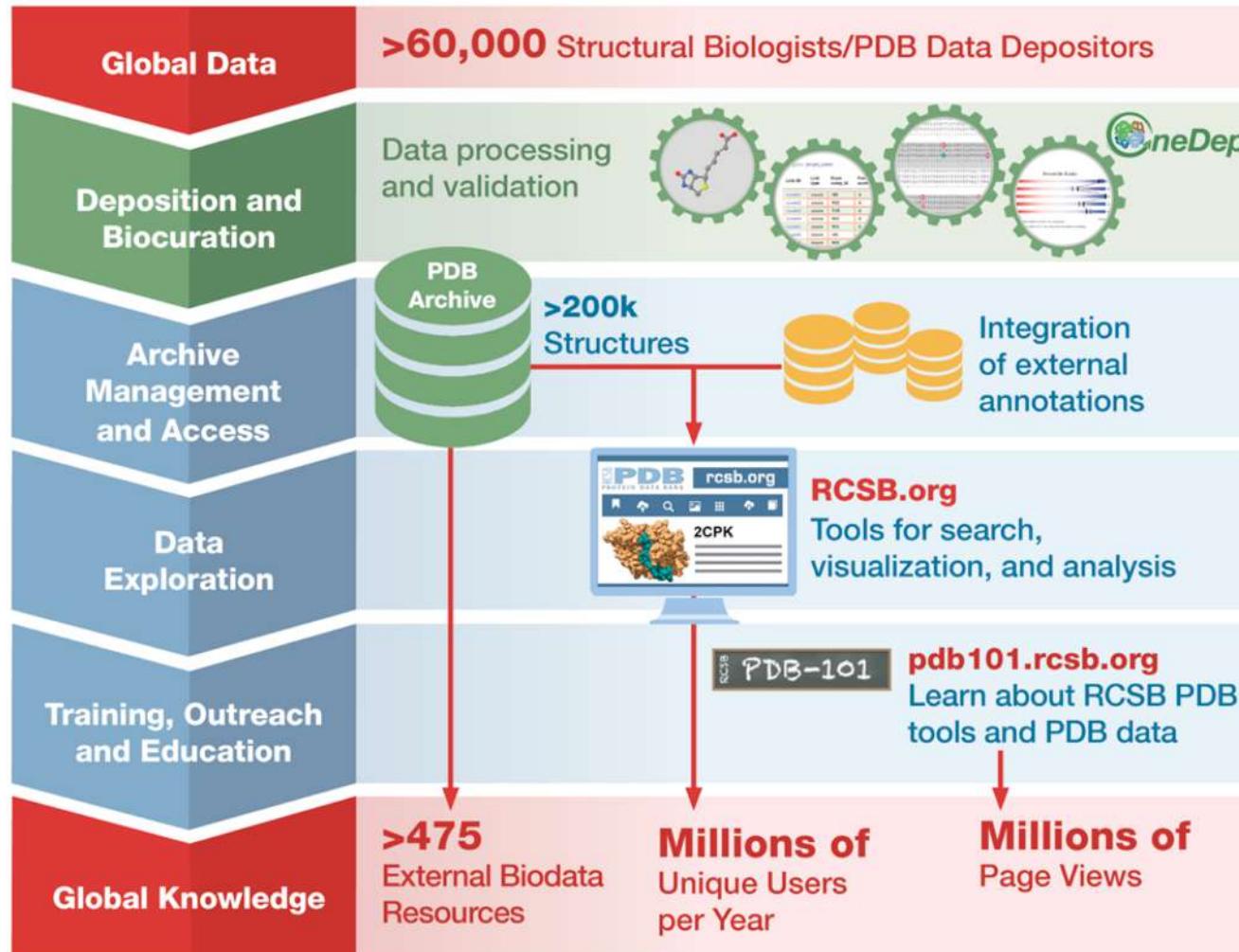
CSM at a Glance 999,251 AlphaFoldDB 69,326 ModelArchive

<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 Crystallographic 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



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
WPDB 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

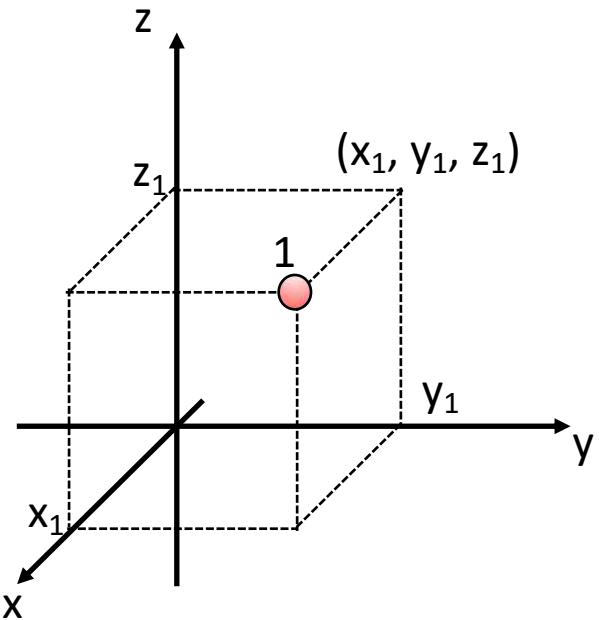
- Contains much of experimental and documentation details
- Is only partially needed depending on the program used to read it
- For pymol:

.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
```

Cartesian coordinates



ATOM 1 C CA 1 A x_1 y_1 z_1

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

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

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
- Model number – for solution NMR structures
- Missing atoms - .cif lists them with ?, .pdb skips them

Most common other entry lines

.pdb

HETATM	4685	01	OXY	A	501	32.806	71.080	19.591	1.00	36.49	0
HETATM	4686	02	OXY	A	501	32.767	71.346	18.525	1.00	30.70	0
HETATM	4687	P	P04	A	502	28.002	69.877	10.767	0.60	36.70	P
HETATM	4688	01	P04	A	502	27.636	69.337	12.136	0.60	34.88	0
HETATM	4689	02	P04	A	502	26.823	70.597	10.165	0.60	36.45	0
HETATM	4690	03	P04	A	502	29.132	70.873	10.871	0.60	37.20	0
HETATM	4691	04	P04	A	502	28.481	68.732	9.914	0.60	35.05	0
HETATM	4692	P	P04	A	503	28.843	73.673	23.642	0.60	26.50	P
HETATM	4693	01	P04	A	503	27.707	73.193	22.759	0.60	28.85	0
HETATM	4694	02	P04	A	503	28.624	75.119	24.044	0.60	28.55	0
HETATM	4695	03	P04	A	503	28.823	72.842	24.897	0.60	27.15	0
HETATM	4696	04	P04	A	503	30.139	73.552	22.905	0.60	26.19	0
HETATM	4697	0	HOH	A	1001	42.027	54.564	33.635	1.00	11.57	0
HETATM	4698	0	HOH	A	1002	41.224	48.003	-7.265	1.00	5.70	0
HETATM	4699	0	HOH	A	1003	50.970	54.849	33.382	1.00	4.39	0
HETATM	4700	0	HOH	A	1004	41.029	57.318	33.353	1.00	5.41	0

.cif

HETATM	4683	0	01	.	OXY	C	3	.	?	32.806	71.080	19.591	1.00	36.49	?	501	OXY	A	01	1
HETATM	4684	0	02	.	OXY	C	3	.	?	32.767	71.346	18.525	1.00	30.70	?	501	OXY	A	02	1
HETATM	4685	P	P	.	P04	D	4	.	?	28.002	69.877	10.767	0.60	36.70	?	502	P04	A	P	1
HETATM	4686	0	01	.	P04	D	4	.	?	27.636	69.337	12.136	0.60	34.88	?	502	P04	A	01	1
HETATM	4687	0	02	.	P04	D	4	.	?	26.823	70.597	10.165	0.60	36.45	?	502	P04	A	02	1
HETATM	4688	0	03	.	P04	D	4	.	?	29.132	70.873	10.871	0.60	37.20	?	502	P04	A	03	1
HETATM	4689	0	04	.	P04	D	4	.	?	28.481	68.732	9.914	0.60	35.05	?	502	P04	A	04	1
HETATM	4690	P	P	.	P04	E	4	.	?	28.843	73.673	23.642	0.60	26.50	?	503	P04	A	P	1
HETATM	4691	0	01	.	P04	E	4	.	?	27.707	73.193	22.759	0.60	28.85	?	503	P04	A	01	1
HETATM	4692	0	02	.	P04	E	4	.	?	28.624	75.119	24.044	0.60	28.55	?	503	P04	A	02	1
HETATM	4693	0	03	.	P04	E	4	.	?	28.823	72.842	24.897	0.60	27.15	?	503	P04	A	03	1
HETATM	4694	0	04	.	P04	E	4	.	?	30.139	73.552	22.905	0.60	26.19	?	503	P04	A	04	1
HETATM	4695	0	0	.	HOH	F	5	.	?	42.027	54.564	33.635	1.00	11.57	?	1001	HOH	A	0	1
HETATM	4696	0	0	.	HOH	F	5	.	?	41.224	48.003	-7.265	1.00	5.70	?	1002	HOH	A	0	1
HETATM	4697	0	0	.	HOH	F	5	.	?	50.970	54.849	33.382	1.00	4.39	?	1003	HOH	A	0	1
HETATM	4698	0	0	.	HOH	F	5	.	?	41.029	57.318	33.353	1.00	5.41	?	1004	HOH	A	0	1
HETATM	4699	0	0	.	HOH	F	5	.	?	47.279	88.807	28.394	1.00	6.48	?	1005	HOH	A	0	1
HETATM	4700	0	0	.	HOH	F	5	.	?	31.581	46.787	-12.605	1.00	8.32	?	1006	HOH	A	0	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

Introduction to protein flexibility

**“Everything that living things do
can be understood in terms of the
wigglings and jigglings of atoms.”**

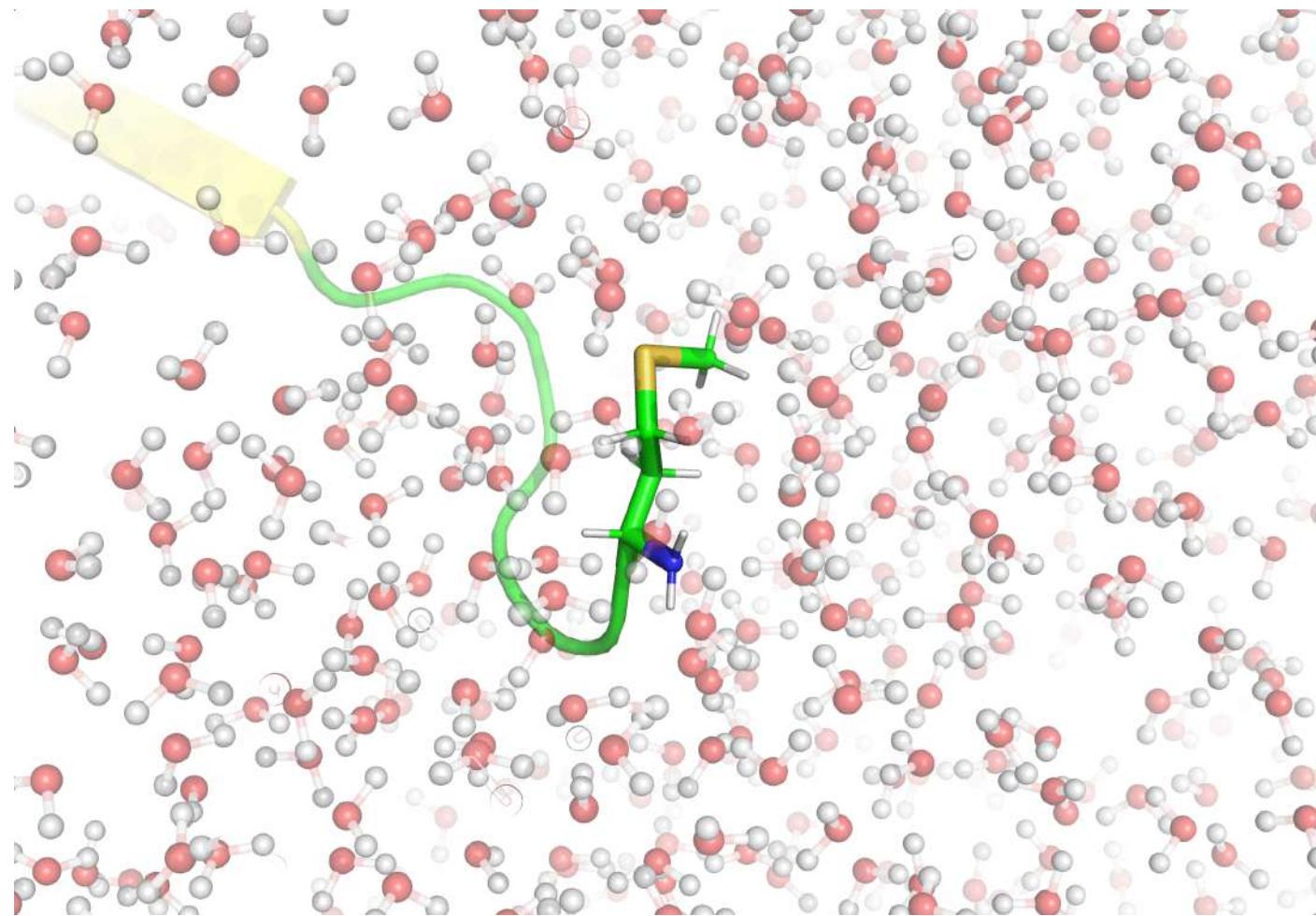


Richard Feynman

Fun to imagine

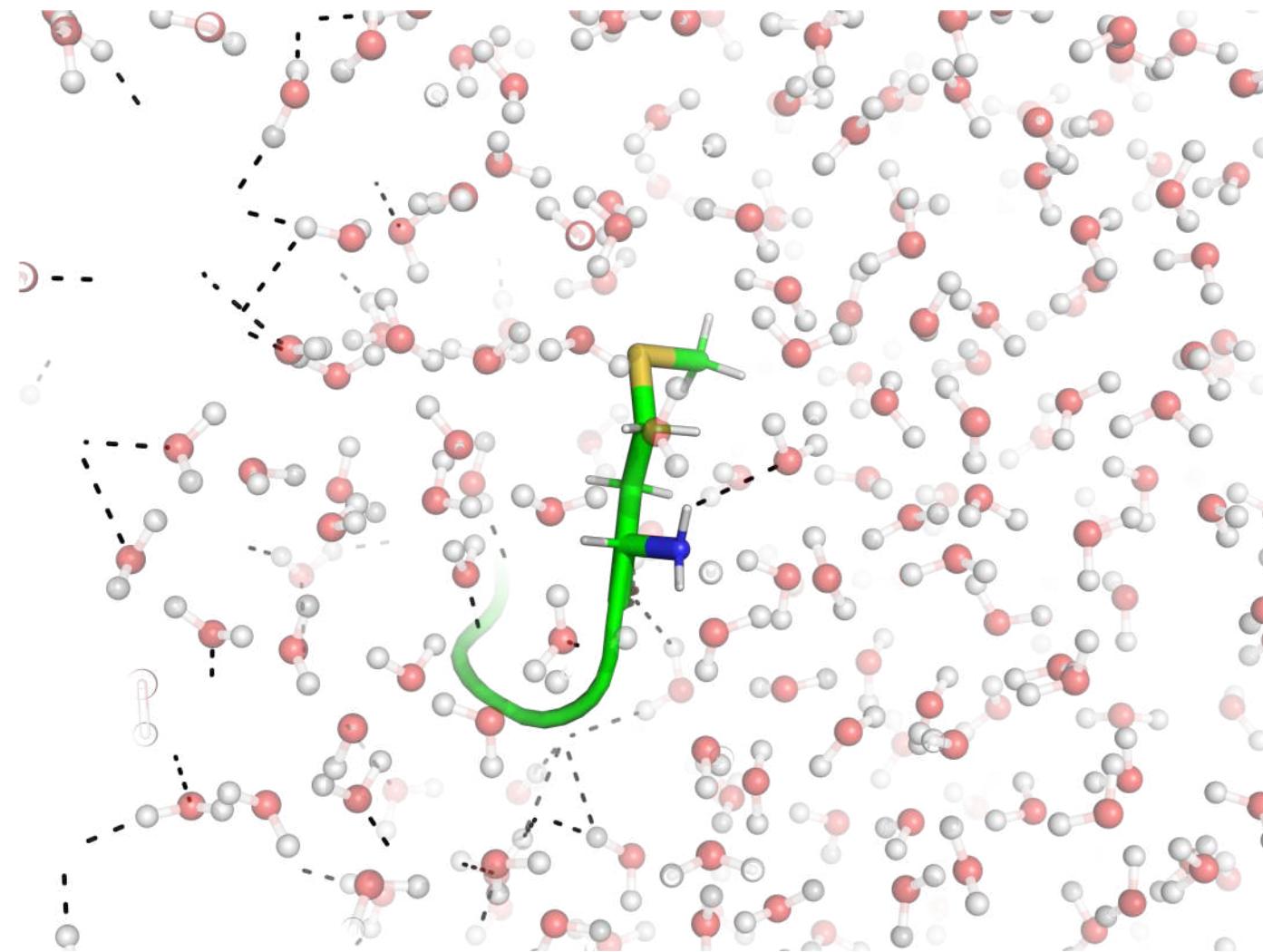
<https://www.youtube.com/watch?v=v3pYRn5j7oI&list=PL04B3F5636096478C>

Wigglings and jigglings of atoms



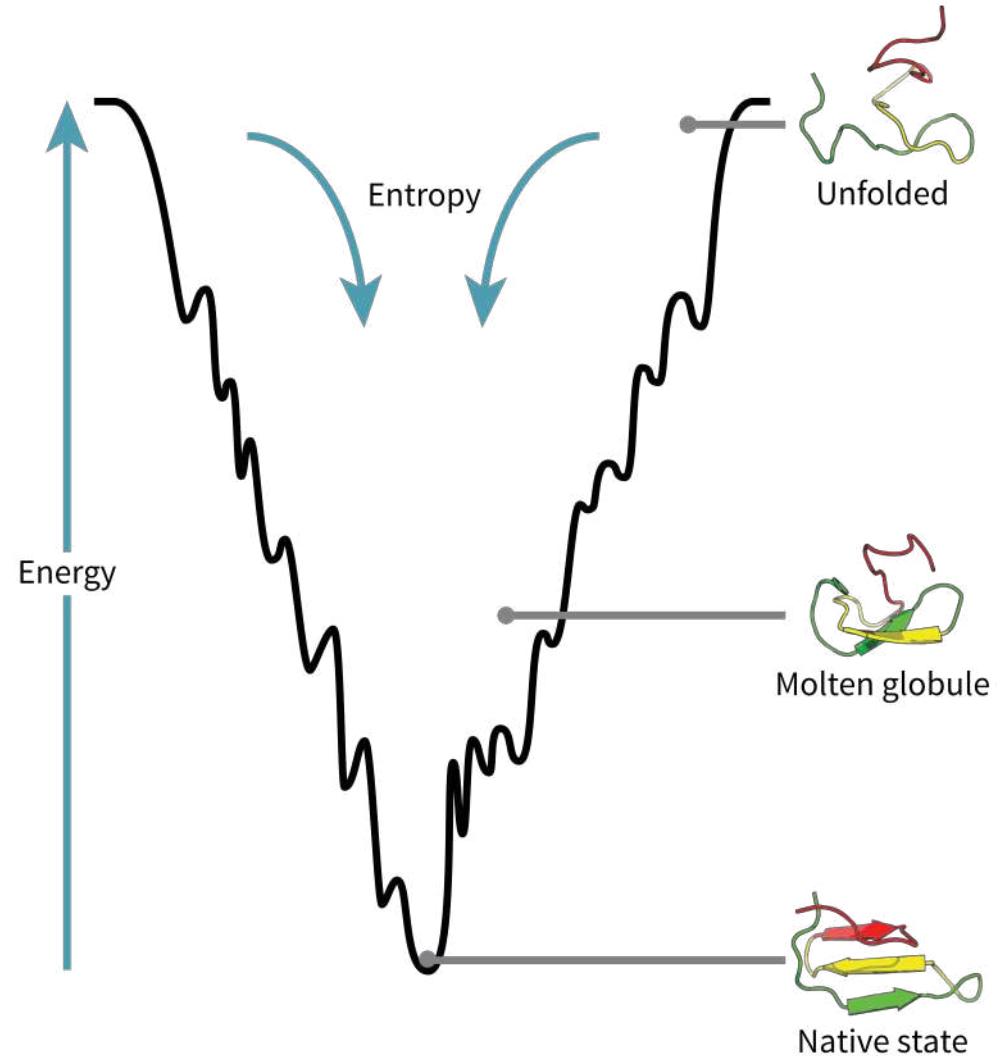
... but 10^{10} times faster!

Atomic forces

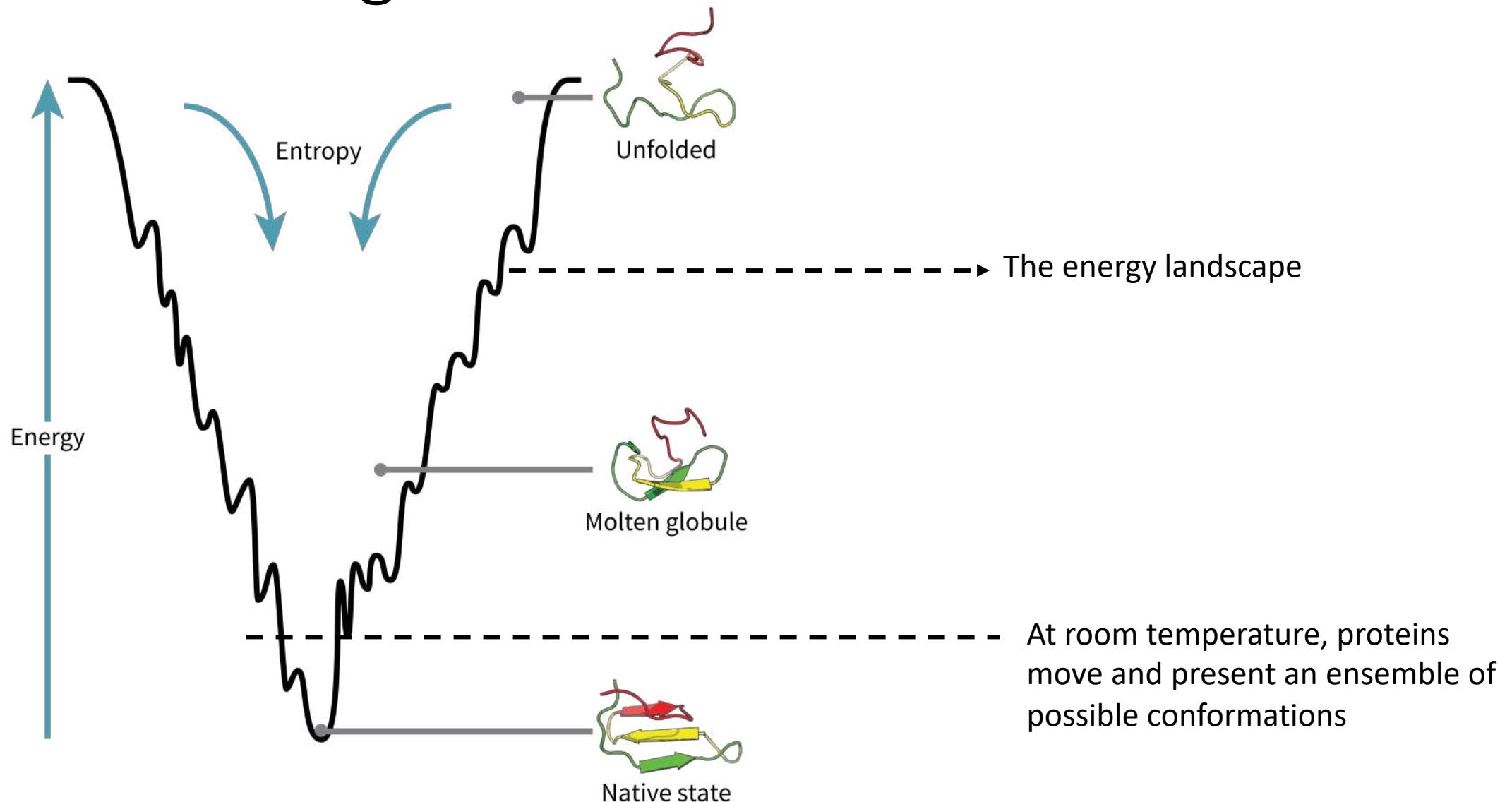


$$\overline{E_K} = \frac{1}{2} m \bar{v}^2 = \frac{3}{2} k_B T$$

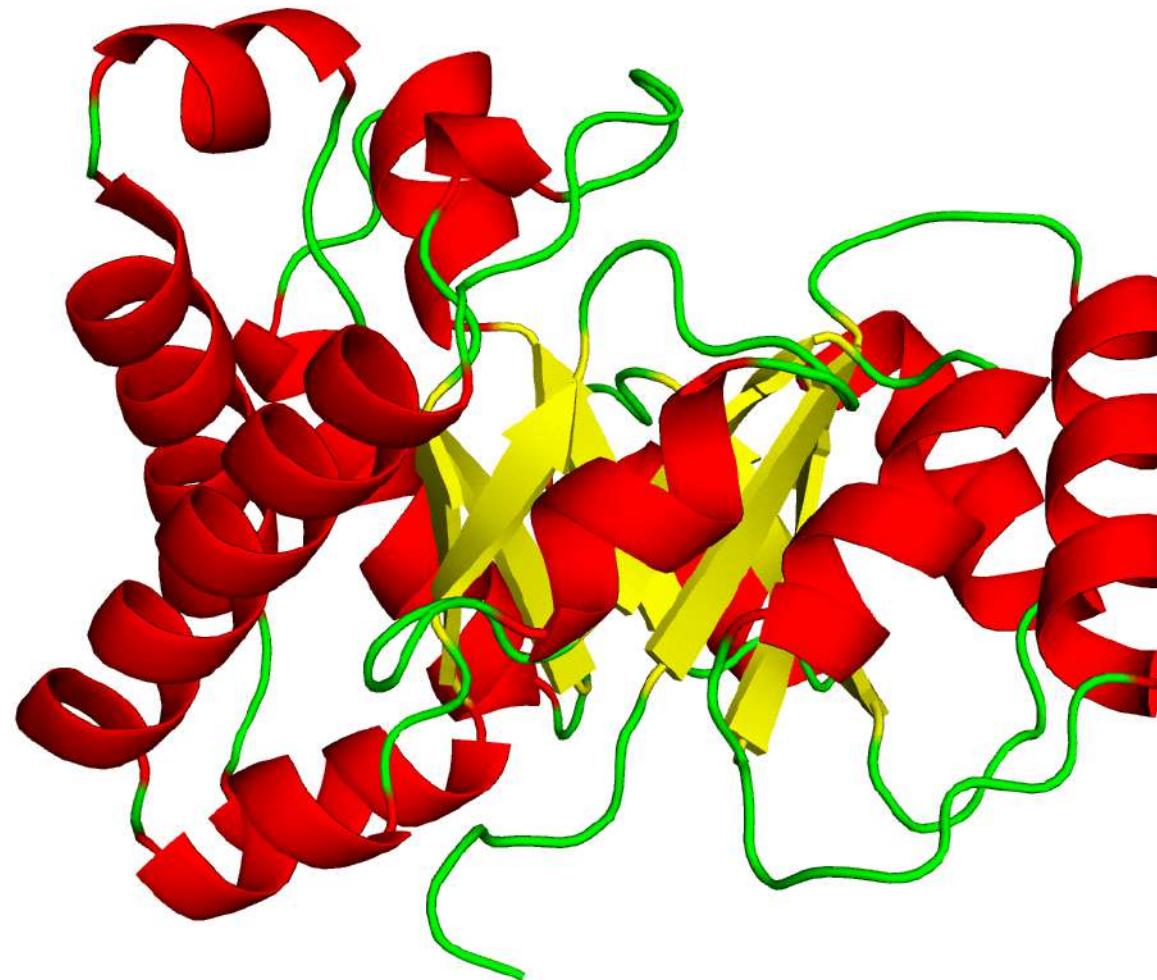
- The higher the temperature the more agitated are the molecules
- Interactions forming secondary structures are lost with higher temperature



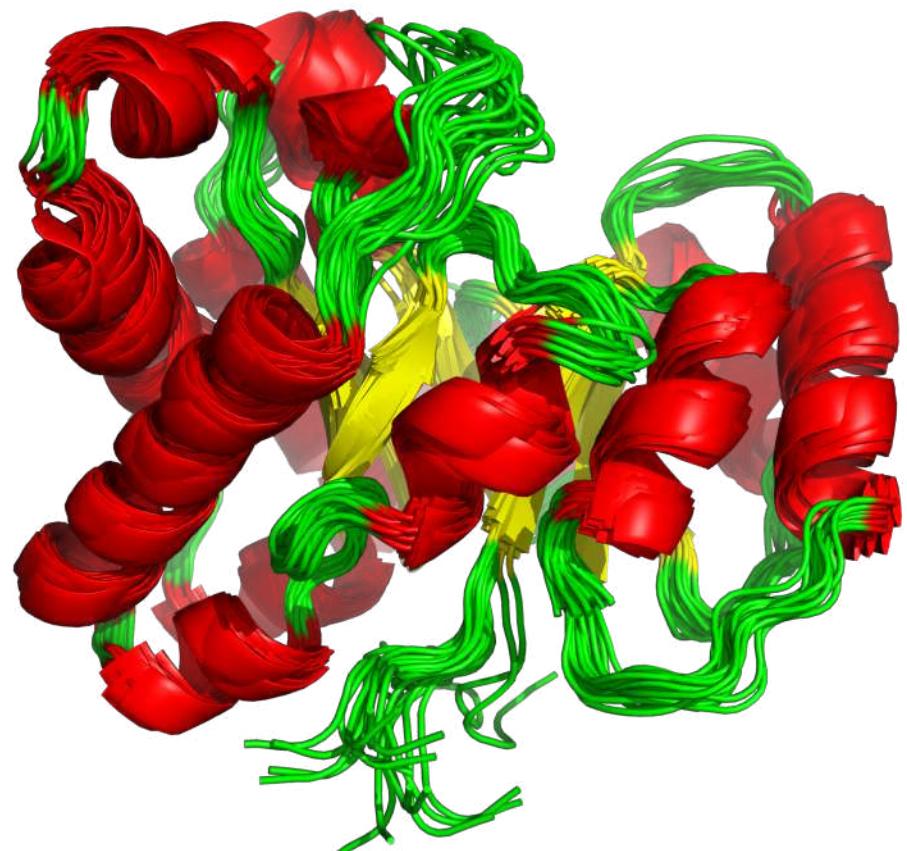
Protein Folding Funnel



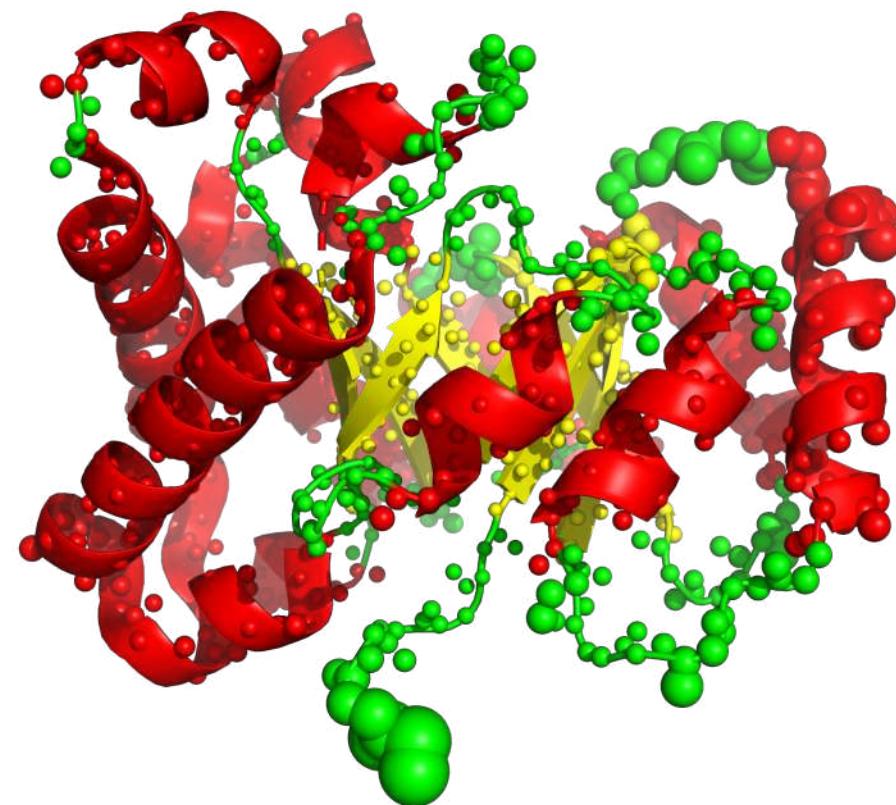
SS are more stable than loops



Considering the dynamics when solving the structure



From Nuclear Magnetic Ressonance



From X-ray diffraction

Atomic Displacement Parameters (ADP)

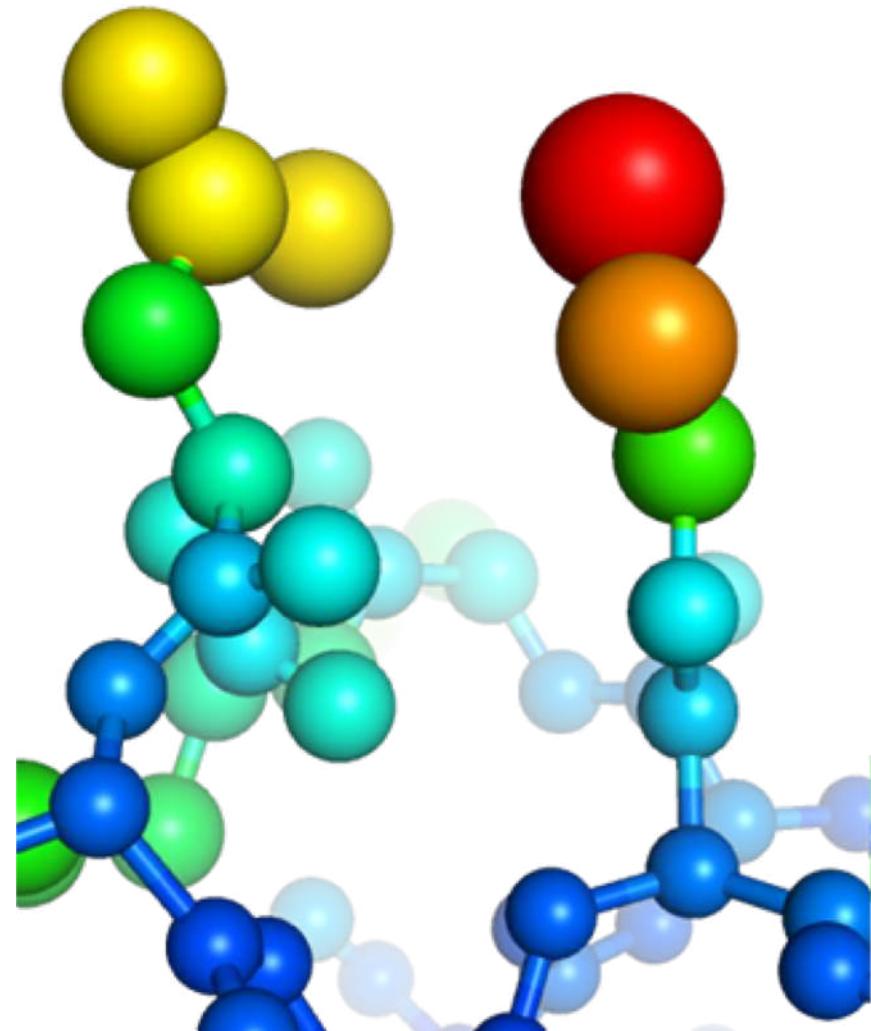
- Or temperature factors or B-factors

.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

.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



Anisotropic ADP at higher resolution

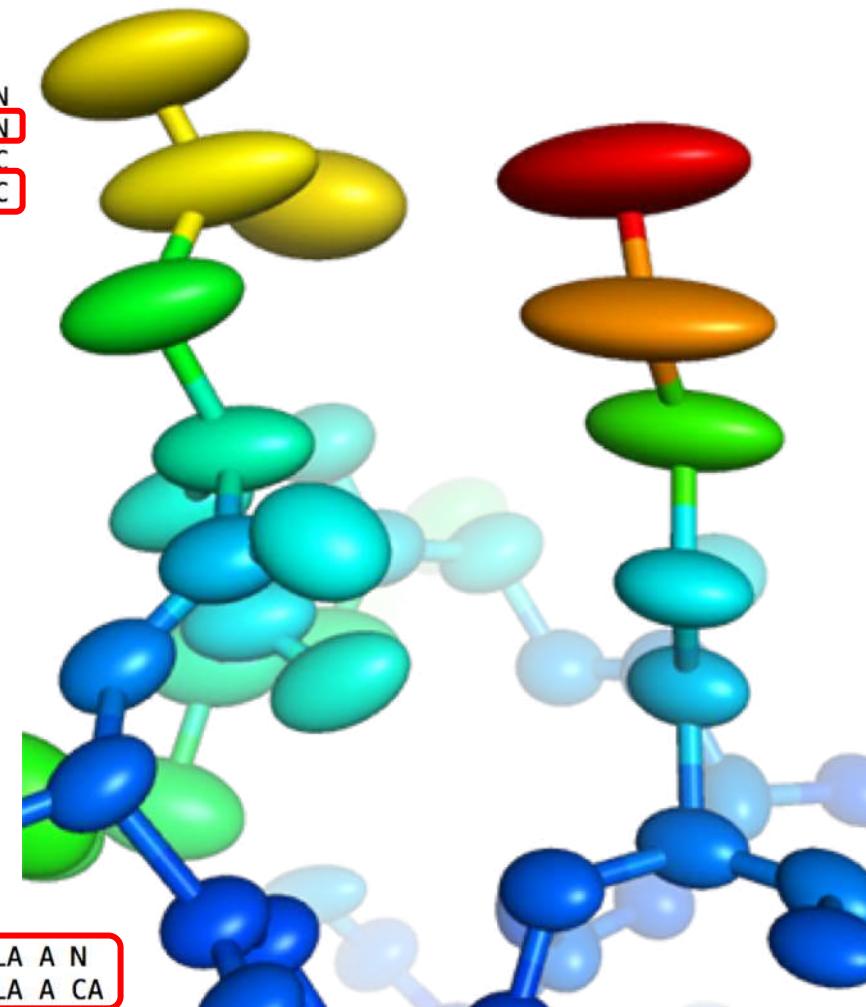
.pdb

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
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

.cif

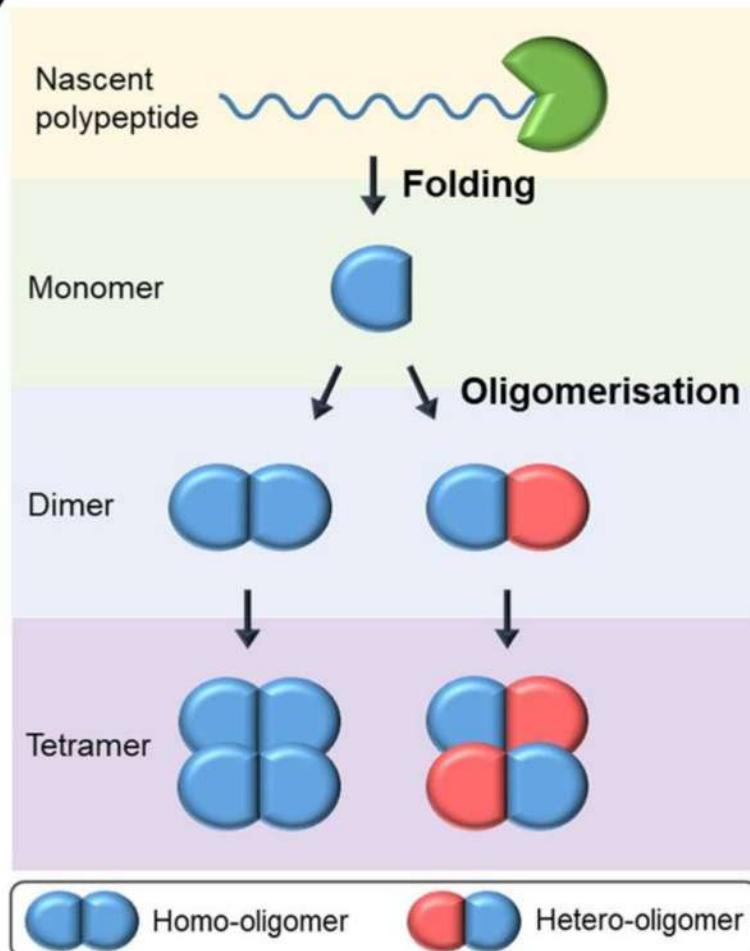
```
#  
loop_  
_atom_site_anisotrop.id  
_atom_site_anisotrop.type_symbol  
_atom_site_anisotrop.pdbx_label_atom_id  
_atom_site_anisotrop.pdbx_label_alt_id  
_atom_site_anisotrop.pdbx_label_comp_id  
_atom_site_anisotrop.pdbx_label_asym_id  
_atom_site_anisotrop.pdbx_label_seq_id  
_atom_site_anisotrop.pdbx_PDB_ins_code  
_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

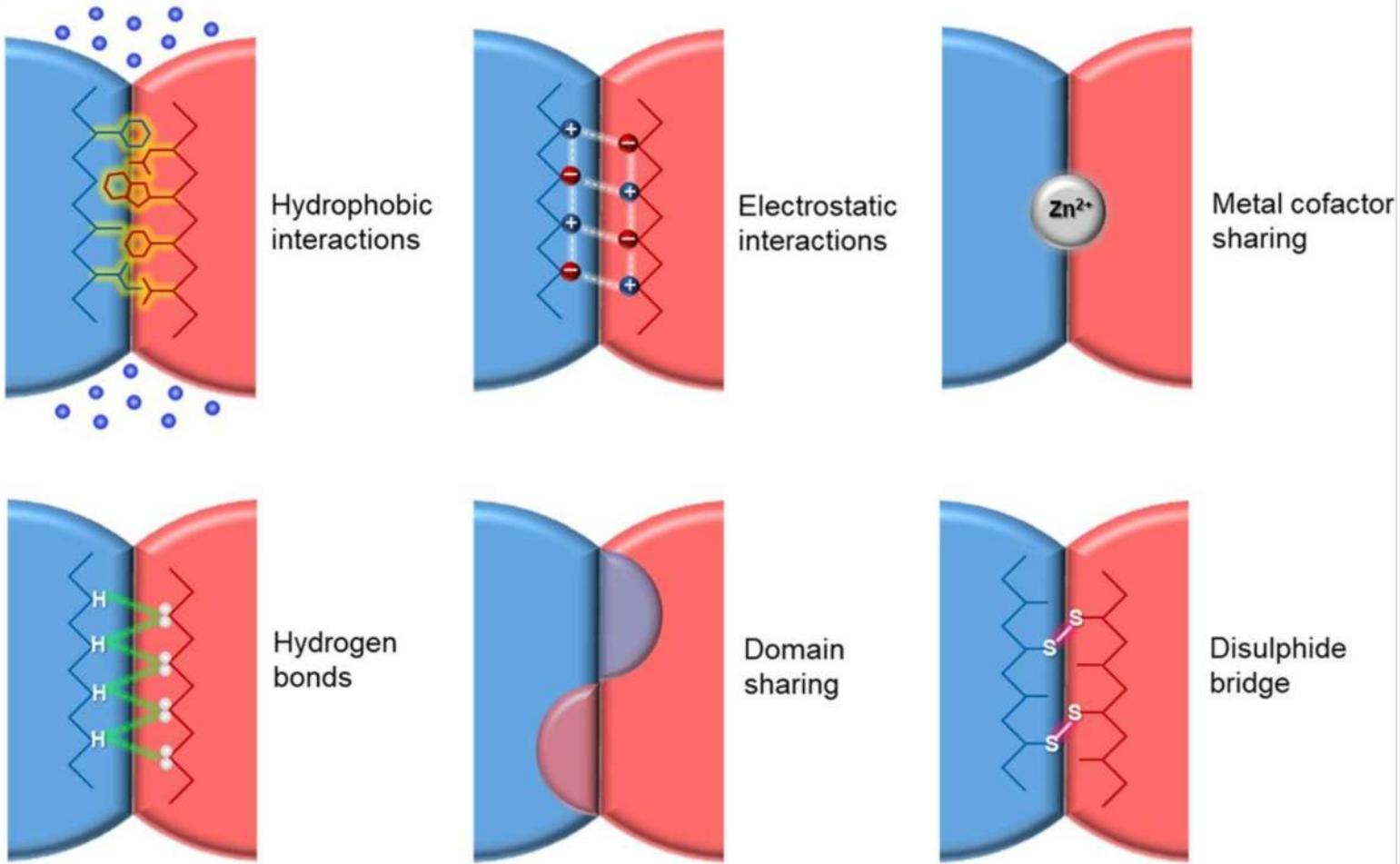


Oligomerization forces

(a)



(b)



Let's analyse the flexibility of a TIM Barril
And the oligomerization of an heptamer

Protein structure prediction

Prediction strategies

- Homology modelling – comparison to homologous proteins

Prediction strategies

- Homology modelling – comparison to homologous proteins
- Fold recognition – comparison to fold of similar sequences

Prediction strategies

- Homology modelling – comparison to homologous proteins
- Fold recognition – comparison to fold of similar sequences
- *De novo* or *ab initio* – prediction from primary structure

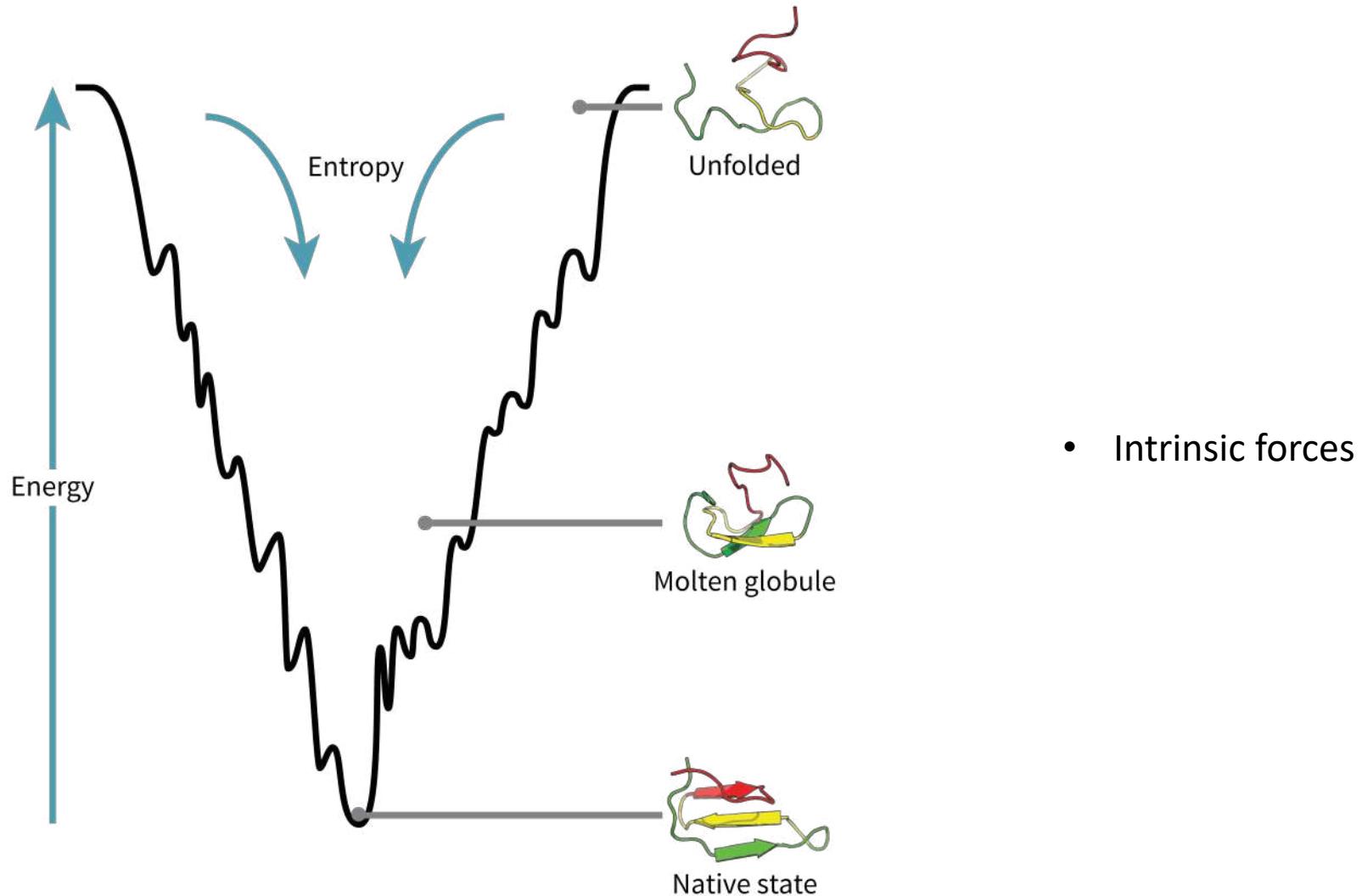
Anfinsen's dogma

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

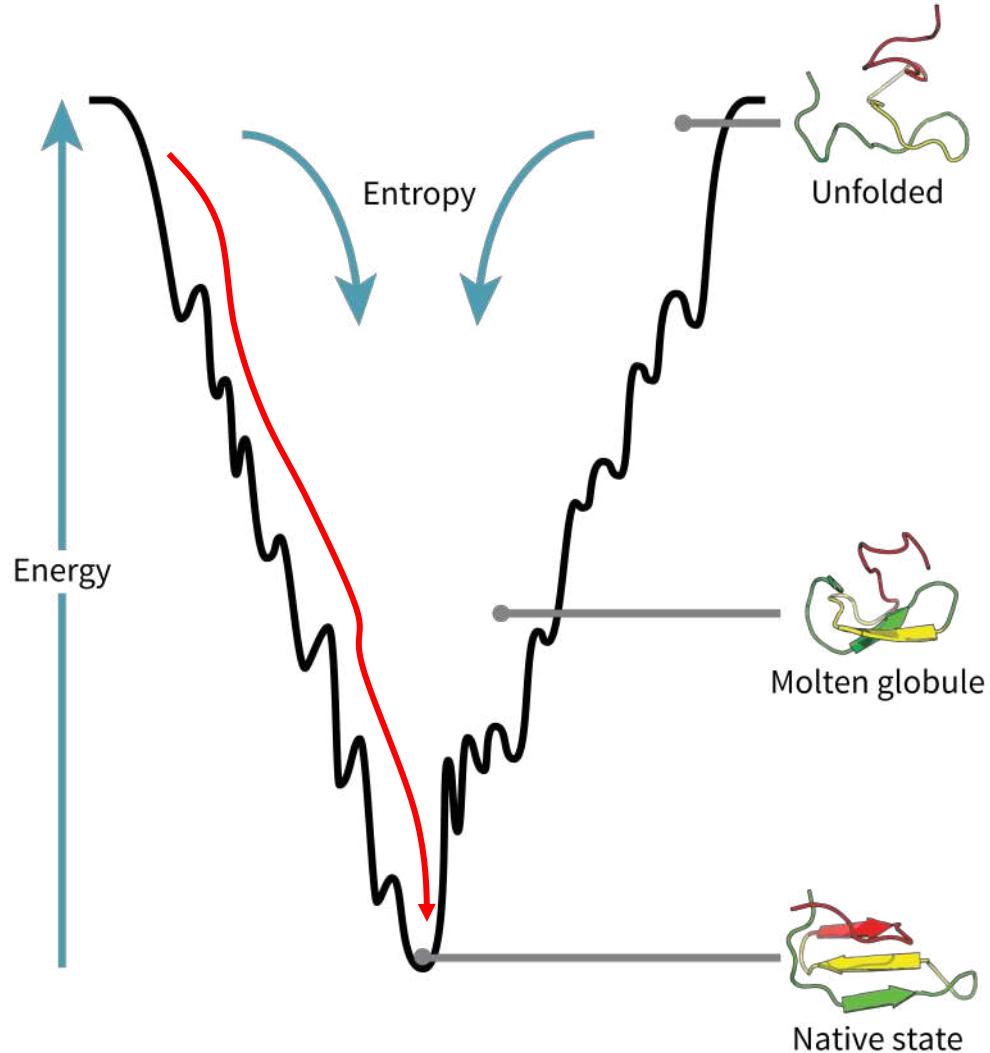


Christian B. Anfinsen

What folds a protein?



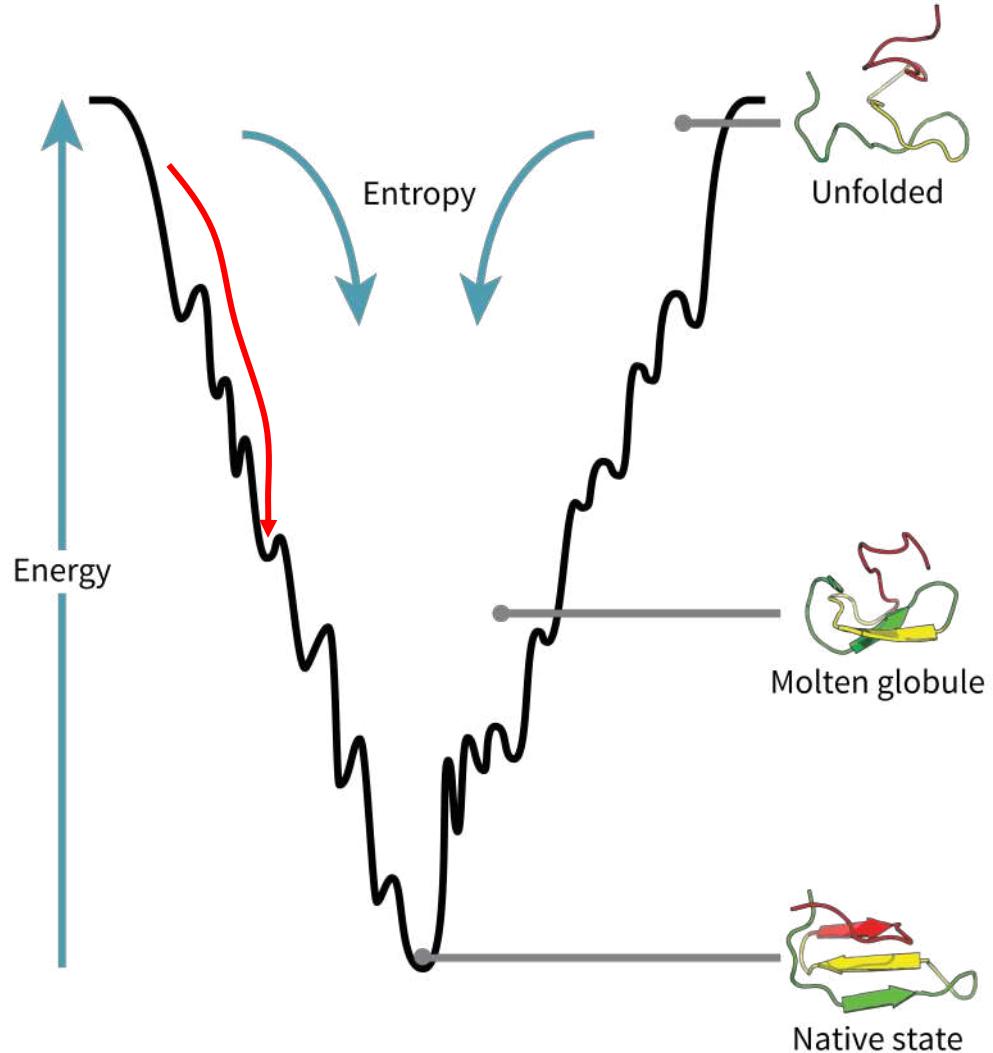
Minimization



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

But it ends up trapped at local minima.

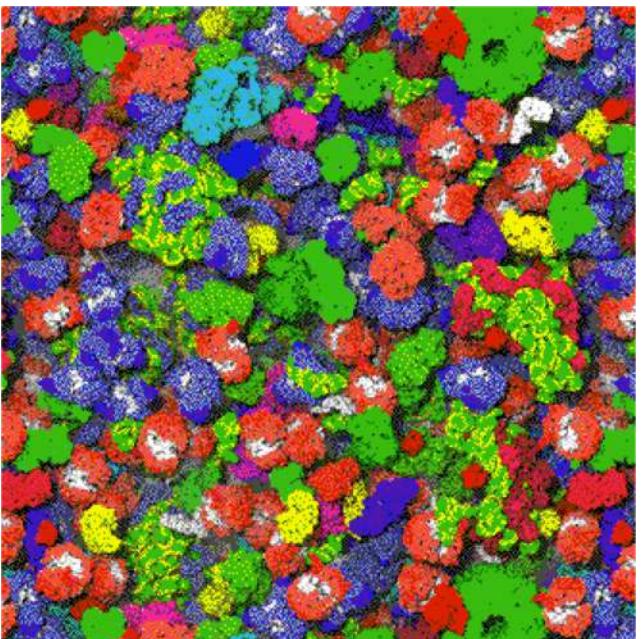
Minimization



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

But it ends up trapped at local minima.

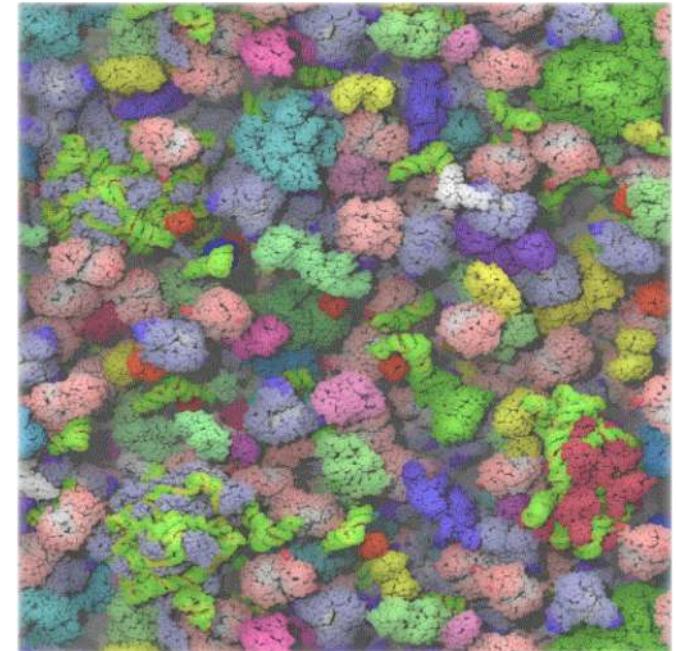
What folds a protein?



McGuffee and Elcock, 2010

15 μ s simulation of cytoplasm

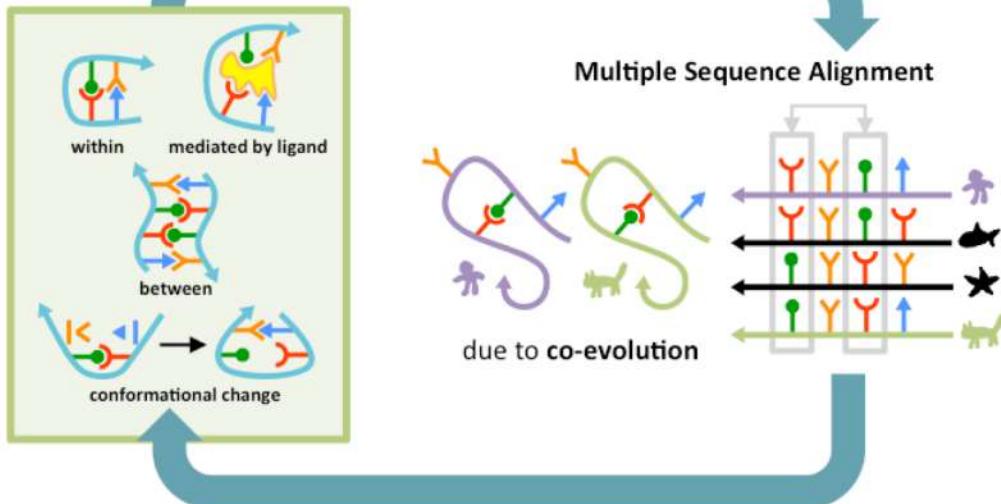
- Environmental conditions



What folds a protein?

What is co-evolution?

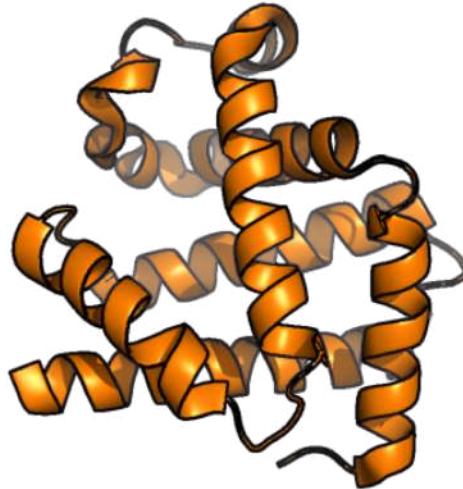
Important **Contacts** in Proteins, are evolutionarily conserved and encoded in a



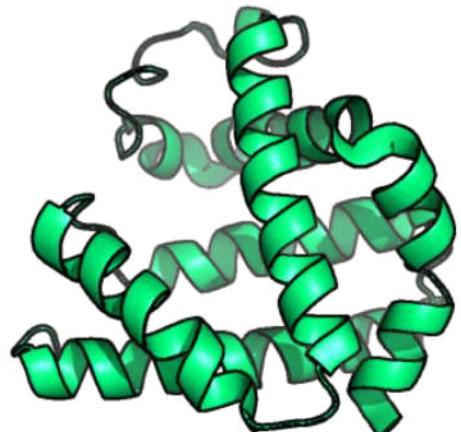
- Evolutionary constraints

Low sequence similarity but high structural homology

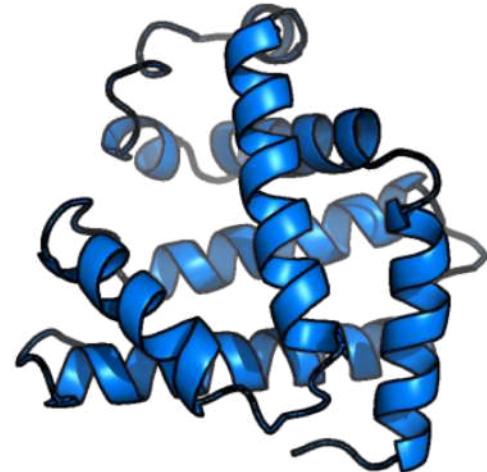
Human mioglobin



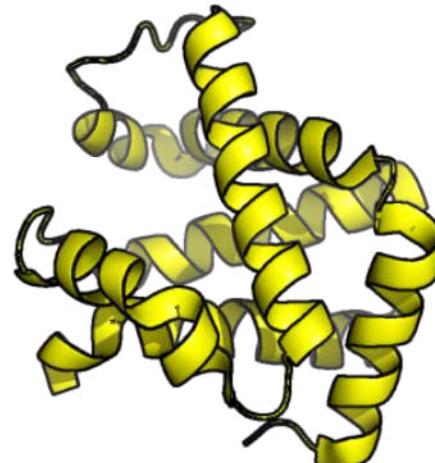
Pigeon mioglobin
25 % sequence identity



African elephant mioglobin
80% sequence identity

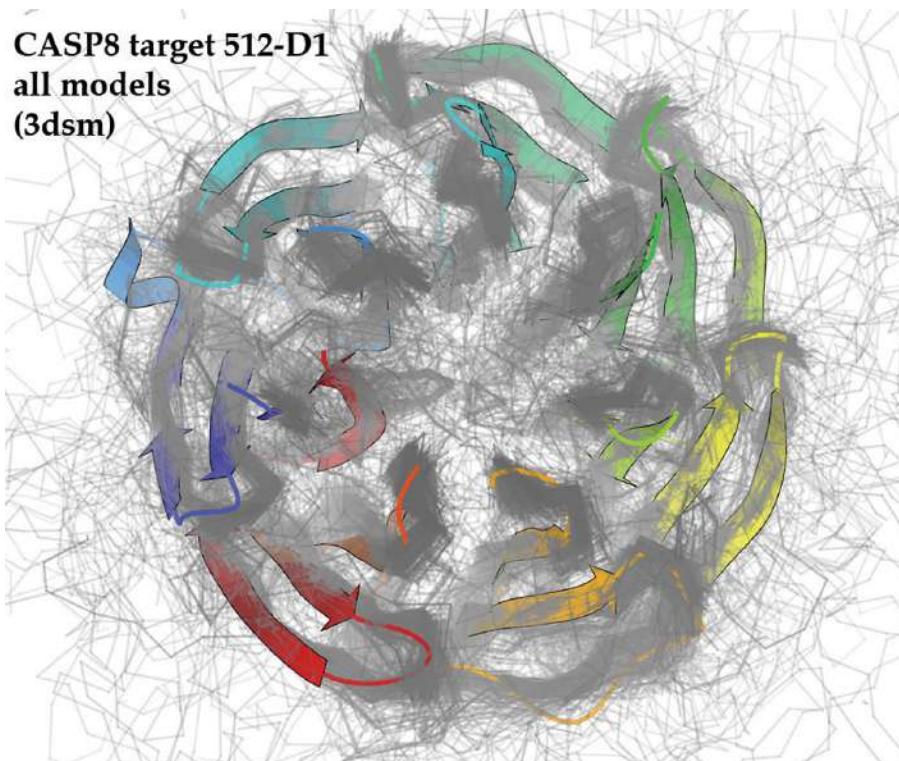


Black-fin tuna mioglobin
45 % sequence identity



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

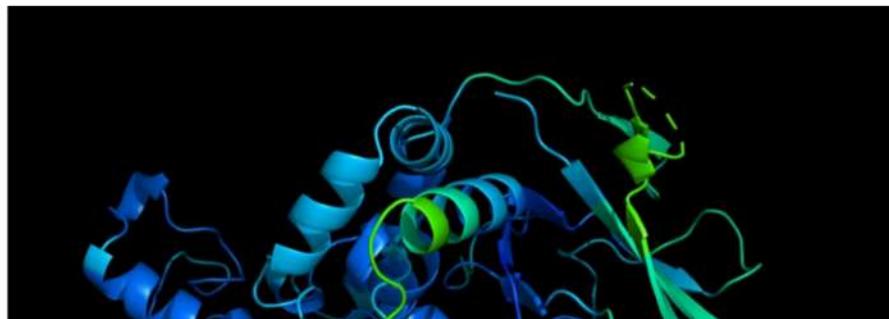
nature > news > article

NEWS | 30 November 2020

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

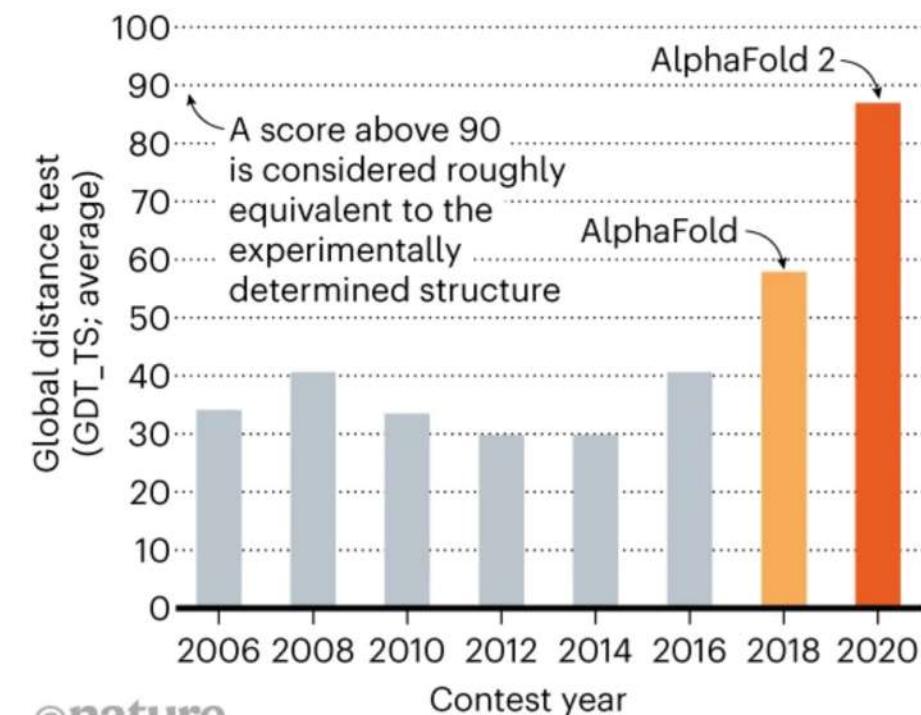
Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway



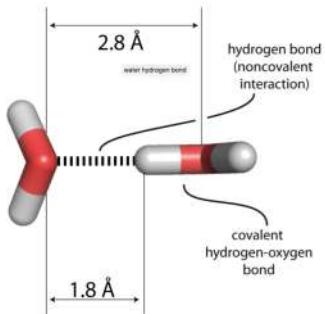
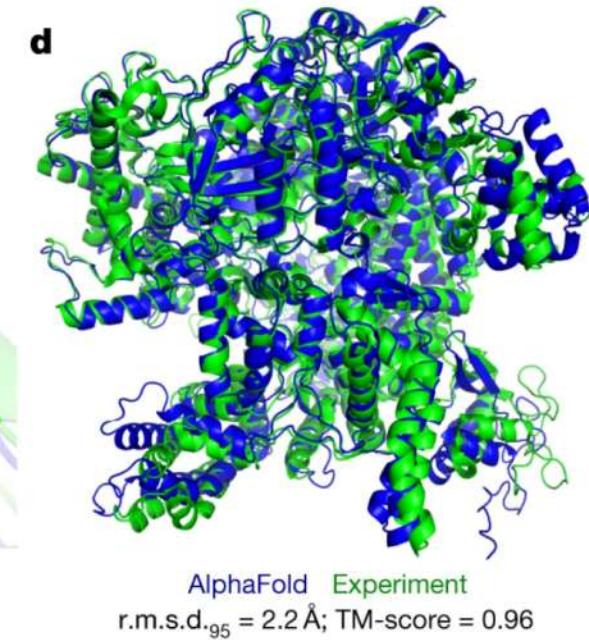
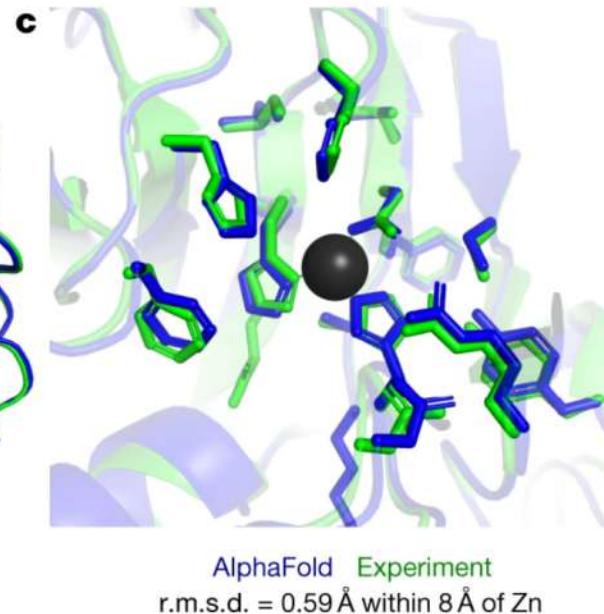
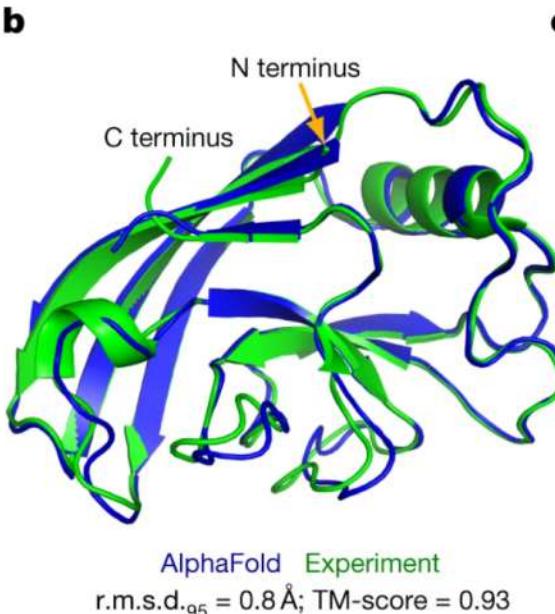
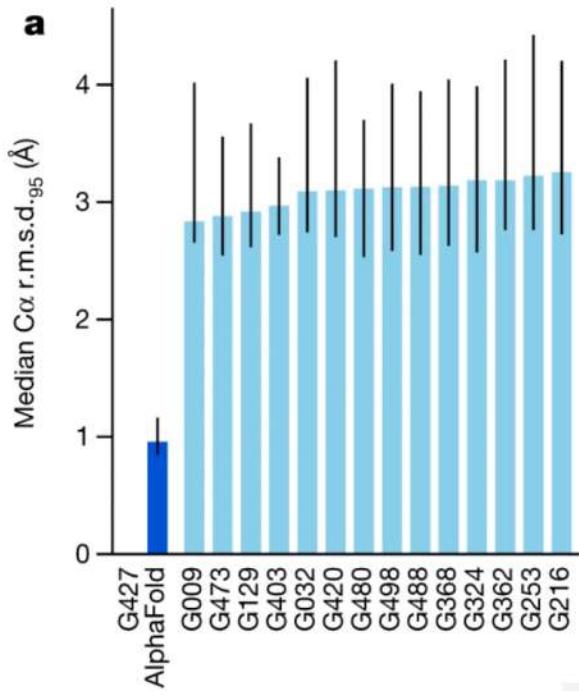
STRUCTURE SOLVER

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.



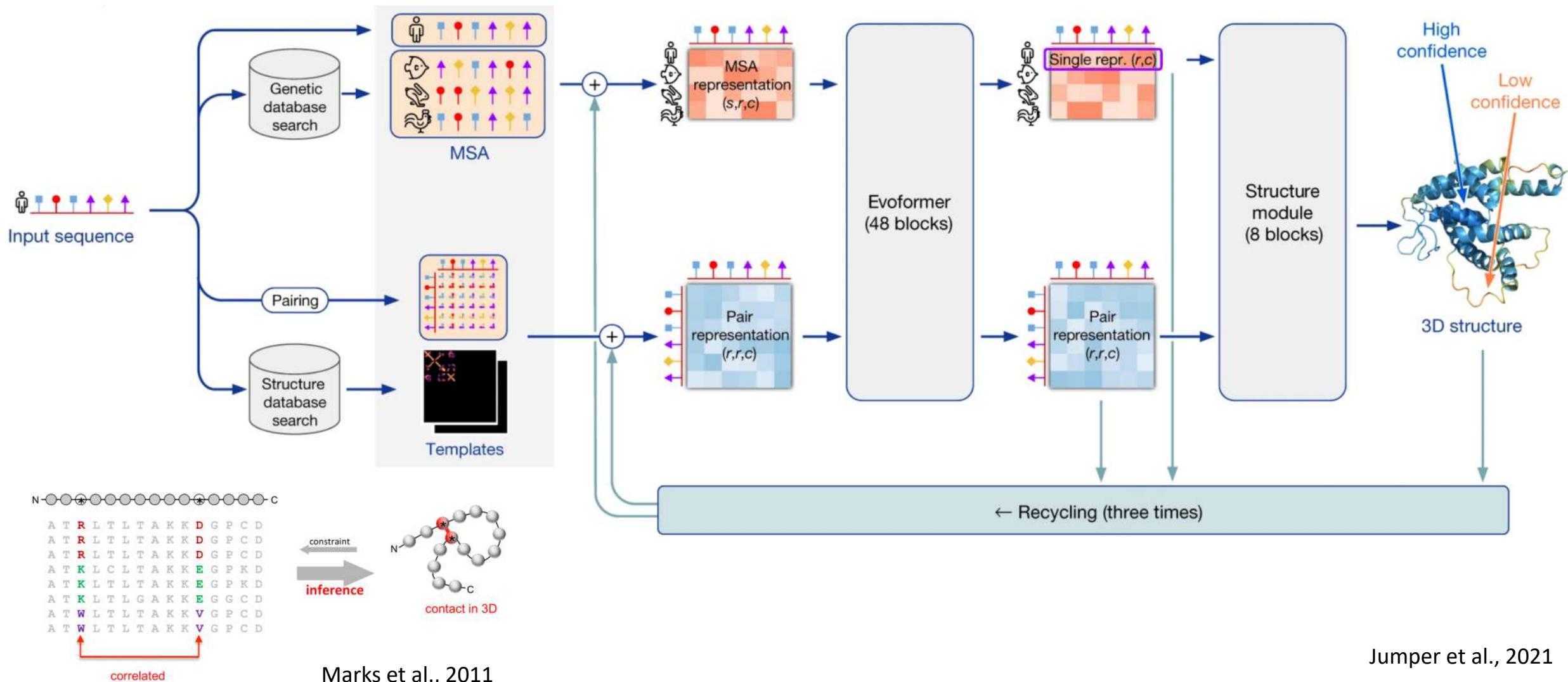
©nature

AlphaFold accuracy



- Compared by RMSD and TM-score

AlphaFold pipeline

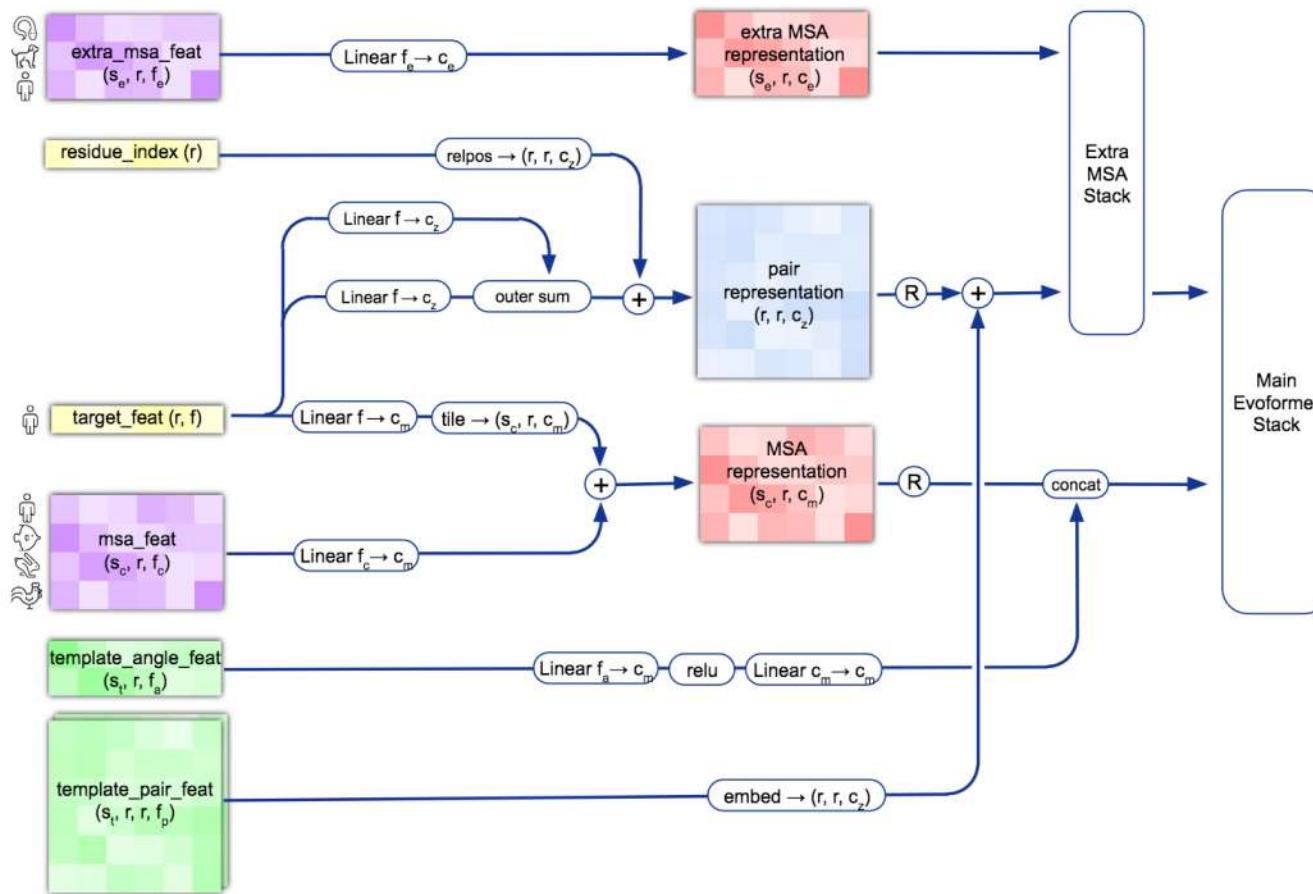


Keys to success of AlphaFold2?

- 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**

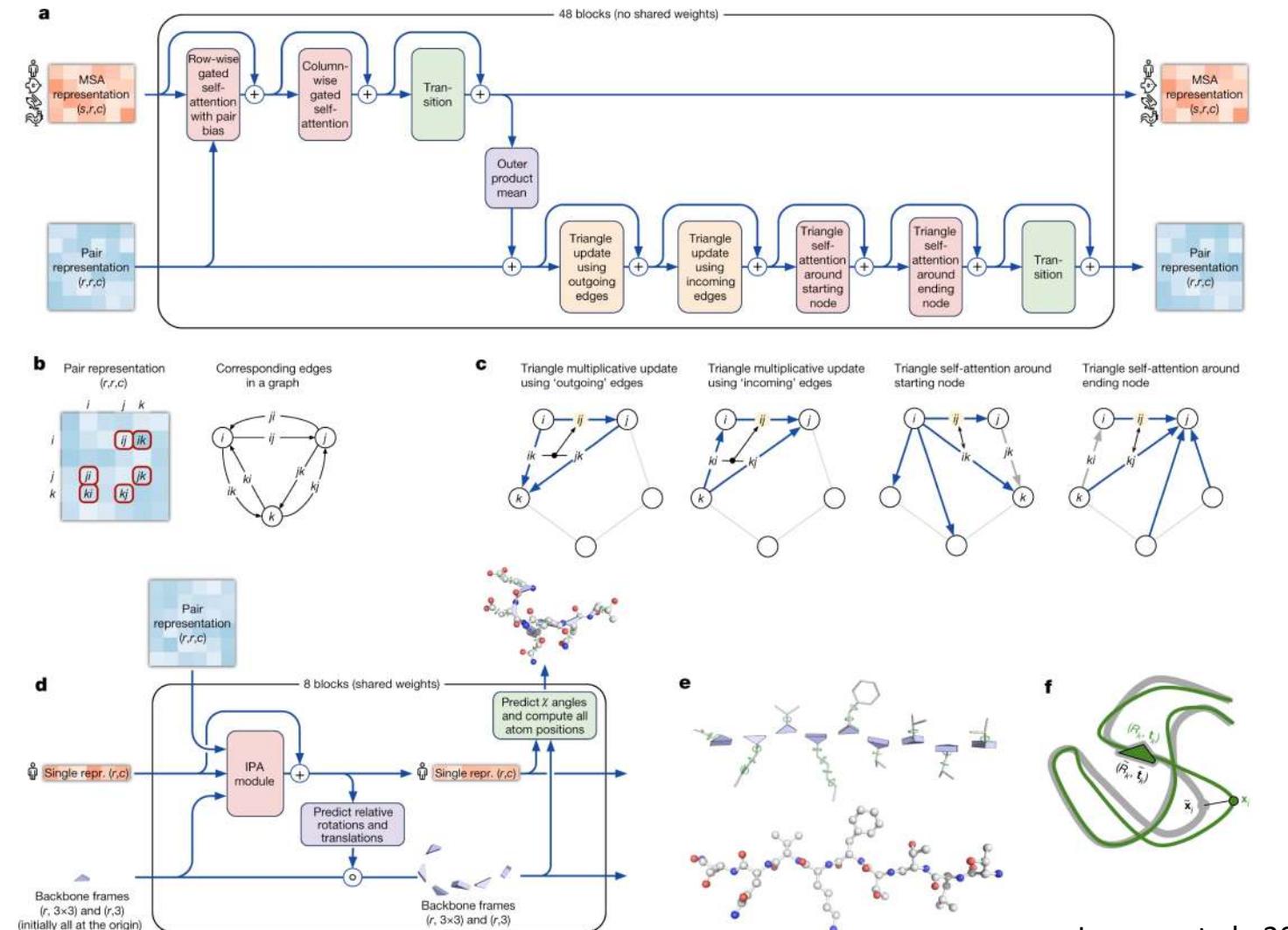
Keys to success of AlphaFold2?

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



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/>

Interpretation of predictions

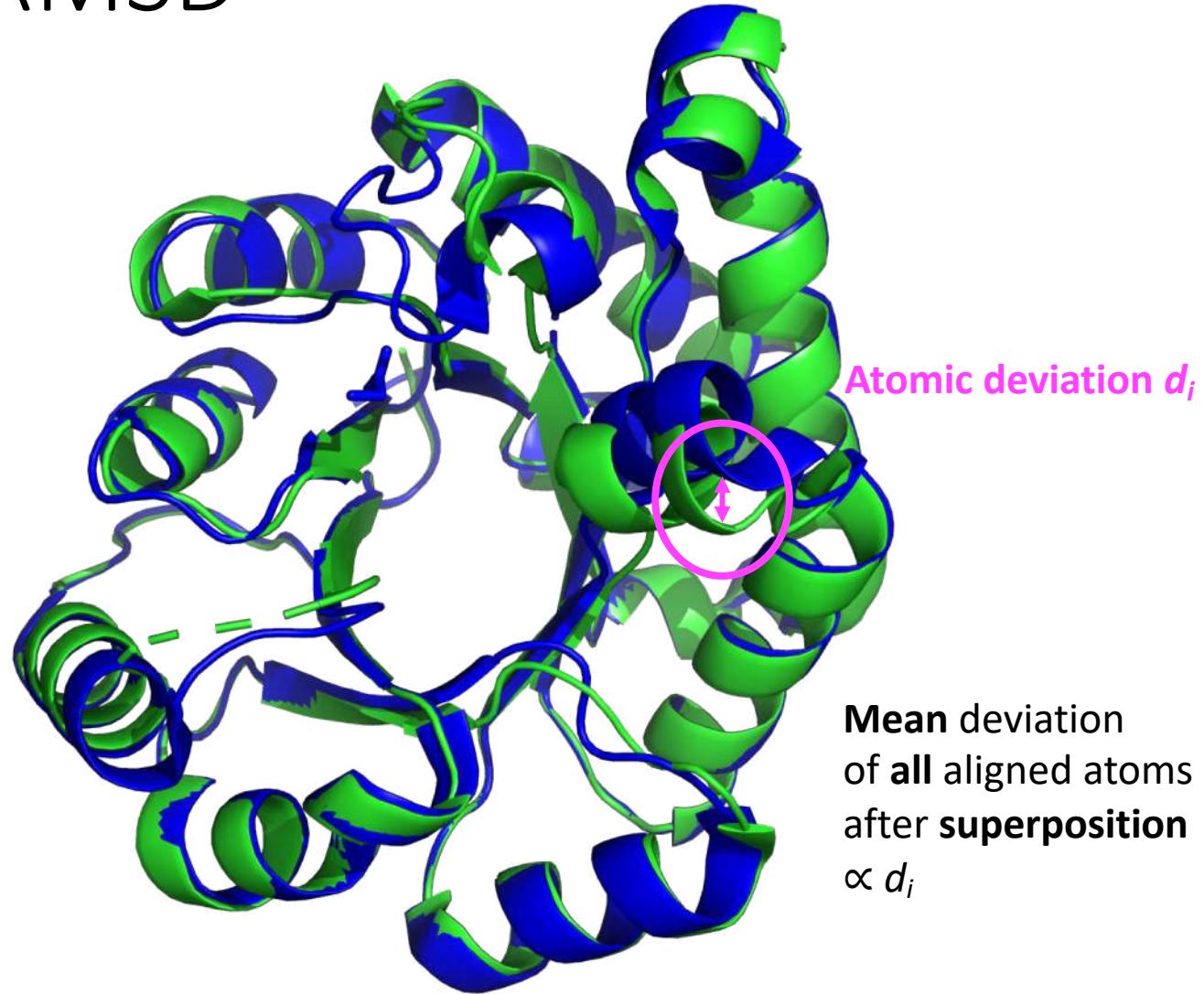
Usual metrics

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

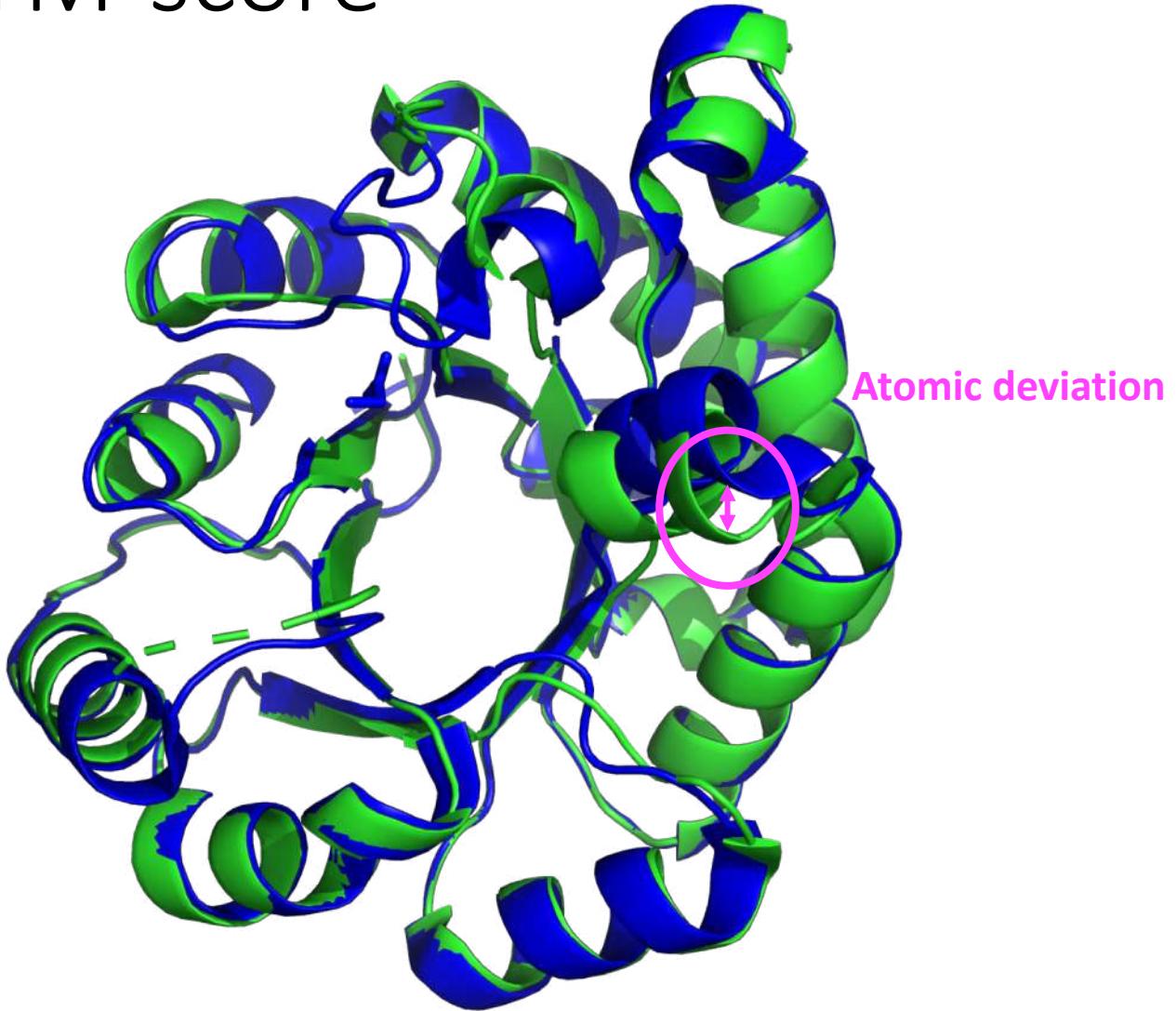
AlphaFold metrics

- pLDDT – predicted LDDT
- PAE – Predicted Aligned Errors

RMSD



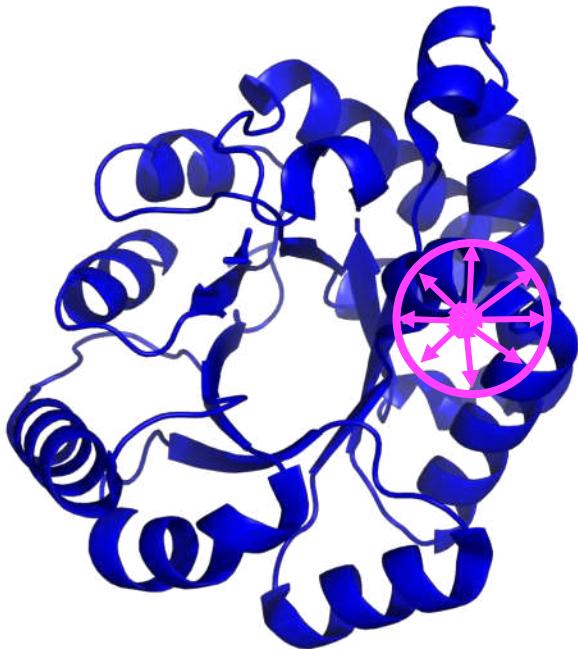
TM-score



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

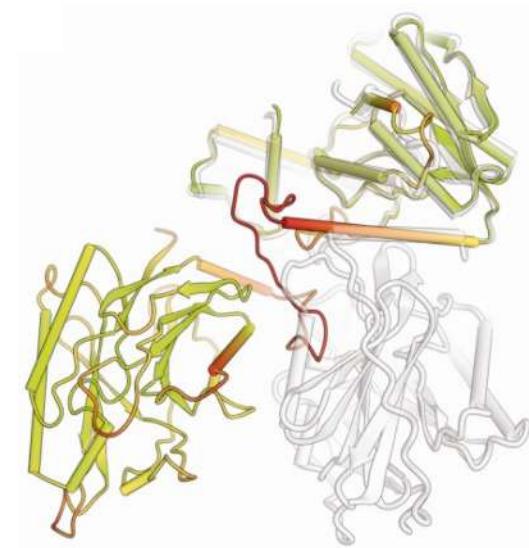
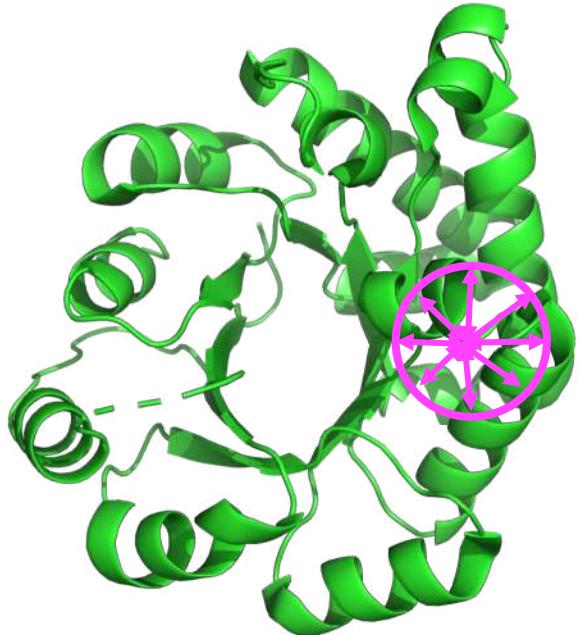
Intended to be more
accurate because smaller
deviations weigh more
than larger deviations.

LDDT



List of d_i within a threshold

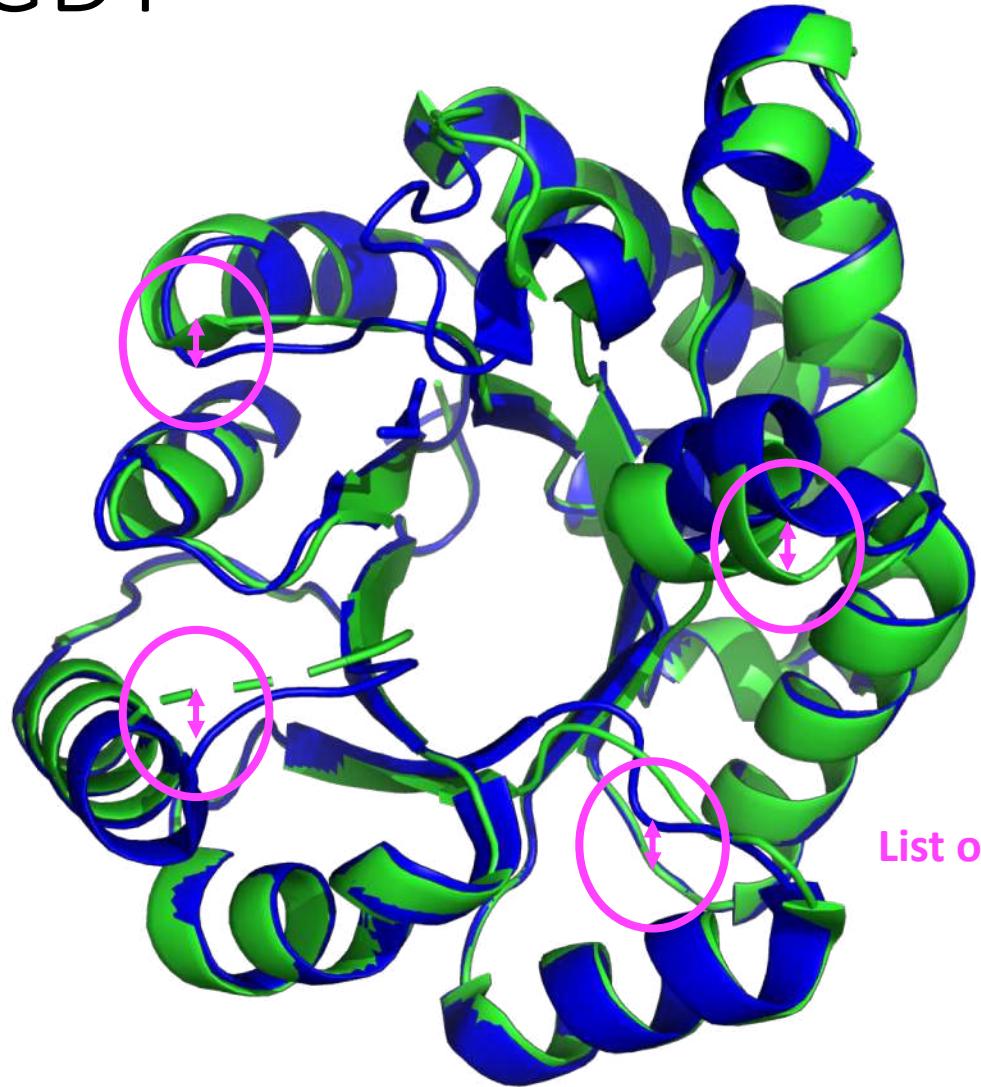
**Superposition free,
percentage of conserved
local inter-atomic distances**



Mariani et al., 2013

Good for comparison of
structures with separate domains

GDT

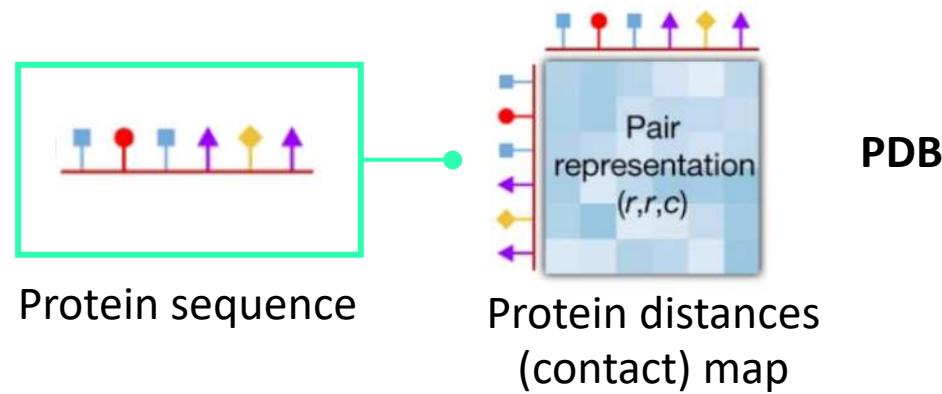
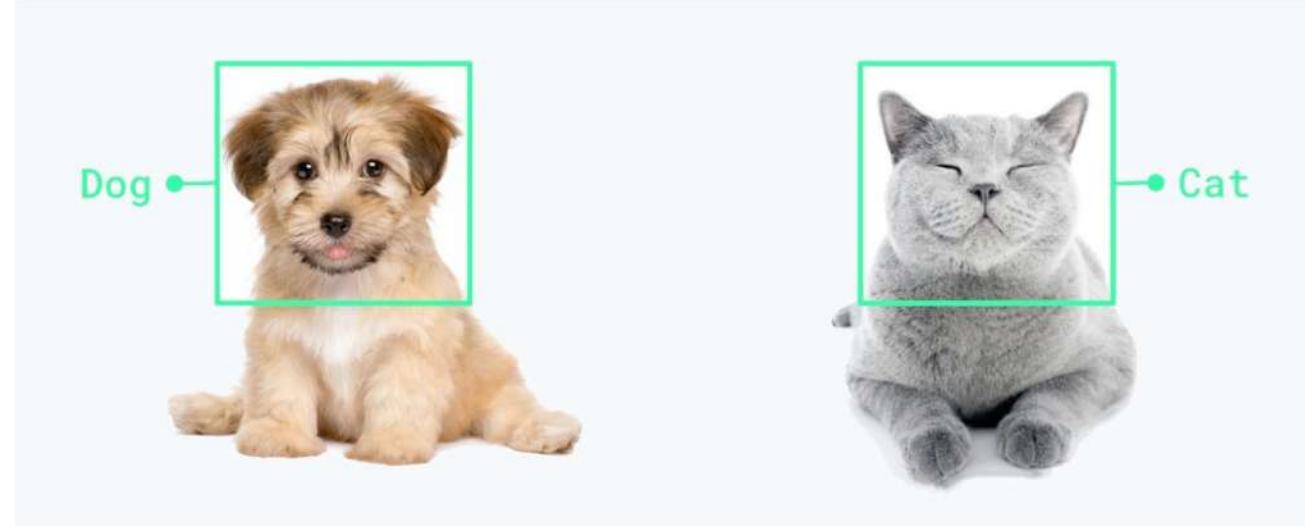


List of deviating residues

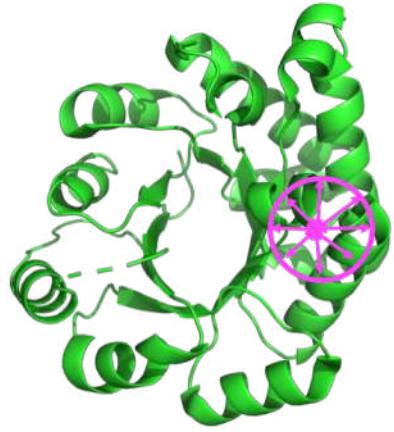
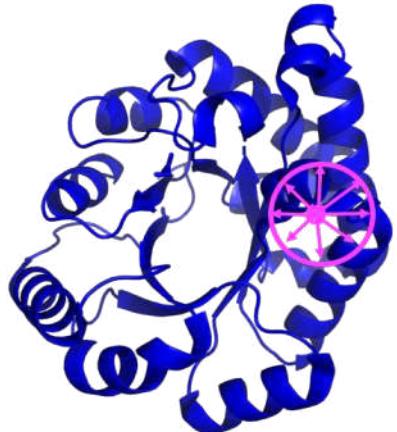
A **percentage** of residues that deviate within a **set of cut-offs** after iterative **superpositions**

Correct in relation to others

Ground truth



pLDDT



Ground truth

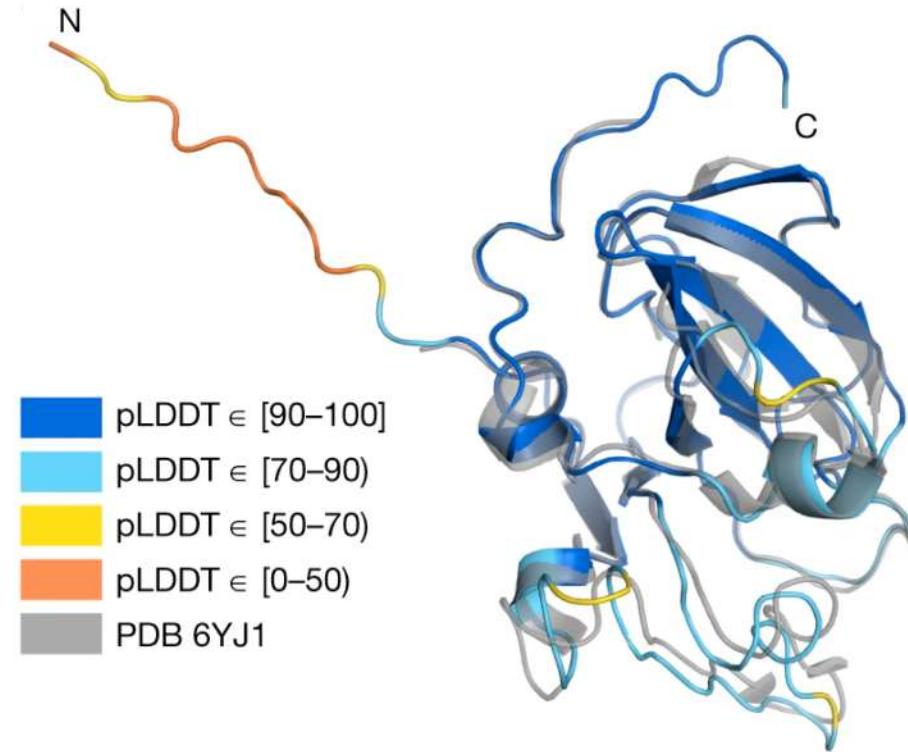
pLDDT > 90 – high accuracy

pLDDT > 70 – generally correct backbone

pLDDT < 70 – low confidence

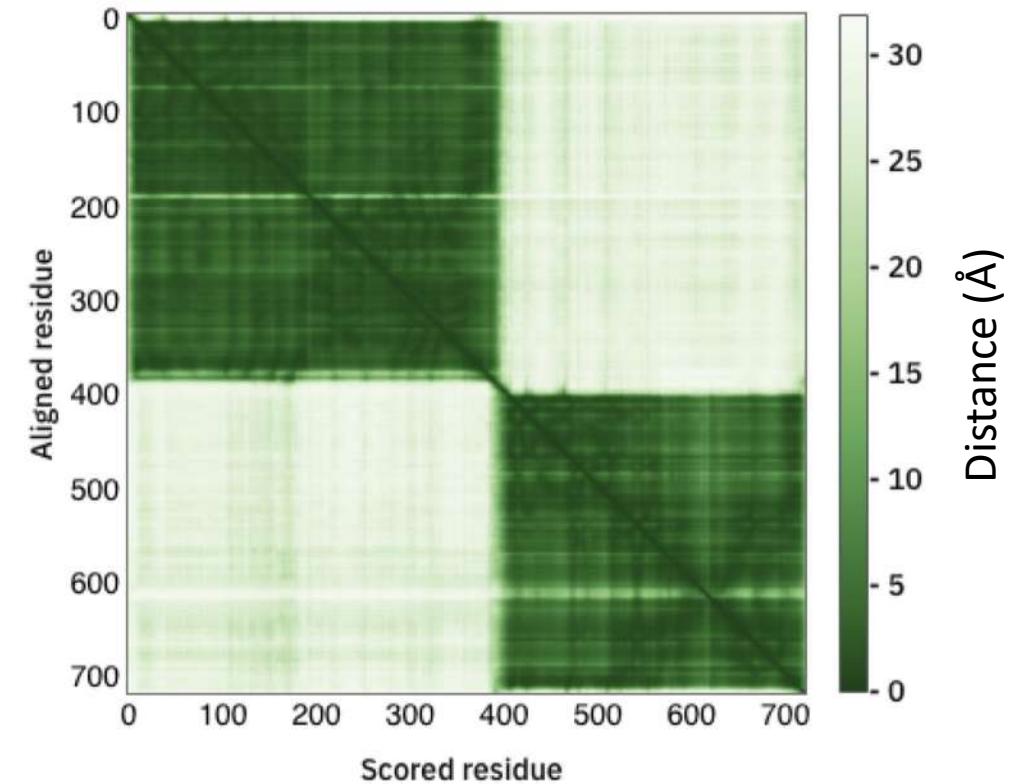
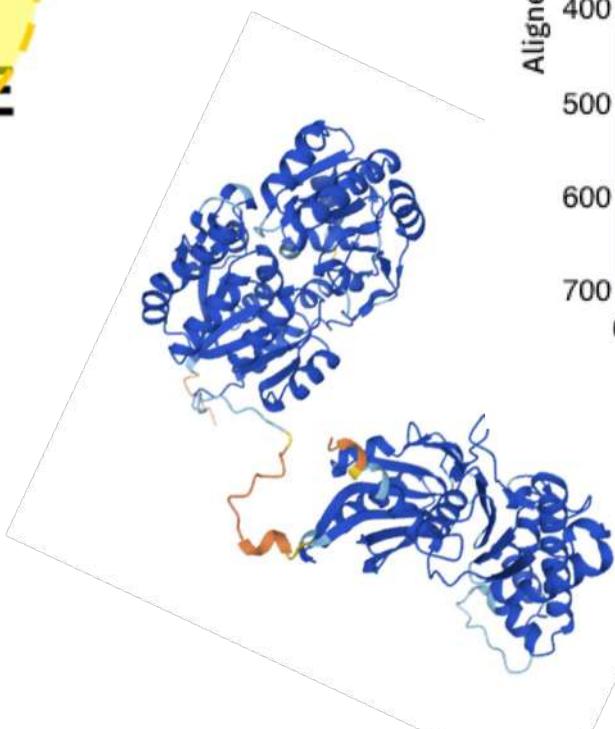
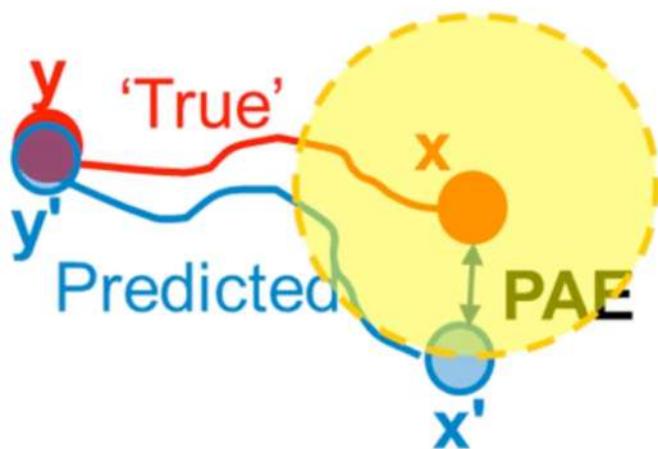
pLDDT < 50 - disordered

Per residue score



PAE

Predicted aligned error (PAE)



Let's do some predictions!

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

Model	initial training	first fine-tuning		second fine-tuning				
	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



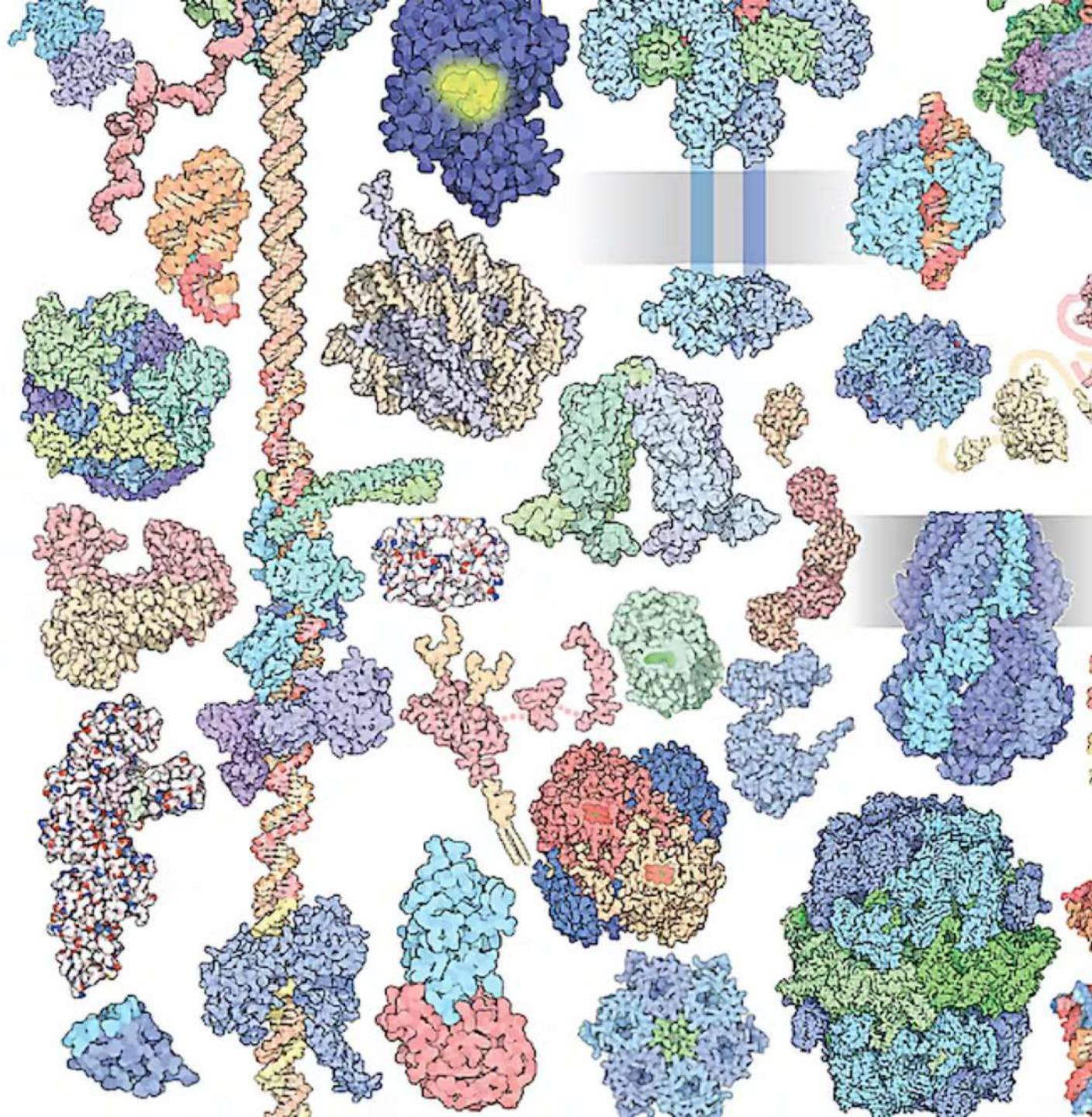
5 models

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

Discussing dynamics

Protein interactions

- Ligand, small molecules
- Proteins and peptides
- Nucleic acids

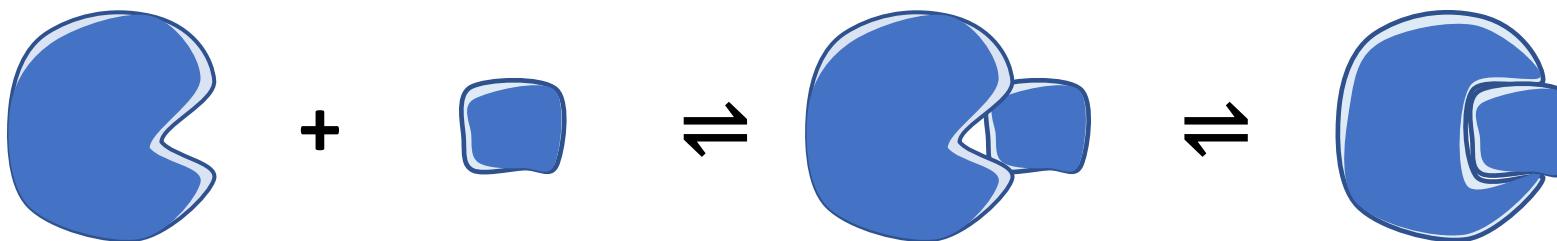


Protein binding mechanisms

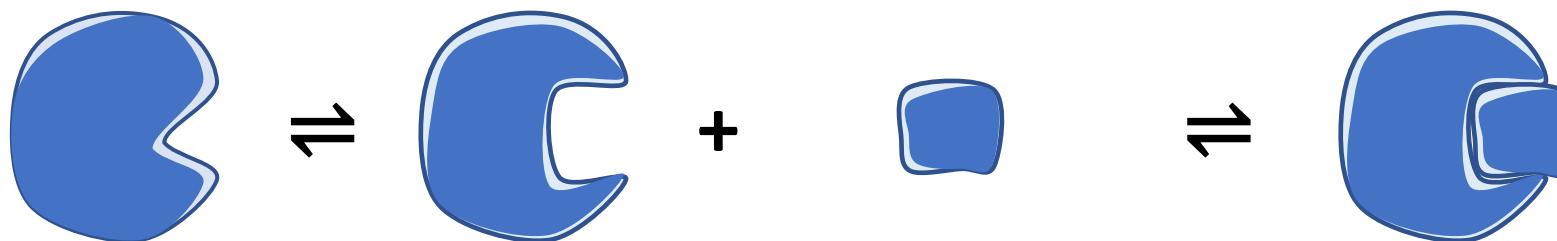
Lock and key



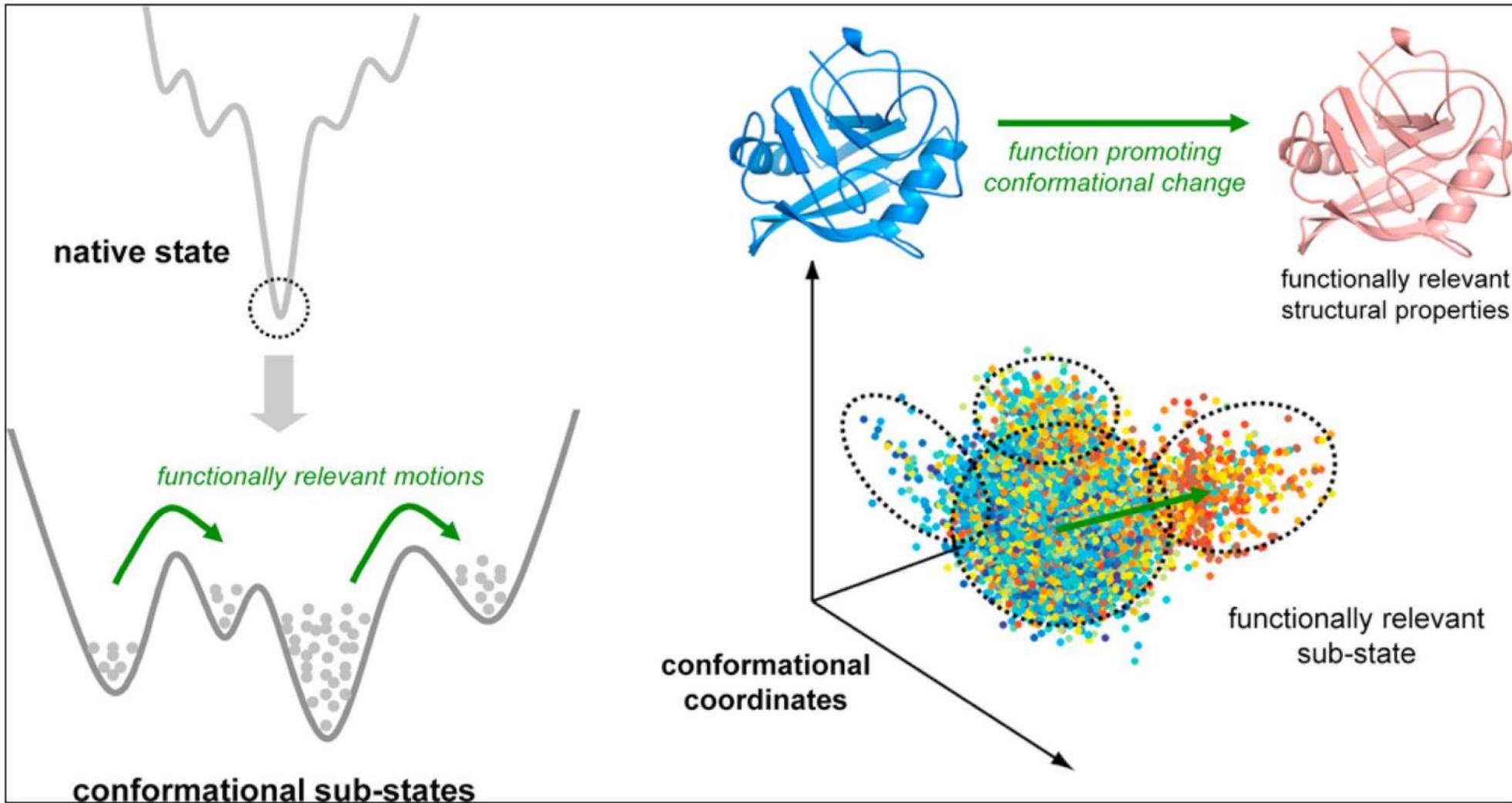
Induced fit



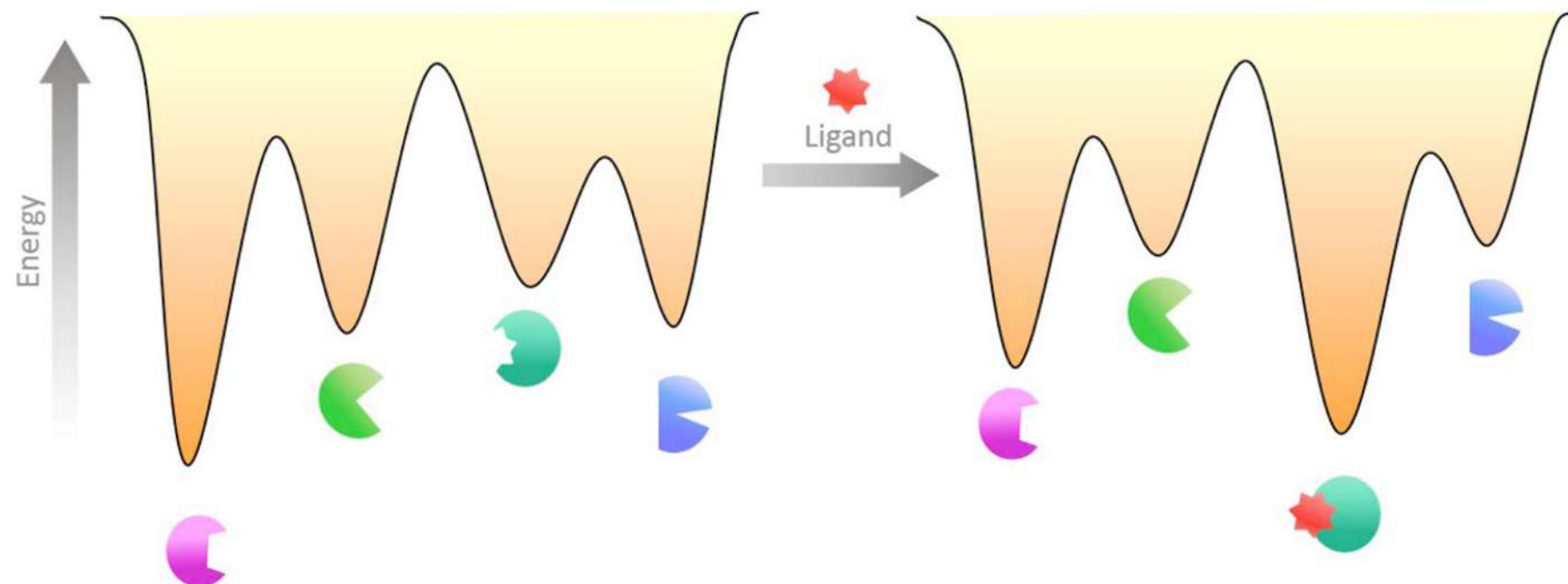
Conformational isomerism



Conformational population and states

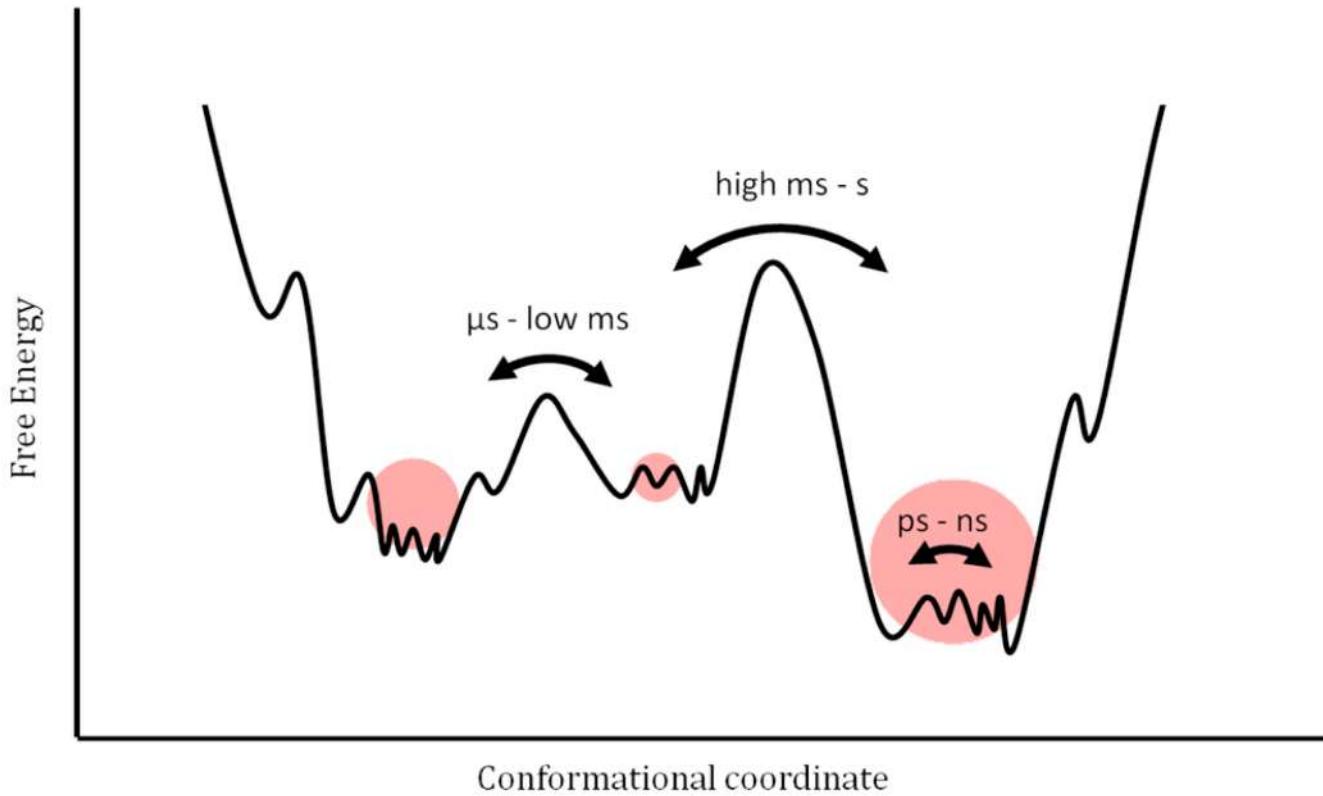


Shift of conformational equilibrium



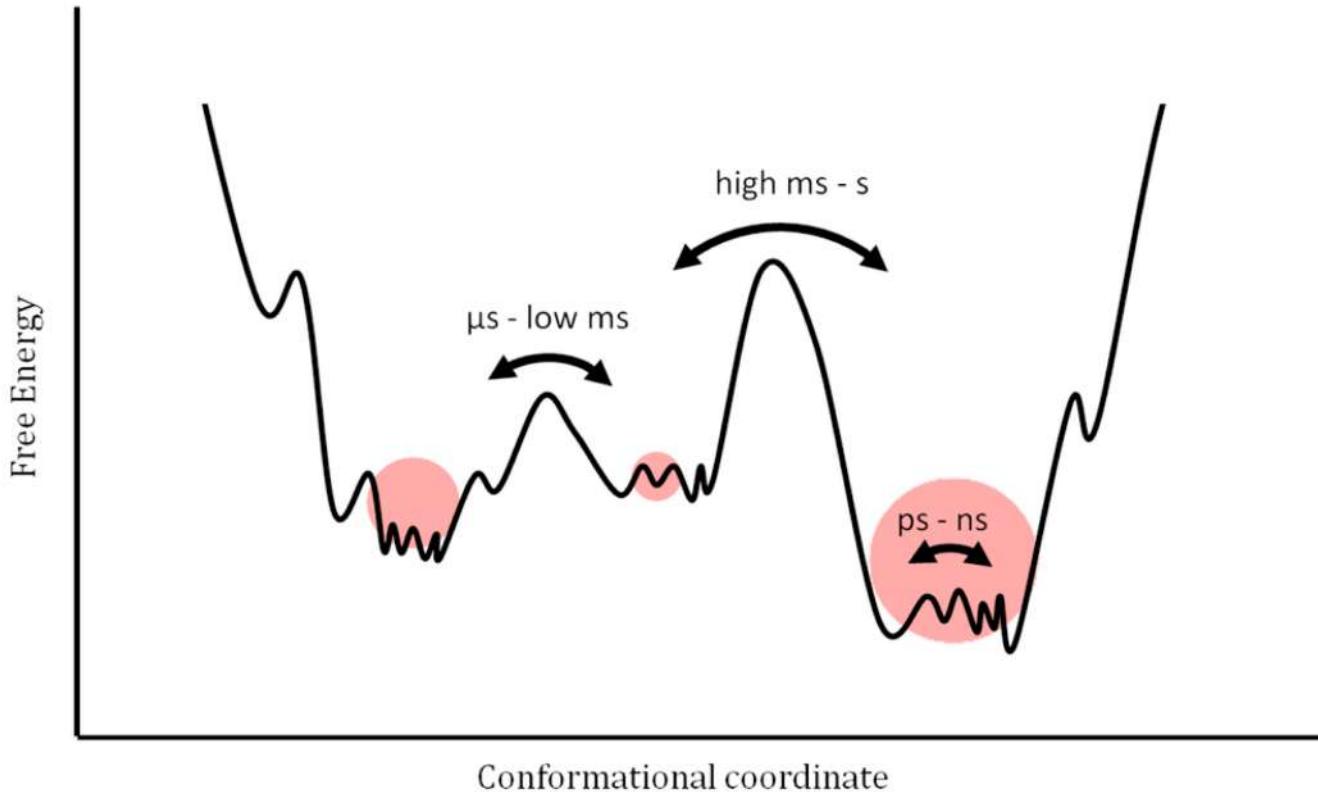
Mishra and Jha, 2022

Conformational states and changes



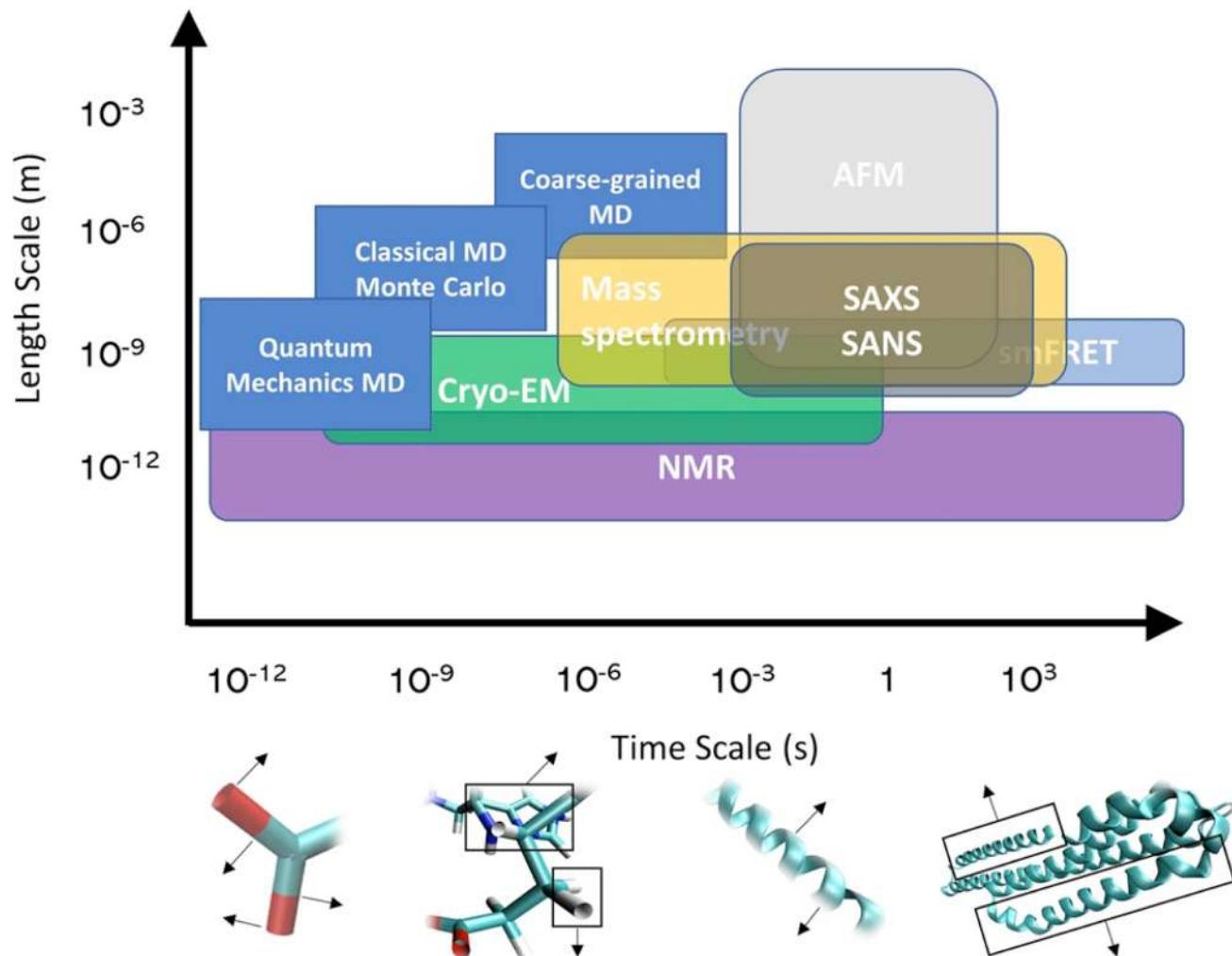
Göbl and Tjandra, 2012

How to sample the conformational space?



Göbl and Tjandra, 2012

How to sample the conformational space?



MD – Molecular Dynamics

AFM – Atomic Force Microscopy

SAXS – Small Angle X-ray scattering

SANS – Small angle neutron scattering

EM – Electron microscopy

NMR – Nuclear resonance spectroscopy

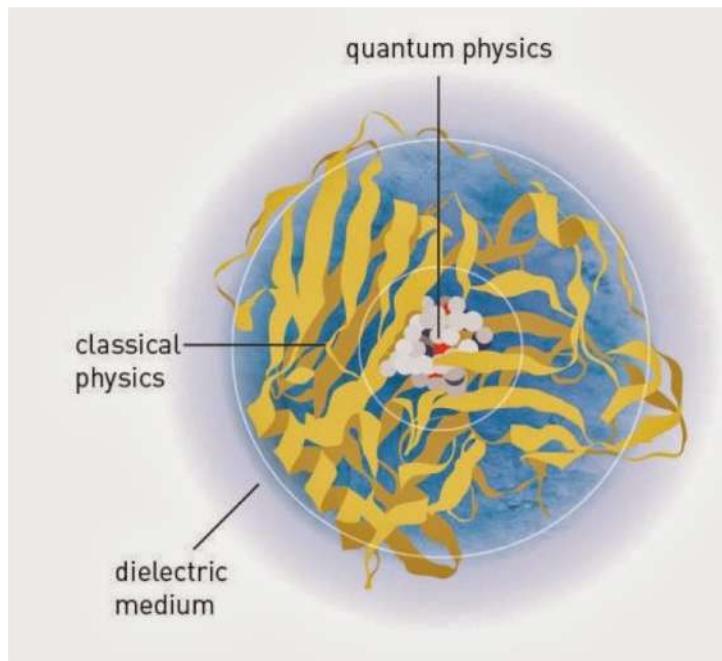
smFRET – single molecule fluorescence

resonance energy transfer

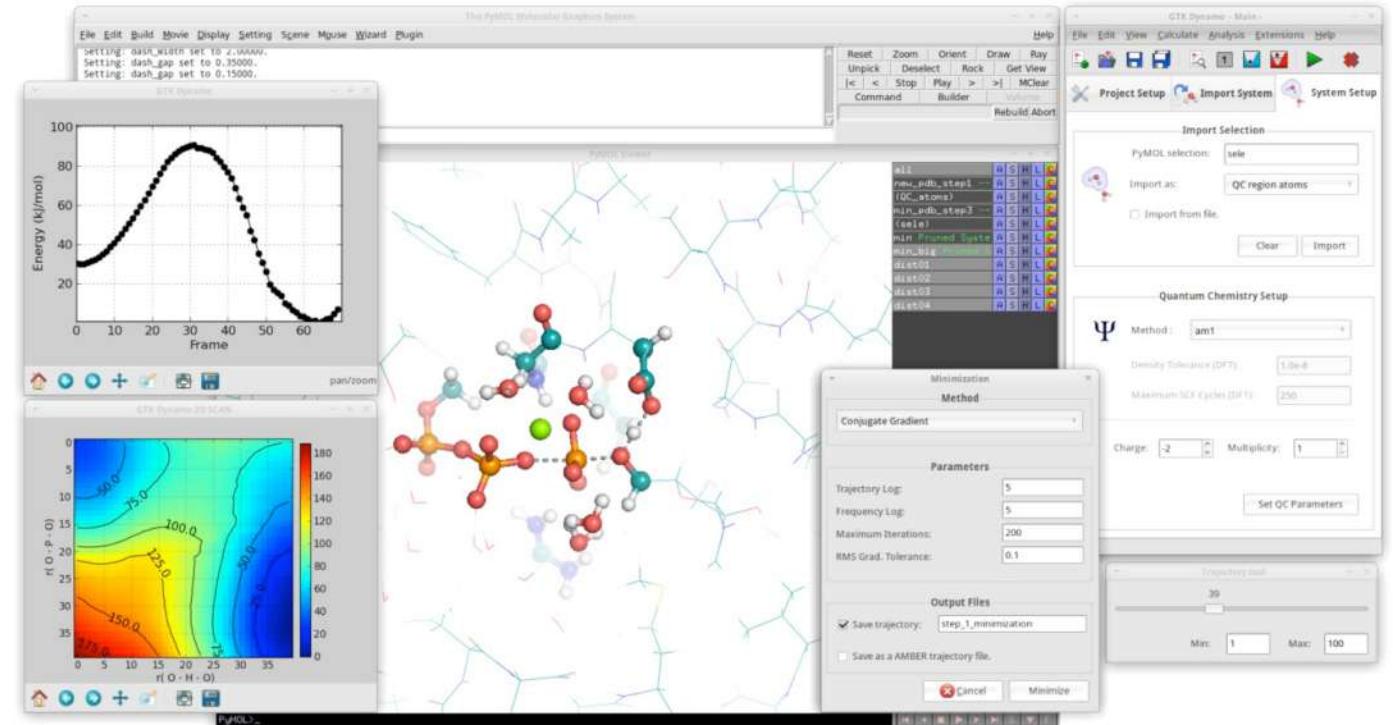
Molecular Dynamics

- Quantum Mechanics
- Classical MD
- Monte Carlo
- Coarse grained MD

Quantum Mechanics/Molecular Dynamics



2013 Nobel Prize in Chemistry



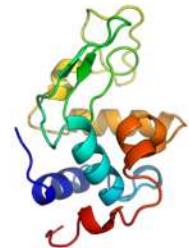
Bachega et al, 2013

Classical MD

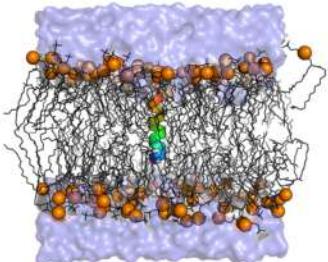
GROMACS Tutorials

Justin A. Lemkul, Ph.D.
Virginia Tech Department of Biochemistry

All tutorials have been updated for GROMACS version 2018!



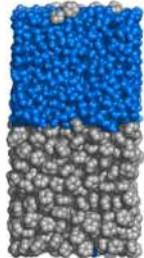
Tutorial 1: Lysozyme in Water



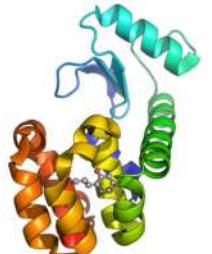
Tutorial 2: KALP₁₅ in DPPC



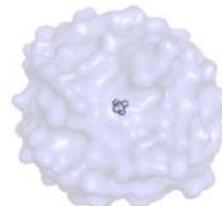
Tutorial 3: Umbrella Sampling



Tutorial 4: Biphasic Systems



Tutorial 5: Protein-Ligand Complex



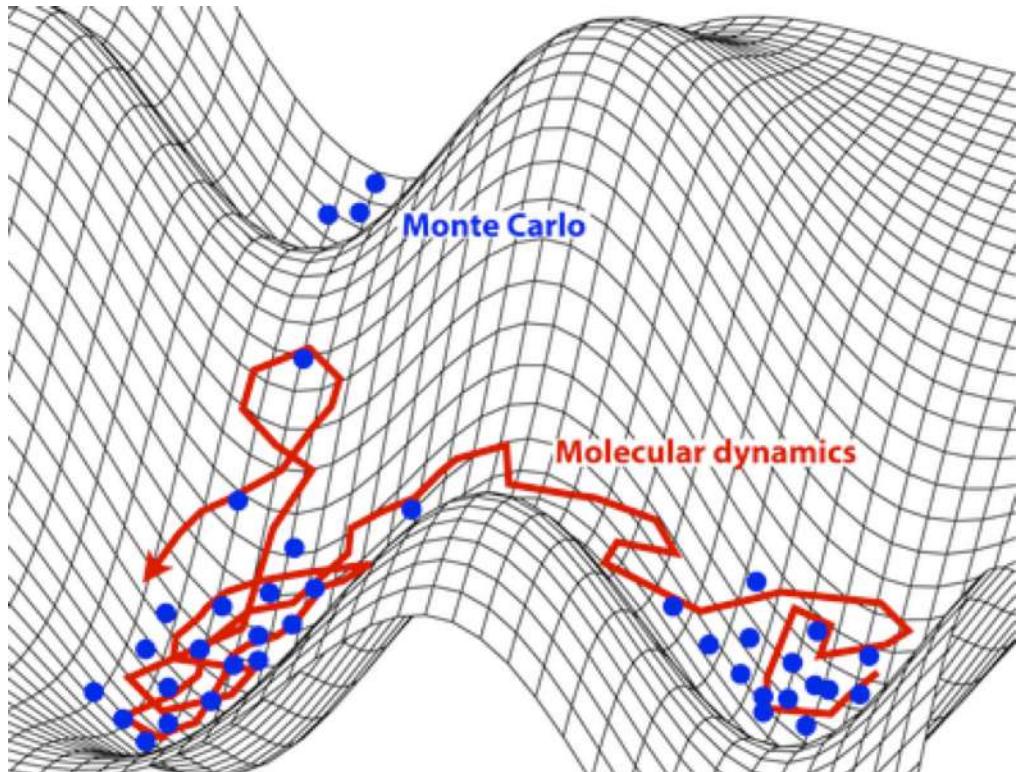
Tutorial 6: Free Energy of Solvation



Tutorial 7: Virtual Sites

Apply a set of forces (force field) and calculate how the molecular interactions evolve with time.

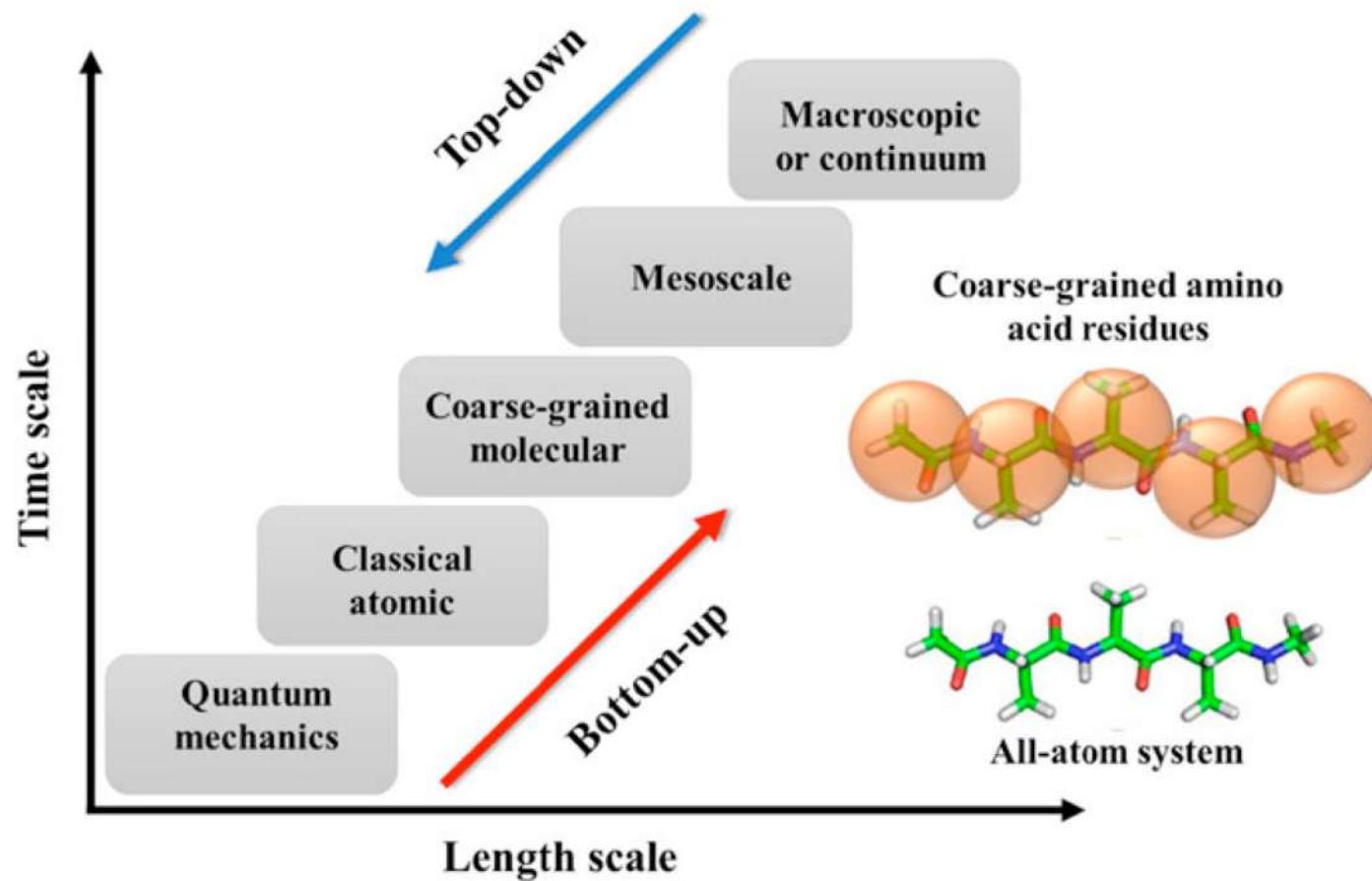
Monte Carlo



wikipedia

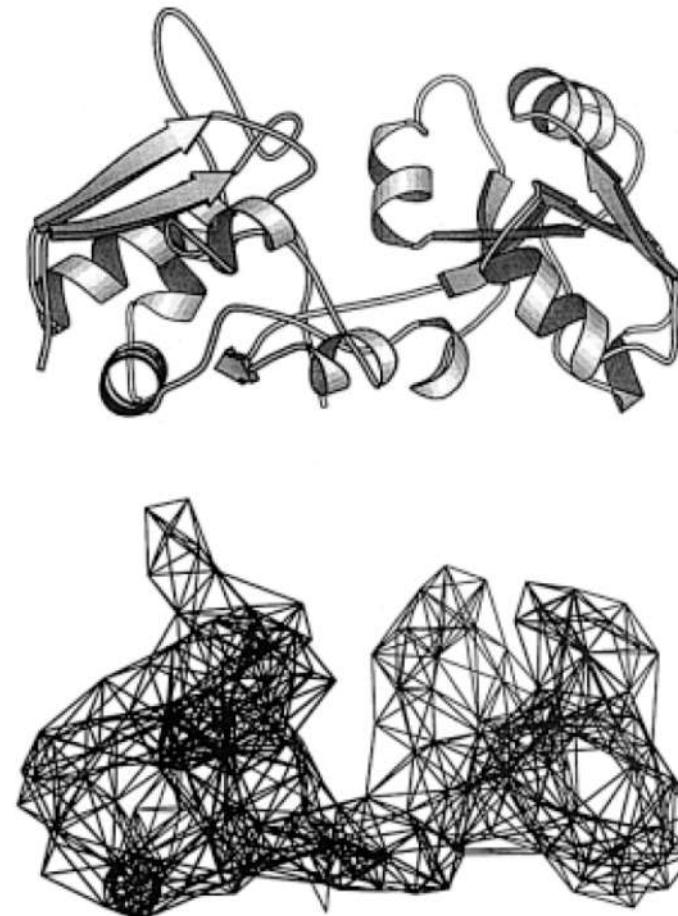
Use of randomness and repeated sampling.
Explores the conformational space not related to time.

Coarse grained MD



Normal Modes or Elastic Network Models

Analyse vibration modes of the entire protein.



Elastic Network Model

A screenshot of the elNémo web interface. At the top, there are logos for CNRS, Nantes Université, and US2B. Below the logos is a navigation bar with links: home, start a new run, job status, references&downloads, examples, and help. A message in green text says "Should you encounter any unexpected behaviour, please let us know." A message in red text says "elNémo will be down on july 17th and 18th. Sorry for the inconvenience." On the right side, there is a decorative graphic of a molecular structure. At the bottom left, there is a button labeled "[START A NEW RUN]".

Welcome to *elNémo* !

elNémo is the Web-interface to the *Elastic Network Model* (ENM), a fast and simple way for computing the low frequency normal modes of a macromolecule ([Tirion, 1996](#)). Note that, thanks to the RTB approximation ([Durand et al., 1994](#); [Tama et al., 2000](#)), this server can perform calculations for all-atom systems.

[START A NEW RUN]

<http://www.sciences.univ-nantes.fr/elnemo/index.html>

Thank you!