UNIVERSITY OF CAMBRIDGE

Looking at Protein Structures

Day 8: Tuesday 29th March

This talk

- Representations of proteins
- Mapping properties onto proteins
- Accessing protein structures
- Software for viewing protein structures
- Analysing structural similarity
- Inspecting protein interfaces

Recap of protein structure

- Proteins are polymers of amino acids
- The sequence of a proteins determines how it folds
- Proteins adopt regular secondary structure
- The arrangement of secondary structural motifs determines protein 3D structure
- Some proteins oligomerise or interact with ligands that are integral to their structure

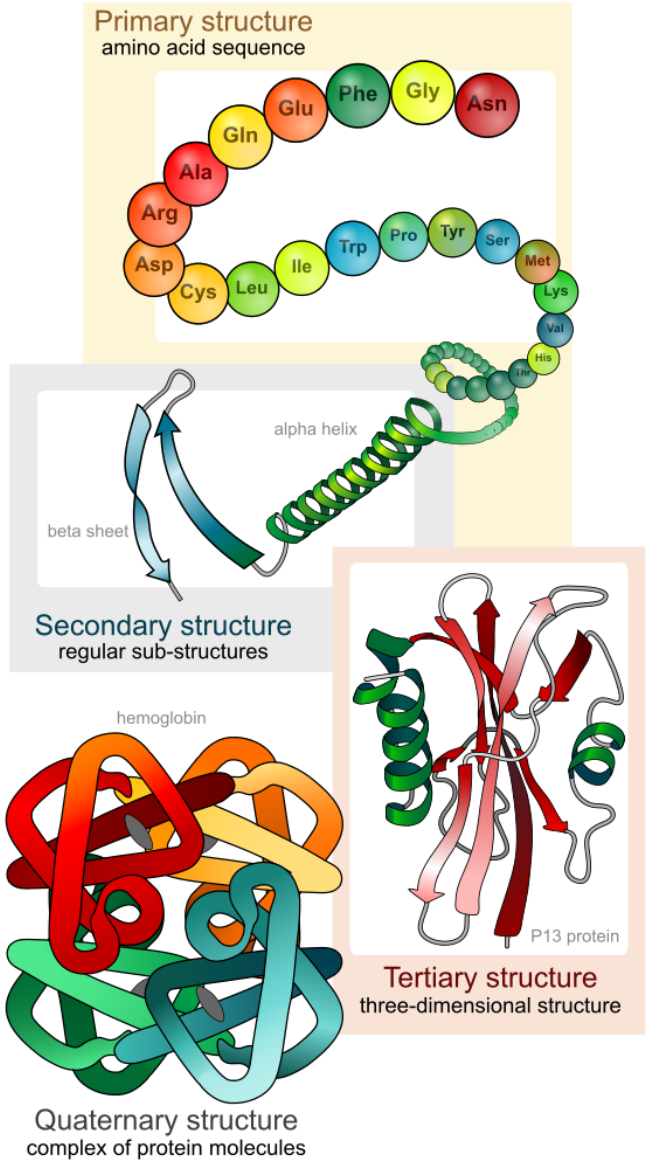


Image: LadyofHats (Wikimedia)

Interactions between amino acids drive protein folding

- Hydrogen bonds
  - Hydrogen shared between electronegative ‘donor’ and ‘acceptor’ atom with lone electron pair

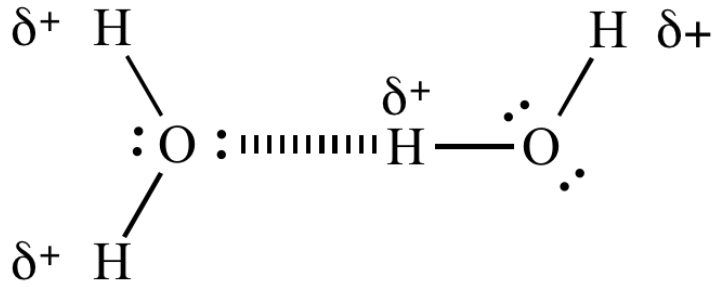


Image: https://www.chem.ucla.edu/~harding/IGOC/H/hydrogen_bond.html

Interactions between amino acids drive protein folding

- Hydrogen bonds
  - Hydrogen shared between electronegative ‘donor’ and ‘acceptor’ atom with lone electron pair
- Salt bridges
  - Between side chains with formal negative and positive charges

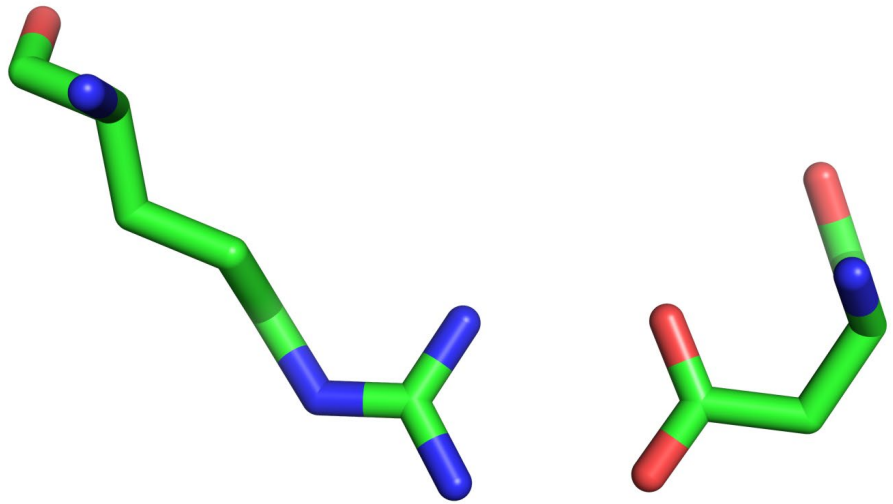


Image: Rubber Duck (Wikimedia)

Interactions between amino acids drive protein folding

- Hydrogen bonds
  - Hydrogen shared between electronegative ‘donor’ and ‘acceptor’ atom with lone electron pair
- Salt bridges
  - Between side chains with formal negative and positive charges
- Hydrophobic interaction
  - Van der Waals interactions
- Hiding hydrophobic residues from water in protein core drives folding

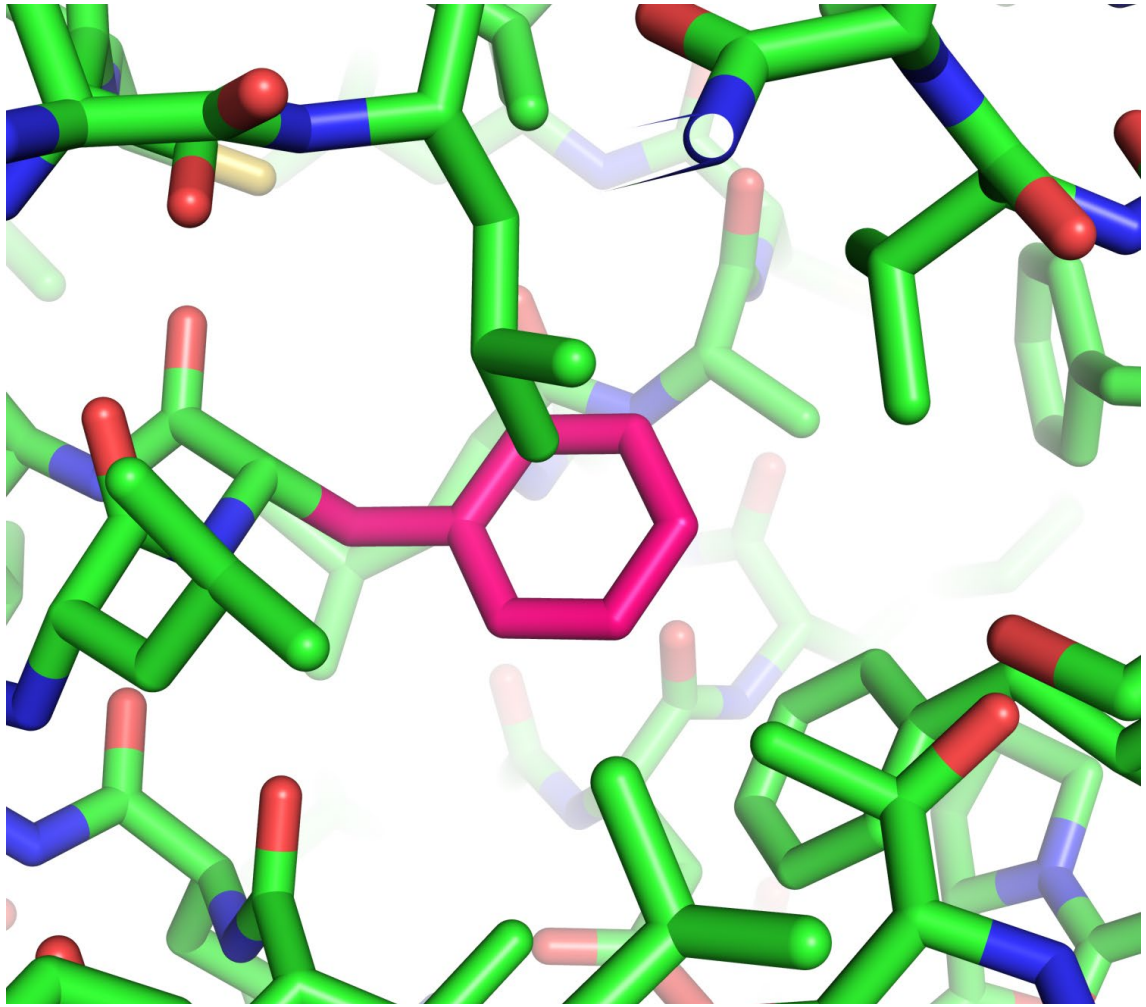
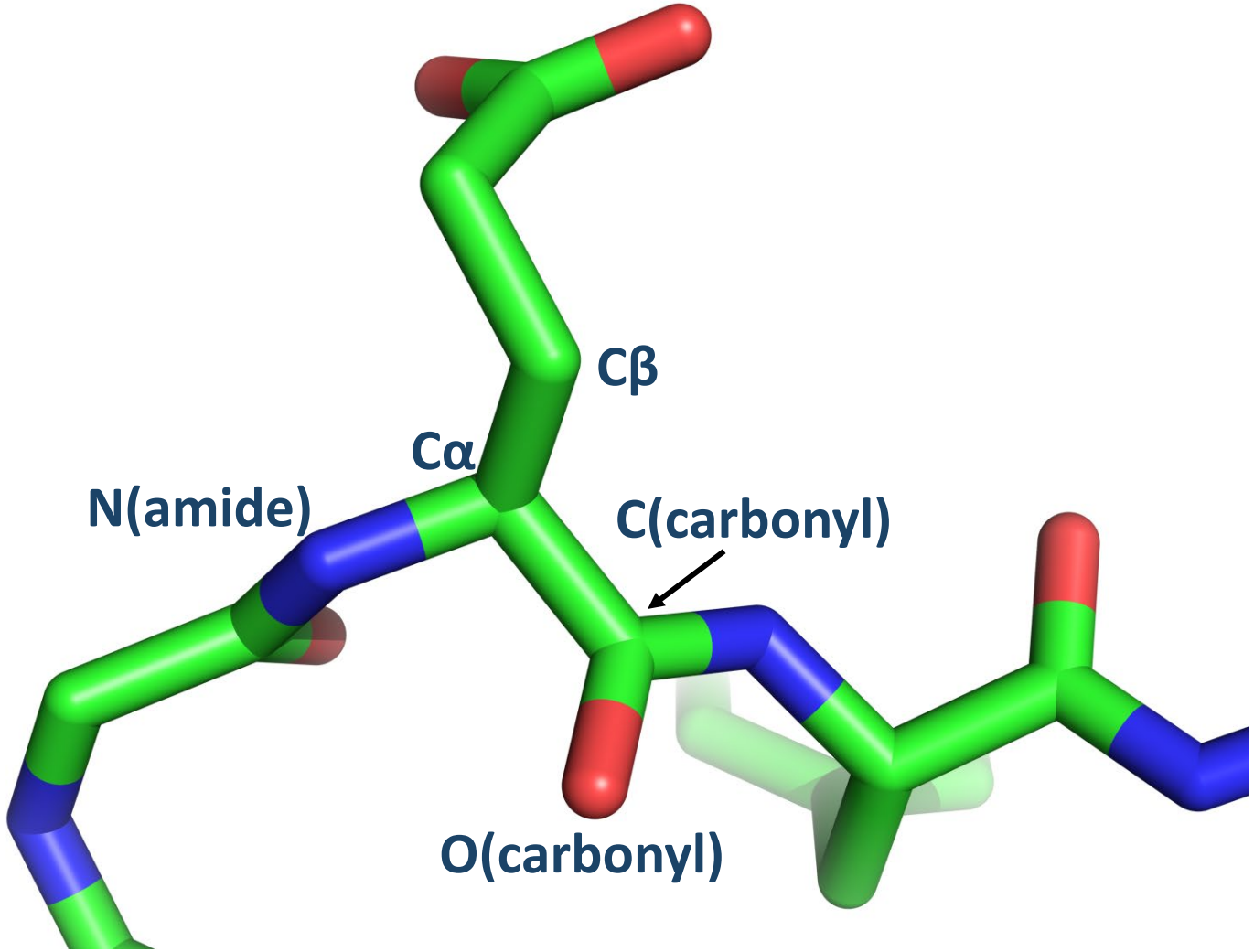


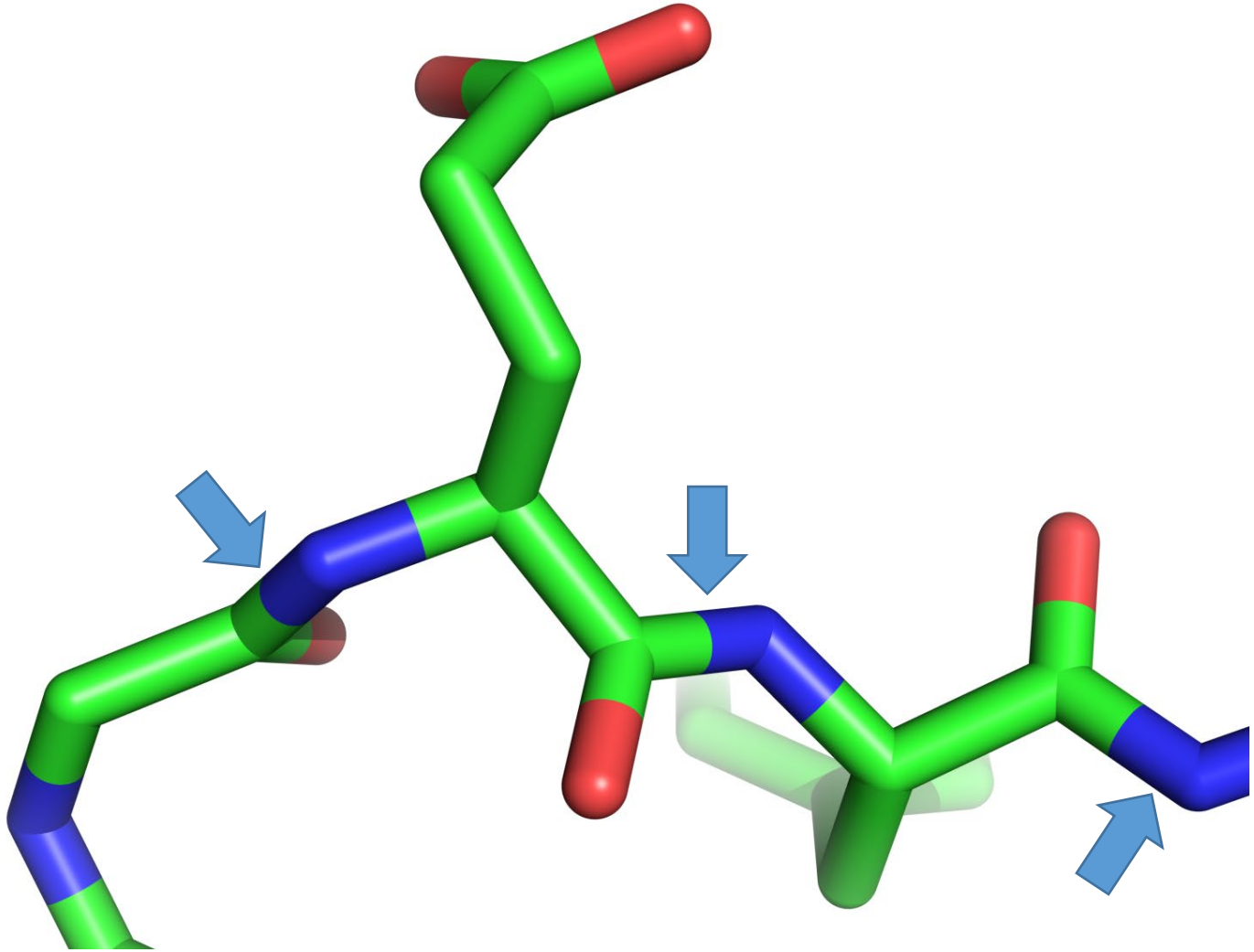
Image: Rubber Duck (Wikimedia)

Amino acid polymers



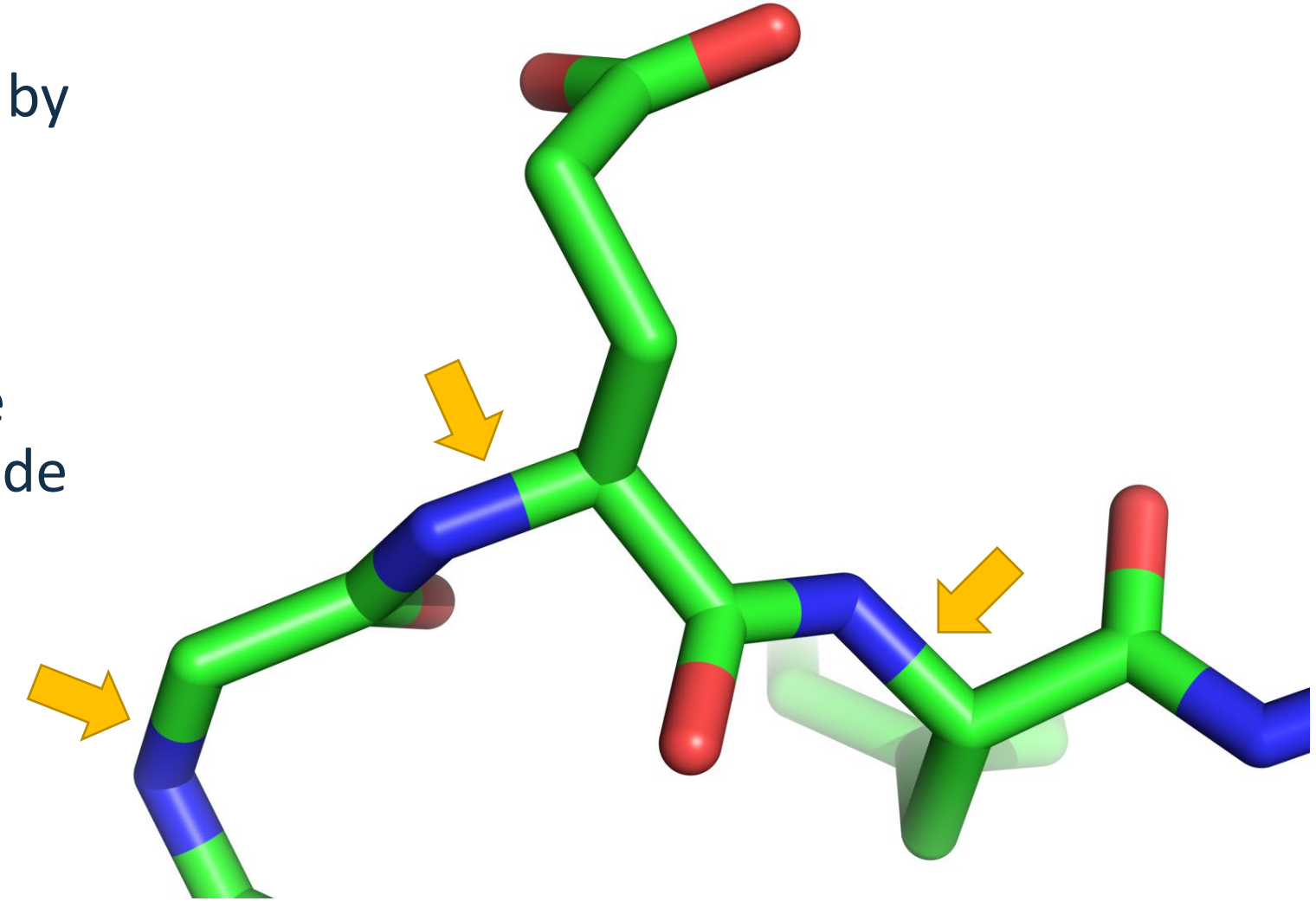
Amino acid polymers

- Amino acids are joined by peptide bonds
  - Planar
  - *trans* (180°) or *cis* (0°)



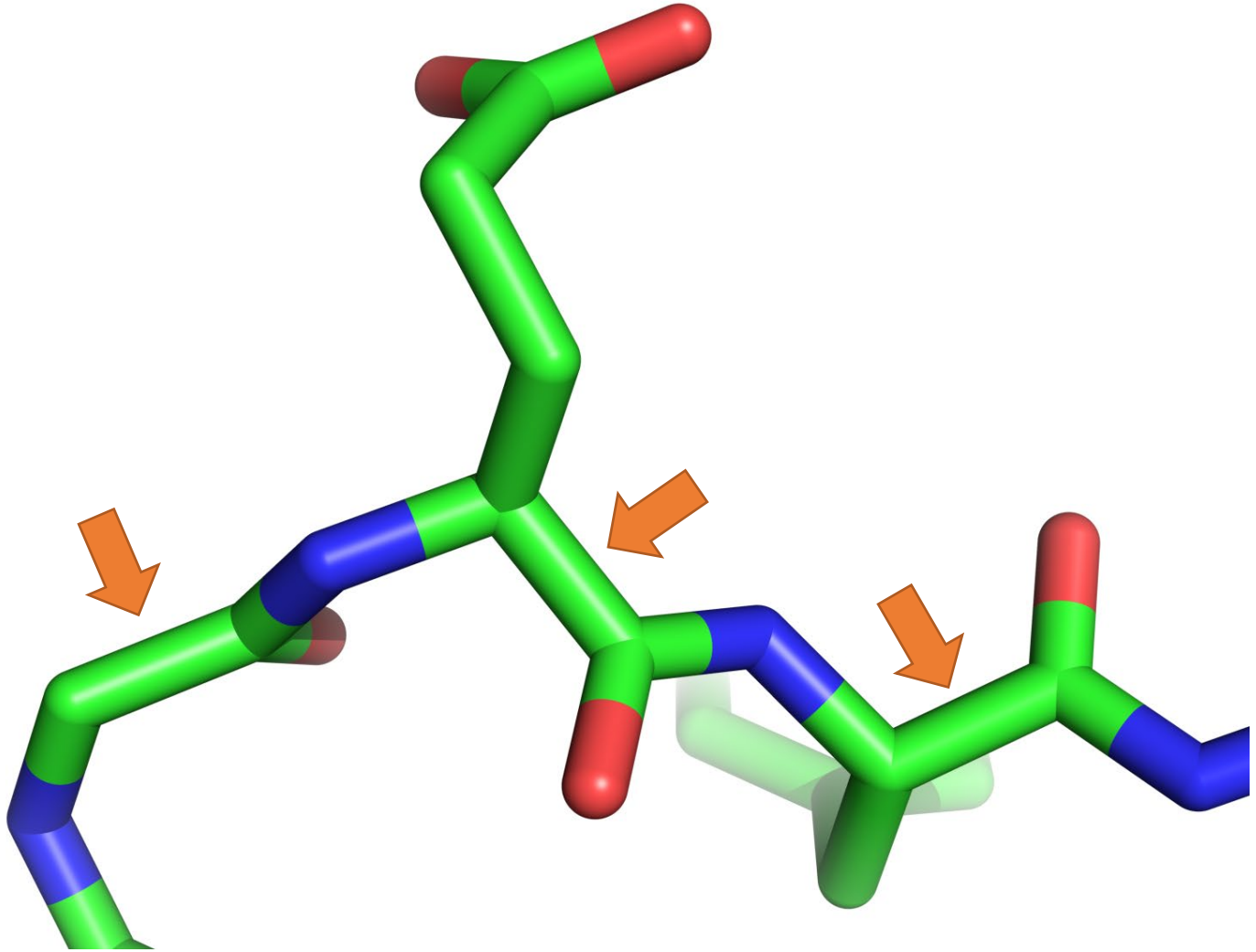
Amino acid polymers

- Amino acids are joined by peptide bonds
  - Planar
  - *trans* (180°) or *cis* (0°)
- Angle between peptide (amide nitrogen) and side chain is phi (ϕ)



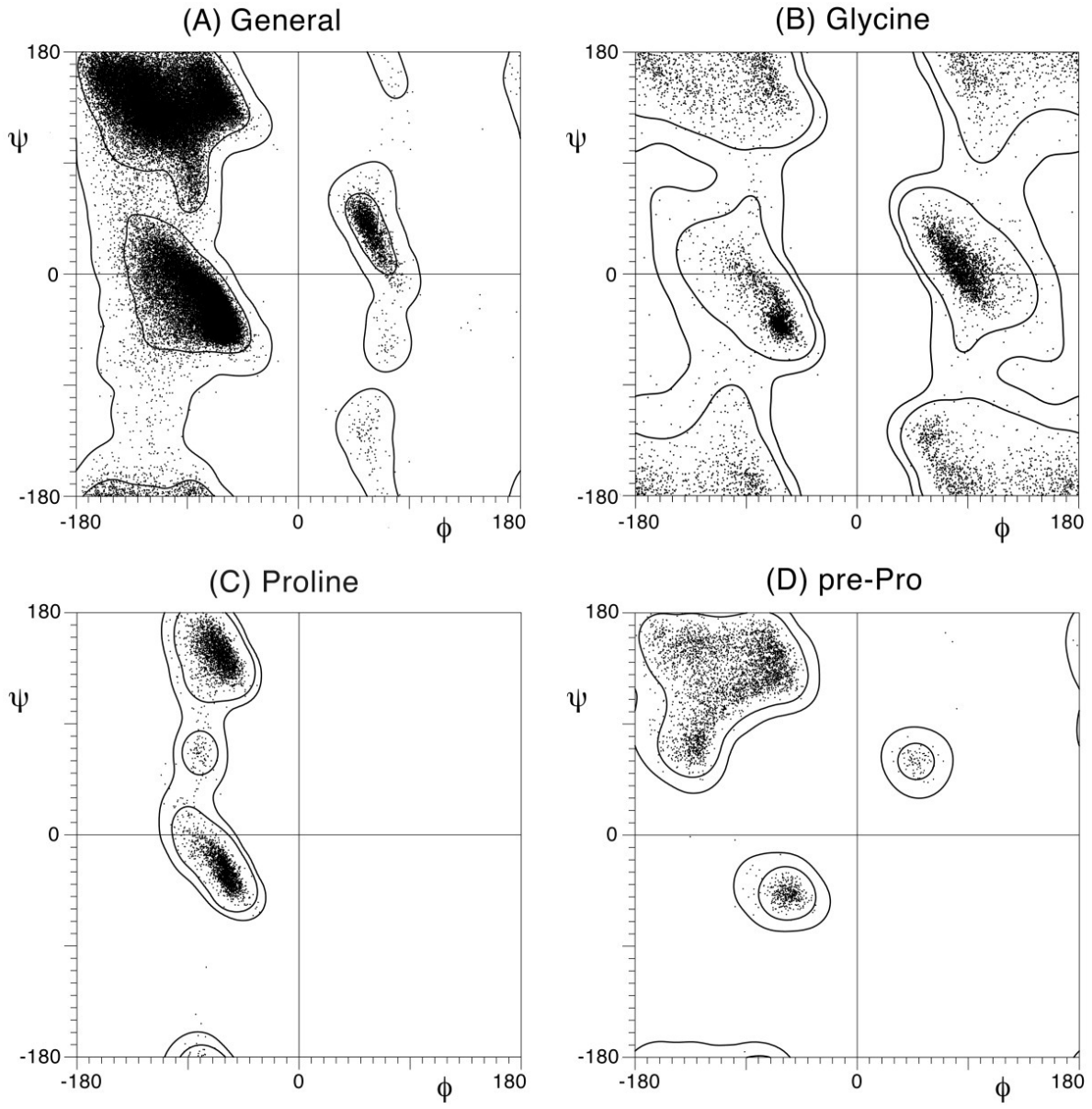
Amino acid polymers

- Amino acids are joined by peptide bonds
  - Planar
  - *trans* (180°) or *cis* (0°)
- Angle between peptide (amide nitrogen) and side chain is phi (ϕ)
- Angle between side chain and peptide (carbonyl oxygen) is psi (ψ)



Ramachandran plot

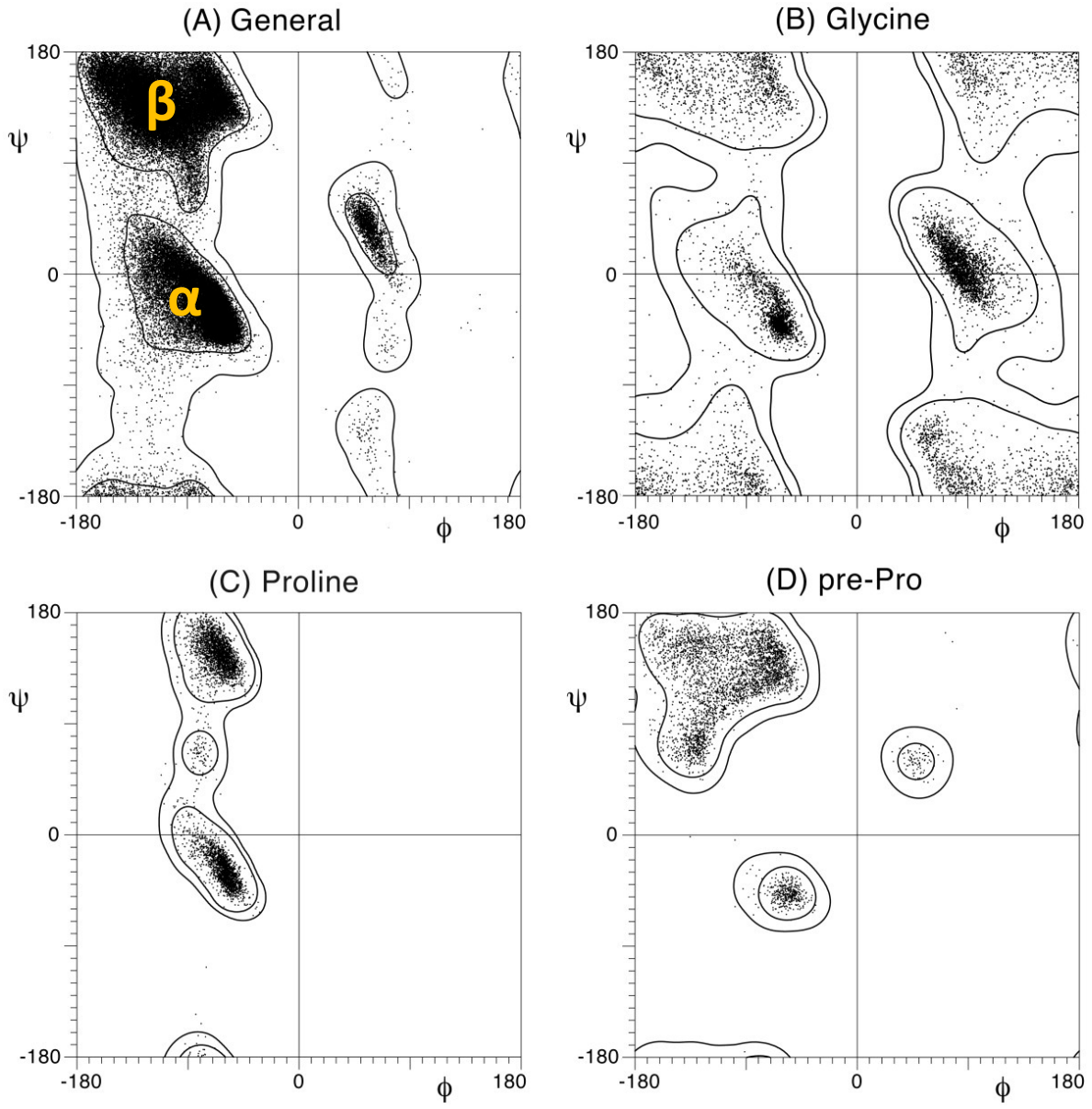
- Only certain ψ/ϕ combinations are allowed
  - Cβ atoms would bump into each other
  - Different for glycine (no Cβ) and proline [Cδ-N(amide) bond]
- Some areas of the Ramachandran plot are highly populated



Lovell et al (2003) Proteins Struct. Func. Bioinf. 50: 437-50

Ramachandran plot

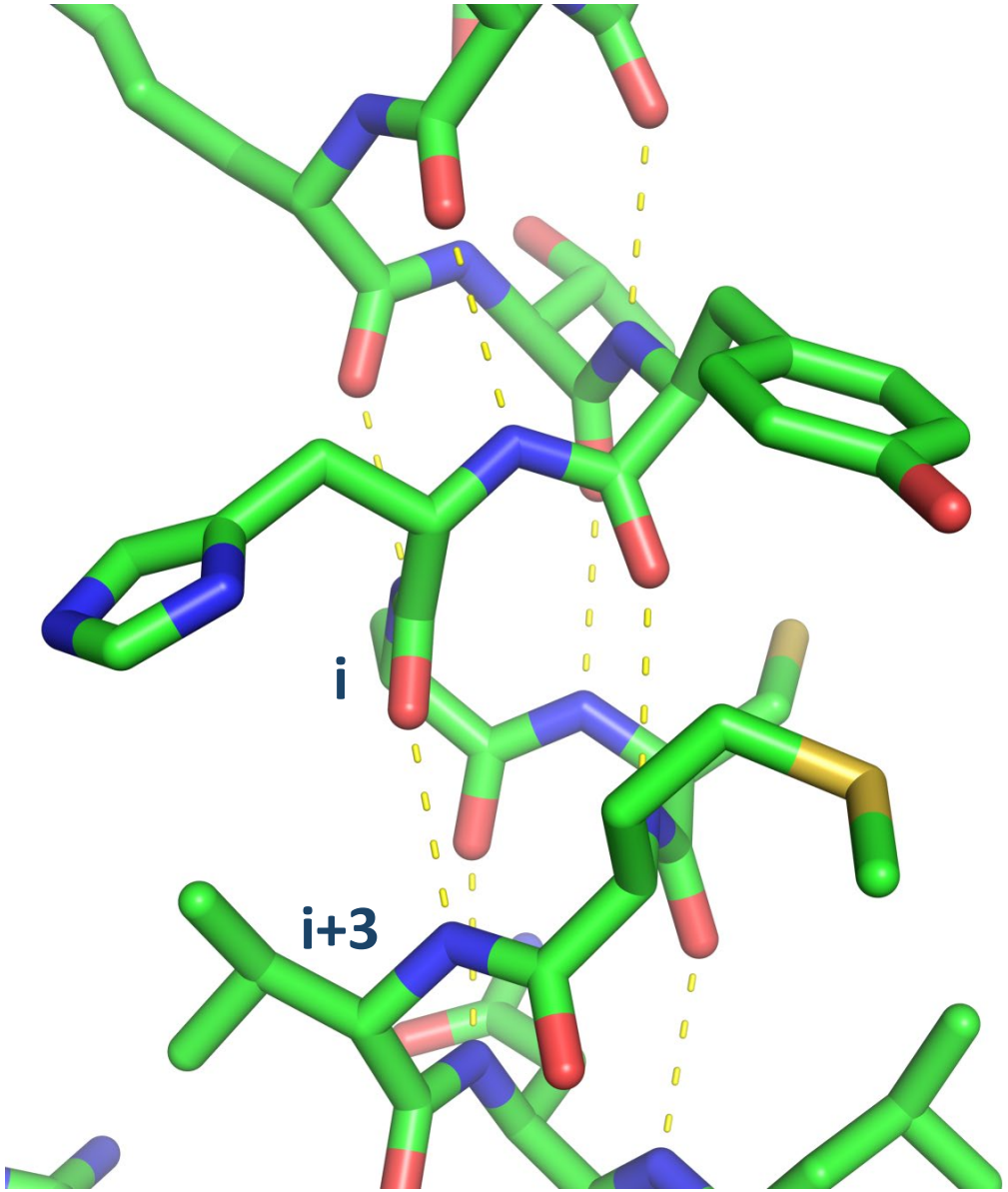
- Only certain ψ/ϕ combinations are allowed
  - Cβ atoms would bump into each other
  - Different for glycine (no Cβ) and proline [Cδ-N(amide) bond]
- Many residues occupy certain regions of the Ramachandran plot:
  - α-helix
  - β-sheet



Lovell et al (2003) Proteins Struct. Func. Bioinf. 50: 437-50

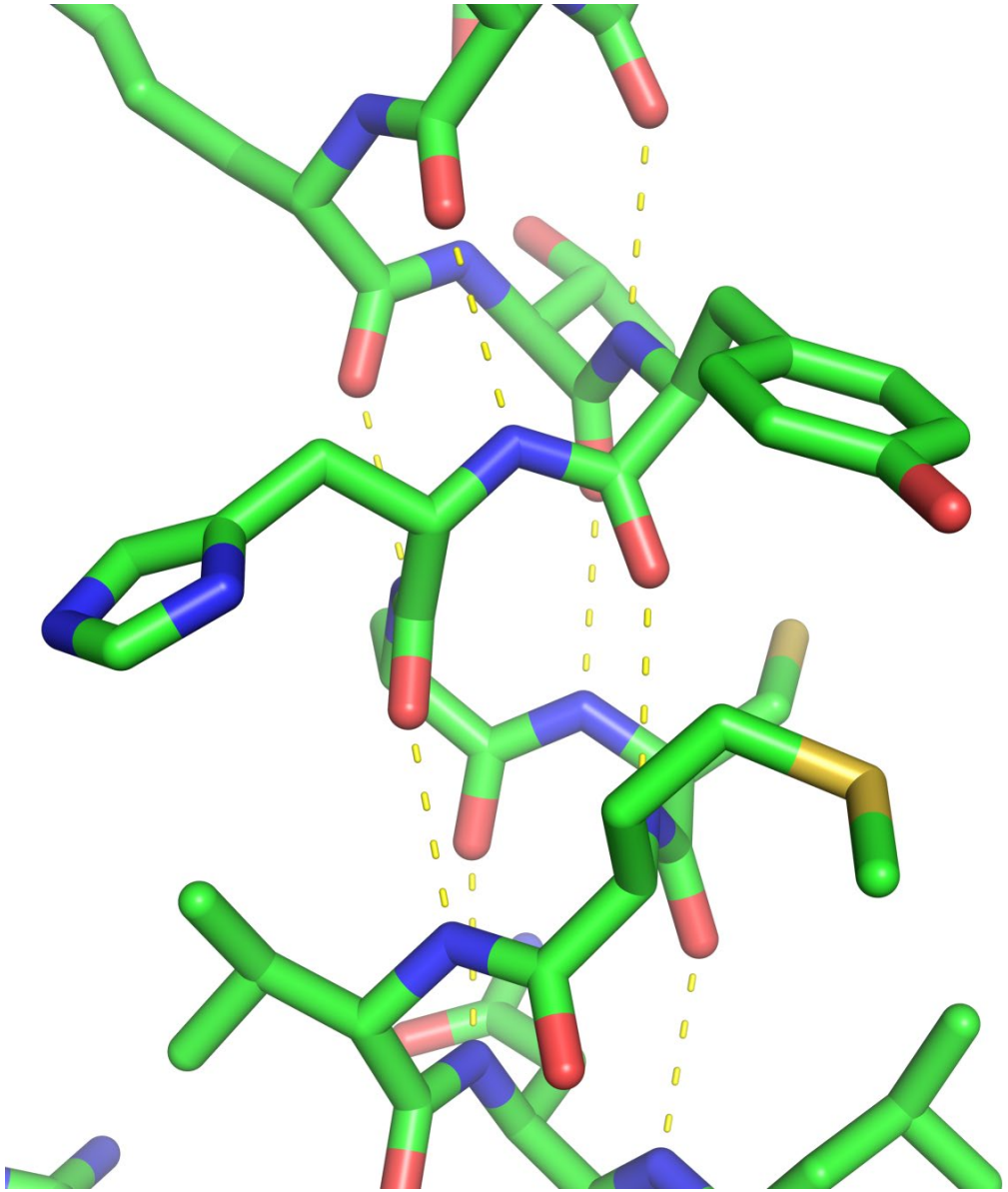
α-helix

- O(carbonyl) makes **hydrogen bond** with N(amide) of i+3 residue



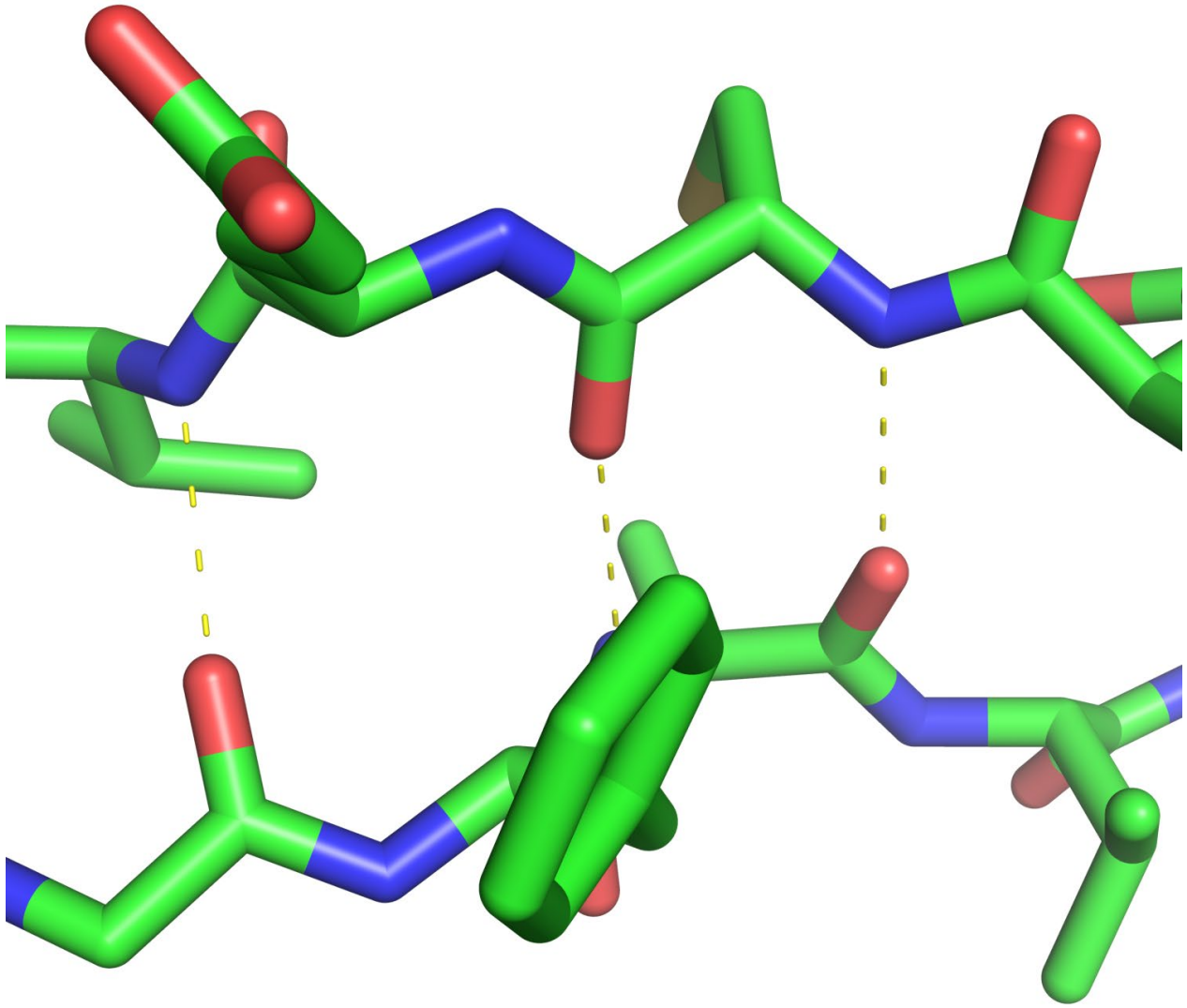
α-helix

- O(carbonyl) makes **hydrogen bond** with N(amide) of i+3 residue
- Helices are **right handed**



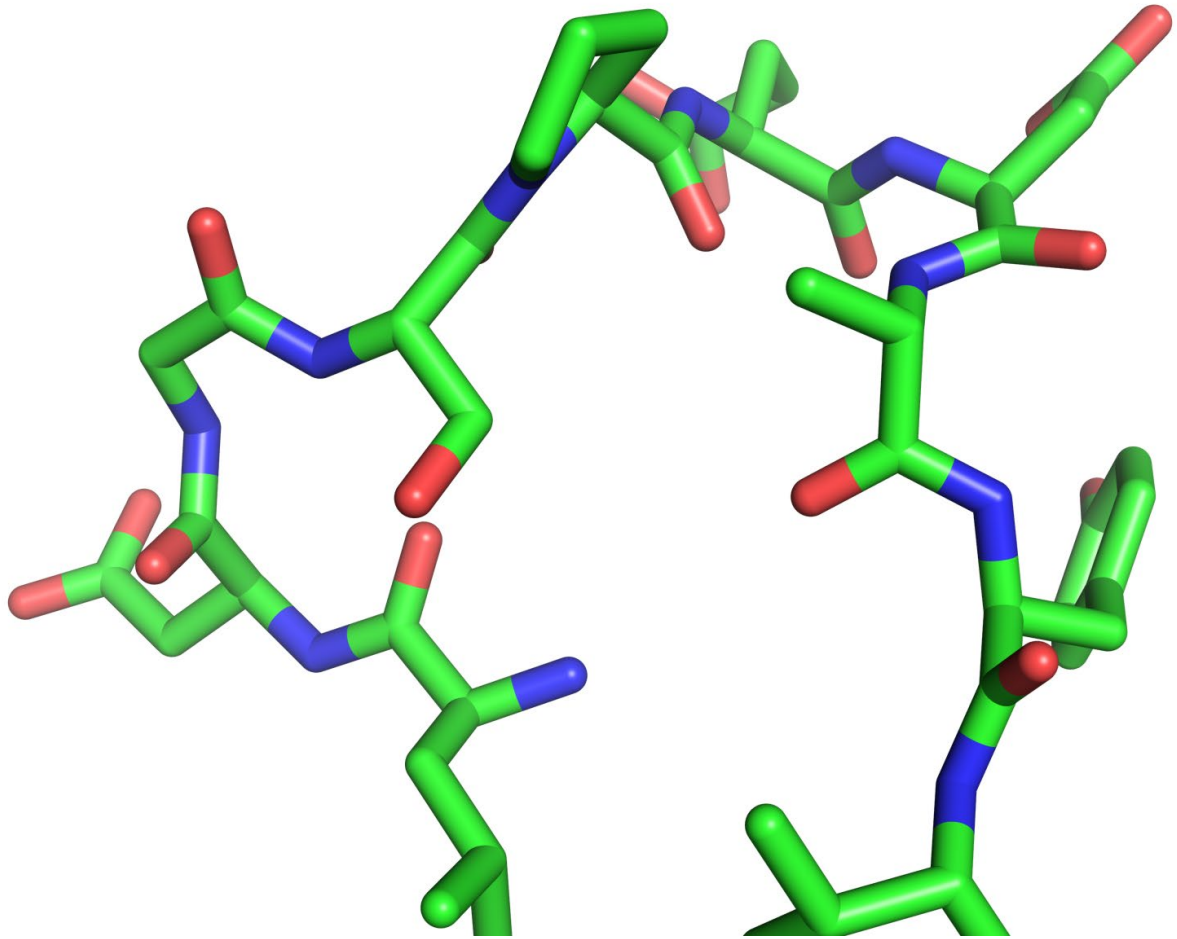
β-sheet

- Hydrogen bonds between adjacent **strands**
- **Parallel:** strands in same direction
- **Anti-parallel** strands in opposite directions



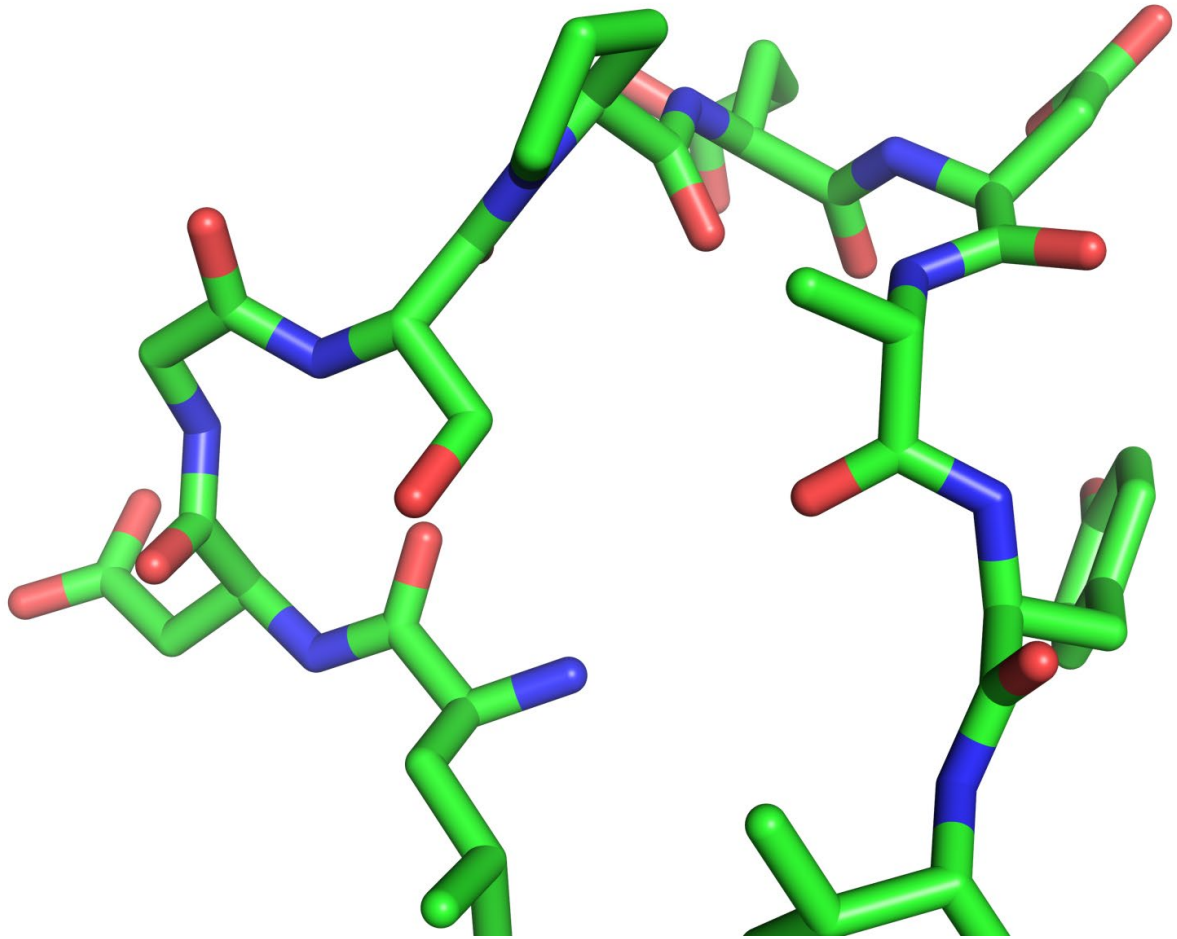
Not all amino acids are in helices or sheets

- Loops (between secondary structural elements)
- Coil (extended regions with no secondary structure)



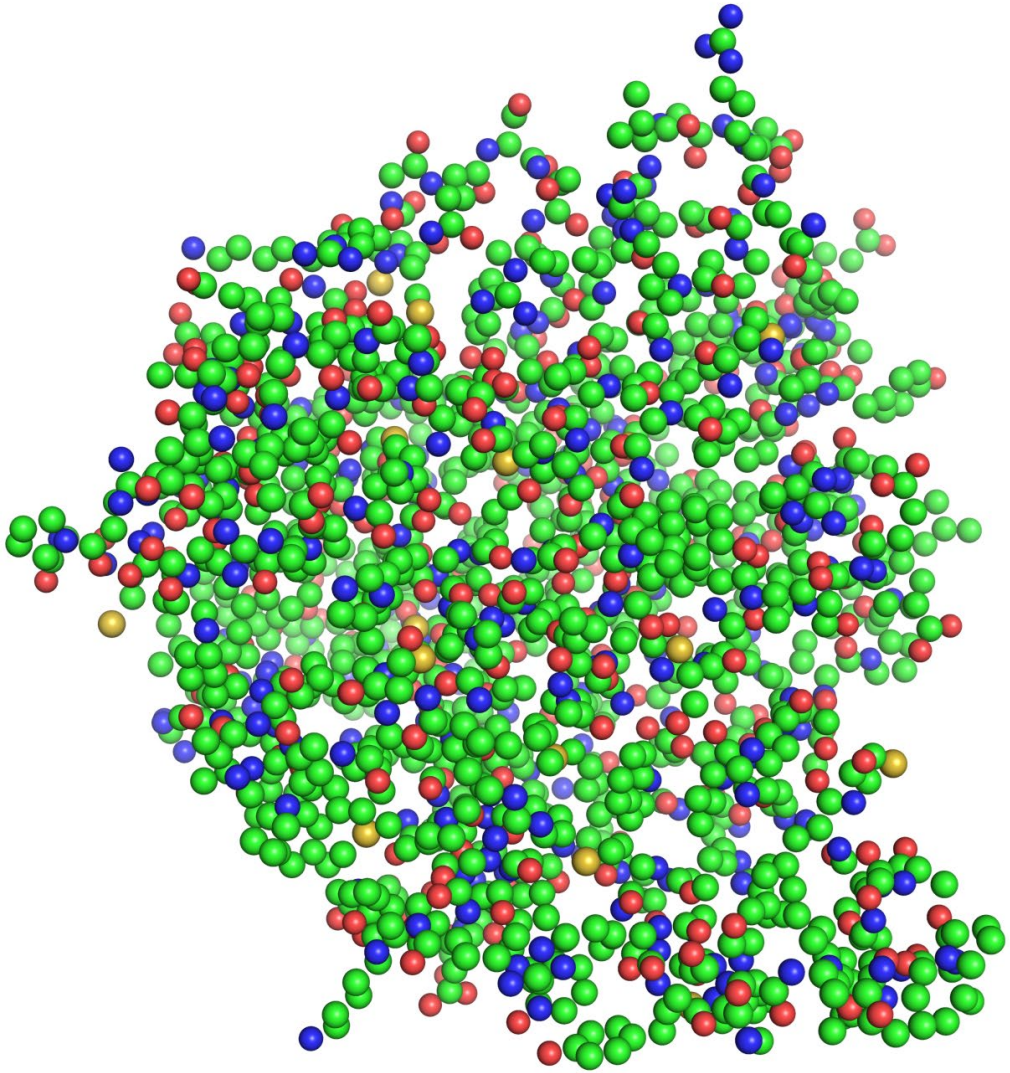
Not all amino acids are in helices or sheets

- Loops (between secondary structural elements)
- Coil (extended regions with no secondary structure)
- *Some regions of proteins have no intrinsic structure*
  - **Adopt conformation** upon binding to partner molecules
  - **Biomolecular condensates** (membrane-less organelles)



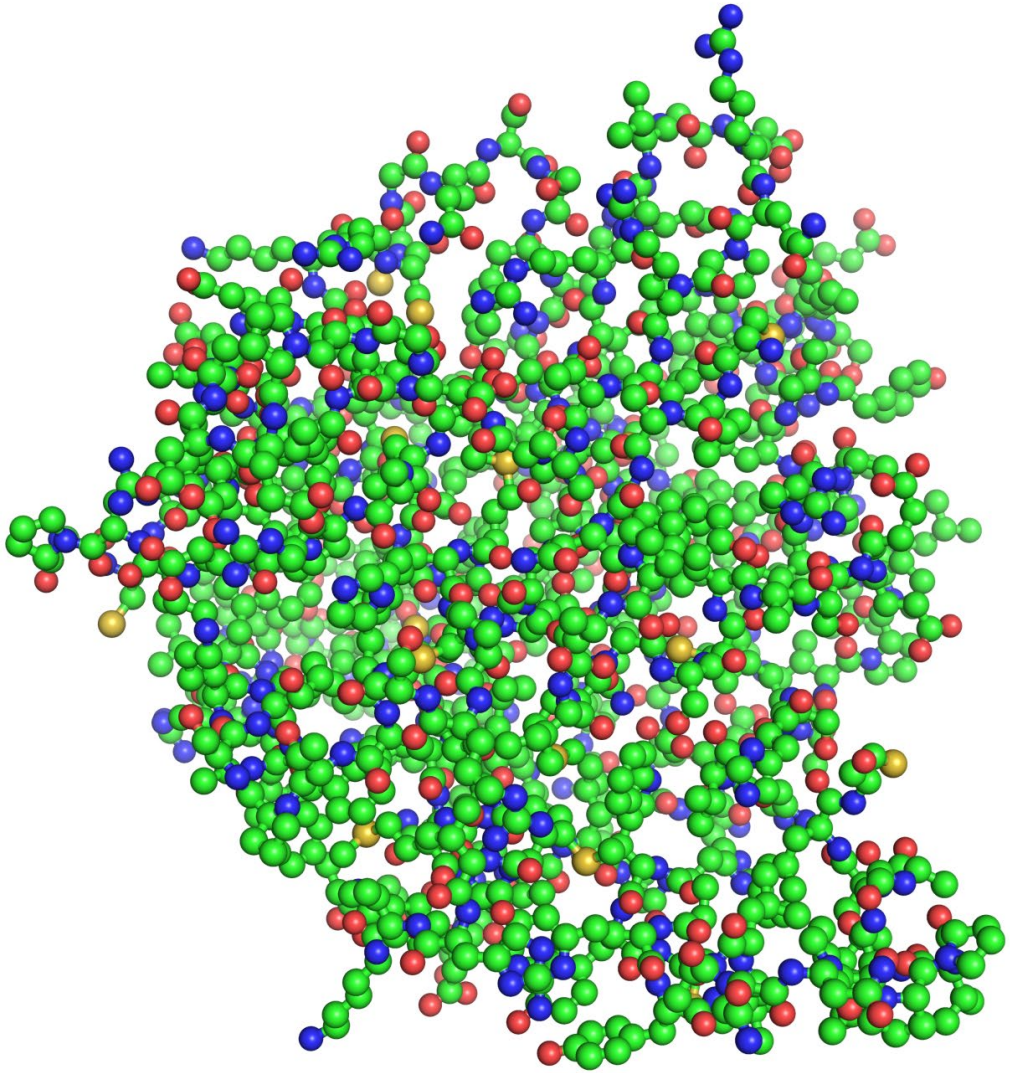
Displaying secondary structure

- Displaying just the atoms as **spheres** is very hard to interpret



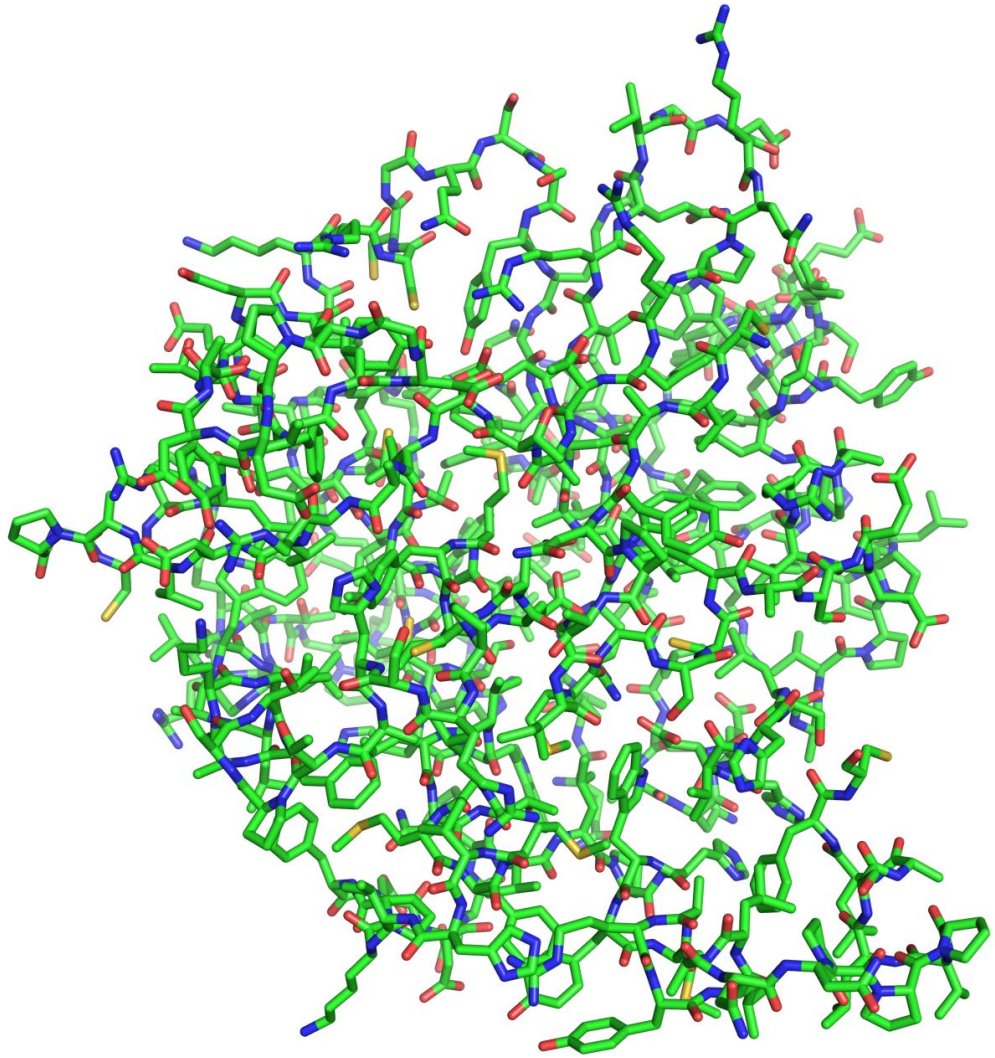
Displaying secondary structure

- Displaying just the atoms as **spheres** is very hard to interpret
- Showing bonds between the atoms as **sticks** helps



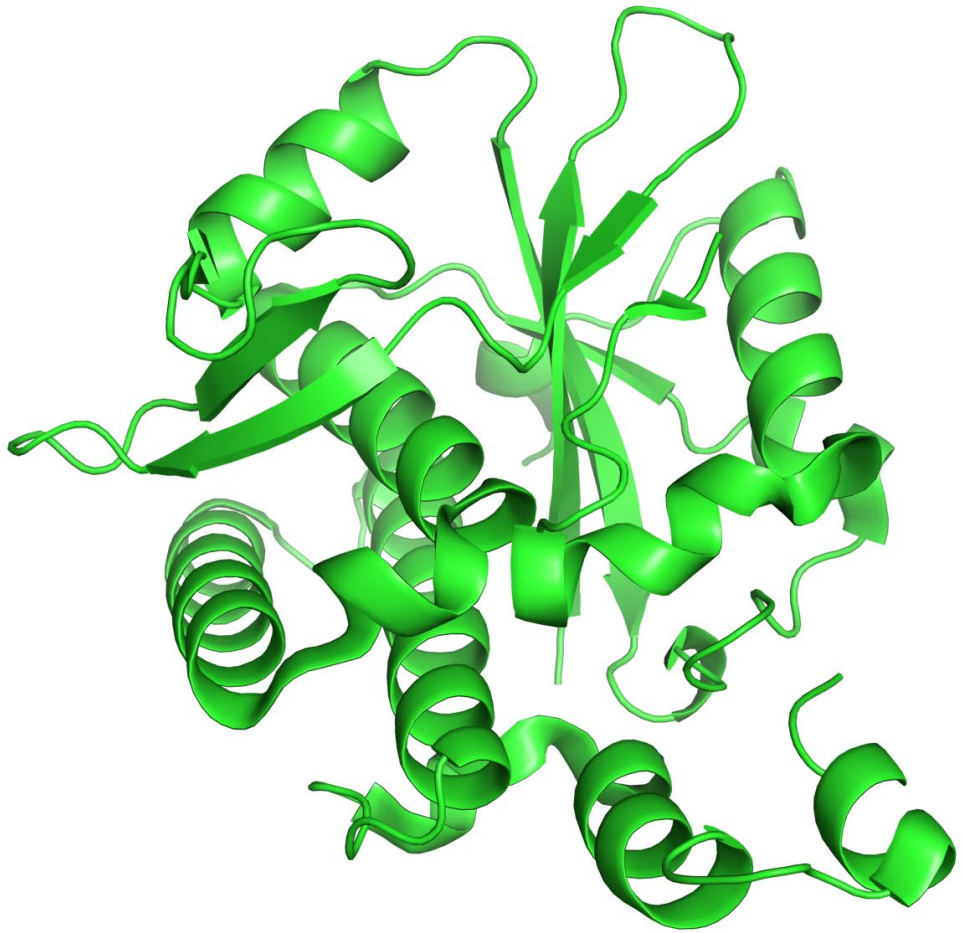
Displaying secondary structure

- Displaying just the atoms as **spheres** is very hard to interpret
- Showing bonds between the atoms as **sticks** helps



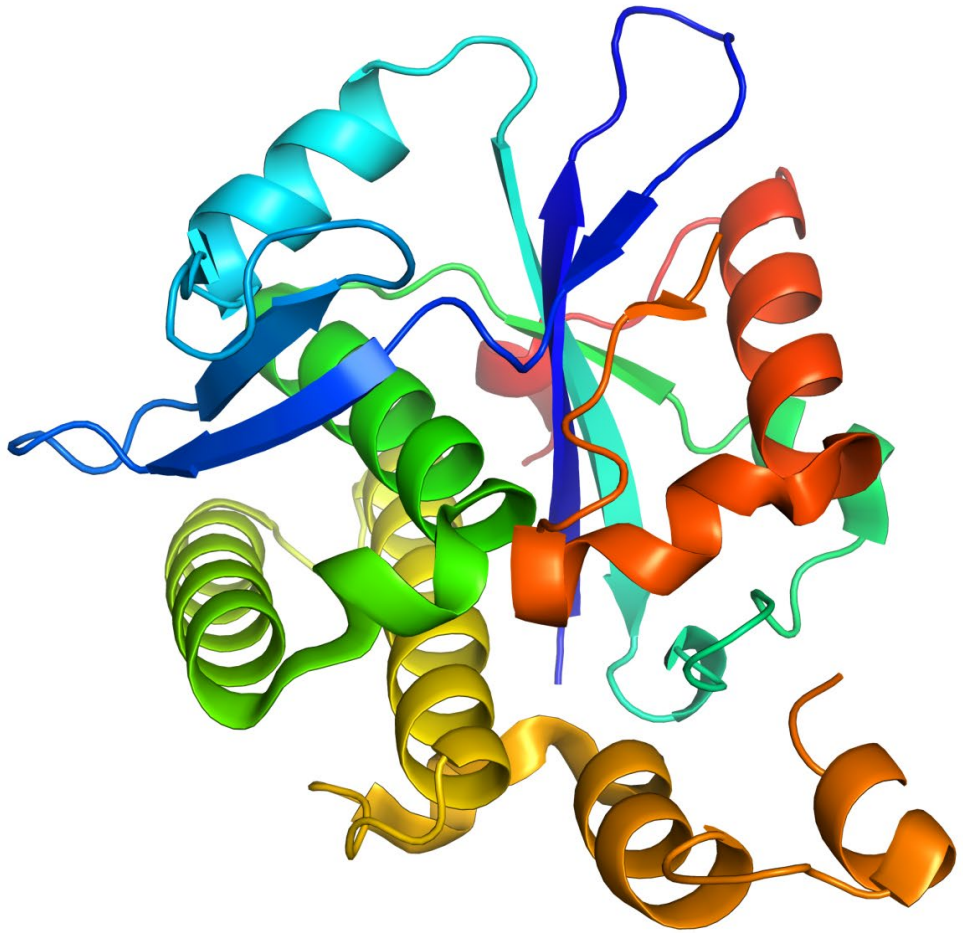
Displaying secondary structure

- Displaying just the atoms as **spheres** is very hard to interpret
- Showing bonds between the atoms as **sticks** helps
- Easier to follow the trace with a **cartoon** (or **ribbon**) representation
  - Helices as spirals
  - Sheets as arrows



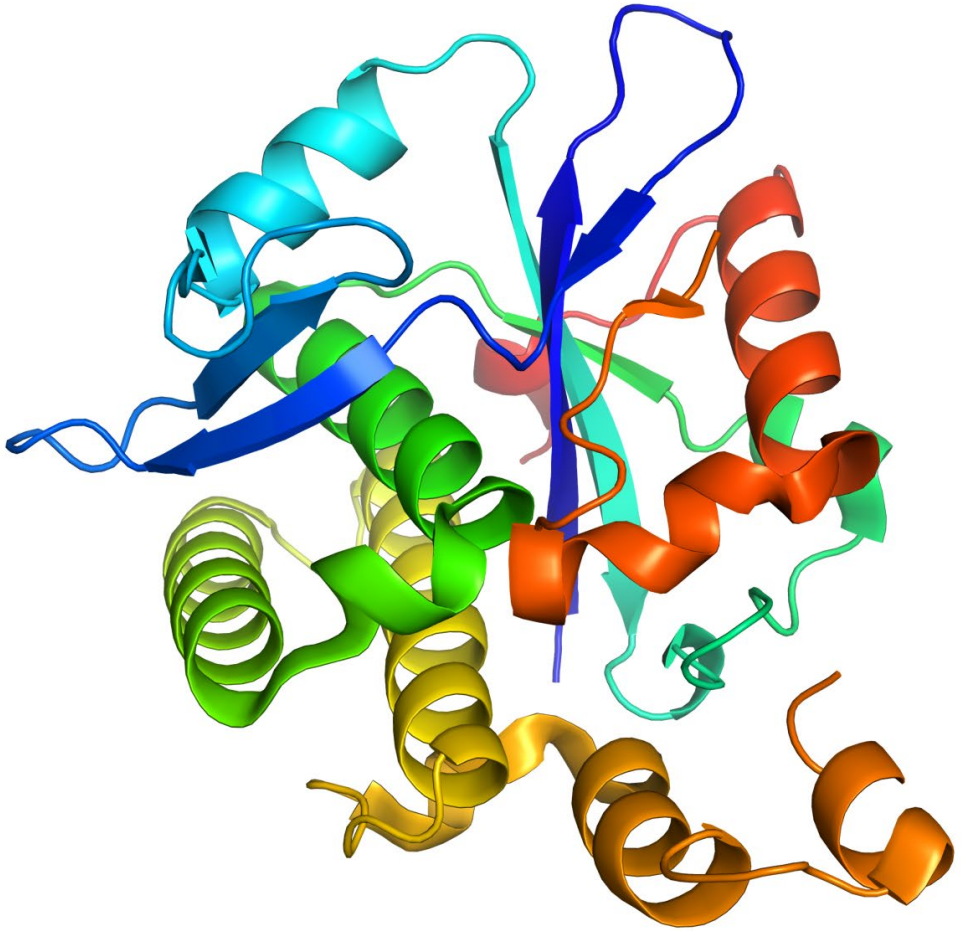
Displaying secondary structure

- Displaying just the atoms as **spheres** is very hard to interpret
- Showing bonds between the atoms as **sticks** helps
- Easier to follow the trace with a **cartoon** (or **ribbon**) representation
  - Helices as spirals
  - Sheets as arrows
- And even better if you colour as a rainbow
  - **Blue** at N terminus to **red** at C terminus



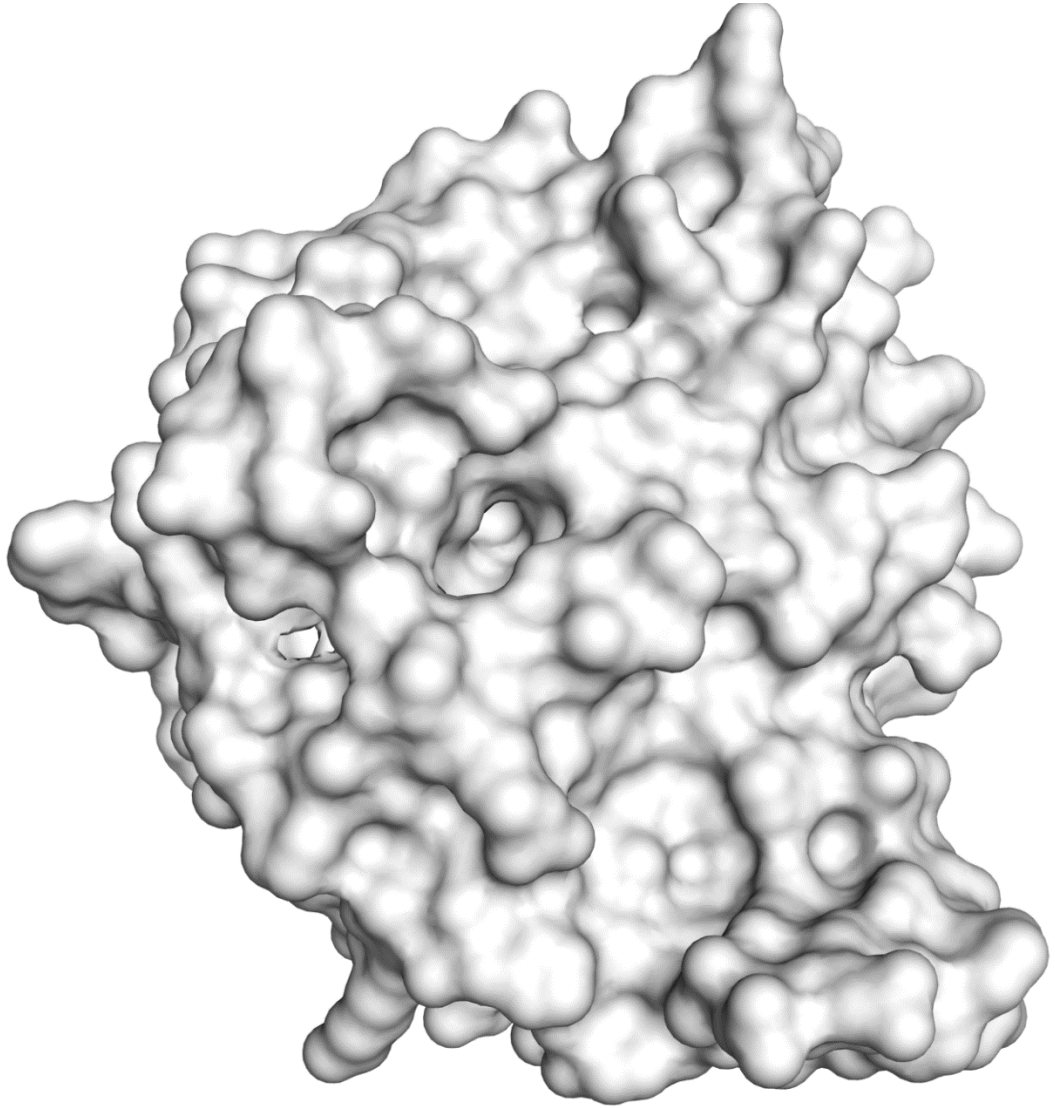
Surfaces

- Cartoons are convenient for following peptide direction
  - But proteins don’t have big gaps
  - The interior of proteins is largely hydrophobic and inaccessible to water



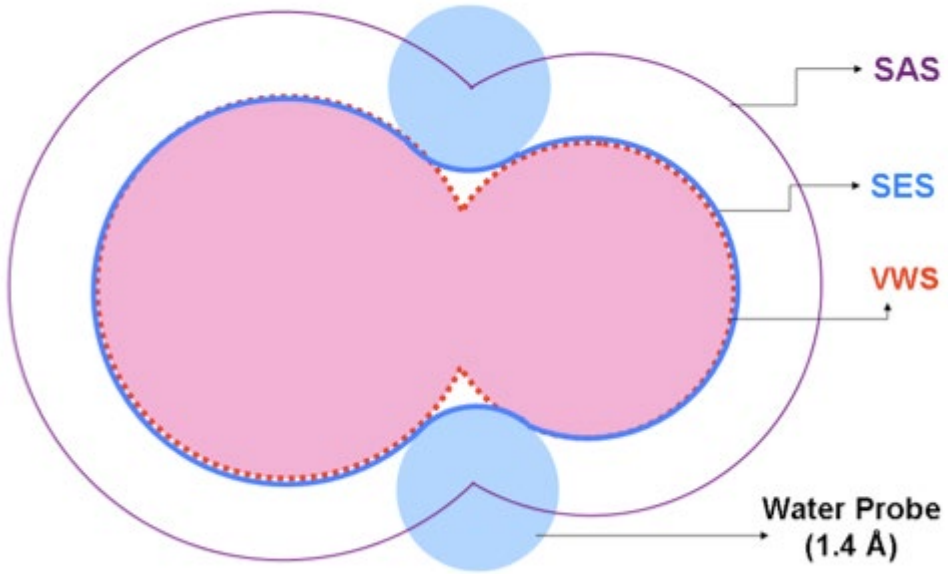
Surfaces

- Cartoons are convenient for following peptide direction
  - But proteins don’t have big gaps
  - The interior of proteins is largely hydrophobic and inaccessible to water
- Can display the surface of proteins



Surfaces

- Cartoons are convenient for following peptide direction
  - But proteins don’t have big gaps
  - The interior of proteins is largely hydrophobic and inaccessible to water
- Can display the surface of proteins
- Most common surface representation is the molecular surface (aka Connelly surface)



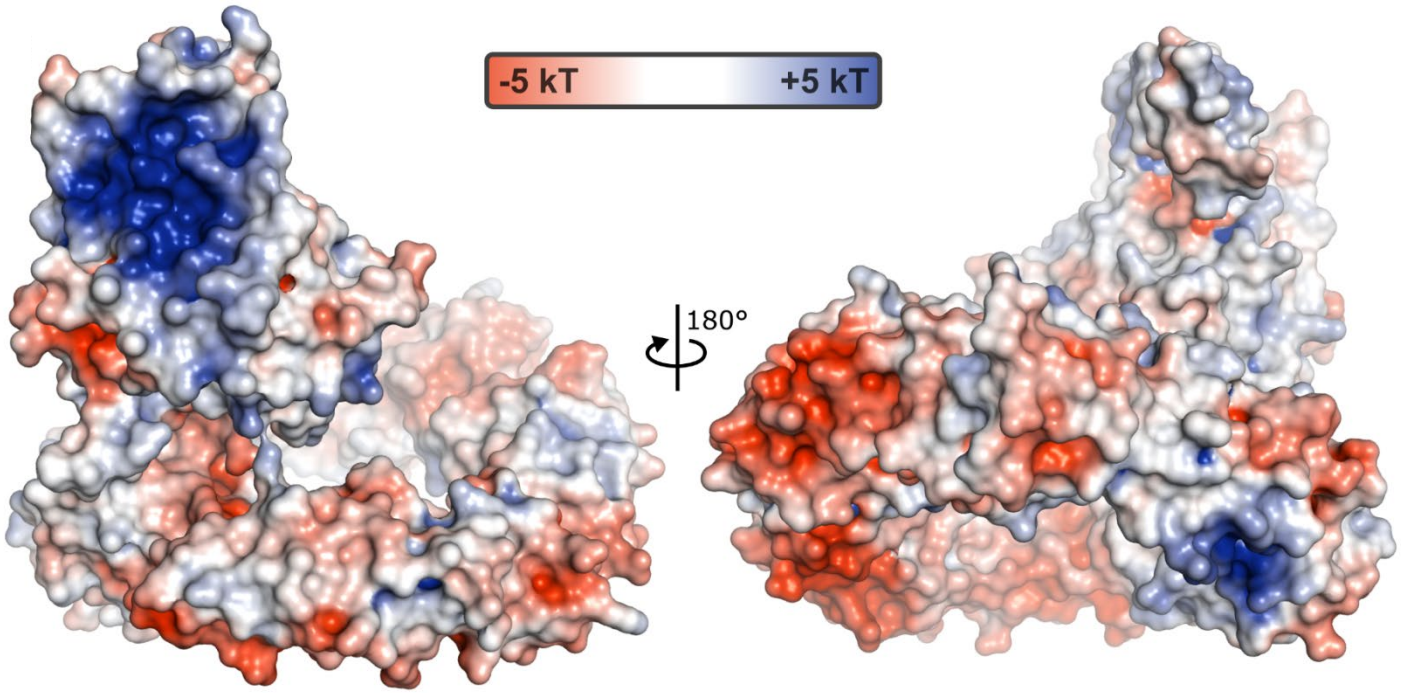
SAS: Solvent-accessible surface
SES: Molecular surface (solvent-excluded surface)
VWS: Van der Waals surface

Image: Pengyu Ren lab, UT Austin (https://biomol.bme.utexas.edu/wiki/index.php/ABS)

Surface representations

Surface representations – Electrostatic potential

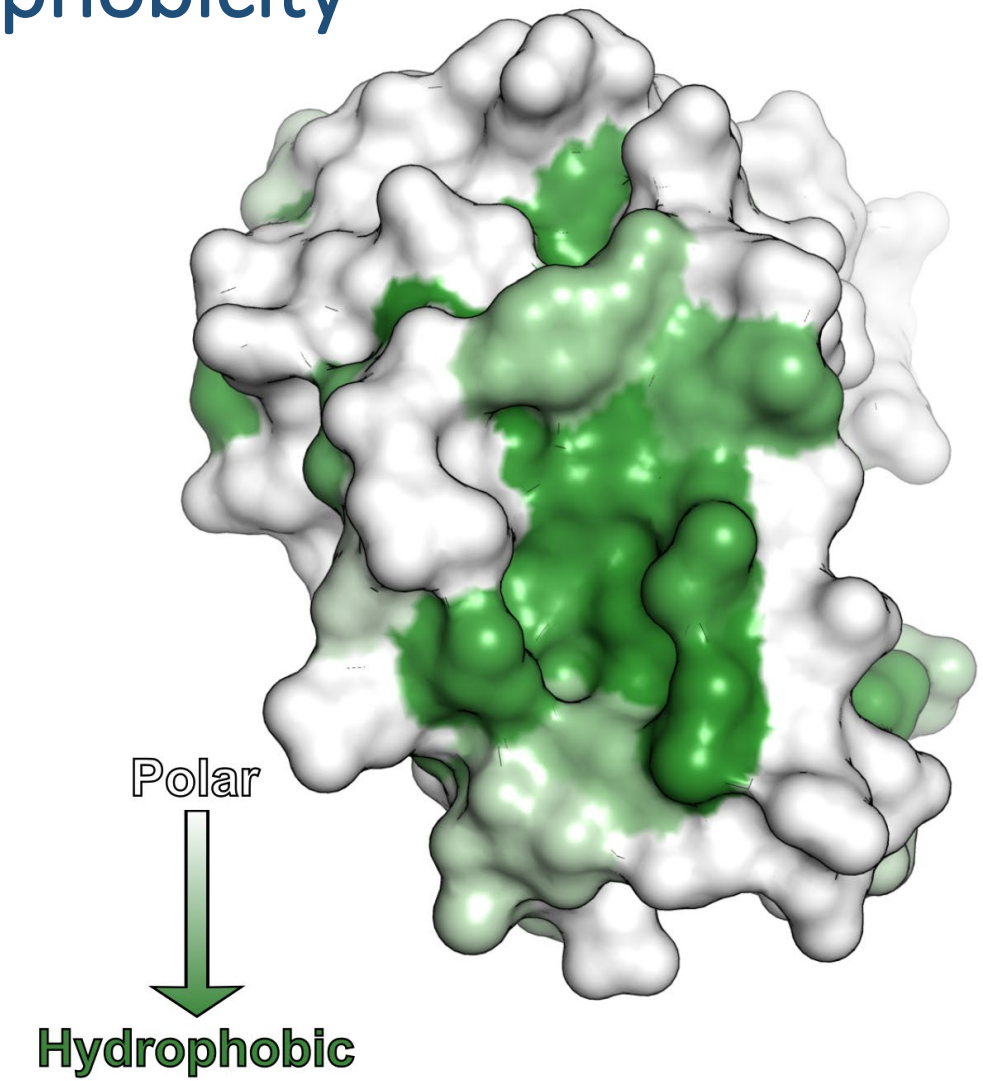
- Can map the electrostatic potential of a protein onto its molecular surface
  - **Blue** = basic = positive charge
  - **Red** = acidic = negative charge
- Understanding protein electrostatic charge can inform biology
  - Long-range interactions between macromolecules
  - Interaction patches for charged small molecules



Gao et al. (2022) *Journal of General Virology*, 103: 001716

Surface representations – Hydrophobicity

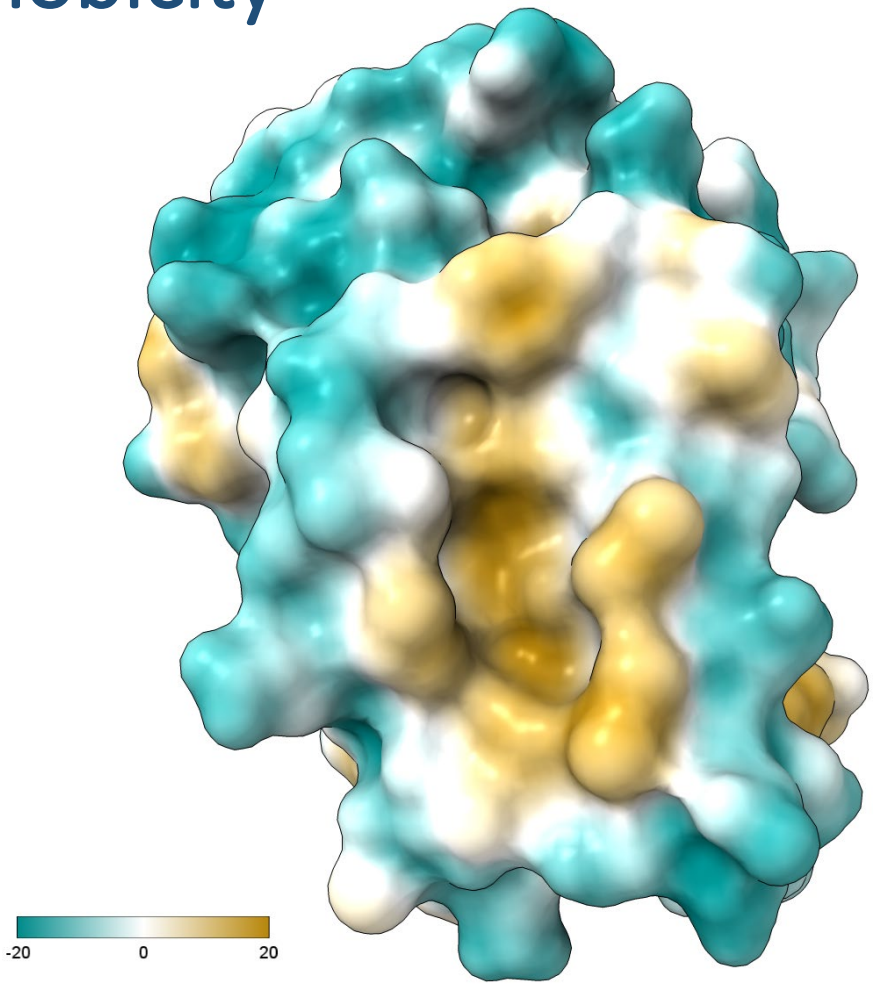
- Surface-exposed hydrophobic patches are energetically unfavourable
- Can colour surfaces of proteins to show the hydrophobicity of each residue at the surface



Neidel et al. (2015) *Journal of Biological Chemistry*, 290: 5991–6002

Surface representations – Hydrophobicity

- Surface-exposed hydrophobic patches are energetically unfavourable
- Can colour surfaces of proteins to show the hydrophobicity of each residue at the surface
- Can also show molecular lipophilicity potential
  - Slightly more sophisticated, per-atom potential



Protein Data Bank (PDB)

- Worldwide repository of biomolecular structural data
  - Established in 1971
- Hold both the atomic models plus the experimental data (maps) used to generate these models
- Hosted by three different institutions:
  - RCSB: https://www.rcsb.org/
  - EBI: https://www.ebi.ac.uk/pdbe/
  - PDBj: https://pdbj.org/
  - Each site has different tools and search, but the underlying data is the same

Protein Data Bank (PDB)

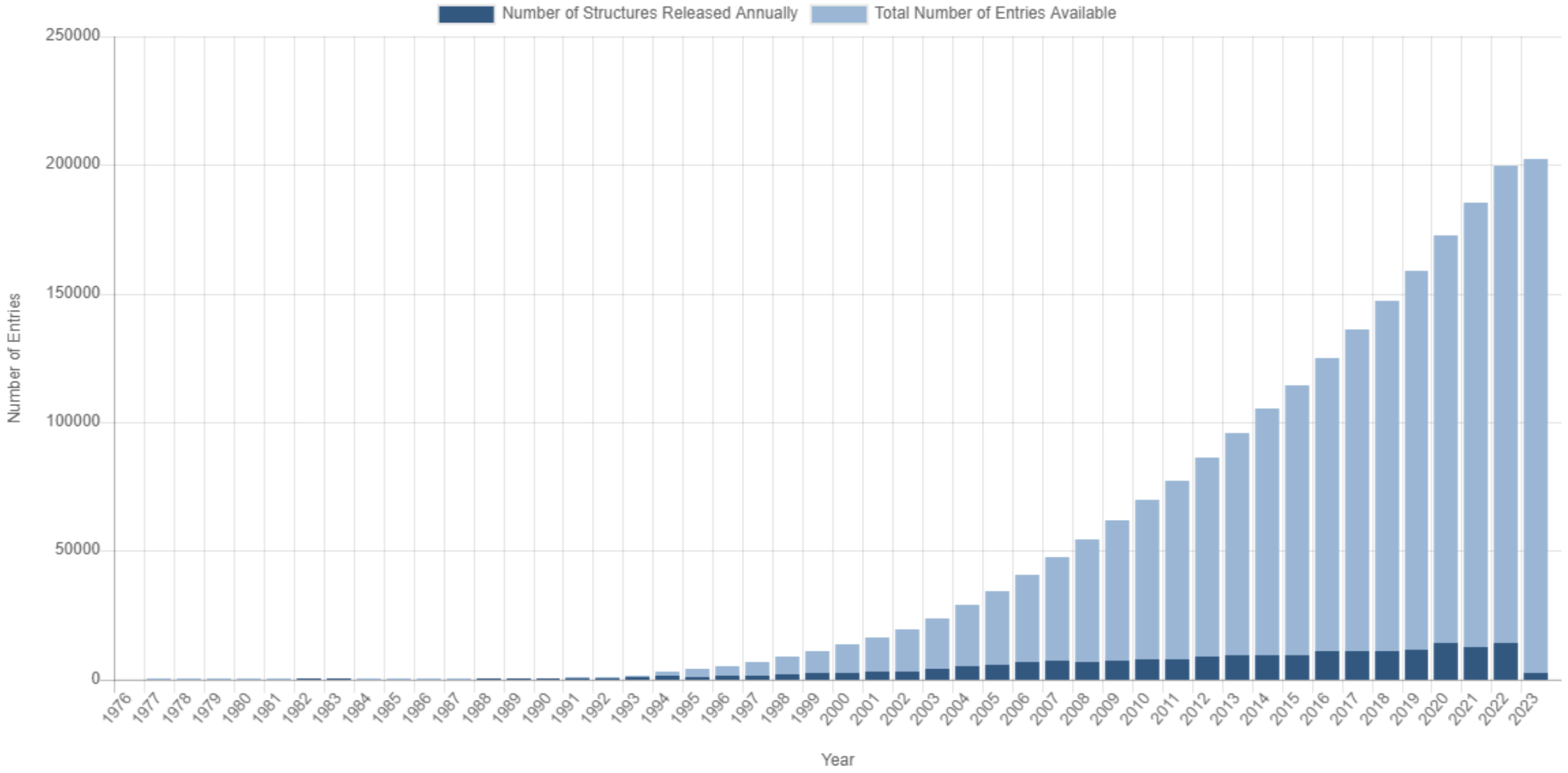
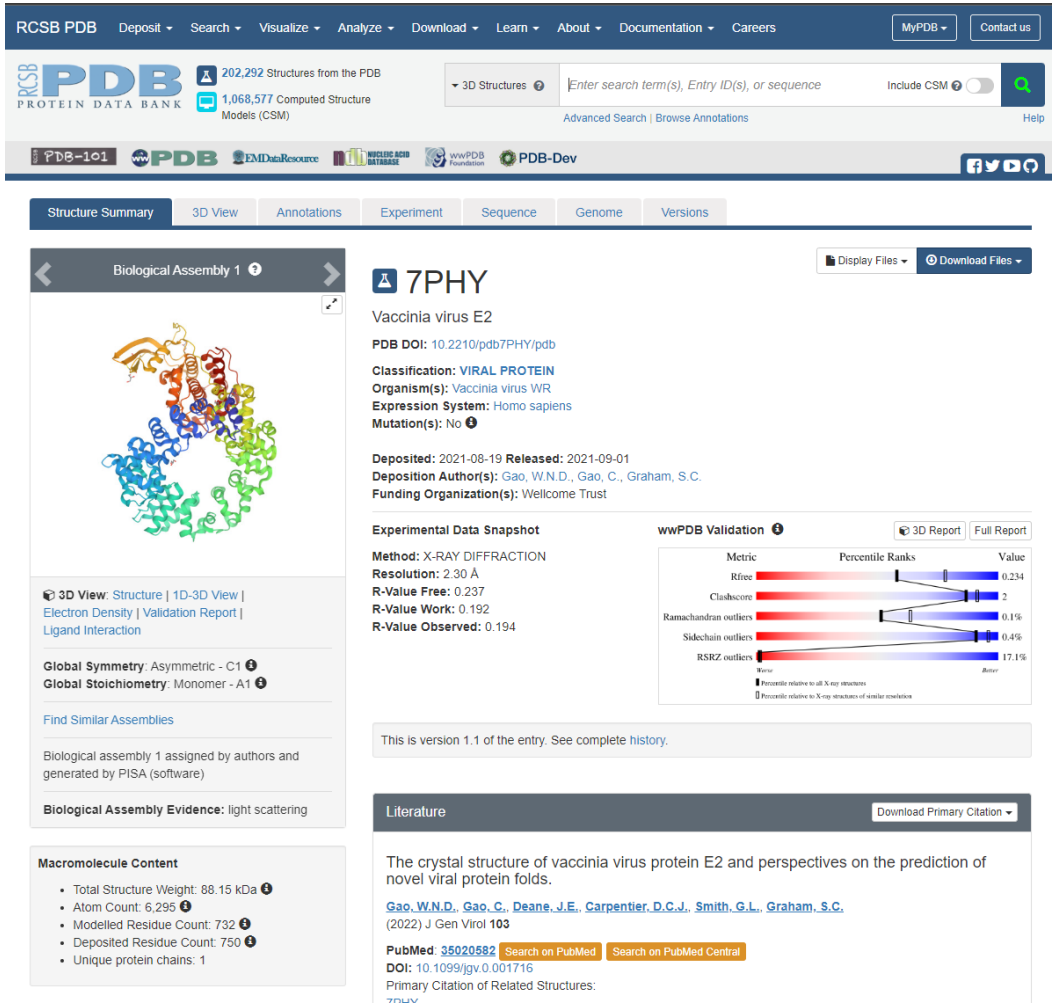
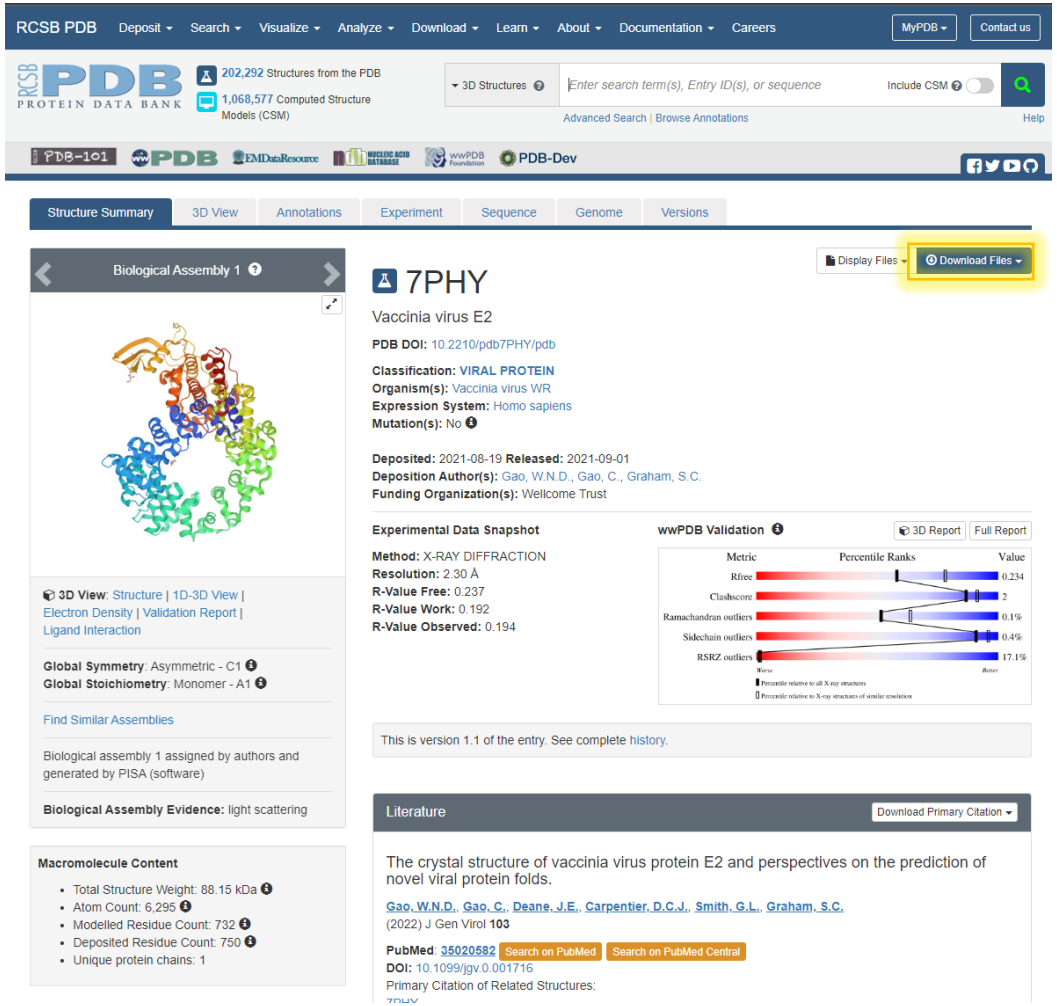


Image: RCSB PDB

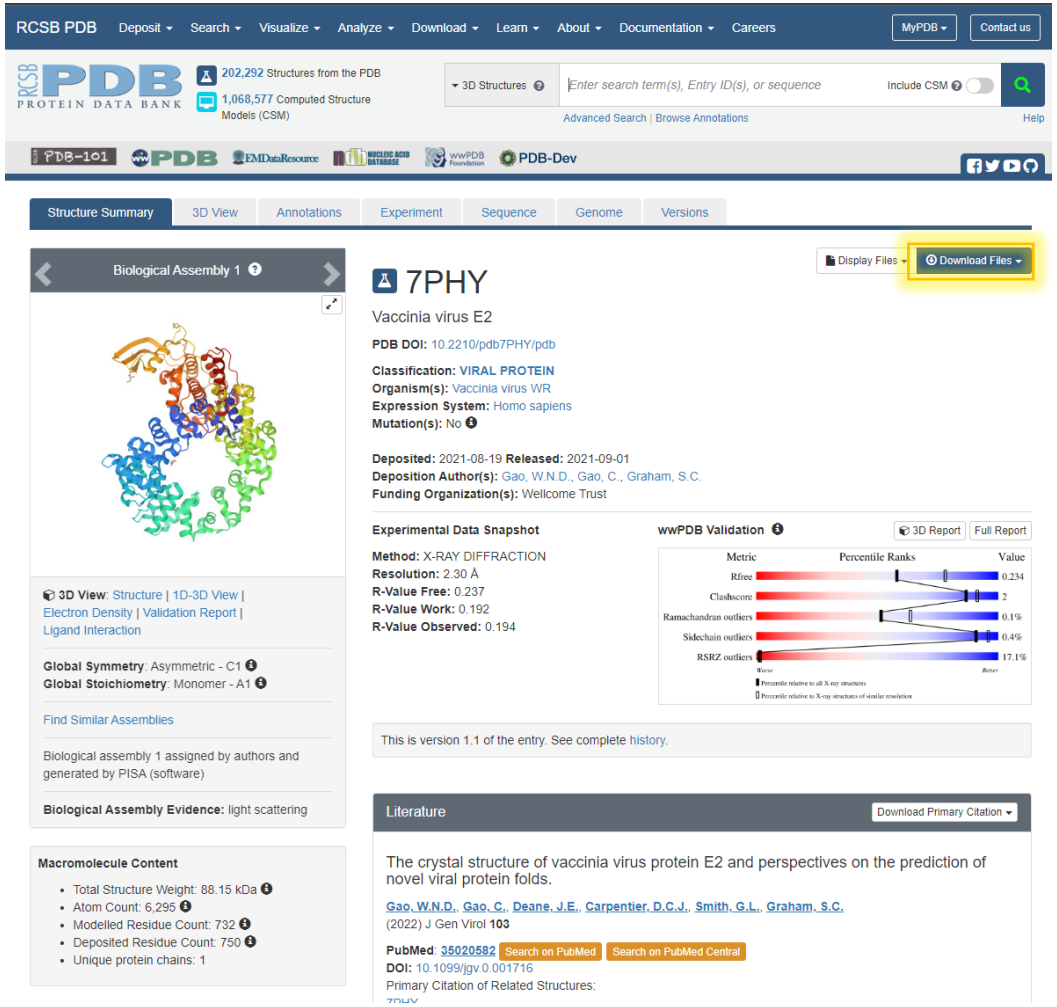
Accessing a protein structure



Accessing a protein structure



Accessing a protein structure



Accessing a protein structure

```
>7PHY_1|Chain A|Protein E2|Vaccinia virus WR (10254)
MISVTDIRRAFLDNECHTITKAFGYLHEDKAIALIKIGFHPTYLPKVLYNNVVEFVPEKLYLFKPRTVAPLDLISTITK
IINDIRRGKIDYYIPYVEDFLEDRTEDLGIYANIFFEDAIDITKLDITKTELEHISKYMNYYTTYIDHIVNIILQNNYI
EENRNIVYKKNNRVLCFDSLENSAFKSLIKIDSIPGLKTYNMKDITYEKSNNIICVRFIPQESIHNEERRIKLQLFDIA
VCVETILDNNQSFKSSKAAAHHHHHHHHH
```

- FASTA sequence

Accessing a protein structure

```
data_7PHY
#
_entry.id   7PHY
#
_audit_conform.dict_name       mmcif_pdbx.dic
_audit_conform.dict_version    5.354
_audit_conform.dict_location   http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
loop_
_database_2.database_id
_database_2.database_code
_database_2.pdbx_database_accession
_database_2.pdbx_DOI
PDB   7PHY         pdb_00007phy 10.2210/pdb7phy/pdb
WWPDB D_1292117740 ?            ?
#
_pdbx_database_status.status_code                     REL
_pdbx_database_status.status_code_sf                  REL
_pdbx_database_status.status_code_mr                  ?
_pdbx_database_status.entry_id                        7PHY
_pdbx_database_status.recvd_initial_deposition_date   2021-08-19
_pdbx_database_status.SG_entry                        N
_pdbx_database_status.deposit_site                    PDBE
_pdbx_database_status.process_site                    PDBE
_pdbx_database_status.status_code_cs                  ?
_pdbx_database_status.status_code_nmr_data            ?
_pdbx_database_status.methods_development_category    ?
_pdbx_database_status.pdb_format_compatible           Y
#
loop_
_audit_author.name
_audit_author.pdbx_ordinal
_audit_author.identifier_ORCID
'Gao, W.N.D.'  1 ?
'Gao, C.'      2 ?
'Graham, S.C.' 3 0000-0003-4547-4034
#
_citation.abstract                  ?
_citation.abstract_id_CAS           ?
_citation.book_id_ISBN              ?
_citation.book_publisher            ?
```

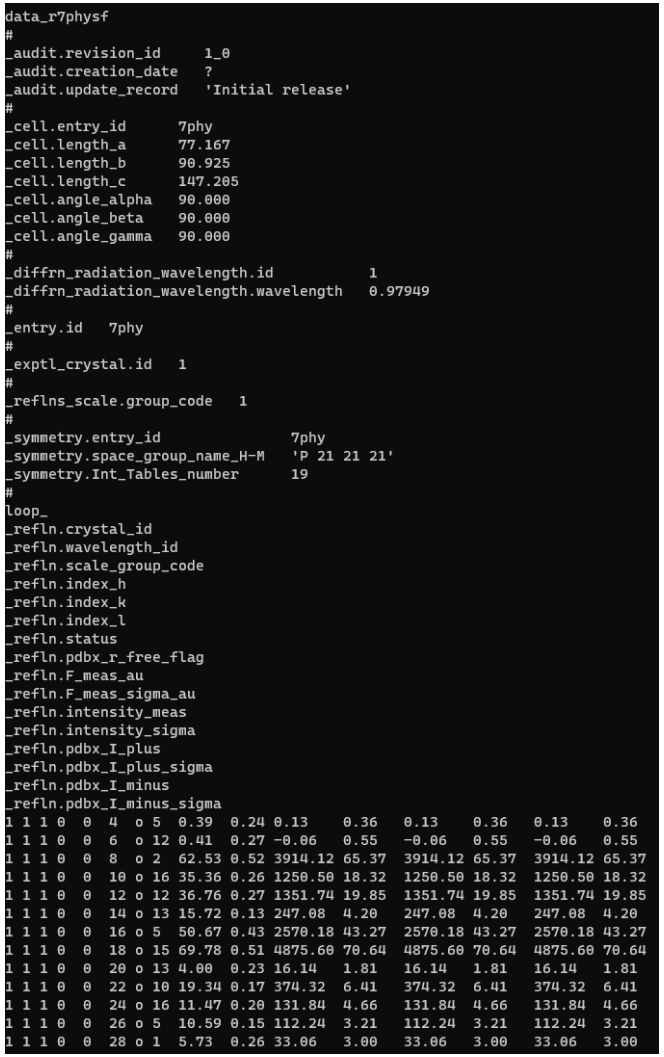
- FASTA sequence
- mmCIF format

Accessing a protein structure

```
HEADER    VIRAL PROTEIN                           19-AUG-21   7PHY
TITLE     VACCINIA VIRUS E2
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: PROTEIN E2;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: VACCINIA VIRUS WR;
SOURCE   3 ORGANISM_TAXID: 10254;
SOURCE   4 GENE: VACWR058, E2L;
SOURCE   5 EXPRESSION_SYSTEM: HOMO SAPIENS;
SOURCE   6 EXPRESSION_SYSTEM_COMMON: HUMAN;
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 9606;
SOURCE   8 EXPRESSION_SYSTEM_CELL_LINE: FREESTYLE 293-F;
SOURCE   9 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE  10 EXPRESSION_SYSTEM_PLASMID: PCDNA3.1
KEYWDS    ASSEMBLY, EGRESS, VIRAL PROTEIN
EXPDTA    X-RAY DIFFRACTION
AUTHOR    W.N.D.GAO,C.GAO,S.C.GRAHAM
REVDAT   2   09-FEB-22 7PHY    1       JRNL
REVDAT   1   01-SEP-21 7PHY    0
JRNL        AUTH   W.N.D.GAO,C.GAO,J.E.DEANE,D.C.J.CARPENTIER,G.L.SMITH,
JRNL        AUTH 2 S.C.GRAHAM
JRNL        TITL   THE CRYSTAL STRUCTURE OF VACCINIA VIRUS PROTEIN E2 AND
JRNL        TITL 2 PERSPECTIVES ON THE PREDICTION OF NOVEL VIRAL PROTEIN FOLDS.
JRNL        REF    J.GEN.VIROL.                  V. 103       2022
JRNL        REFN                   ESSN 1465-2099
JRNL        PMID   35020582
JRNL        DOI    10.1099/JGV.0.001716
REMARK   2
REMARK   2 RESOLUTION.    2.30 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : PHENIX 1.19.2_4158
REMARK   3   AUTHORS     : PAUL ADAMS,PAVEL AFONINE,VINCENT CHEN,IAN
REMARK   3               : DAVIS,KRESHNA GOPAL,RALF GROSSE-KUNSTLEVE,
REMARK   3               : LI-WEI HUNG,ROBERT IMMORMINO,TOM IOERGER,
```

- FASTA sequence
- mmCIF format
- PDB format

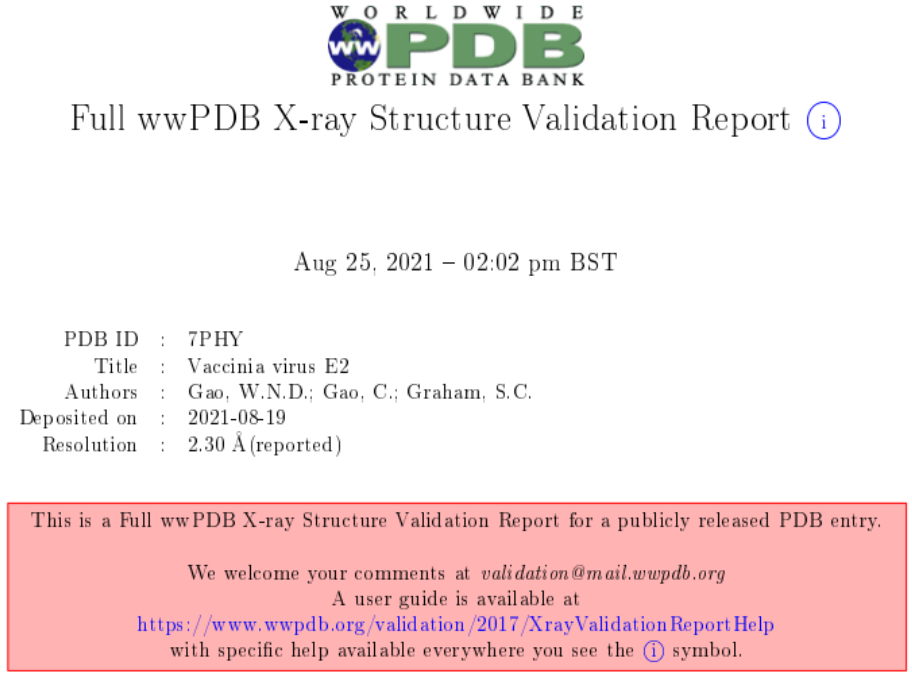
Accessing a protein structure



```
data_r7physf
#
_audit.revision_id     1_0
_audit.creation_date   ?
_audit.update_record   'Initial release'
#
_cell.entry_id      7phy
_cell.length_a      77.167
_cell.length_b      90.925
_cell.length_c      147.205
_cell.angle_alpha   90.000
_cell.angle_beta    90.000
_cell.angle_gamma   90.000
#
_diffrn_radiation_wavelength.id           1
_diffrn_radiation_wavelength.wavelength   0.97949
#
_entry.id   7phy
#
_exptl_crystal.id   1
#
_reflns_scale.group_code   1
#
_symmetry.entry_id              7phy
_symmetry.space_group_name_H-M   'P 21 21 21'
_symmetry.Int_Tables_number      19
#
loop_
_refln.crystal_id
_refln.wavelength_id
_refln.scale_group_code
_refln.index_h
_refln.index_k
_refln.index_l
_refln.status
_refln.pdbx_r_free_flag
_refln.F_meas_au
_refln.F_meas_sigma_au
_refln.intensity_meas
_refln.intensity_sigma
_refln.pdbx_I_plus
_refln.pdbx_I_plus_sigma
_refln.pdbx_I_minus
_refln.pdbx_I_minus_sigma
1 1 1 0  0  4  o 5  0.39  0.24 0.13    0.36   0.13    0.36   0.13    0.36
1 1 1 0  0  6  o 12 0.41  0.27 -0.06   0.55   -0.06   0.55   -0.06   0.55
1 1 1 0  0  8  o 2  62.53 0.52 3914.12 65.37  3914.12 65.37  3914.12 65.37
1 1 1 0  0  10 o 16 35.36 0.26 1250.50 18.32  1250.50 18.32  1250.50 18.32
1 1 1 0  0  12 o 12 36.76 0.27 1351.74 19.85  1351.74 19.85  1351.74 19.85
1 1 1 0  0  14 o 13 15.72 0.13 247.08  4.20   247.08  4.20   247.08  4.20
1 1 1 0  0  16 o 5  50.67 0.43 2570.18 43.27  2570.18 43.27  2570.18 43.27
1 1 1 0  0  18 o 15 69.78 0.51 4875.60 70.64  4875.60 70.64  4875.60 70.64
1 1 1 0  0  20 o 13 4.00  0.23 16.14   1.81   16.14   1.81   16.14   1.81
1 1 1 0  0  22 o 10 19.34 0.17 374.32  6.41   374.32  6.41   374.32  6.41
1 1 1 0  0  24 o 16 11.47 0.20 131.84  4.66   131.84  4.66   131.84  4.66
1 1 1 0  0  26 o 5  10.59 0.15 112.24  3.21   112.24  3.21   112.24  3.21
1 1 1 0  0  28 o 1  5.73  0.26 33.06   3.00   33.06   3.00   33.06   3.00
```

- FASTA sequence
- mmCIF format
- PDB format
- Structure factors (experimental data)

Accessing a protein structure



WORLDWIDE PDB
PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Aug 25, 2021 – 02:02 pm BST

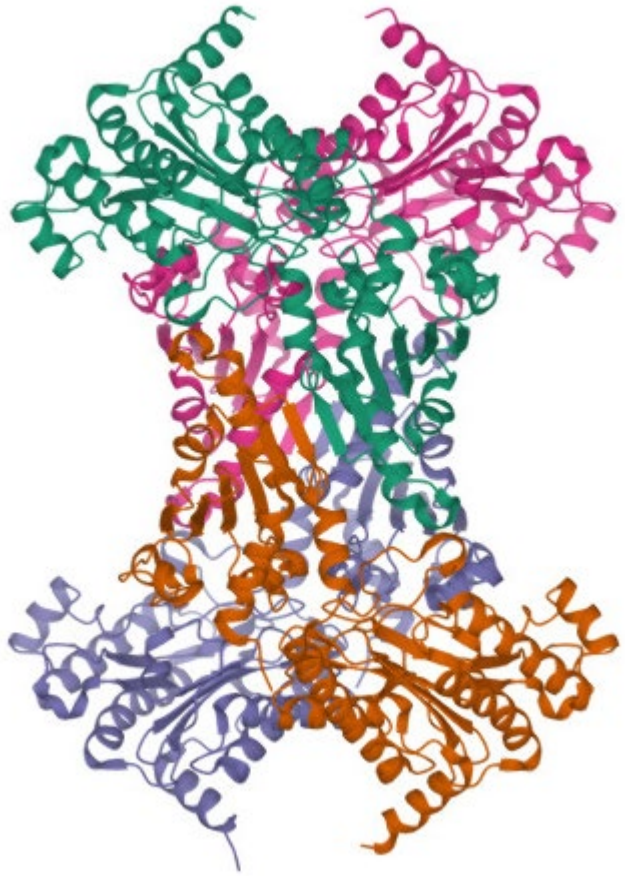
PDB ID	:	7PHY
Title	:	Vaccinia virus E2
Authors	:	Gao, W.N.D.; Gao, C.; Graham, S.C.
Deposited on	:	2021-08-19
Resolution	:	2.30 Å (reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

- FASTA sequence
- mmCIF format
- PDB format
- Structure factors (experimental data)
- Validation report

Accessing a protein structure



Display Files ▾ | Download Files ▾

FASTA Sequence

PDBx/mmCIF Format
PDBx/mmCIF Format (gz)

PDB Format
PDB Format (gz)

PDBML/XML Format (gz)

Structure Factors (CIF)
Structure Factors (CIF - gz)

Validation Full PDF
Validation XML
Validation CIF

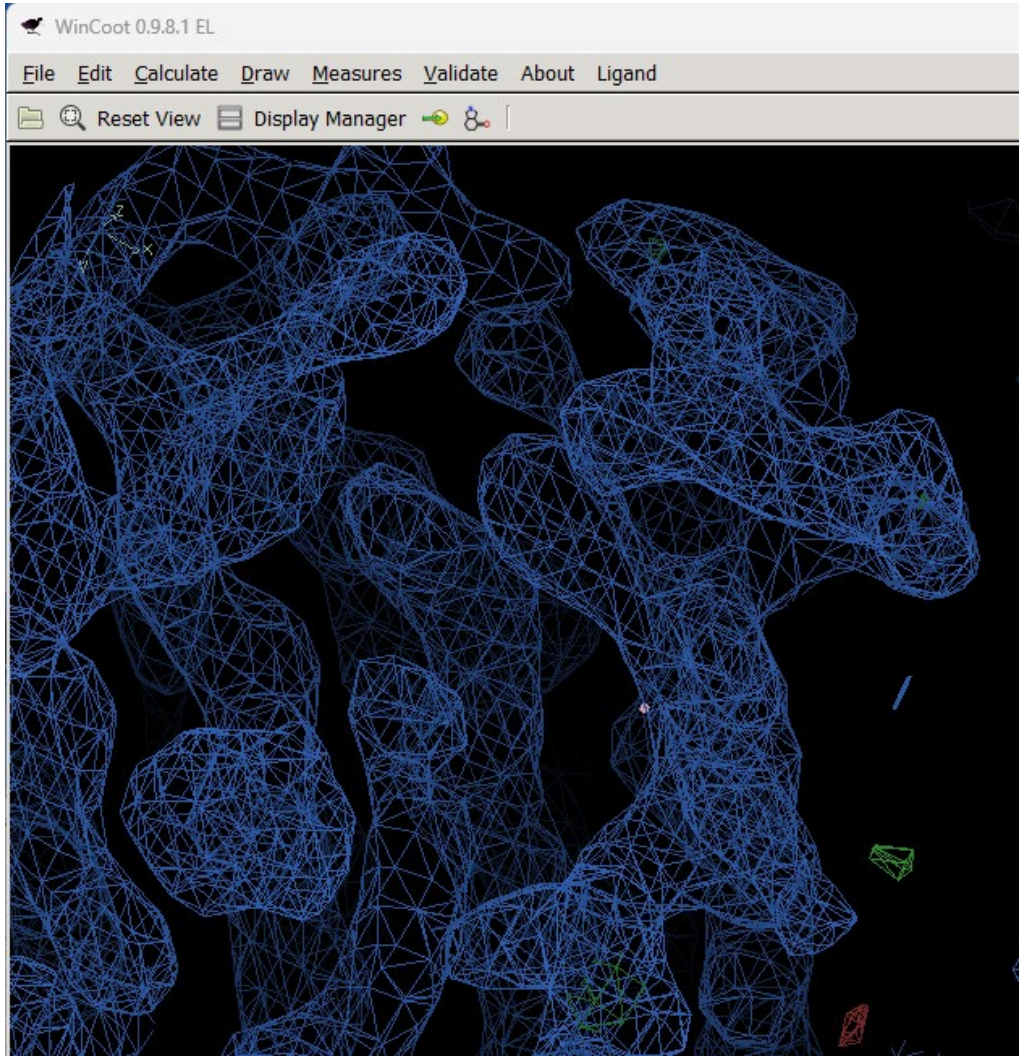
Biological Assembly 1 (CIF - gz)

Biological Assembly 1 (PDB - gz)

fo-fc Map (DSN6)
2fo-fc Map (DSN6)
Map Coefficients (MTZ format)

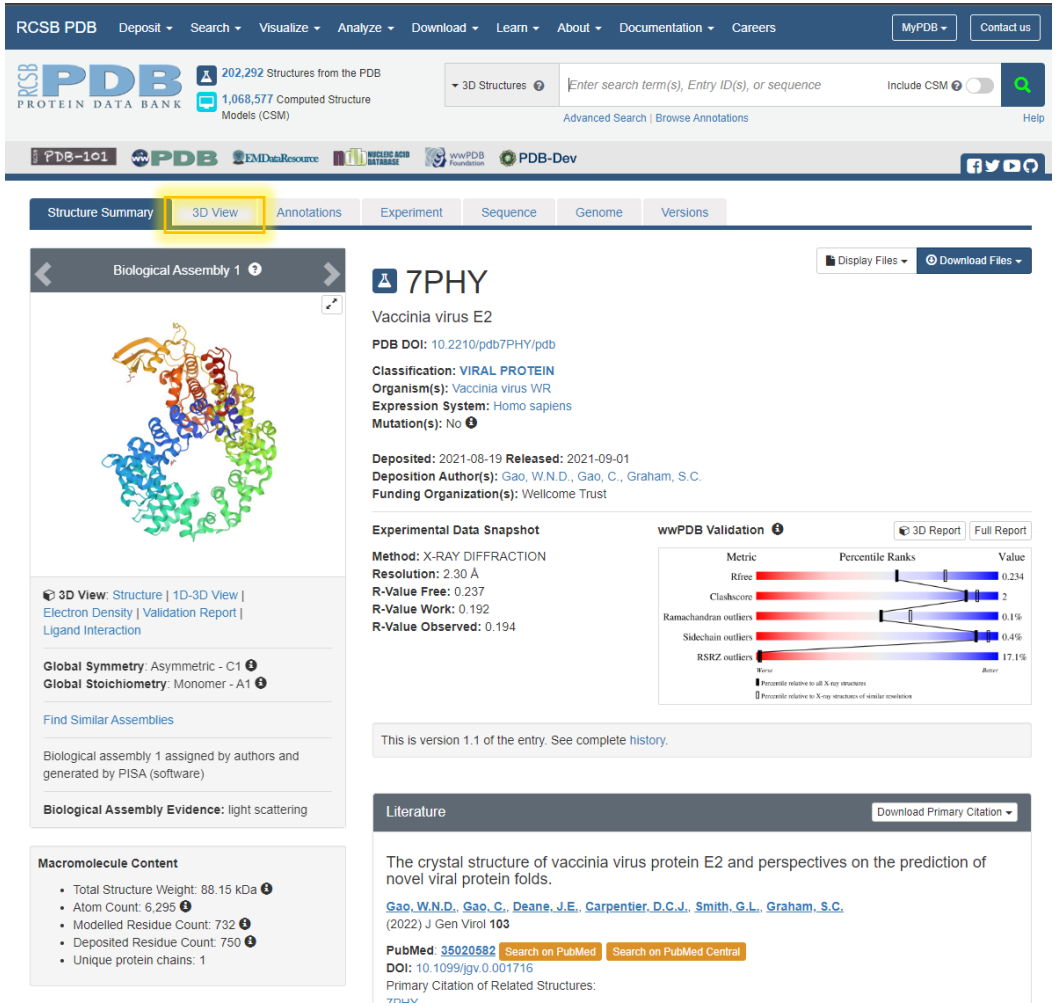
- FASTA sequence
- mmCIF format
- PDB format
- Structure factors (experimental data)
- Validation report
- Assembly

Accessing a protein structure

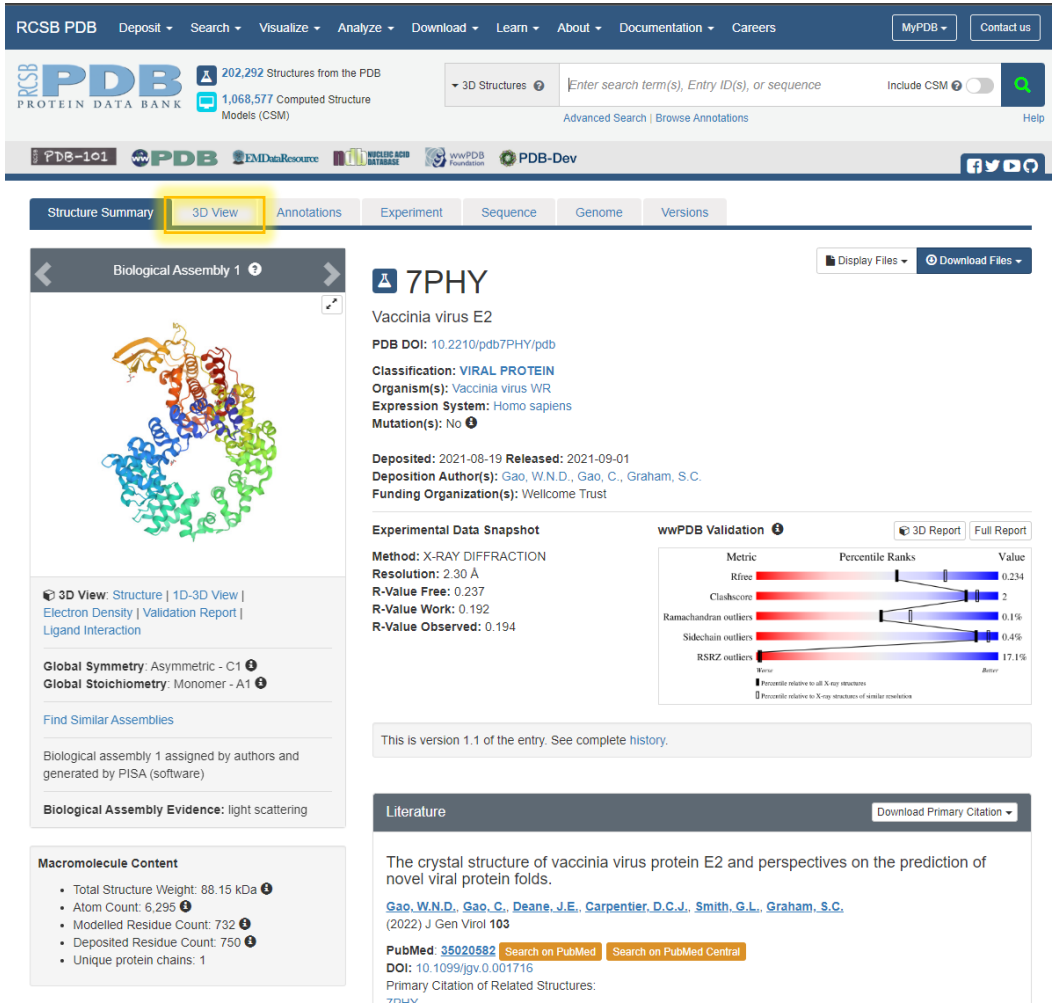


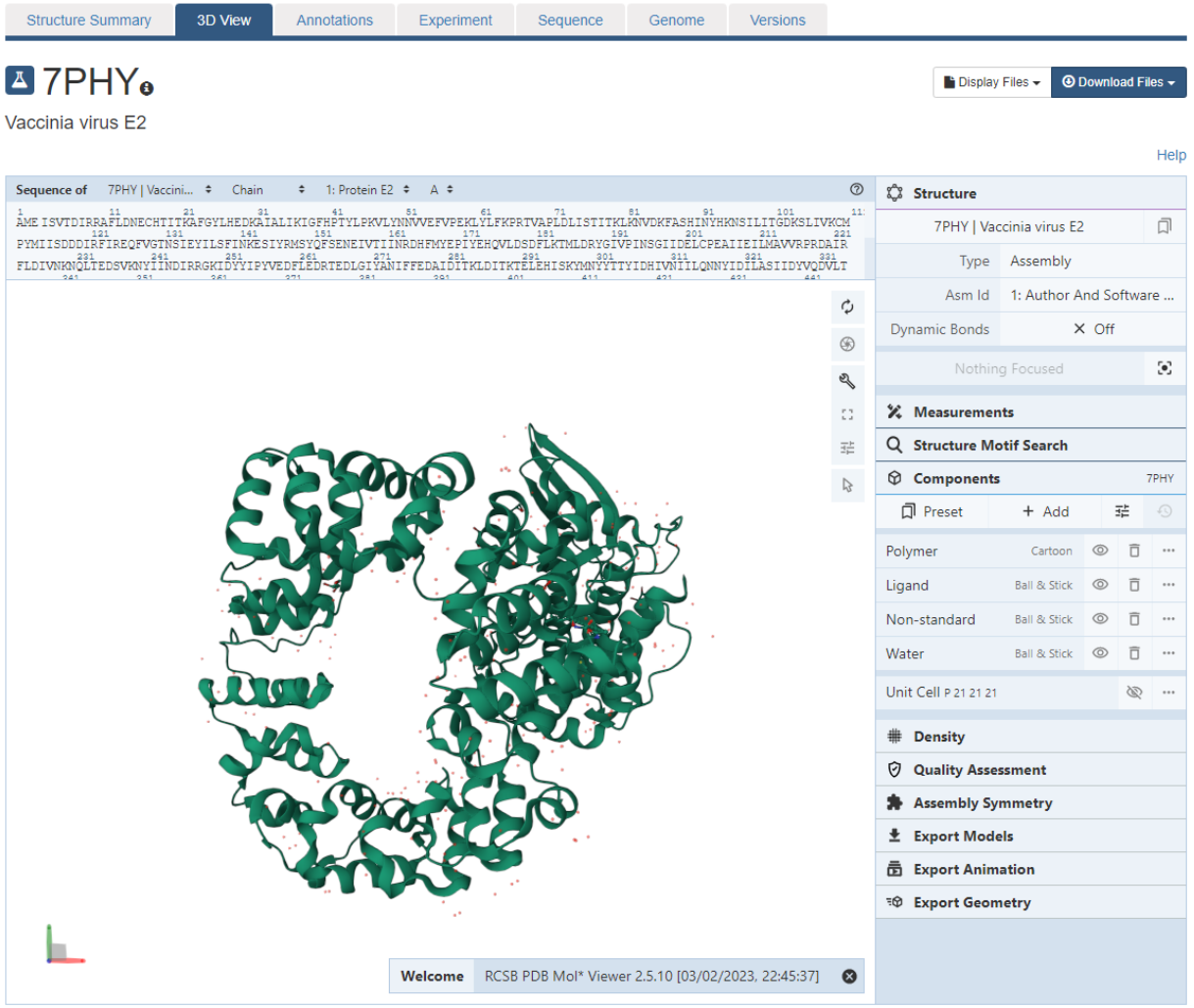
- FASTA sequence
- mmCIF format
- PDB format
- Structure factors (experimental data)
- Validation report
- Assembly
- Maps

Accessing a protein structure

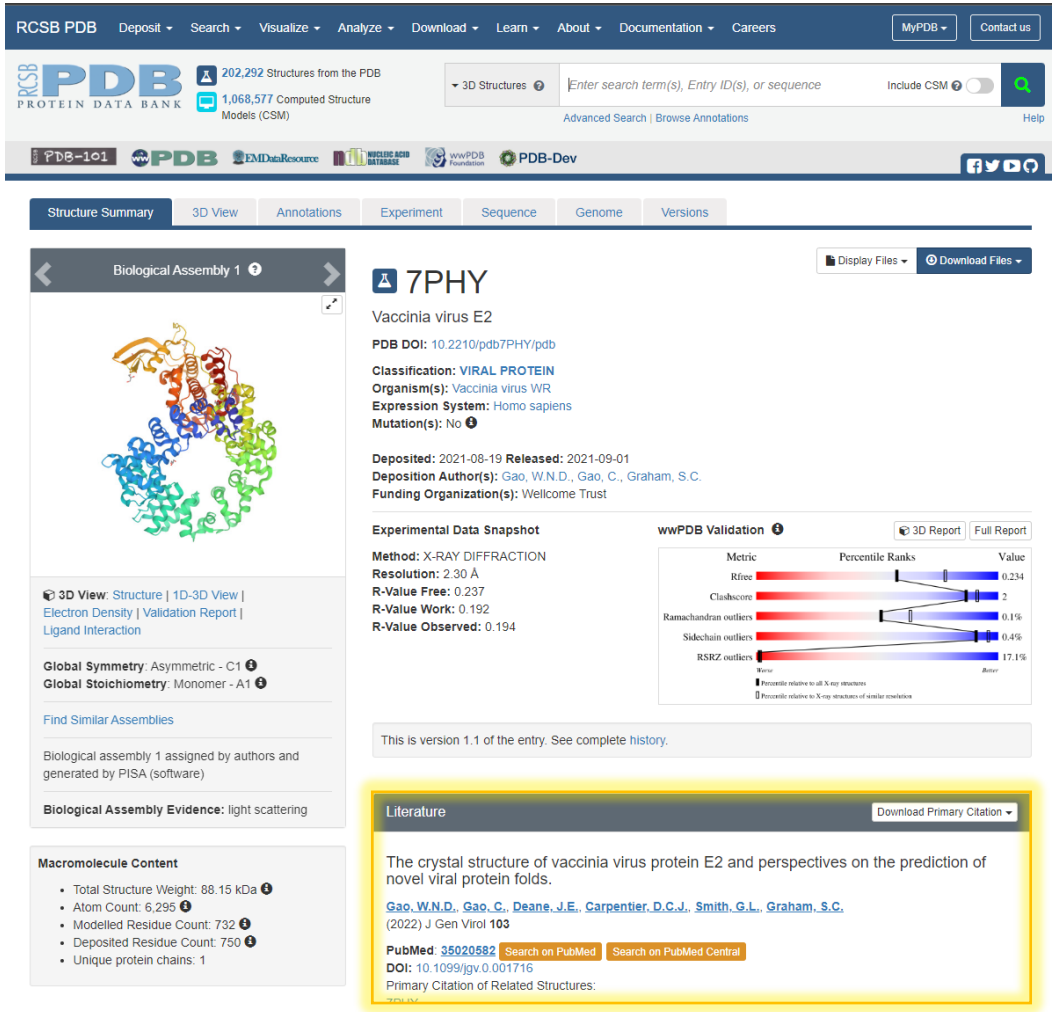


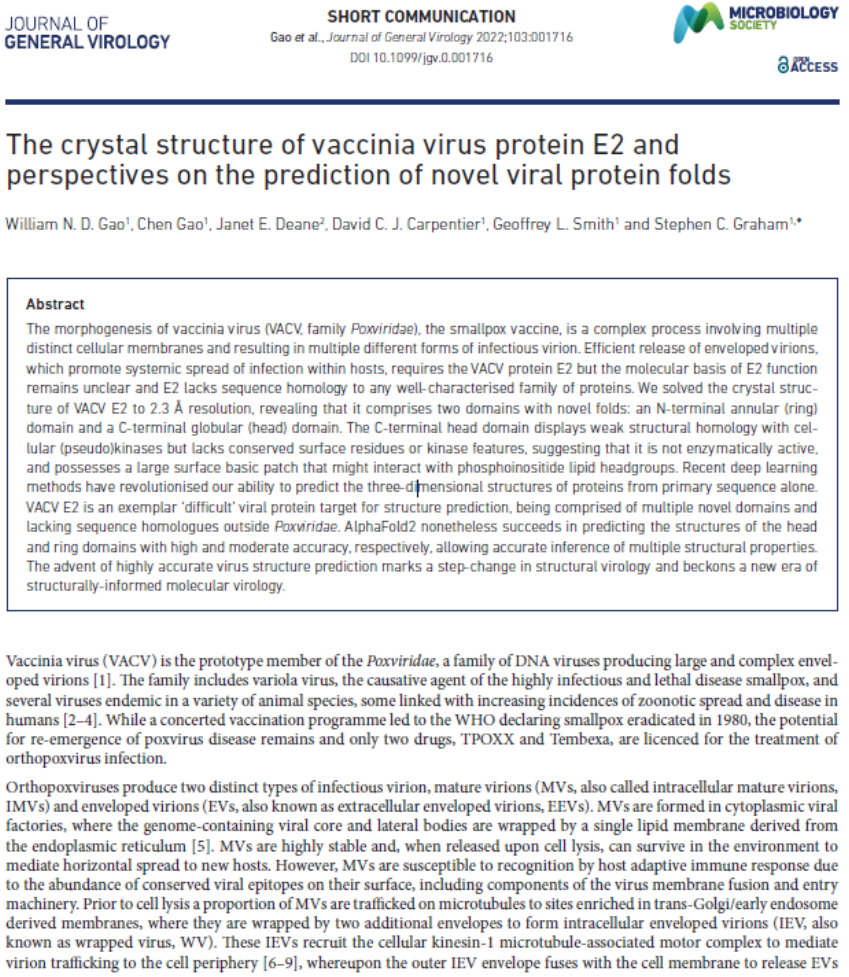
Accessing a protein structure





Accessing a protein structure





JOURNAL OF GENERAL VIROLOGY

SHORT COMMUNICATION

Gao et al., *Journal of General Virology* 2022;103:001716
DOI 10.1099/jgv.0.001716

MICROBIOLOGY SOCIETY

OPEN ACCESS

The crystal structure of vaccinia virus protein E2 and perspectives on the prediction of novel viral protein folds

William N. D. Gao[1], Chen Gao[1], Janet E. Deane[2], David C. J. Carpentier[1], Geoffrey L. Smith[1] and Stephen C. Graham[1,*]

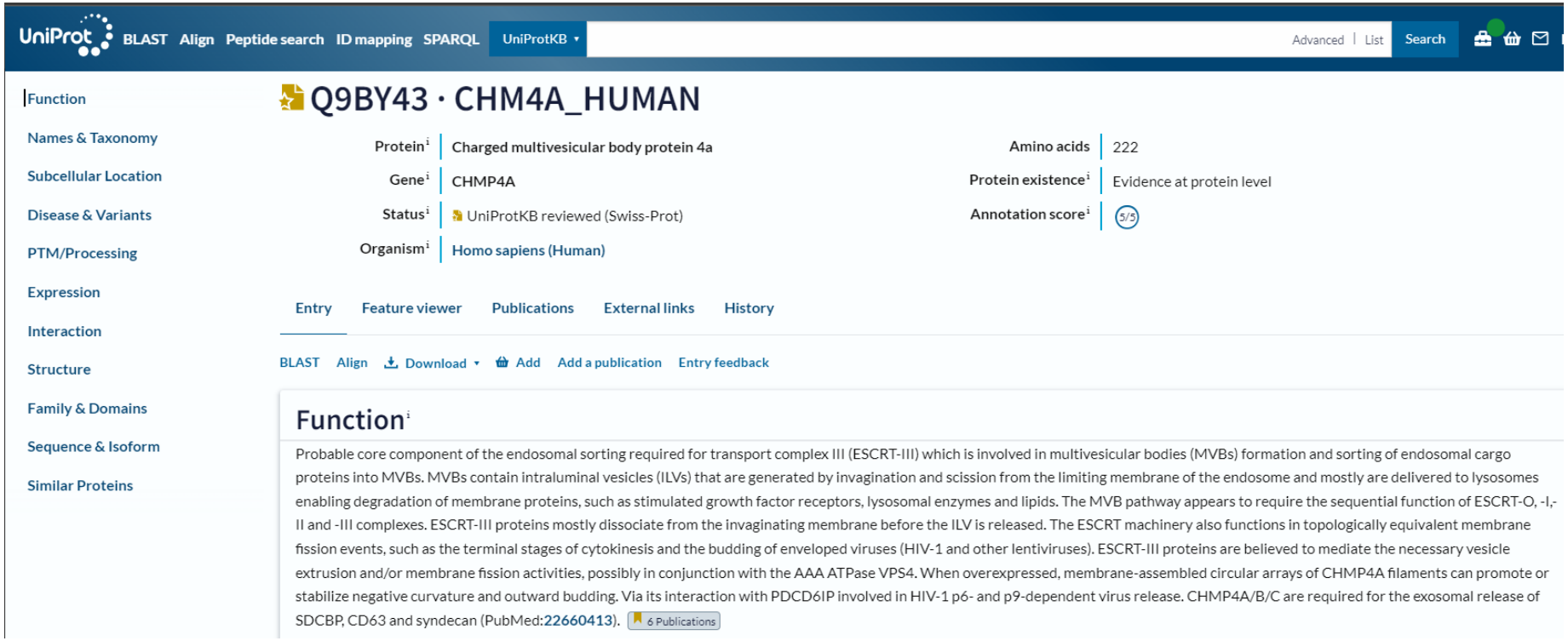
Abstract

The morphogenesis of vaccinia virus (VACV, family *Poxviridae*), the smallpox vaccine, is a complex process involving multiple distinct cellular membranes and resulting in multiple different forms of infectious virion. Efficient release of enveloped virions, which promote systemic spread of infection within hosts, requires the VACV protein E2 but the molecular basis of E2 function remains unclear and E2 lacks sequence homology to any well-characterised family of proteins. We solved the crystal structure of VACV E2 to 2.3 Å resolution, revealing that it comprises two domains with novel folds: an N-terminal annular (ring) domain and a C-terminal globular (head) domain. The C-terminal head domain displays weak structural homology with cellular (pseudo)kinases but lacks conserved surface residues or kinase features, suggesting that it is not enzymatically active, and possesses a large surface basic patch that might interact with phosphoinositide lipid headgroups. Recent deep learning methods have revolutionised our ability to predict the three-dimensional structures of proteins from primary sequence alone. VACV E2 is an exemplar 'difficult' viral protein target for structure prediction, being comprised of multiple novel domains and lacking sequence homologues outside *Poxviridae*. AlphaFold2 nonetheless succeeds in predicting the structures of the head and ring domains with high and moderate accuracy, respectively, allowing accurate inference of multiple structural properties. The advent of highly accurate virus structure prediction marks a step-change in structural virology and beckons a new era of structurally-informed molecular virology.

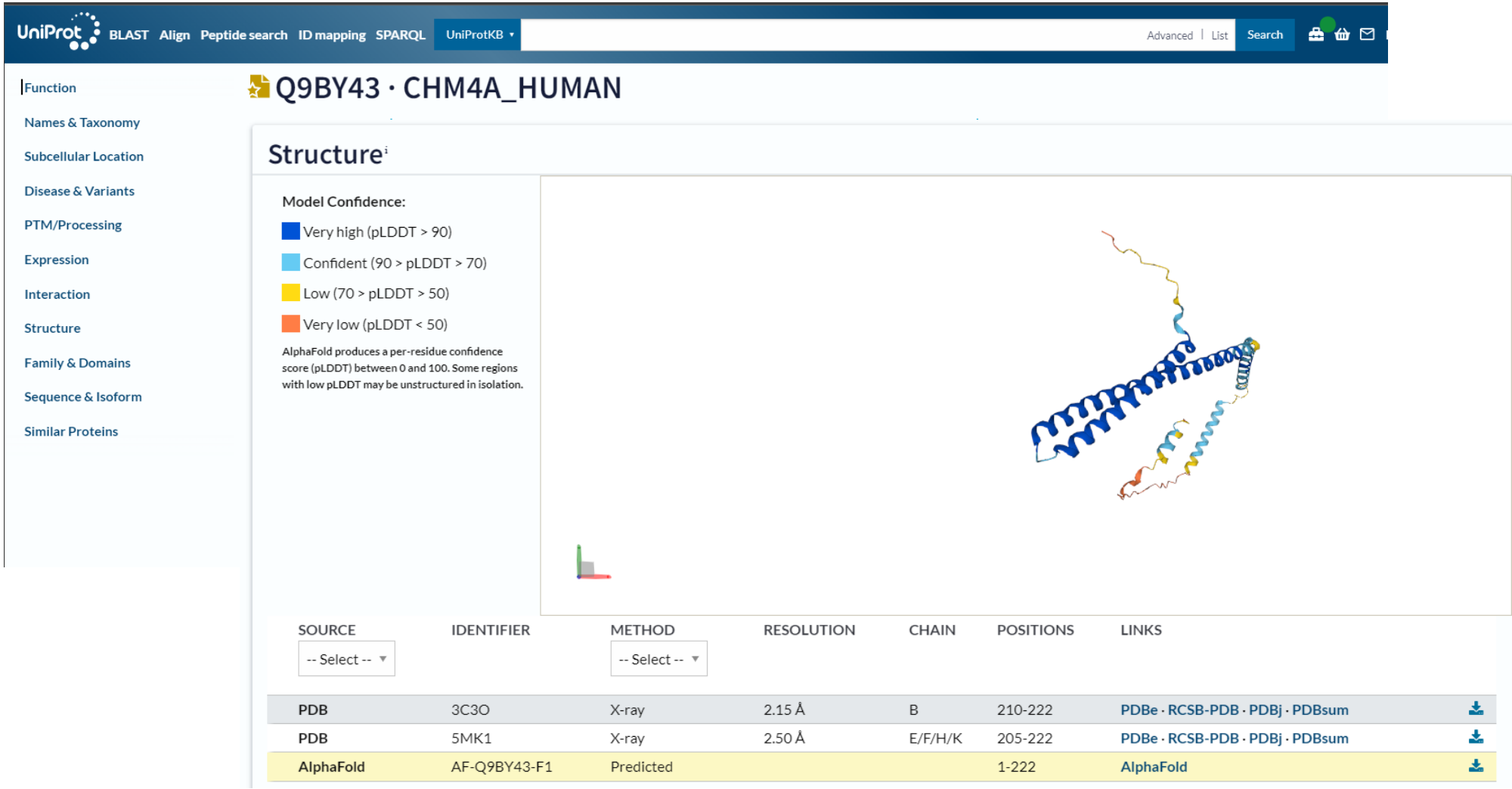
Vaccinia virus (VACV) is the prototype member of the *Poxviridae*, a family of DNA viruses producing large and complex enveloped virions [1]. The family includes variola virus, the causative agent of the highly infectious and lethal disease smallpox, and several viruses endemic in a variety of animal species, some linked with increasing incidences of zoonotic spread and disease in humans [2–4]. While a concerted vaccination programme led to the WHO declaring smallpox eradicated in 1980, the potential for re-emergence of poxvirus disease remains and only two drugs, TPOXX and Tembexa, are licenced for the treatment of orthopoxvirus infection.

Orthopoxviruses produce two distinct types of infectious virion, mature virions (MVs, also called intracellular mature virions, IMVs) and enveloped virions (EVs, also known as extracellular enveloped virions, EEVs). MVs are formed in cytoplasmic viral factories, where the genome-containing viral core and lateral bodies are wrapped by a single lipid membrane derived from the endoplasmic reticulum [5]. MVs are highly stable and, when released upon cell lysis, can survive in the environment to mediate horizontal spread to new hosts. However, MVs are susceptible to recognition by host adaptive immune response due to the abundance of conserved viral epitopes on their surface, including components of the virus membrane fusion and entry machinery. Prior to cell lysis a proportion of MVs are trafficked on microtubules to sites enriched in trans-Golgi/early endosome derived membranes, where they are wrapped by two additional envelopes to form intracellular enveloped virions (IEV, also known as wrapped virus, WV). These IEVs recruit the cellular kinesin-1 microtubule-associated motor complex to mediate virion trafficking to the cell periphery [6–9], whereupon the outer IEV envelope fuses with the cell membrane to release EVs

Accessing protein structural models



Accessing protein structural models



Software for viewing protein structures

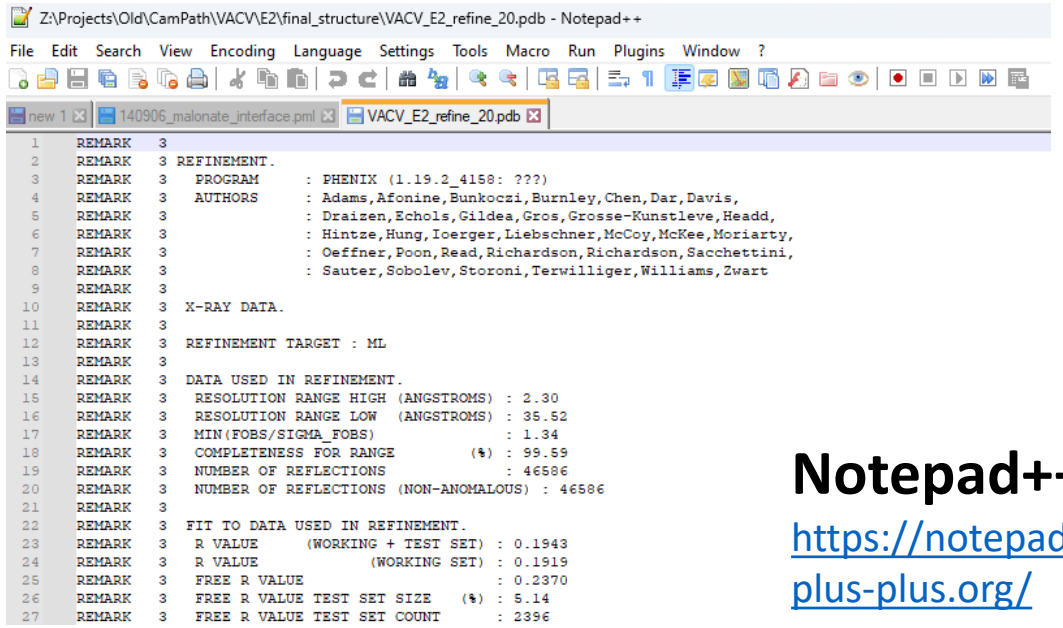
- Many, many different programs for viewing proteins structures for analysis and illustration:

Software for viewing protein structures

- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors

```
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3   PROGRAM     : PHENIX (1.19.2_4158: ???)
REMARK   3   AUTHORS     : Adams,Afonine,Bunkoczi,Burnley,Chen,Dar,Davis,
REMARK   3               : Draizen,Echols,Gildea,Gros,Grosse-Kunstleve,Headd,
REMARK   3               : Hintze,Hung,Ioerger,Liebschner,McCoy,McKee,Moriarty,
REMARK   3               : Oeffner,Poon,Read,Richardson,Richardson,Sacchettini,
REMARK   3               : Sauter,Sobolev,Storoni,Terwilliger,Williams,Zwart
REMARK   3
REMARK   3  X-RAY DATA.
REMARK   3
REMARK   3  REFINEMENT TARGET : ML
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.30
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 35.52
:
```

Vim

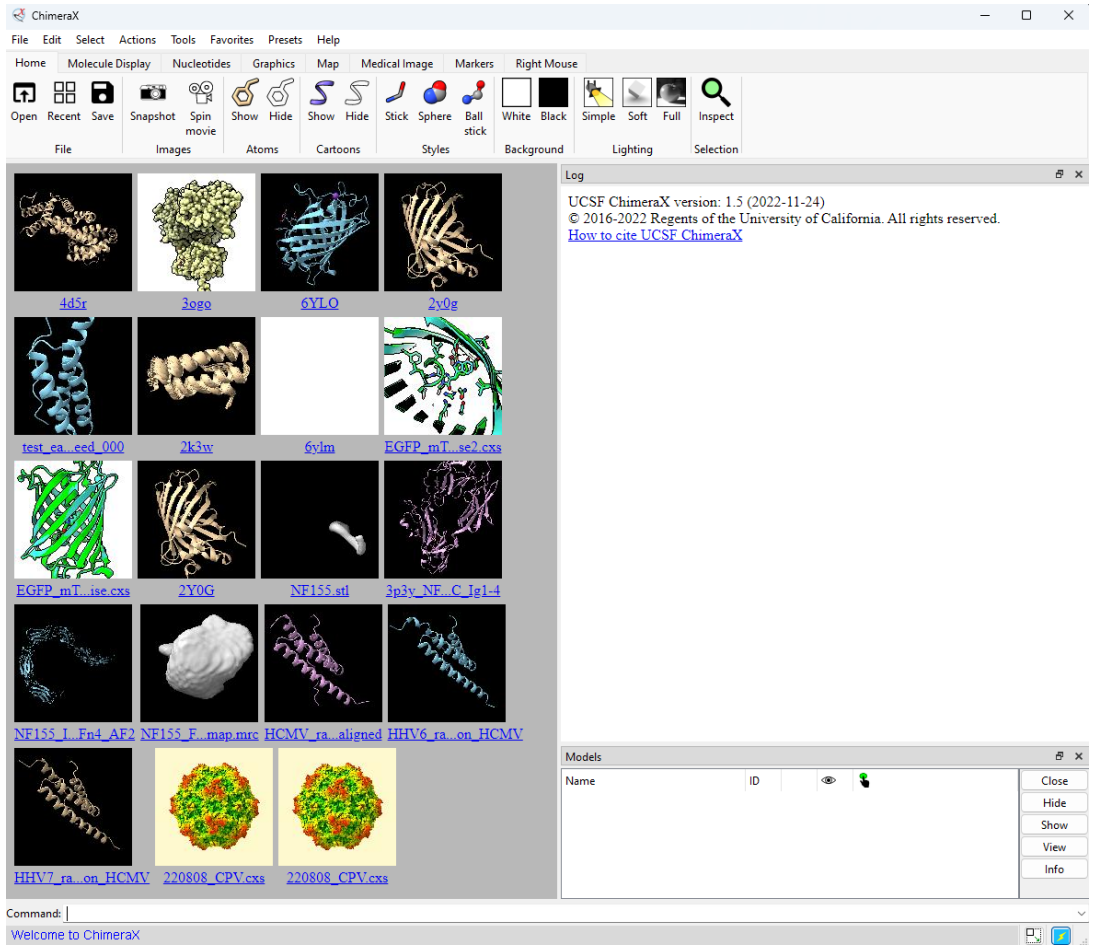


Notepad++

https://notepad-plus-plus.org/

Software for viewing protein structures

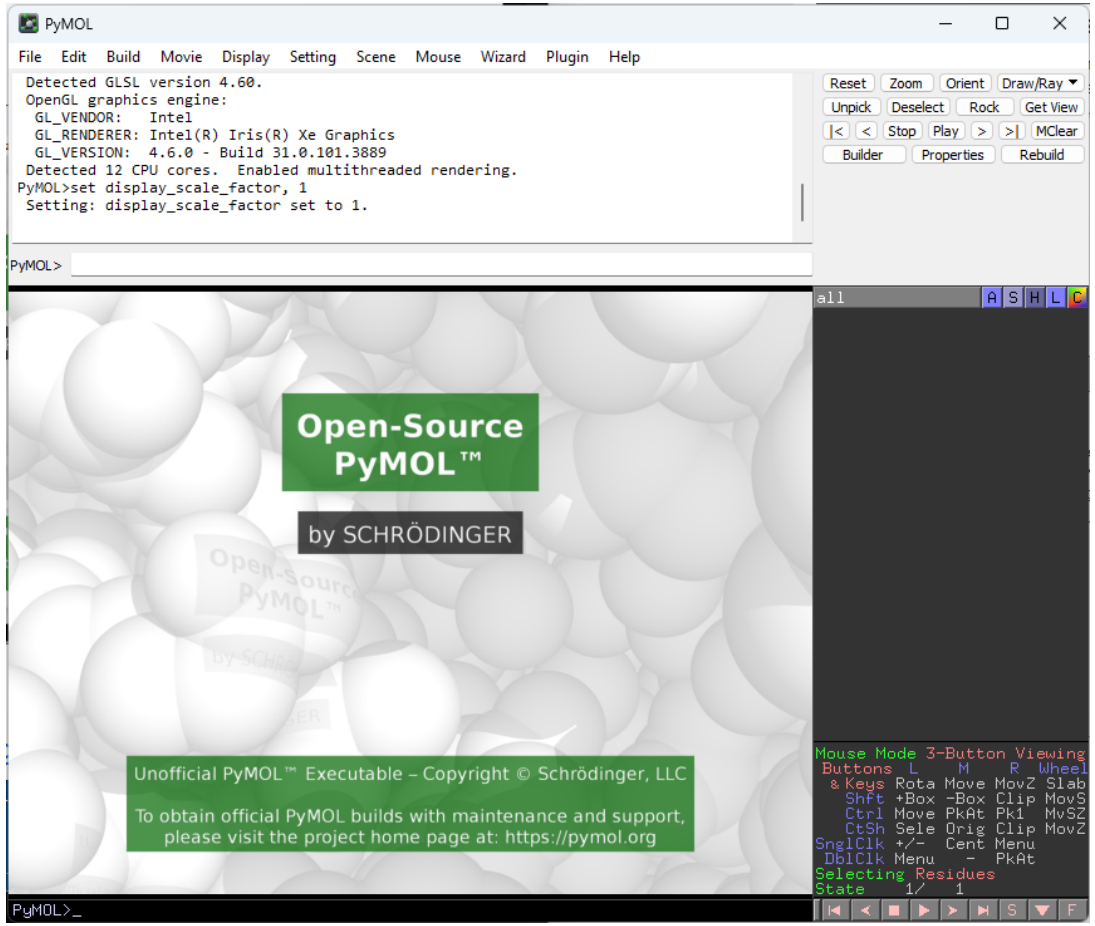
- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX



https://www.cgl.ucsf.edu/chimerax/

Software for viewing protein structures

- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL



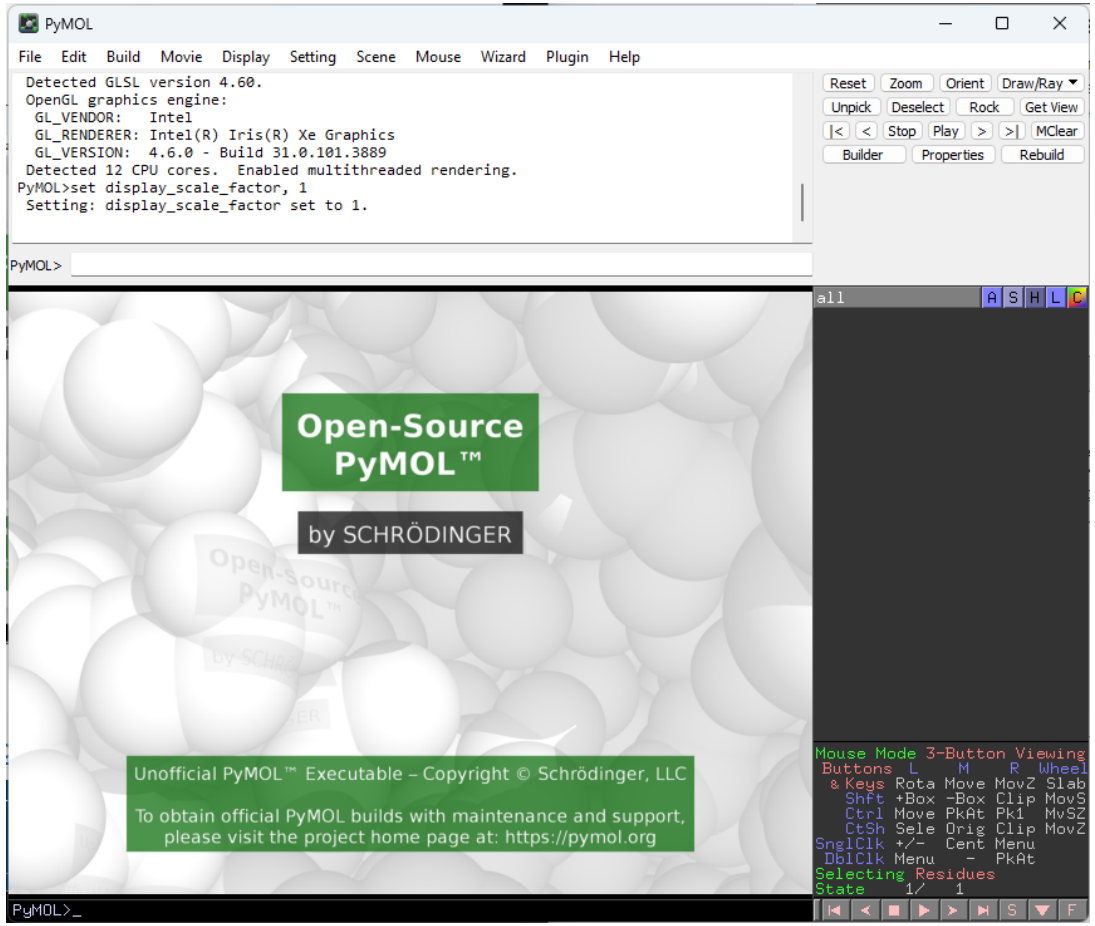
https://github.com/schrodinger/pymol-open-source

Software for viewing protein structures

- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL
- There are open source and paid-for versions of PyMOL. It's easy to install the free open source version:

  https://pymolwiki.org/index.php/Windows_Install

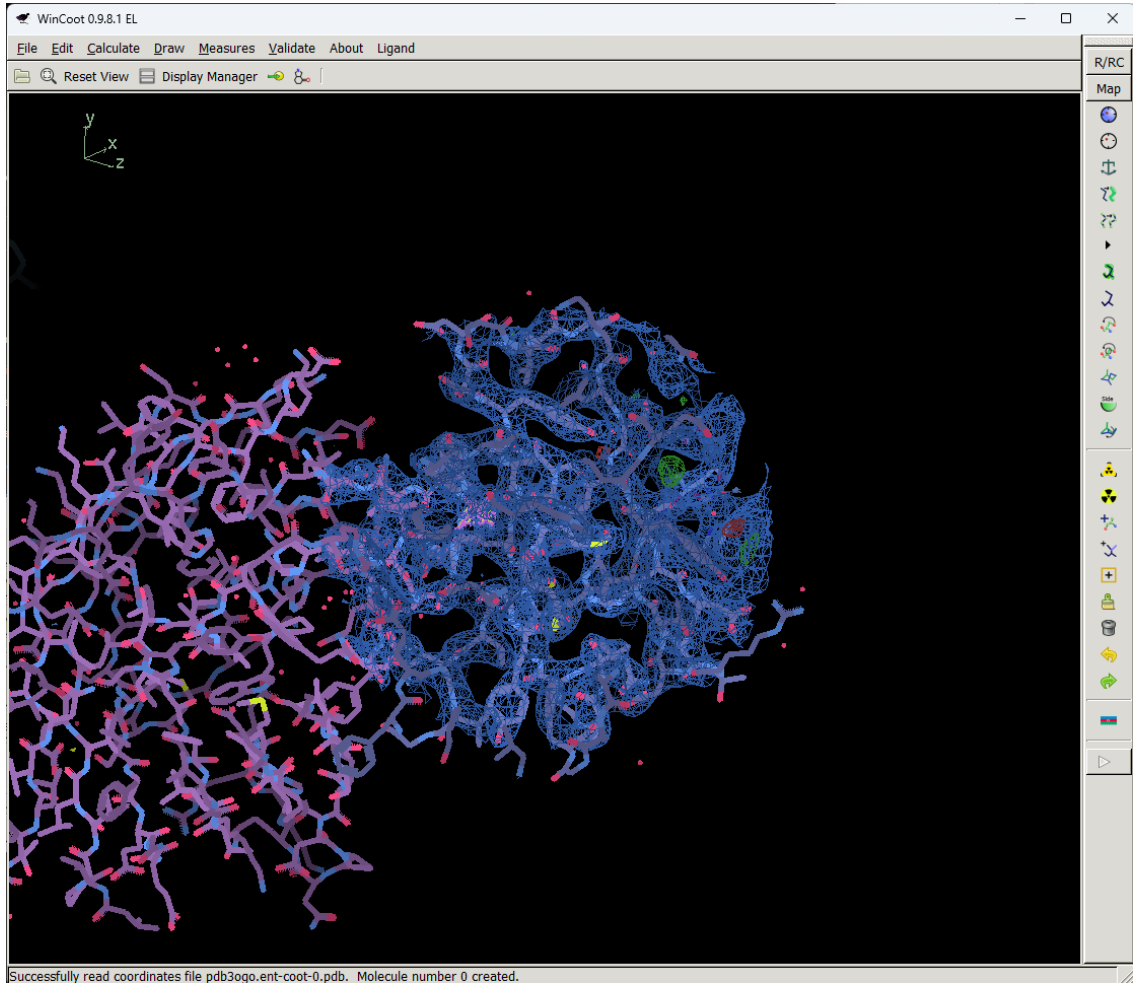
  https://pymolwiki.org/index.php/MAC_Install



https://github.com/schrodinger/pymol-open-source

Software for viewing protein structures

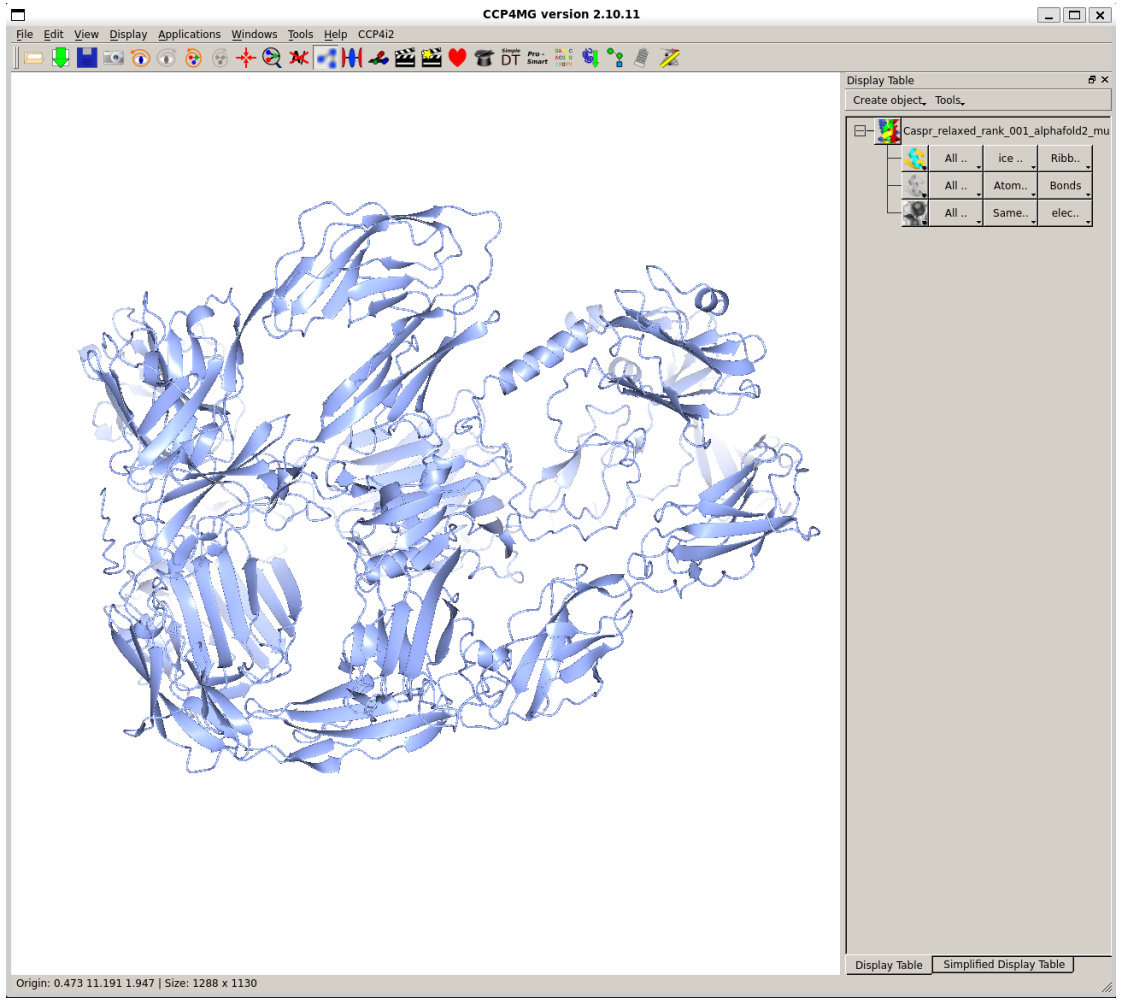
- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL
  - COOT



https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/

Software for viewing protein structures

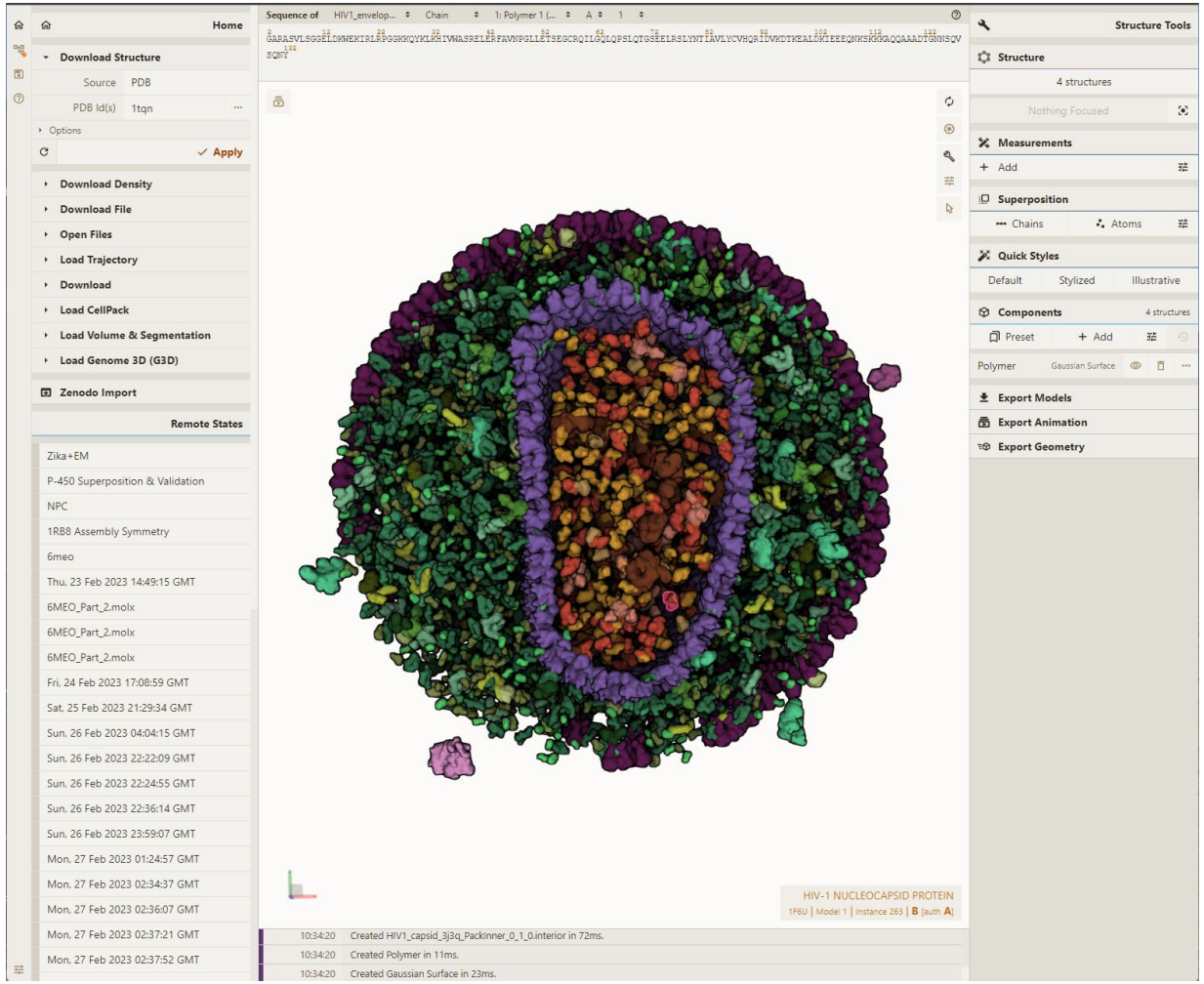
- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL
  - COOT
  - CCP4mg



https://www.ccp4.ac.uk/MG/

Software for viewing protein structures

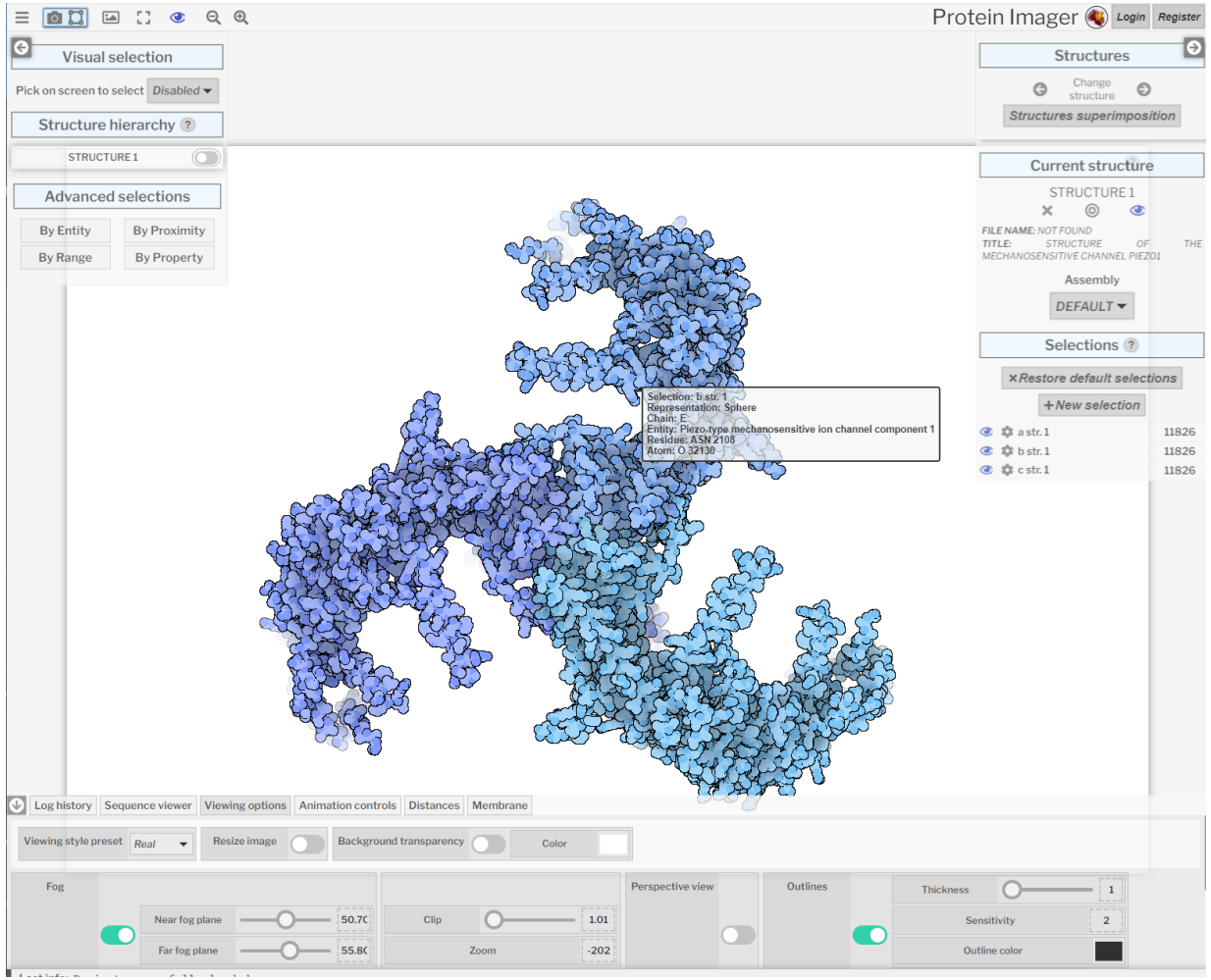
- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL
  - COOT
  - CCP4mg
  - Mol*



https://molstar.org/

Software for viewing protein structures

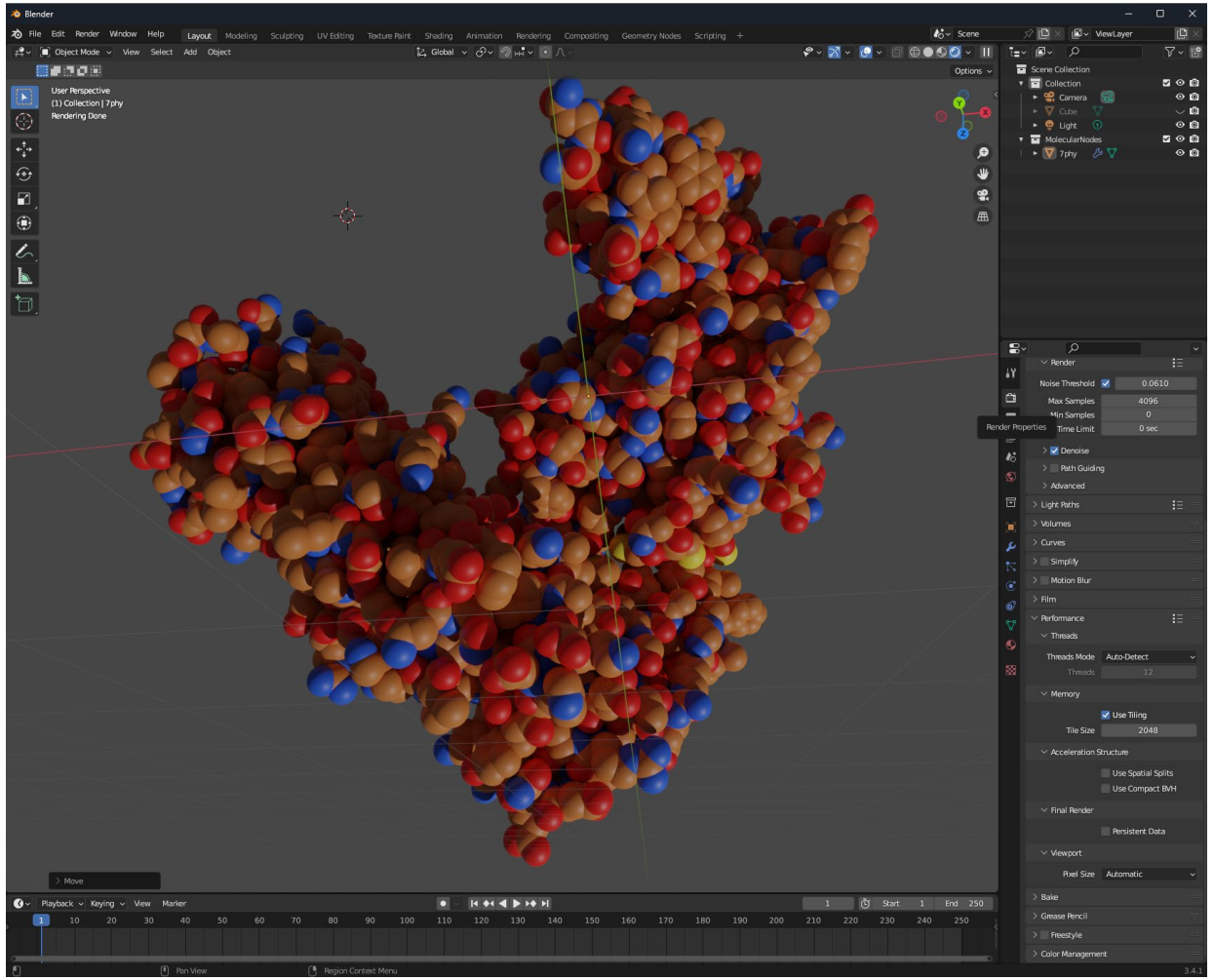
- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL
  - COOT
  - CCP4mg
  - Mol*
  - Protein Imager



https://3dproteinimaging.com/protein-imager/

Software for viewing protein structures

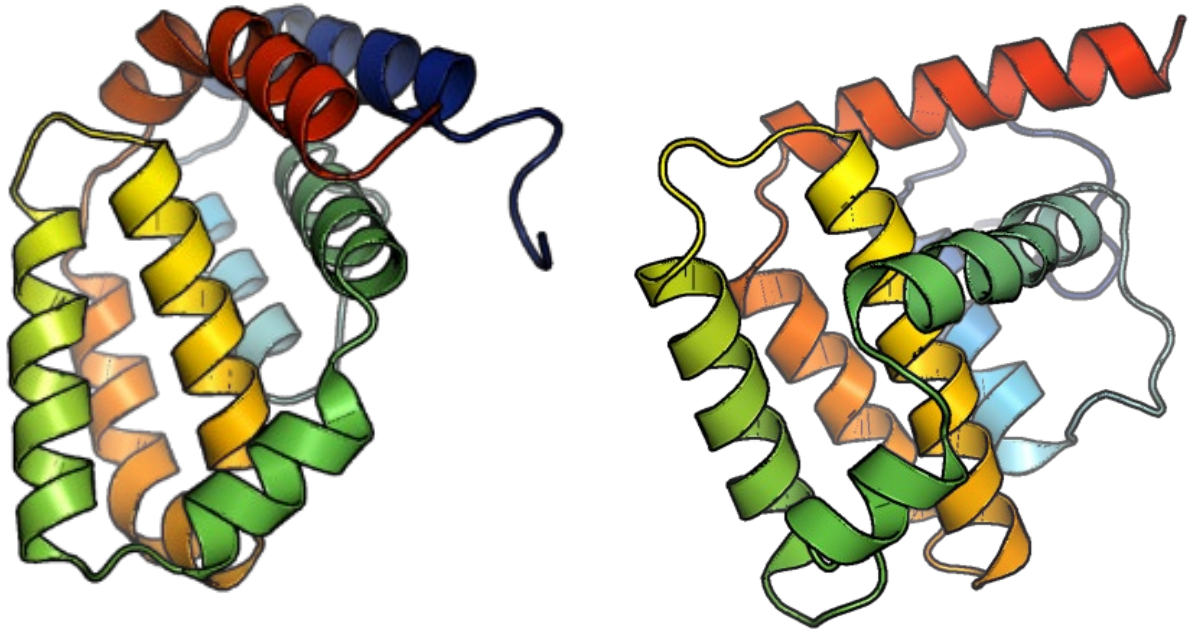
- Many, many different programs for viewing proteins structures for analysis and illustration:
  - Text editors (notepad!)
  - ChimeraX
  - PyMOL
  - COOT
  - CCP4mg
  - Mol*
  - Protein Imager
  - Blender (with Molecular Nodes)



https://bradyajohnston.github.io/MolecularNodes/

Measures of protein structural similarity

- Sequence identity is a measure of amino acid sequence similarity
  - High identity suggests evolutionary similarity
- Proteins can have similar structures in absence of similar sequences
- How do we measure protein structure similarity?



Vaccinia virus proteins A49 (left) and A52 (right)
Very similar structures but no sequence identity was unidentifiable

Root mean squared deviation (RMSD)

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

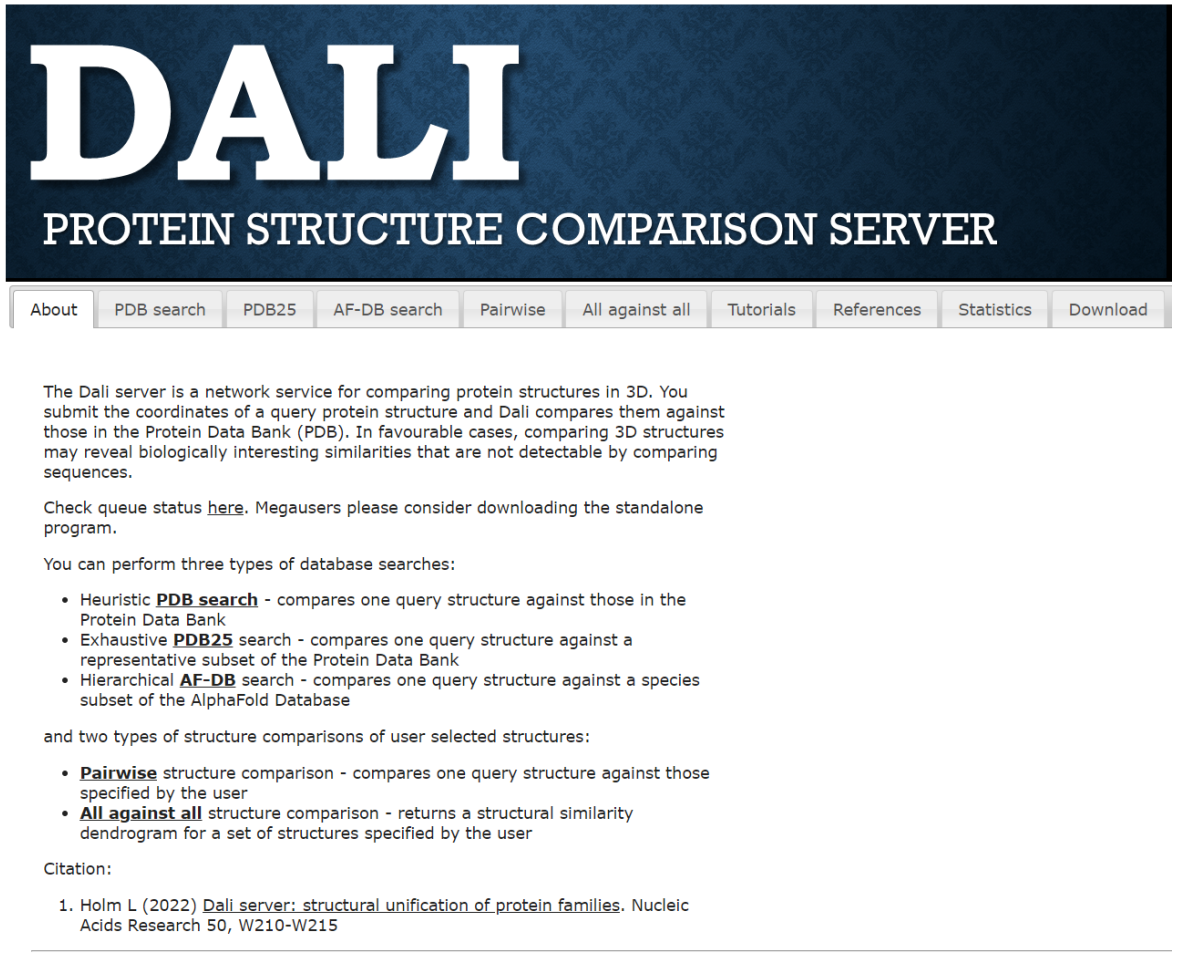
- Where δ is the distance between two equivalent atoms
- Generally measured between Cα atoms
- When quoting, need to specify both RMSD and number of matched atoms
- Exist several more advanced measures (normalised RMSD, GDT_TS, …)
- Can calculate in ChimeraX, PyMOL, COOT, …

How can you find other structures similar to your protein of interest?

Structure-based searches: DALI

http://ekhidna2.biocenter.helsinki.fi/dali/

- Can search databases of experimental structures and AlphaFold2 models



Structure-based searches: DALI

http://ekhidna2.biocenter.helsinki.fi/dali/

- Can search databases of experimental structures and AlphaFold2 models
- Search results order by Z score (statistical significance)
- Uses all residue pairwise distance matrices, not RMSD
- Z scores above 7 or so are likely to be meaningful

Results: 7phyA

Query: 7phyA

MOLECULE: PROTEIN E2;

Select neighbours (check boxes) for viewing as multiple structural alignment or 3D superimposition. The list of neighbours is sorted by Z-score. Similarities with a Z-score lower than 2 are spurious. Each neighbour has links to pairwise structural alignment with the query structure, and to the PDB format coordinate file where the neighbour is superimposed onto the query structure.

Structural Alignment | ☑ Expand gaps | 3D Superimposition (PV) | SANS | PANZ | Pfam | Reset Selection

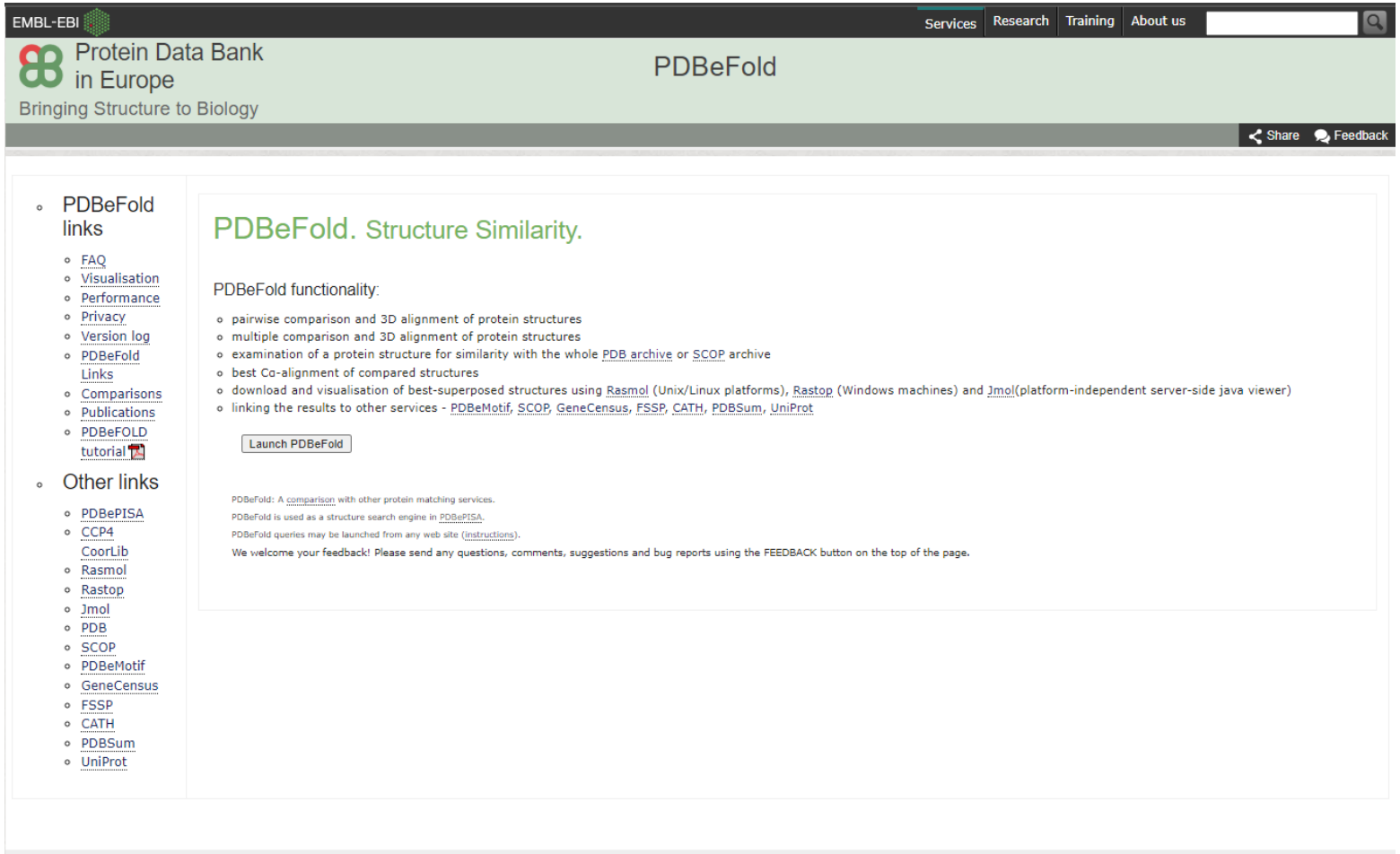
Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
☐ 1:	7phy-A	54.5	0.0	732	732	100	PDB	MOLECULE: PROTEIN E2;
☐ 2:	6plm-B	10.3	4.8	256	752	10	PDB	MOLECULE: SIDJ PROTEIN;
☐ 3:	7r5s-I	4.9	15.1	176	622	7	PDB	MOLECULE: CENTROMERE PROTEIN H;
☐ 4:	1z2c-B	4.8	18.7	127	346	8	PDB	MOLECULE: RHO-RELATED GTP-BINDING PROTEIN RHOC;
☐ 5:	6yle-D	4.7	9.8	139	547	5	PDB	MOLECULE: PRE-RRNA-PROCESSING PROTEIN IPI3;
☐ 6:	4lnb-A	4.7	10.0	156	339	6	PDB	MOLECULE: CAAX FARNESYLTRANSFERASE ALPHA SUBUNIT RAM2;
☐ 7:	3dad-A	4.5	10.1	104	324	13	PDB	MOLECULE: FH1/FH2 DOMAIN-CONTAINING PROTEIN 1;
☐ 8:	8e2f-A	4.4	12.0	140	771	4	PDB	MOLECULE: BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 6;
☐ 9:	7uwf-C	4.3	12.1	119	517	5	PDB	MOLECULE: WD REPEAT-CONTAINING PROTEIN 18;
☐ 10:	6dee-A	4.2	13.7	125	382	6	PDB	MOLECULE: NCK-INTERACTING PROTEIN WITH SH3 DOMAIN;
☐ 11:	4d0l-E	4.2	14.3	110	479	10	PDB	MOLECULE: PHOSPHATIDYLINOSITOL 4-KINASE BETA;
☐ 12:	7zkq-C	4.2	8.0	122	416	9	PDB	MOLECULE: NADH DEHYDROGENASE SUBUNIT 2;
☐ 13:	4imj-A	4.1	14.4	139	333	5	PDB	MOLECULE: SYMPLEKIN;
☐ 14:	7x9r-A	4.1	19.4	153	448	8	PDB	MOLECULE: GLYCOSYL TRANSFERASE FAMILY 2;
☐ 15:	4cem-B	4.1	3.8	122	309	4	PDB	MOLECULE: REGULATOR OF NONSENSE TRANSCRIPTS 2;
☐ 16:	4ww9-A	4.1	3.7	125	238	11	PDB	MOLECULE: EKC/KEOPS COMPLEX SUBUNIT BUD32;
☐ 17:	6yai-M	4.0	16.8	168	518	13	PDB	MOLECULE: CLATHRIN HEAVY CHAIN;
☐ 18:	6j6g-v	3.9	24.8	153	722	7	PDB	MOLECULE: PRE-MRNA-SPLICING FACTOR 8;
☐ 19:	5wy3-B	3.9	7.5	118	366	4	PDB	MOLECULE: PUTATIVE UNCHARACTERIZED PROTEIN;
☐ 20:	5l4k-P	3.9	17.4	138	456	3	PDB	MOLECULE: 26S PROTEASOME NON-ATPASE REGULATORY SUBUNIT 4;

Structure-based searches: PDBeFold

https://www.ebi.ac.uk/msd-srv/ssm/

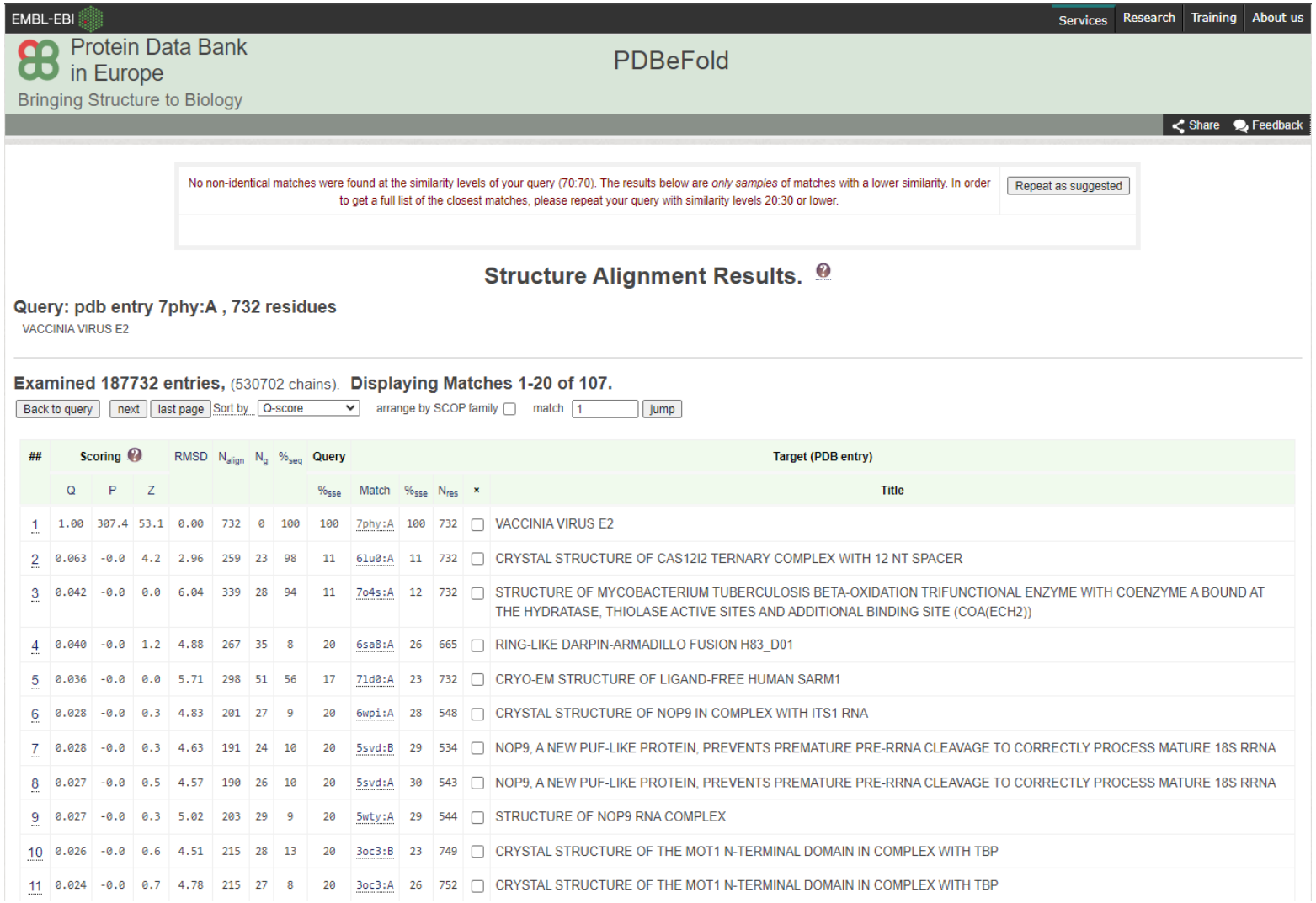
- Searches against PDB or a list of structures you supply
  - Can do all-vs-all pairwise superposition



Structure-based searches: PDBeFold

https://www.ebi.ac.uk/msd-srv/ssm/

- Searches against PDB or a list of structures you supply
  - Can do all-vs-all pairwise superposition
- Uses RMSD
- Reports Q score (1 is perfect) an Z score (higher is better)



EMBL-EBI Services Research Training About us

Protein Data Bank in Europe
Bringing Structure to Biology

PDBeFold

Share Feedback

No non-identical matches were found at the similarity levels of your query (70:70). The results below are *only samples* of matches with a lower similarity. In order to get a full list of the closest matches, please repeat your query with similarity levels 20:30 or lower.

Repeat as suggested

Structure Alignment Results.

Query: pdb entry 7phy:A , 732 residues

VACCINIA VIRUS E2

Examined 187732 entries, (530702 chains). **Displaying Matches 1-20 of 107.**

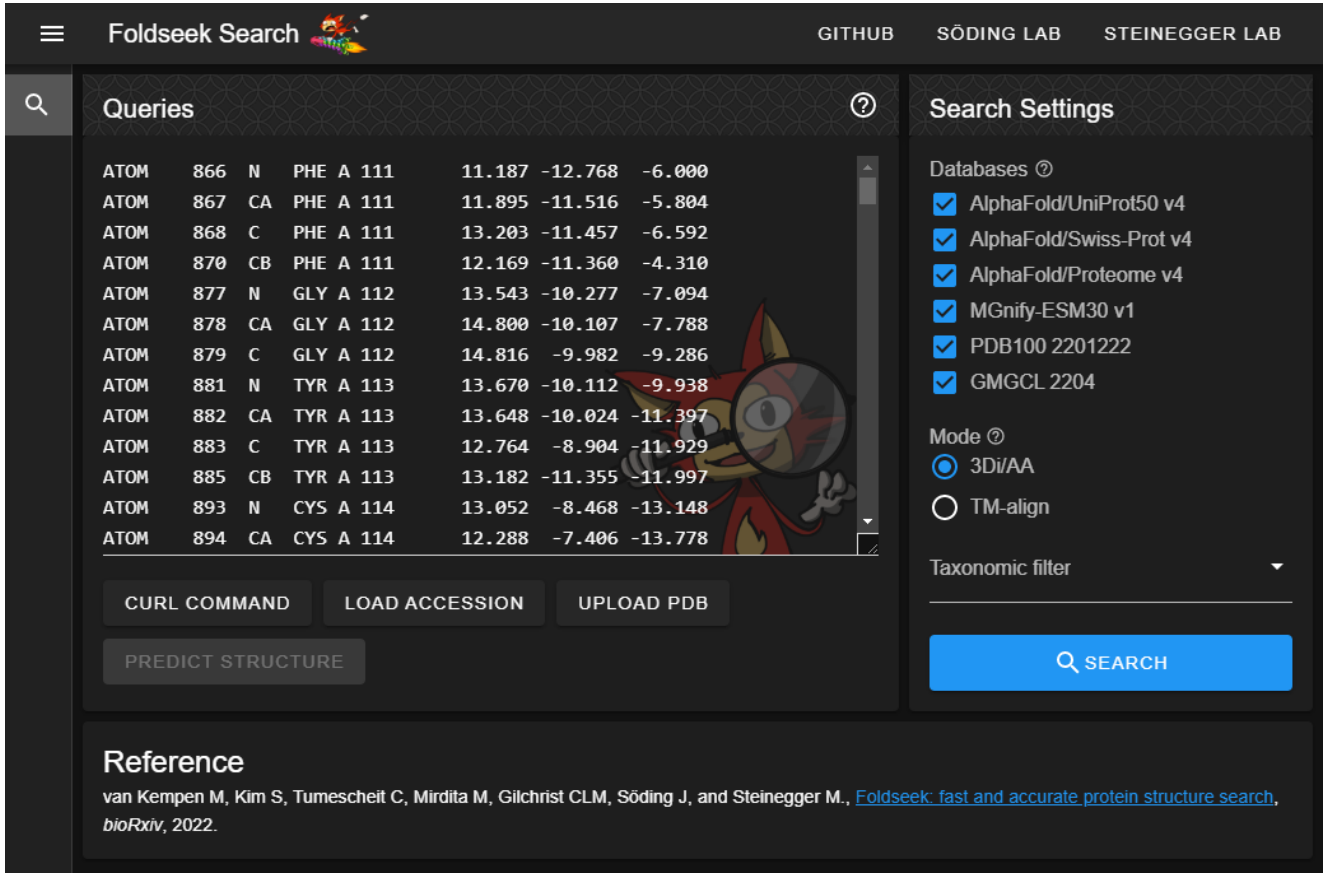
Back to query | next | last page | Sort by Q-score | arrange by SCOP family | match 1 | jump

##	Scoring			RMSD	N_{align}	N_g	$\%_{seq}$	Query	Target (PDB entry)				
	Q	P	Z					$\%_{sse}$	Match	$\%_{sse}$	N_{res}	×	Title
1	1.00	307.4	53.1	0.00	732	0	100	100	7phy:A	100	732	☐	VACCINIA VIRUS E2
2	0.063	-0.0	4.2	2.96	259	23	98	11	6lu0:A	11	732	☐	CRYSTAL STRUCTURE OF CAS12I2 TERNARY COMPLEX WITH 12 NT SPACER
3	0.042	-0.0	0.0	6.04	339	28	94	11	7o4s:A	12	732	☐	STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS BETA-OXIDATION TRIFUNCTIONAL ENZYME WITH COENZYME A BOUND AT THE HYDRATASE, THIOLASE ACTIVE SITES AND ADDITIONAL BINDING SITE (COA(ECH2))
4	0.040	-0.0	1.2	4.88	267	35	8	20	6sa8:A	26	665	☐	RING-LIKE DARPIN-ARMADILLO FUSION H83_D01
5	0.036	-0.0	0.0	5.71	298	51	56	17	7ld0:A	23	732	☐	CRYO-EM STRUCTURE OF LIGAND-FREE HUMAN SARM1
6	0.028	-0.0	0.3	4.83	201	27	9	20	6wpi:A	28	548	☐	CRYSTAL STRUCTURE OF NOP9 IN COMPLEX WITH ITS1 RNA
7	0.028	-0.0	0.3	4.63	191	24	10	20	5svd:B	29	534	☐	NOP9, A NEW PUF-LIKE PROTEIN, PREVENTS PREMATURE PRE-RRNA CLEAVAGE TO CORRECTLY PROCESS MATURE 18S RRNA
8	0.027	-0.0	0.5	4.57	190	26	10	20	5svd:A	30	543	☐	NOP9, A NEW PUF-LIKE PROTEIN, PREVENTS PREMATURE PRE-RRNA CLEAVAGE TO CORRECTLY PROCESS MATURE 18S RRNA
9	0.027	-0.0	0.3	5.02	203	29	9	20	5wty:A	29	544	☐	STRUCTURE OF NOP9 RNA COMPLEX
10	0.026	-0.0	0.6	4.51	215	28	13	20	3oc3:B	23	749	☐	CRYSTAL STRUCTURE OF THE MOT1 N-TERMINAL DOMAIN IN COMPLEX WITH TBP
11	0.024	-0.0	0.7	4.78	215	27	8	20	3oc3:A	26	752	☐	CRYSTAL STRUCTURE OF THE MOT1 N-TERMINAL DOMAIN IN COMPLEX WITH TBP

Sequence/Structure based search: Foldseek

https://search.foldseek.com/search

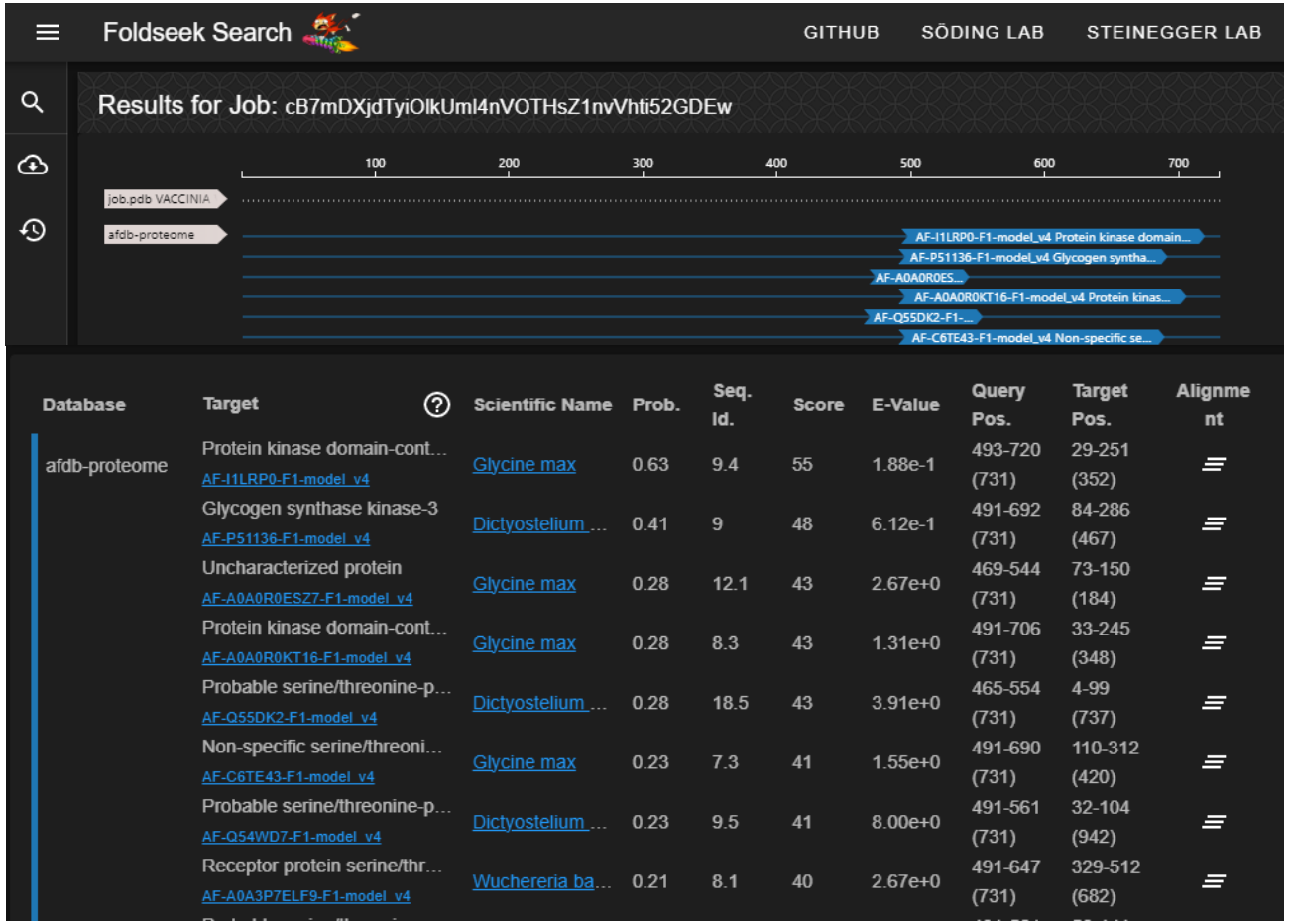
- Combines sequence analysis and structure analysis using deep learning
- Much faster than other techniques, so can search larger databases (database with AlphaFold2 models of all proteins)



Sequence/Structure based search: Foldseek

https://search.foldseek.com/search

- Combines sequence analysis and structure analysis using deep learning
- Much faster than other techniques, so can search larger databases (database with AlphaFold2 models of all proteins)
- Scored by E-value (probability of significance, lower is better)

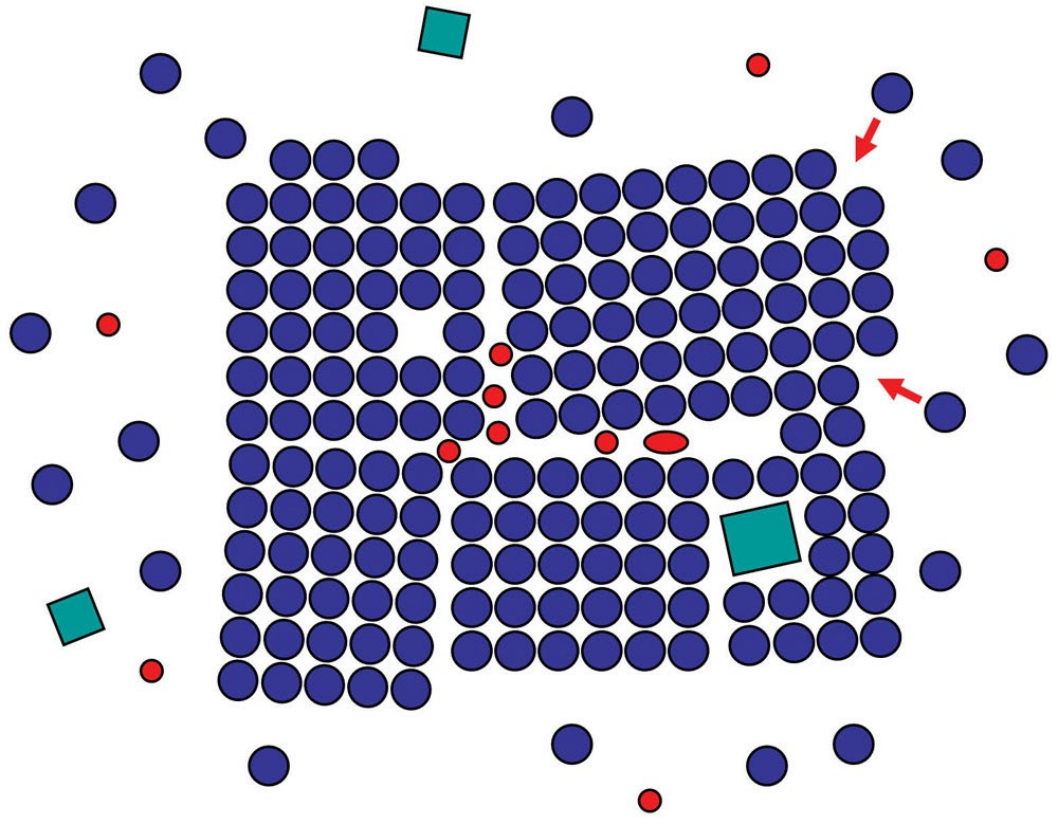


Database	Target	Scientific Name	Prob.	Seq. Id.	Score	E-Value	Query Pos.	Target Pos.	Alignment
afdb-proteome	Protein kinase domain-cont... AF-I1LRP0-F1-model_v4	Glycine max	0.63	9.4	55	1.88e-1	493-720 (731)	29-251 (352)	
	Glycogen synthase kinase-3 AF-P51136-F1-model_v4	Dictyostelium ...	0.41	9	48	6.12e-1	491-692 (731)	84-286 (467)	
	Uncharacterized protein AF-A0A0R0ESZ7-F1-model_v4	Glycine max	0.28	12.1	43	2.67e+0	469-544 (731)	73-150 (184)	
	Protein kinase domain-cont... AF-A0A0R0KT16-F1-model_v4	Glycine max	0.28	8.3	43	1.31e+0	491-706 (731)	33-245 (348)	
	Probable serine/threonine-p... AF-Q55DK2-F1-model_v4	Dictyostelium ...	0.28	18.5	43	3.91e+0	465-554 (731)	4-99 (737)	
	Non-specific serine/threoni... AF-C6TE43-F1-model_v4	Glycine max	0.23	7.3	41	1.55e+0	491-690 (731)	110-312 (420)	
	Probable serine/threonine-p... AF-Q54WD7-F1-model_v4	Dictyostelium ...	0.23	9.5	41	8.00e+0	491-561 (731)	32-104 (942)	
	Receptor protein serine/thr... AF-A0A3P7ELF9-F1-model_v4	Wuchereria ba...	0.21	8.1	40	2.67e+0	491-647 (731)	329-512 (682)	

Analysing protein complexes

Inspecting protein interaction interfaces

- Many (most) proteins function as part of macromolecular assemblies
- Crystals are large arrays of protein molecules
  - Some interactions between adjacent molecules will be biologically meaningful
  - Other interactions will be *crystallisation artefacts* (don't occur in nature)
- How can we identify and characterise ***biologically meaningful*** protein interactions?

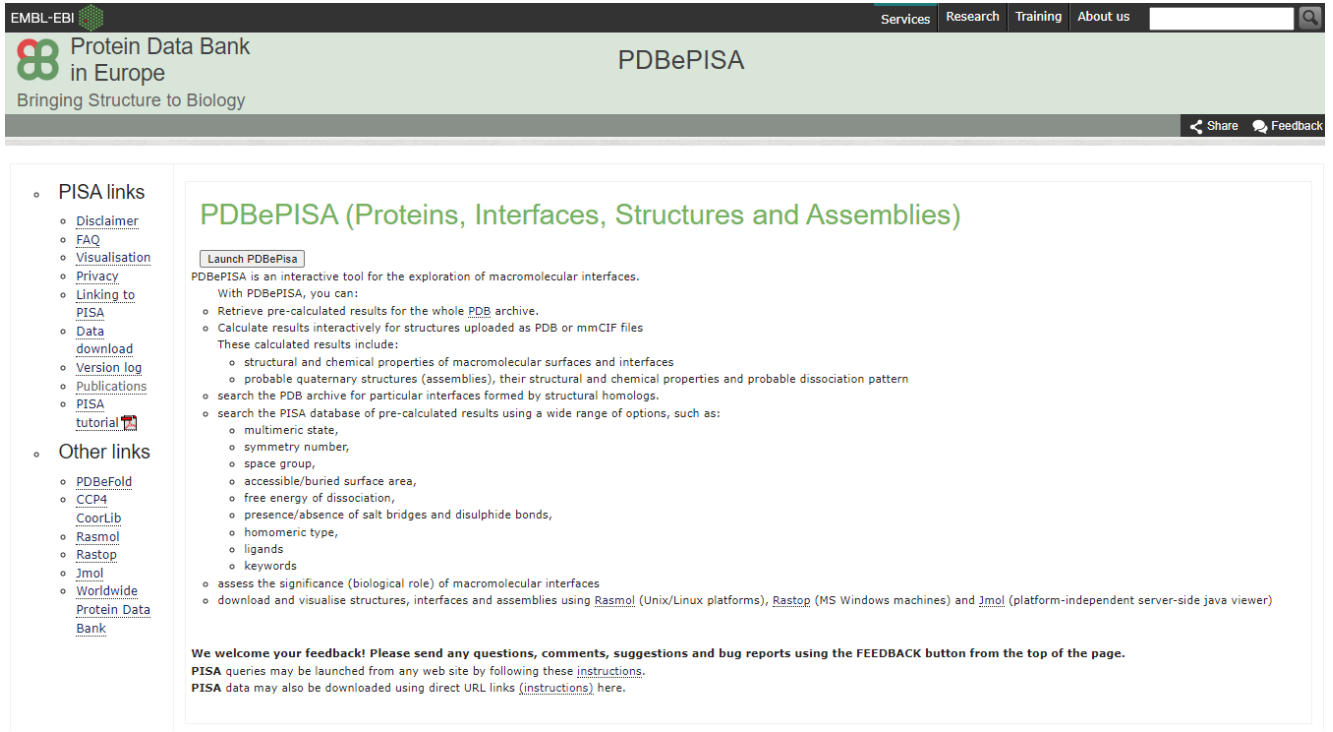


Inspecting protein interaction interfaces – PDBePISA

- PDBePISA server

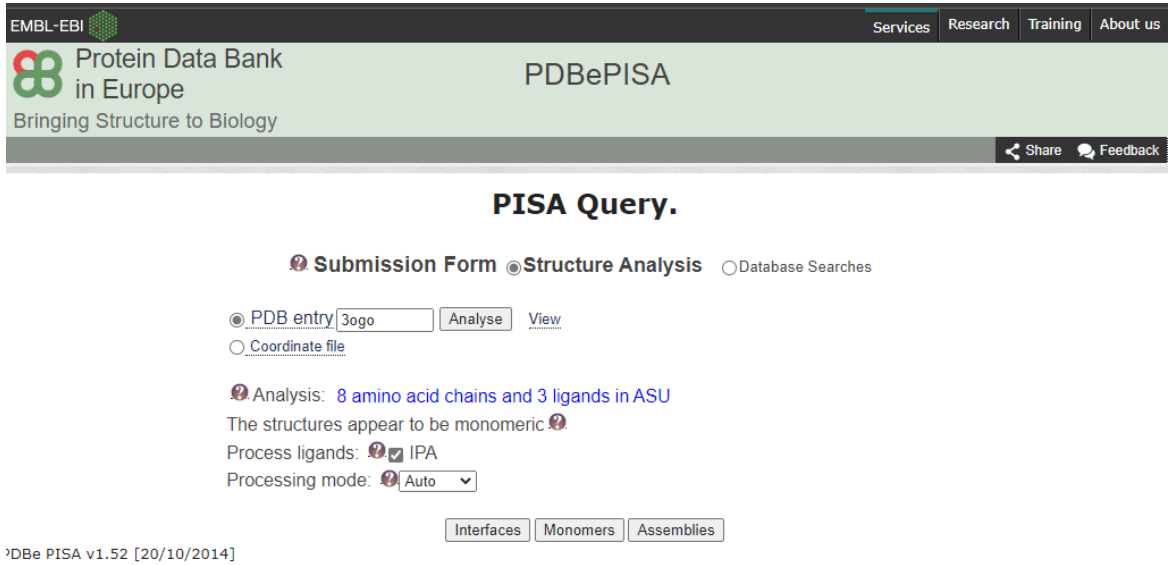
https://www.ebi.ac.uk/pdbe/pisa/pistart.html

- Analyses the Gibbs free energy (ΔG) gained or lost by complex formation, plus the size and shape of the interaction



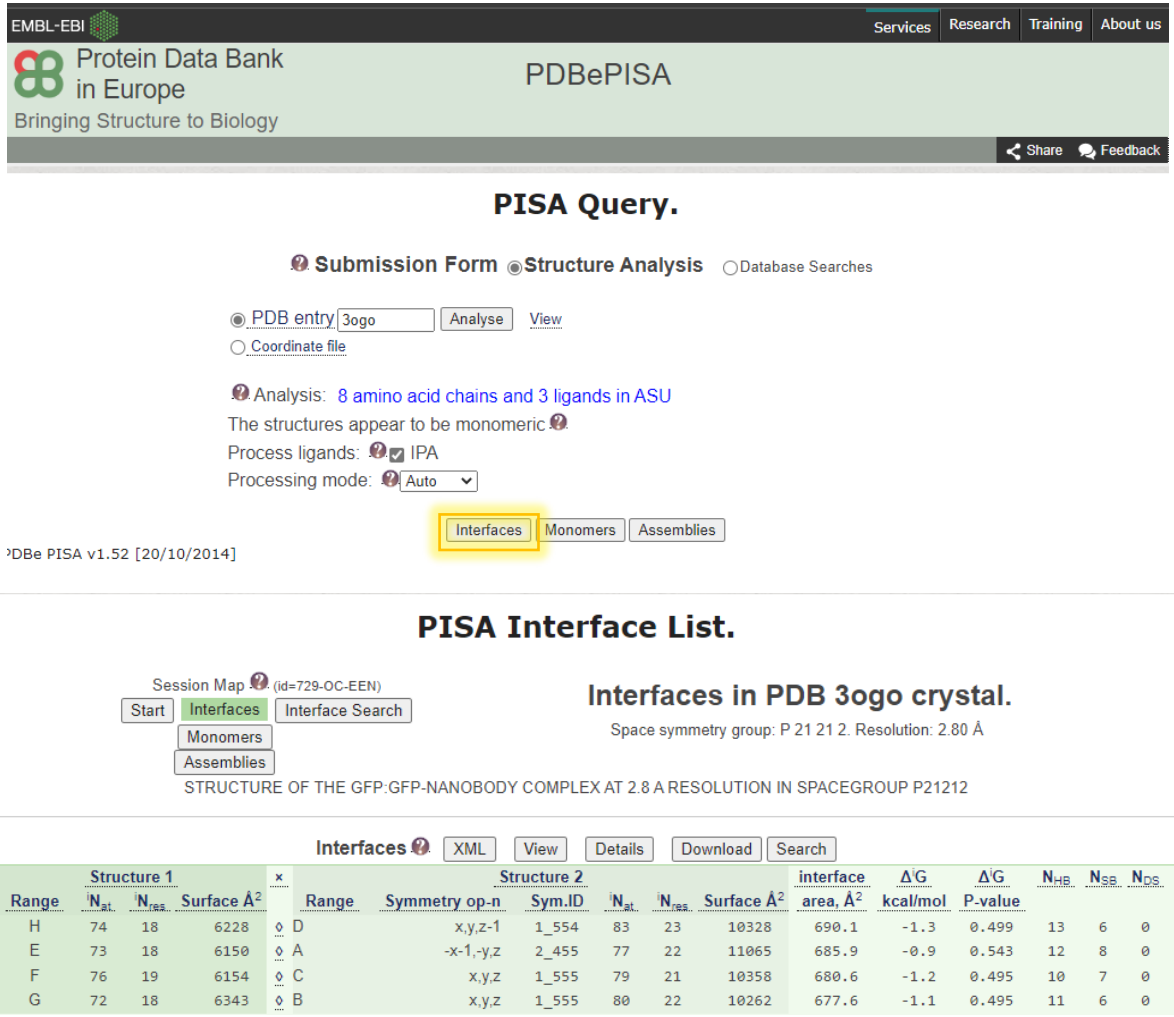
Inspecting protein interaction interfaces – PDBePISA

- Analyse any structure from PDB



Inspecting protein interaction interfaces – PDBePISA

- Analyse any structure from PDB
- Analysing ***interfaces***
  - Extent of interaction
  - Number of hydrogen bonds, salt bridges and disulfide bonds
  - Amount of energy gained/lost by burying interface
    - Hiding residues from solvent
  - Suggests which interactions are biologically meaningful CSS score (not always correct)



Id	## NN	«»	Structure 1 Range	N_{at}	N_{res}	Surface Å²	×	Structure 2 Range	Symmetry op-n	Sym.ID	N_{at}	N_{res}	Surface Å²	interface area, Å²	ΔG kcal/mol	ΔG P-value	N_{HB}	N_{SB}	N_{DS}	CSS
1	1	◉	H	74	18	6228	◊	D	x,y,z-1	1_554	83	23	10328	690.1	-1.3	0.499	13	6	0	0.000
	2	○	E	73	18	6150	◊	A	-x-1,-y,z	2_455	77	22	11065	685.9	-0.9	0.543	12	8	0	0.000
	3	○	F	76	19	6154	◊	C	x,y,z	1_555	79	21	10358	680.6	-1.2	0.495	10	7	0	0.000
	4	○	G	72	18	6343	◊	B	x,y,z	1_555	80	22	10262	677.6	-1.1	0.495	11	6	0	0.000

Inspecting protein interaction interfaces – PDBePISA

- Can view the interaction interface interactively

PISA Interface List.

Session Map (id=729-OC-EEN)

Start | Interfaces | Interface Search | Monomers | Assemblies

Interfaces in PDB 3ogo crystal.

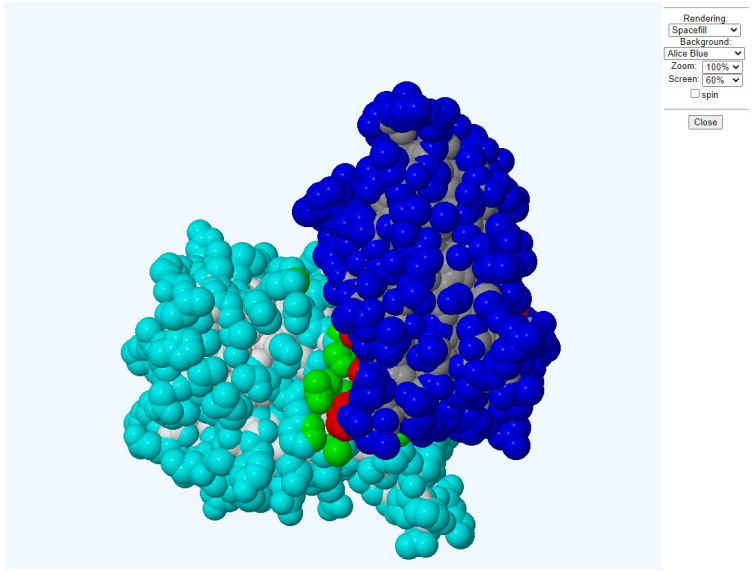
Space symmetry group: P 21 21 2. Resolution: 2.80 Å

STRUCTURE OF THE GFP:GFP-NANOBODY COMPLEX AT 2.8 A RESOLUTION IN SPACEGROUP P21212

Interfaces XML | View | Details | Download | Search

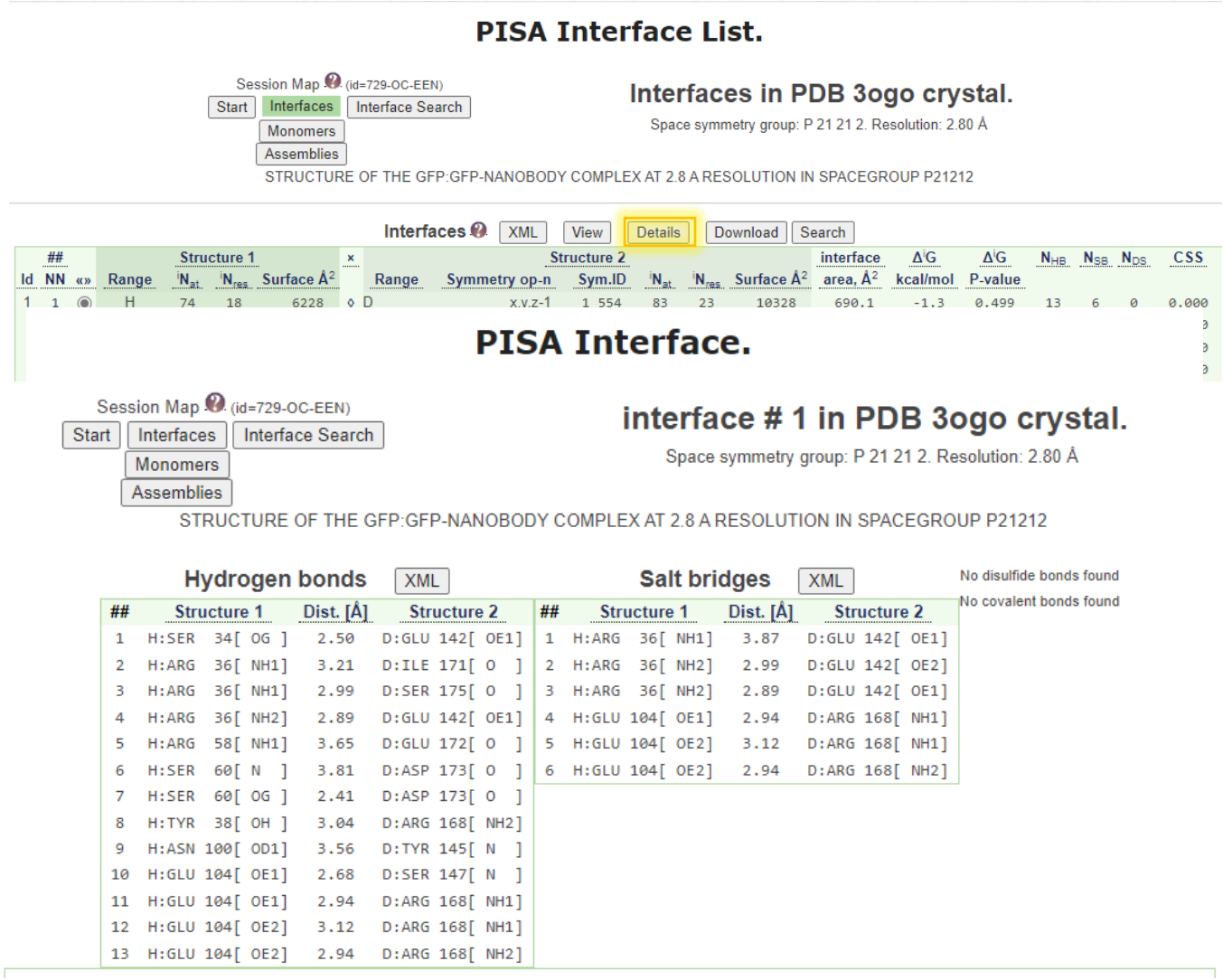
Id	## NN	«»	Structure 1 Range	N_{at}	N_{res}	Surface Å²	×	Structure 2 Range	Symmetry op-n	Sym.ID	N_{at}	N_{res}	Surface Å²	interface area, Å²	Δ^iG kcal/mol	Δ^iG P-value	N_{HB}	N_{SB}	N_{DS}	CSS
1	1	◉	H	74	18	6228	◊	D	x,y,z-1	1_554	83	23	10328	690.1	-1.3	0.499	13	6	0	0.000
	2	○	E	73	18	6150	◊	A	-x-1,-y,z	2_455	77	22	11065	685.9	-0.9	0.543	12	8	0	0.000
	3	○	F	76	19	6154	◊	C	x,y,z	1_555	79	21	10358	680.6	-1.2	0.495	10	7	0	0.000
	4	○	G	72	18	6343	◊	B	x,y,z	1_555	80	22	10262	677.6	-1.1	0.495	11	6	0	0.000

Interface #1 in 3ogo//H:D



Inspecting protein interaction interfaces – PDBePISA

- Can view the interaction interface interactively
- Can view the details of the interaction
  - Handy list of hydrogen bonds



PISA Interface List.

Session Map (id=729-OC-EEN)

Start | Interfaces | Interface Search | Monomers | Assemblies

Interfaces in PDB 3ogo crystal.

Space symmetry group: P 21 21 2. Resolution: 2.80 Å

STRUCTURE OF THE GFP:GFP-NANOBODY COMPLEX AT 2.8 A RESOLUTION IN SPACEGROUP P21212

Interfaces | XML | View | Details | Download | Search

Id	## NN	«»	Structure 1 Range	$^{I}N_{at}$	$^{I}N_{res}$	Surface Å²	×	Structure 2 Range	Symmetry op-n	Sym.ID	$^{I}N_{at}$	$^{I}N_{res}$	Surface Å²	interface area, Å²	Δ^{I}G kcal/mol	Δ^{I}G P-value	N_{HB}	N_{SB}	N_{DS}	CSS
1	1	◉	H	74	18	6228	◊	D	x,y,z-1	1 554	83	23	10328	690.1	-1.3	0.499	13	6	0	0.000

PISA Interface.

Session Map (id=729-OC-EEN)

Start | Interfaces | Interface Search | Monomers | Assemblies

interface # 1 in PDB 3ogo crystal.

Space symmetry group: P 21 21 2. Resolution: 2.80 Å

STRUCTURE OF THE GFP:GFP-NANOBODY COMPLEX AT 2.8 A RESOLUTION IN SPACEGROUP P21212

Hydrogen bonds XML

##	Structure 1	Dist. [Å]	Structure 2
1	H:SER 34[OG]	2.50	D:GLU 142[OE1]
2	H:ARG 36[NH1]	3.21	D:ILE 171[O]
3	H:ARG 36[NH1]	2.99	D:SER 175[O]
4	H:ARG 36[NH2]	2.89	D:GLU 142[OE1]
5	H:ARG 58[NH1]	3.65	D:GLU 172[O]
6	H:SER 60[N]	3.81	D:ASP 173[O]
7	H:SER 60[OG]	2.41	D:ASP 173[O]
8	H:TYR 38[OH]	3.04	D:ARG 168[NH2]
9	H:ASN 100[OD1]	3.56	D:TYR 145[N]
10	H:GLU 104[OE1]	2.68	D:SER 147[N]
11	H:GLU 104[OE1]	2.94	D:ARG 168[NH1]
12	H:GLU 104[OE2]	3.12	D:ARG 168[NH1]
13	H:GLU 104[OE2]	2.94	D:ARG 168[NH2]

Salt bridges XML

##	Structure 1	Dist. [Å]	Structure 2
1	H:ARG 36[NH1]	3.87	D:GLU 142[OE1]
2	H:ARG 36[NH2]	2.99	D:GLU 142[OE2]
3	H:ARG 36[NH2]	2.89	D:GLU 142[OE1]
4	H:GLU 104[OE1]	2.94	D:ARG 168[NH1]
5	H:GLU 104[OE2]	3.12	D:ARG 168[NH1]
6	H:GLU 104[OE2]	2.94	D:ARG 168[NH2]

No disulfide bonds found

No covalent bonds found

Inspecting protein interaction interfaces – PDBePISA

- Can view the interaction interface interactively
- Can view the details of the interaction
  - Handy list of hydrogen bonds
  - And list of which residues are buried (hidden from solvent) by the interaction interface

PISA Interface List.

Session Map (id=729-OC-EEN)

Start | Interfaces | Interface Search | Monomers | Assemblies

Interfaces in PDB 3ogo crystal.

Space symmetry group: P 21 21 2. Resolution: 2.80 Å

STRUCTURE OF THE GFP:GFP-NANOBODY COMPLEX AT 2.8 A RESOLUTION IN SPACEGROUP P21212

Interfaces | XML | View | Details | Download | Search

Id	## NN	«»	Structure 1 Range	$^{i}N_{at}$	$^{i}N_{res}$	Surface Å^2	×	Structure 2 Range	Symmetry op-n	Sym.ID	$^{i}N_{at}$	$^{i}N_{res}$	Surface Å^2	interface area, Å^2	$\Delta^{i}G$ kcal/mol	$\Delta^{i}G$ P-value	N_{HB}	N_{SB}	N_{DS}	CSS
1	1	◉	H	74	18	6228	◊	D	x,y,z-1	1_554	83	23	10328	690.1	-1.3	0.499	13	6	0	0.000
	2	○	E	73	18	6150	◊	A	-x-1,-y,z	2_455	77	22	11065	685.9	-0.9	0.543	12	8	0	0.000
	3	○	F	76	19	6154	◊	C	x,y,z	1_555	79	21	10358	680.6	-1.2	0.495	10	7	0	0.000
	4	○	G	72	18	6343	◊	B	x,y,z	1_555	80	22	10262	677.6	-1.1	0.495	11	6	0	0.000

33	H:SER 34	H	31.48	22.17 |||||||	-0.02
34	H:MET 35		0.00	0.00		0.00
35	H:ARG 36	HS	88.67	88.08 ||||||||	-1.91
36	H:TRP 37		0.00	0.00		0.00
37	H:TYR 38	H	46.25	44.79 ||||||||	-0.13
38	H:ARG 39		14.04	0.00		0.00
39	H:GLN 40		55.42	0.00		0.00
40	H:ALA 41		25.98	0.00		0.00
41	H:PRO 42		116.06	0.00		0.00
42	H:GLY 43		87.82	0.00		0.00
43	H:LYS 44		142.93	0.00		0.00
44	H:GLU 45		156.25	31.40 |||	-0.38
45	H:ARG 46		89.45	6.16		-0.01
46	H:GLU 47		66.28	0.00		0.00
47	H:TRP 48		83.49	48.83 ||||||	0.78
48	H:VAL 49		0.00	0.00		0.00
49	H:ALA 50		0.00	0.00		0.00
50	H:GLY 51		4.94	4.94 ||||||||	0.08
51	H:MET 52		22.86	3.78 ||	0.05
52	H:SER 53		26.17	21.48 |||||||	0.26
53	H:SER 54		48.27	0.15		-0.00
54	H:ALA 55		79.34	0.00		0.00

33	D:GLY 35	0.63	0.00	0.00
34	D:ASP 36	23.04	0.00	0.00
35	D:ALA 37	11.42	0.00	0.00
36	D:THR 38	71.76	0.00	0.00
37	D:TYR 39	145.65	0.00	0.00
38	D:GLY 40	1.33	0.00	0.00
39	D:LYS 41	69.66	0.00	0.00
40	D:LEU 42	4.85	0.00	0.00
41	D:THR 43	57.41	0.00	0.00
42	D:LEU 44	6.18	0.00	0.00
43	D:LYS 45	58.58	0.00	0.00
44	D:PHE 46	0.00	0.00	0.00
45	D:ILE 47	20.59	0.00	0.00
46	D:CYS 48	0.98	0.00	0.00
47	D:THR 49	76.33	0.00	0.00
48	D:THR 50	72.72	0.00	0.00
49	D:GLY 51	44.87	0.00	0.00
50	D:LYS 52	148.11	0.00	0.00
51	D:LEU 53	1.55	0.00	0.00
52	D:PRO 54	24.33	0.00	0.00
53	D:VAL 55	0.98	0.00	0.00
54	D:PRO 56	10.55	0.00	0.00

Inspecting protein interaction interfaces – PDBePISA

- Can view the interaction interface interactively
- Can view the details of the interaction
  - Handy list of hydrogen bonds
  - And list of which residues are buried (hidden from solvent) by the interaction interface
- And can download the PDB file (structure) of the complex

PISA Interface List.

Session Map (id=729-OC-EEN)

Start | Interfaces | Interface Search | Monomers | Assemblies

Interfaces in PDB 3ogo crystal.

Space symmetry group: P 21 21 2. Resolution: 2.80 Å

STRUCTURE OF THE GFP:GFP-NANOBODY COMPLEX AT 2.8 A RESOLUTION IN SPACEGROUP P21212

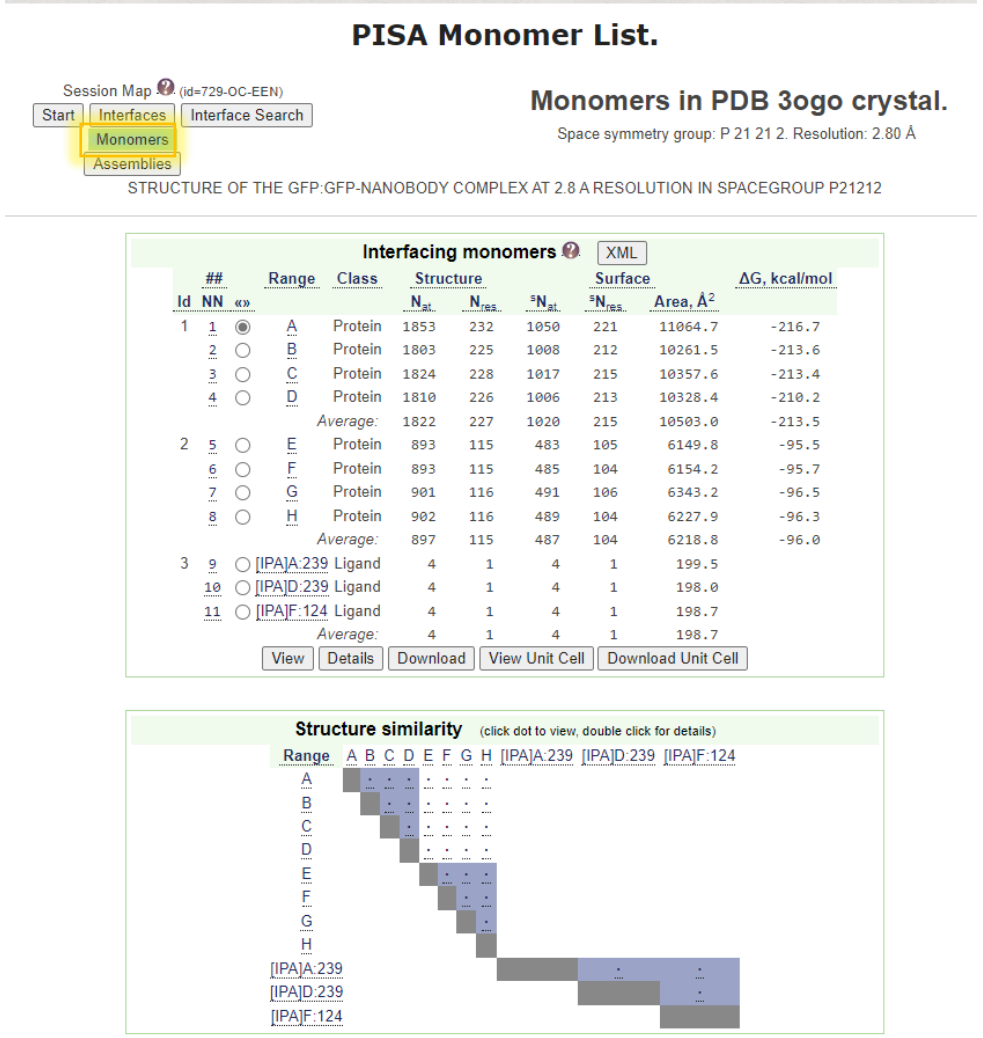
Interfaces | XML | View | Details | Download | Search

Id	## NN	«»	Structure 1 Range	$^{I}N_{at}$	$^{I}N_{res}$	Surface Å^2	×	Structure 2 Range	Symmetry op-n	Sym.ID	$^{I}N_{at}$	$^{I}N_{res}$	Surface Å^2	interface area, Å^2	$\Delta^{I}G$ kcal/mol	$\Delta^{I}G$ P-value	N_{HB}	N_{SB}	N_{DS}	CSS
1	1	◉	H	74	18	6228	◊	D	x,y,z-1	1_554	83	23	10328	690.1	-1.3	0.499	13	6	0	0.000
	2	○	E	73	18	6150	◊	A	-x-1,-y,z	2_455	77	22	11065	685.9	-0.9	0.543	12	8	0	0.000
	3	○	F	76	19	6154	◊	C	x,y,z	1_555	79	21	10358	680.6	-1.2	0.495	10	7	0	0.000
	4	○	G	72	18	6343	◊	B	x,y,z	1_555	80	22	10262	677.6	-1.1	0.495	11	6	0	0.000

```
ATOM  10288  N   GLN H   2      -32.681  26.405 -54.523  1.00 39.05           N
ATOM  10289  CA  GLN H   2      -33.107  27.175 -55.708  1.00 58.43           C
ATOM  10290  C   GLN H   2      -32.760  28.658 -55.664  1.00 59.86           C
ATOM  10291  O   GLN H   2      -32.635  29.254 -54.588  1.00 51.82           O
ATOM  10292  CB  GLN H   2      -34.605  27.044 -55.955  1.00 72.43           C
ATOM  10293  CG  GLN H   2      -35.214  28.205 -56.750  1.00 43.04           C
ATOM  10294  CD  GLN H   2      -36.144  29.063 -55.902  1.00 72.43           C
ATOM  10295  OE1 GLN H   2      -37.023  29.763 -56.426  1.00 55.72           O
ATOM  10296  NE2 GLN H   2      -35.958  29.008 -54.581  1.00 57.70           N
ATOM  10297  N   VAL H   3      -32.664  29.258 -56.848  1.00 52.53           N
ATOM  10298  CA  VAL H   3      -31.974  30.542 -56.997  1.00 42.84           C
ATOM  10299  C   VAL H   3      -32.737  31.771 -56.507  1.00 37.43           C
ATOM  10300  O   VAL H   3      -33.933  31.915 -56.738  1.00 31.85           O
ATOM  10301  CB  VAL H   3      -31.421  30.754 -58.433  1.00 36.56           C
ATOM  10302  CG1 VAL H   3      -31.941  29.682 -59.384  1.00 41.08           C
ATOM  10303  CG2 VAL H   3      -31.734  32.153 -58.931  1.00 25.64           C
ATOM  10304  N   GLN H   4      -32.010  32.661 -55.838  1.00 38.34           N
ATOM  10305  CA  GLN H   4      -32.609  33.805 -55.171  1.00 34.03           C
ATOM  10306  C   GLN H   4      -31.593  34.930 -55.001  1.00 36.47           C
```

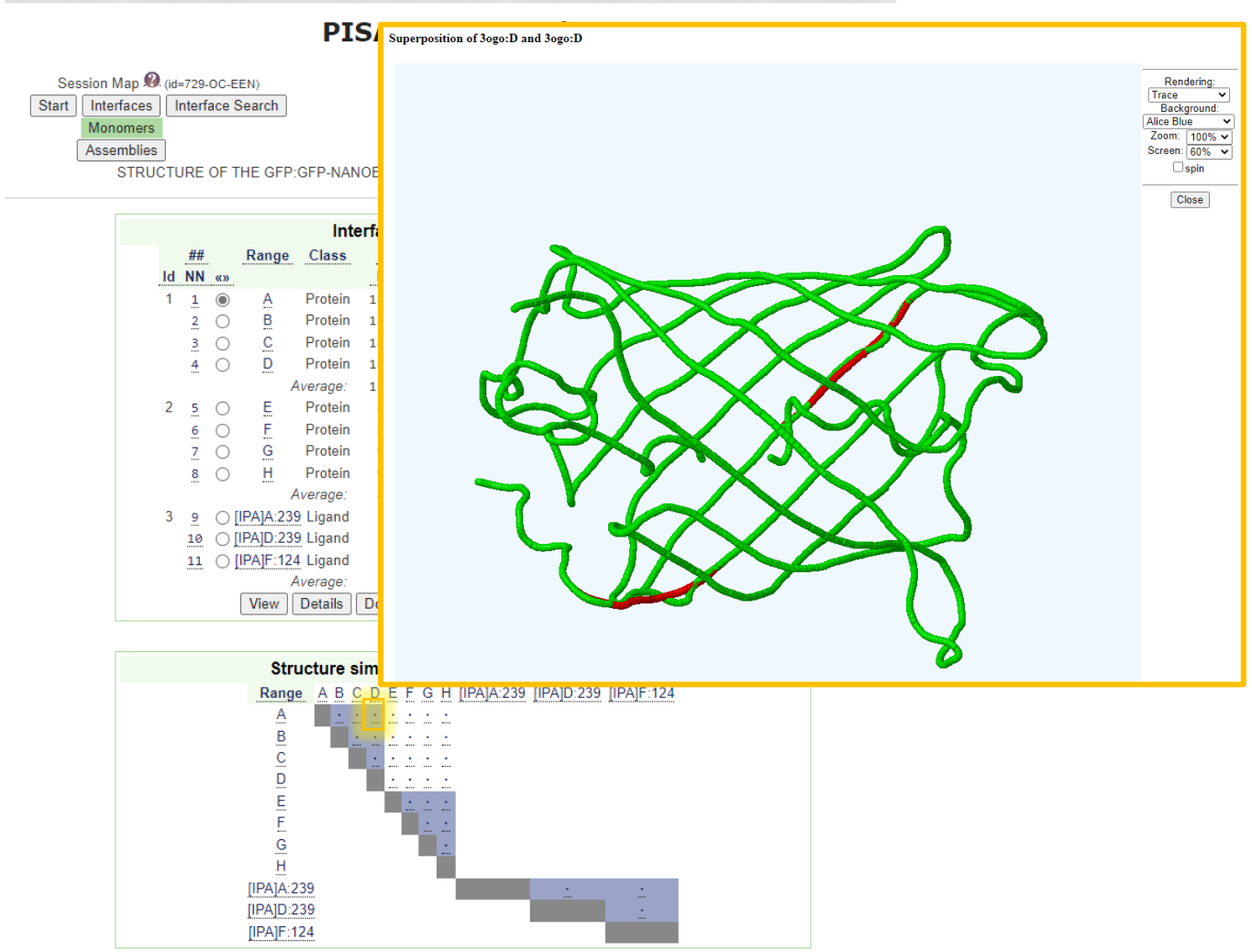
Inspecting protein interaction interfaces – PDBePISA

- The ***Monomers*** page also has a handy tool to highlight differences between different copies of the same protein in the crystal structure



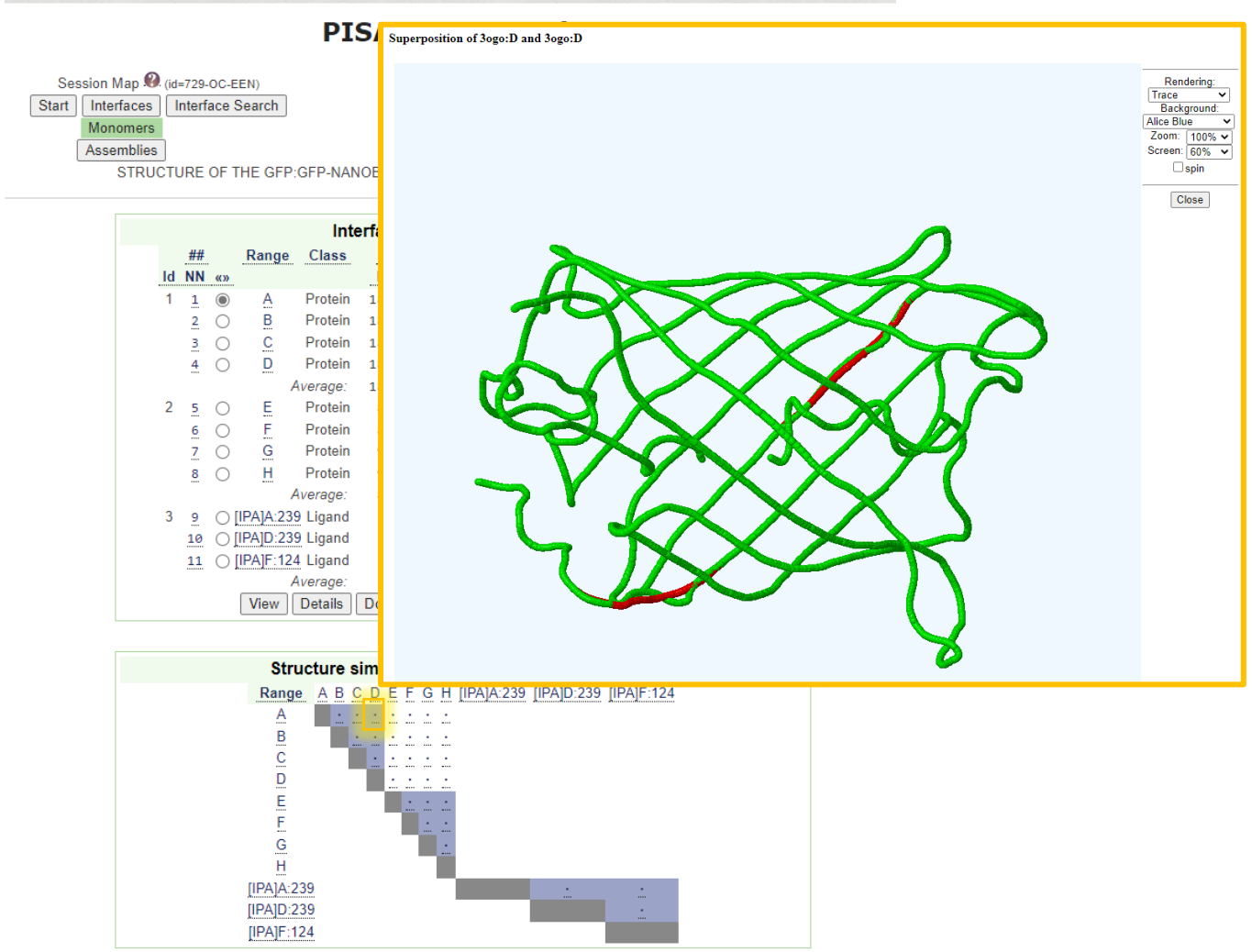
Inspecting protein interaction interfaces – PDBePISA

- The ***Monomers*** page also has a handy tool to highlight differences between different copies of the same protein in the crystal structure



Inspecting protein interaction interfaces – PDBePISA

- The ***Monomers*** page also has a handy tool to highlight differences between different copies of the same protein in the crystal structure
- Also works for cryoEM structures
  - To analyse interfaces



This talk - Recap

- Representations of proteins
- Mapping properties onto proteins
- Accessing protein structures
- Software for viewing protein structures
- Analysing structural similarity
- Inspecting protein interfaces

Tomorrow: How do we determine protein structures?