

UNIVERSITY OF  
CAMBRIDGE

# Looking at Protein Structures

Day 8: Tuesday 29<sup>th</sup> March

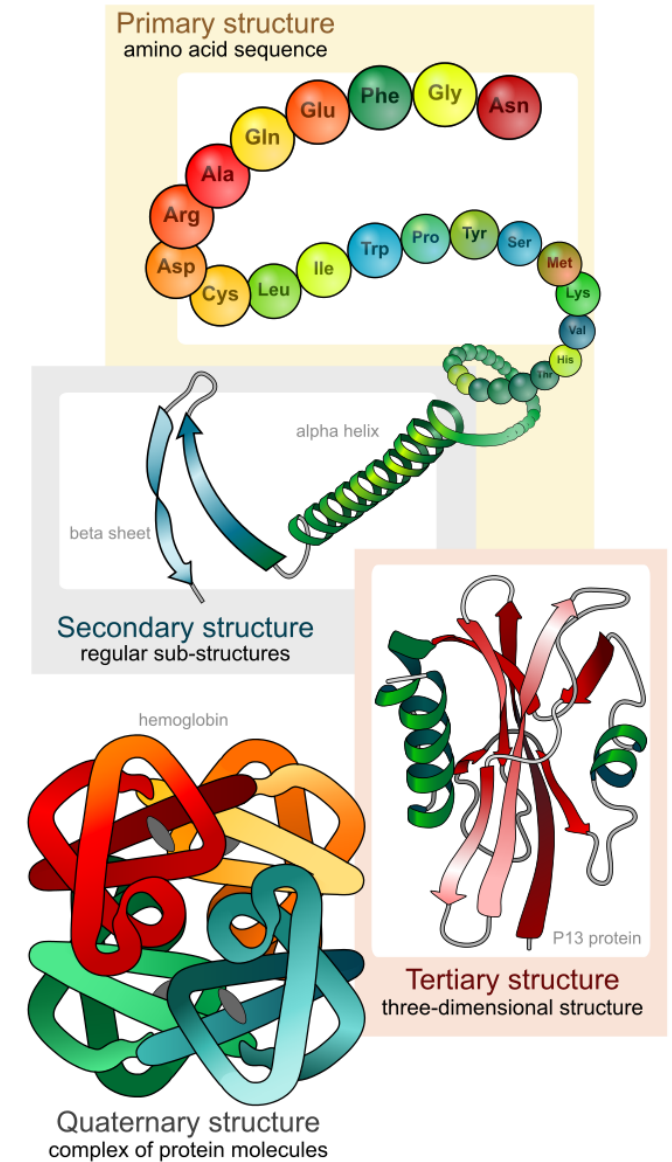
# This talk

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



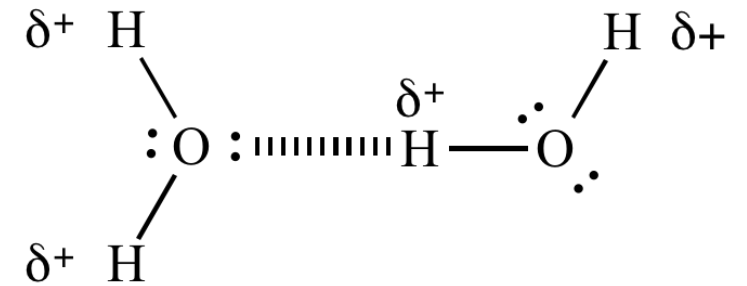
# Recap of protein structure

- 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



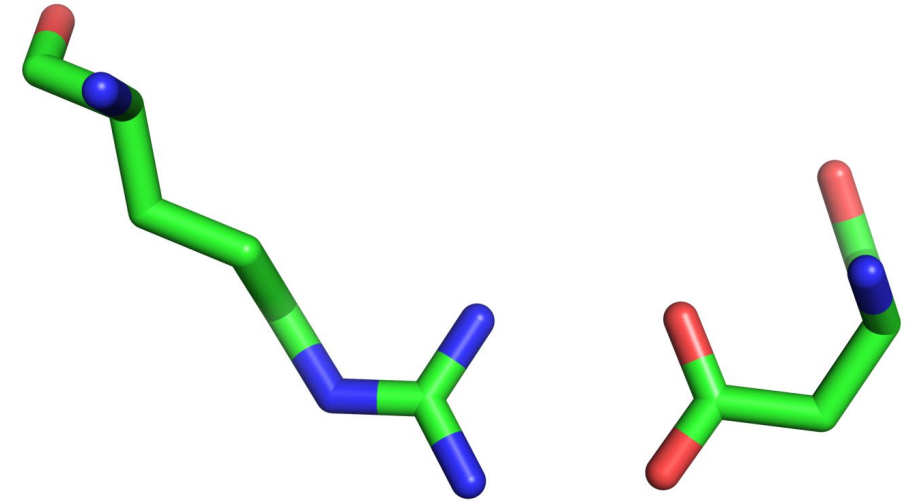
# Interactions between amino acids drive protein folding

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



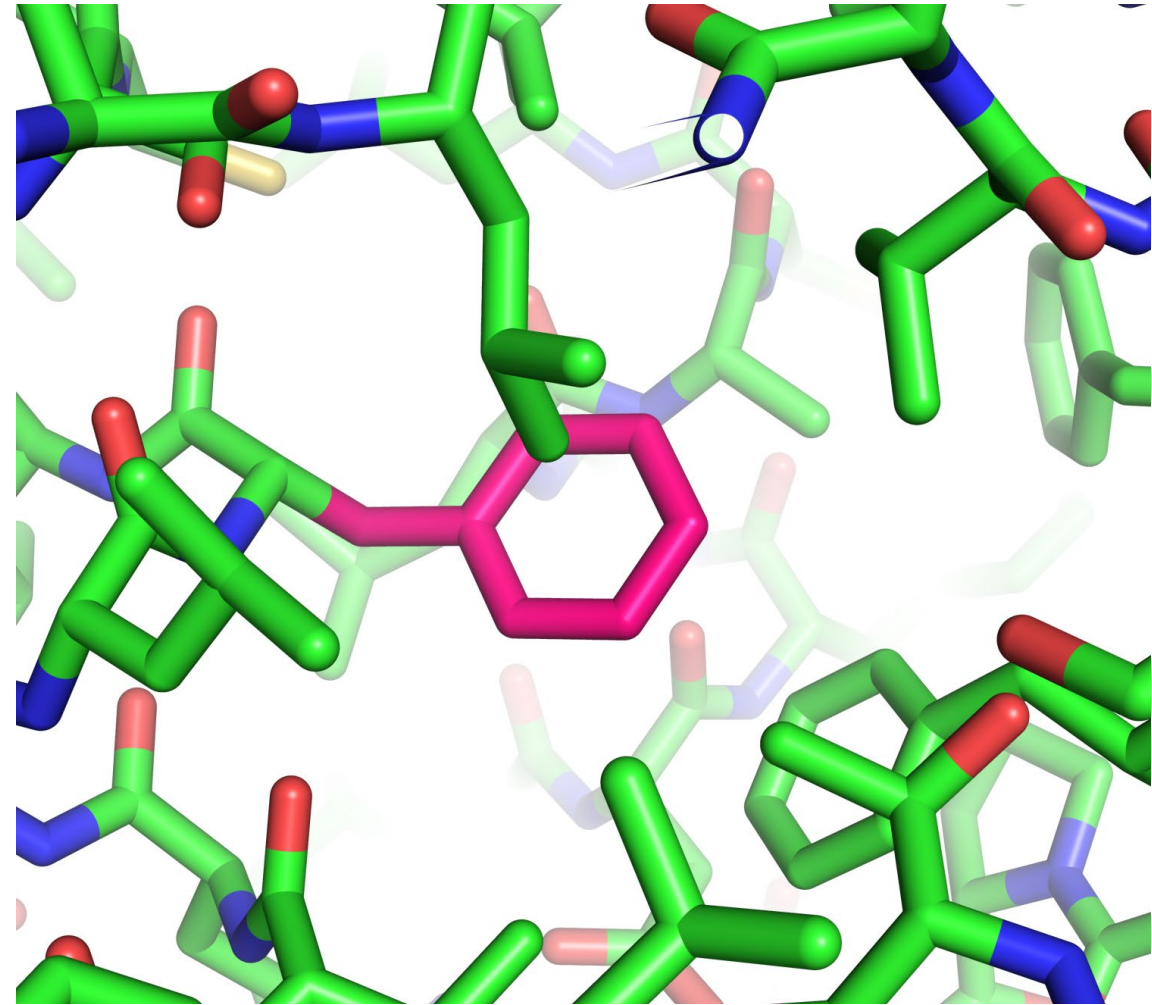
# Interactions between amino acids drive protein folding

- 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

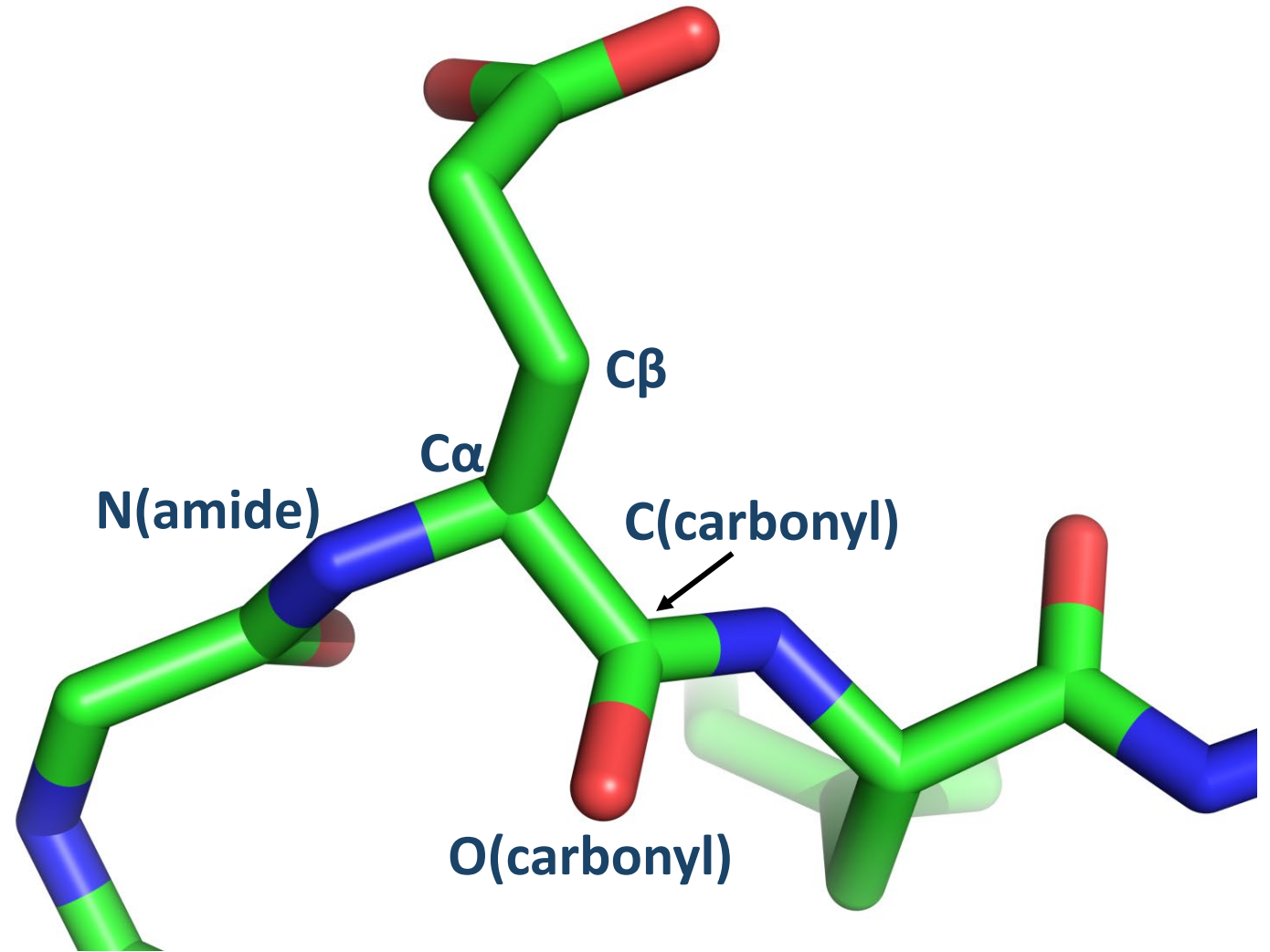


# Interactions between amino acids drive protein folding

- 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

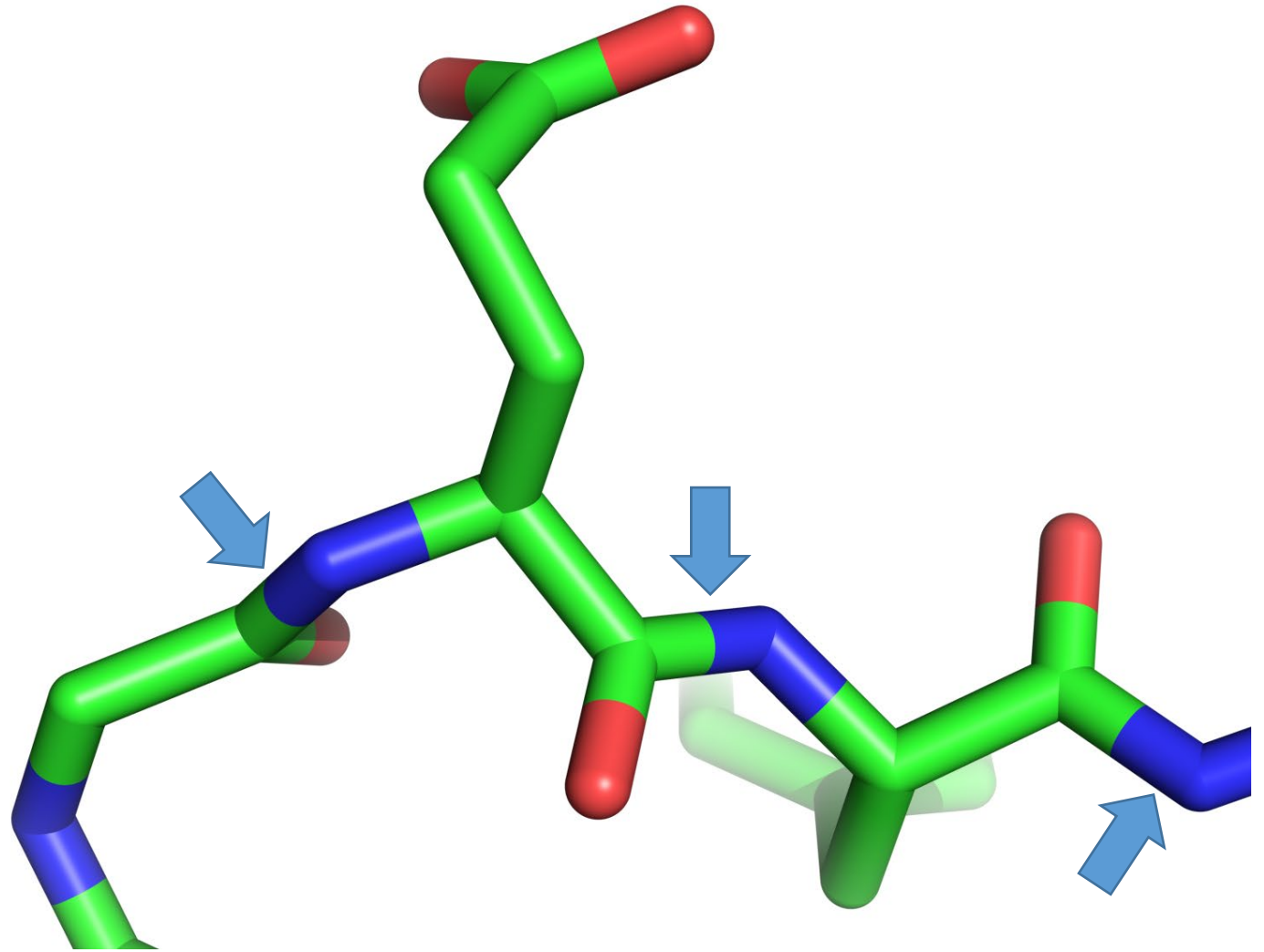


# Amino acid polymers



# Amino acid polymers

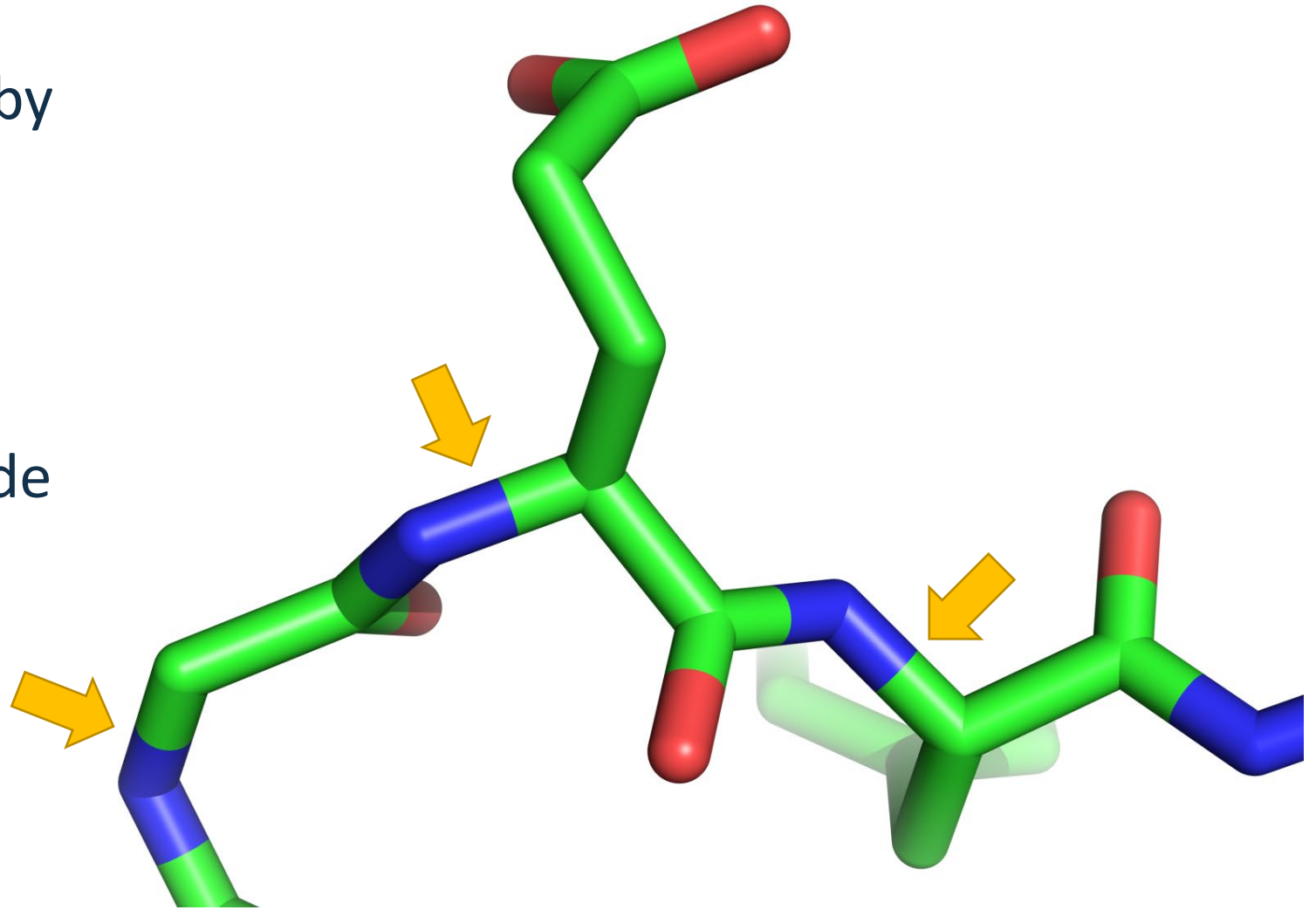
- Amino acids are joined by peptide bonds
  - Planar
  - *trans* ( $180^\circ$ ) or *cis* ( $0^\circ$ )





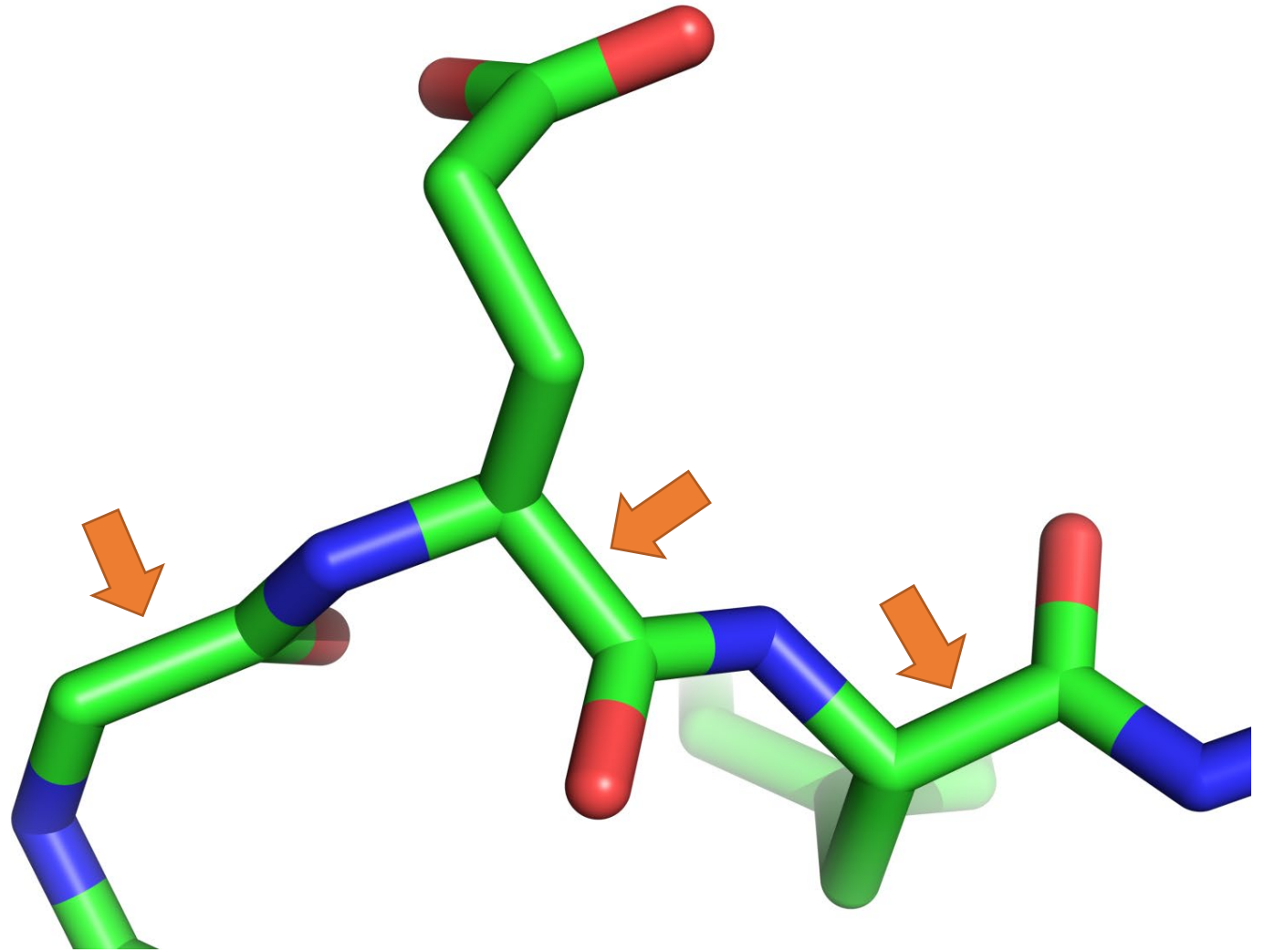
# Amino acid polymers

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



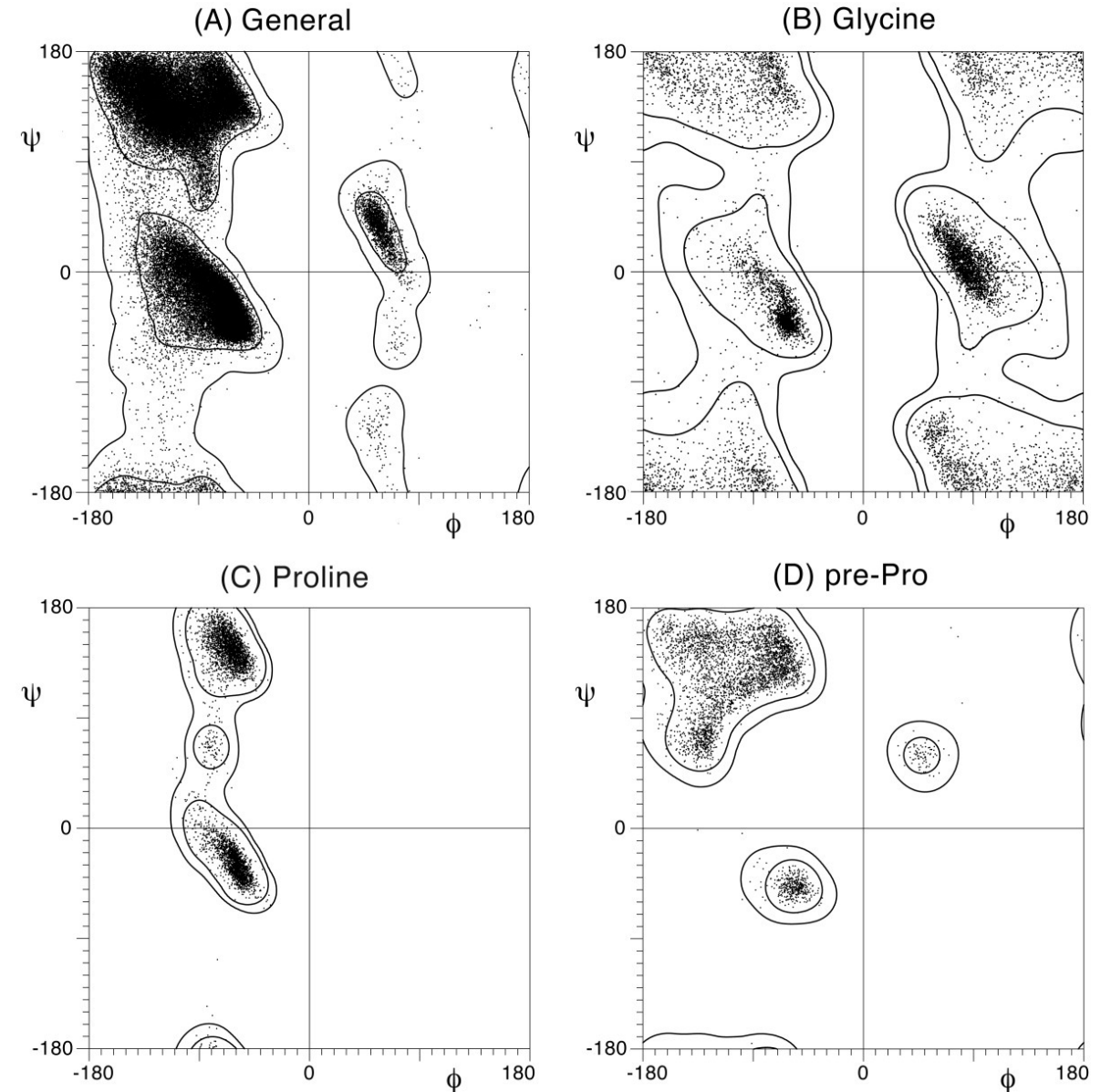
# Amino acid polymers

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



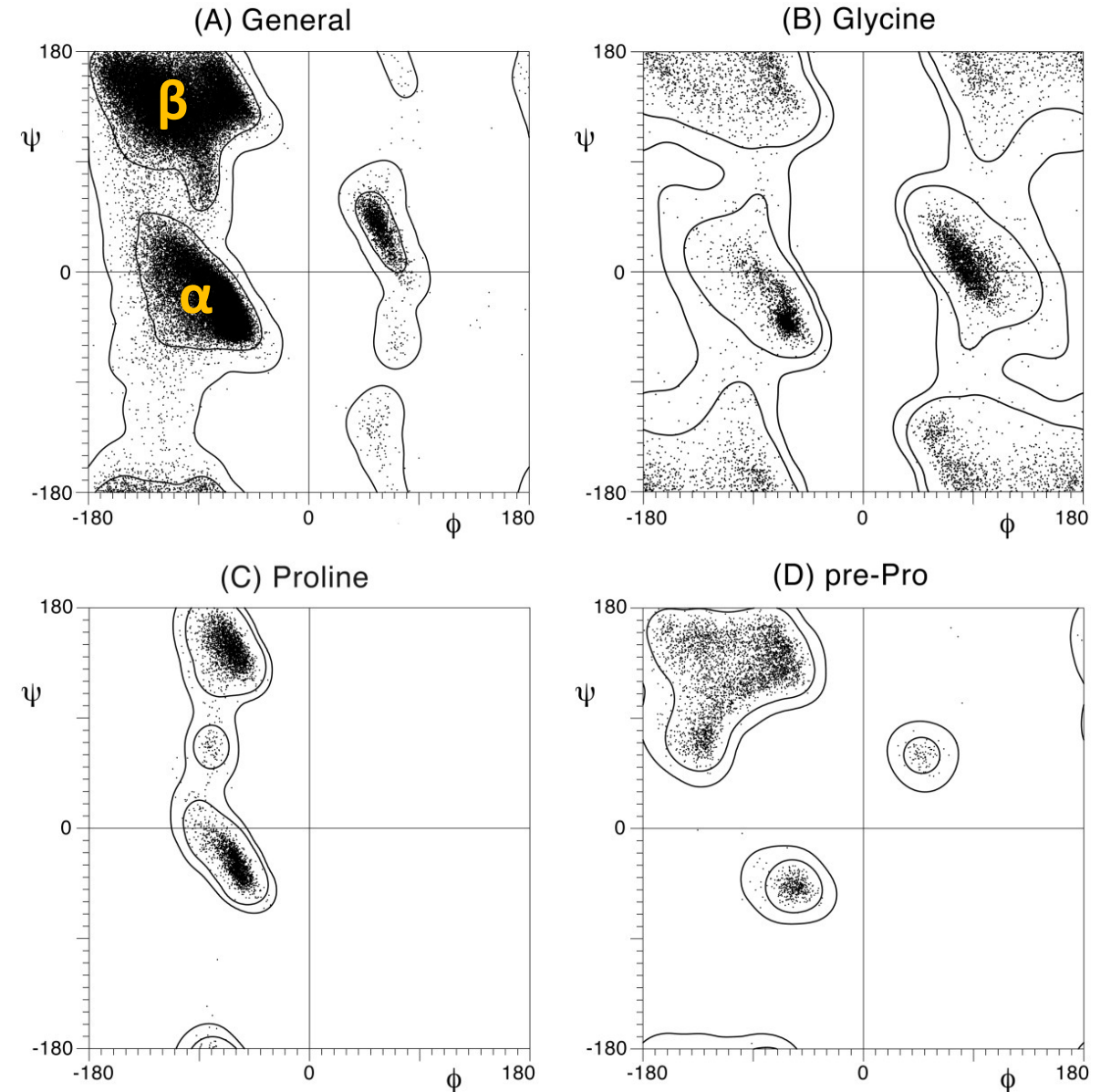
# Ramachandran plot

- Only certain  $\psi/\phi$  combinations are allowed
  - C $\beta$  atoms would bump into each other
  - Different for glycine (no C $\beta$ ) and proline [C $\delta$ -N(amide) bond]
- Some areas of the Ramachandran plot are highly populated



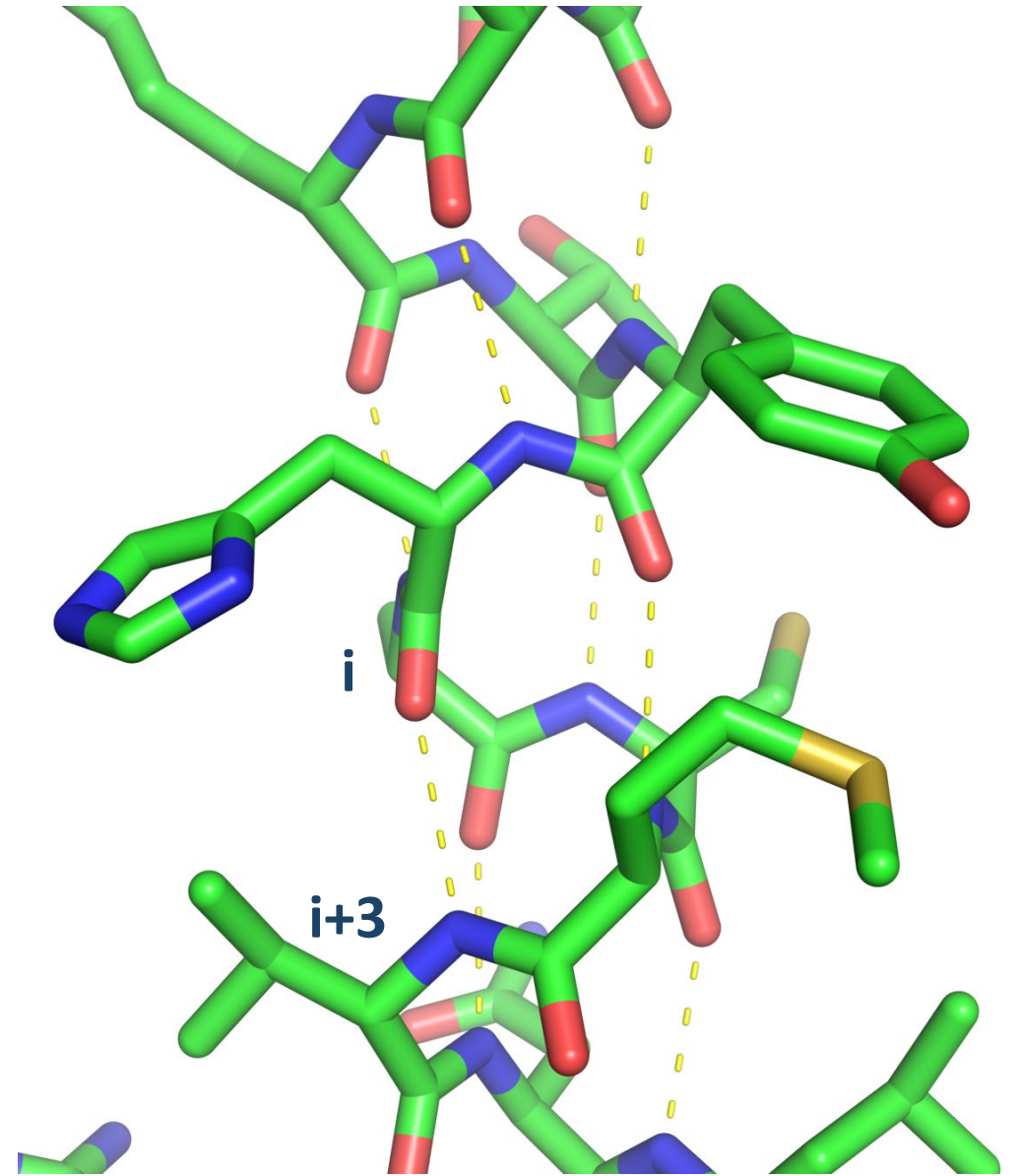
# Ramachandran plot

- Only certain  $\psi/\phi$  combinations are allowed
  - C $\beta$  atoms would bump into each other
  - Different for glycine (no C $\beta$ ) and proline [C $\delta$ -N(amide) bond]
- Many residues occupy certain regions of the Ramachandran plot:
  - $\alpha$ -helix
  - $\beta$ -sheet



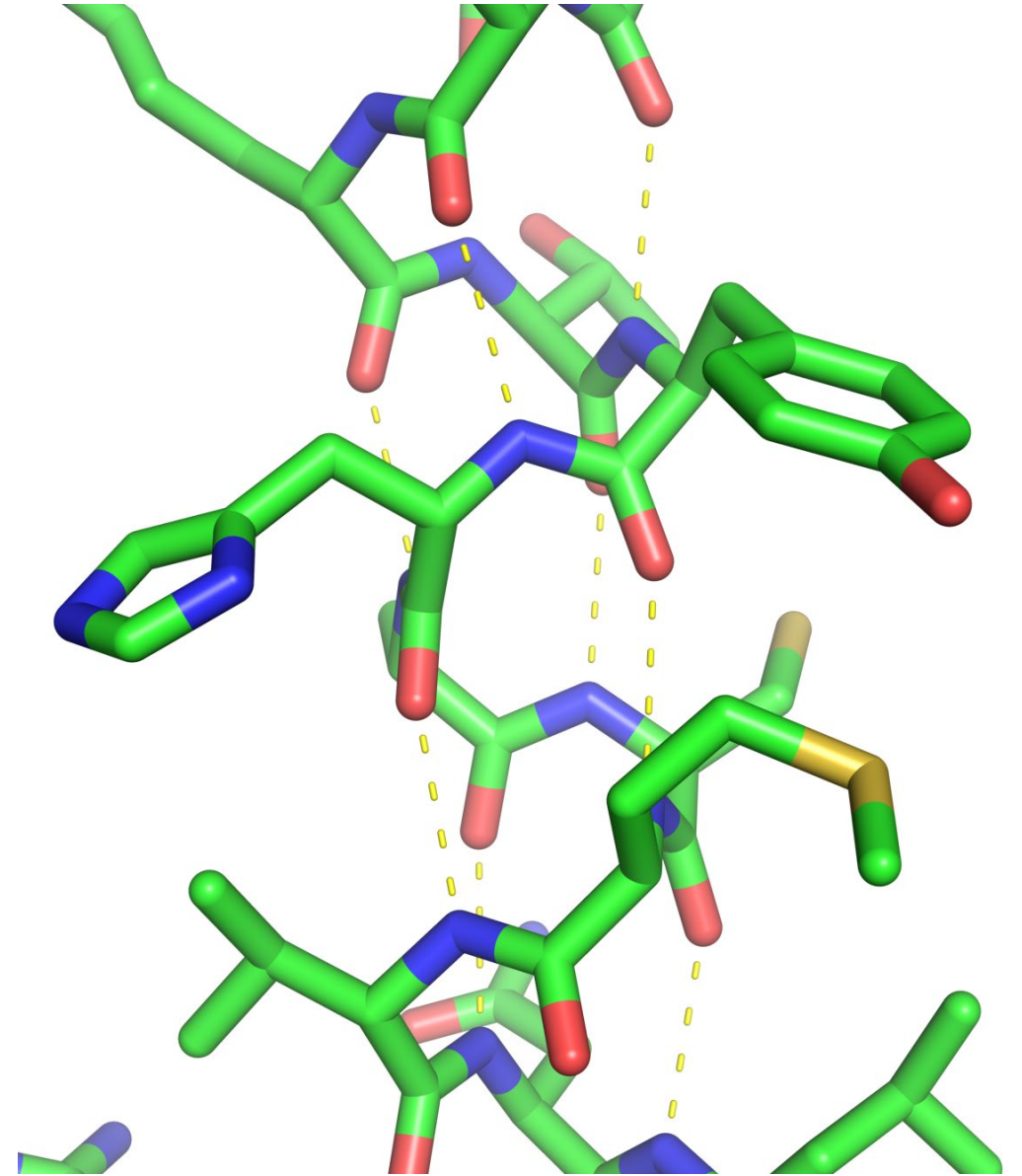
# $\alpha$ -helix

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



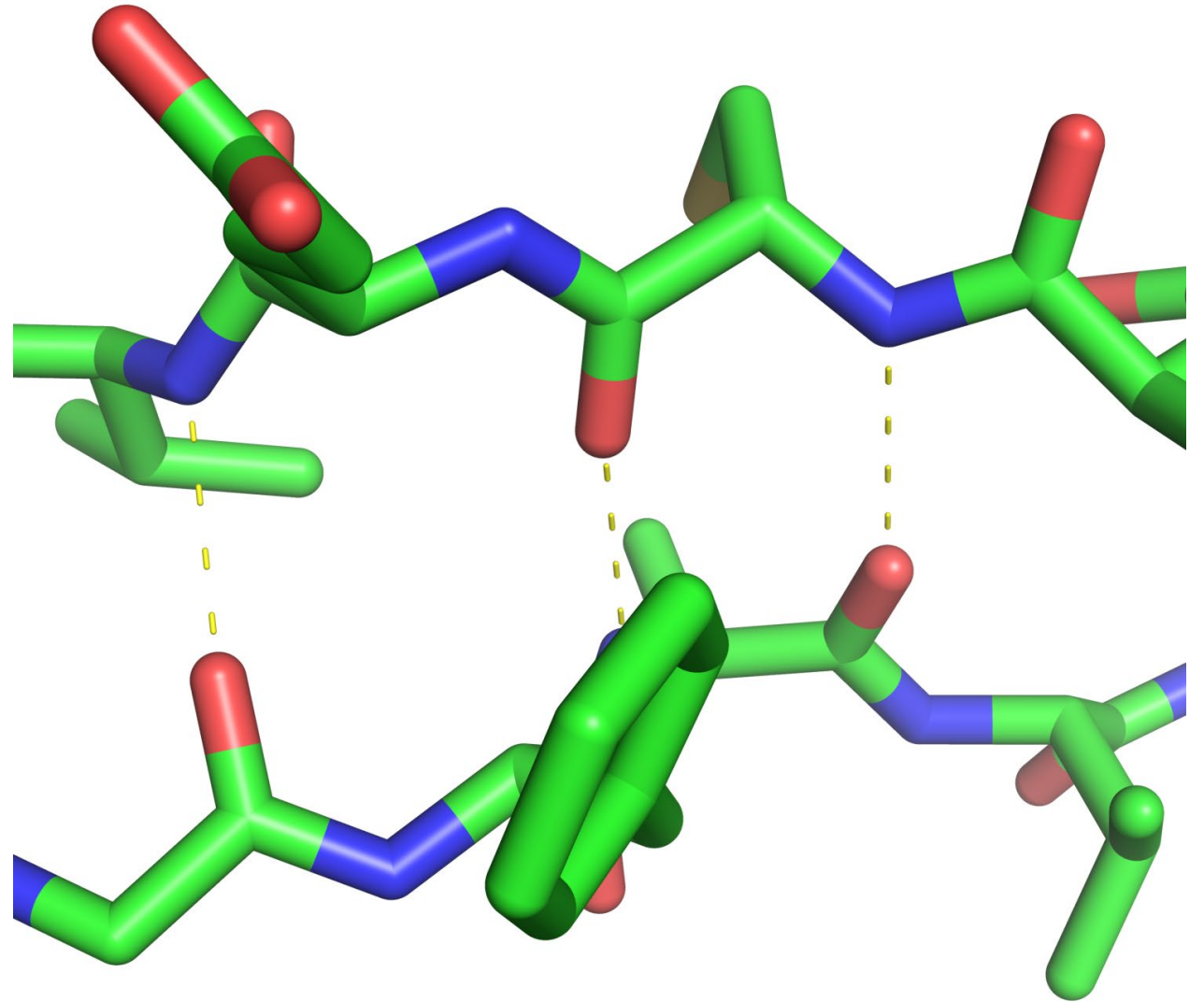
# $\alpha$ -helix

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



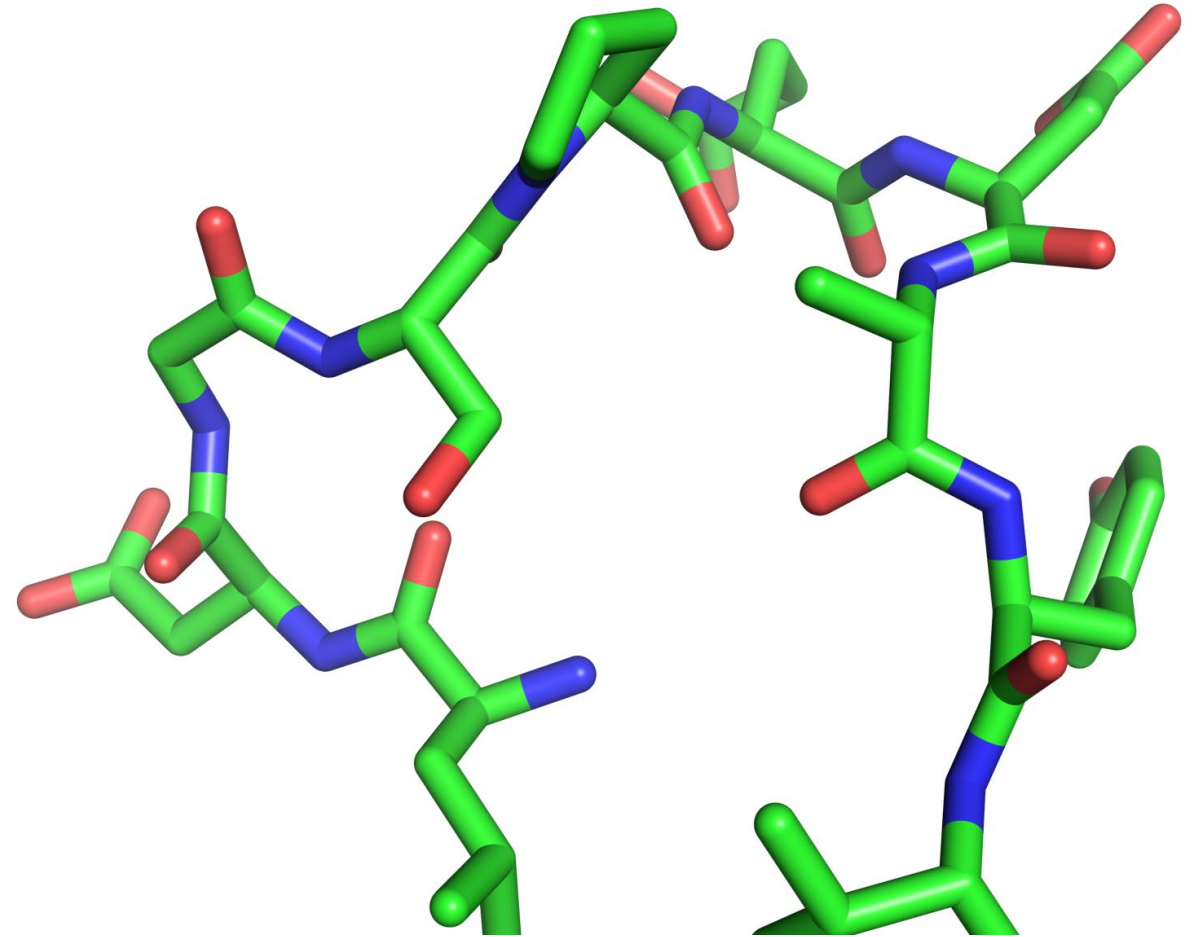
# $\beta$ -sheet

- 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

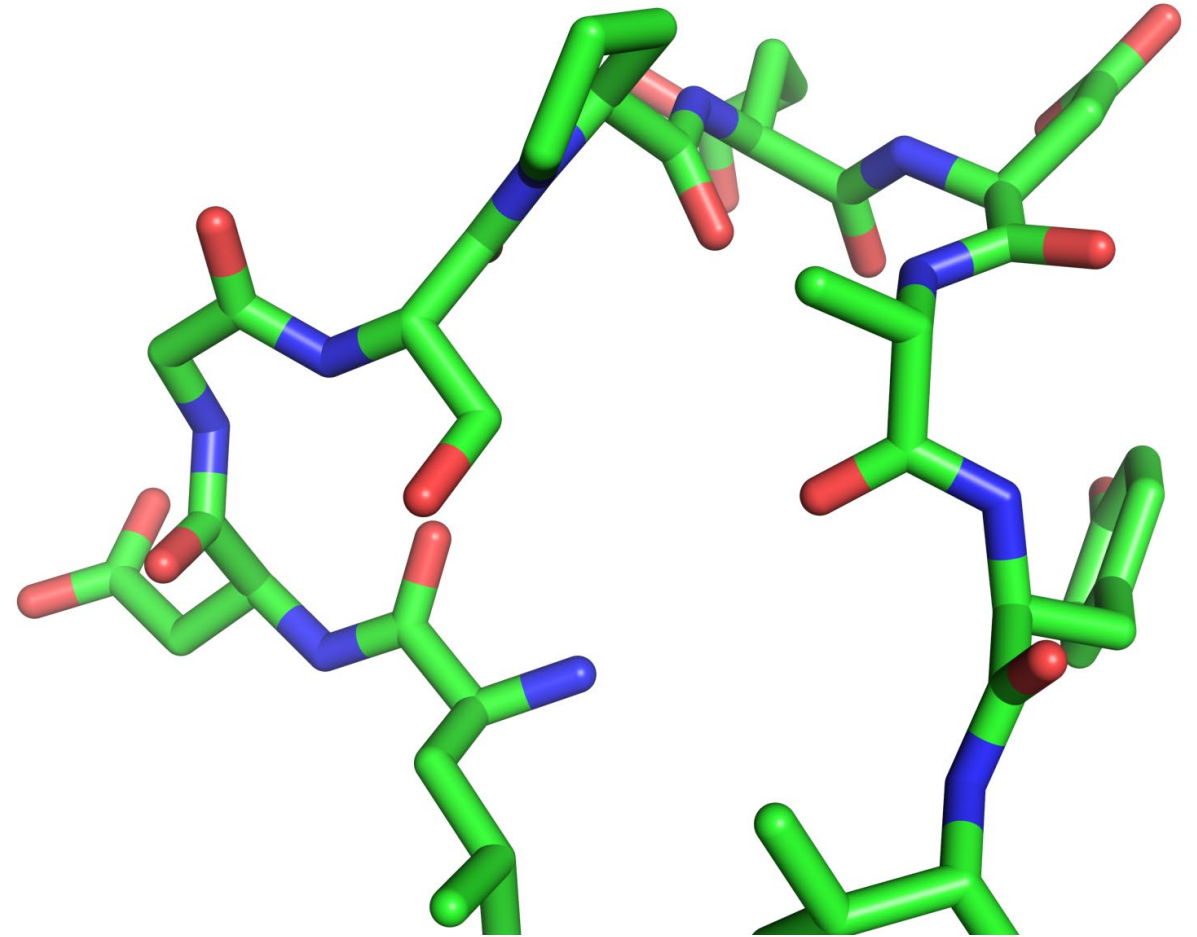
- Loops (between secondary structural elements)
- Coil (extended regions with no secondary structure)





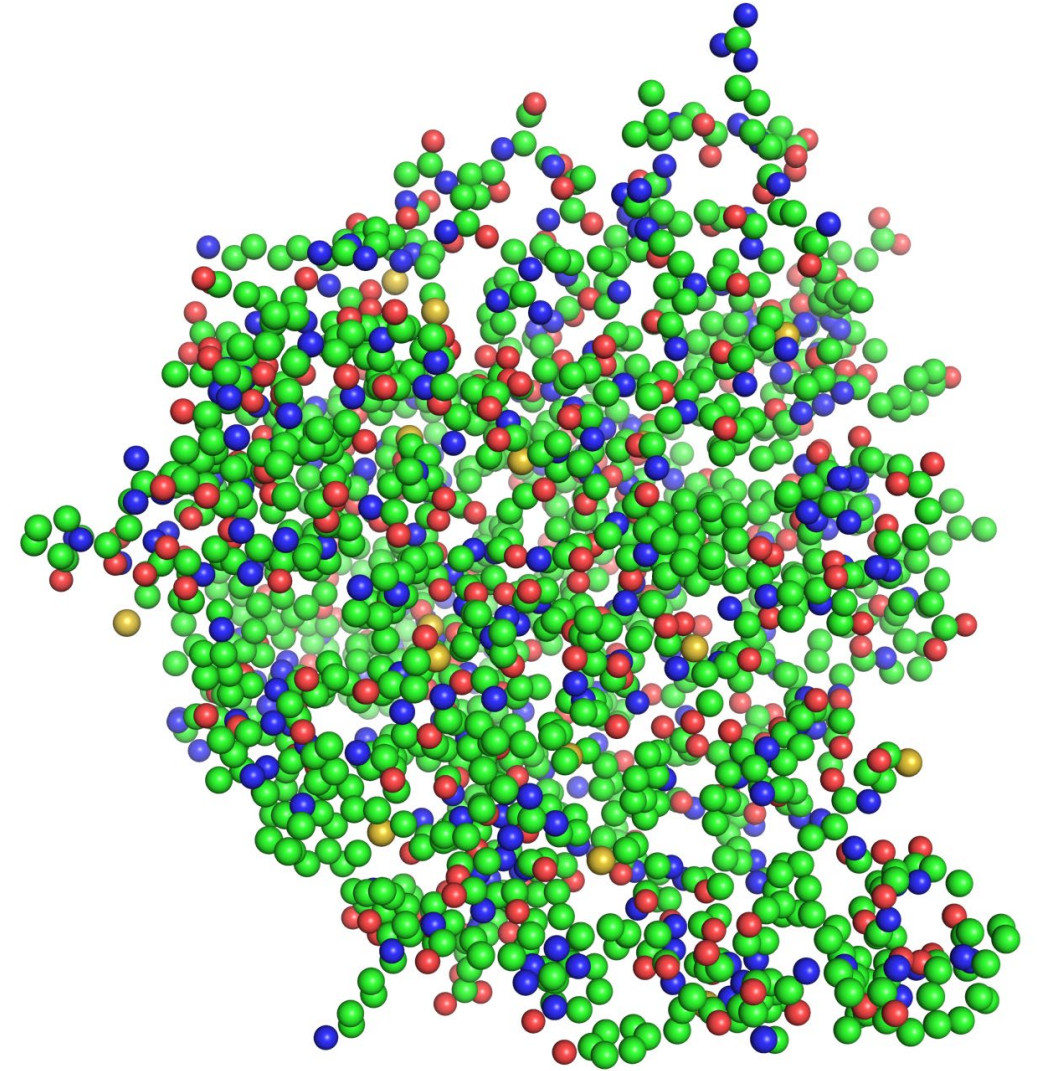
# Not all amino acids are in helices or sheets

- 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)



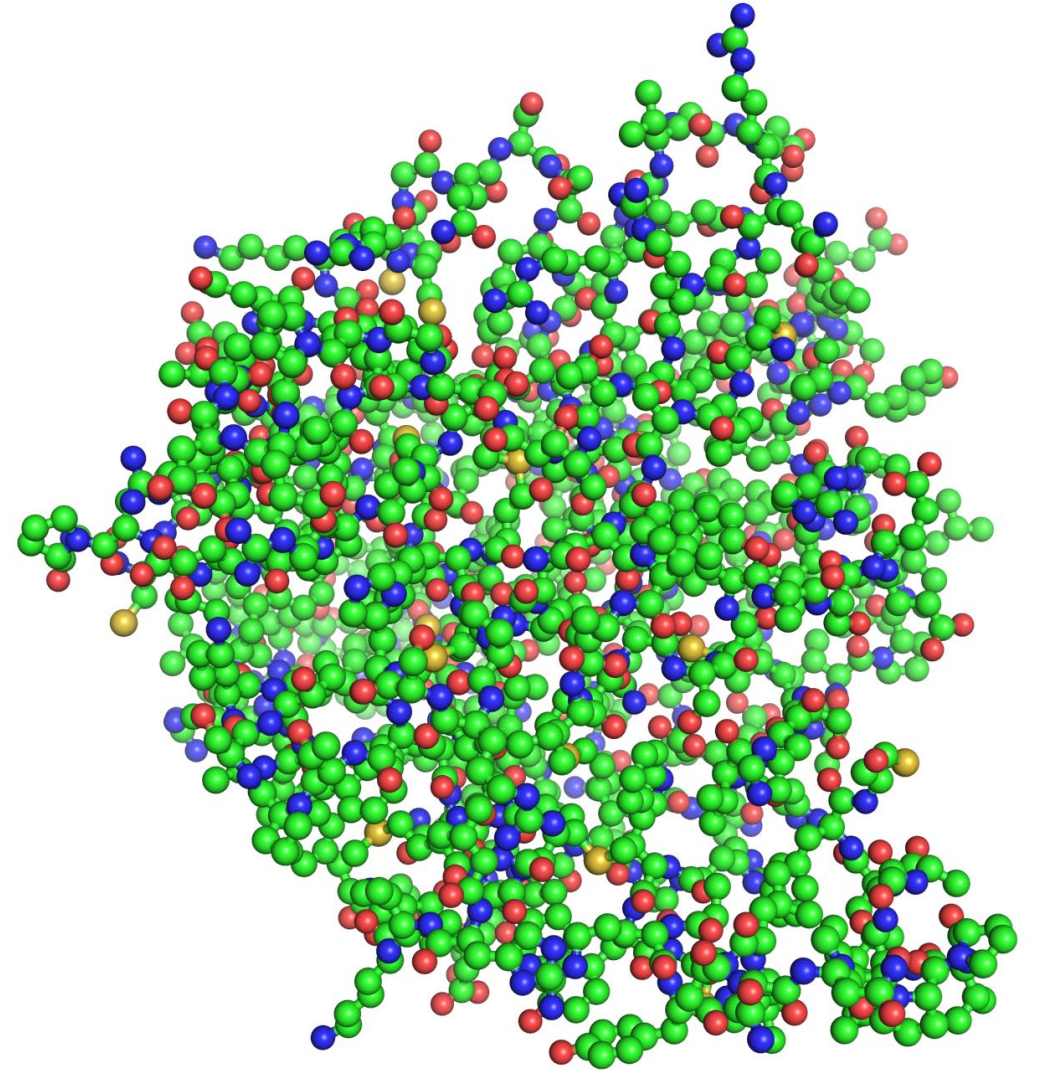
# Displaying secondary structure

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



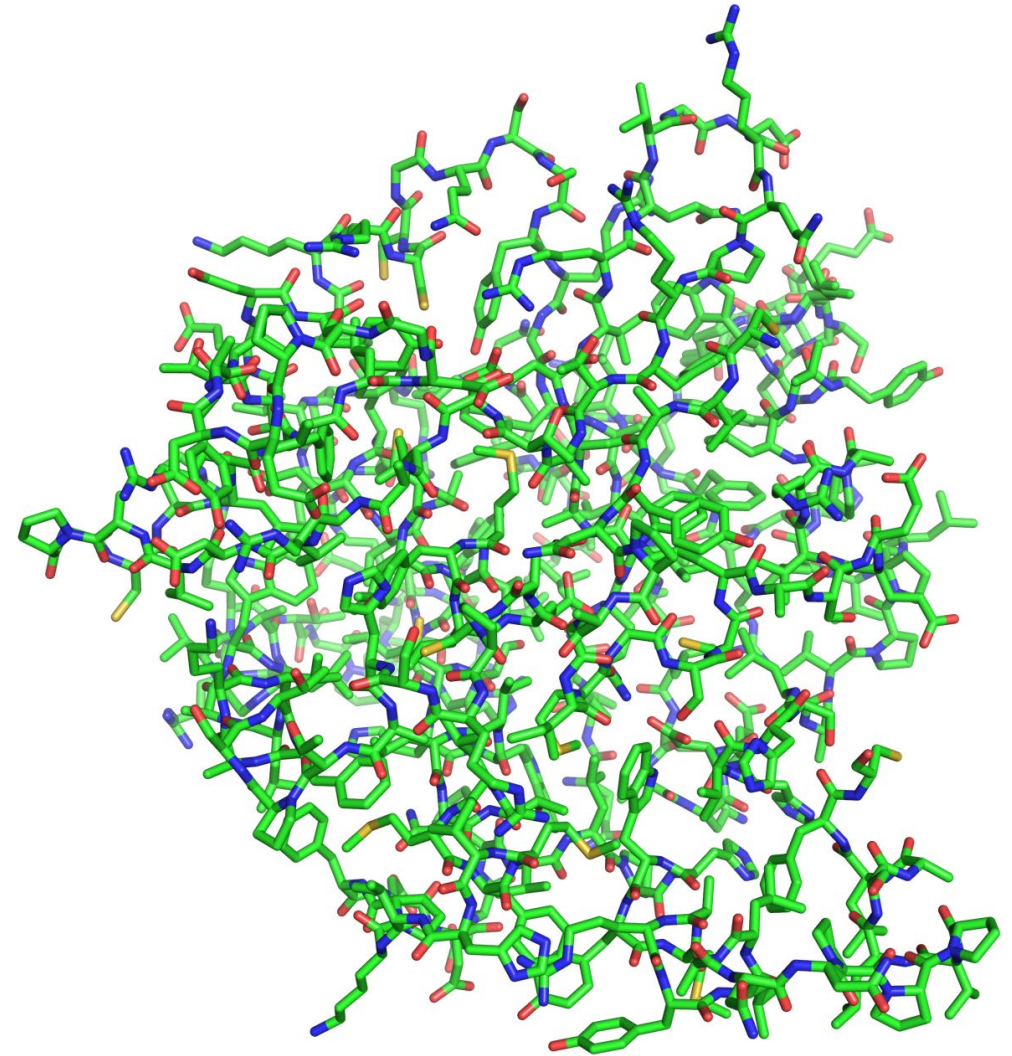
# Displaying secondary structure

- 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



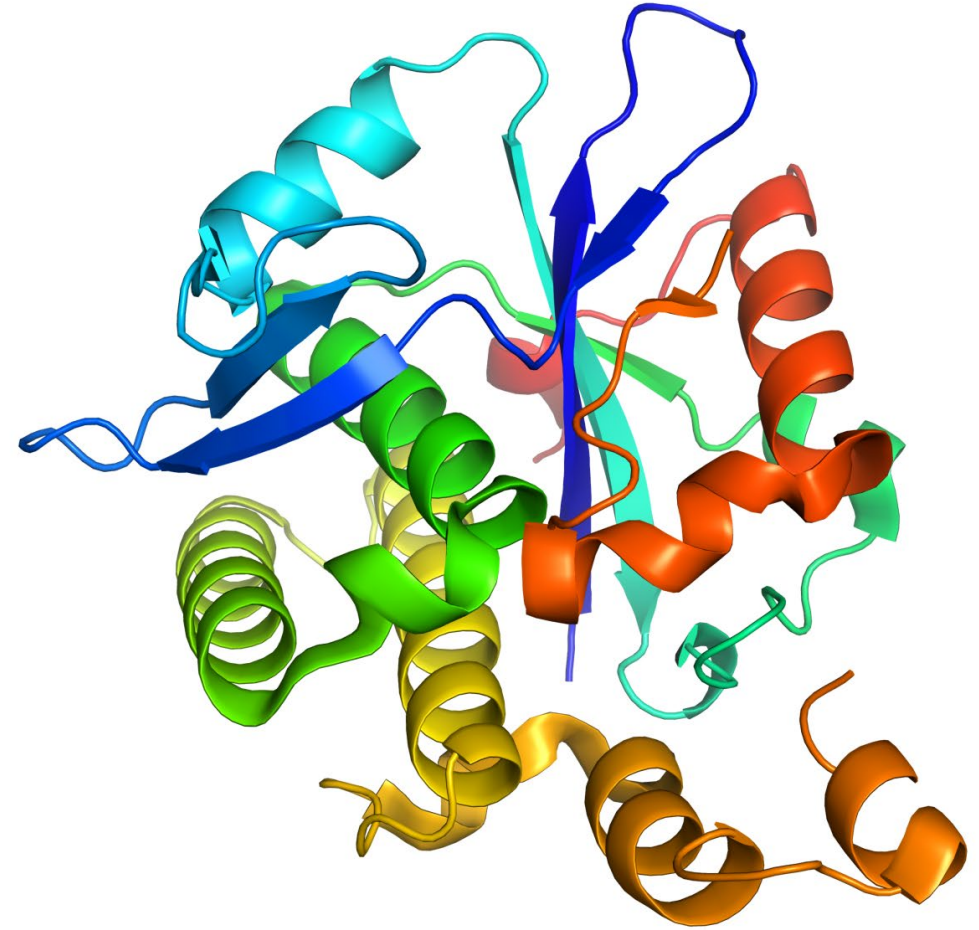
# Displaying secondary structure

- 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



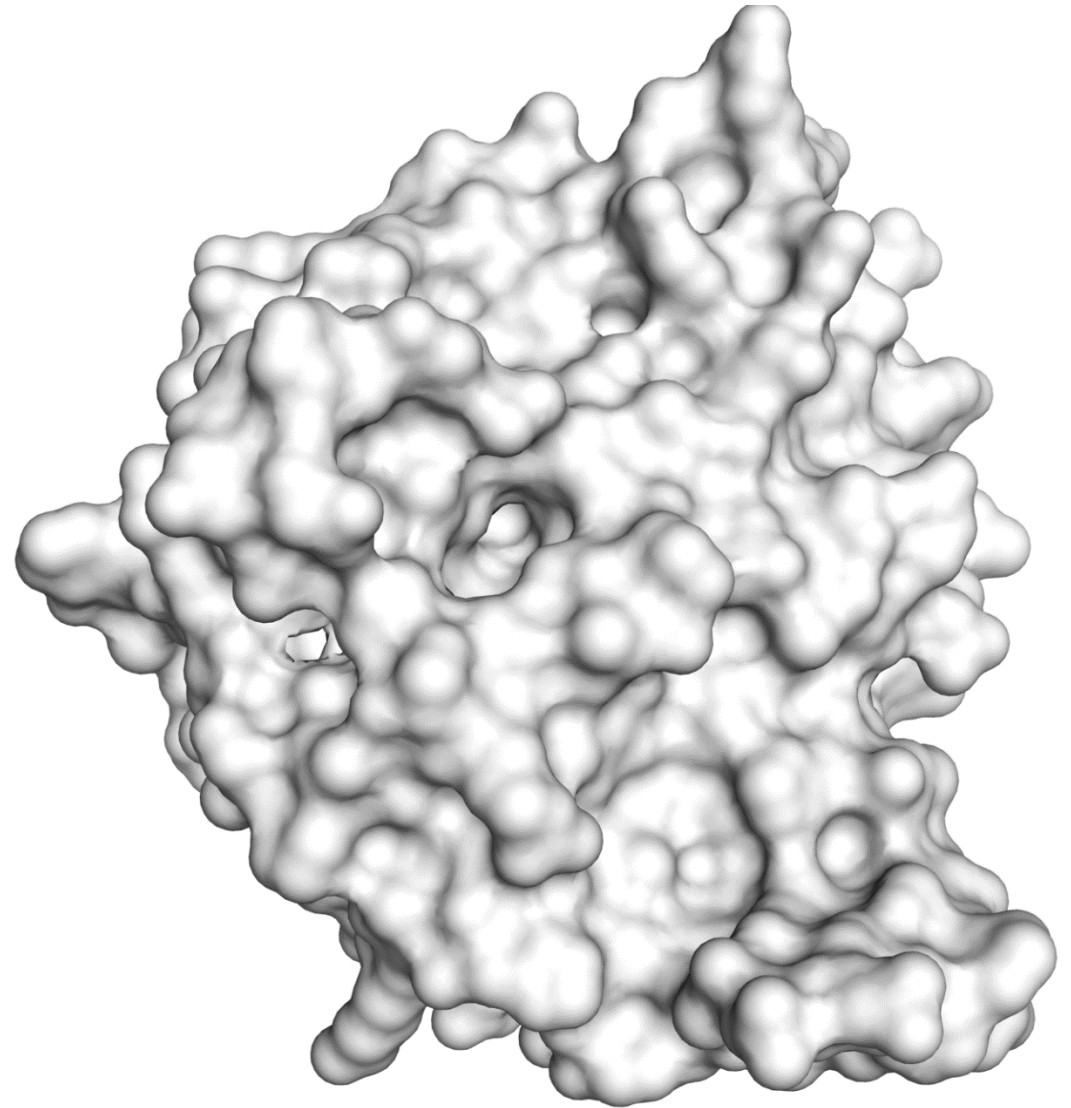
# Surfaces

- 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

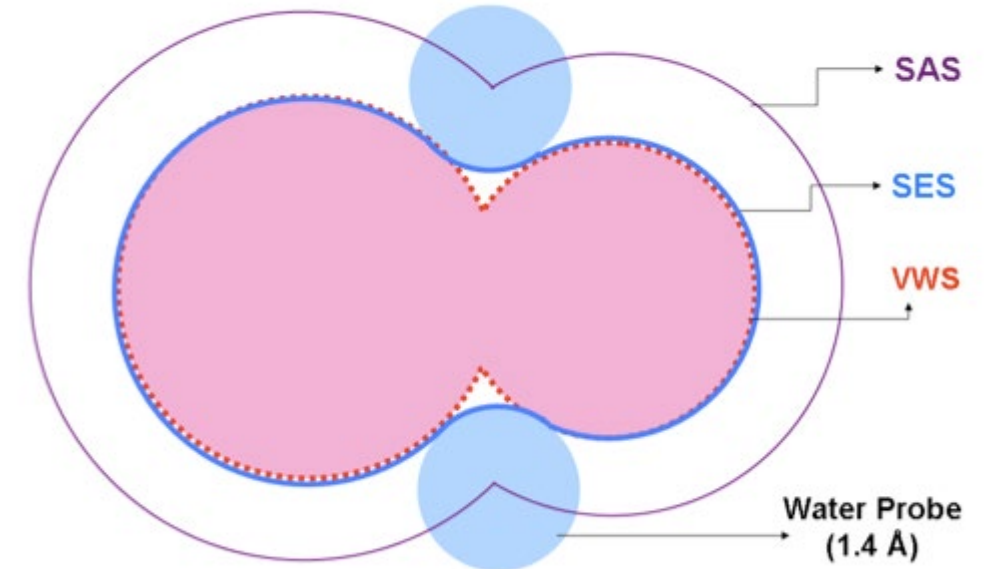
- Cartoons are convenient for following peptide direction
  - But proteins don't have big gaps
  - The interior of proteins is largely hydrophobic and inaccessible to water
- Can display the surface of proteins





# 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 Connolly surface)



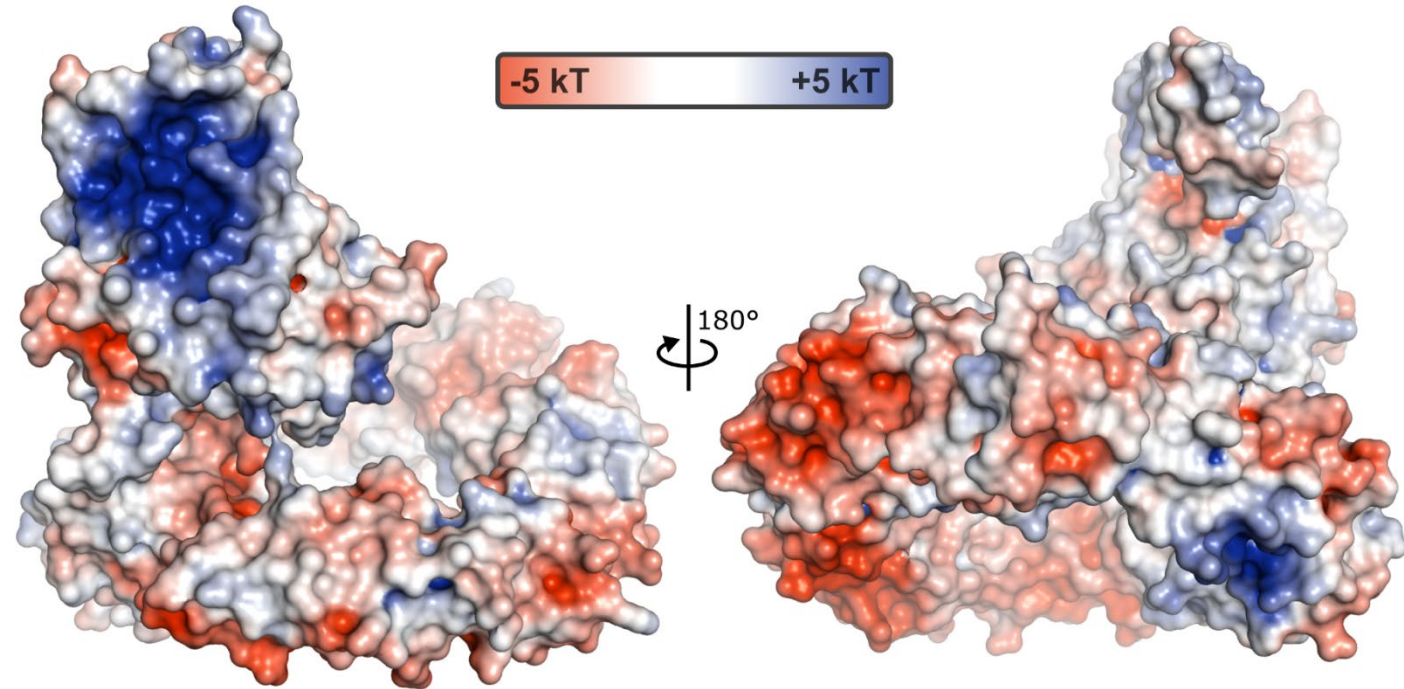
**SAS**: Solvent-accessible surface  
**SES**: Molecular surface (solvent-excluded surface)  
**VWS**: Van der Waals surface

# Surface representations



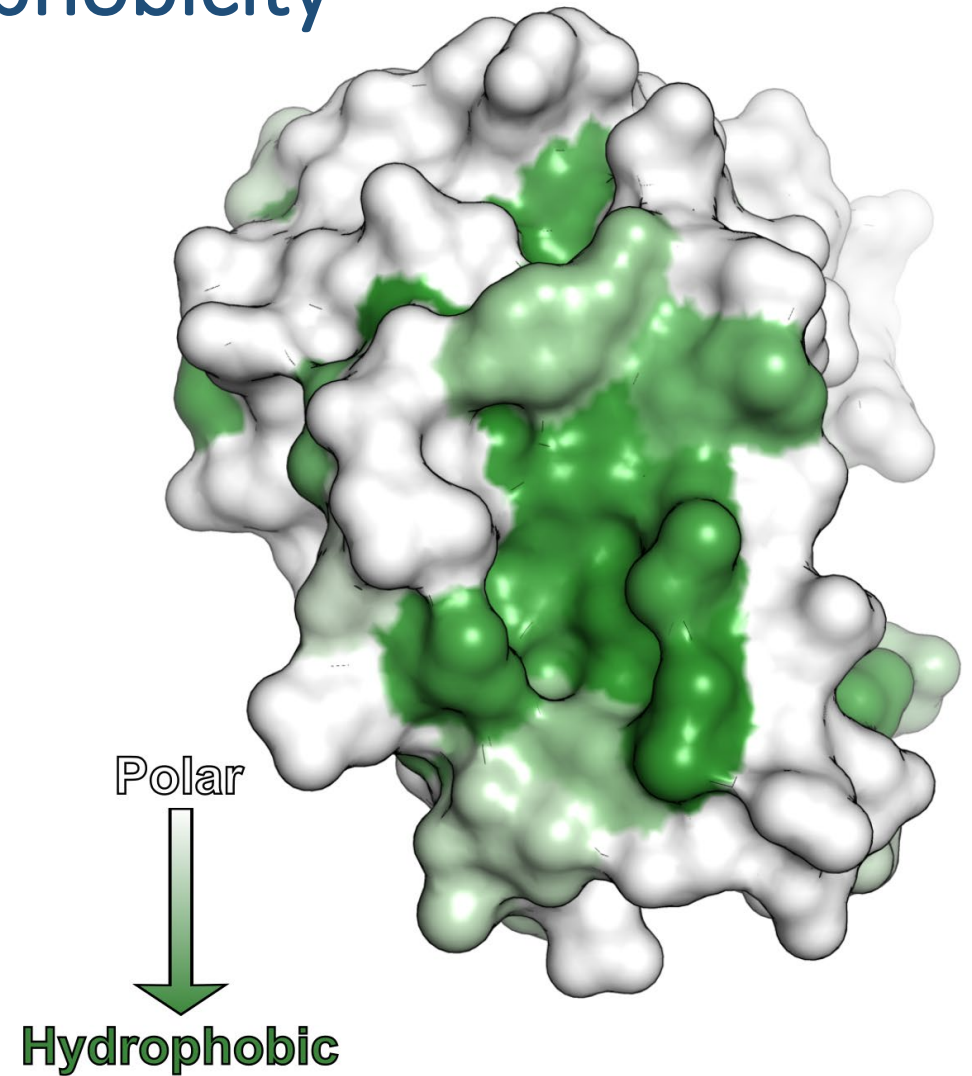
# Surface representations – Electrostatic potential

- 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



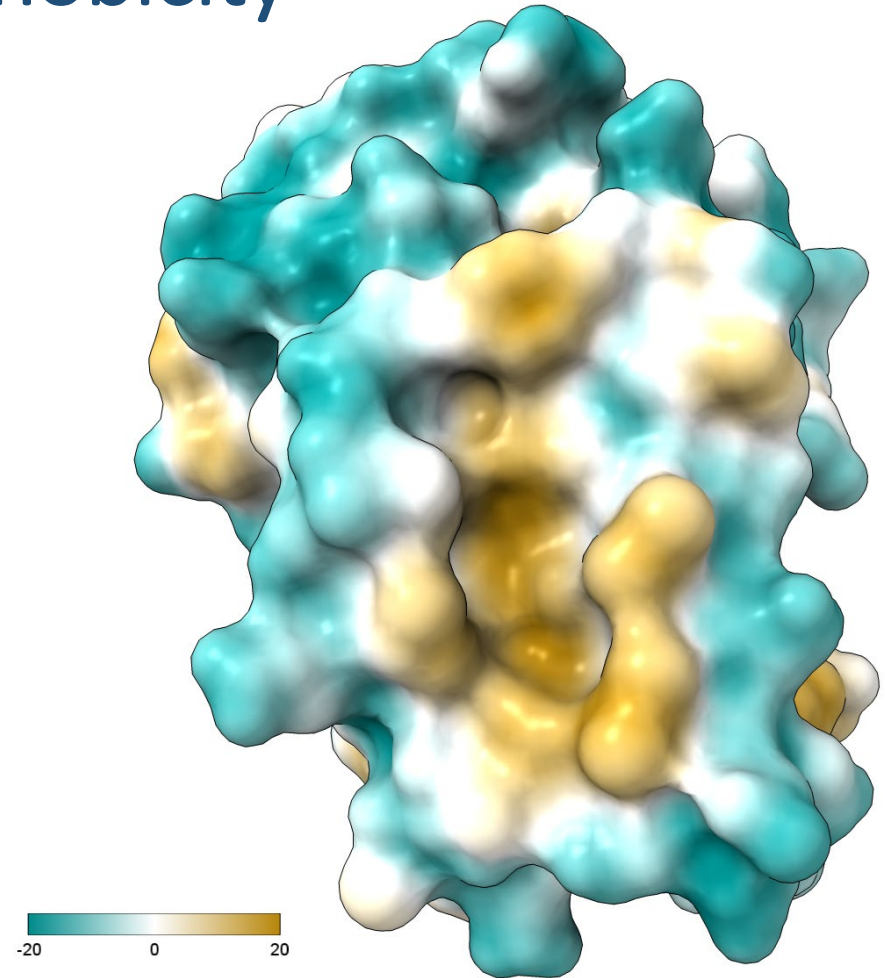
# Surface representations – Hydrophobicity

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



# 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

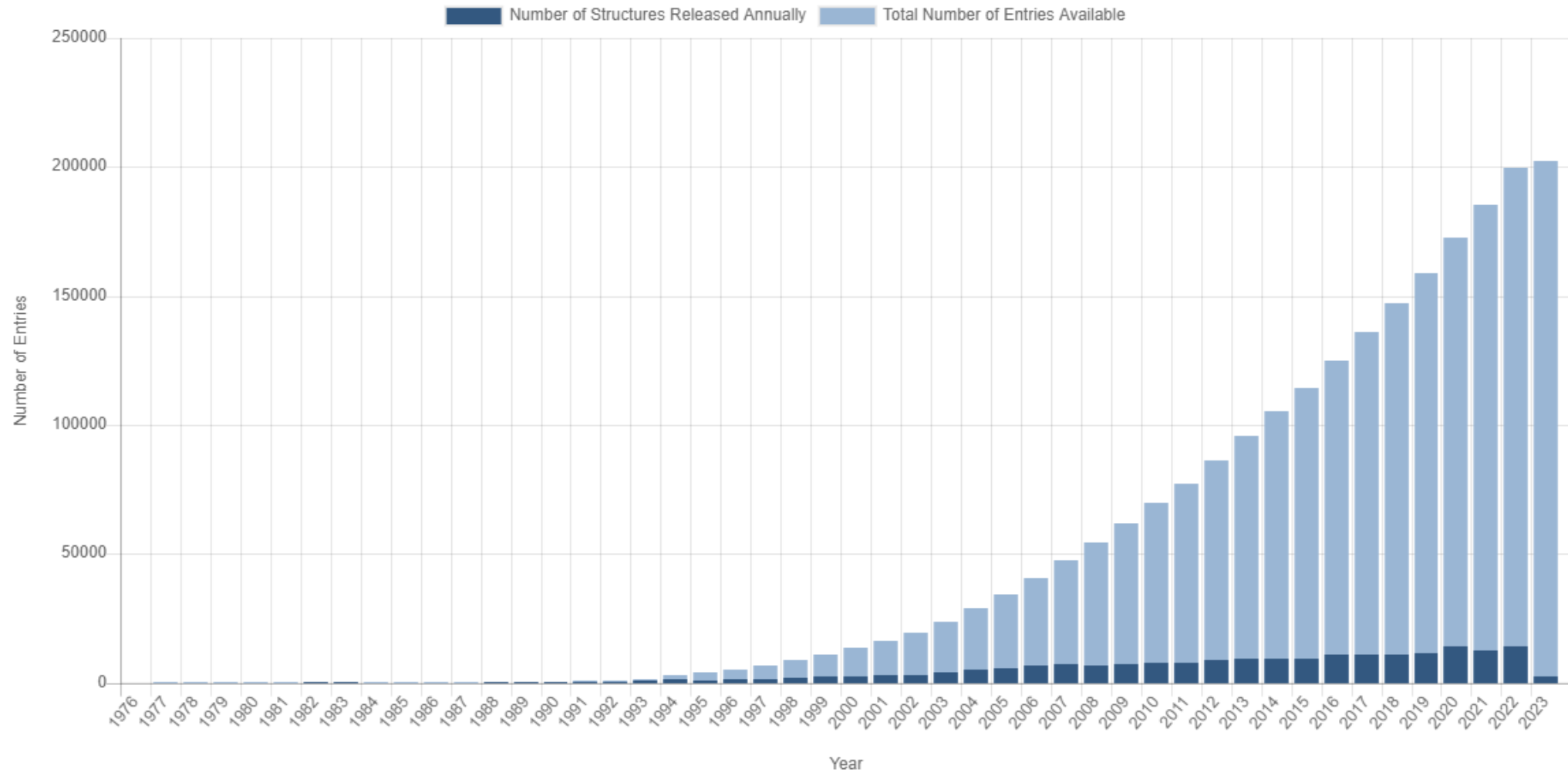


# Protein Data Bank (PDB)

- 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

The screenshot shows the RCSB PDB website homepage. At the top, there is a navigation bar with links for Deposit, Search, Visualize, Analyze, Download, Learn, About, Documentation, and Careers. Below this, the PDB logo is displayed alongside statistics: 202,292 Structures from the PDB and 1,068,577 Computed Structure Models (CSM). A search bar is present with the text 'Enter search term(s), Entry ID(s), or sequence'. Below the search bar, there are links for 'Advanced Search' and 'Browse Annotations'. A banner for 'New: More Computed Structure Models (CSM) available' is visible. The main content area features a 'Welcome' message, a 'Deposit' button, and a list of search options. A 'March Molecule of the Month' section highlights the Anaphase-Promoting Complex / Cyclosome. Below this, there are sections for 'Latest Entries' (showing entry 8DTH), 'Features & Highlights' (including news about PDB ID distribution changes and SDSC/SingAREN commitments), and 'News' (including a new coloring page and a paper published in PDB).

# Protein Data Bank (PDB)



# Accessing a protein structure

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

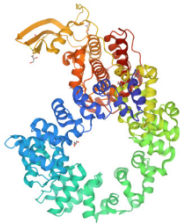
RCSB PDB PROTEIN DATA BANK 202,292 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Help

PDB-101 PDB EMDataResource NUCLEIC ACID DATABASE wwPDB Foundation PDB-Dev

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

Biological Assembly 1



## 7PHY

Vaccinia virus E2

PDB DOI: 10.2210/pdb7PHY/pdb

Classification: VIRAL PROTEIN

Organism(s): Vaccinia virus WR

Expression System: Homo sapiens

Mutation(s): No

Deposited: 2021-08-19 Released: 2021-09-01

Deposition Author(s): Gao, W.N.D., Gao, C., Graham, S.C.

Funding Organization(s): Wellcome Trust

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.30 Å

R-Value Free: 0.237

R-Value Work: 0.192

R-Value Observed: 0.194

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		17.1%

This is version 1.1 of the entry. See complete history.

Literature Download Primary Citation

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

Gao, W.N.D., Gao, C., Deane, J.E., Carpentier, D.C.J., Smith, G.L., Graham, S.C. (2022) J Gen Virol 103

PubMed: 35020582 Search on PubMed Search on PubMed Central

DOI: 10.1099/jgv.0.001716

Primary Citation of Related Structures: 7PLV

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: light scattering

Macromolecule Content

- Total Structure Weight: 88.15 kDa
- Atom Count: 6,295
- Modelled Residue Count: 732
- Deposited Residue Count: 750
- Unique protein chains: 1





# Accessing a protein structure

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

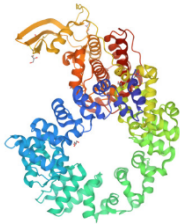
RCSB PDB PROTEIN DATA BANK 202,292 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Help

PDB-101 PDB EMDataResource NUCLEIC ACID DATABASE wwPDB Foundation PDB-Dev

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

Biological Assembly 1



## 7PHY

Vaccinia virus E2

PDB DOI: 10.2210/pdb7PHY/pdb

Classification: VIRAL PROTEIN

Organism(s): Vaccinia virus WR

Expression System: Homo sapiens

Mutation(s): No

Deposited: 2021-08-19 Released: 2021-09-01

Deposition Author(s): Gao, W.N.D., Gao, C., Graham, S.C.

Funding Organization(s): Wellcome Trust

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.30 Å

R-Value Free: 0.237

R-Value Work: 0.192

R-Value Observed: 0.194

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		17.1%

This is version 1.1 of the entry. See complete history.

Literature Download Primary Citation

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

Gao, W.N.D., Gao, C., Deane, J.E., Carpentier, D.C.J., Smith, G.L., Graham, S.C. (2022) J Gen Virol 103

PubMed: 35020582 Search on PubMed Search on PubMed Central

DOI: 10.1099/jgv.0.001716

Primary Citation of Related Structures: 7PLV

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: light scattering

Macromolecule Content

- Total Structure Weight: 88.15 kDa
- Atom Count: 6,295
- Modelled Residue Count: 732
- Deposited Residue Count: 750
- Unique protein chains: 1



# Accessing a protein structure

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

202,292 Structures from the PDB  
1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Help

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

Biological Assembly 1

## 7PHY

Vaccinia virus E2  
PDB DOI: 10.2210/pdb7PHY/pdb  
Classification: VIRAL PROTEIN  
Organism(s): Vaccinia virus WR  
Expression System: Homo sapiens  
Mutation(s): No

Deposited: 2021-08-19 Released: 2021-09-01  
Deposition Author(s): Gao, W.N.D., Gao, C., Graham, S.C.  
Funding Organization(s): Wellcome Trust

**Experimental Data Snapshot**  
Method: X-RAY DIFFRACTION  
Resolution: 2.30 Å  
R-Value Free: 0.237  
R-Value Work: 0.192  
R-Value Observed: 0.194

**wwPDB Validation**

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		17.1%

This is version 1.1 of the entry. See complete history.

**Literature** Download Primary Citation

The crystal structure of vaccinia virus protein E2 and perspectives on the prediction of novel viral protein folds.  
Gao, W.N.D., Gao, C., Deane, J.E., Carpentier, D.C.J., Smith, G.L., Graham, S.C.  
(2022) J Gen Virol 103  
PubMed: 35020582 Search on PubMed Search on PubMed Central  
DOI: 10.1099/jgv.0.001716  
Primary Citation of Related Structures:  
7PLV

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1  
Global Stoichiometry: Monomer - A1

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: light scattering

Macromolecule Content

- Total Structure Weight: 88.15 kDa
- Atom Count: 6,295
- Modelled Residue Count: 732
- Deposited Residue Count: 750
- Unique protein chains: 1

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)



# Accessing a protein structure

```
>7PHY_1|Chain A|Protein E2|Vaccinia virus WR (10254)
MISVTDIRRAFLDNECHTITKAFGYLHEDKAIALIKIGFHPTYLPKVLNMMVVEFVPEKLYLFKPRTVAPLDLISTITK
IINDIRRGKIDYYIPYVEDFLEDRTEDLGIYANIFFEDAIDITKLDITKTELEHISKYMMYYTTYIDHIVNIILONNYI
EENRNIVYKKNRVLCFDSENSAFKSLIKIDSIPGLKTYMMKDITYEKSNNIICVRFIPQESIHNERRIKLQLFDIA
VCVETILDNNQSFKSSKAAAHHHHHHHHH
```

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

# 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 ?
```

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



# 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,
```

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



# 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 0 5 0.39 0.24 0.13 0.36 0.13 0.36 0.13 0.36
1 1 1 0 0 6 0 12 0.41 0.27 -0.06 0.55 -0.06 0.55 -0.06 0.55
1 1 1 0 0 8 0 2 62.53 0.52 3914.12 65.37 3914.12 65.37 3914.12 65.37
1 1 1 0 0 10 0 16 35.36 0.26 1250.50 18.32 1250.50 18.32 1250.50 18.32
1 1 1 0 0 12 0 12 36.76 0.27 1351.74 19.85 1351.74 19.85 1351.74 19.85
1 1 1 0 0 14 0 13 15.72 0.13 247.08 4.20 247.08 4.20 247.08 4.20
1 1 1 0 0 16 0 5 50.67 0.43 2570.18 43.27 2570.18 43.27 2570.18 43.27
1 1 1 0 0 18 0 15 69.78 0.51 4875.60 70.64 4875.60 70.64 4875.60 70.64
1 1 1 0 0 20 0 13 4.00 0.23 16.14 1.81 16.14 1.81 16.14 1.81
1 1 1 0 0 22 0 10 19.34 0.17 374.32 6.41 374.32 6.41 374.32 6.41
1 1 1 0 0 24 0 16 11.47 0.20 131.84 4.66 131.84 4.66 131.84 4.66
1 1 1 0 0 26 0 5 10.59 0.15 112.24 3.21 112.24 3.21 112.24 3.21
1 1 1 0 0 28 0 1 5.73 0.26 33.06 3.00 33.06 3.00 33.06 3.00
```

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)



# Accessing a protein structure



Full wwPDB X-ray Structure Validation Report ⓘ

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)

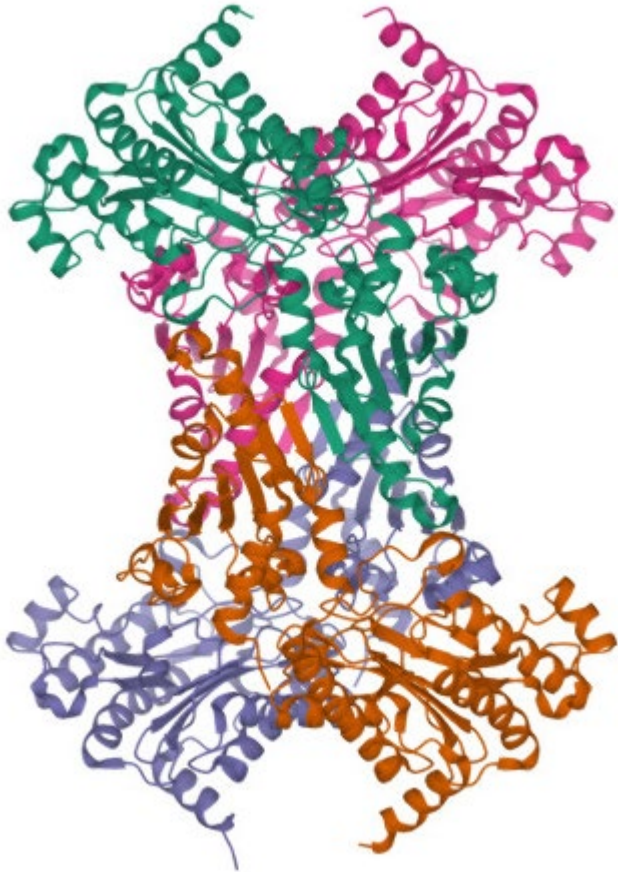
This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ symbol.

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

# Accessing a protein structure



Display Files ▾ Download Files ▾

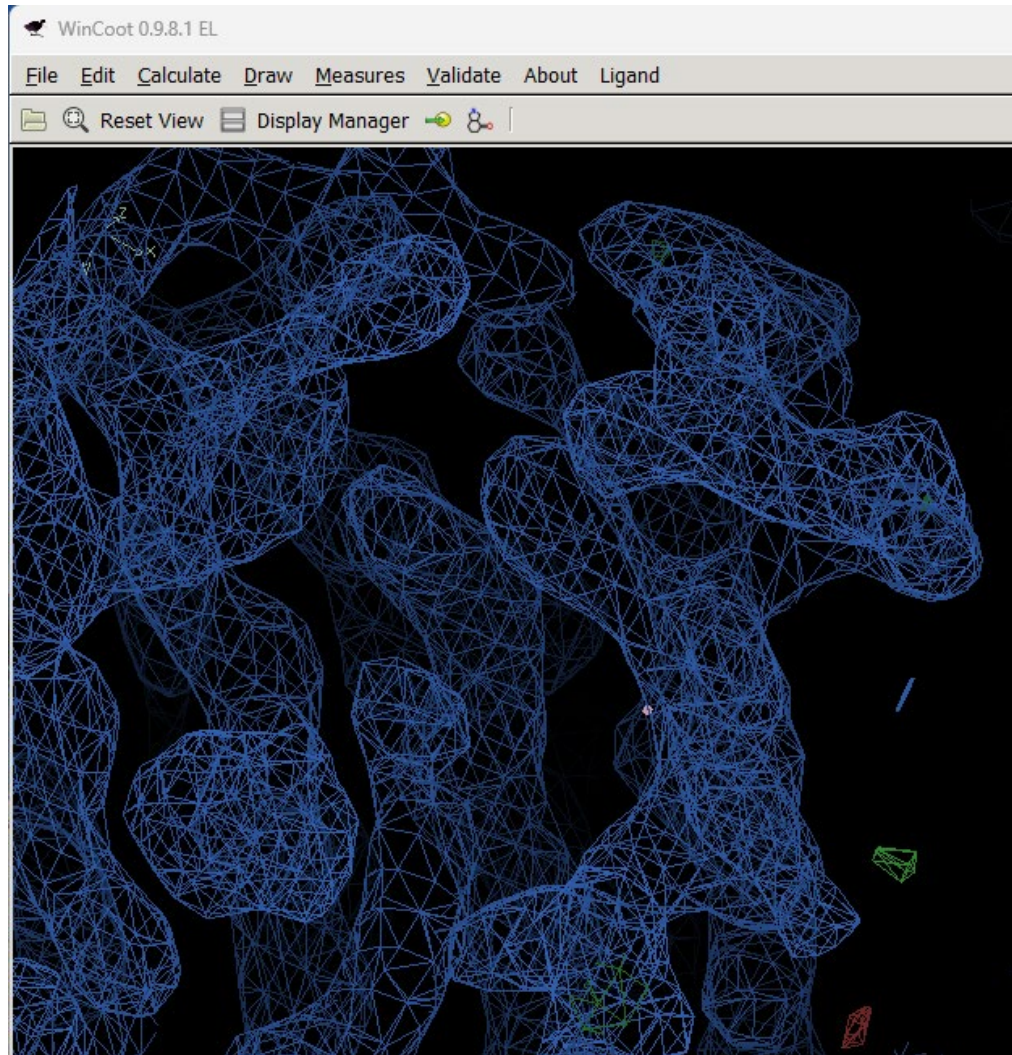
- FASTA Sequence
- PDBx/mmCIF Format  
PDBx/mmCIF Format (gz)
- PDB Format  
PDB Format (gz)
- PDBML/XML Format (gz)
- Structure Factors (CIF)  
Structure Factors (CIF - gz)
- Validation Full PDF  
Validation XML  
Validation CIF
- Biological Assembly 1 (CIF - gz) ⓘ
- Biological Assembly 1 (PDB - gz)
- 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



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
- Maps



# Accessing a protein structure

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

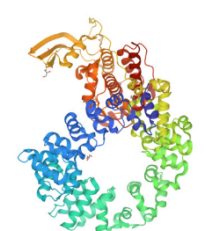
RCSB PDB PROTEIN DATA BANK 202,292 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Help

PDB-101 PDB EMDataResource NUCLEIC ACID DATABASE wwPDB Foundation PDB-Dev

Structure Summary **3D View** Annotations Experiment Sequence Genome Versions

Biological Assembly 1



## 7PHY

Vaccinia virus E2

PDB DOI: 10.2210/pdb7PHY/pdb

Classification: VIRAL PROTEIN

Organism(s): Vaccinia virus WR

Expression System: Homo sapiens

Mutation(s): No

Deposited: 2021-08-19 Released: 2021-09-01

Deposition Author(s): Gao, W.N.D., Gao, C., Graham, S.C.

Funding Organization(s): Wellcome Trust

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.30 Å

R-Value Free: 0.237

R-Value Work: 0.192

R-Value Observed: 0.194

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		17.1%

This is version 1.1 of the entry. See complete history.

Literature Download Primary Citation

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

Gao, W.N.D., Gao, C., Deane, J.E., Carpentier, D.C.J., Smith, G.L., Graham, S.C. (2022) J Gen Virol 103

PubMed: 35020582 Search on PubMed Search on PubMed Central

DOI: 10.1099/jgv.0.001716

Primary Citation of Related Structures: 7PLV

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1

Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: light scattering

Macromolecule Content

- Total Structure Weight: 88.15 kDa
- Atom Count: 6,295
- Modelled Residue Count: 732
- Deposited Residue Count: 750
- Unique protein chains: 1



# Accessing a protein structure

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

RCSB PDB PROTEIN DATA BANK 202,292 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Help

Advanced Search | Browse Annotations

PDB-101 PDB EMDataResource Nucleic Acid Database wwPDB Foundation PDB-Dev

Structure Summary **3D View** Annotations Experiment Sequence Genome Versions

Biological Assembly 1

## 7PHY

Vaccinia virus E2

PDB DOI: 10.2210/pdb7PHY/pdb

Classification: VIRAL PROTEIN  
Organism(s): Vaccinia virus WR  
Expression System: Homo sapiens  
Mutation(s): No

Deposited: 2021-08-19 Released: 2021-09-01  
Deposition Author(s): Gao, W.N.D., Gao, C., Graham, S.C.  
Funding Organization(s): Wellcome Trust

Experimental Data Snapshot

Method: X-RAY DIFFRACTION  
Resolution: 2.30 Å  
R-Value Free: 0.237  
R-Value Work: 0.192  
R-Value Observed: 0.194

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		17.1%

This is version 1.1 of the entry. See complete history.

Literature [Download Primary Citation](#)

The crystal structure of vaccinia virus protein E2 and perspectives on the prediction of novel viral protein folds.  
[Gao, W.N.D., Gao, C., Deane, J.E., Carpentier, D.C.J., Smith, G.L., Graham, S.C. \(2022\) J Gen Virol 103](#)  
PubMed: [35020582](#) [Search on PubMed](#) [Search on PubMed Central](#)  
DOI: 10.1099/igv.0.001716  
Primary Citation of Related Structures: 7PHY

3D View: Structure | 1D-3D View | Electron Density | Validation Report | Ligand Interaction

Global Symmetry: Asymmetric - C1  
Global Stoichiometry: Monomer - A1

Find Similar Assemblies

Biological assembly 1 assigned by authors and generated by PISA (software)

Biological Assembly Evidence: light scattering

Macromolecule Content

- Total Structure Weight: 88.15 kDa
- Atom Count: 6,295
- Modelled Residue Count: 732
- Deposited Residue Count: 750
- Unique protein chains: 1

Structure Summary **3D View** Annotations Experiment Sequence Genome Versions

## 7PHY

Vaccinia virus E2

Display Files Download Files

Sequence of 7PHY | Vaccini... Chain 1: Protein E2 A

```
AME ISVDIRRAFLDNECHTITKAFGLYHEDKALIKIGFPTLYLKVLYNWFVPEKLVLPKPRVAPLDLSTIITKLRNVDFKASHINHYKNSILITGDKSLIVKCM  
1 21 41 61 81 101 121 141 161 181 201 221  
PYMIISDDDIRFIREQFVGNSTSEYILSFNKESTVMSYQSENEIVTIINRDFMYPETVEHVDLSDFLKTMMLDRVGIWVINSGLIDELCPALIEILMAWRPRDAIF  
231 241 261 281 301 321 341 361 381 401 421 441 461 481 501  
FLDIVNRNQLTDSVKVYIINDIRRGKLDYYPVVEDFLDRTEDLGIYANIFFEDAIDITKLDITKLEHISKIMNYTTYIDHIVNITLQNNVIDILASIIDVQDVLIT  
261 281 301 321 341 361 381 401 421 441 461 481 501
```

Structure

7PHY | Vaccinia virus E2

Type Assembly

Asm Id 1: Author And Software ...

Dynamic Bonds X Off

Nothing Focused

Measurements

Structure Motif Search

Components 7PHY

Preset + Add

Polymer Cartoon

Ligand Ball & Stick

Non-standard Ball & Stick

Water Ball & Stick

Unit Cell P 21 21 21

Density

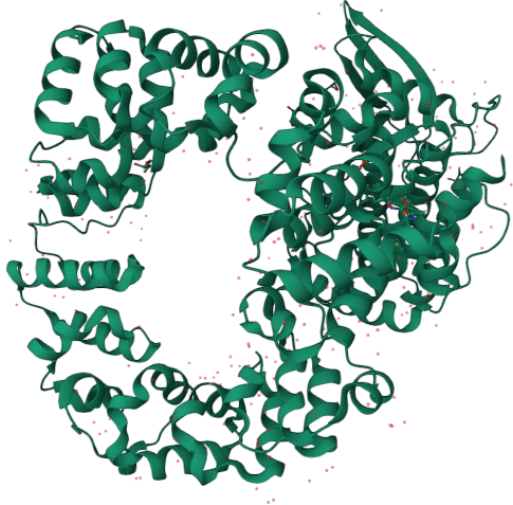
Quality Assessment

Assembly Symmetry

Export Models

Export Animation

Export Geometry



Welcome RCSB PDB Mol\* Viewer 2.5.10 [03/02/2023, 22:45:37]

# Accessing a protein structure

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

RCSB PDB PROTEIN DATA BANK 202,292 Structures from the PDB 1,068,577 Computed Structure Models (CSM)

3D Structures Enter search term(s), Entry ID(s), or sequence Include CSM Help

Structure Summary 3D View Annotations Experiment Sequence Genome Versions

Biological Assembly 1

## 7PHY

Vaccinia virus E2

PDB DOI: 10.2210/pdb7PHY/pdb

Classification: VIRAL PROTEIN

Organism(s): Vaccinia virus WR

Expression System: Homo sapiens

Mutation(s): No

Deposited: 2021-08-19 Released: 2021-09-01

Deposition Author(s): Gao, W.N.D., Gao, C., Graham, S.C.

Funding Organization(s): Wellcome Trust

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.30 Å

R-Value Free: 0.237

R-Value Work: 0.192

R-Value Observed: 0.194

wwPDB Validation

Metric	Percentile Ranks	Value
Rfree		0.234
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.4%
RSRZ outliers		17.1%

This is version 1.1 of the entry. See complete history.

Literature

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

Gao, W.N.D., Gao, C., Deane, J.E., Carpentier, D.C.J., Smith, G.L., Graham, S.C. (2022) J Gen Virol 103

PubMed: 35020582 Search on PubMed Search on PubMed Central

DOI: 10.1099/jgv.0.001716

Primary Citation of Related Structures:

JOURNAL OF  
GENERAL VIROLOGY

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



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

William N. D. Gao<sup>1</sup>, Chen Gao<sup>1</sup>, Janet E. Deane<sup>2</sup>, David C. J. Carpentier<sup>1</sup>, Geoffrey L. Smith<sup>1</sup> and Stephen C. Graham<sup>1\*</sup>

### 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

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search

## Q9BY43 · CHM4A\_HUMAN

Protein <sup>i</sup>	Charged multivesicular body protein 4a	Amino acids	222
Gene <sup>i</sup>	CHMP4A	Protein existence <sup>i</sup>	Evidence at protein level
Status <sup>i</sup>	UniProtKB reviewed (Swiss-Prot)	Annotation score <sup>i</sup>	5/5
Organism <sup>i</sup>	Homo sapiens (Human)		




Entry Feature viewer Publications External links History

BLAST Align Download Add Add a publication Entry feedback

### Function<sup>i</sup>

Probable core component of the endosomal sorting required for transport complex III (ESCRT-III) which is involved in multivesicular bodies (MVBs) formation and sorting of endosomal cargo proteins into MVBs. MVBs contain intraluminal vesicles (ILVs) that are generated by invagination and scission from the limiting membrane of the endosome and mostly are delivered to lysosomes enabling degradation of membrane proteins, such as stimulated growth factor receptors, lysosomal enzymes and lipids. The MVB pathway appears to require the sequential function of ESCRT-O, -I, -II and -III complexes. ESCRT-III proteins mostly dissociate from the invaginating membrane before the ILV is released. The ESCRT machinery also functions in topologically equivalent membrane fission events, such as the terminal stages of cytokinesis and the budding of enveloped viruses (HIV-1 and other lentiviruses). ESCRT-III proteins are believed to mediate the necessary vesicle extrusion and/or membrane fission activities, possibly in conjunction with the AAA ATPase VPS4. When overexpressed, membrane-assembled circular arrays of CHMP4A filaments can promote or stabilize negative curvature and outward budding. Via its interaction with PDCD6IP involved in HIV-1 p6- and p9-dependent virus release. CHMP4A/B/C are required for the exosomal release of SDCBP, CD63 and syndecan (PubMed:22660413). [6 Publications](#)

# Accessing protein structural models

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB  Advanced | List Search   

Function **Q9BY43 · CHM4A\_HUMAN**


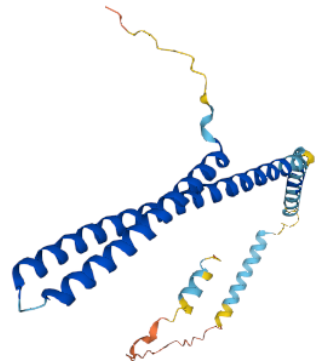
Names & Taxonomy  
Subcellular Location  
Disease & Variants  
PTM/Processing  
Expression  
Interaction  
Structure  
Family & Domains  
Sequence & Isoform  
Similar Proteins



### Structure<sup>i</sup>

**Model Confidence:**

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

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions with low pLDDT may be unstructured in isolation.



SOURCE	IDENTIFIER	METHOD	RESOLUTION	CHAIN	POSITIONS	LINKS
-- Select --		-- Select --				
PDB	3C30	X-ray	2.15 Å	B	210-222	<a href="#">PDBe</a> · <a href="#">RCSB-PDB</a> · <a href="#">PDBj</a> · <a href="#">PDBsum</a> 
PDB	5MK1	X-ray	2.50 Å	E/F/H/K	205-222	<a href="#">PDBe</a> · <a href="#">RCSB-PDB</a> · <a href="#">PDBj</a> · <a href="#">PDBsum</a> 
AlphaFold	AF-Q9BY43-F1	Predicted			1-222	<a href="#">AlphaFold</a> 

# 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
REMARK 3
REMARK 3
```

Vim

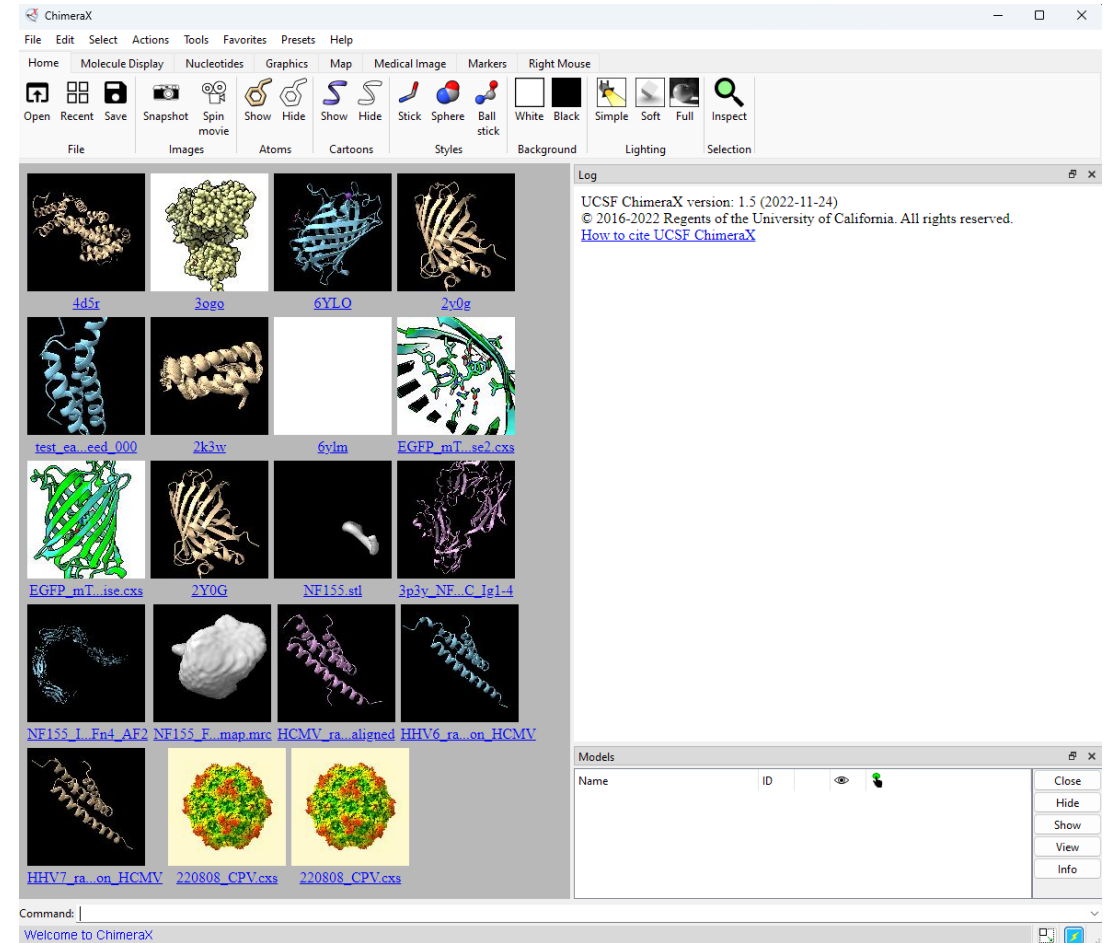
```
Z:\Projects\Old\CamPath\VACV\E2\final_structure\VACV_E2_refine_20.pdb - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
new 1 140906_malonate_interface.pml VACV_E2_refine_20.pdb
1 REMARK 3
2 REMARK 3 REFINEMENT.
3 REMARK 3 PROGRAM : PHENIX (1.19.2_4158: ???)
4 REMARK 3 AUTHORS : Adams, Afonine, Bunkoczi, Burnley, Chen, Dar, Davis,
5 REMARK 3 : Draizen, Echols, Gildea, Gros, Grosse-Kunstleve, Headd,
6 REMARK 3 : Hintze, Hung, Ioerger, Liebschner, McCoy, McKee, Moriarty,
7 REMARK 3 : Oeffner, Poon, Read, Richardson, Richardson, Sacchettini,
8 REMARK 3 : Sauter, Sobolev, Storoni, Terwilliger, Williams, Zwart
9 REMARK 3
10 REMARK 3 X-RAY DATA.
11 REMARK 3
12 REMARK 3 REFINEMENT TARGET : ML
13 REMARK 3
14 REMARK 3 DATA USED IN REFINEMENT.
15 REMARK 3 RESOLUTION RANGE HIGH (ANGSTROMS) : 2.30
16 REMARK 3 RESOLUTION RANGE LOW (ANGSTROMS) : 35.52
17 REMARK 3 MIN(FOBS/SIGMA_FOBS) : 1.34
18 REMARK 3 COMPLETENESS FOR RANGE (%) : 99.59
19 REMARK 3 NUMBER OF REFLECTIONS : 46586
20 REMARK 3 NUMBER OF REFLECTIONS (NON-ANOMALOUS) : 46586
21 REMARK 3
22 REMARK 3 FIT TO DATA USED IN REFINEMENT.
23 REMARK 3 R VALUE (WORKING + TEST SET) : 0.1943
24 REMARK 3 R VALUE (WORKING SET) : 0.1919
25 REMARK 3 FREE R VALUE : 0.2370
26 REMARK 3 FREE R VALUE TEST SET SIZE (%) : 5.14
27 REMARK 3 FREE R VALUE TEST SET COUNT : 2396
```

**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

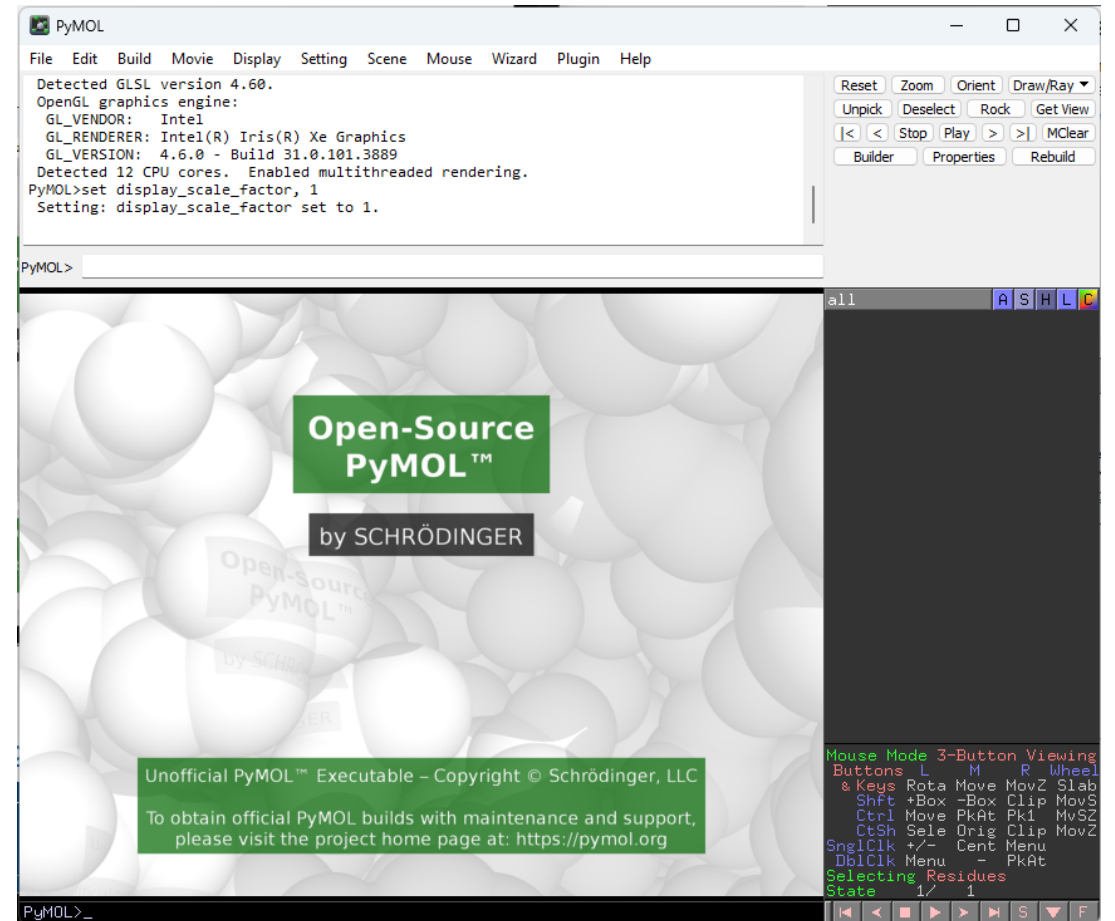


<https://www.cgl.ucsf.edu/chimerax/>



# Software for viewing protein structures

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

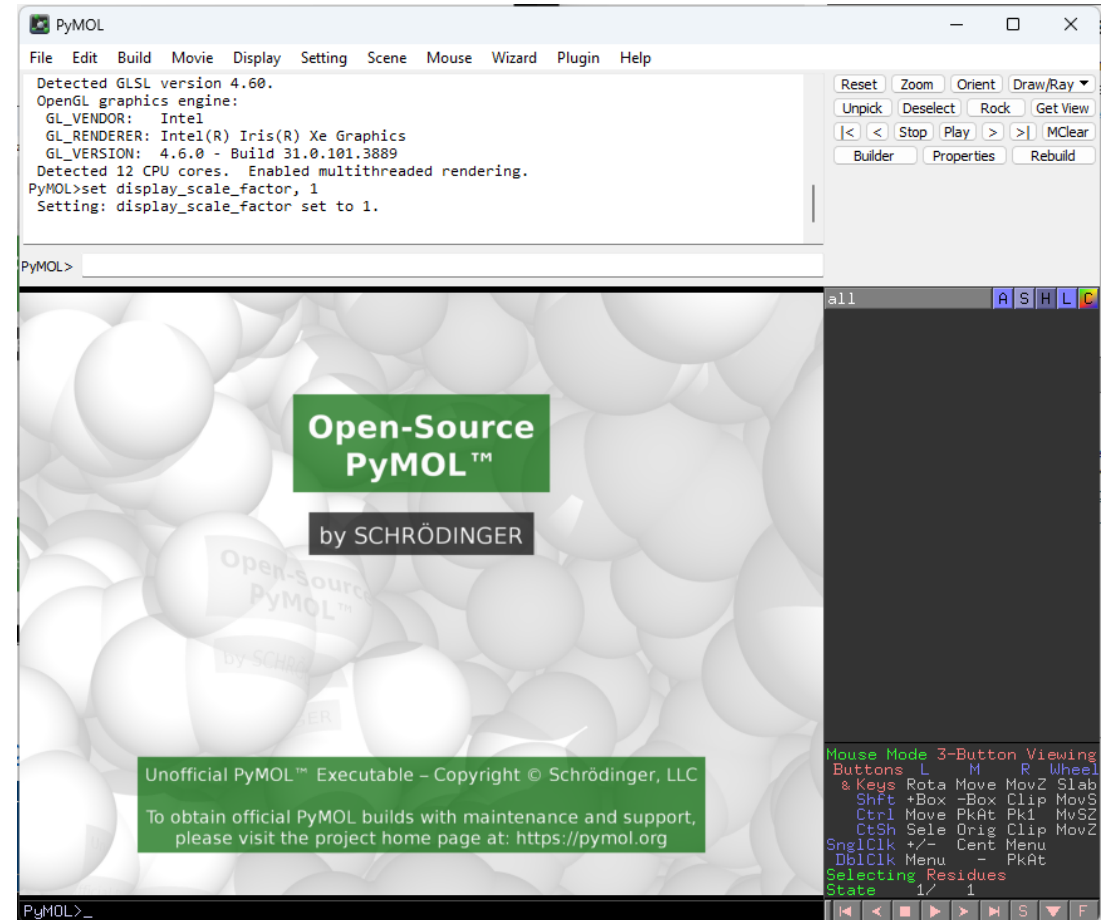


<https://github.com/schrodinger/pymol-open-source>



# 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/Windows_Install)  
[https://pymolwiki.org/index.php/MAC\\_Install](https://pymolwiki.org/index.php/MAC_Install)

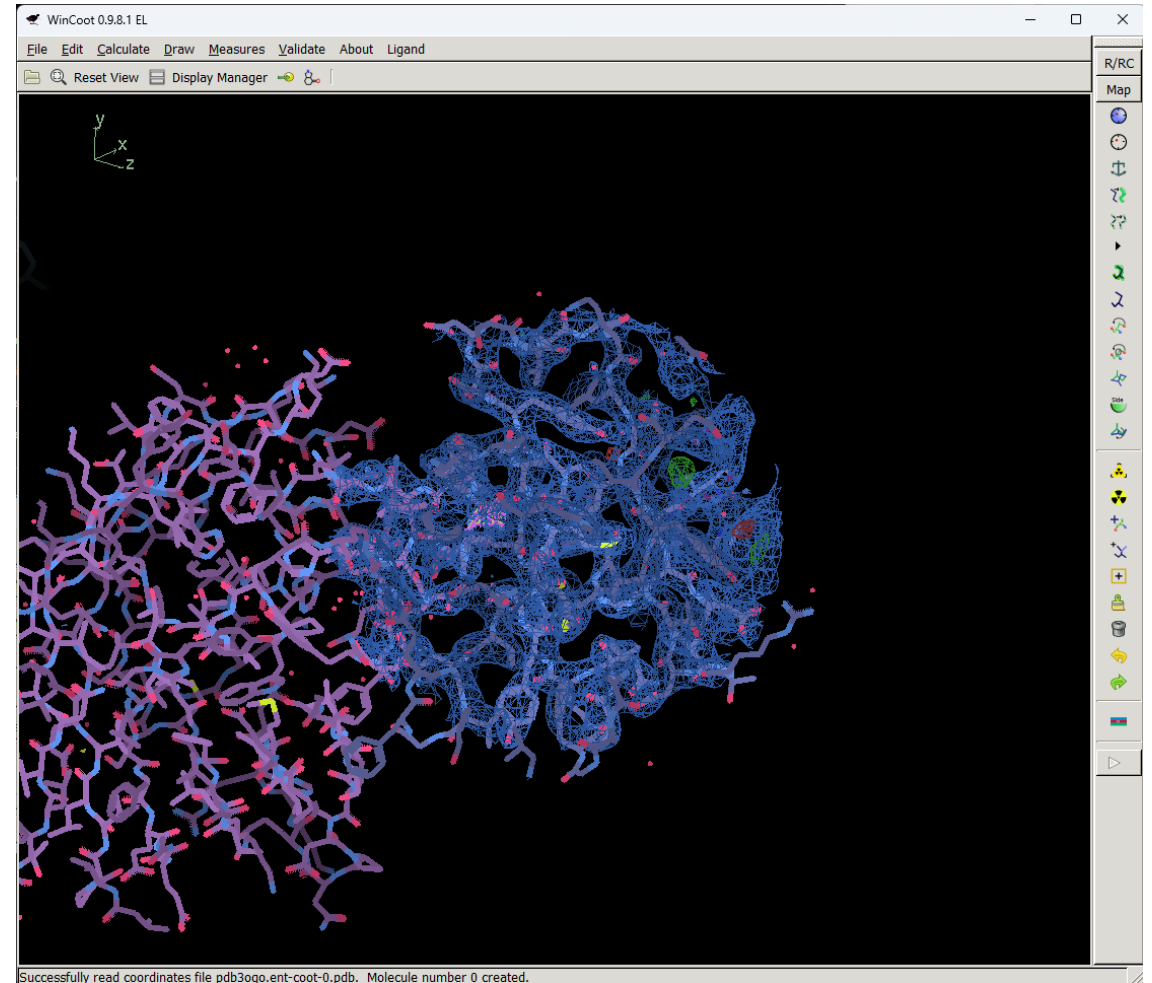


<https://github.com/schrodinger/pymol-open-source>



# Software for viewing protein structures

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

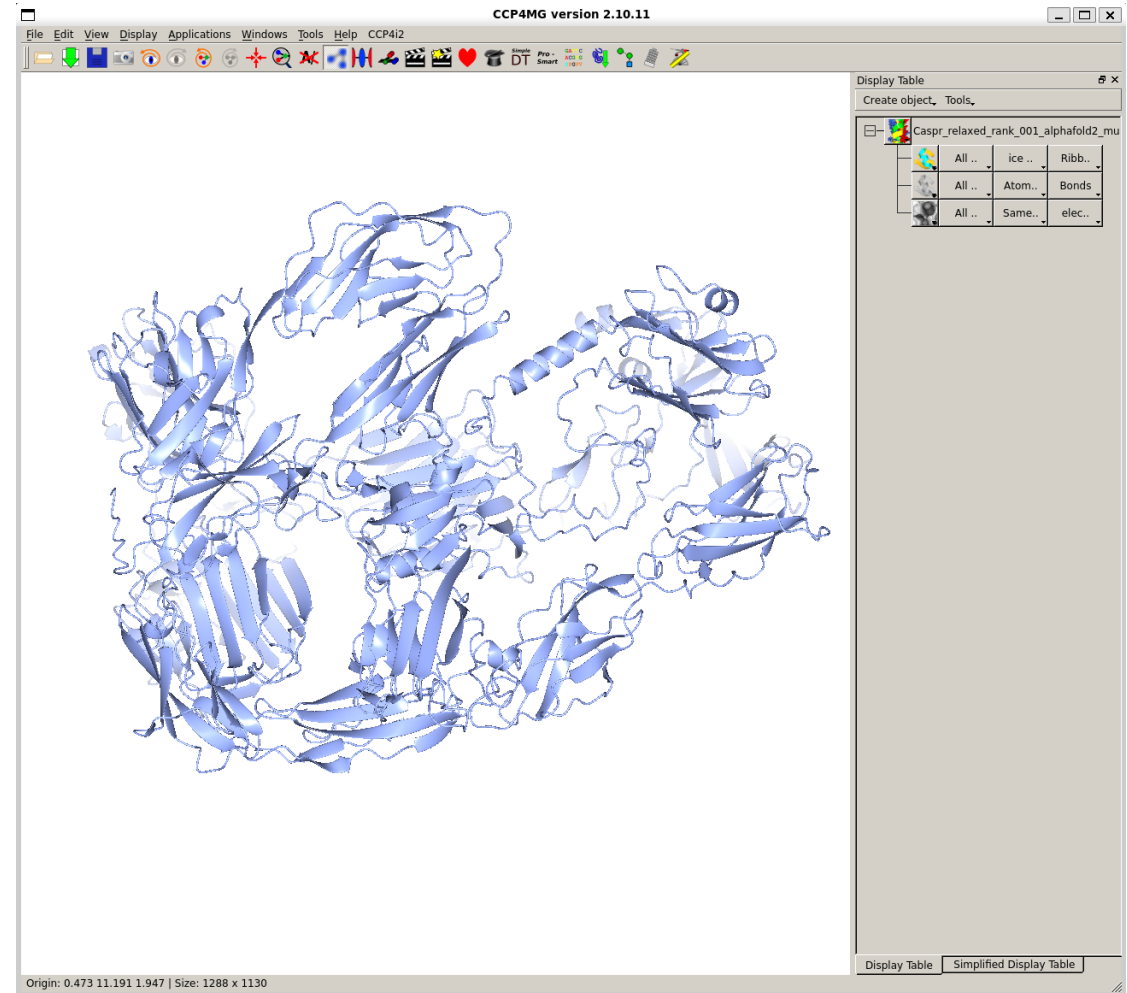


<https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/>



# Software for viewing protein structures

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

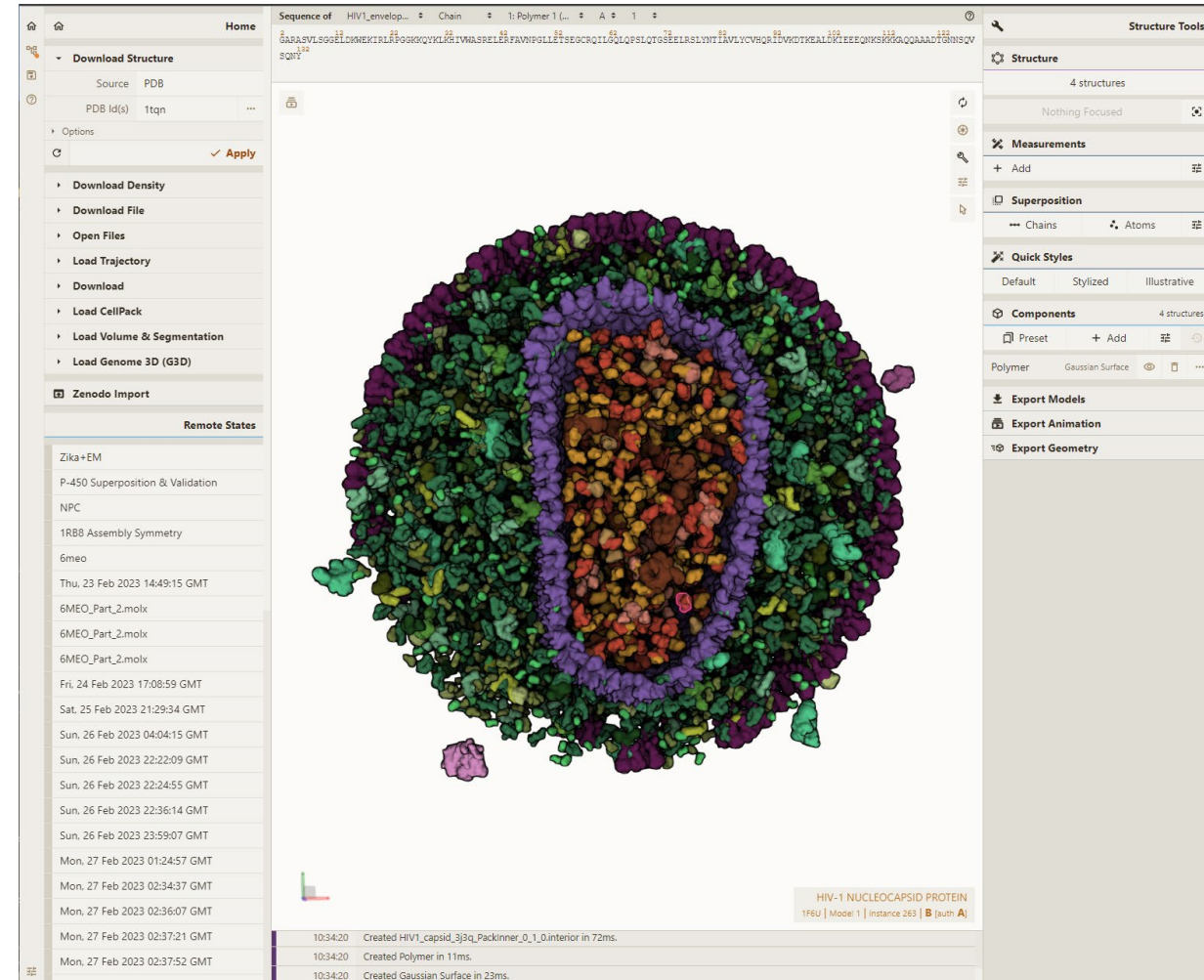


<https://www.ccp4.ac.uk/MG/>



# 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\*

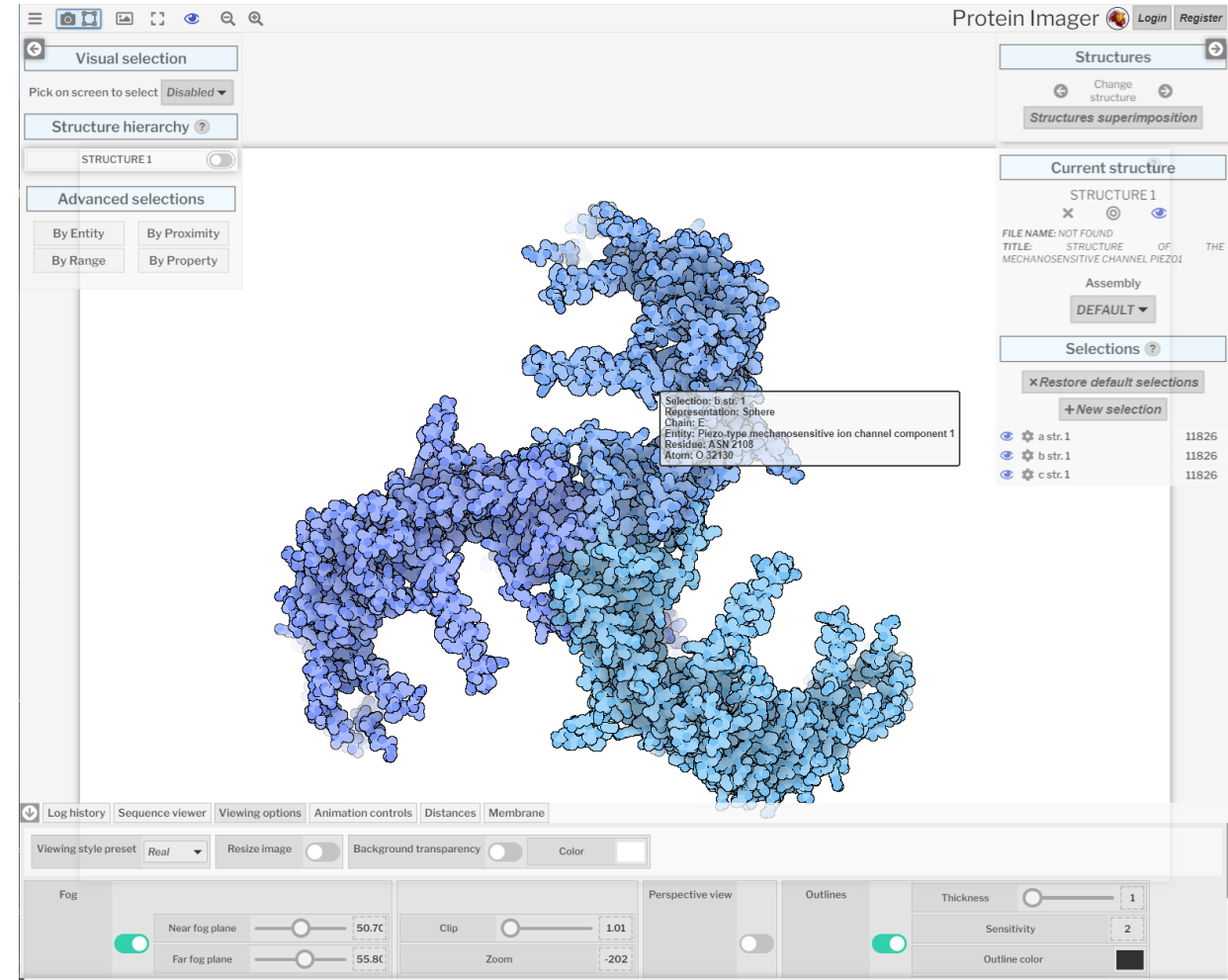


<https://molstar.org/>



# 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

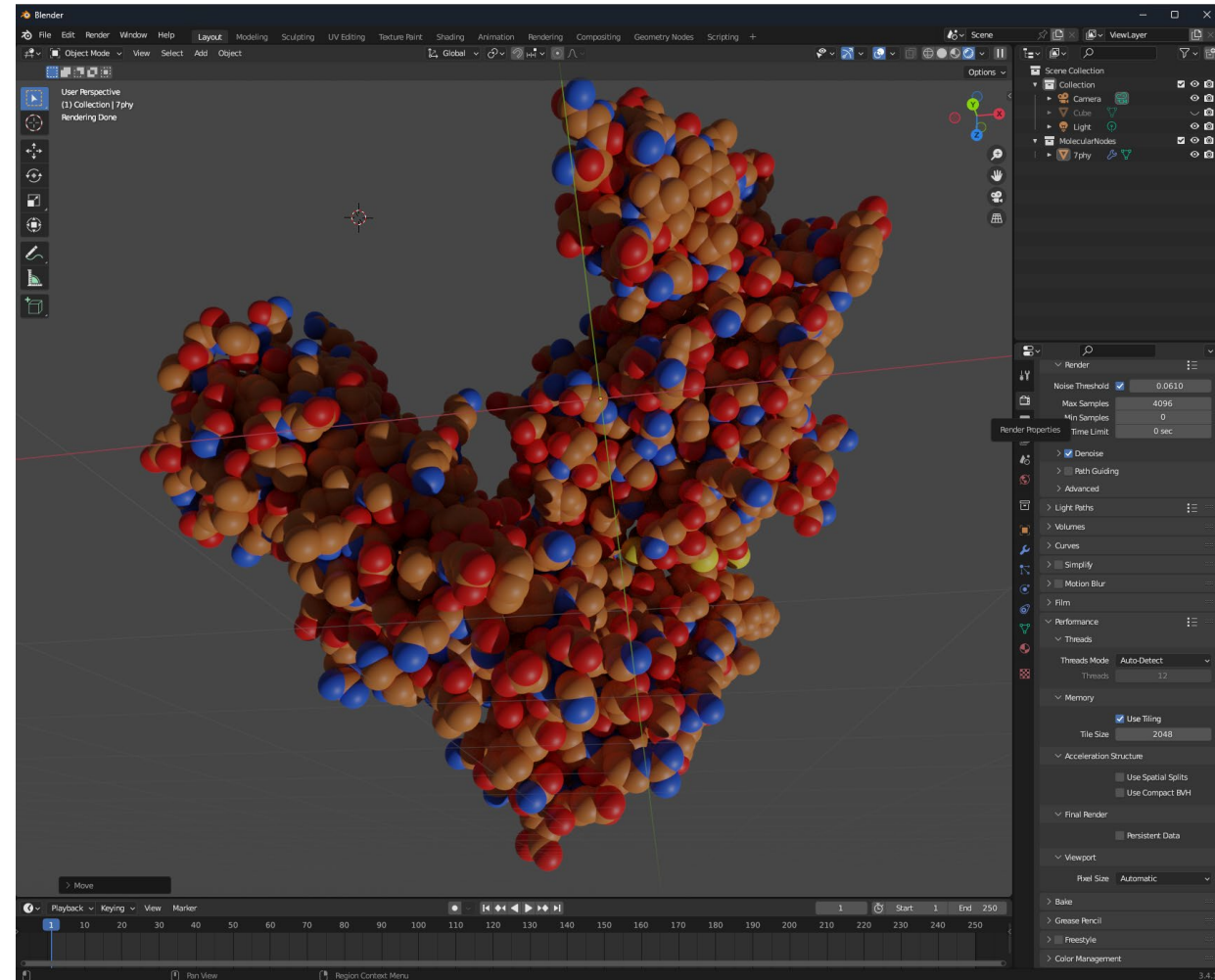


<https://3dproteinimaging.com/protein-imager/>



# 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)



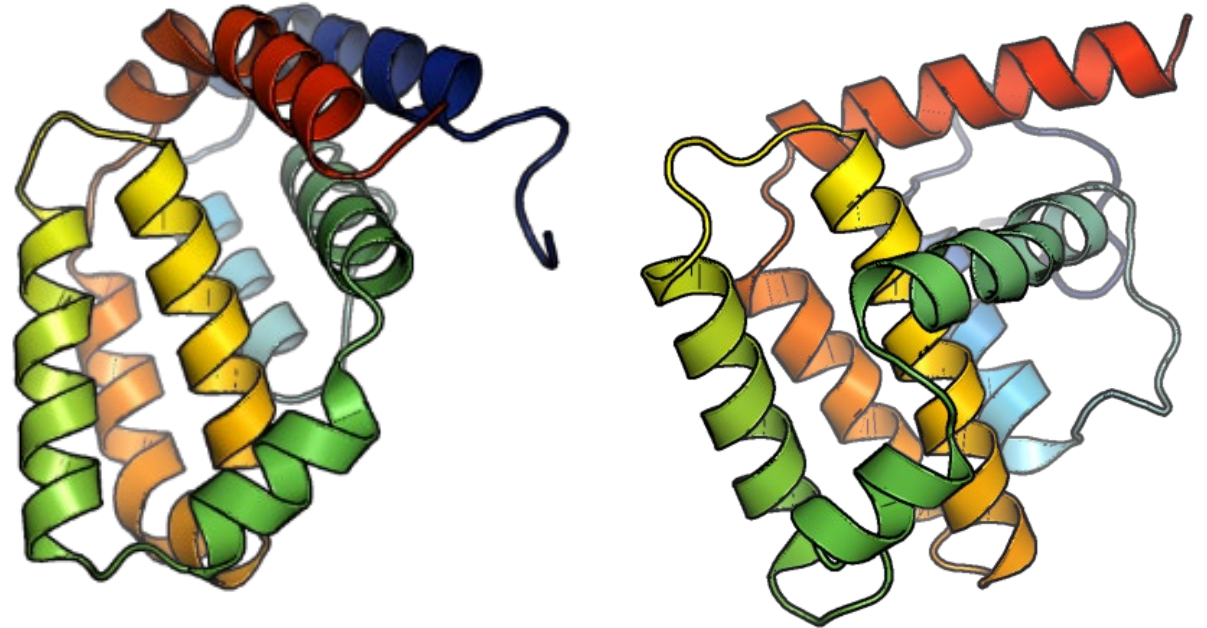
<https://bradyajohnston.github.io/MolecularNodes/>





# 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)

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

- Where  $\delta$  is the distance between two equivalent atoms
- Generally measured between C $\alpha$  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, ...



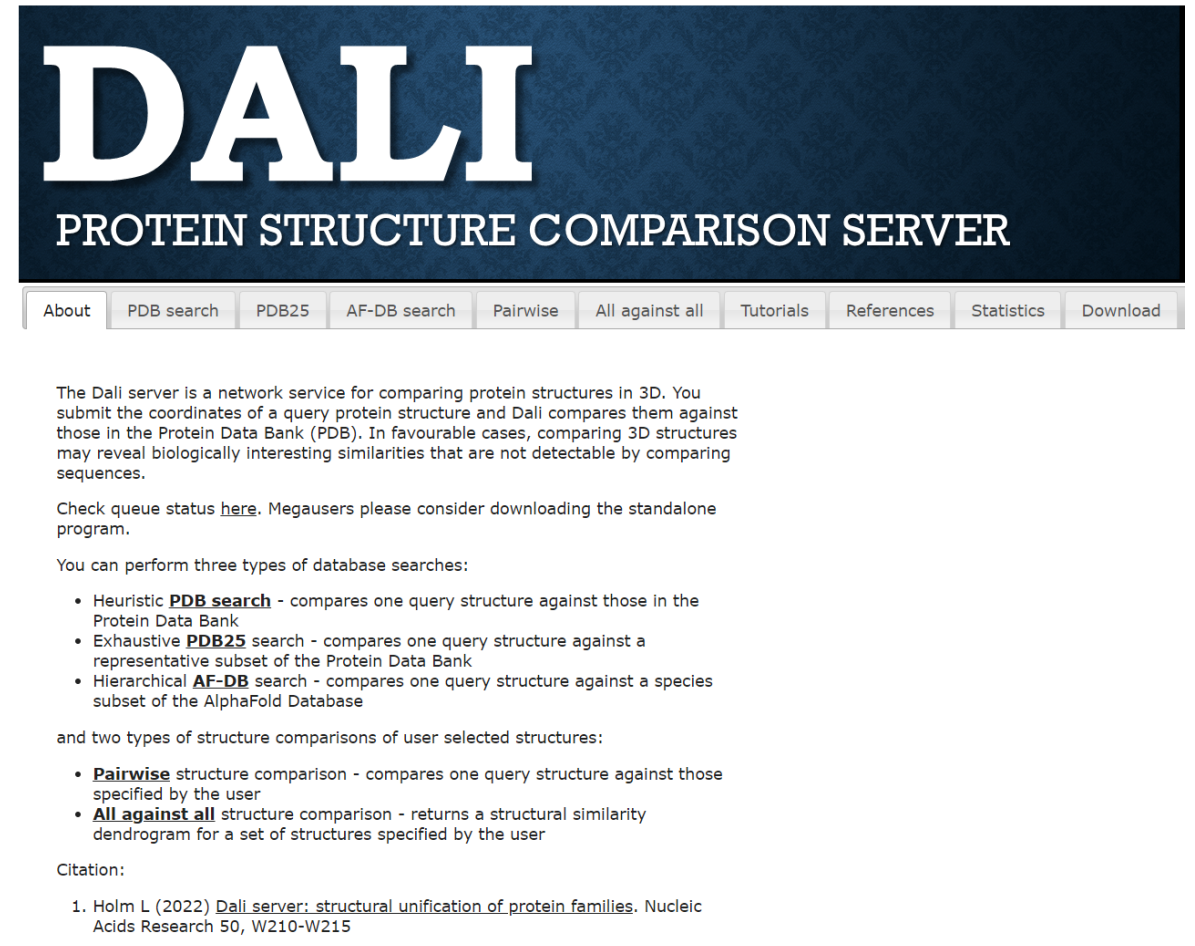
*How can you find other structures similar to your protein of interest?*



# Structure-based searches: DALI

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

- Can search databases of experimental structures and AlphaFold2 models



The screenshot shows the DALI Protein Structure Comparison Server website. The header features the DALI logo in large white letters on a dark blue background, with the text "PROTEIN STRUCTURE COMPARISON SERVER" below it. A navigation bar contains links for "About", "PDB search", "PDB25", "AF-DB search", "Pairwise", "All against all", "Tutorials", "References", "Statistics", and "Download". The main content area describes the server's purpose: comparing protein structures in 3D against the Protein Data Bank (PDB) and the AlphaFold Database. It lists three types of database searches: Heuristic PDB search, Exhaustive PDB25 search, and Hierarchical AF-DB search. It also lists two types of structure comparisons: Pairwise and All against all. A citation is provided at the bottom: Holm L (2022) Dali server: structural unification of protein families. Nucleic Acids Research 50, W210-W215.

**DALI**  
PROTEIN STRUCTURE COMPARISON SERVER

About PDB search PDB25 AF-DB search Pairwise All against all Tutorials References Statistics Download

The Dali server is a network service for comparing protein structures in 3D. You submit the coordinates of a query protein structure and Dali compares them against those in the Protein Data Bank (PDB). In favourable cases, comparing 3D structures may reveal biologically interesting similarities that are not detectable by comparing sequences.

Check queue status [here](#). Megausers please consider downloading the standalone program.

You can perform three types of database searches:

- Heuristic **PDB search** - compares one query structure against those in the Protein Data Bank
- Exhaustive **PDB25** search - compares one query structure against a representative subset of the Protein Data Bank
- Hierarchical **AF-DB** search - compares one query structure against a species subset of the AlphaFold Database

and two types of structure comparisons of user selected structures:

- **Pairwise** structure comparison - compares one query structure against those specified by the user
- **All against all** structure comparison - returns a structural similarity dendrogram for a set of structures specified by the user

Citation:

1. Holm L (2022) [Dali server: structural unification of protein families](#). Nucleic Acids Research 50, W210-W215



# 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.

Expand gaps

## Summary

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
<input type="checkbox"/> 1:	7phy-A	54.5	0.0	732	732	100	<a href="#">PDB</a>	MOLECULE: PROTEIN E2;
<input type="checkbox"/> 2:	6plm-B	10.3	4.8	256	752	10	<a href="#">PDB</a>	MOLECULE: SIDJ PROTEIN;
<input type="checkbox"/> 3:	7r5s-I	4.9	15.1	176	622	7	<a href="#">PDB</a>	MOLECULE: CENTROMERE PROTEIN H;
<input type="checkbox"/> 4:	1z2c-B	4.8	18.7	127	346	8	<a href="#">PDB</a>	MOLECULE: RHO-RELATED GTP-BINDING PROTEIN RHOC;
<input type="checkbox"/> 5:	6yle-D	4.7	9.8	139	547	5	<a href="#">PDB</a>	MOLECULE: PRE-RRNA-PROCESSING PROTEIN IPI3;
<input type="checkbox"/> 6:	4lnb-A	4.7	10.0	156	339	6	<a href="#">PDB</a>	MOLECULE: CAAX FARNESYLTRANSFERASE ALPHA SUBUNIT RAM2;
<input type="checkbox"/> 7:	3dad-A	4.5	10.1	104	324	13	<a href="#">PDB</a>	MOLECULE: FH1/FH2 DOMAIN-CONTAINING PROTEIN 1;
<input type="checkbox"/> 8:	8e2f-A	4.4	12.0	140	771	4	<a href="#">PDB</a>	MOLECULE: BACULOVIRAL IAP REPEAT-CONTAINING PROTEIN 6;
<input type="checkbox"/> 9:	7uwf-C	4.3	12.1	119	517	5	<a href="#">PDB</a>	MOLECULE: WD REPEAT-CONTAINING PROTEIN 18;
<input type="checkbox"/> 10:	6dee-A	4.2	13.7	125	382	6	<a href="#">PDB</a>	MOLECULE: NCK-INTERACTING PROTEIN WITH SH3 DOMAIN;
<input type="checkbox"/> 11:	4d0l-E	4.2	14.3	110	479	10	<a href="#">PDB</a>	MOLECULE: PHOSPHATIDYLINOSITOL 4-KINASE BETA;
<input type="checkbox"/> 12:	7zkq-C	4.2	8.0	122	416	9	<a href="#">PDB</a>	MOLECULE: NADH DEHYDROGENASE SUBUNIT 2;
<input type="checkbox"/> 13:	4imj-A	4.1	14.4	139	333	5	<a href="#">PDB</a>	MOLECULE: SYMPLEKIN;
<input type="checkbox"/> 14:	7x9r-A	4.1	19.4	153	448	8	<a href="#">PDB</a>	MOLECULE: GLYCOSYL TRANSFERASE FAMILY 2;
<input type="checkbox"/> 15:	4cem-B	4.1	3.8	122	309	4	<a href="#">PDB</a>	MOLECULE: REGULATOR OF NONSENSE TRANSCRIPTS 2;
<input type="checkbox"/> 16:	4ww9-A	4.1	3.7	125	238	11	<a href="#">PDB</a>	MOLECULE: EKC/KEOPS COMPLEX SUBUNIT BUD32;
<input type="checkbox"/> 17:	6yai-M	4.0	16.8	168	518	13	<a href="#">PDB</a>	MOLECULE: CLATHRIN HEAVY CHAIN;
<input type="checkbox"/> 18:	6j6g-v	3.9	24.8	153	722	7	<a href="#">PDB</a>	MOLECULE: PRE-MRNA-SPLICING FACTOR 8;
<input type="checkbox"/> 19:	5wy3-B	3.9	7.5	118	366	4	<a href="#">PDB</a>	MOLECULE: PUTATIVE UNCHARACTERIZED PROTEIN;
<input type="checkbox"/> 20:	514k-P	3.9	17.4	138	456	3	<a href="#">PDB</a>	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

The screenshot shows the PDBeFold website interface. At the top, there is a navigation bar with 'EMBL-EBI' logo, 'Protein Data Bank in Europe' text, and 'Bringing Structure to Biology' tagline. The 'PDBeFold' title is prominently displayed. A search bar is visible in the top right corner. Below the navigation bar, there is a sidebar with a list of links under 'PDBeFold links' and 'Other links'. The main content area features the heading 'PDBeFold. Structure Similarity.' followed by a section titled 'PDBeFold functionality:' which lists several search capabilities. A 'Launch PDBeFold' button is present below the functionality list. At the bottom of the main content area, there is a small section titled 'PDBeFold: A comparison with other protein matching services.' and a feedback request.

EMBL-EBI Protein Data Bank in Europe Bringing Structure to Biology

Services Research Training About us

PDBeFold

Share Feedback

- PDBeFold links
  - [FAQ](#)
  - [Visualisation](#)
  - [Performance](#)
  - [Privacy](#)
  - [Version log](#)
  - [PDBeFold Links](#)
  - [Comparisons](#)
  - [Publications](#)
  - [PDBeFOLD tutorial](#)
- Other links
  - [PDBePISA](#)
  - [CCP4](#)
  - [CoorLib](#)
  - [Rasmol](#)
  - [Rastop](#)
  - [Jmol](#)
  - [PDB](#)
  - [SCOP](#)
  - [PDBeMotif](#)
  - [GeneCensus](#)
  - [FSSP](#)
  - [CATH](#)
  - [PDBSum](#)
  - [UniProt](#)

## PDBeFold. Structure Similarity.

PDBeFold functionality:

- pairwise comparison and 3D alignment of protein structures
- multiple comparison and 3D alignment of protein structures
- examination of a protein structure for similarity with the whole [PDB archive](#) or [SCOP archive](#)
- best Co-alignment of compared structures
- download and visualisation of best-superposed structures using [Rasmol](#) (Unix/Linux platforms), [Rastop](#) (Windows machines) and [Jmol](#) (platform-independent server-side java viewer)
- linking the results to other services - [PDBeMotif](#), [SCOP](#), [GeneCensus](#), [FSSP](#), [CATH](#), [PDBSum](#), [UniProt](#)

[Launch PDBeFold](#)

PDBeFold: A [comparison](#) with other protein matching services.  
PDBeFold is used as a structure search engine in [PDBePISA](#).  
PDBeFold queries may be launched from any web site ([instructions](#)).  
We welcome your feedback! Please send any questions, comments, suggestions and bug reports using the FEEDBACK button on the top of the page.

# 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)

The screenshot shows the PDBeFold web interface. At the top, it says "EMBL-EBI Protein Data Bank in Europe Bringing Structure to Biology" and "PDBeFold". There are navigation links for "Services", "Research", "Training", and "About us". A message box states: "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." Below this is a "Repeat as suggested" button. The main heading is "Structure Alignment Results." followed by a help icon. The query is "Query: pdb entry 7phy:A , 732 residues VACCINIA VIRUS E2". It indicates "Examined 187732 entries, (530702 chains). Displaying Matches 1-20 of 107." There are navigation buttons: "Back to query", "next", "last page", "Sort by..." (set to "Q-score"), "arrange by SCOP family" (checkbox), "match" (input field with "1"), and "jump". A table of results is shown with columns for "#", "Scoring", "RMSD", "N<sub>align</sub>", "N<sub>g</sub>", "%<sub>seq</sub>", "Query", "Target (PDB entry)", and "Title".

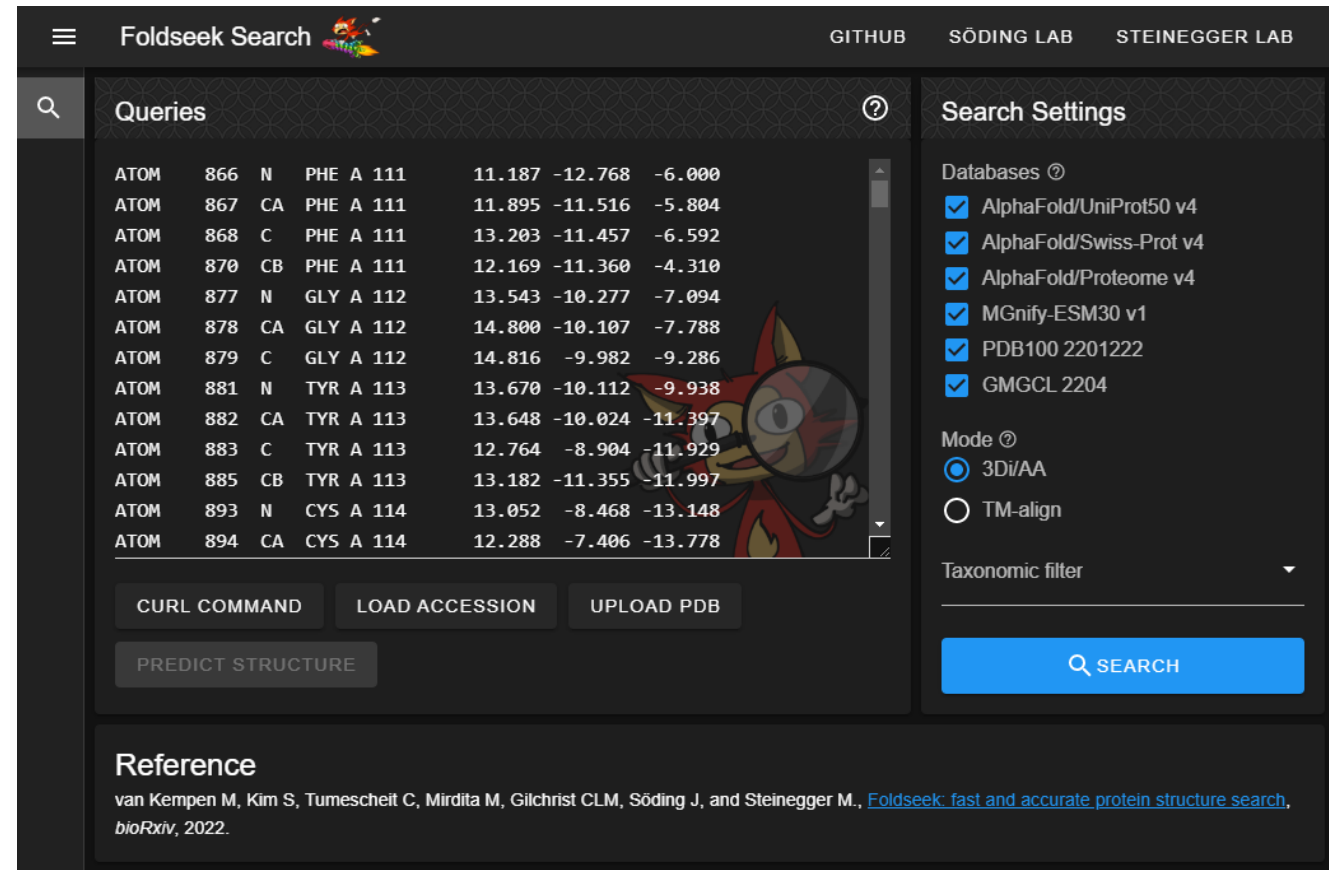
#	Scoring			RMSD	N <sub>align</sub>	N <sub>g</sub>	% <sub>seq</sub>	Query				Target (PDB entry)	
	Q	P	Z					% <sub>sse</sub>	Match	% <sub>sse</sub>	N <sub>res</sub>	*	Title
<a href="#">1</a>	1.00	307.4	53.1	0.00	732	0	100	100	<a href="#">7phy:A</a>	100	732	<input type="checkbox"/>	VACCINIA VIRUS E2
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



Foldseek Search

GITHUB SÖDING LAB STEINEGGER LAB

Queries

ATOM	866	N	PHE	A	111	11.187	-12.768	-6.000
ATOM	867	CA	PHE	A	111	11.895	-11.516	-5.804
ATOM	868	C	PHE	A	111	13.203	-11.457	-6.592
ATOM	870	CB	PHE	A	111	12.169	-11.360	-4.310
ATOM	877	N	GLY	A	112	13.543	-10.277	-7.094
ATOM	878	CA	GLY	A	112	14.800	-10.107	-7.788
ATOM	879	C	GLY	A	112	14.816	-9.982	-9.286
ATOM	881	N	TYR	A	113	13.670	-10.112	-9.938
ATOM	882	CA	TYR	A	113	13.648	-10.024	-11.397
ATOM	883	C	TYR	A	113	12.764	-8.904	-11.929
ATOM	885	CB	TYR	A	113	13.182	-11.355	-11.997
ATOM	893	N	CYS	A	114	13.052	-8.468	-13.148
ATOM	894	CA	CYS	A	114	12.288	-7.406	-13.778

CURL COMMAND LOAD ACCESSION UPLOAD PDB

PREDICT STRUCTURE

Search Settings

Databases

- AlphaFold/UniProt50 v4
- AlphaFold/Swiss-Prot v4
- AlphaFold/Proteome v4
- MGnify-ESM30 v1
- PDB100 2201222
- GMGCL 2204

Mode

- 3Di/AA
- TM-align

Taxonomic filter

SEARCH

Reference

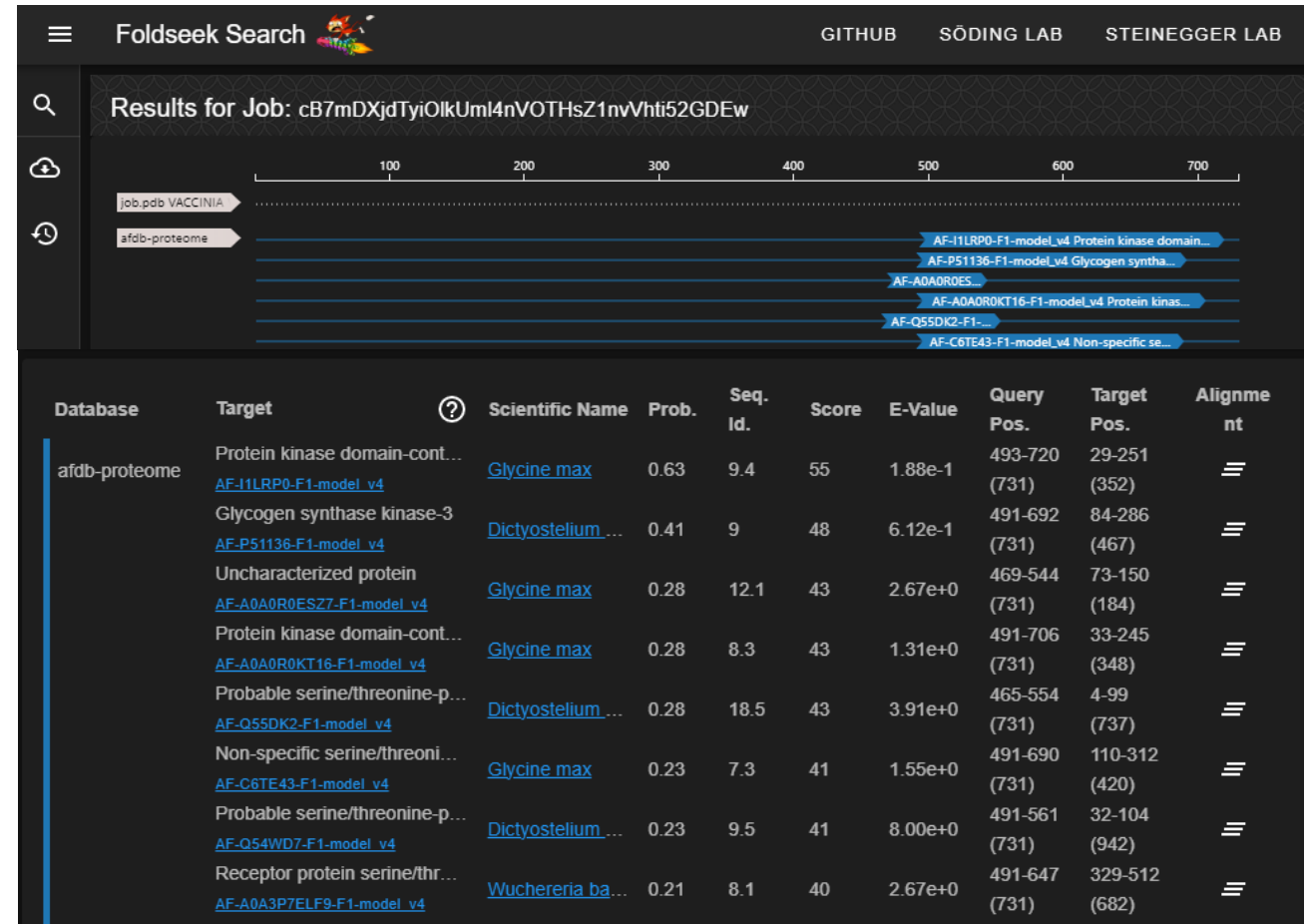
van Kempen M, Kim S, Tumescheit C, Mirdita M, Gilchrist CLM, Söding J, and Steinegger M., [Foldseek: fast and accurate protein structure search](#), *bioRxiv*, 2022.



# 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)



Foldseek Search

Results for Job: cB7mDXjdTyiOIkUml4nVOTHsZ1nvVhti52GDEw

job.pdb VACCINIA

afdb-proteome

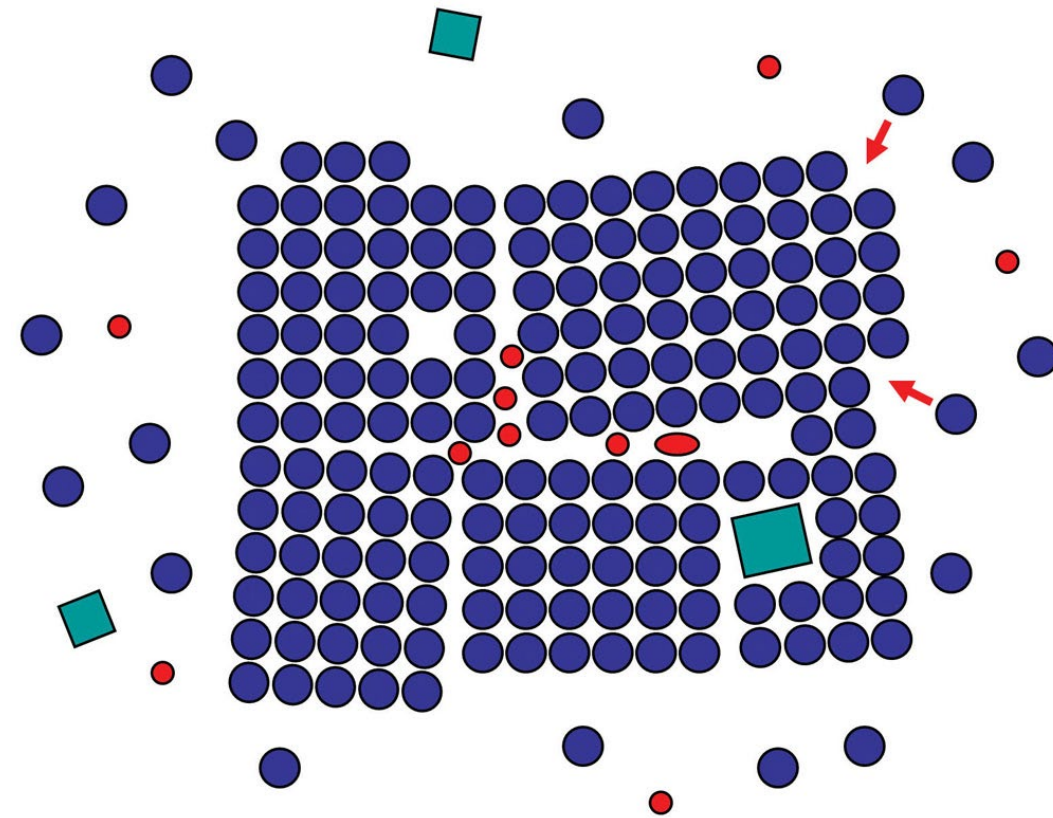
Database	Target	Scientific Name	Prob.	Seq. Id.	Score	E-Value	Query Pos.	Target Pos.	Alignment
afdb-proteome	Protein kinase domain-cont...	<a href="#">Glycine max</a>	0.63	9.4	55	1.88e-1	493-720 (731)	29-251 (352)	≡
	<a href="#">AF-I1LRP0-F1-model_v4</a>								
	Glycogen synthase kinase-3	<a href="#">Dictyostelium...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

# *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?



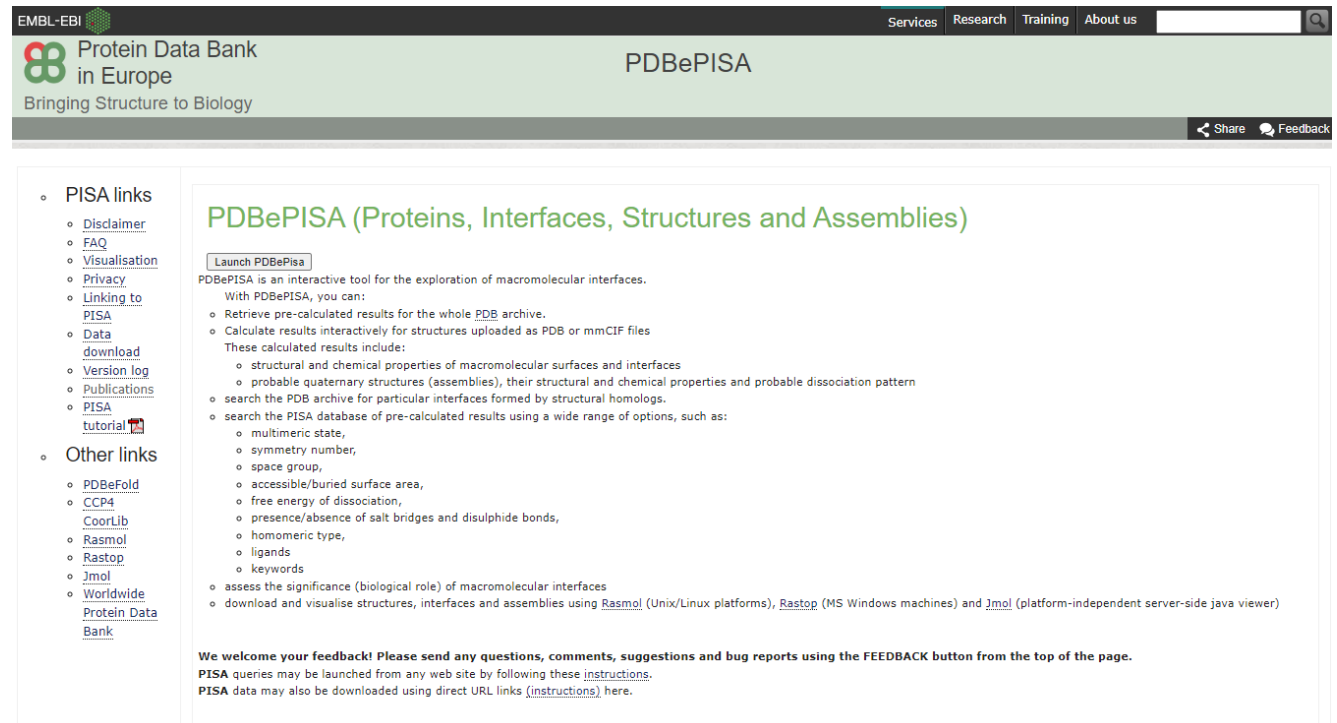
© Garland Science 2010

# Inspecting protein interaction interfaces – PDBePISA

- PDBePISA server

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

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



The screenshot shows the PDBePISA website interface. At the top, there is a navigation bar with 'Services', 'Research', 'Training', and 'About us' links. Below this is the EMBL-EBI logo and the text 'Protein Data Bank in Europe' and 'Bringing Structure to Biology'. The main content area is titled 'PDBePISA (Proteins, Interfaces, Structures and Assemblies)'. It features a 'Launch PDBePisa' button and a list of capabilities: 'Retrieve pre-calculated results for the whole PDB archive', 'Calculate results interactively for structures uploaded as PDB or mmCIF files', and 'search the PDB archive for particular interfaces formed by structural homologs'. A detailed list of calculated results includes: 'structural and chemical properties of macromolecular surfaces and interfaces', 'probable quaternary structures (assemblies), their structural and chemical properties and probable dissociation pattern', and 'search the PISA database of pre-calculated results using a wide range of options, such as: multimeric state, symmetry number, space group, accessible/buried surface area, free energy of dissociation, presence/absence of salt bridges and disulphide bonds, homomeric type, ligands, and keywords'. It also mentions 'assess the significance (biological role) of macromolecular interfaces' and 'download and visualise structures, interfaces and assemblies using Rasmol (Unix/Linux platforms), Rastop (MS Windows machines) and Jmol (platform-independent server-side java viewer)'. A footer section encourages feedback and provides instructions on how to launch queries and download PISA data.

- PISA links
  - [Disclaimer](#)
  - [FAQ](#)
  - [Visualisation](#)
  - [Privacy](#)
  - [Linking to PISA](#)
  - [Data download](#)
  - [Version log](#)
  - [Publications](#)
  - [PISA tutorial](#)
- Other links
  - [PDBeFold](#)
  - [CCP4](#)
  - [CoorLib](#)
  - [Rasmol](#)
  - [Rastop](#)
  - [Jmol](#)
  - [Worldwide Protein Data Bank](#)

## PDBePISA (Proteins, Interfaces, Structures and Assemblies)

[Launch PDBePisa](#)

PDBePISA is an interactive tool for the exploration of macromolecular interfaces.

With PDBePISA, you can:

- Retrieve pre-calculated results for the whole PDB archive.
- Calculate results interactively for structures uploaded as PDB or mmCIF files

These calculated results include:

- structural and chemical properties of macromolecular surfaces and interfaces
- probable quaternary structures (assemblies), their structural and chemical properties and probable dissociation pattern

- search the PDB archive for particular interfaces formed by structural homologs.
- search the PISA database of pre-calculated results using a wide range of options, such as:
  - multimeric state,
  - symmetry number,
  - space group,
  - accessible/buried surface area,
  - free energy of dissociation,
  - presence/absence of salt bridges and disulphide bonds,
  - homomeric type,
  - ligands
  - keywords

- assess the significance (biological role) of macromolecular interfaces
- download and visualise structures, interfaces and assemblies using [Rasmol](#) (Unix/Linux platforms), [Rastop](#) (MS Windows machines) and [Jmol](#) (platform-independent server-side java viewer)

**We welcome your feedback! Please send any questions, comments, suggestions and bug reports using the FEEDBACK button from the top of the page.**

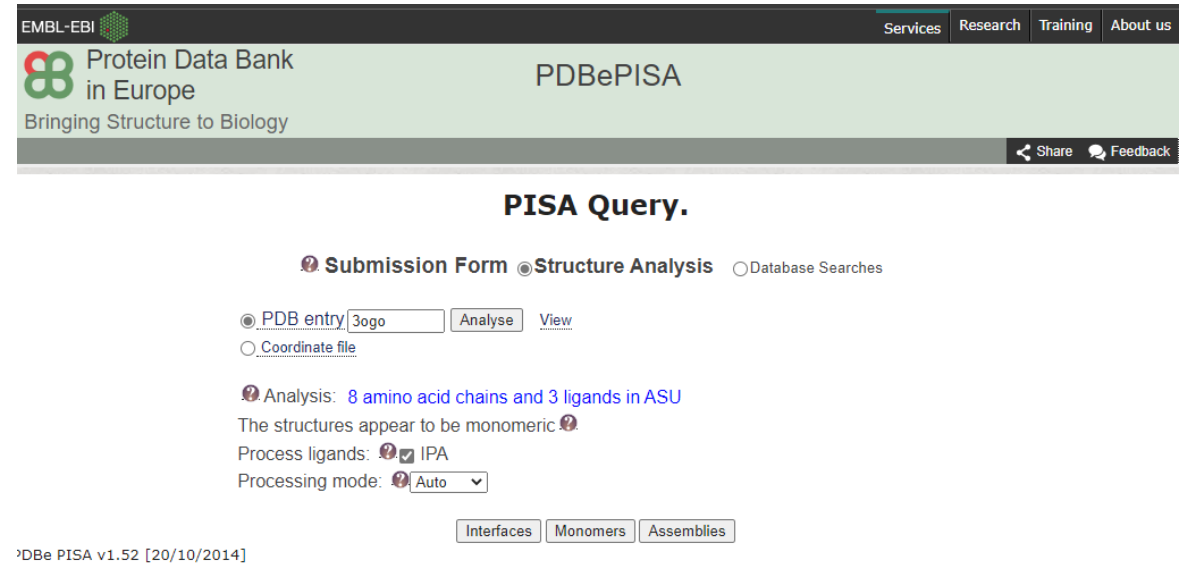
PISA queries may be launched from any web site by following these [instructions](#).

PISA data may also be downloaded using direct URL links ([instructions](#)) here.



# Inspecting protein interaction interfaces – PDBePISA

- Analyse any structure from PDB



The screenshot shows the PDBePISA web interface. At the top, there is a navigation bar with links for Services, Research, Training, and About us. Below this, the Protein Data Bank logo and name are displayed, along with the tagline 'Bringing Structure to Biology'. The main heading is 'PDBePISA'. A 'PISA Query.' section is active, with radio buttons for 'Submission Form', 'Structure Analysis', and 'Database Searches'. Under 'Structure Analysis', there are radio buttons for 'PDB entry' (selected) and 'Coordinate file'. A text input field contains '3ogo', followed by 'Analyse' and 'View' buttons. Below this, the analysis results are shown: 'Analysis: 8 amino acid chains and 3 ligands in ASU'. A message states 'The structures appear to be monomeric.' There are checkboxes for 'Process ligands' (checked) and 'IPPA'. A dropdown menu for 'Processing mode' is set to 'Auto'. At the bottom, there are buttons for 'Interfaces', 'Monomers', and 'Assemblies'. The footer text reads 'PDBe PISA v1.52 [20/10/2014]'.

# 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)

The screenshot shows the PDBePISA web interface. At the top, there is a navigation bar with 'Services', 'Research', 'Training', and 'About us'. Below this is the 'Protein Data Bank in Europe' logo and the text 'Bringing Structure to Biology'. The main heading is 'PDBePISA'. Below the heading, there are tabs for 'Submission Form', 'Structure Analysis', and 'Database Searches'. The 'Structure Analysis' tab is selected. Under this tab, there are options for 'PDB entry' (3ogo) and 'Coordinate file'. The 'PDB entry' option is selected, and there are 'Analyse' and 'View' buttons. Below these, there is a section for 'Analysis' showing '8 amino acid chains and 3 ligands in ASU'. The text 'The structures appear to be monomeric' is displayed. There are also options for 'Process ligands' (checked) and 'Processing mode' (Auto). Below this, there are buttons for 'Interfaces', 'Monomers', and 'Assemblies'. The 'Interfaces' button is highlighted. Below the buttons, there is a section for 'PDBe PISA v1.52 [20/10/2014]'. The main heading for this section is 'PISA Interface List.'. Below this, there is a 'Session Map' section with buttons for 'Start', 'Interfaces', and 'Interface Search'. The 'Interfaces' button is selected. Below this, there is a section for 'Interfaces in PDB 3ogo crystal.' with the text 'Space symmetry group: P 21 21 2. Resolution: 2.80 Å'. Below this, there is a table of interfaces. The table has columns for '#', 'Id', 'NN', 'Range', 'Structure 1' (N<sub>at</sub>, N<sub>res</sub>, Surface Å<sup>2</sup>), 'x', 'Structure 2' (Symmetry op-n, Sym.ID, N<sub>at</sub>, N<sub>res</sub>, Surface Å<sup>2</sup>), 'interface area, Å<sup>2</sup>', 'ΔG kcal/mol', 'ΔG P-value', 'N<sub>HB</sub>', 'N<sub>SB</sub>', 'N<sub>DS</sub>', and 'CSS'. The table contains 4 rows of data.

#	Id	NN	Range	Structure 1	x	Structure 2	interface	ΔG	ΔG	N <sub>HB</sub>	N <sub>SB</sub>	N <sub>DS</sub>	CSS
				N <sub>at</sub> N <sub>res</sub> Surface Å <sup>2</sup>		Symmetry op-n Sym.ID N <sub>at</sub> N <sub>res</sub> Surface Å <sup>2</sup>	area, Å <sup>2</sup>	kcal/mol	P-value				
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	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	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	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.

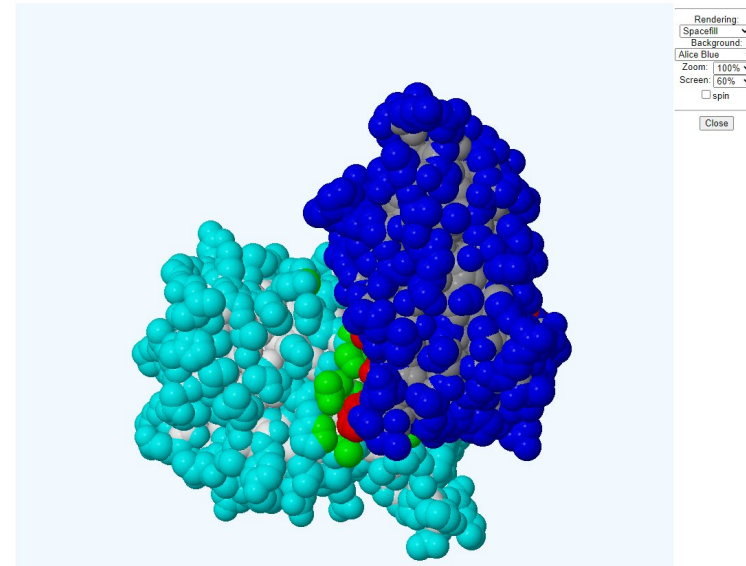
Space symmetry group: P 21 21 2. Resolution: 2.80 Å

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

Interfaces XML View Details Download Search

##		Structure 1					x	Structure 2					interface	$\Delta G$	$\Delta G$	$N_{HB}$	$N_{SB}$	$N_{DS}$	CSS
Id	NN	Range	$N_{at}$	$N_{res}$	Surface Å <sup>2</sup>		Range	Symmetry op-n	Sym.ID	$N_{at}$	$N_{res}$	Surface Å <sup>2</sup>	area, Å <sup>2</sup>	kcal/mol	P-value				
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	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	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	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 Å RESOLUTION IN SPACEGROUP P21212

---

**Interfaces** XML View **Details** Download Search

##	Structure 1	x	Structure 2	interface	$\Delta G$	$\Delta G$	$N_{HB}$	$N_{SB}$	$N_{CS}$	CSS		
Id NN	Range	$N_{at}$ $N_{res}$	Surface $\text{Å}^2$	Range	Symmetry op-n	Sym.ID	$N_{at}$ $N_{res}$	Surface $\text{Å}^2$	area, $\text{Å}^2$	kcal/mol	P-value	
1 1	H	74 18	6228	D	x.v.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 Å 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 Å RESOLUTION IN SPACEGROUP P21212

### Interfaces XML View Details Download Search

##	Structure 1	x	Structure 2	interface	$\Delta G$	$\Delta G$	$N_{HB}$	$N_{SB}$	$N_{DS}$	CSS	
Id	NN	Range	$N_{at}$ $N_{res}$ Surface $\text{\AA}^2$	Range	Symmetry op-n	Sym.ID	$N_{at}$ $N_{res}$ Surface $\text{\AA}^2$	area, $\text{\AA}^2$	kcal/mol	P-value	
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	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	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	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	33	D:GLY 35	0.63	0.00	0.00
34	H:MET 35		0.00	0.00		0.00	34	D:ASP 36	23.04	0.00	0.00
35	H:ARG 36	HS	88.67	88.08		-1.91	35	D:ALA 37	11.42	0.00	0.00
36	H:TRP 37		0.00	0.00		0.00	36	D:THR 38	71.76	0.00	0.00
37	H:TYR 38	H	46.25	44.79		-0.13	37	D:TYR 39	145.65	0.00	0.00
38	H:ARG 39		14.04	0.00		0.00	38	D:GLY 40	1.33	0.00	0.00
39	H:GLN 40		55.42	0.00		0.00	39	D:LYS 41	69.66	0.00	0.00
40	H:ALA 41		25.98	0.00		0.00	40	D:LEU 42	4.85	0.00	0.00
41	H:PRO 42		116.06	0.00		0.00	41	D:THR 43	57.41	0.00	0.00
42	H:GLY 43		87.82	0.00		0.00	42	D:LEU 44	6.18	0.00	0.00
43	H:LYS 44		142.93	0.00		0.00	43	D:LYS 45	58.58	0.00	0.00
44	H:GLU 45		156.25	31.40		-0.38	44	D:PHE 46	0.00	0.00	0.00
45	H:ARG 46		89.45	6.16		-0.01	45	D:ILE 47	20.59	0.00	0.00
46	H:GLU 47		66.28	0.00		0.00	46	D:CYS 48	0.98	0.00	0.00
47	H:TRP 48		83.49	48.83		0.78	47	D:THR 49	76.33	0.00	0.00
48	H:VAL 49		0.00	0.00		0.00	48	D:THR 50	72.72	0.00	0.00
49	H:ALA 50		0.00	0.00		0.00	49	D:GLY 51	44.87	0.00	0.00
50	H:GLY 51		4.94	4.94		0.08	50	D:LYS 52	148.11	0.00	0.00
51	H:MET 52		22.86	3.78		0.05	51	D:LEU 53	1.55	0.00	0.00
52	H:SER 53		26.17	21.48		0.26	52	D:PRO 54	24.33	0.00	0.00
53	H:SER 54		48.27	0.15		-0.00	53	D:VAL 55	0.98	0.00	0.00
54	H:ALA 55		79.34	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 Å RESOLUTION IN SPACEGROUP P21212

---

Interfaces XML View Details Download Search

##	Structure 1			x	Structure 2			interface area, Å <sup>2</sup>	ΔG kcal/mol	ΔG P-value	N <sub>HB</sub>	N <sub>SB</sub>	N <sub>CS</sub>	CSS						
	Id	NN	Range		N <sub>at</sub>	N <sub>res</sub>	Surface Å <sup>2</sup>								Range	Symmetry op-n	Sym.ID	N <sub>at</sub>	N <sub>res</sub>	Surface Å <sup>2</sup>
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	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	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	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
ATOM	10307	O	GLN	H	4	-30.438	34.688	-54.658	1.00	31.43										O

# 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

**PISA Monomer List.**

Session Map (id=729-OC-EEN)

Start **Interfaces** Interface Search

**Monomers**

Assemblies

**Monomers 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 Å RESOLUTION IN SPACEGROUP P21212

---

**Interfacing monomers** XML

##	Id	NN	↔	Range	Class	Structure			Surface		ΔG, kcal/mol
						N <sub>st.</sub>	N <sub>res.</sub>	±N <sub>st.</sub>	±N <sub>res.</sub>	Area, Å <sup>2</sup>	
1	1	⊙		A	Protein	1853	232	1050	221	11064.7	-216.7
	2	○		B	Protein	1803	225	1008	212	10261.5	-213.6
	3	○		C	Protein	1824	228	1017	215	10357.6	-213.4
	4	○		D	Protein	1810	226	1006	213	10328.4	-210.2
	Average:						1822	227	1020	215	10503.0
2	5	○		E	Protein	893	115	483	105	6149.8	-95.5
	6	○		F	Protein	893	115	485	104	6154.2	-95.7
	7	○		G	Protein	901	116	491	106	6343.2	-96.5
	8	○		H	Protein	902	116	489	104	6227.9	-96.3
Average:						897	115	487	104	6218.8	-96.0
3	9	○		[IPA]A:239	Ligand	4	1	4	1	199.5	
	10	○		[IPA]D:239	Ligand	4	1	4	1	198.0	
	11	○		[IPA]F:124	Ligand	4	1	4	1	198.7	
Average:						4	1	4	1	198.7	

[View](#) [Details](#) [Download](#) [View Unit Cell](#) [Download Unit Cell](#)

**Structure similarity** (click dot to view, double click for details)

Range	A	B	C	D	E	F	G	H	[IPA]A:239	[IPA]D:239	[IPA]F:124
A	•	•	•	•	•	•	•	•			
B		•	•	•	•	•	•	•			
C			•	•	•	•	•	•			
D				•	•	•	•	•			
E					•	•	•	•			
F						•	•	•			
G							•	•			
H								•			
[IPA]A:239									•	•	
[IPA]D:239										•	•
[IPA]F:124											•

# 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

PISA: Superposition of 3go:D and 3go:D

Session Map (id=729-OC-EEN)  
Start Interfaces Interface Search  
Monomers  
Assemblies

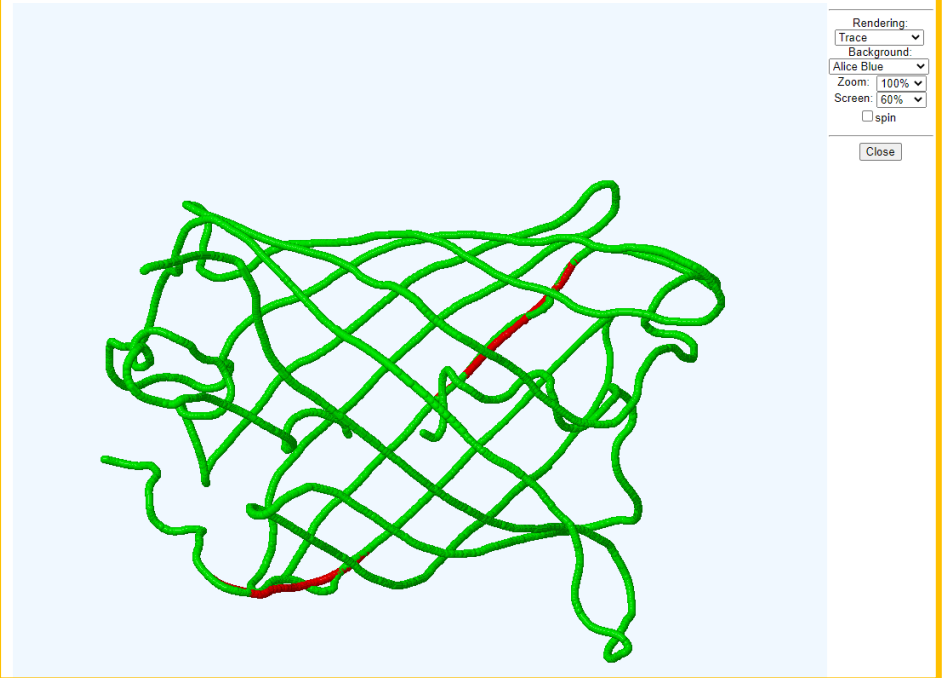
STRUCTURE OF THE GFP:GFP-NANOE

##		Range	Class
1	1	A	Protein
	2	B	Protein
	3	C	Protein
	4	D	Protein
Average: 1			
2	5	E	Protein
	6	F	Protein
	7	G	Protein
	8	H	Protein
Average:			
3	9	[IPA]A:239	Ligand
	10	[IPA]D:239	Ligand
	11	[IPA]F:124	Ligand
Average:			

View Details

Structure sim

Range	A	B	C	D	E	F	G	H	[IPA]A:239	[IPA]D:239	[IPA]F:124
A	..	..	..	..	..	..	..	..	..	..	..
B	..	..	..	..	..	..	..	..	..	..	..
C	..	..	..	..	..	..	..	..	..	..	..
D	..	..	..	..	..	..	..	..	..	..	..
E	..	..	..	..	..	..	..	..	..	..	..
F	..	..	..	..	..	..	..	..	..	..	..
G	..	..	..	..	..	..	..	..	..	..	..
H	..	..	..	..	..	..	..	..	..	..	..
[IPA]A:239	..	..	..	..	..	..	..	..	..	..	..
[IPA]D:239	..	..	..	..	..	..	..	..	..	..	..
[IPA]F:124	..	..	..	..	..	..	..	..	..	..	..



Rendering:  
Trace  
Background: Alice Blue  
Zoom: 100%  
Screen: 60%  
spin  
Close

# 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

PISA: Superposition of 3ogo:D and 3ogo:D

Session Map (id=729-OC-EEN)  
Start Interfaces Interface Search  
Monomers  
Assemblies

STRUCTURE OF THE GFP:GFP-NANOE

##		Range	Class
1	1	A	Protein
	2	B	Protein
	3	C	Protein
	4	D	Protein
Average: 1			
2	5	E	Protein
	6	F	Protein
	7	G	Protein
	8	H	Protein
Average:			
3	9	[IPA]A:239	Ligand
	10	[IPA]D:239	Ligand
	11	[IPA]F:124	Ligand
Average:			

View Details

Structure sim

Range	A	B	C	D	E	F	G	H	[IPA]A:239	[IPA]D:239	[IPA]F:124
A	..	..	..	..	..	..	..	..	..	..	..
B	..	..	..	..	..	..	..	..	..	..	..
C	..	..	..	..	..	..	..	..	..	..	..
D	..	..	..	..	..	..	..	..	..	..	..
E	..	..	..	..	..	..	..	..	..	..	..
F	..	..	..	..	..	..	..	..	..	..	..
G	..	..	..	..	..	..	..	..	..	..	..
H	..	..	..	..	..	..	..	..	..	..	..
[IPA]A:239	..	..	..	..	..	..	..	..	..	..	..
[IPA]D:239	..	..	..	..	..	..	..	..	..	..	..
[IPA]F:124	..	..	..	..	..	..	..	..	..	..	..



# 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?*

