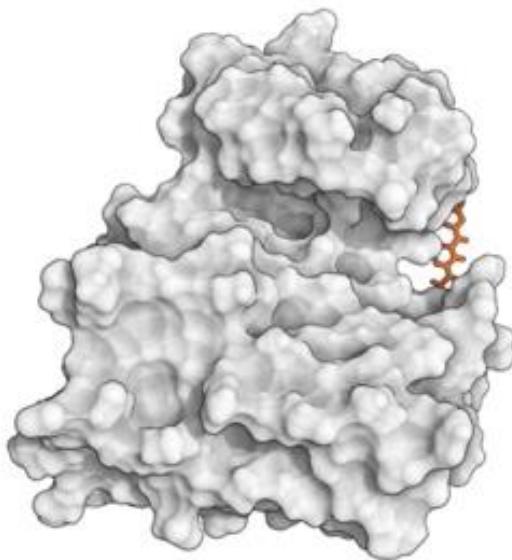
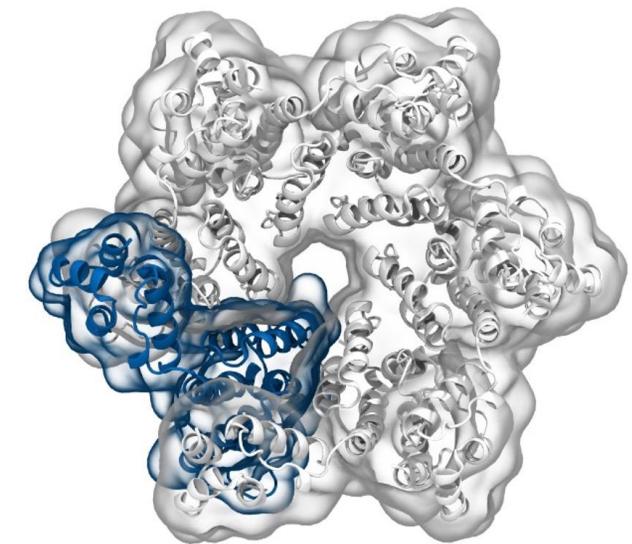


Simulation of Biomolecules



Protein preparation



Dr Matteo Degiacomi

Durham University

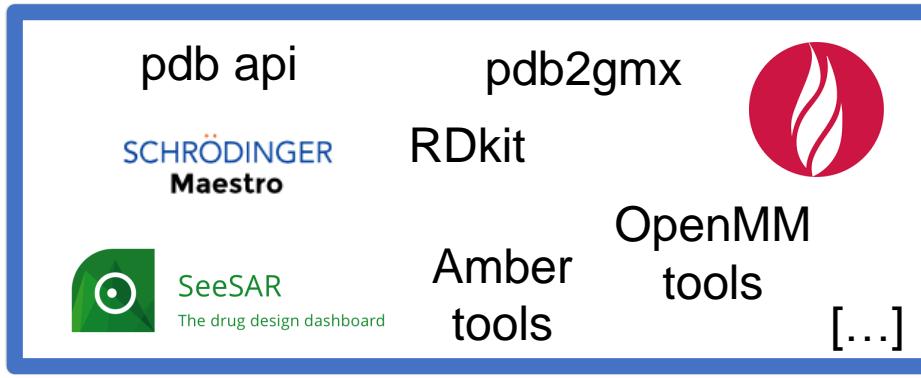
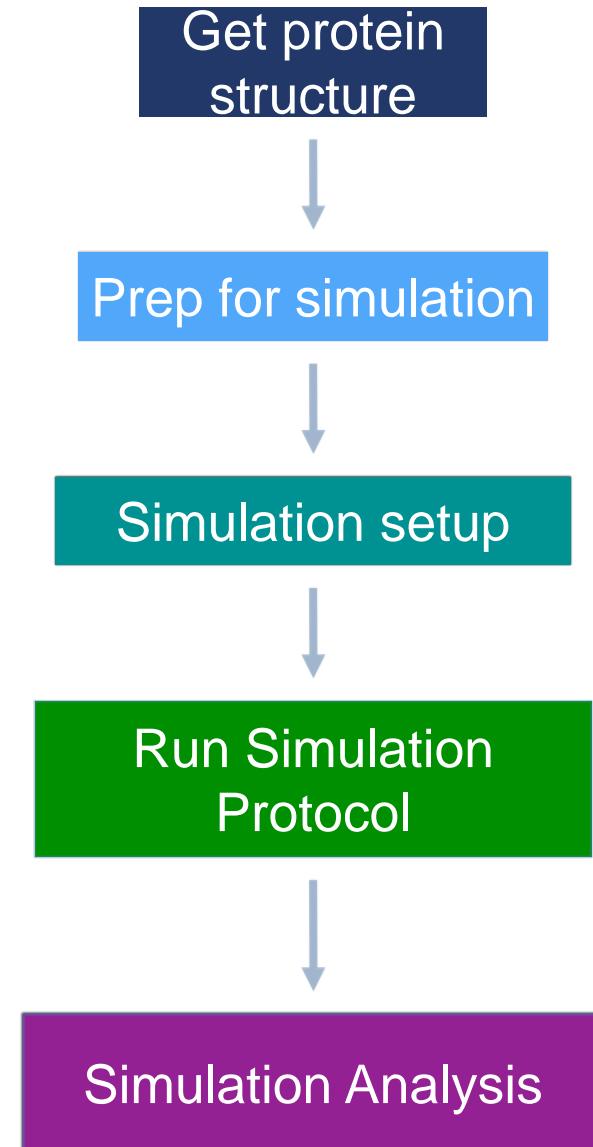
matteo.t.degiacomi@durham.ac.uk

Dr Antonia Mey

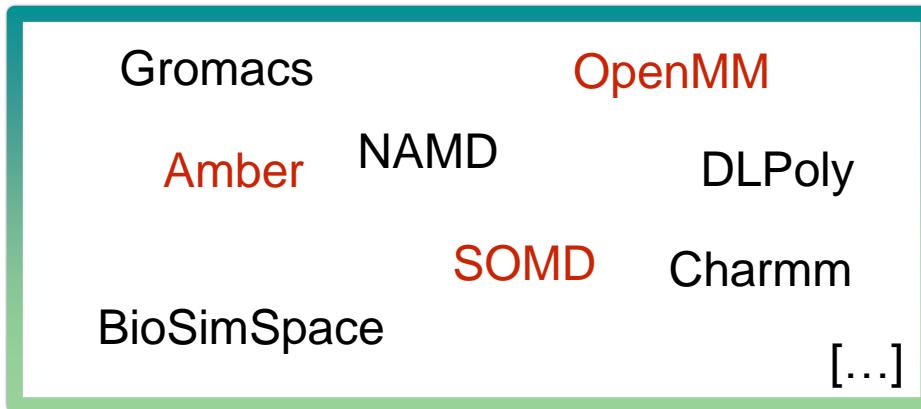
University of Edinburgh

antonia.mey@ed.ac.uk

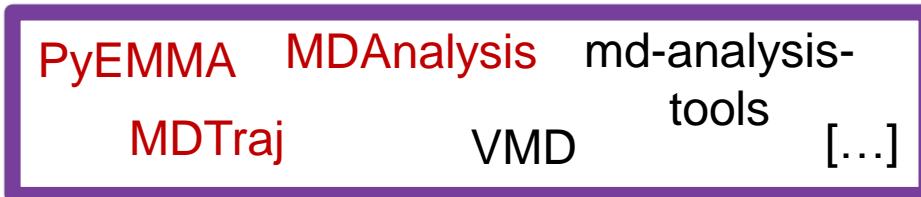
A typical workflow for molecular dynamics



*GUI, TCL, bash,
Python, Perl ...
get creative*

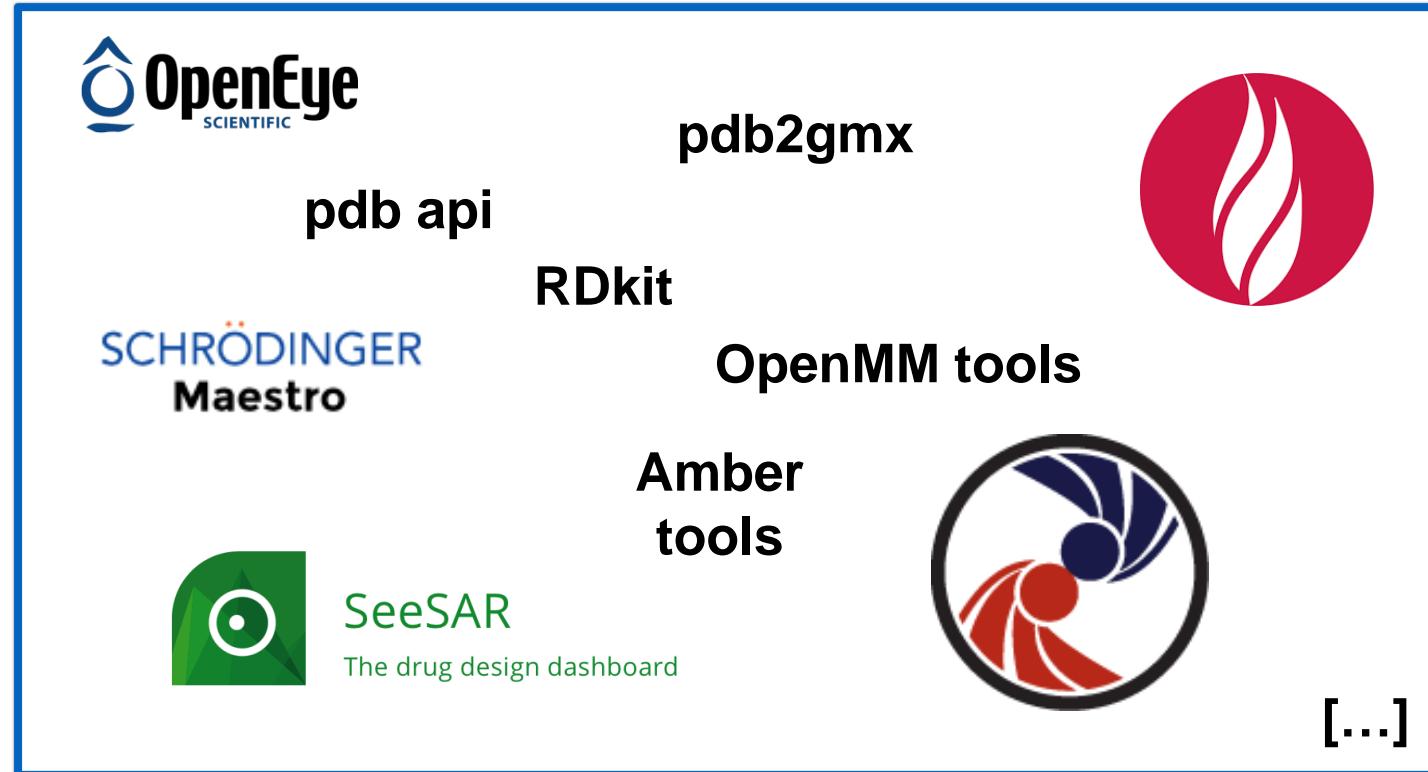
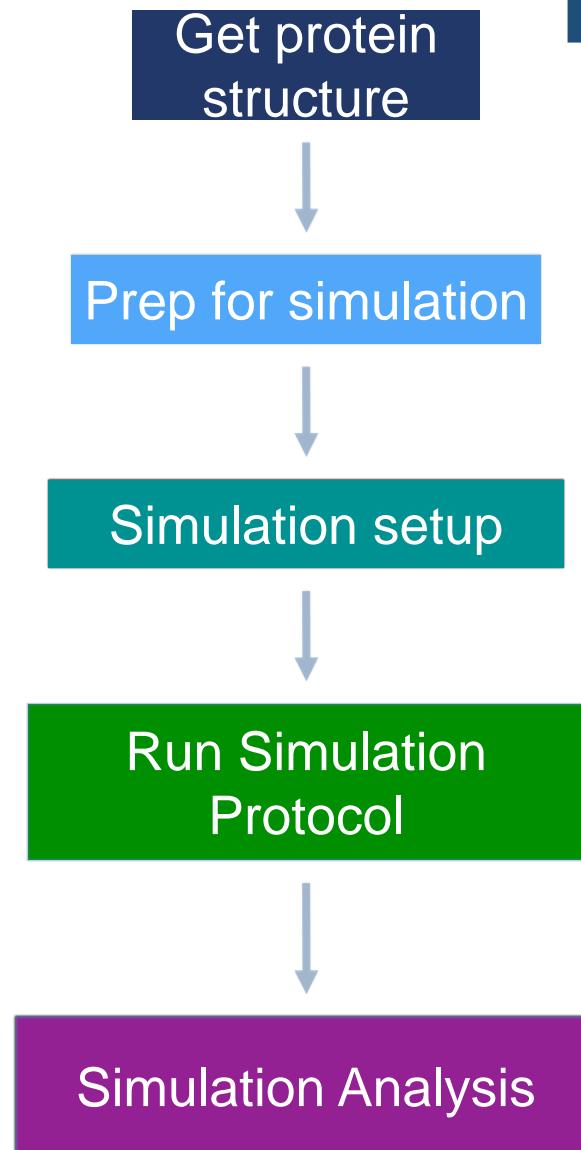


*mostly C++
command line*



*GUI, TCL, bash,
Python, Perl, ...
get creative*

Let's get started with understanding protein structures



GUI, TCL, bash, python, Perl, ... get creative

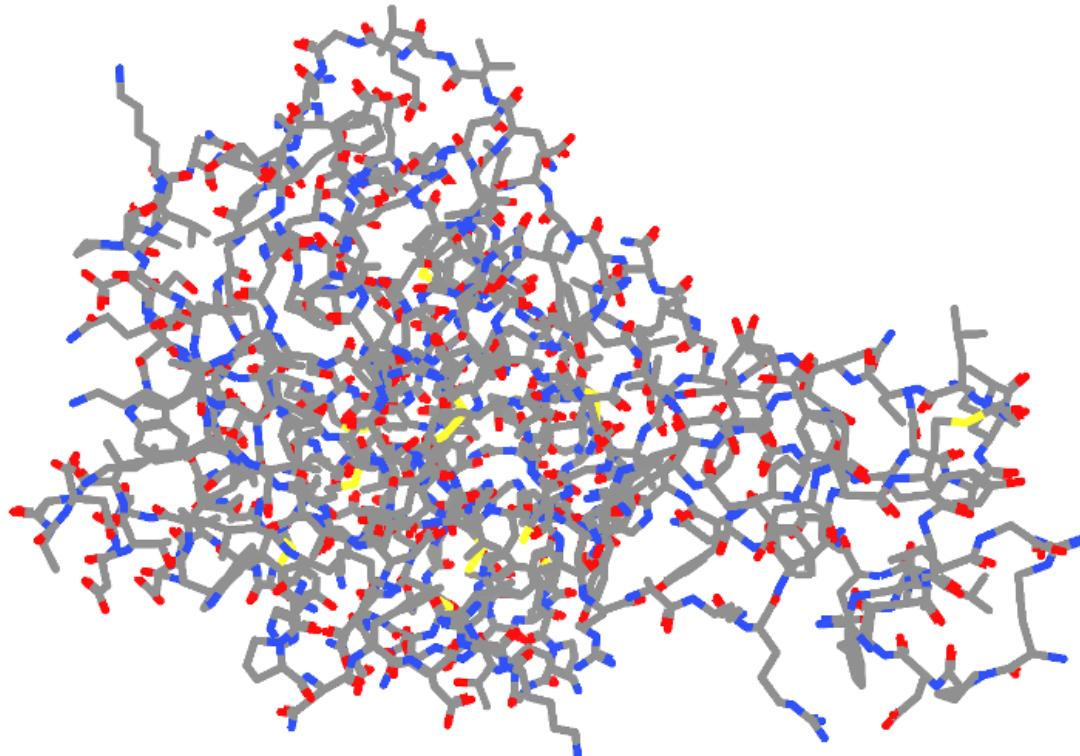
What things are important to look out for in protein prep?

<https://www.menti.com/alc5aozap924>

58950948



Crystal structures provide electron densities



Crystal structures are models with potential errors

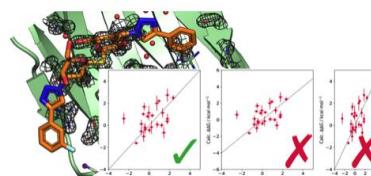
[HOME](#) / [ARCHIVES](#) / VOL. 4 NO. 1 (2022) / [Articles](#)

Best Practices for Constructing, Preparing, and Evaluating Protein-Ligand Binding Affinity Benchmarks [Article v1.0]

David F. Hahn

Computational Chemistry, Janssen Research & Development, Turnhoutseweg 30,
Beerse B-2340, Belgium

<https://orcid.org/0000-0003-2830-6880>



Examples of errors/oddities in PDB structures

<https://swift.cmbi.umcn.nl/teach/pdbad/>

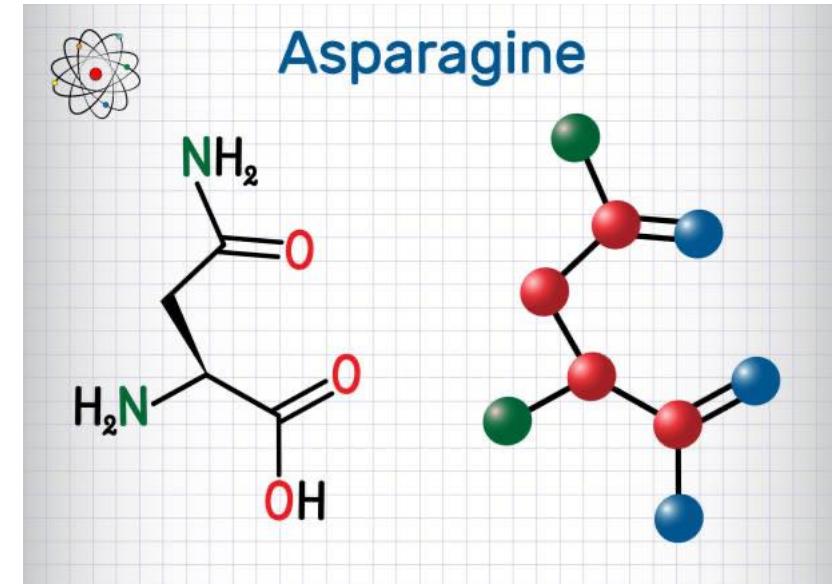


Gert Vriend (author of WHAT_CHECK)

Crystal structures are models derived from electron densities

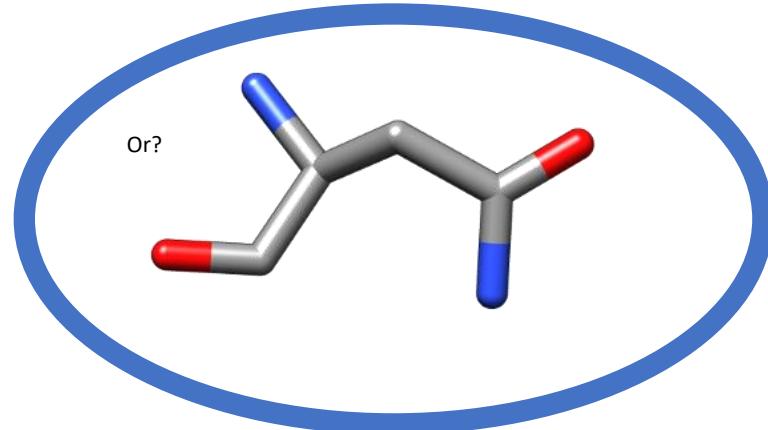
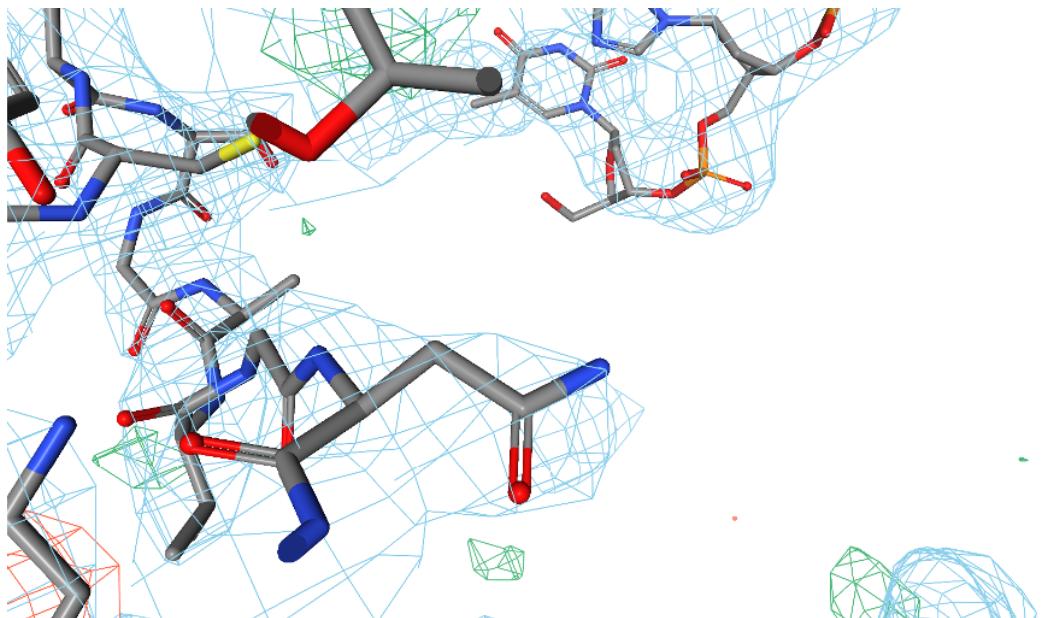
1T38: HUMAN O6-ALKYLGUANINE-DNA ALKYLTRANSFERASE

ASN157



Crystal structures are models derived from electron densities

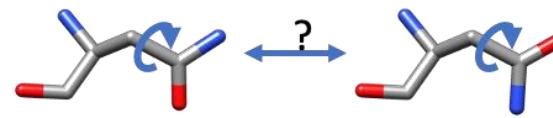
1T38: HUMAN O6-ALKYLGUANINE-DNA ALKYLTRANSFERASE



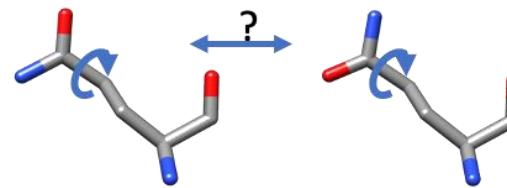
Alternative conformations of side chains - NGH flips

Typically, the crystallographers will have assigned the orientation based on potential H-bonding interactions, etc.

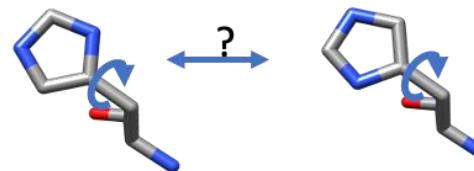
No standard protocol for this: always worth double-checking.



Asparagine (N)



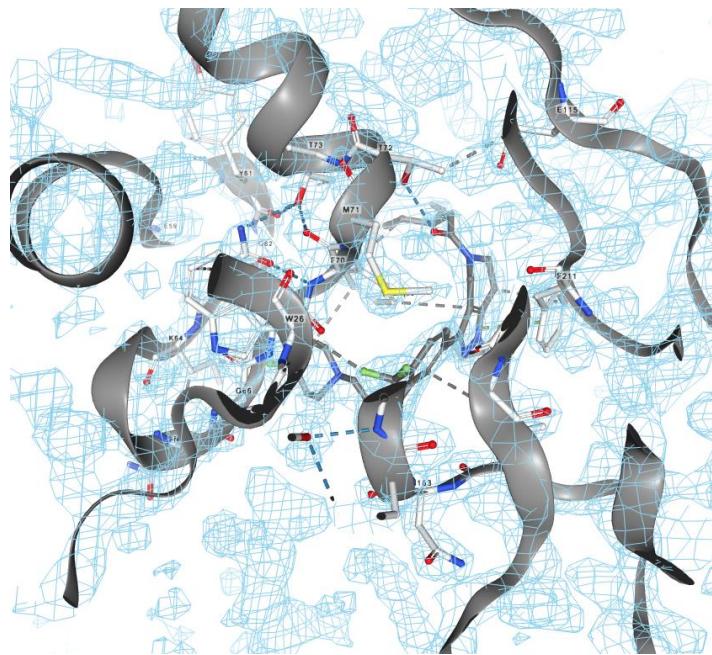
Glutamine (Q)



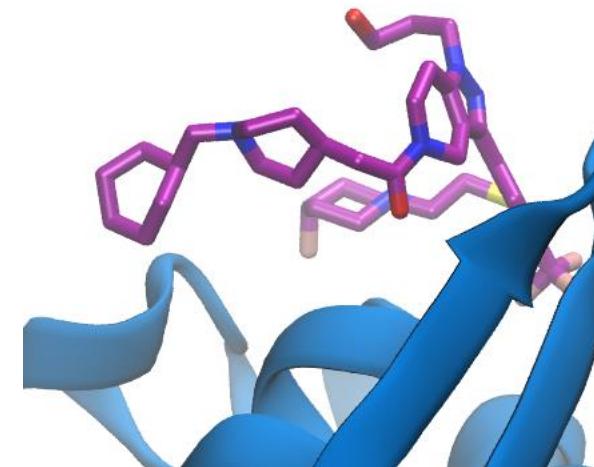
Histidine (H)

Ligand densities can also have creative input from the crystallographer

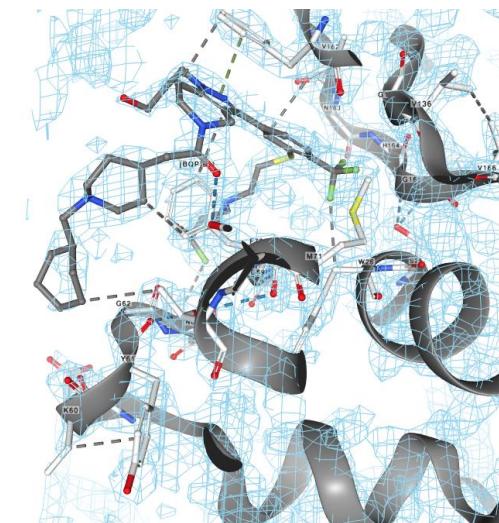
Cathepsin S



Cathepsin S
with ligand

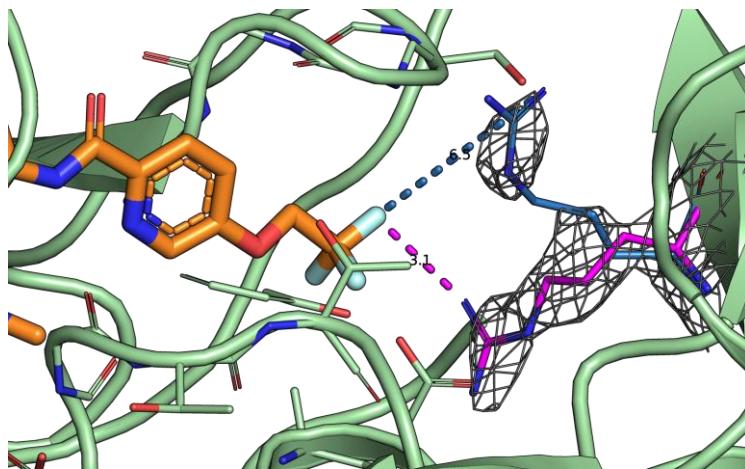


boat/chair?

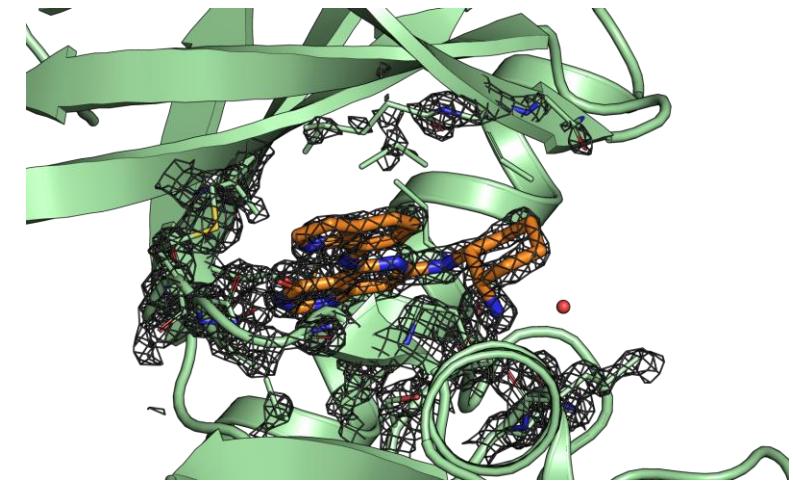
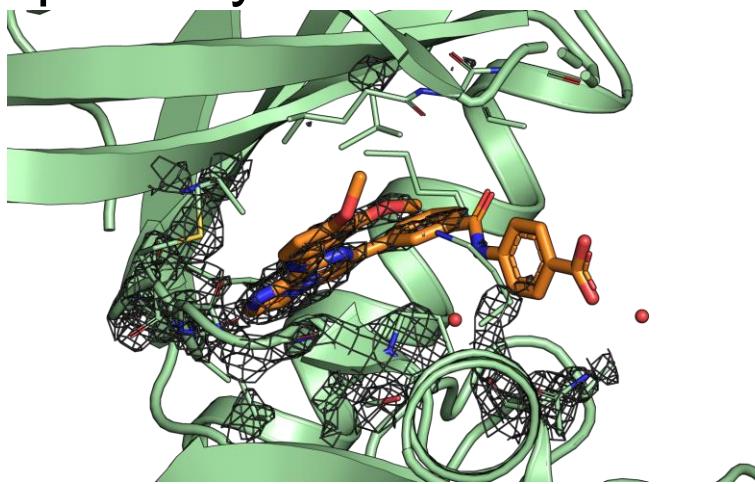


Picking the best crystal structure requires care

BACE (Hunt)



spleen tyrosine kinase

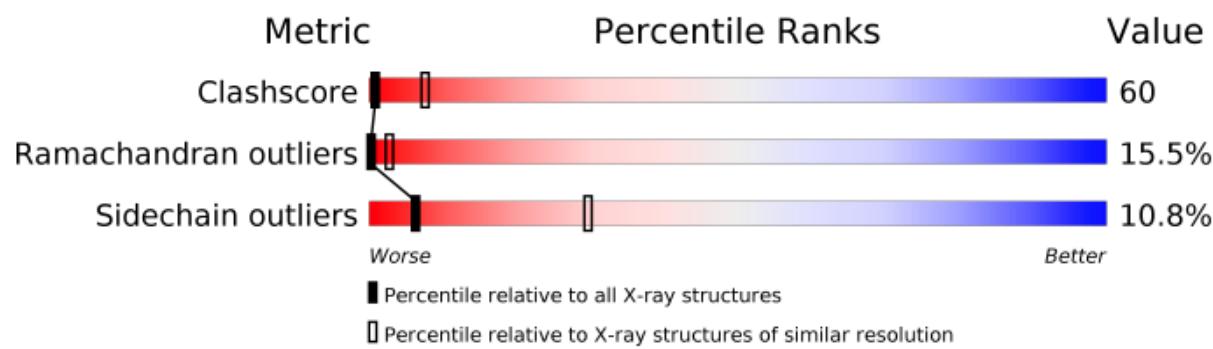


Which protein structure?

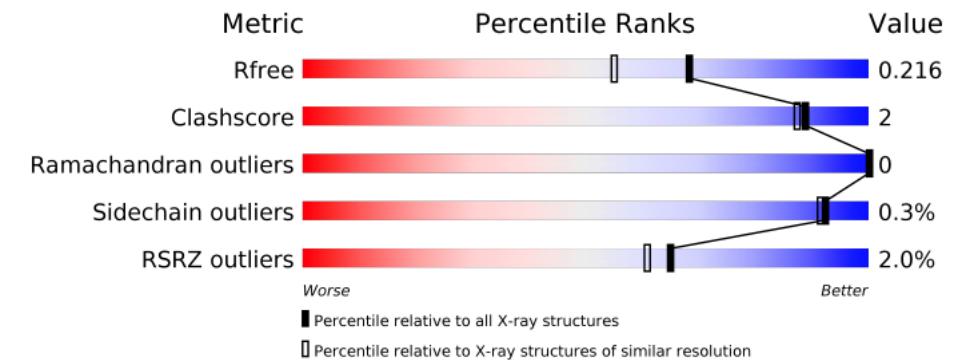
- Alternative side chain conformations need to be assessed carefully
- Active site residue densities are important for choosing the right crystal structure

The RCSB PDB report can help with choosing structures

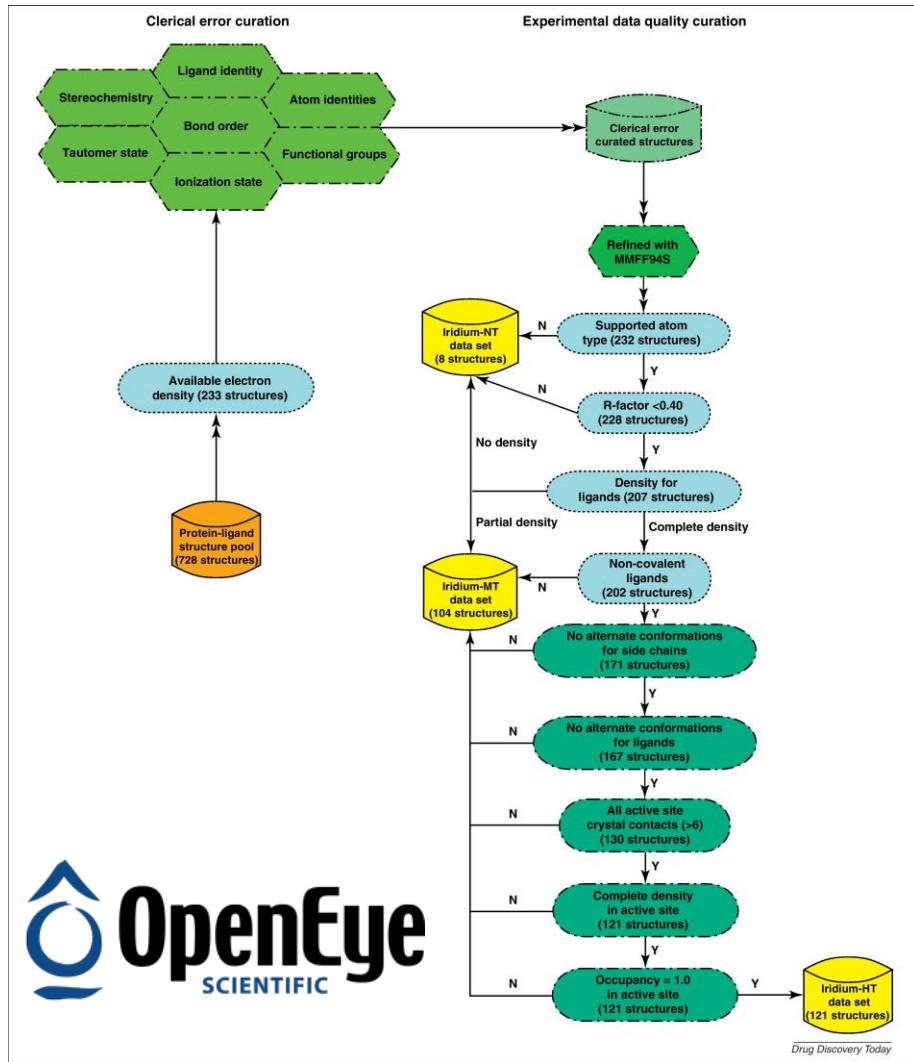
Jnk1 - 2GMX



Jnk1 - 3ELJ



The Iridium score can help assess the trustworthiness of an X-ray structure



Warren et al., *Drug Discovery Today*, 2012

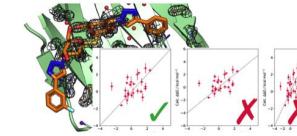
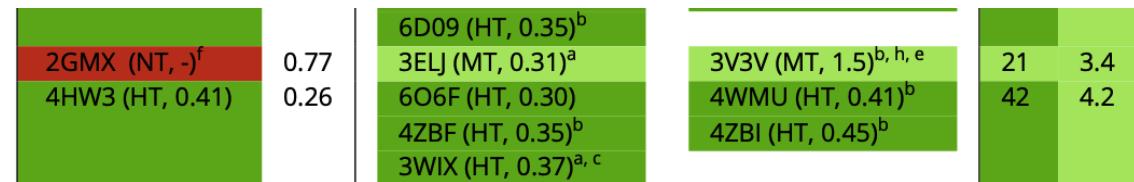
HOME / ARCHIVES / VOL. 4 NO. 1 (2022) / Articles

Best Practices for Constructing, Preparing, and Evaluating Protein-Ligand Binding Affinity Benchmarks [Article v1.0]

David F. Hahn

Computational Chemistry, Janssen Research & Development, Turnhoutseweg 30, Beerse B-2340, Belgium
<https://orcid.org/0000-0003-2830-6880>

Jnk1[[74](#), [86](#)]
MCL1[[74](#), [87](#)]



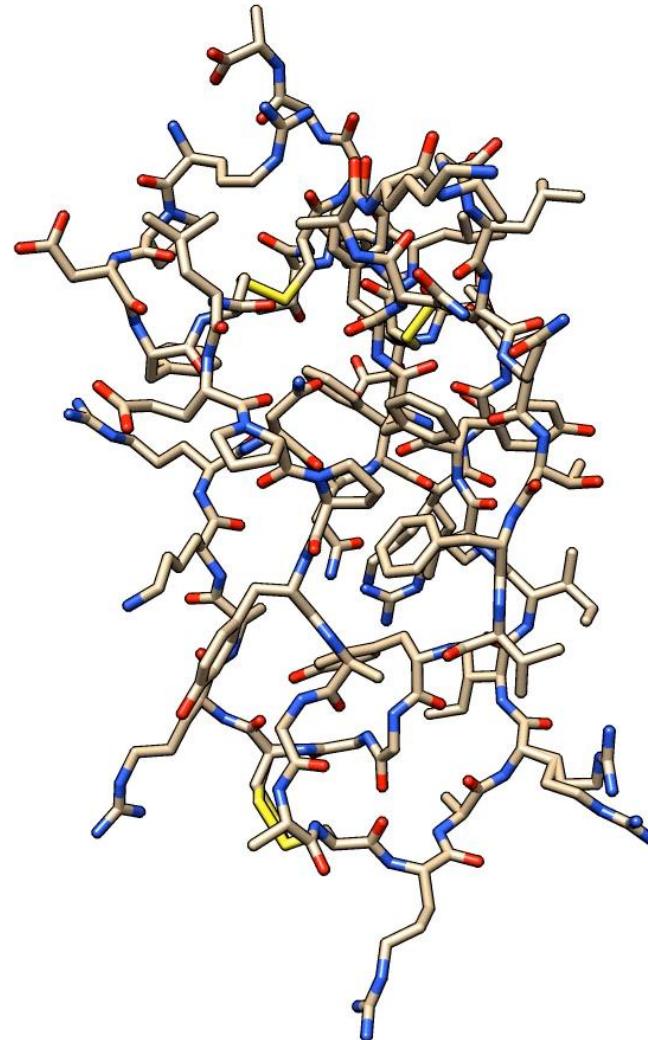
12 Features to inform the Iridium Score:

- R-free value
- Resolution
- Density coverage of ligand heavy atoms
- Active site density coverage
- Alternative locations active site/ligands
- [...]

	HT	MT	NT
Ligand	>0.9	<0.9 and >0.5	<0.5
Active Site	>0.95	<0.95 and >0.5	<0.5

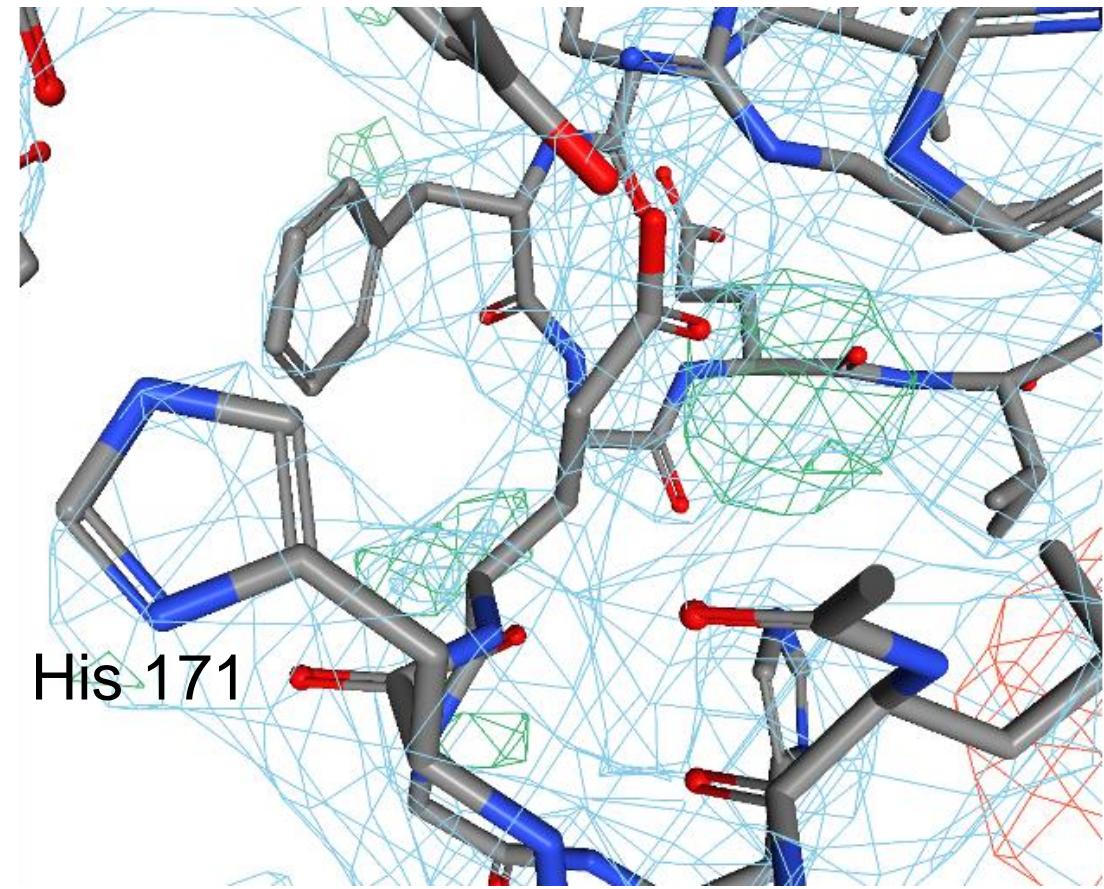
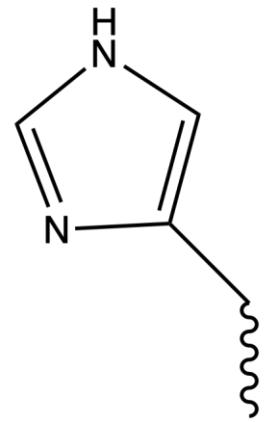
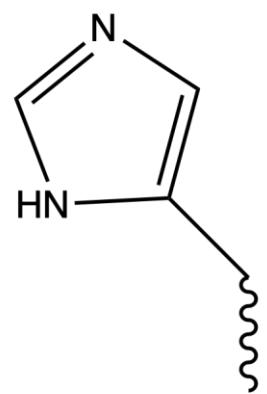
Missing information in structures: Hydrogen atoms

- Missing from most crystal structures
- Needed for molecular modelling
- Most MD packages (AMBER, CHARMM, GROMACS, etc.) include tools to “automatically” add H-atoms.
- Any issues?



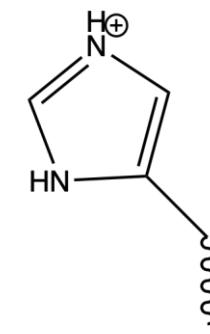
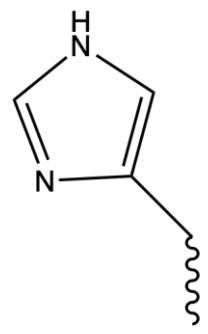
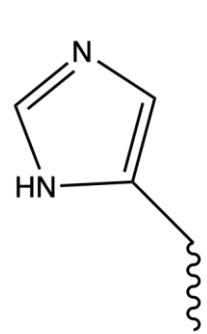
How to choose the right tautomer?

Which tautomer?



How to choose the right tautomer and protonation state?

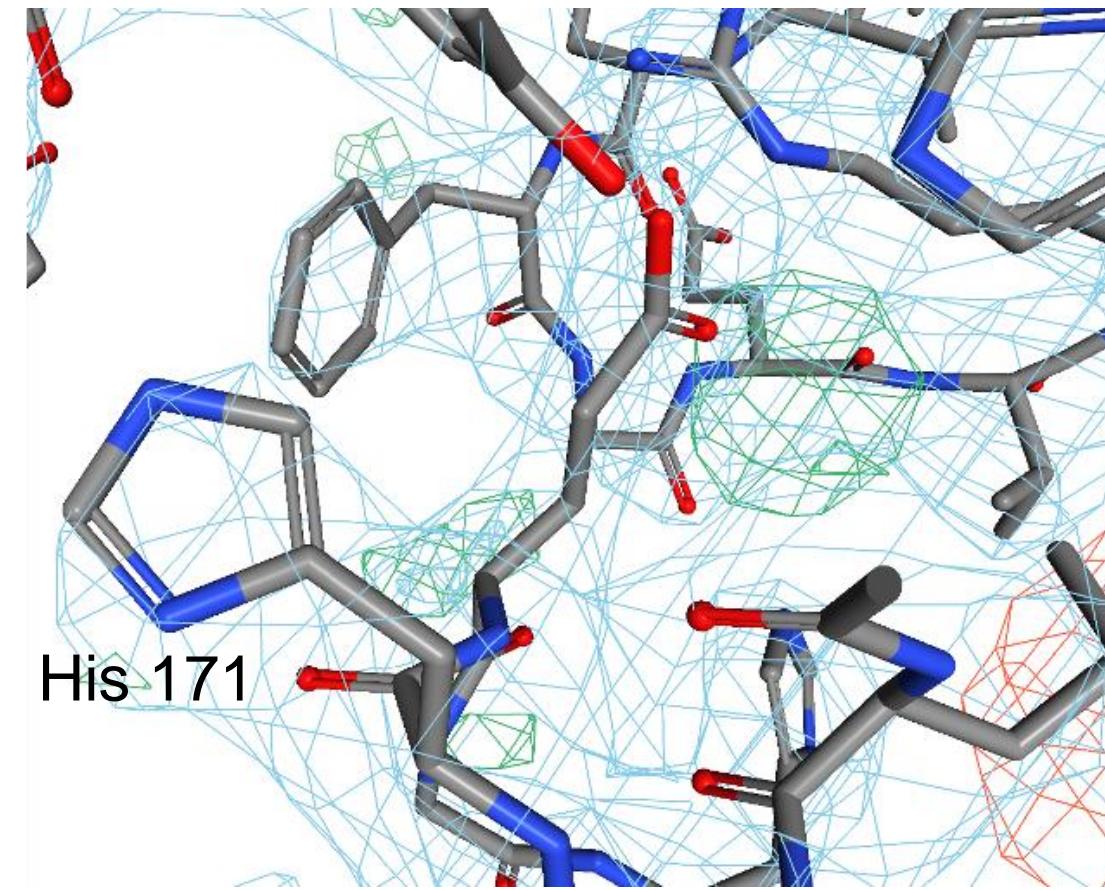
Which tautomer and protonation?



δ -tautomer

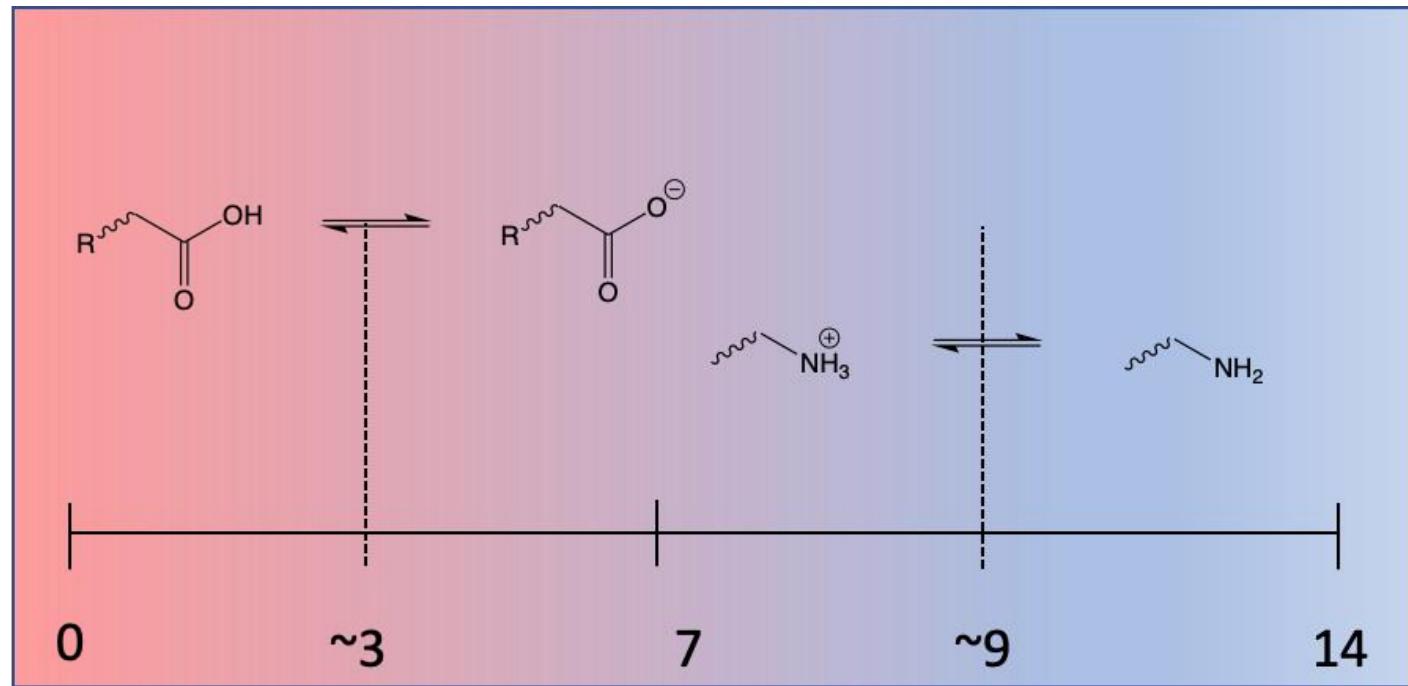
ϵ -tautomer

Protonated



We need to worry about pKa to decide protonation

- pKa: pH at which an acidic/basic group is 50% protonated/deprotonated.
- pKas are not fixed things!
- “Standard” values refer to the situation when the group is in dilute aqueous solution.
- Groups buried in the centre of hydrophobic proteins or close to other charged groups can show large pKa shifts.

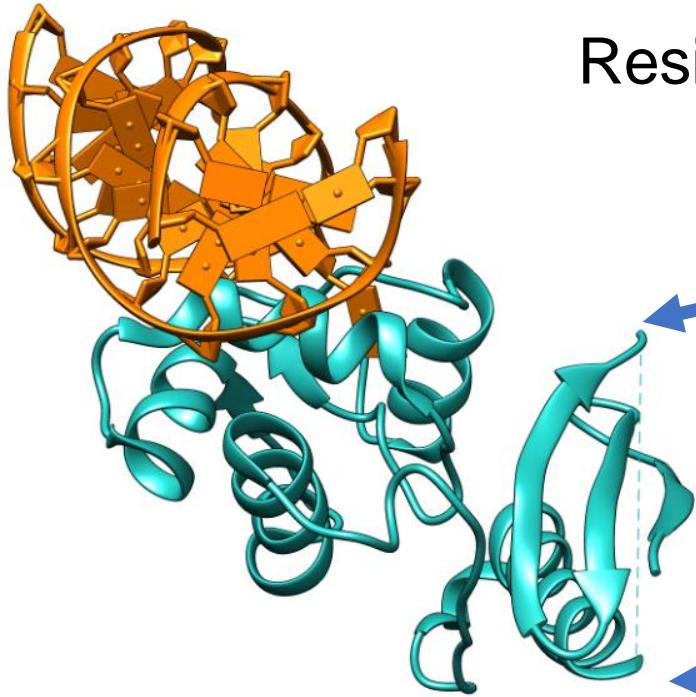


Amino acids that need protonation state consideration

Amino acid	pKa	options	Significance*
Aspartic acid	3.65	-COOH instead of COO ⁻ ?	possible
Glutamic acid	4.25	-COOH instead of COO ⁻ ?	possible
Histidine	6.00	protonated instead of neutral?	very possible
Cysteine	8.18	-S ⁻ instead of SH?	very possible
Tyrosine	10.07	-O ⁻ instead of OH?	possible
Lysine	10.53	-NH ₂ instead of NH ₃ ⁺ ?	possible
Arginine	12.48	neutral instead of protonated?	unlikely

*for a simulation around physiological pH

What if the best structure has missing residues?



Residues 36 to 55 missing

```
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465 M RES C SSSEQI
```

```
REMARK 465      LYS A    36
REMARK 465      GLY A    37
REMARK 465      THR A    38
REMARK 465      SER A    39
REMARK 465      ALA A    40
REMARK 465      ALA A    41
REMARK 465      ASP A    42
REMARK 465      ALA A    43
REMARK 465      VAL A    44
REMARK 465      GLU A    45
REMARK 465      VAL A    46
REMARK 465      PRO A    47
REMARK 465      ALA A    48
REMARK 465      PRO A    49
REMARK 465      ALA A    50
REMARK 465      ALA A    51
REMARK 465      VAL A    52
REMARK 465      LEU A    53
REMARK 465      GLY A    54
REMARK 465      GLY A    55
```

Modeller

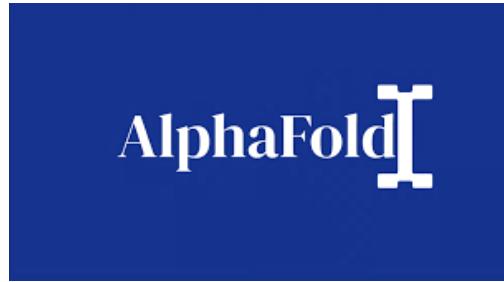
Program for Comparative Protein
Structure Modelling by Satisfaction
of Spatial Restraints



**PDB
FIXER**

AlphaFold

AlphaFold2 structures have no missing residues or atoms

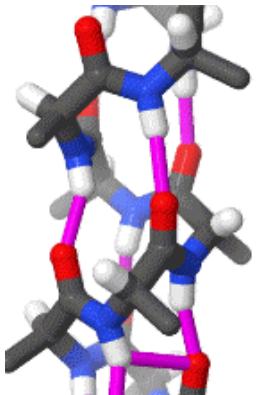


G T R Y A G K V V

G T R K V V



No missing residues

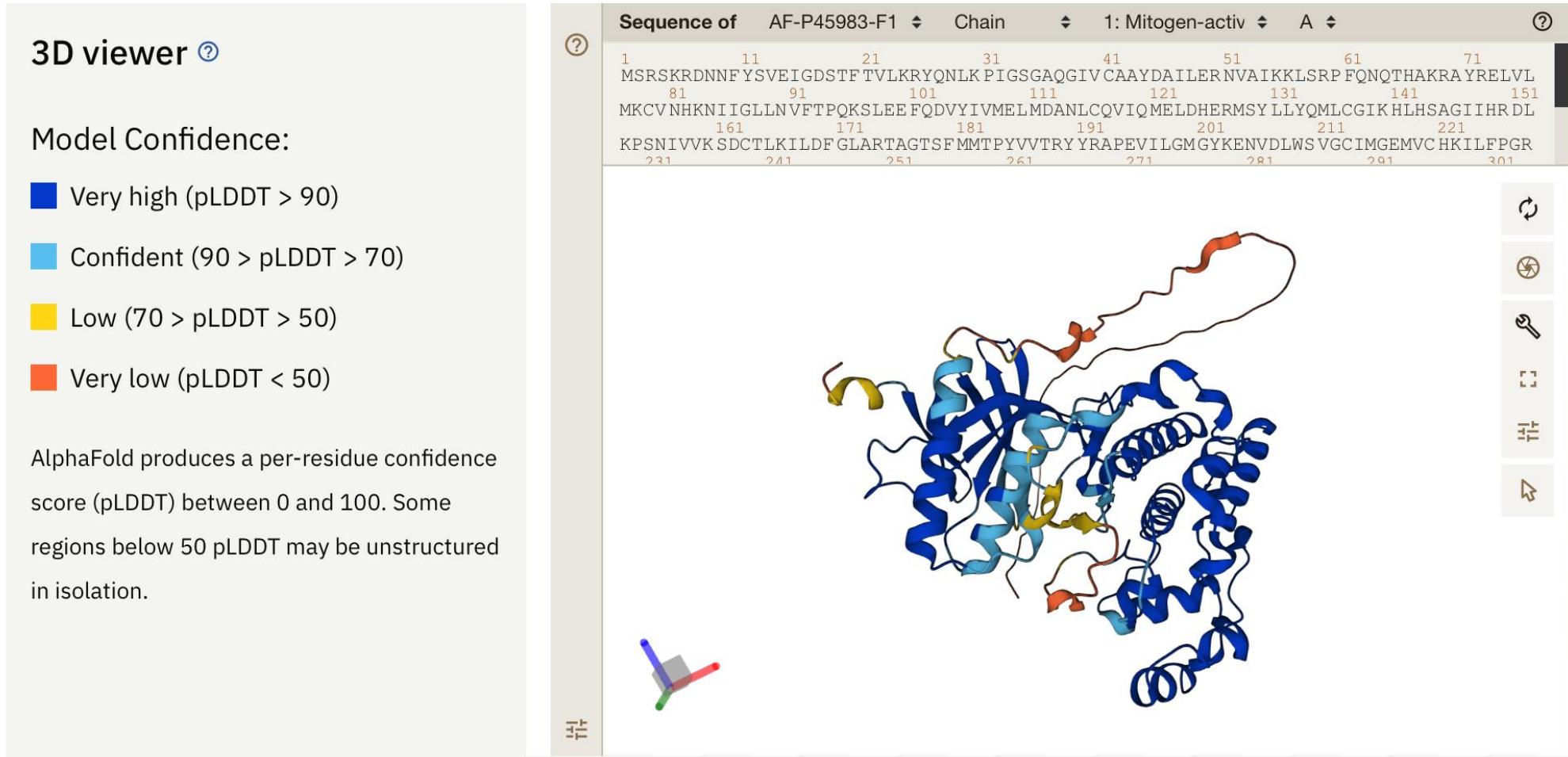


Has hydrogens



AlphaFold2 structures can have unreliable areas

JNK1 — again



What is missing in this very good structure?

Alcohol dehydrogenase

3D viewer ⓘ

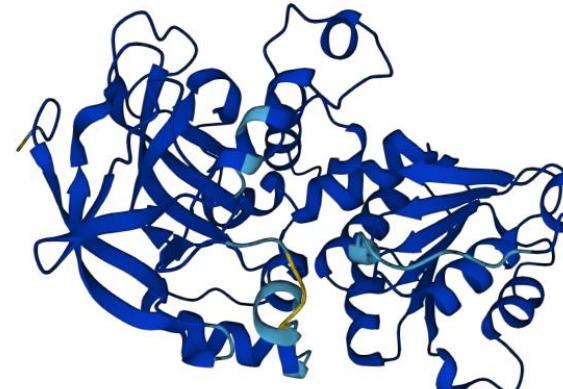
Model Confidence:

- 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 below 50 pLDDT may be unstructured in isolation.

Sequence of AF-P28332-F1 ⌂ Chain ⌂ 1: Alcohol dehyd ⌂ A ⌂ ⓘ

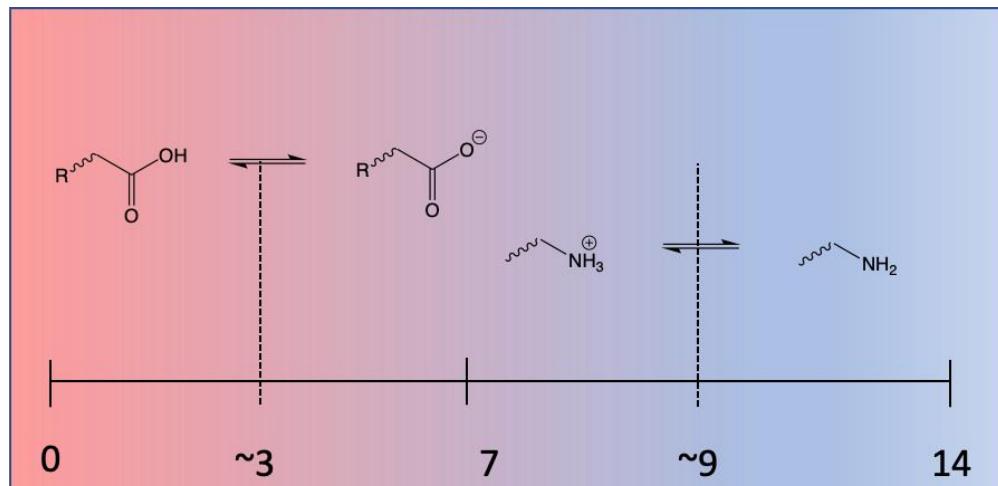
1	11	21	31	41	51	61	71
MSTTGQVIRCKAAILWKPGAPFSIEEVEA	PPKAKEVRIKV	VVATGLCGTE	MKVVLGSKHLD	LLYPTILGHE	GAGIVE		
81	91	101	111	121	131	141	151
SIGEGVSTVKPGDKVITLFLPQC	ECTSCLNSEG	NFCIQFKQSK	TQLMSDGTSR	FTCKGKS	YHFGNTSTFCEY	TV	
161	171	181	191	201	211	221	
IKEISVAKIDAVAPLEKV	CLISCGFSTG	FGAAINTAKV	TPGSTCAVFG	LGGVGLSVVM	GCKAAGAARI	IGVDVNKE	
231	241	251	261	271	281	291	301



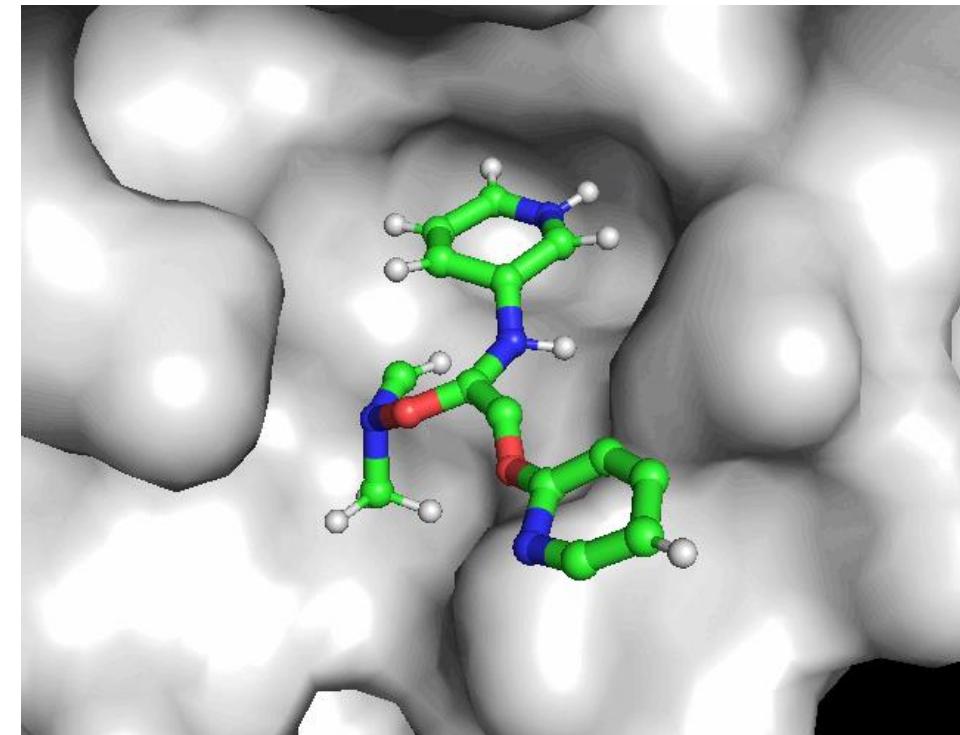
The 3D viewer interface shows a detailed ribbon model of the Alcohol dehydrogenase protein. The model consists of several blue-colored alpha-helices and beta-sheets forming a complex tertiary structure. A small yellow segment is visible near the bottom center, likely representing a region with lower confidence. The interface includes various controls for rotation, zoom, and selection on the right side.

What about ligands and co-factors?

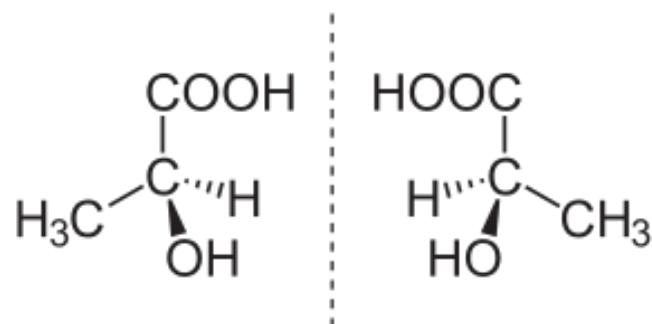
pKa for ligands is also important!



How do you get a ligand position when there is no crystal structure?



Choosing the right enantiomer



Trypsin: 1TRN

Task 1:

You have been told you to simulate Trypsin starting from the PDB code *1TRN*. Have a look at the *1TRN* entry in the pdb and note down what you see and choices you make. Think about:

- Stoichiometry (monomer or dimer protein?)
- PTMs (does it have post-translational modifications?)
- Non-protein molecules (simulate or remove?)
- Disulphides (does it have potential disulphide bridges?)
- Histidines (what will be the protonation state of histidines?)

<https://www.rcsb.org/structure/1TRN>

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

Biological Assembly 1 ?

Display Files Download Files Data API

1TRN

CRYSTAL STRUCTURE OF HUMAN TRYPSIN 1: UNEXPECTED PHOSPHORYLATION OF TYROSINE 151

PDB DOI: <https://doi.org/10.2214/pdb1TRN/pdb>

Classification: HYDROLASE (SERINE PROTEINASE)

Organism(s): Homo sapiens

Mutation(s): No ⓘ

Deposited: 1995-03-16 Released: 1995-06-03

Deposition Author(s): Gaboriaud, C., Fontecilla-Camps, J.C.

Explore in 3D: Structure | Sequence Annotations | Electron Density | Validation Report | Ligand Interaction (ISP)

Global Symmetry: Asymmetric - C1 ⓘ

Global Stoichiometry: Homo 2-mer - A2 ⓘ

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.20 Å

R-Value Work: 0.177

R-Value Observed: 0.177

wwPDB Validation ⓘ

3D Report Full Report

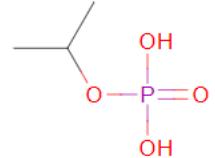
Metric	Percentile Ranks	Value
Clashscore	7	7
Ramachandran outliers	0	0
Sidechain outliers	5.4%	5.4%
RSRZ outliers	0.2%	0.2%

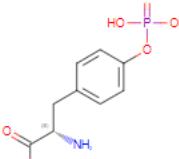
Worse Better

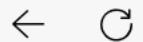
Percentile relative to all X-ray structures

Percentile relative to X-ray structures of similar resolution

Small Molecules

Ligands 1 Unique				
ID	Chains ⓘ	Name / Formula / InChI Key	2D Diagram	3D Interactions
ISP Query on ISP	C [auth A], D [auth B]	PHOSPHORYLISOPROPANE C ₃ H ₉ O ₄ P QPPQHRDVVPBTVEV-UHFFFAOYSA-N		Interactions ▾ Interactions & Density ▾
Download Ideal Coordinates CCD File ⓘ				

Modified Residues ⓘ 1 Unique					
ID	Chains ⓘ	Type	Formula	2D Diagram	Parent
PTR Query on PTR	A, B	L-PEPTIDE LINKING	C ₉ H ₁₂ N O ₆ P		TYR



https://www.rcsb.org/structure/1TRN



RCSB PDB

Deposit ▾ Search ▾ Visualize ▾ Analyze ▾ Download ▾ Learn ▾ About ▾ Documentation ▾ Careers COVID-19

MyPDB ▾

Contact us

Find proteins for [P07477](#) (*Homo sapiens*)

Explore [P07477](#) ⓘ

Go to UniProtKB: [P07477](#)

GTEX: [ENSG00000204983](#)

Entity Groups ⓘ

Sequence Clusters

[30% Identity](#) [50% Identity](#) [70% Identity](#) [90% Identity](#) [95% Identity](#) [100% Identity](#)

UniProt Group

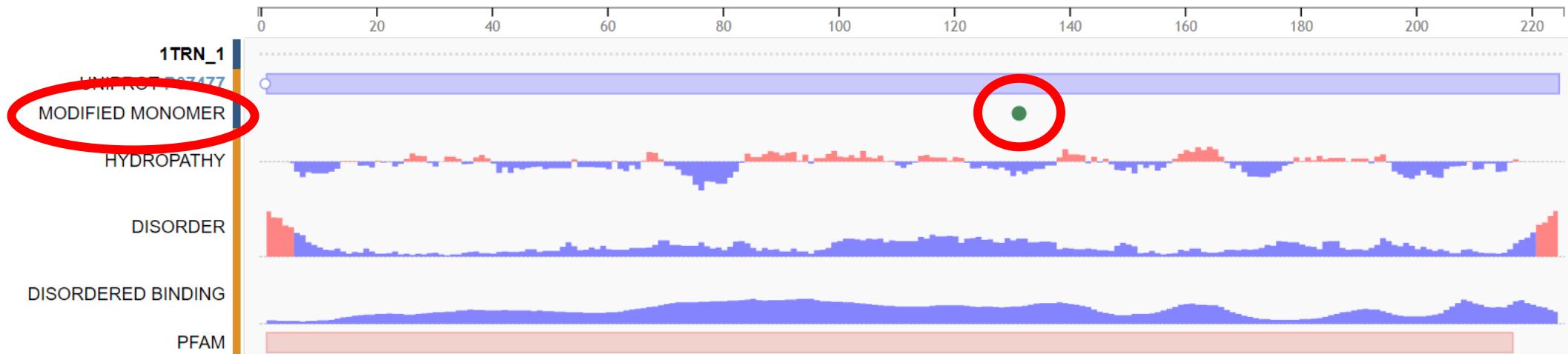
[P07477](#)

Sequence Annotations

[Expand](#)

Reference Sequence

1TRN_1



Function	Protein ⁱ	Serine protease 1	Amino acids	247 (go to sequence)
Names & Taxonomy	Gene ⁱ	PRSS1	Protein existence ⁱ	Evidence at protein level
Subcellular Location	Status ⁱ	 UniProtKB reviewed (Swiss-Prot)	Annotation score ⁱ	5/5
	Organism ⁱ	Homo sapiens (Human)		

Disease & Variants

PTM/Processing

Expression

Interaction

Structure

Family & Domains

Sequence

Similar Proteins

Entry Variant viewer 524 Feature viewer Genomic coordinates Publications External links Hi

[BLAST](#) [Download](#) [Add](#) [Add a publication](#) [Entry feedback](#)

Functionⁱ

Has activity against the synthetic substrates Boc-Phe-Ser-Arg-Mec, Boc-Leu-Thr-Arg-Mec, Boc-Gln-Ala-Arg-Mec and Boc-Val-Pro-Arg-Mec. The single-chain form is more active than the two-chain form against all of these substrates.

1 Publication

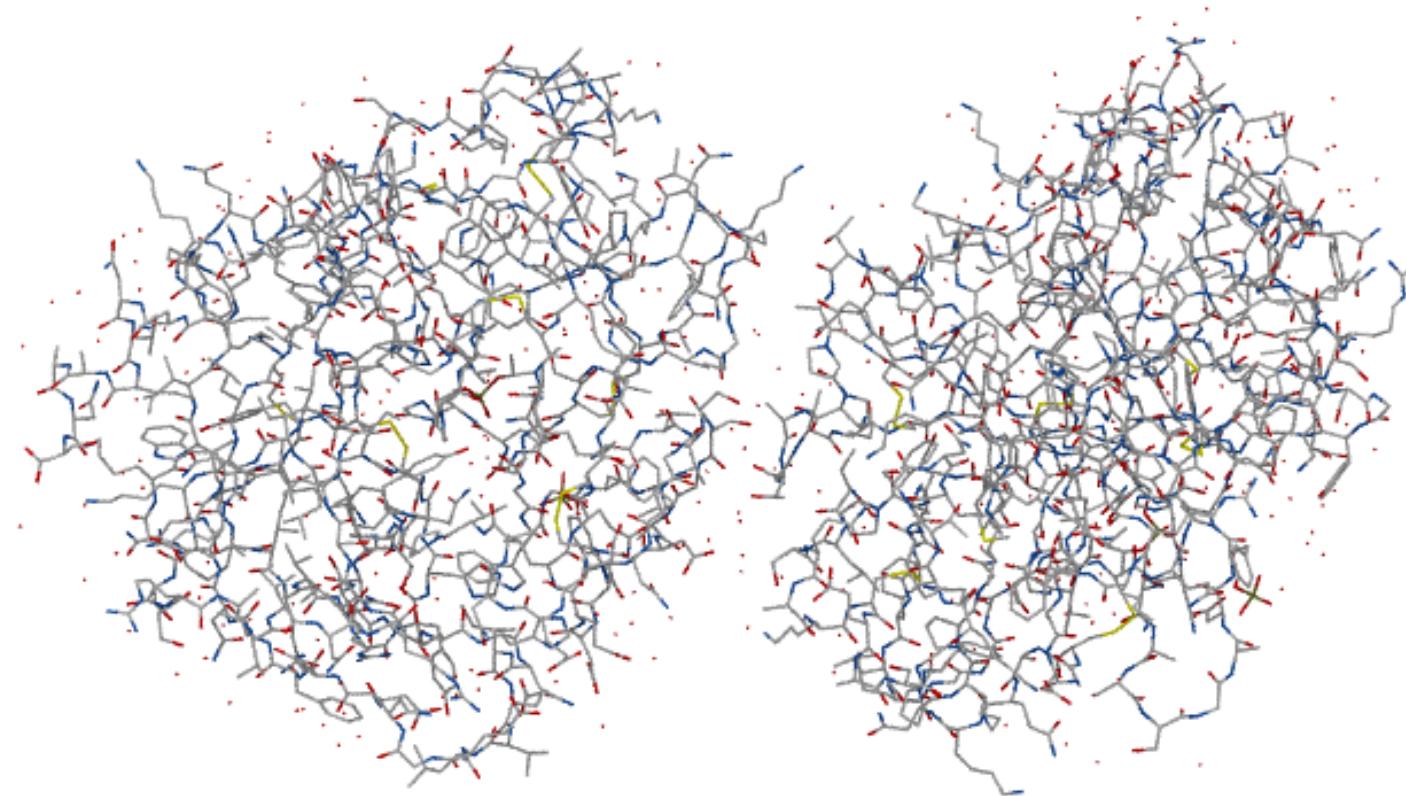
Caution

Tyr-154 was proposed to be phosphorylated (PubMed:[8683601](#)) but it has been shown (PubMed:[17087724](#)) to be sulfated instead.

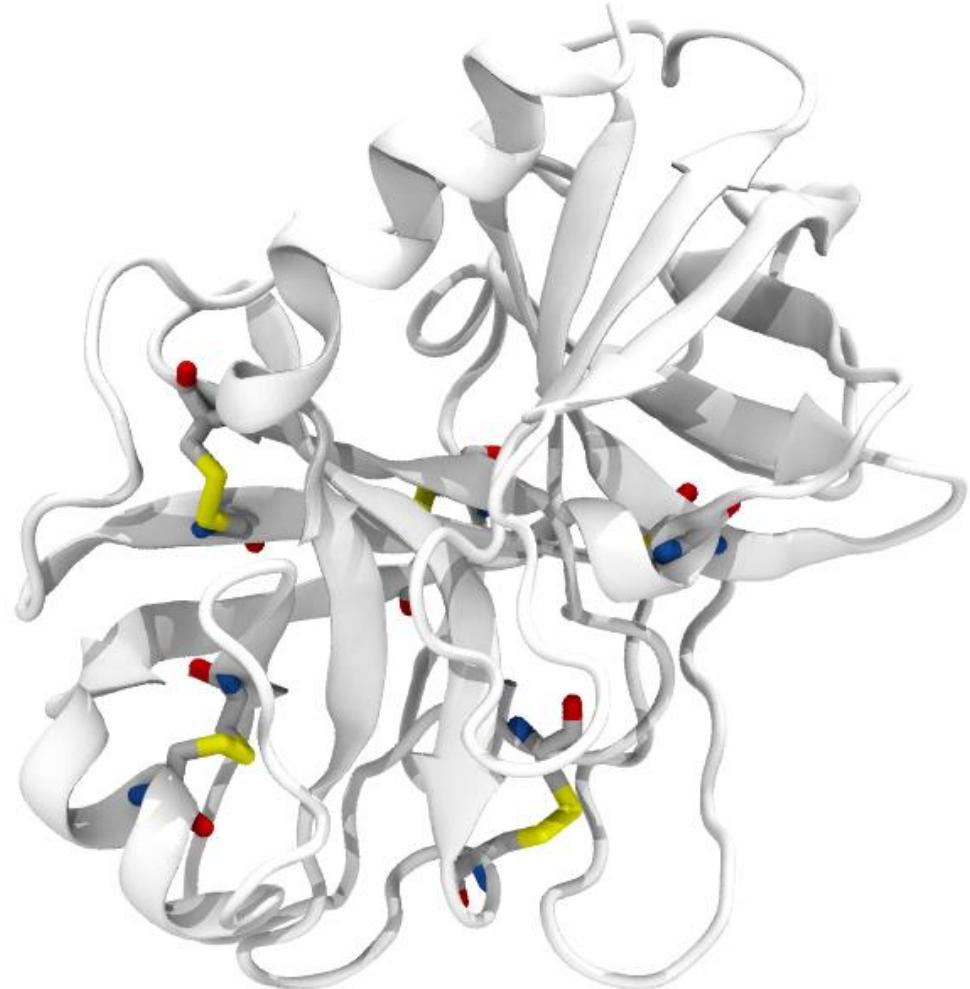
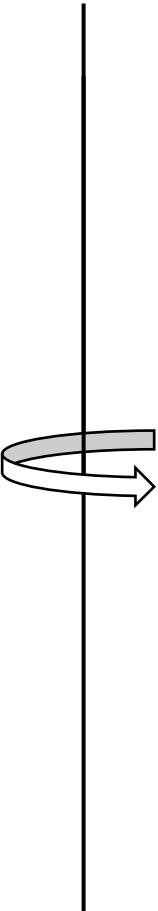
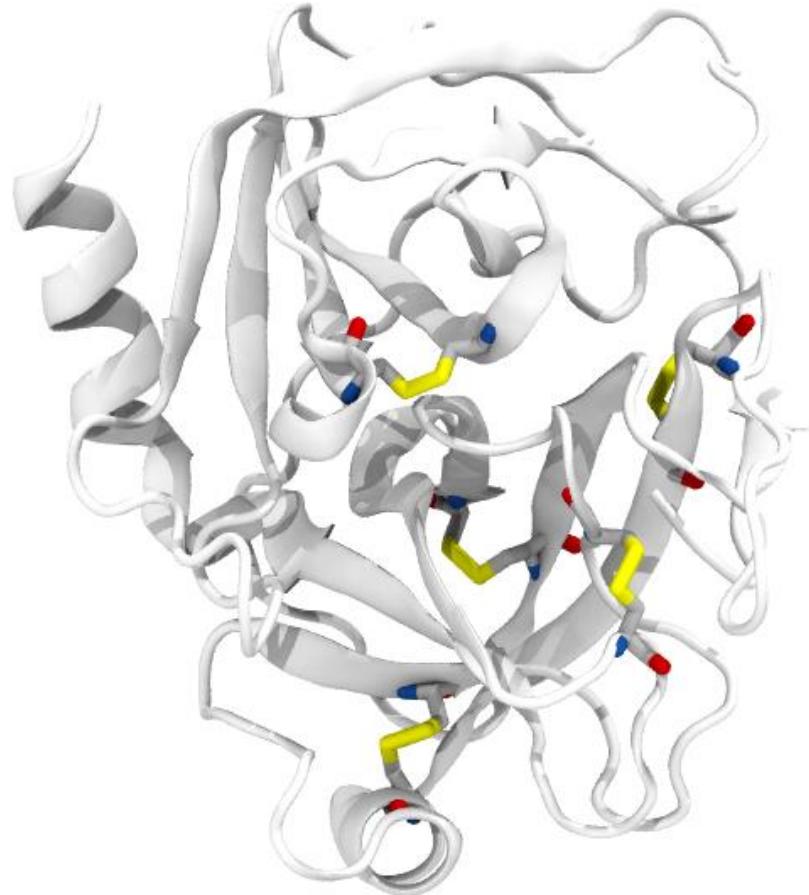
Phosphate and sulfate groups are similar in mass and size, and this can lead to erroneous interpretation of the results.

2 Publications

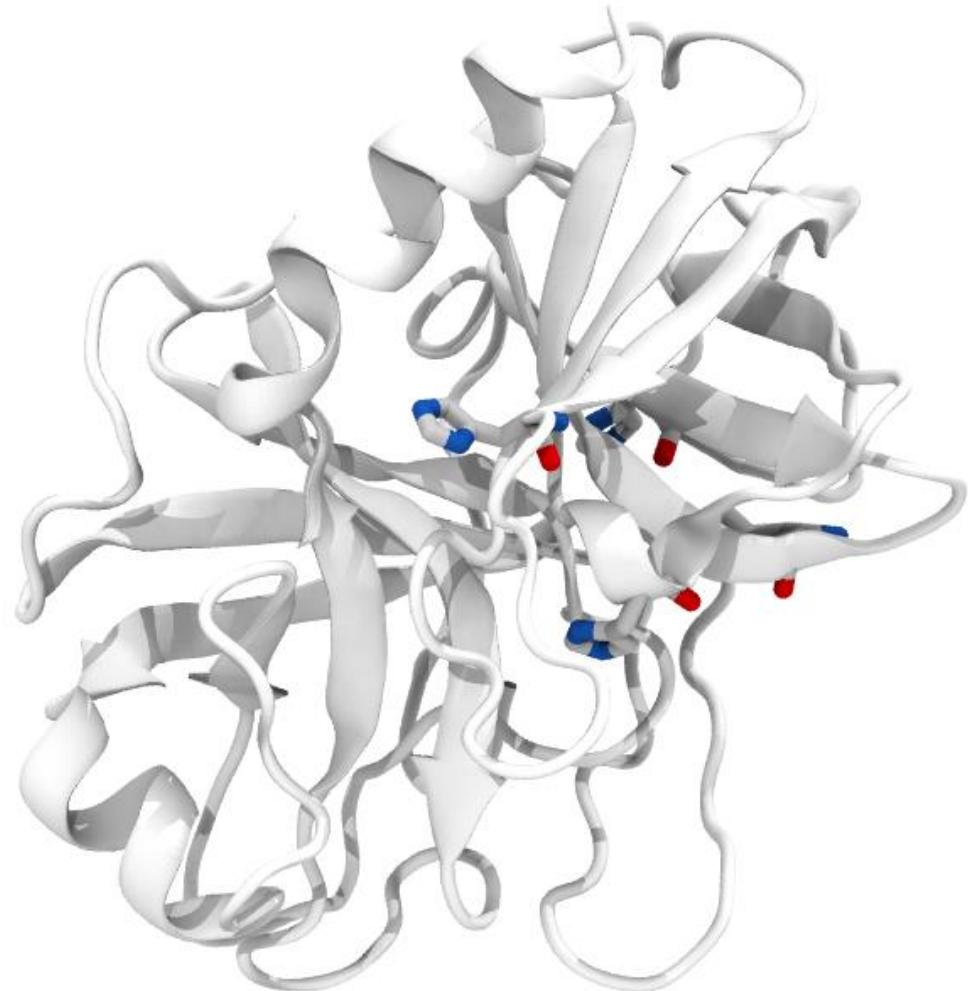
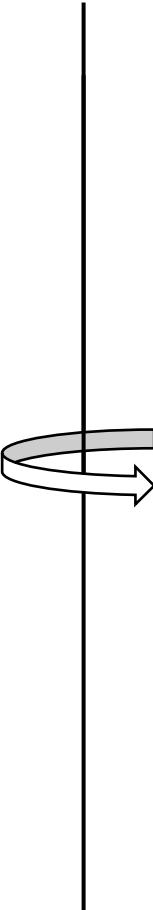
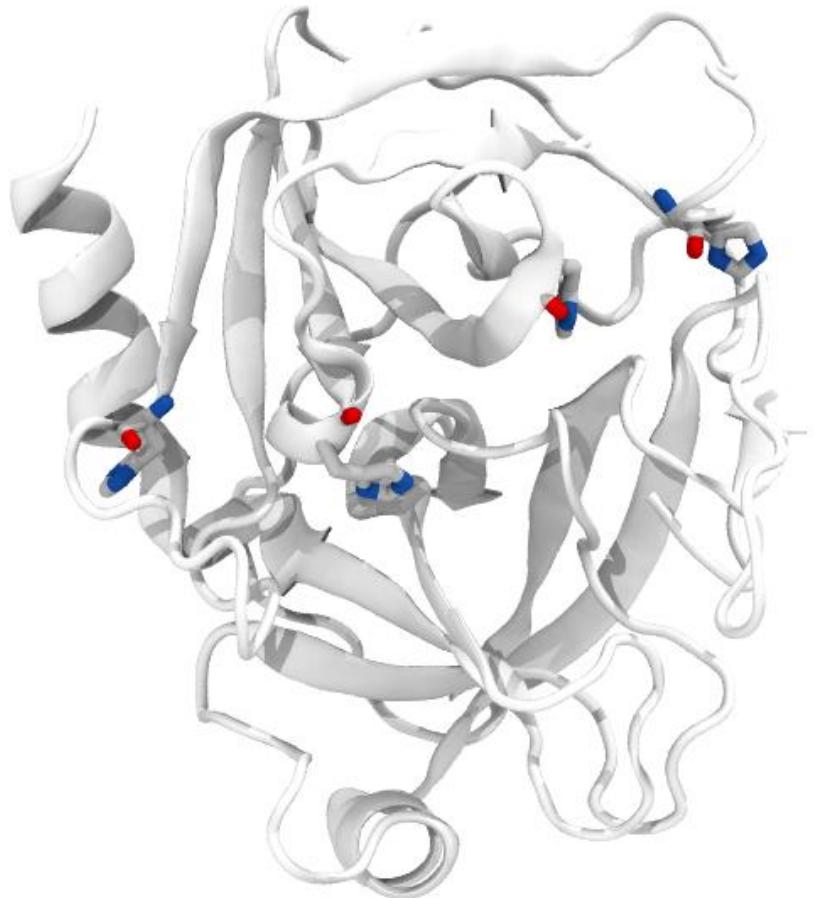
Non-protein molecules



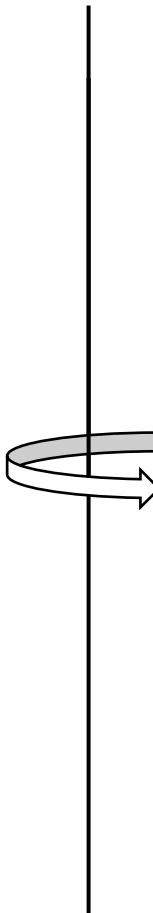
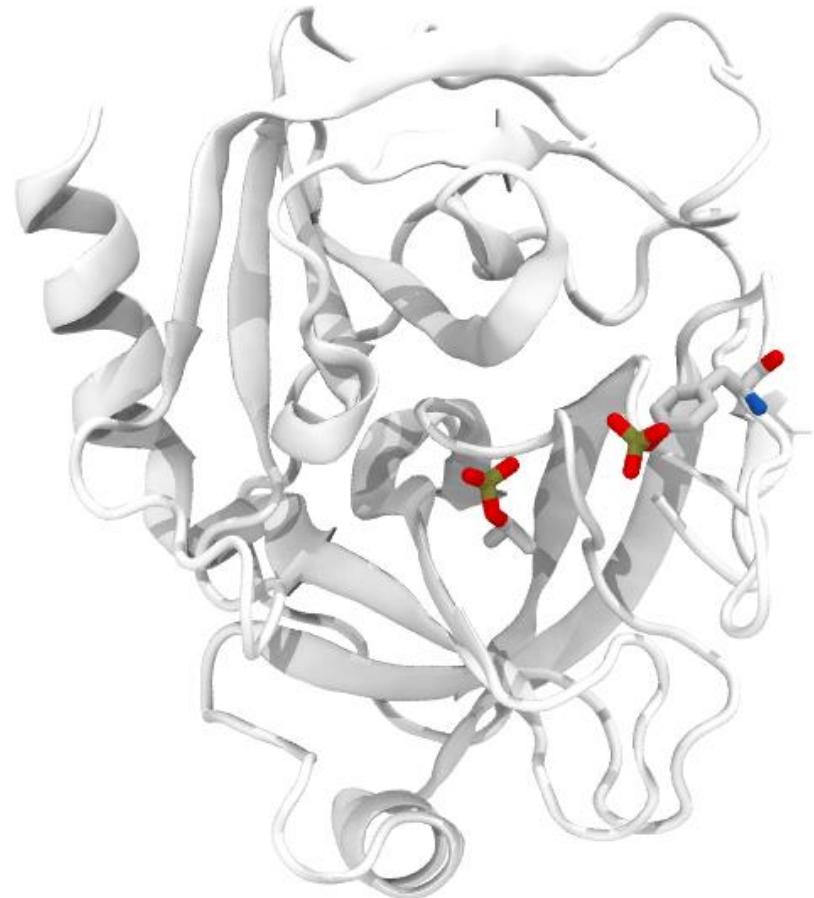
Disulphide bridges



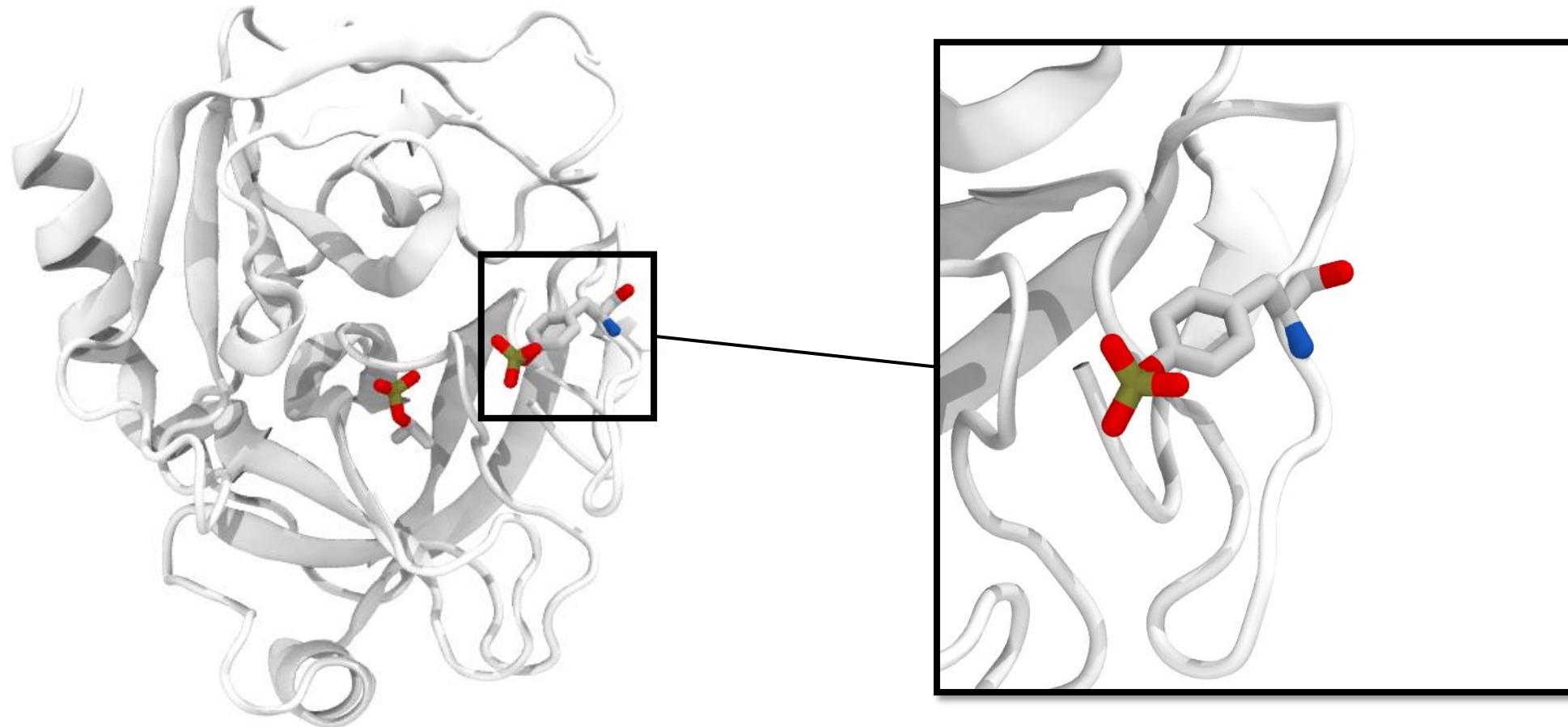
Histidines



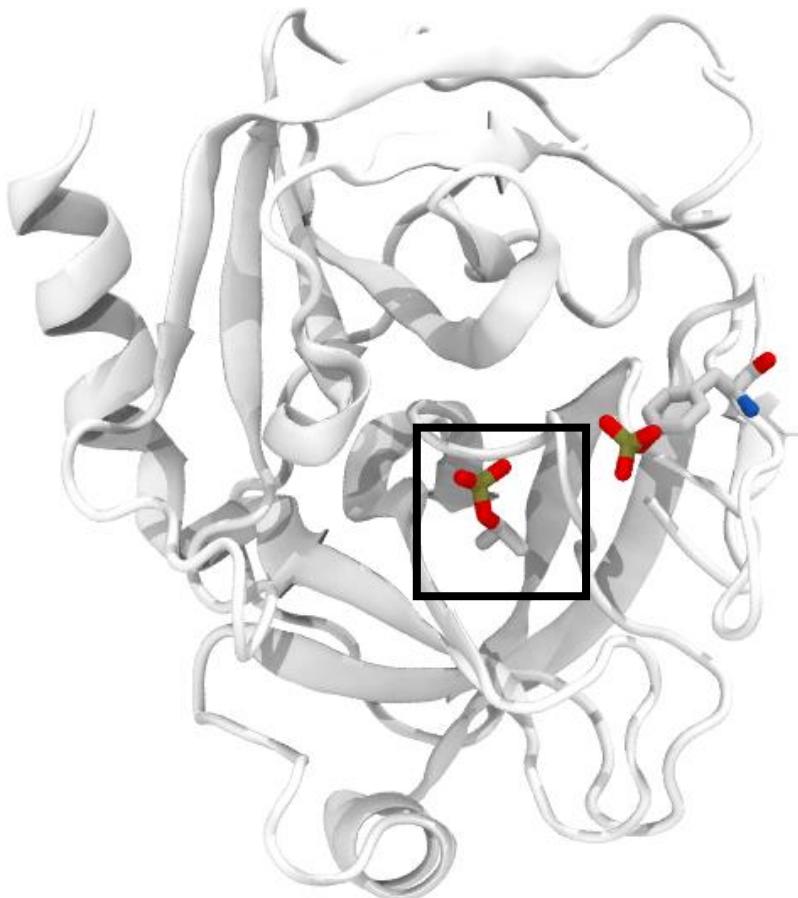
Amino acid modifications: TYR151



Amino acid modifications: TYR151



Ligand?



https://en.wikipedia.org/wiki/TAPS_(buffer)

WIKIPEDIA
The Free Encyclopedia

TAPS (buffer)

From Wikipedia, the free encyclopedia

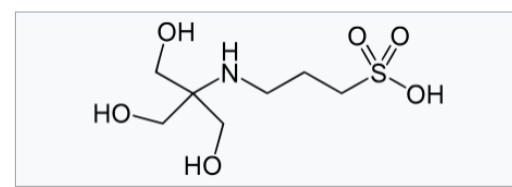
TAPS
([tris(hydroxymethyl)methylamino]propanesulfonic acid) is a chemical compound commonly used to make buffer solutions.

It can bind **divalent cations**, including **Co(II)** and **Ni(II)**.^[1]

TAPS is effective to make buffer solutions in the **pH** range 7.7–9.1, since it has a **pK_a** value of 8.44 (ionic strength $I = 0$, 25 °C).^[2]

The pH (and **pK_a** at $I \neq 0$) of the buffer solution changes with concentration and temperature, and this effect may be predicted e.g. using online calculators.^[3]

TAPS



Names

Preferred IUPAC name
3-[{1,3-Dihydroxy-2-(hydroxymethyl)propan-2-yl}amino]propane-1-sulfonic acid

Other names
N-Tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid

Identifiers

CAS Number | 29915-38-6 ✓

Tools that will help with all of this

Most simulation packages contain tools to help add missing H-atoms. Not all work in the same way.

Where there is more than one possible answer, not all packages will **make the same decision**.

There are two independent, well-established tools designed specifically to look at these issues, and a webserver that puts them together:

Reduce

Adds hydrogens, and tests for
'NQH flips'

Download:

<http://kinemage.biochem.duke.edu/software/reduce.php>

PropKa

Predicts amino acid
protonation states

Download:

<https://github.com/jensengroup/propka-3.1>

PDB2PQR server

Interface to both Reduce and
proPKa

Website:

<http://server.poissonboltzmann.org/pdb2pqr>

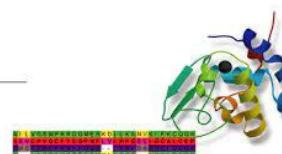
H++

 OpenEye
SCIENTIFIC



 PDB
FIXER


Modeller
Program for Comparative Protein
Structure Modelling by Satisfaction
of Spatial Restraints



...

Let's try some protein prep out!



[uniprot.org](https://www.uniprot.org)



[rcsb.org](https://www.rcsb.org)



<https://server.poissonboltzmann.org>



Task 2:

Identify a good crystal structure for **Carboxypeptidase B2** and use the Poisson Boltzmann server to produce output for amber.
Starting structure PDB: 3D67 use UNIPROT to find out more!

Task 3:

Look at the AlphaFold 2 structure, what do you need to watch out for here?