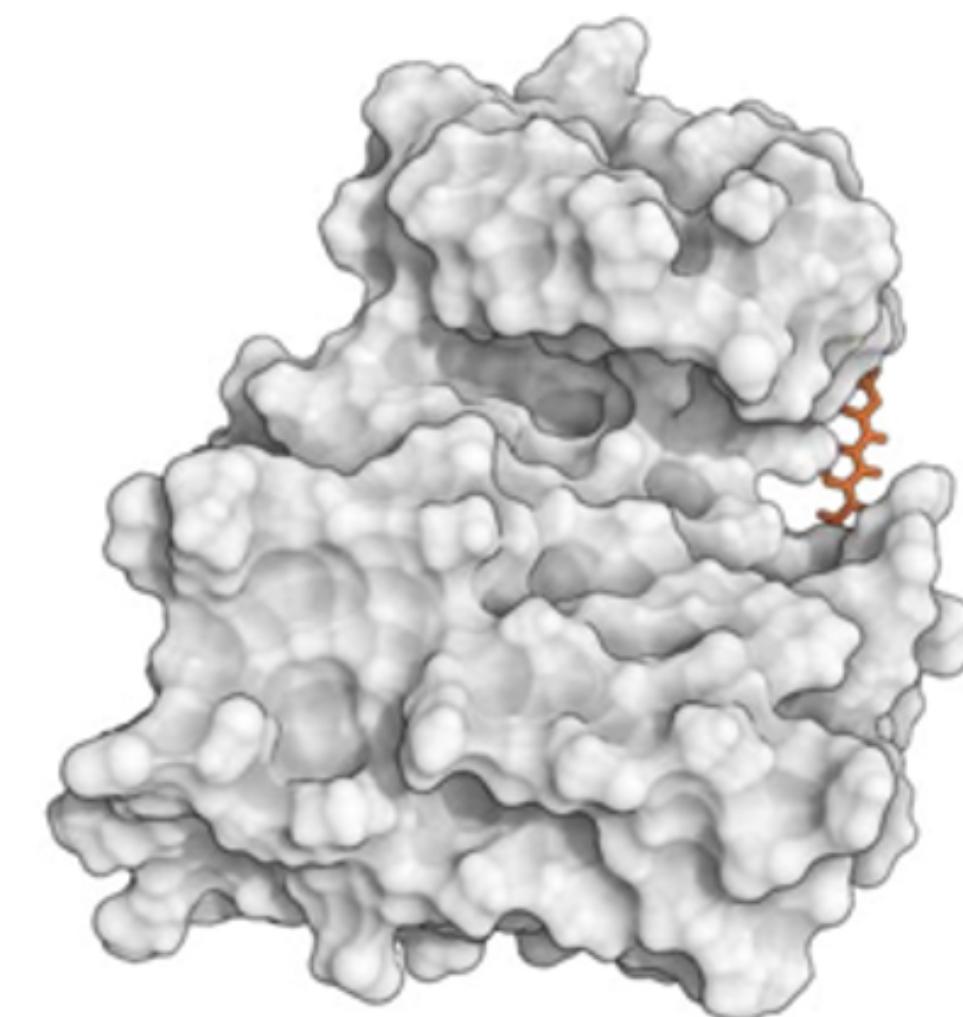
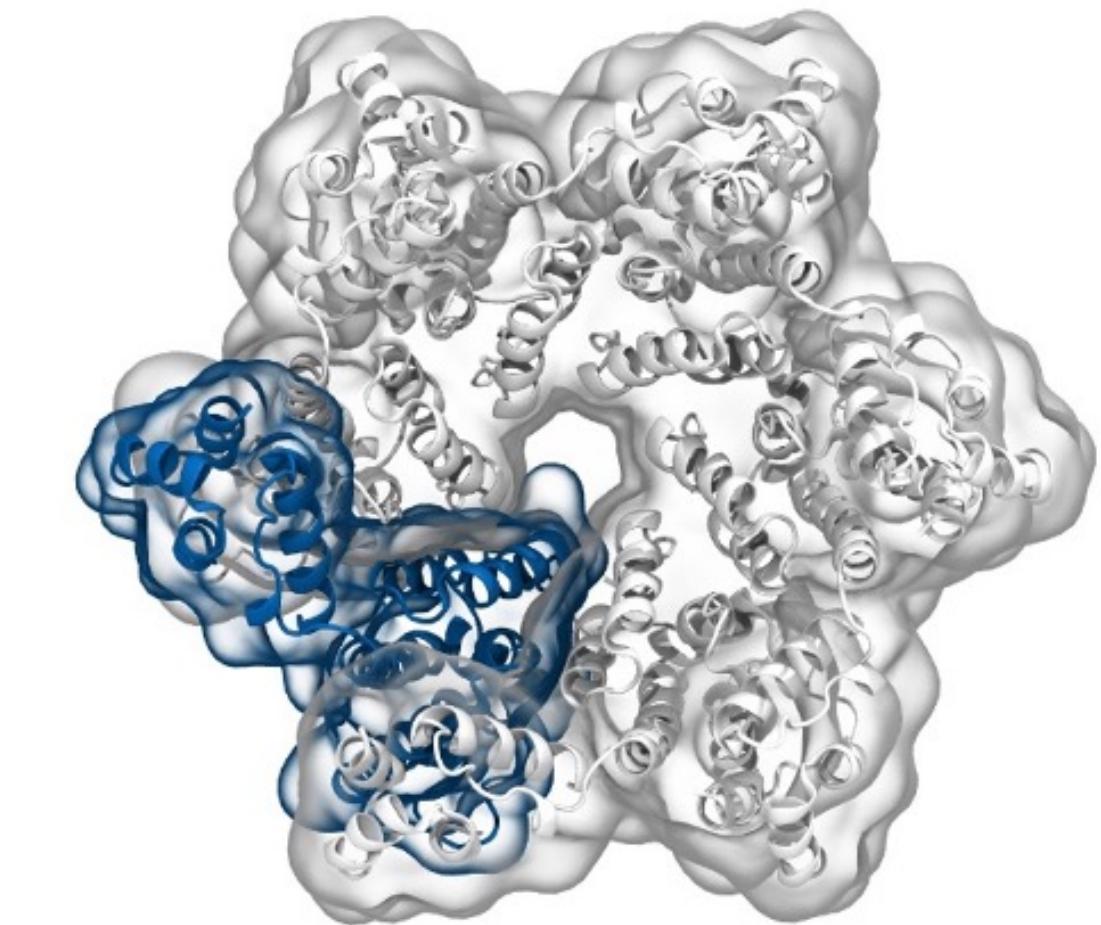


# Simulation of Biomolecules



## Lecture 2: Prepping proteins for simulations

2023 CCP5 Summer School



Dr Matteo Degiacomi  
Durham University  
[matteo.t.degiacomi@durham.ac.uk](mailto:matteo.t.degiacomi@durham.ac.uk)

Dr Antonia Mey  
University of Edinburgh  
[antonia.mey@ed.ac.uk](mailto:antonia.mey@ed.ac.uk)

# General Information



CCPBioSim Training week 25 - 28 September 2023 at the University of Leeds and online.

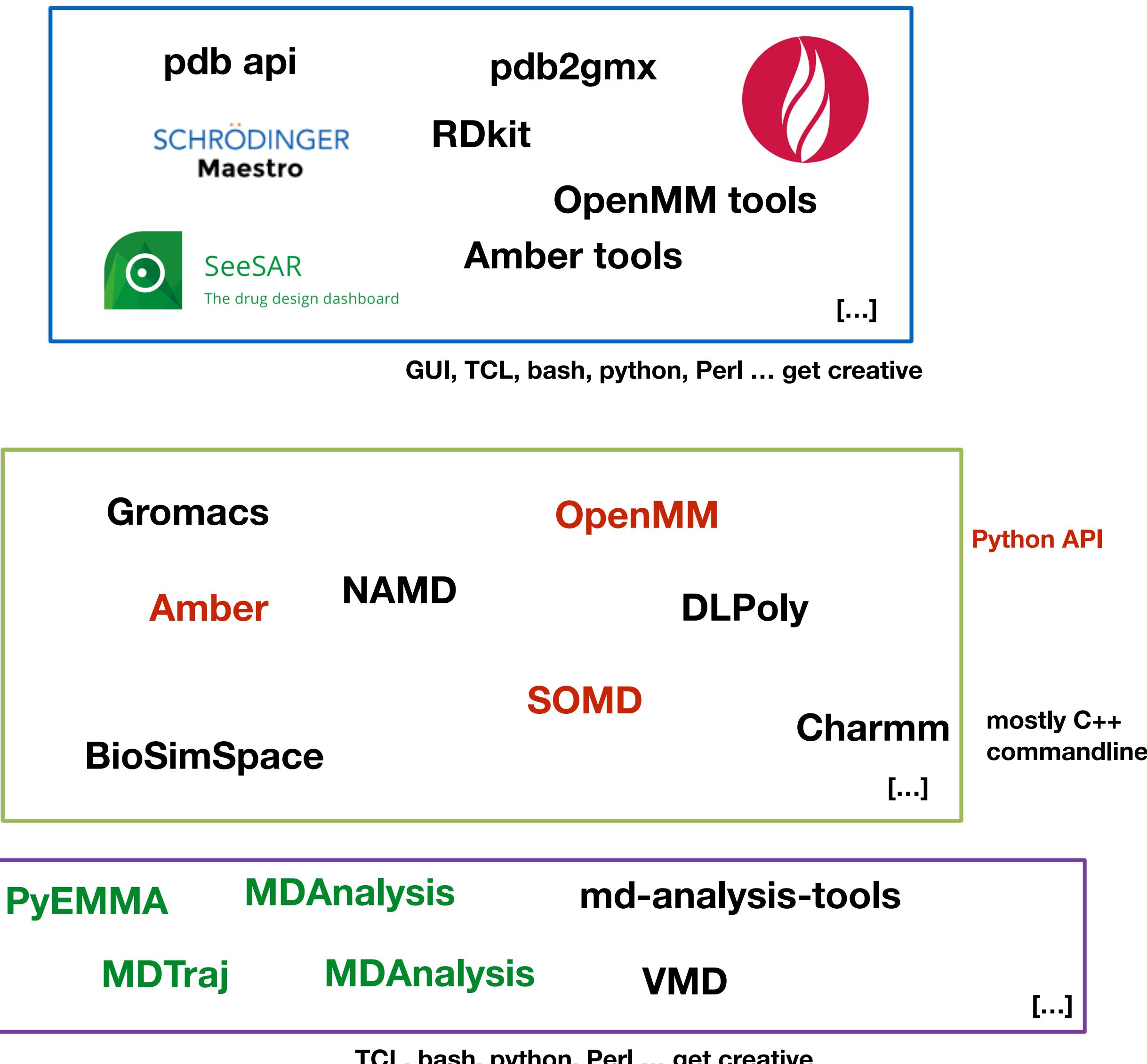
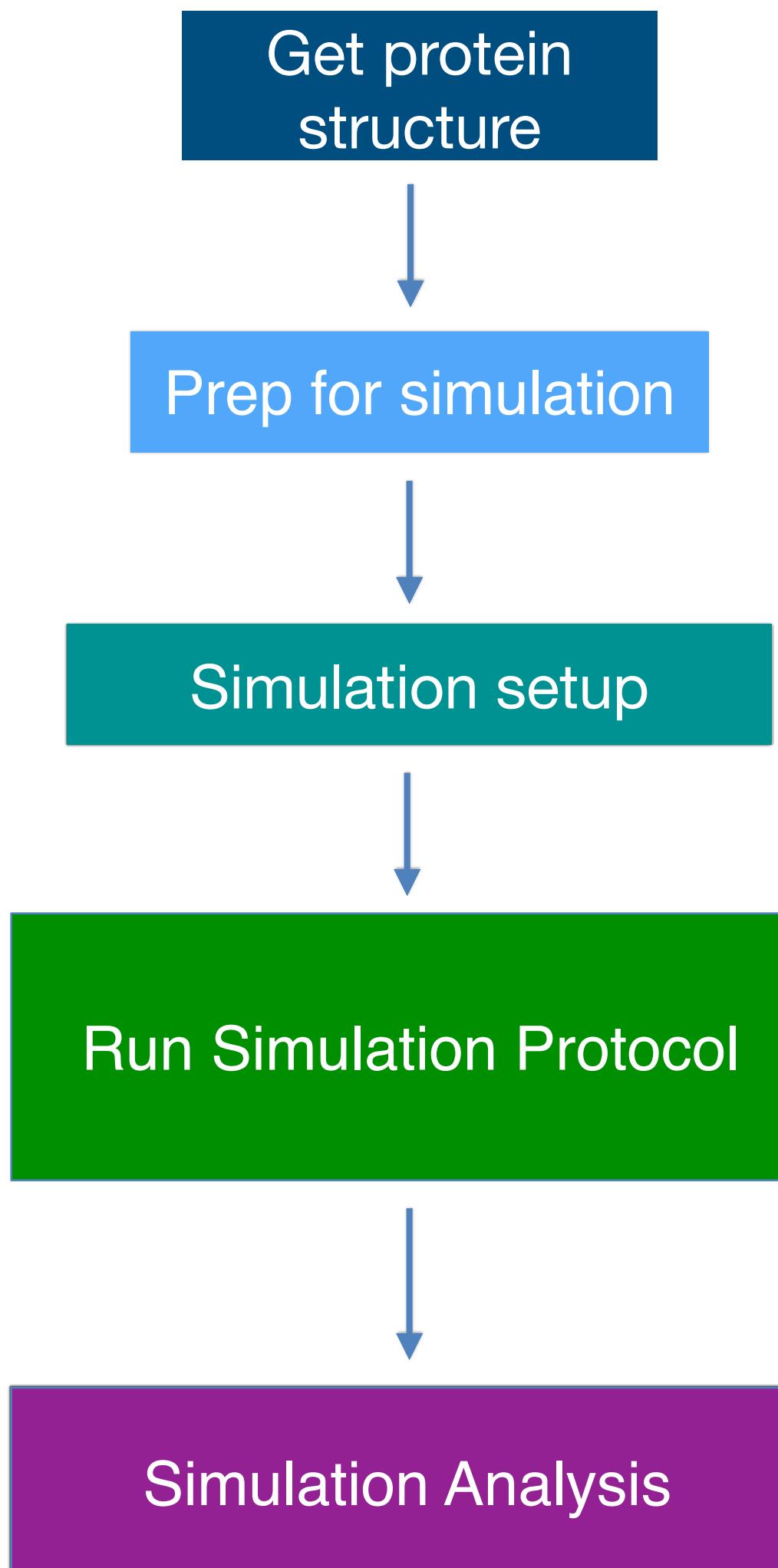
**Registration closes: 1st September 2023**

**Cost: £0**

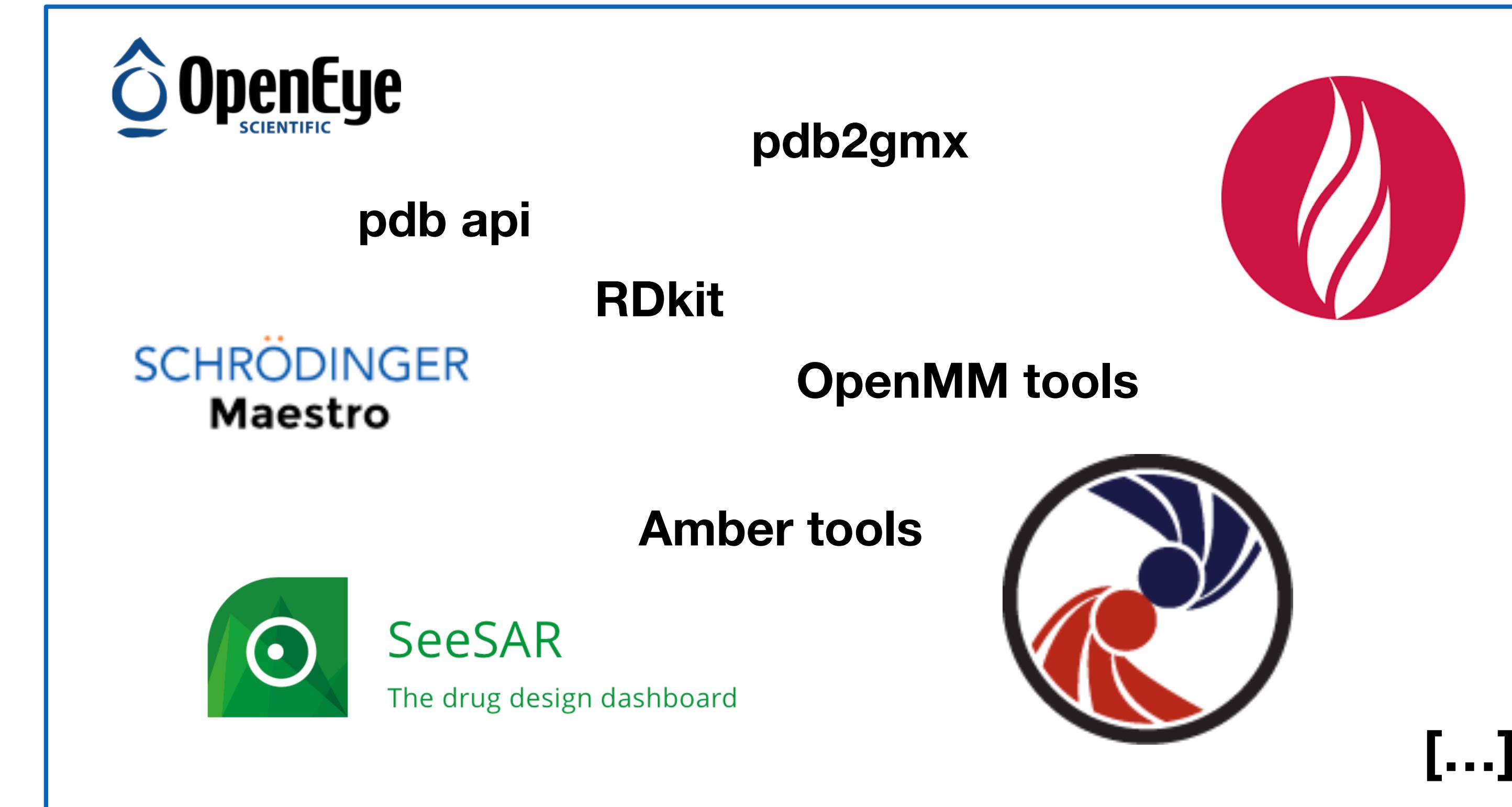
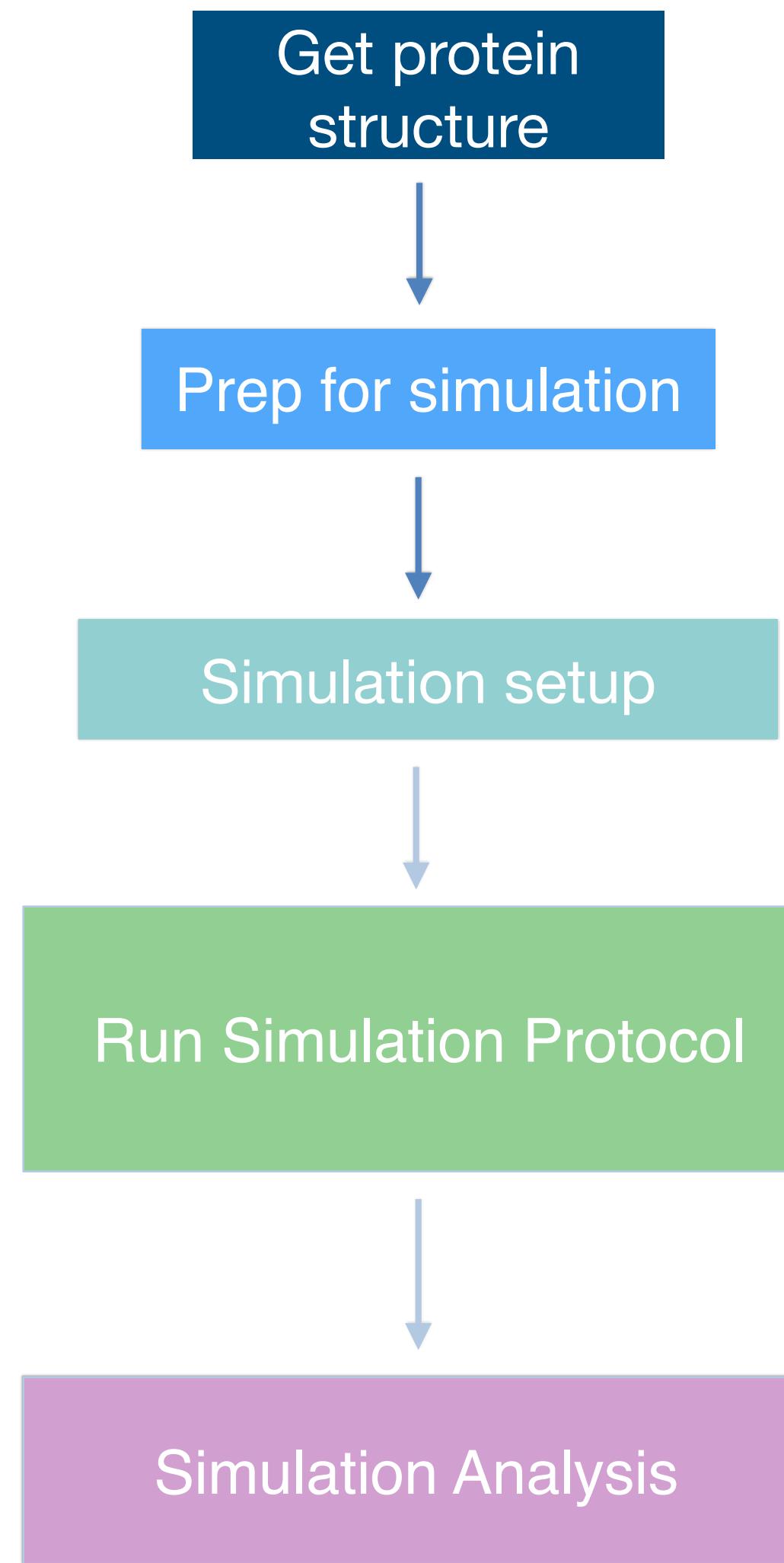


Some of the material presented is adapted from Prof. Charlie Laughton

# A typical workflow for molecular dynamics



# Let's get started with understanding protein structures

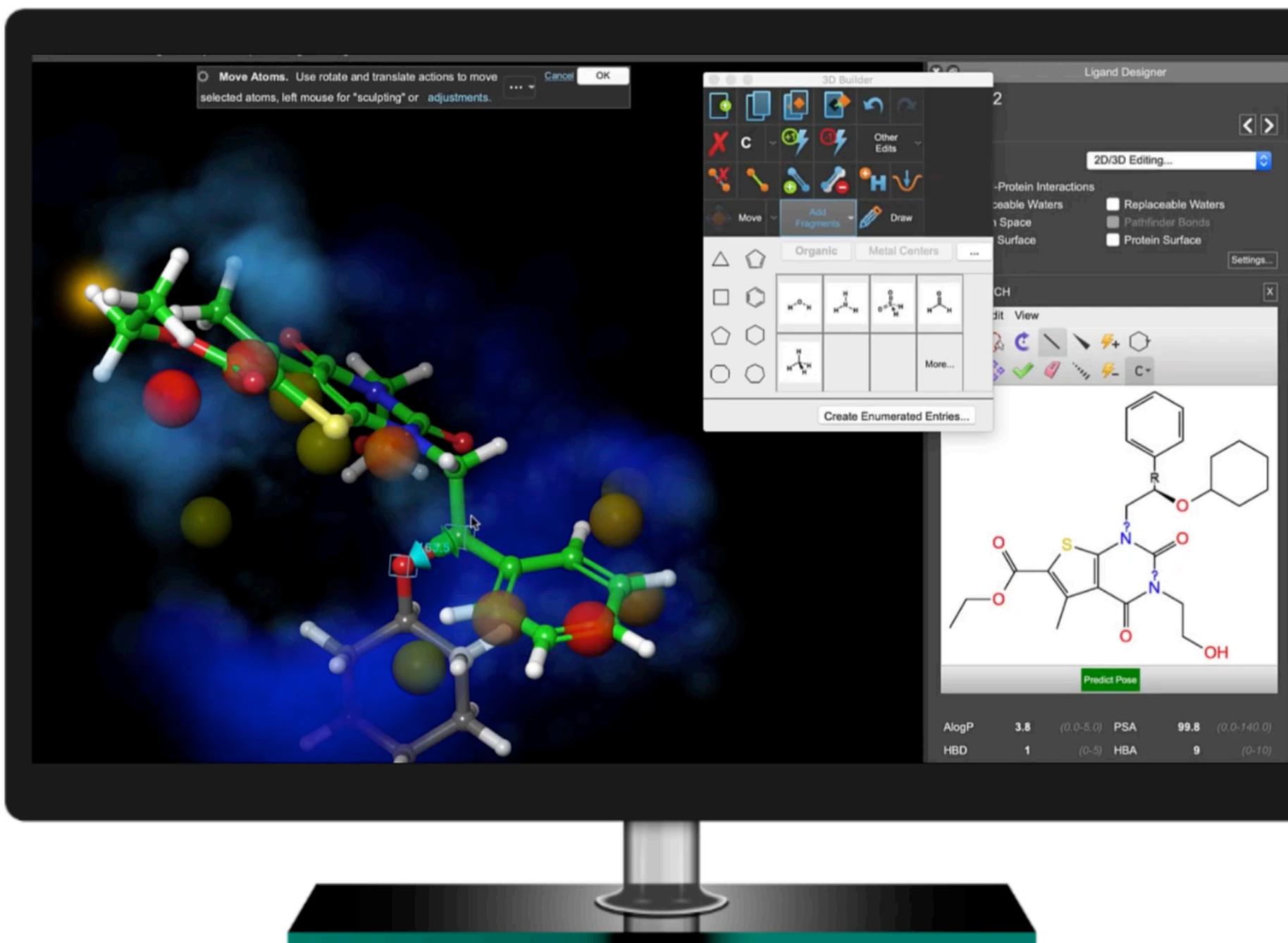


GUI, TCL, bash, python, Perl ... get creative

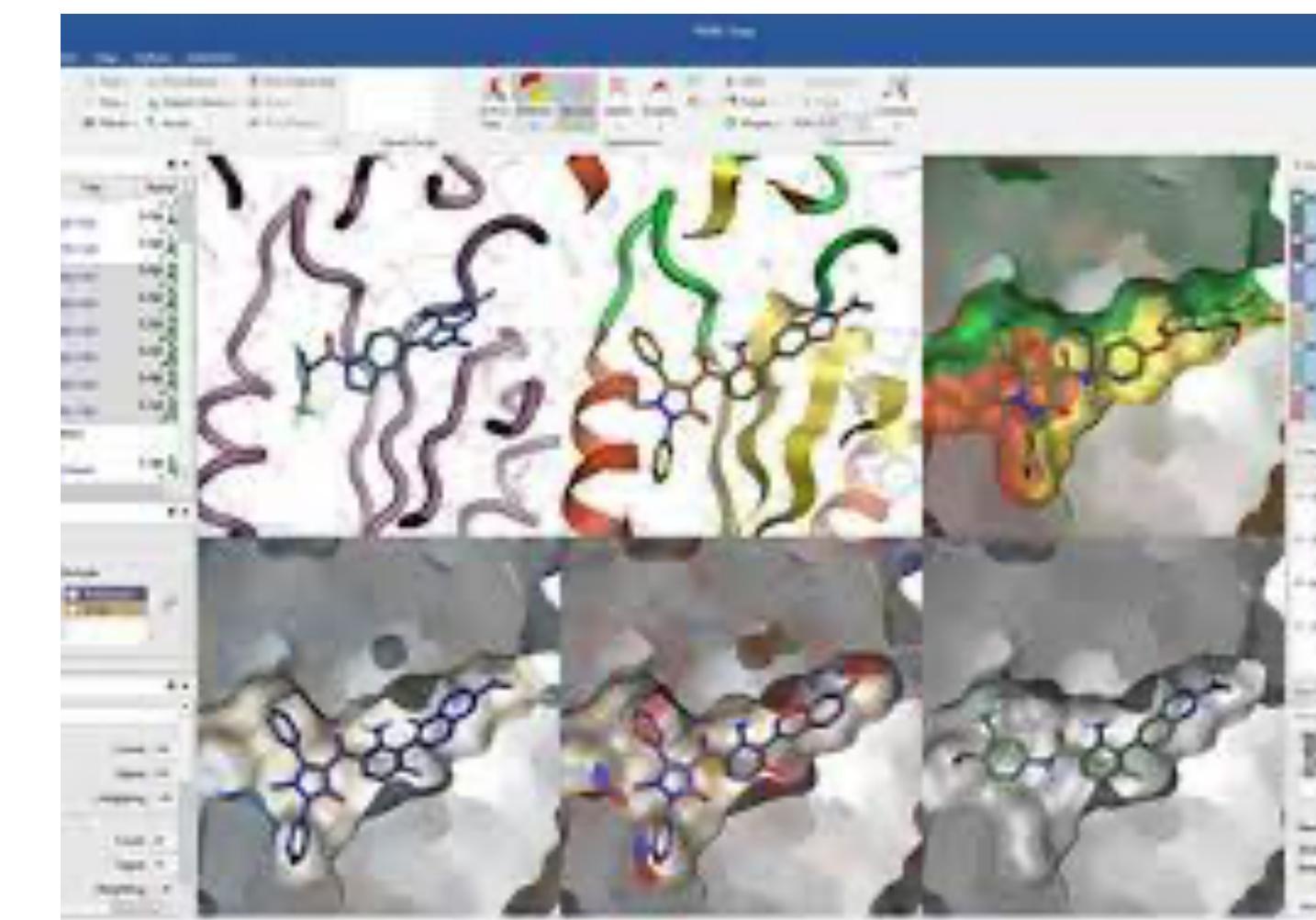
**Part 1: I have a protein I want to model, now what?**

# Black-box protein setup tools can help with protein prep

SCHRÖDINGER  
**Maestro**



**Flare**



**SeeSAR**  
The drug design dashboard

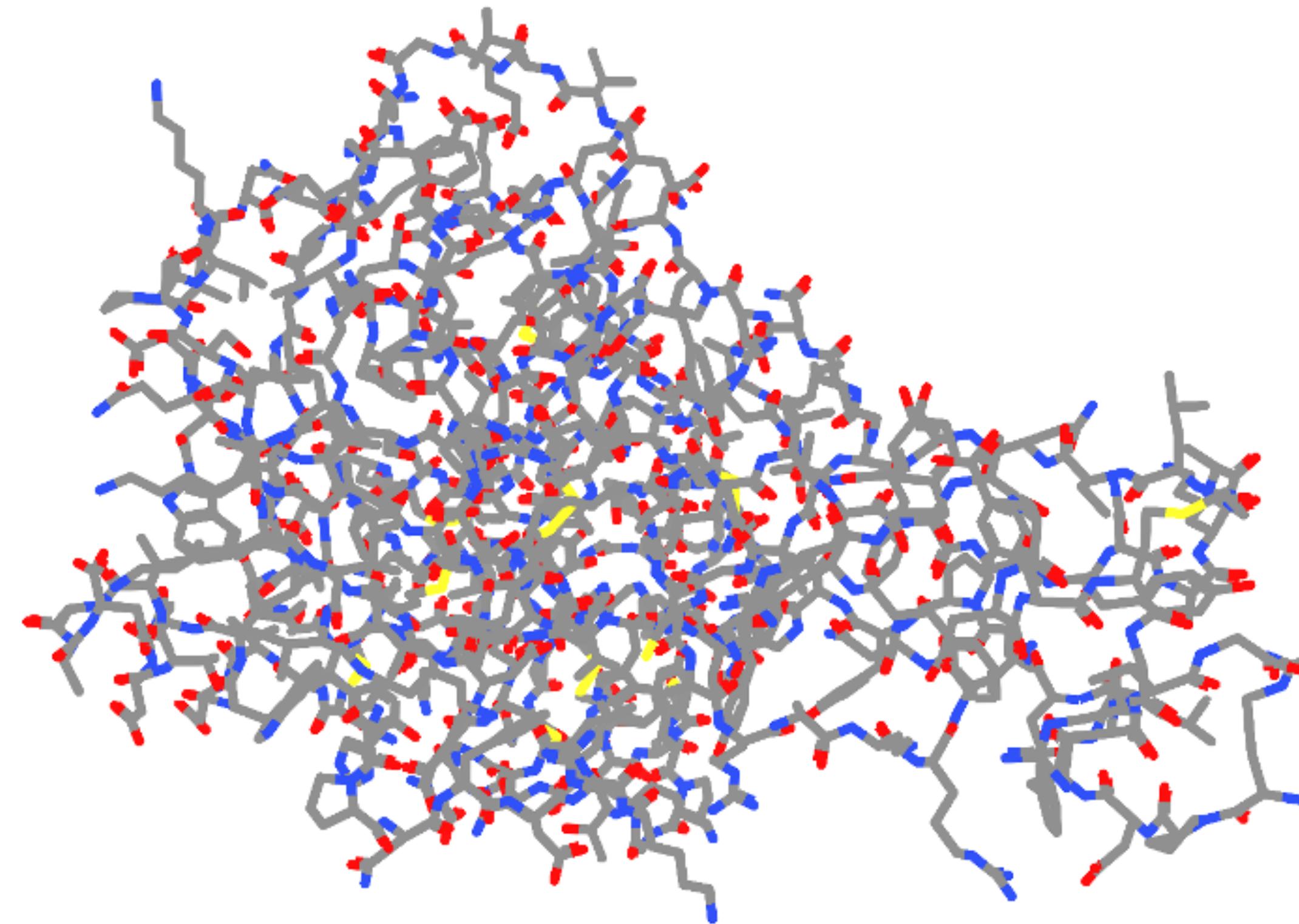
**More on this in  
the afternoon**

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

Go to Menti.com: and enter 2627 8018



# Crystal structures provide electron densities



# Crystal structures are models with potential errors

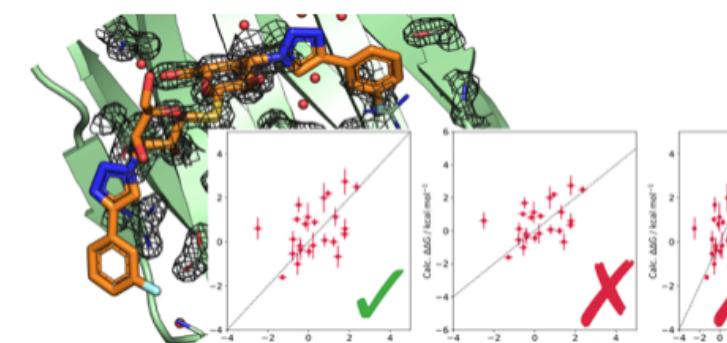
[HOME](#) / [ARCHIVES](#) / [VOL. 4 NO. 1 \(2022\)](#) / [Articles](#)

## 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>



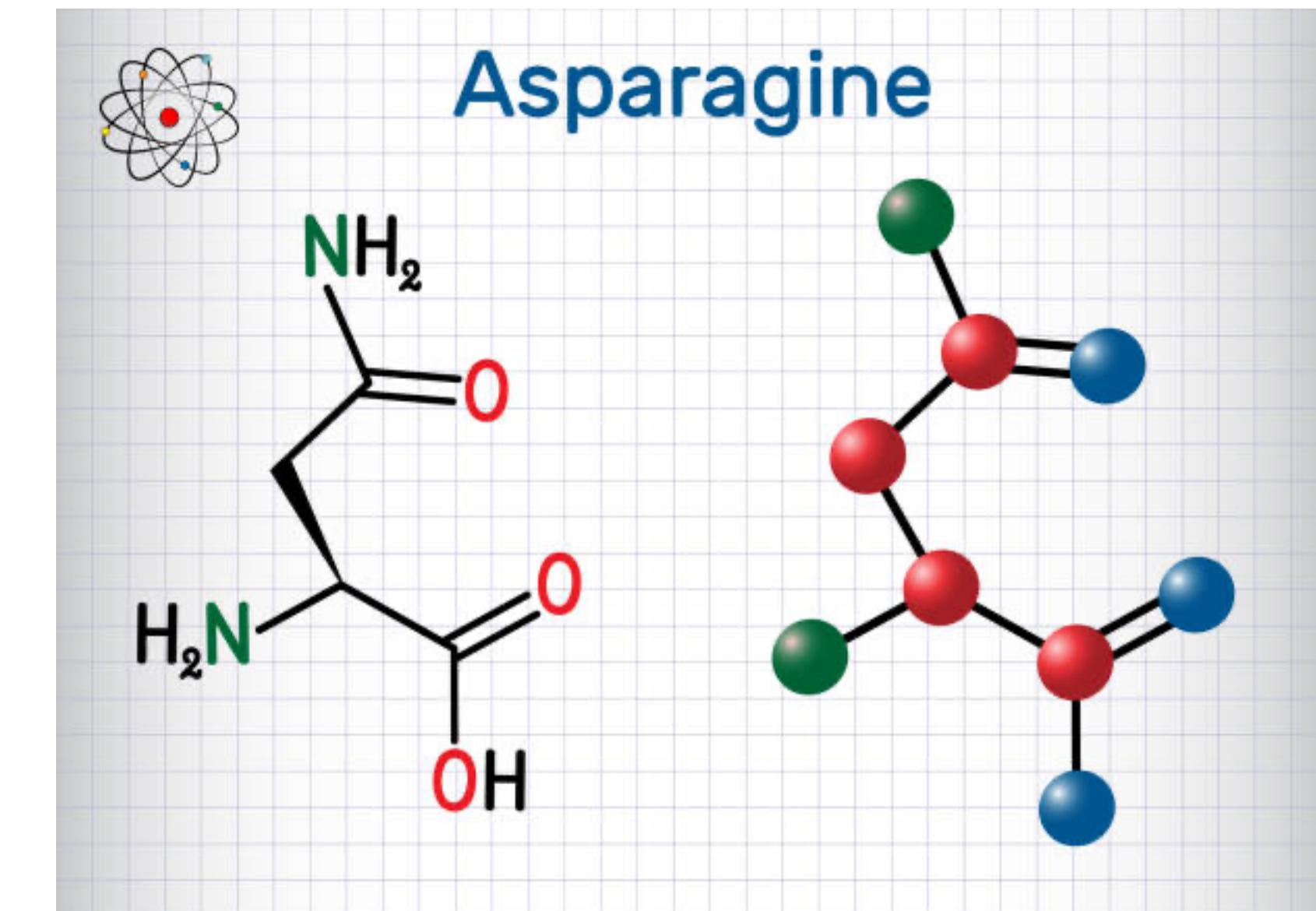
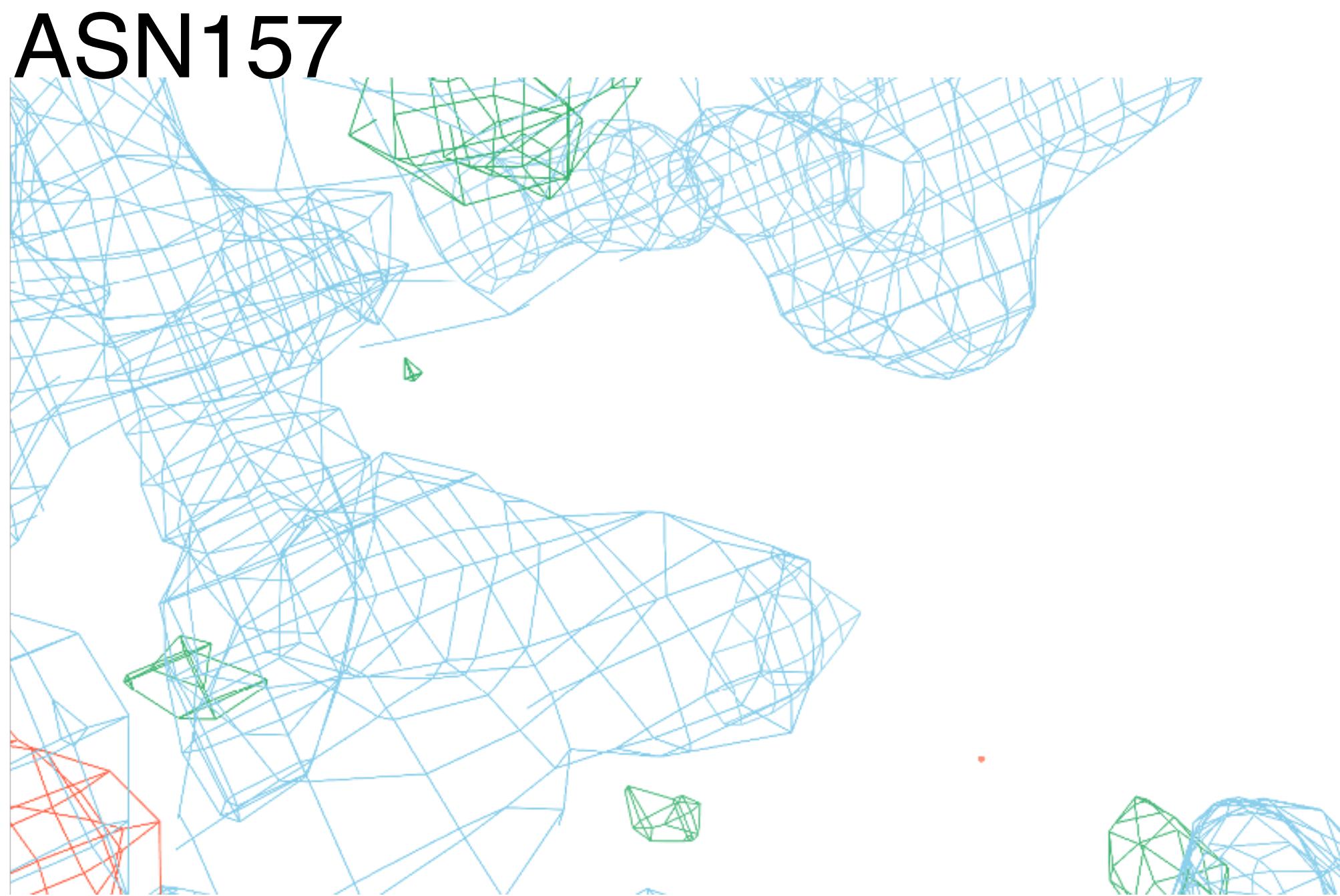
## Examples of errors/oddities in PDB structures

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

Gert Vriend (author of WHAT\_CHECK)

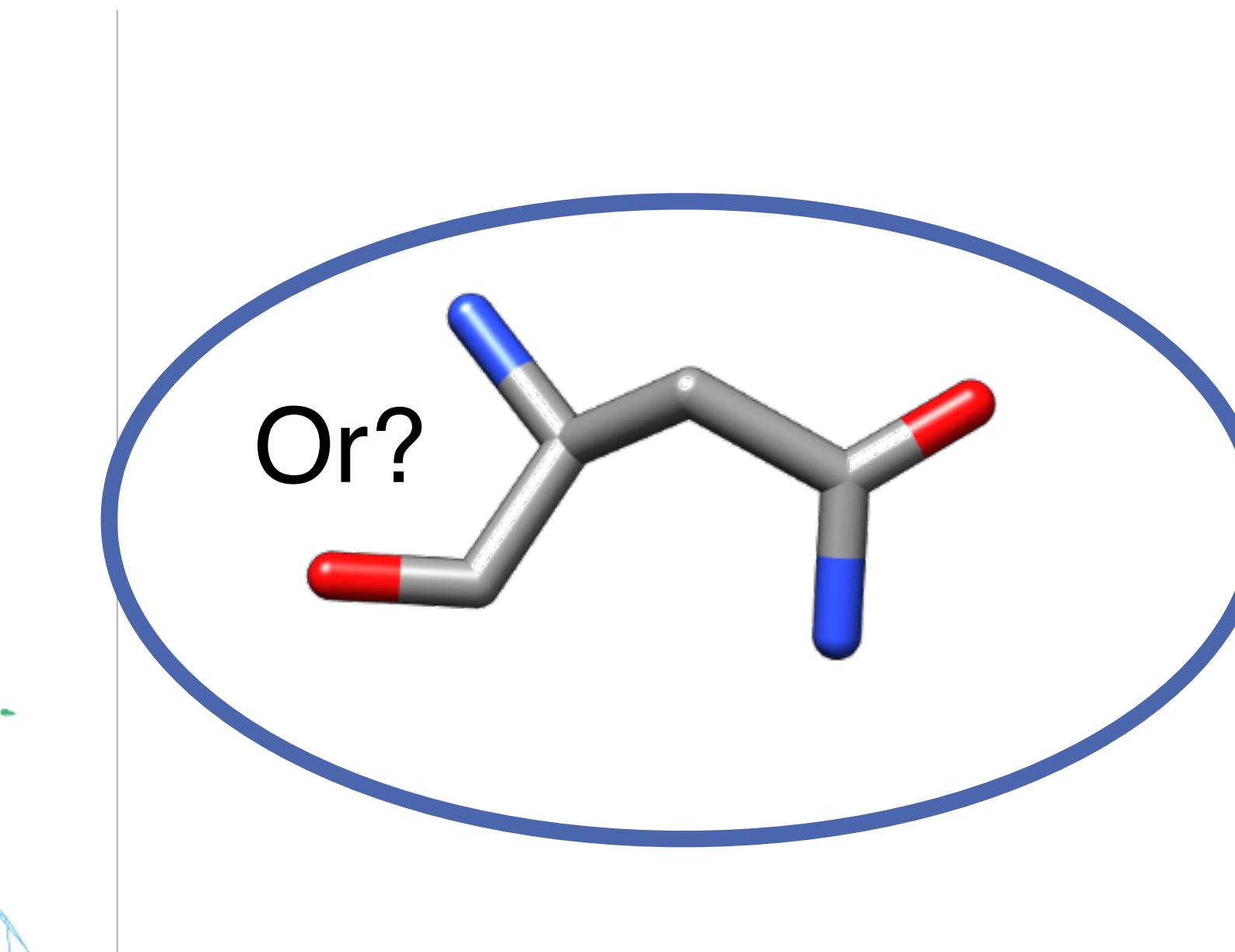
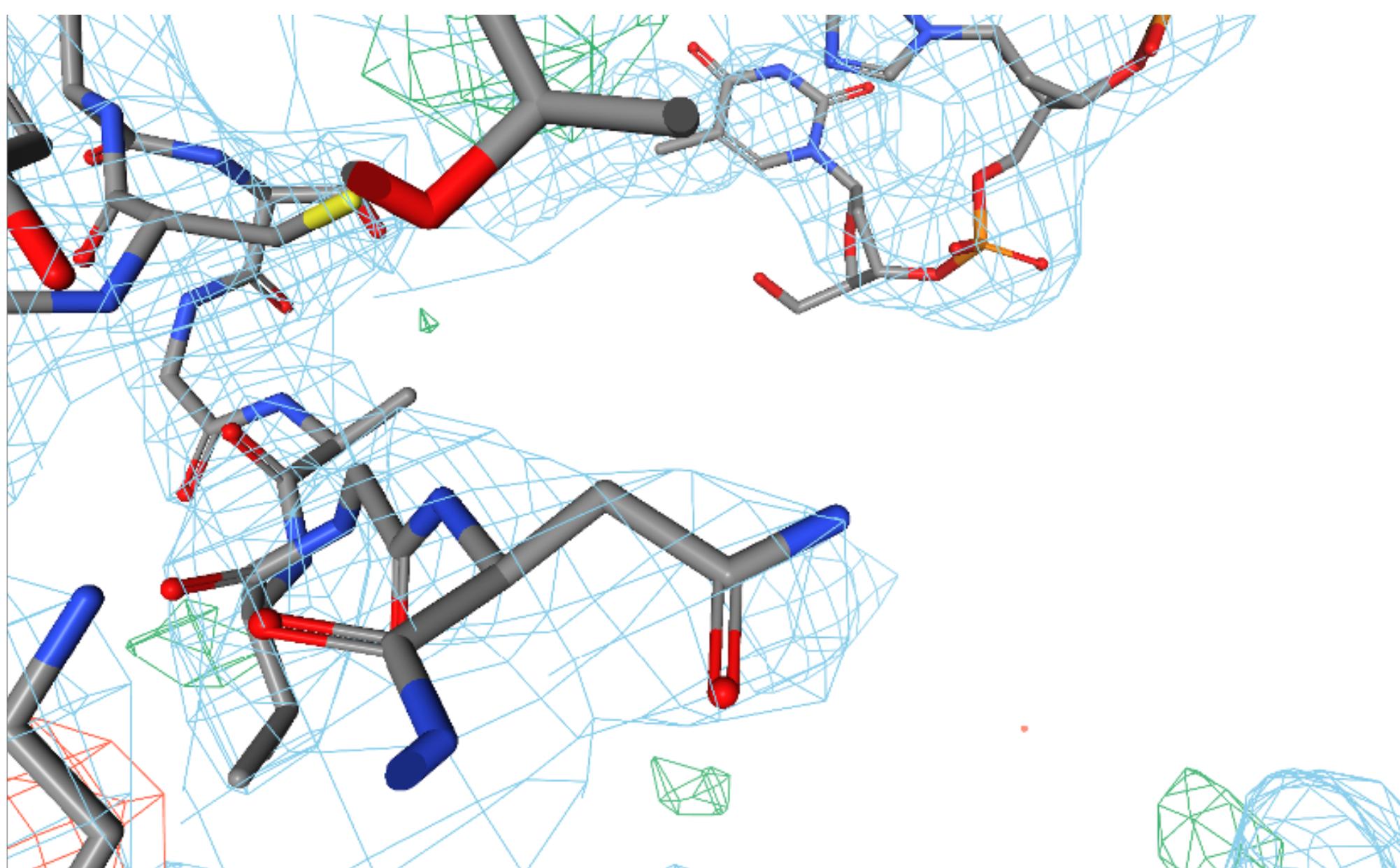
# Crystal structures are models derived from electron densities

1T38: HUMAN O6-ALKYLGUANINE-DNA ALKYLTRANSFERASE



# Crystal structures are models derived from electron densities

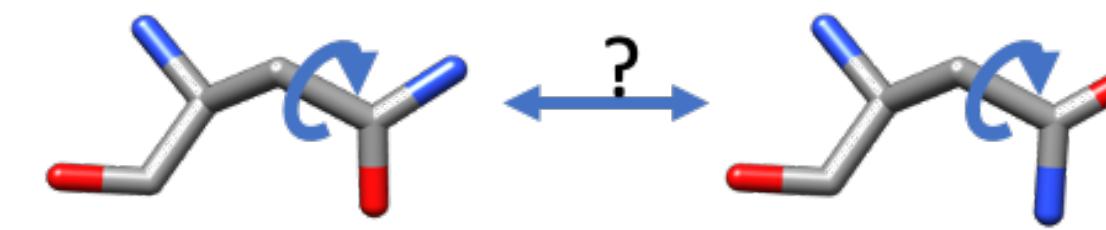
1T38: HUMAN O6-ALKYLGUANINE-DNA ALKYLTRANSFERASE



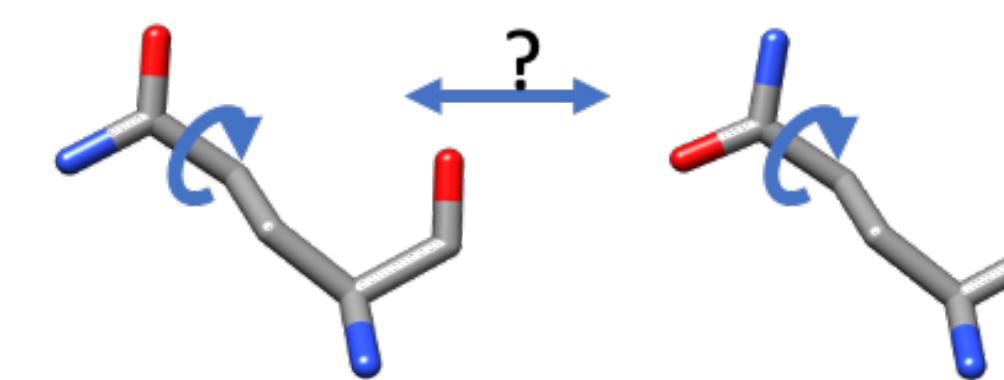
# Alternative conformations of side chains - NGH flips

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

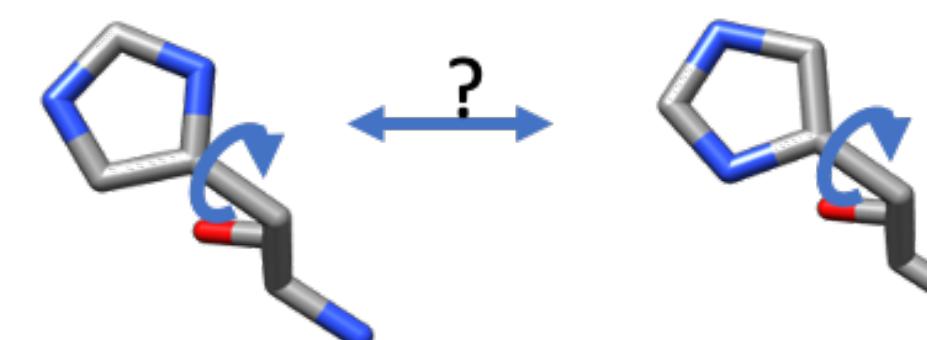
No standard protocol for this: always worth double-checking.



**Asparagine (N)**



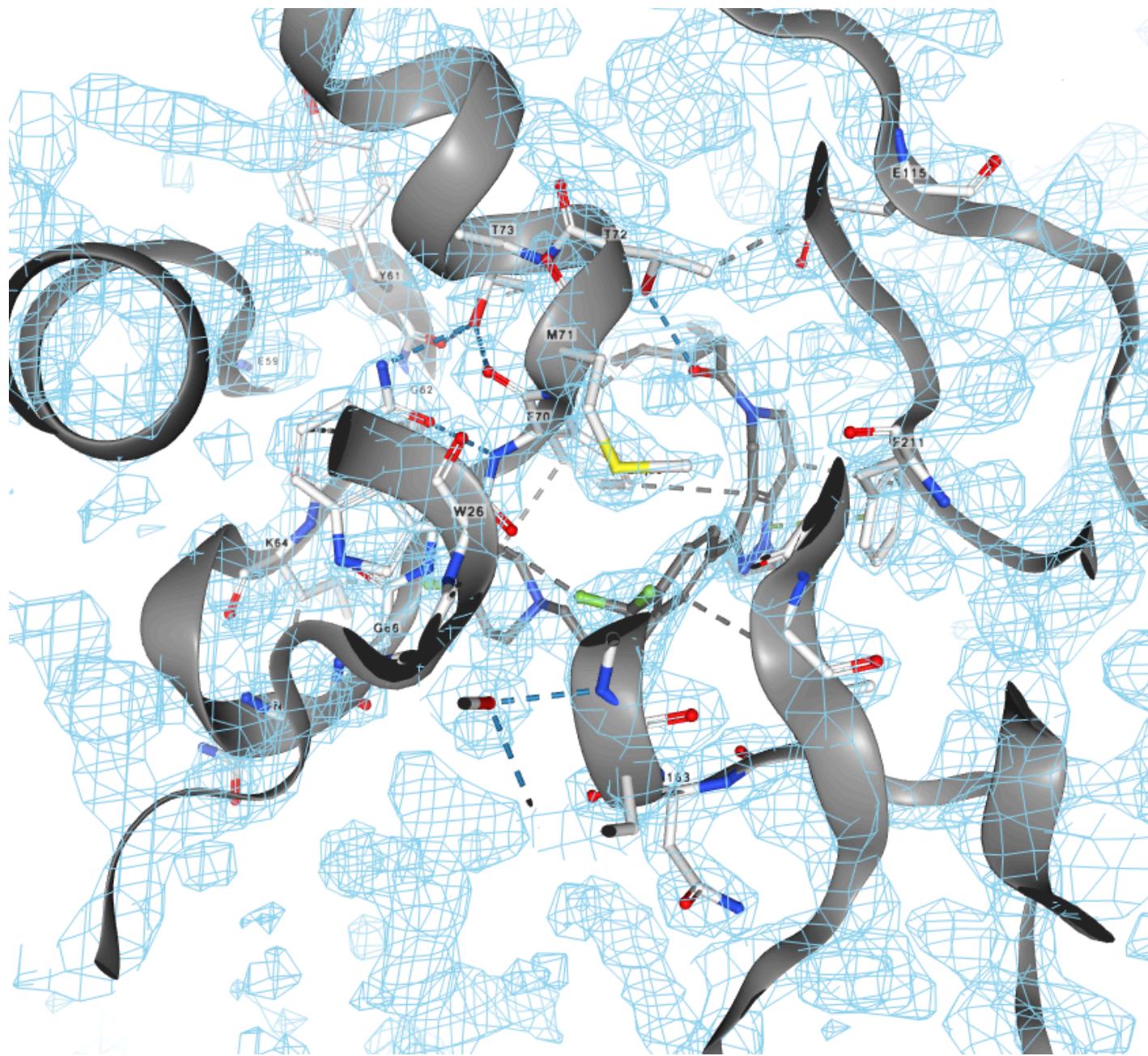
**Glutamine (Q)**



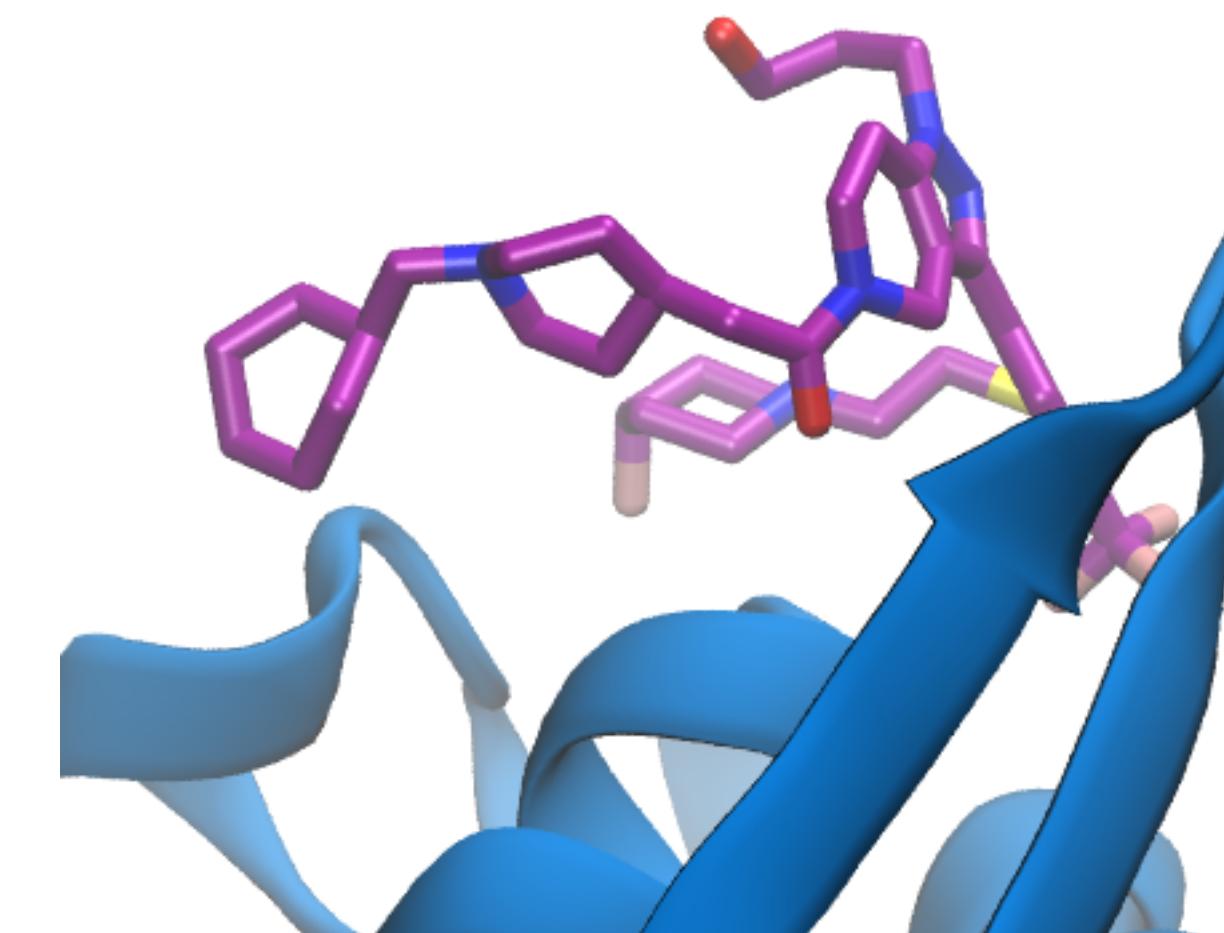
**Histidine (H)**

# Ligand densities can also have creative input from the crystallographer

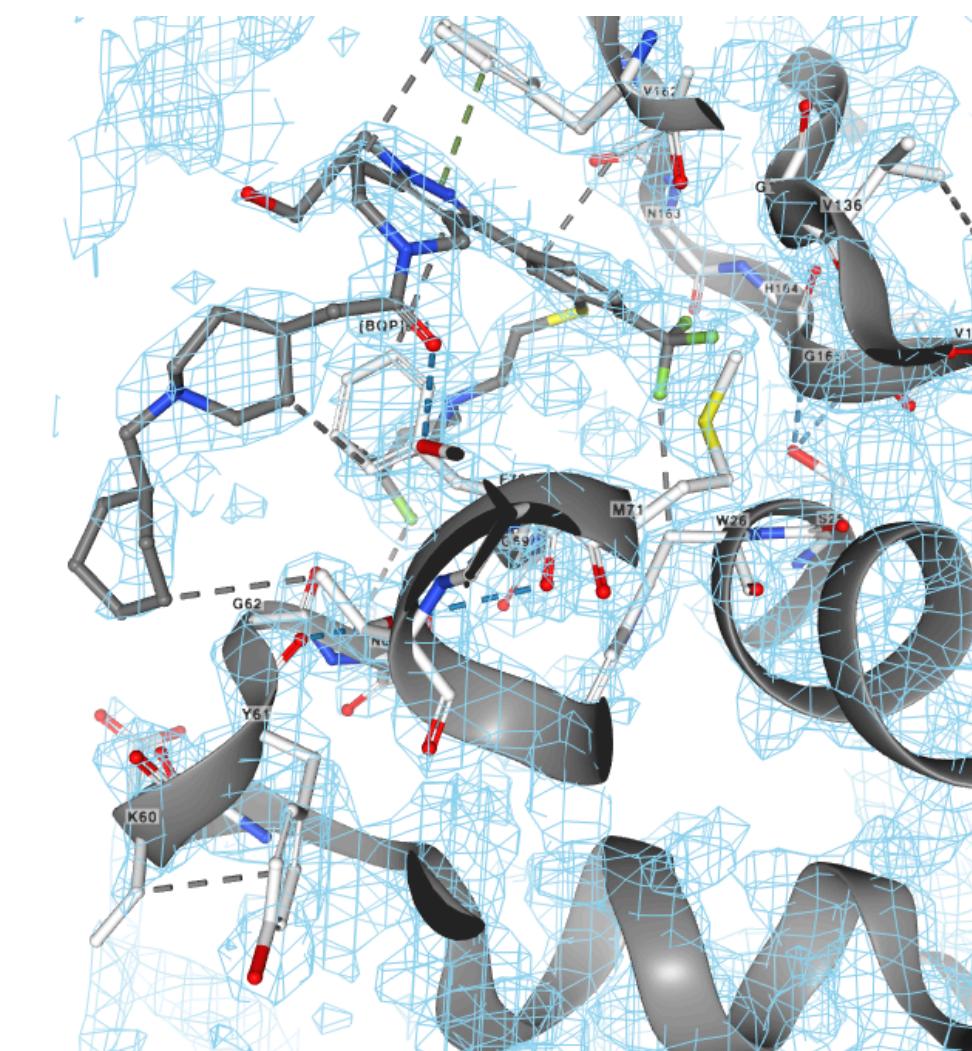
Cathepsin S



Cathepsin S  
with ligand

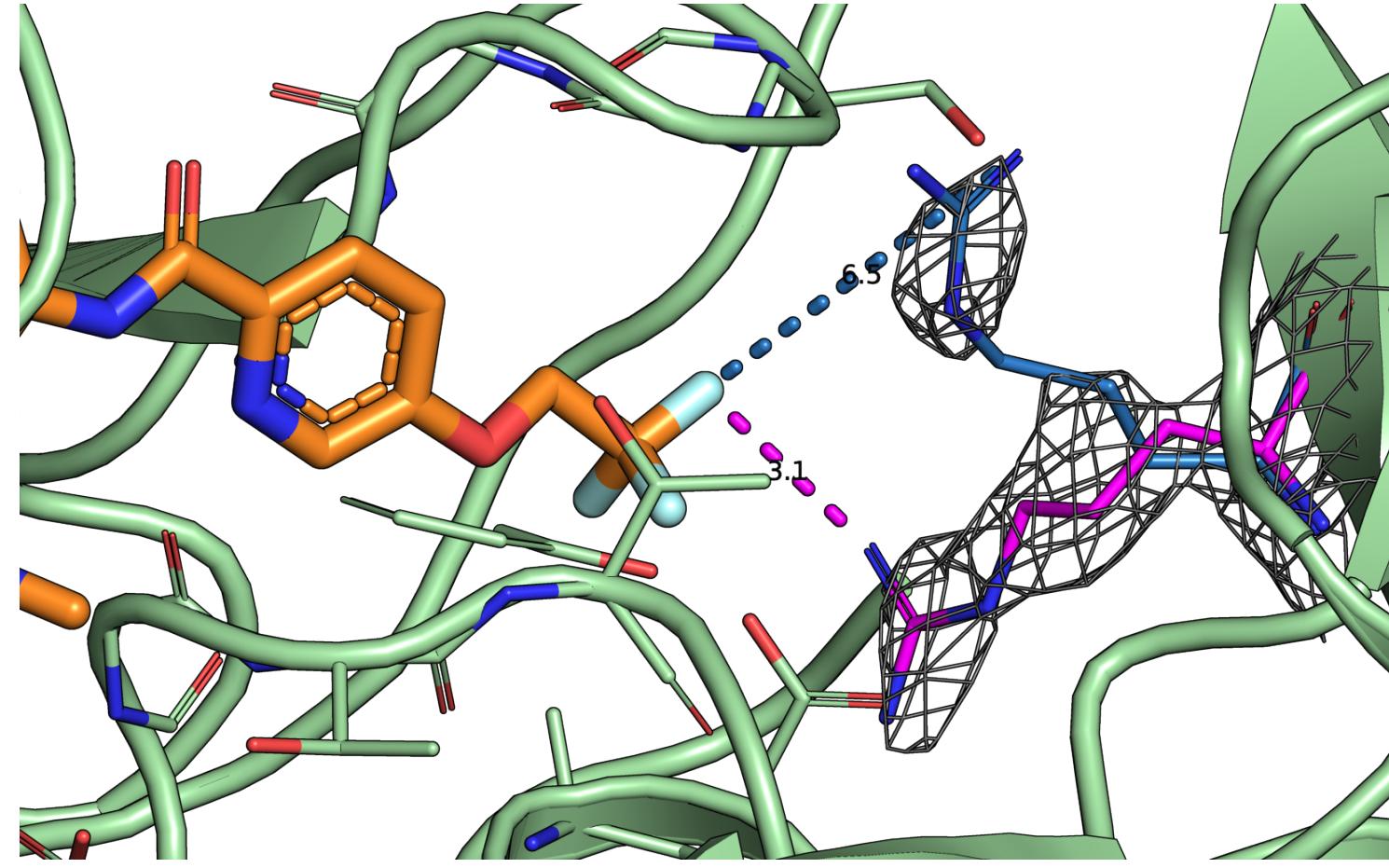


boat/chair?

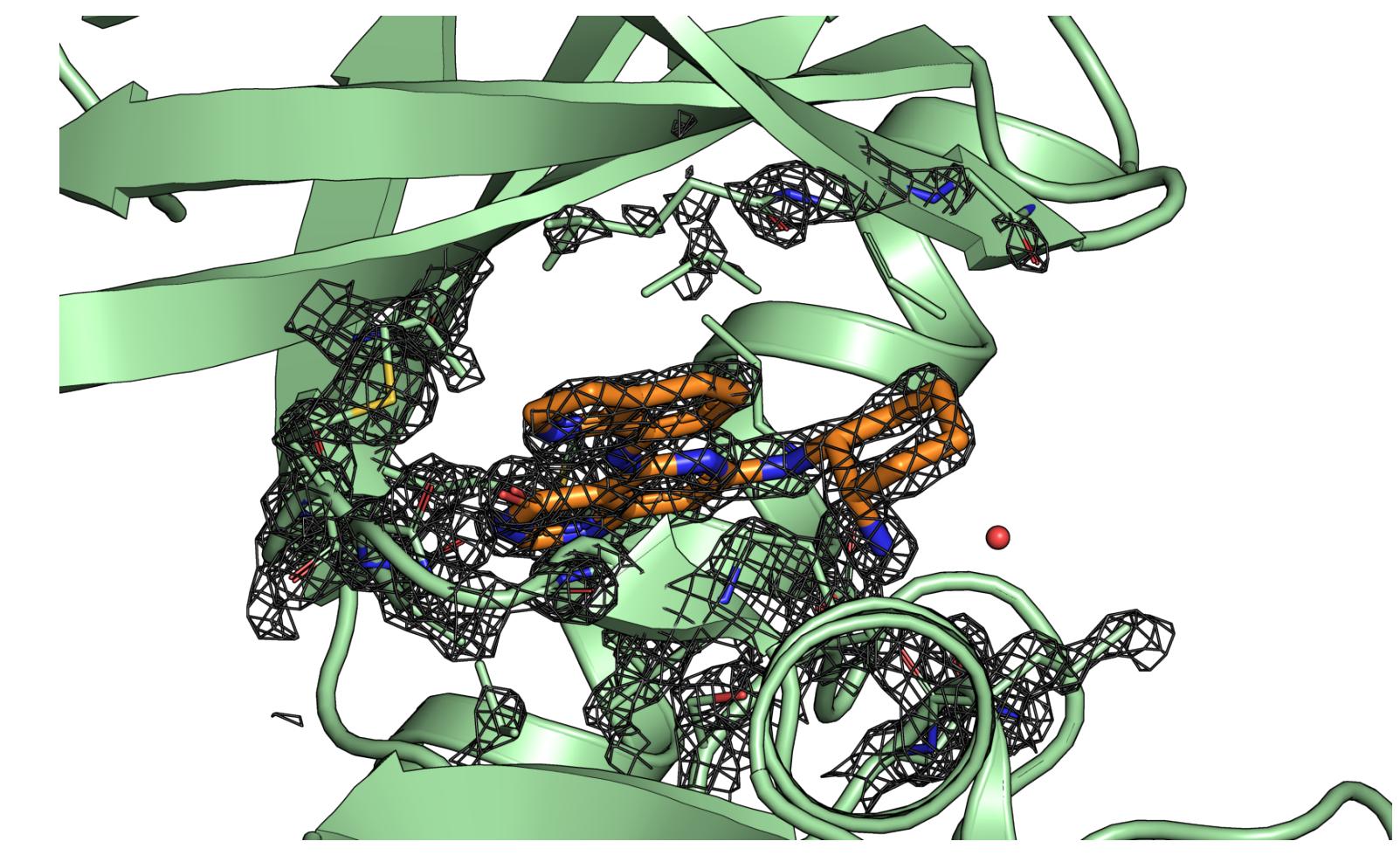
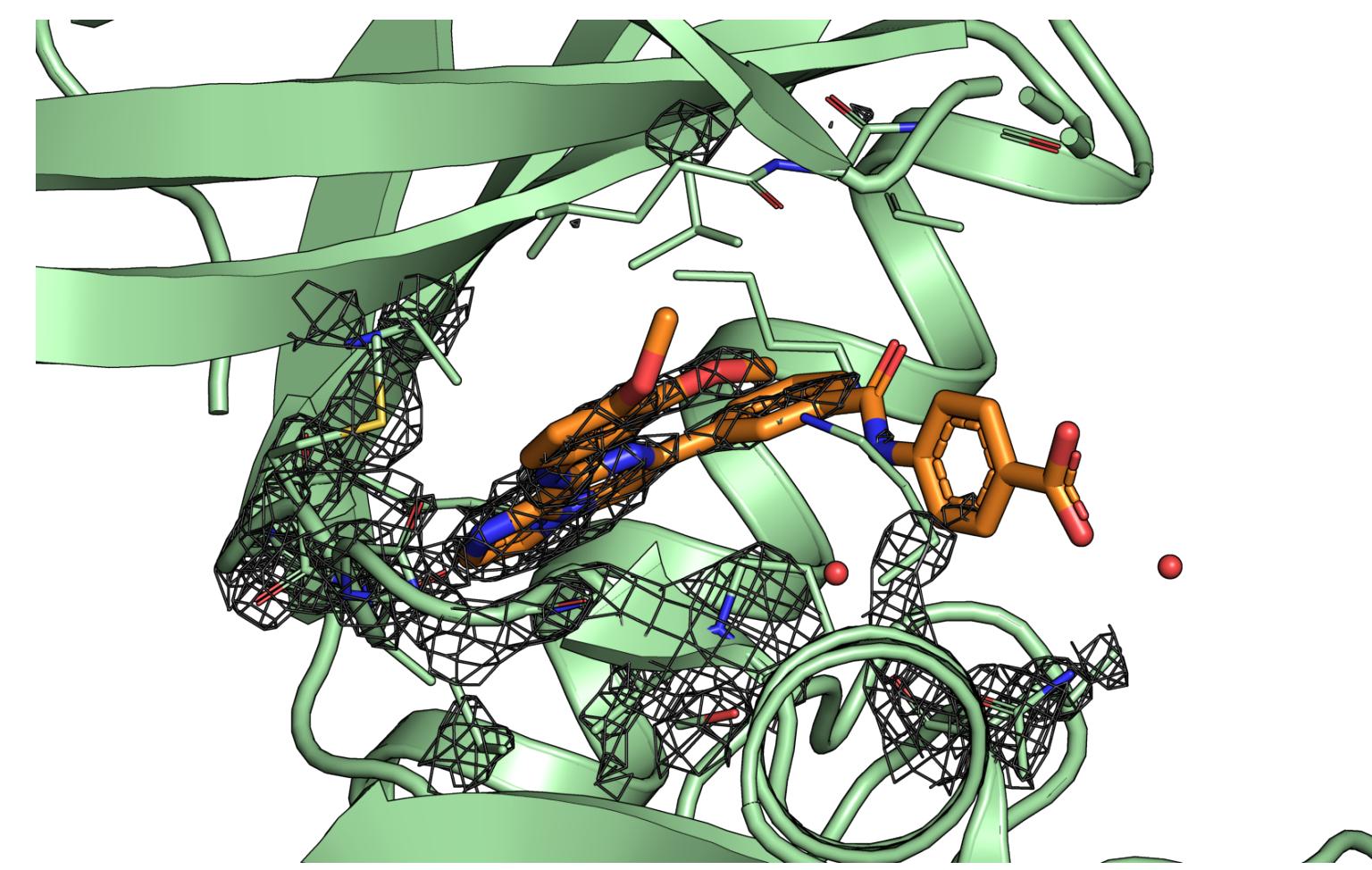


# Picking the best crystal structure requires care

BACE (Hunt)



spleen tyrosine kinase

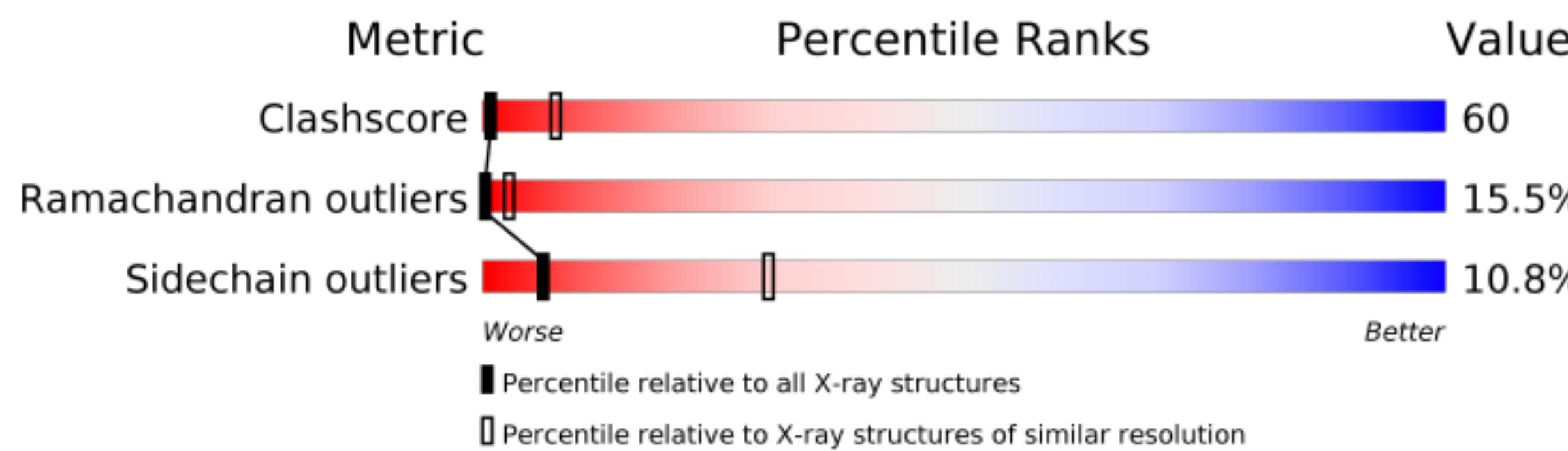


Which protein structure?

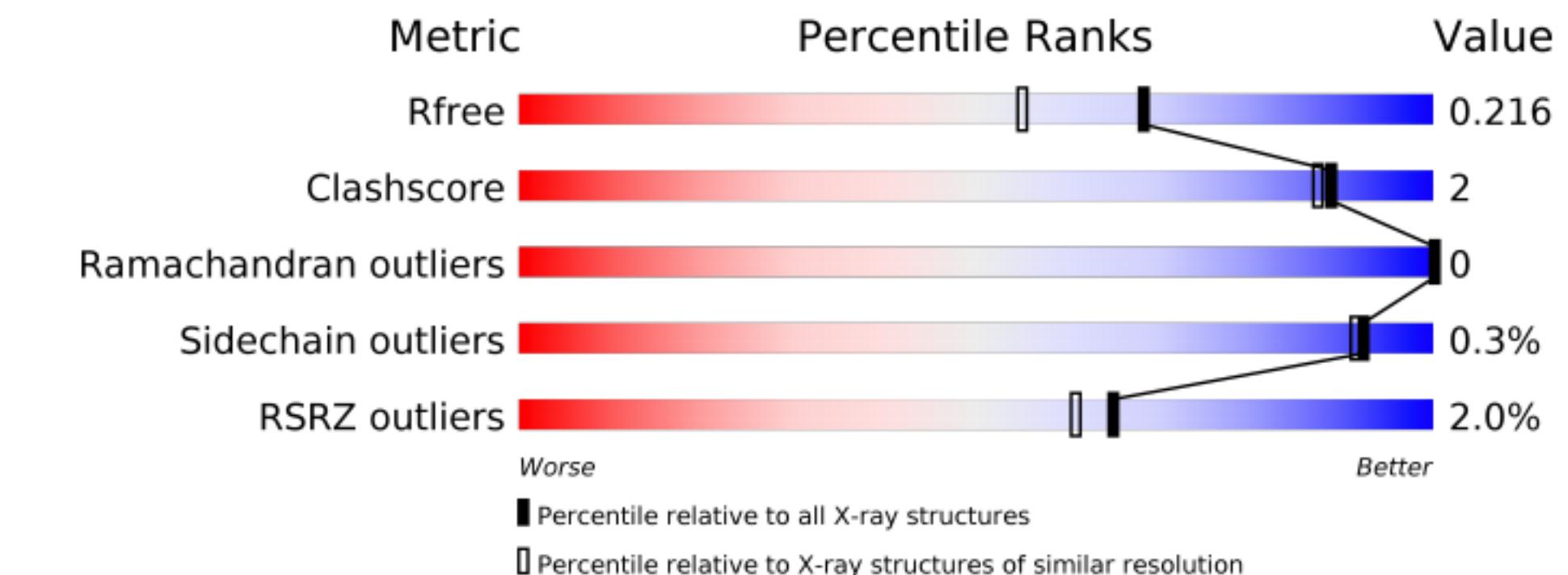
- Alternative side chain conformations need to be assessed carefully
- Active site residue densities are important for choosing the right crystal structure

# The RCSB PDB report can help with choosing structures

Jnk1 - 2GMX

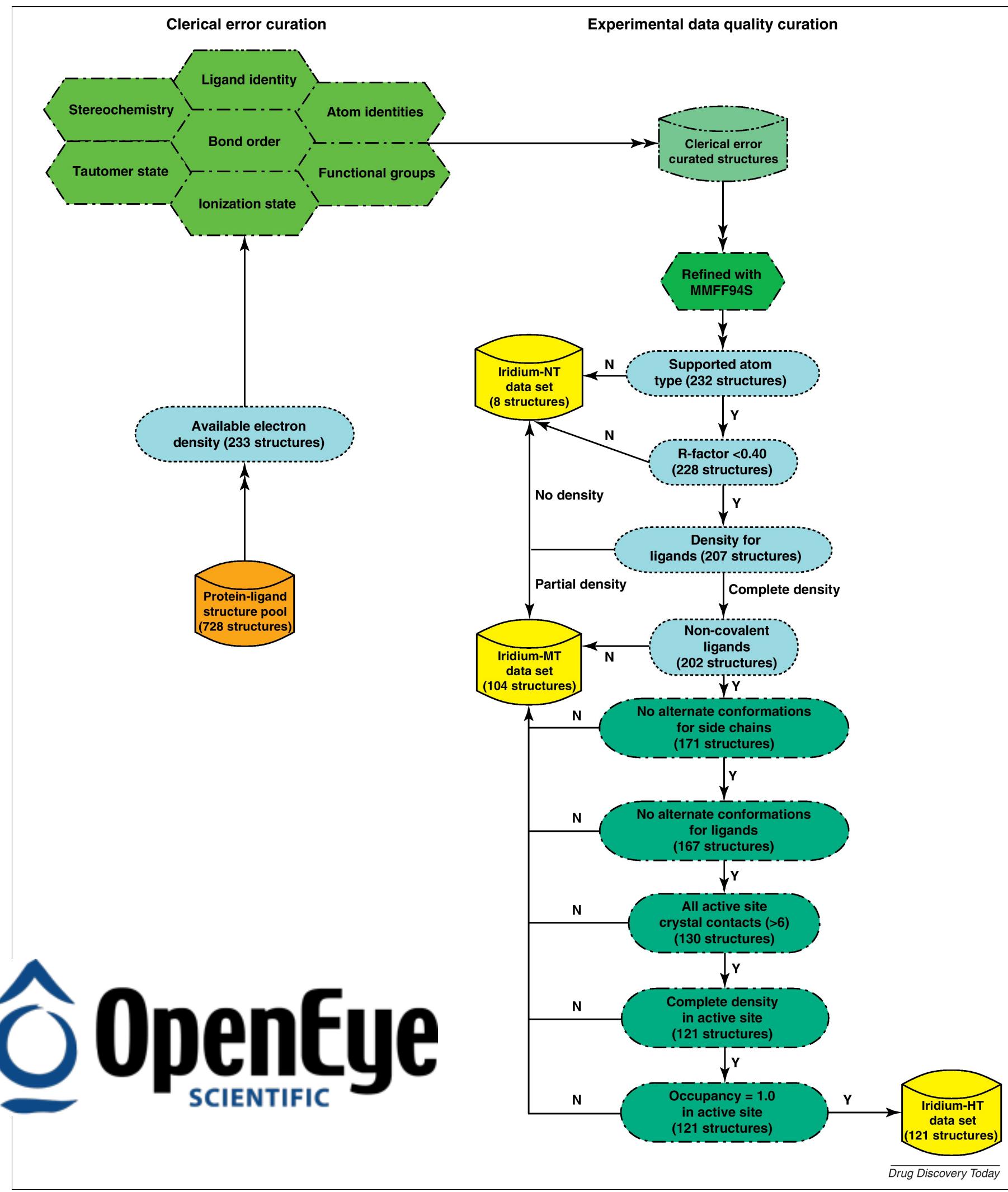


Jnk1 - 3ELJ



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

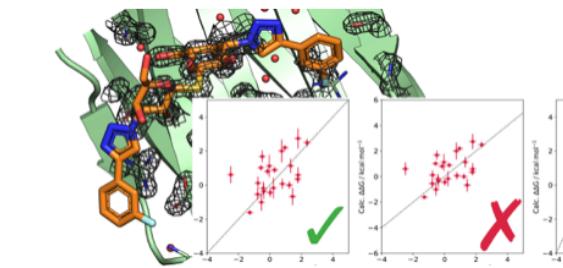
HOME / ARCHIVES / VOL. 4 NO. 1 (2022) / Articles



## 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](#)]

	6D09 (HT, 0.35) <sup>b</sup>	2GMX (NT, -) <sup>f</sup>	3ELJ (MT, 0.31) <sup>a</sup>	6O6F (HT, 0.30)	3V3V (MT, 1.5) <sup>b, h, e</sup>	4HW3 (HT, 0.41)	4ZBF (HT, 0.35) <sup>b</sup>	4WMU (HT, 0.41) <sup>b</sup>	4ZBI (HT, 0.45) <sup>b</sup>	21	3.4
	0.77	0.26								42	4.2

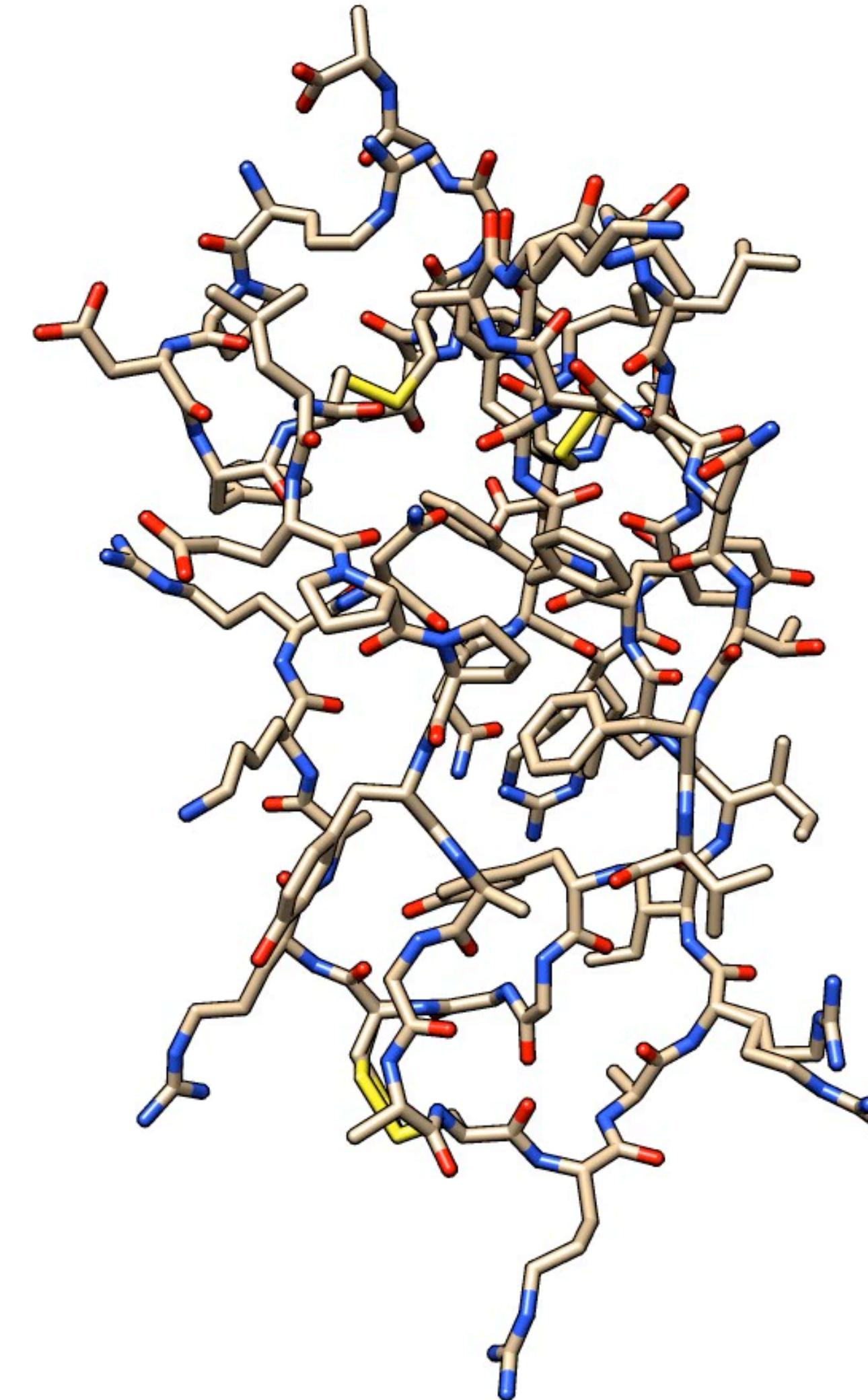
## 12 Features to inform the Iridium Score:

- 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
Active Site	> 0.95	< 0.95 and >	< 0.5

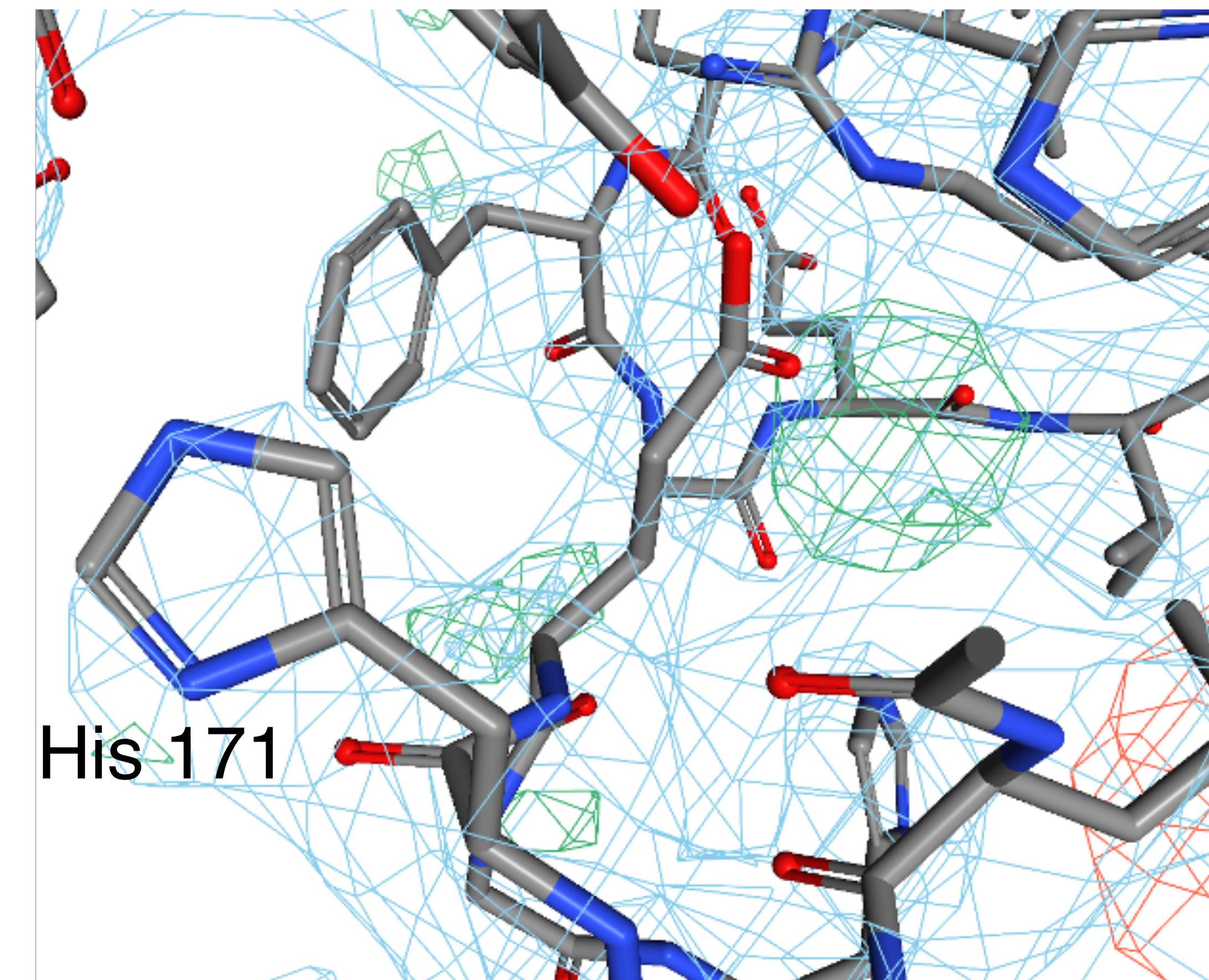
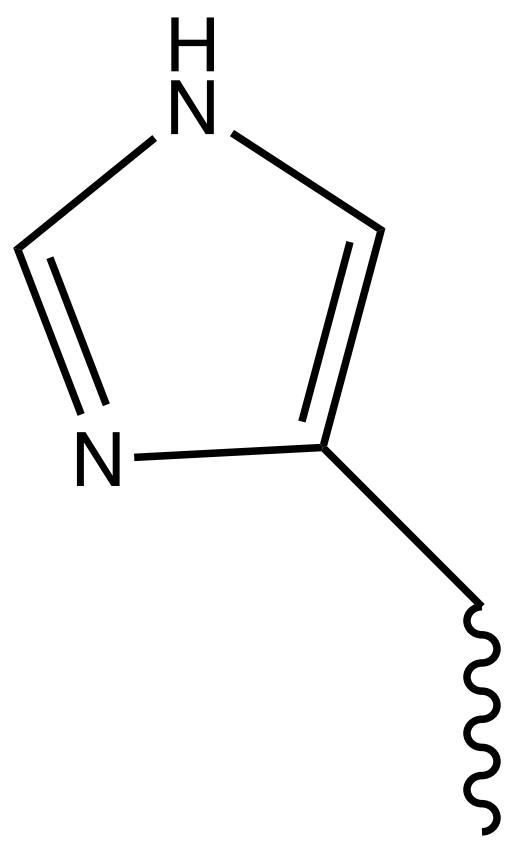
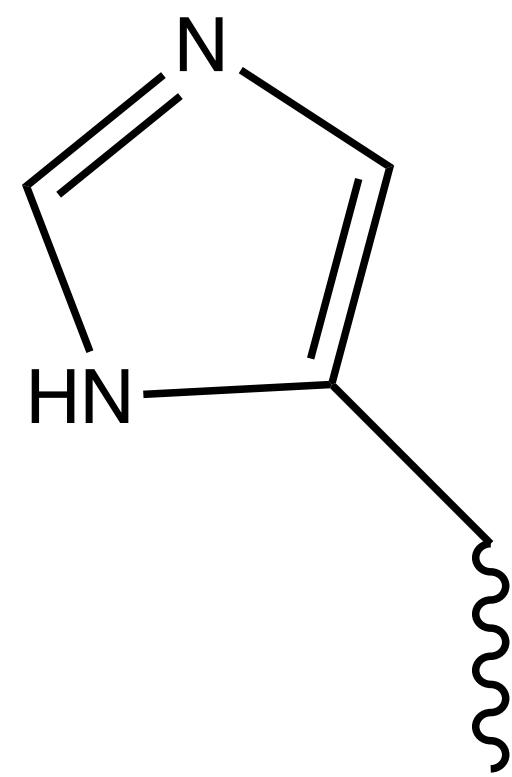
# Missing information in structures: Hydrogen atoms

- 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?



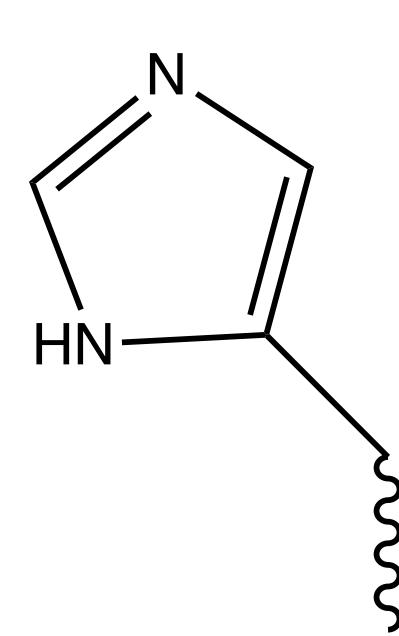
# How to choose the right tautomer?

Which tautomer?

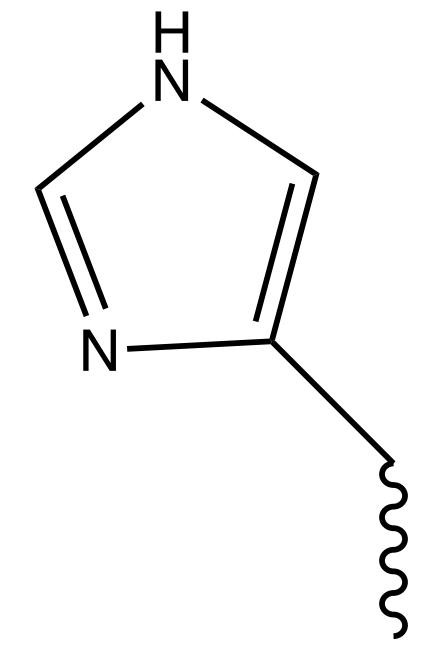


# How to choose the right tautomer and protonation state?

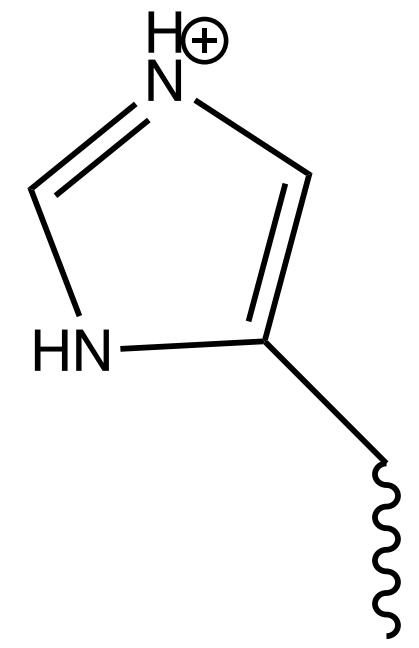
Which tautomer and protonation?



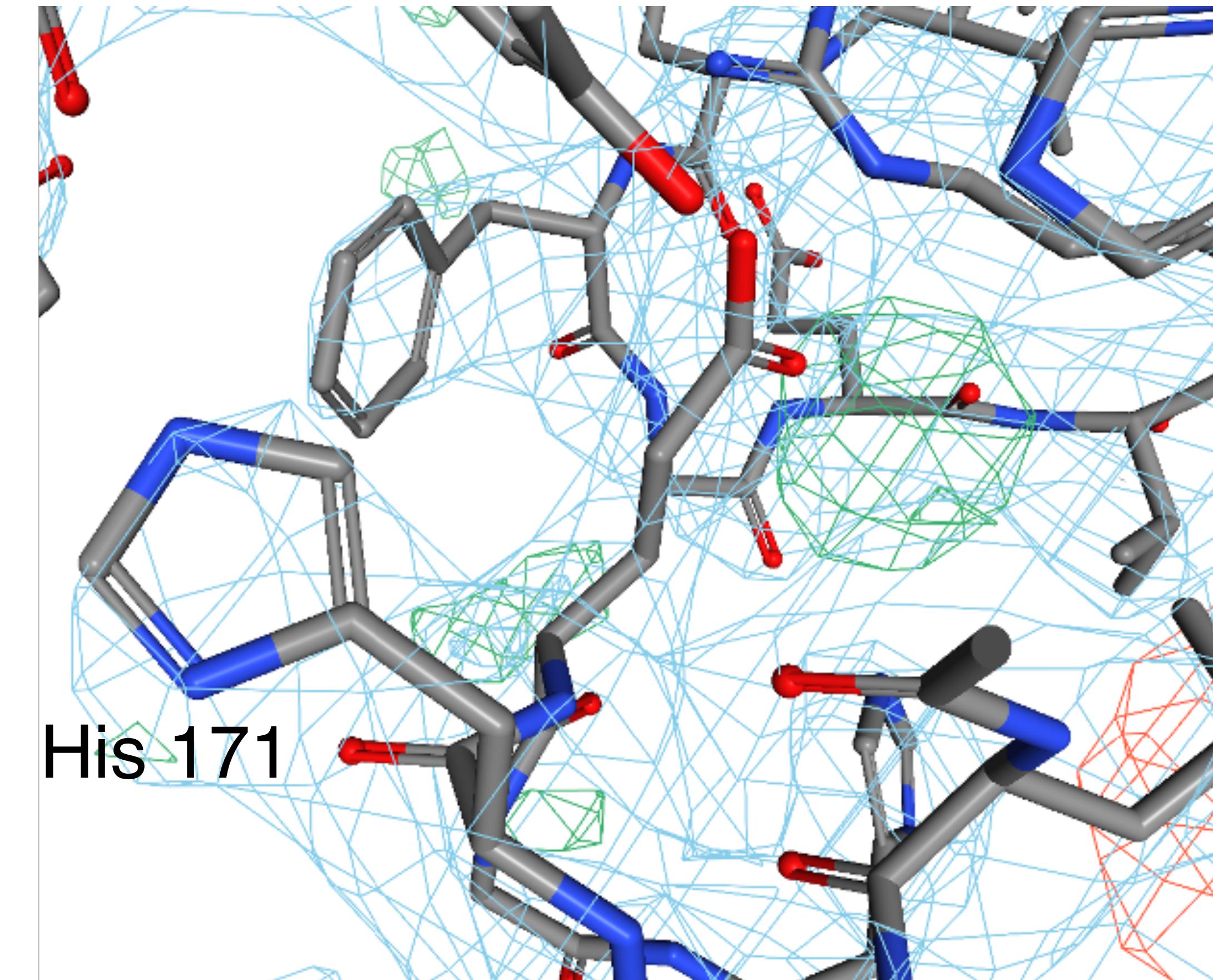
$\delta$ -tautomer



$\epsilon$ -tautomer

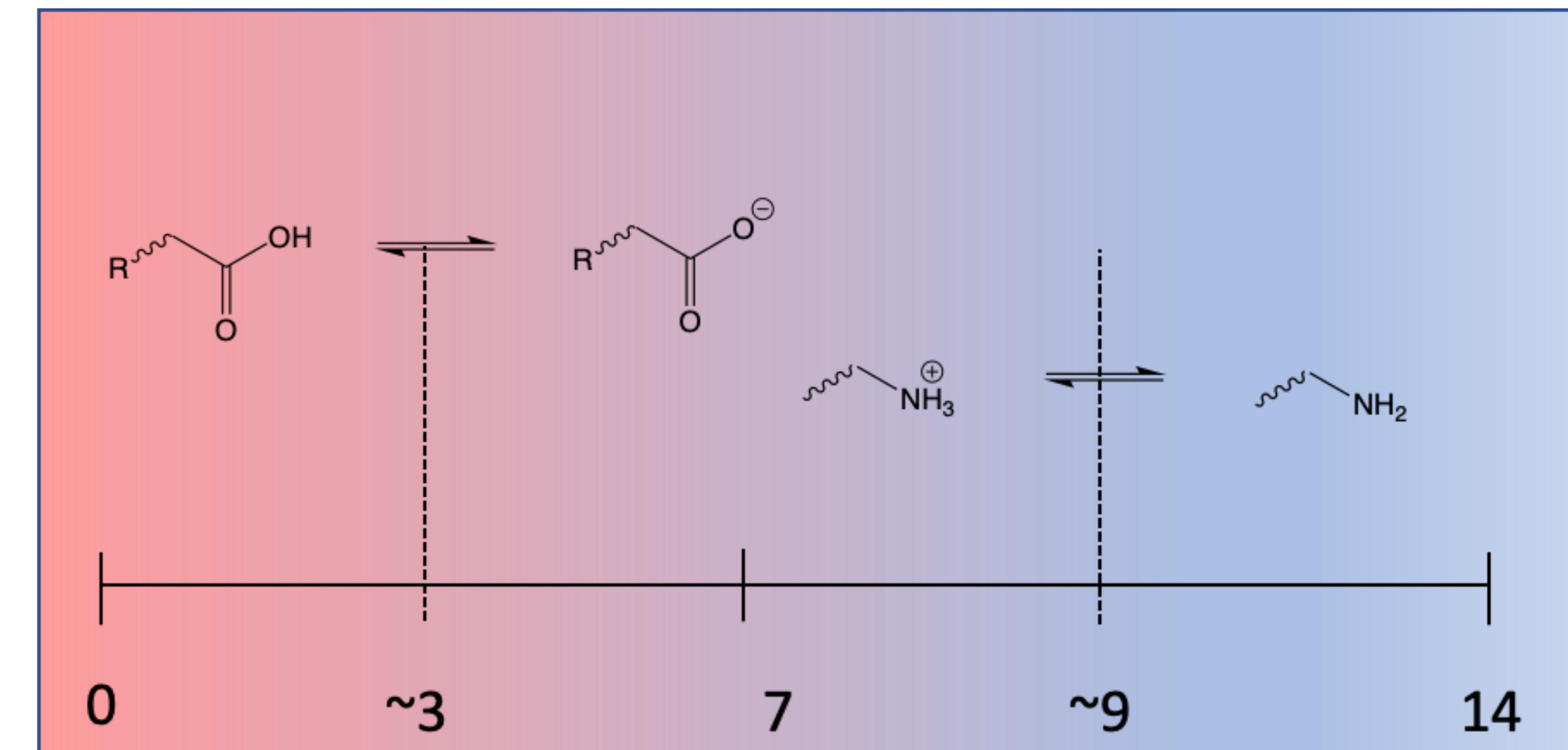


Protonated



# 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.

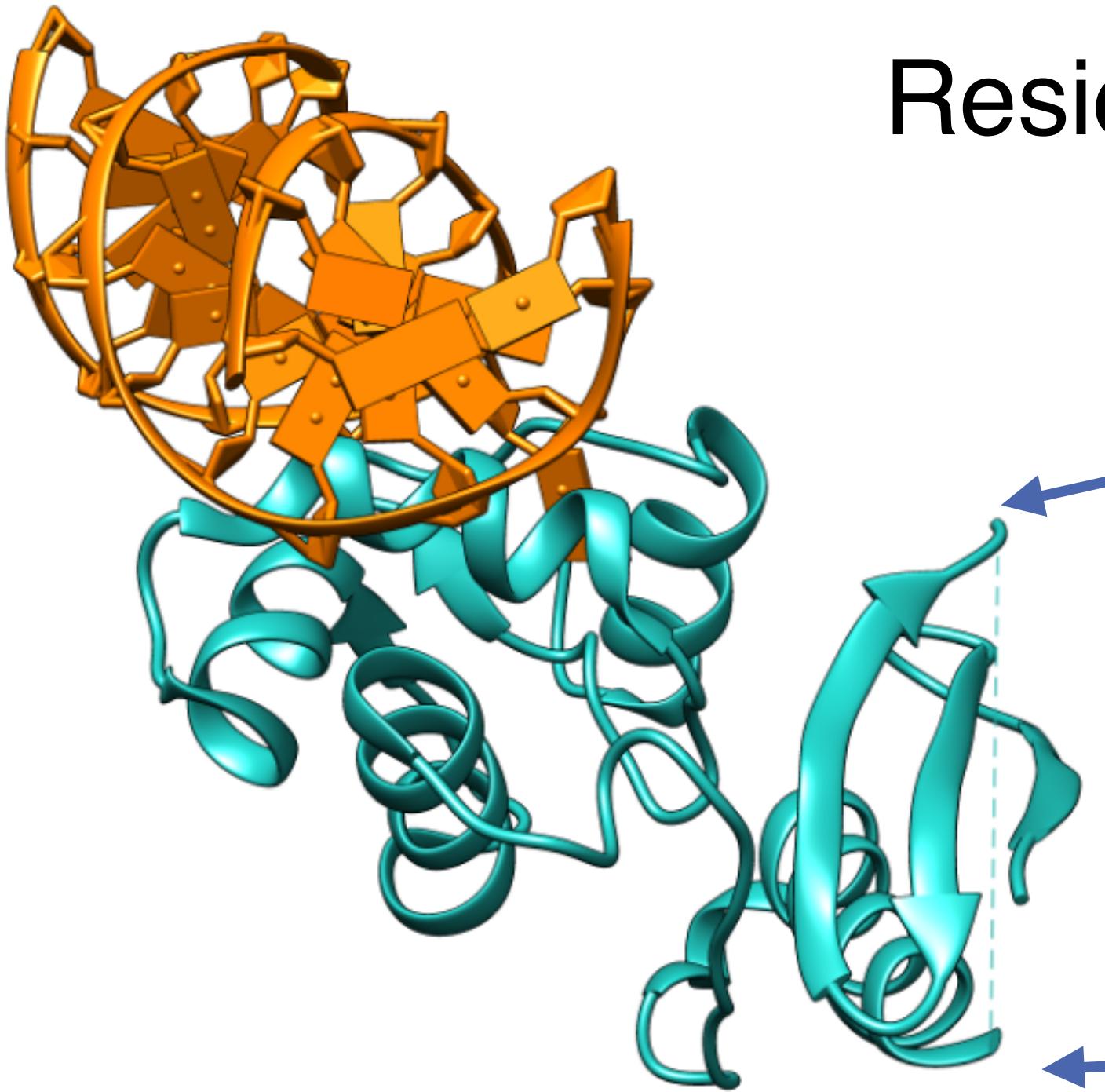


# Amino acids that need protonation state consideration

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 <sub>2</sub> instead of NH <sub>3</sub> <sup>+</sup> ?	possible
Arginine	12.48	neutral instead of protonated?	unlikely

\*for a simulation around physiological pH

# What if the best structure has missing residues?

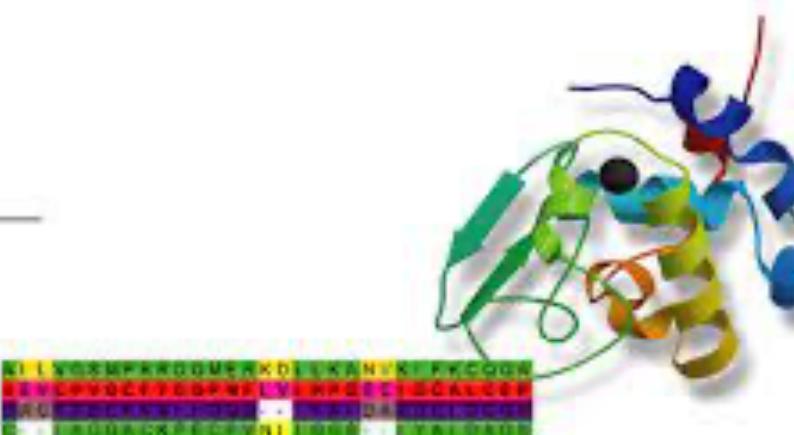


REMARK 465 MISSING RESIDUES  
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE  
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN  
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)  
REMARK 465  
REMARK 465 M RES C SSSEQI

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**

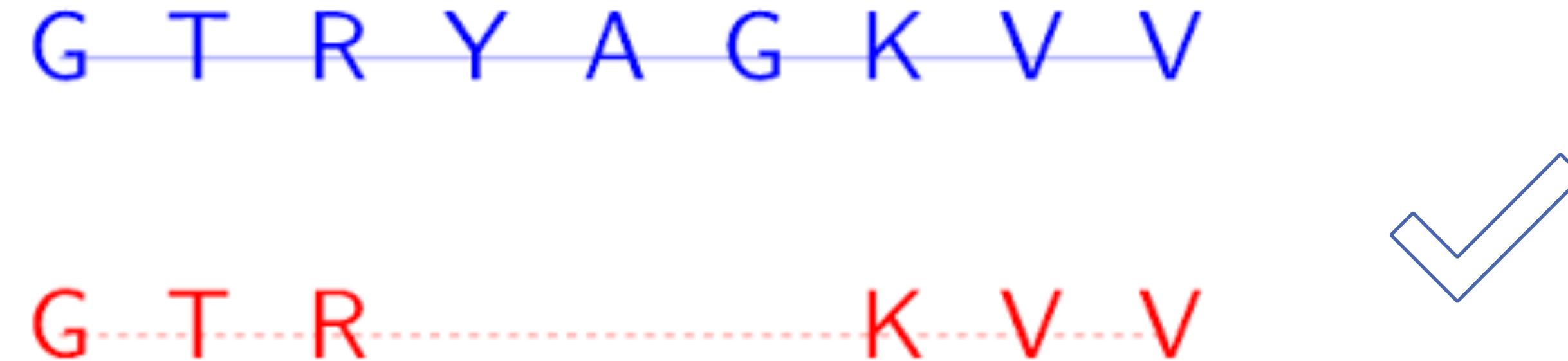
Program for Comparative Protein  
Structure Modelling by Satisfaction  
of Spatial Restraints



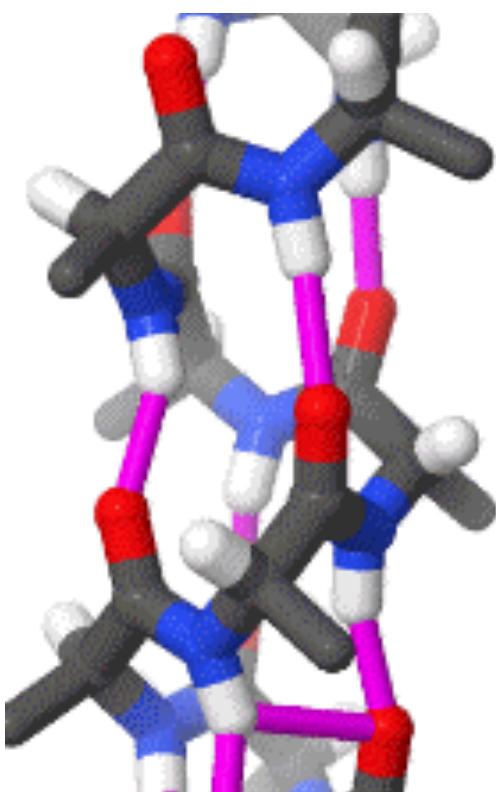
**PDB  
FIXER**

**AlphaFold**

# AlphaFold2 structures have no missing residues or atoms



No missing residues



Has hydrogens



# AlphaFold2 structures can have unreliable areas

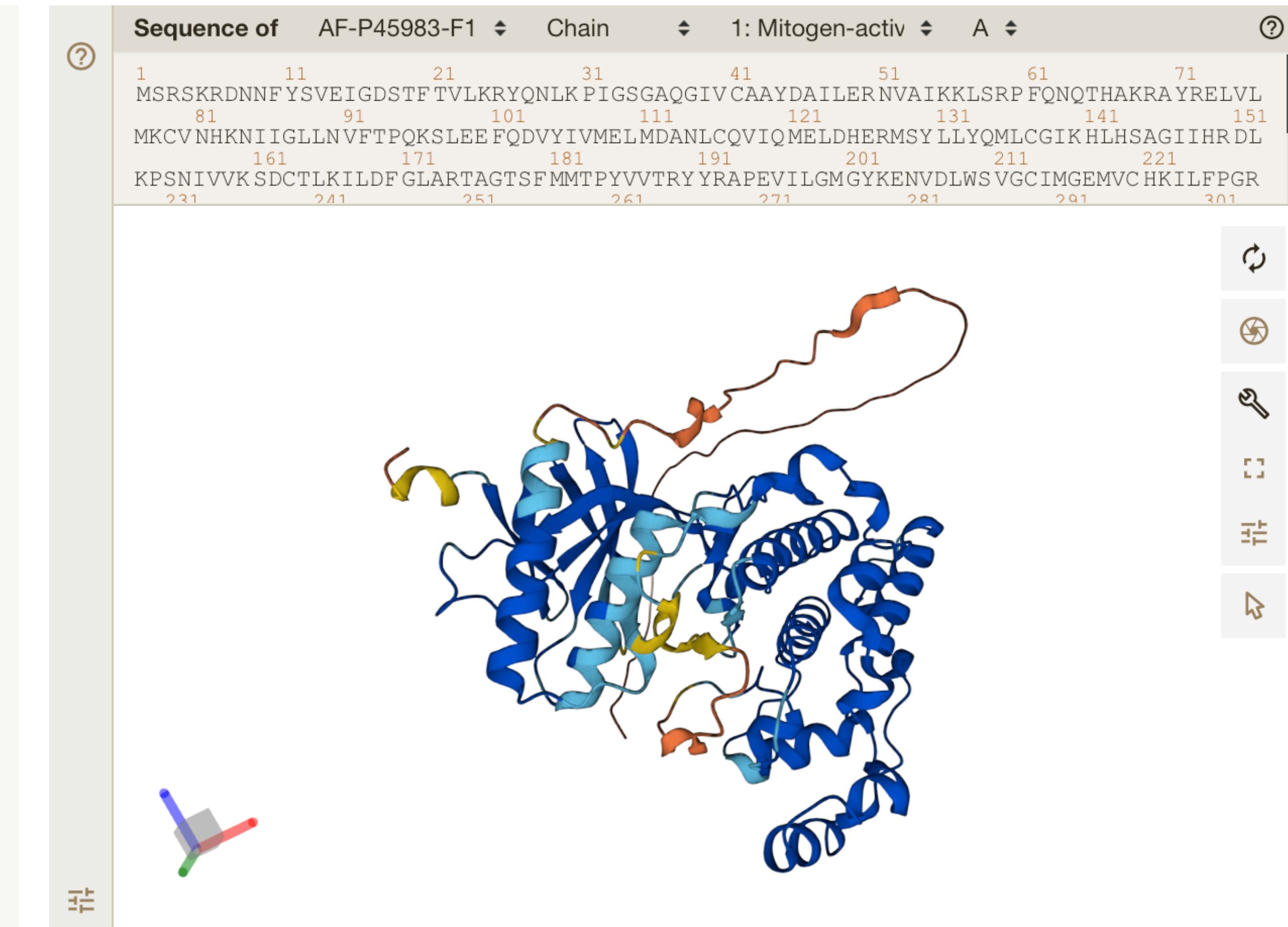
## JNK1 – again

### 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.



# What is missing in this very good structure?

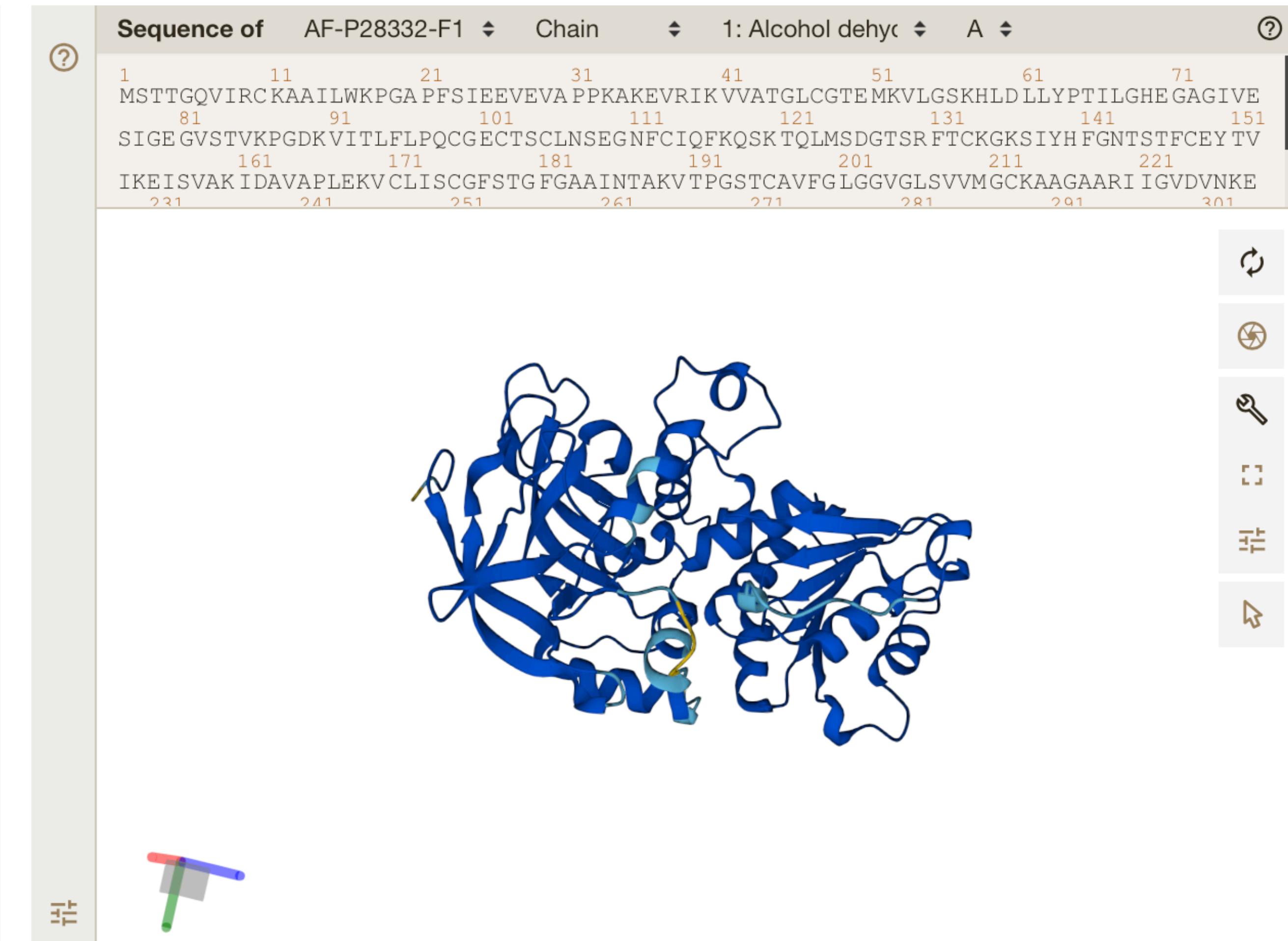
## Alcohol dehydrogenase

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.



# 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:

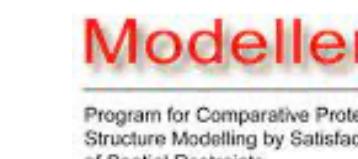
<https://github.com/jensengroup/propka-3.1>

## PDB2PQR server

Interface to both Reduce and  
proPka

Website:

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

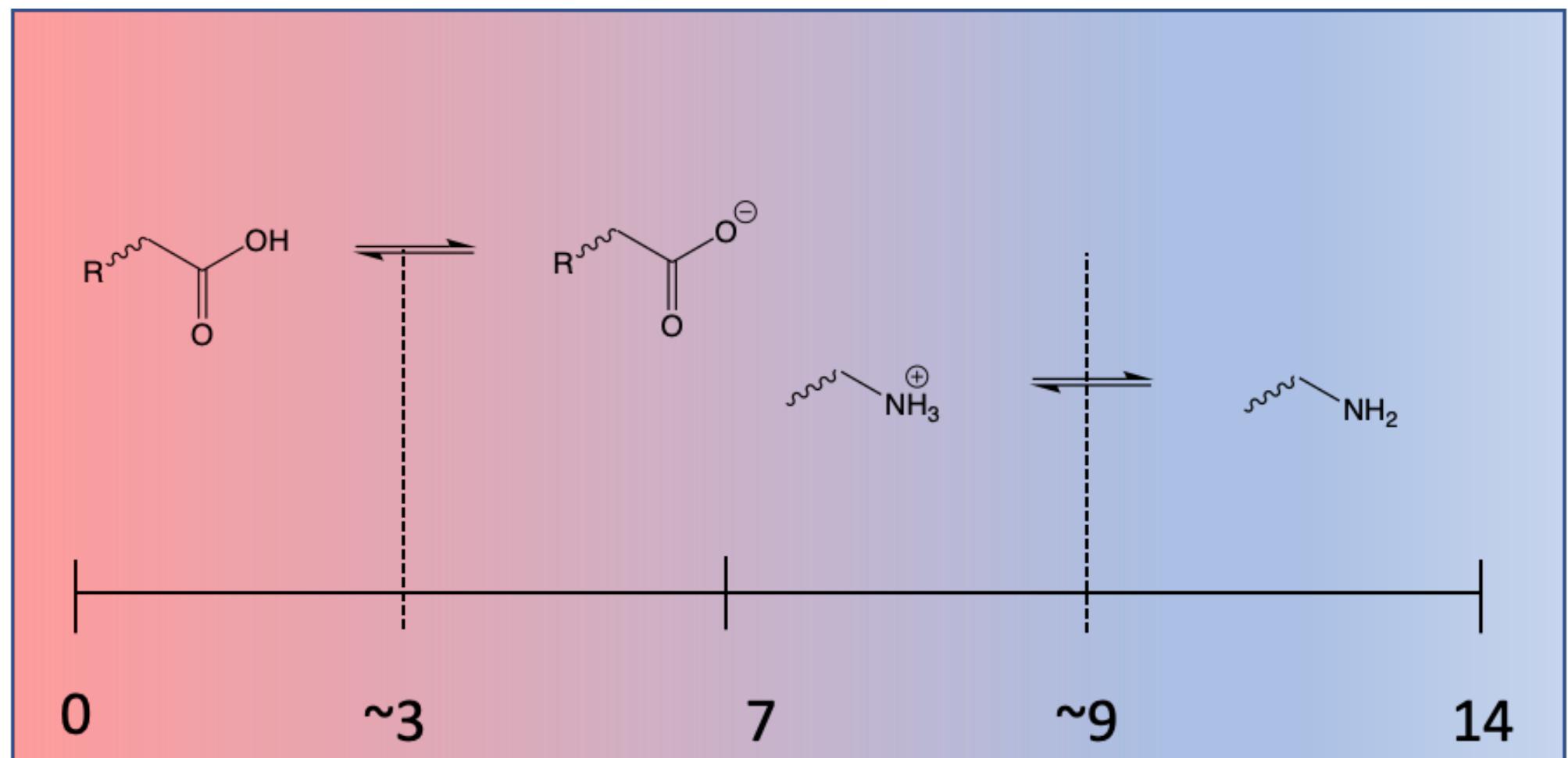


...

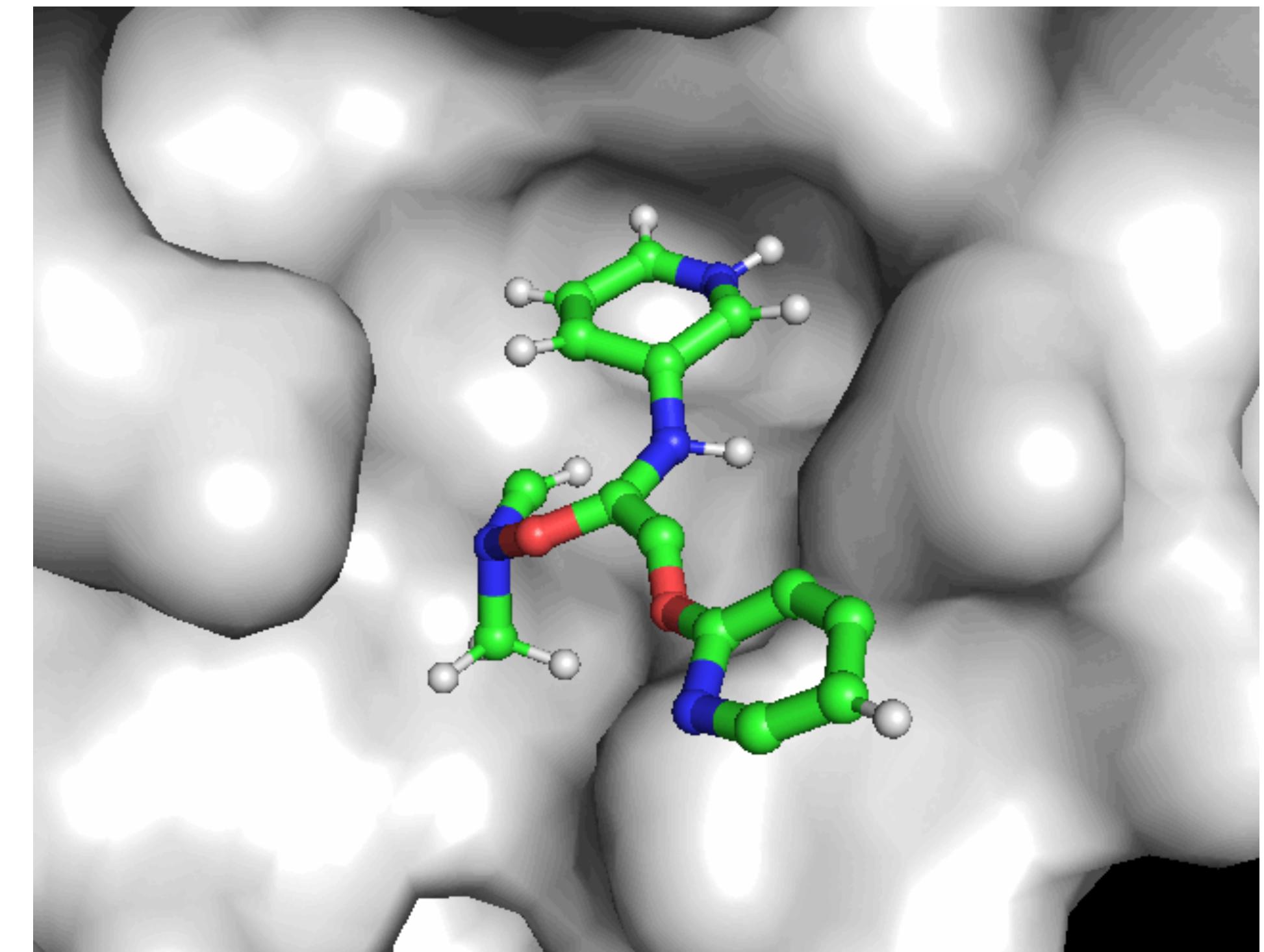
**Let's try some protein prep out!**

# What about ligands and co-factors?

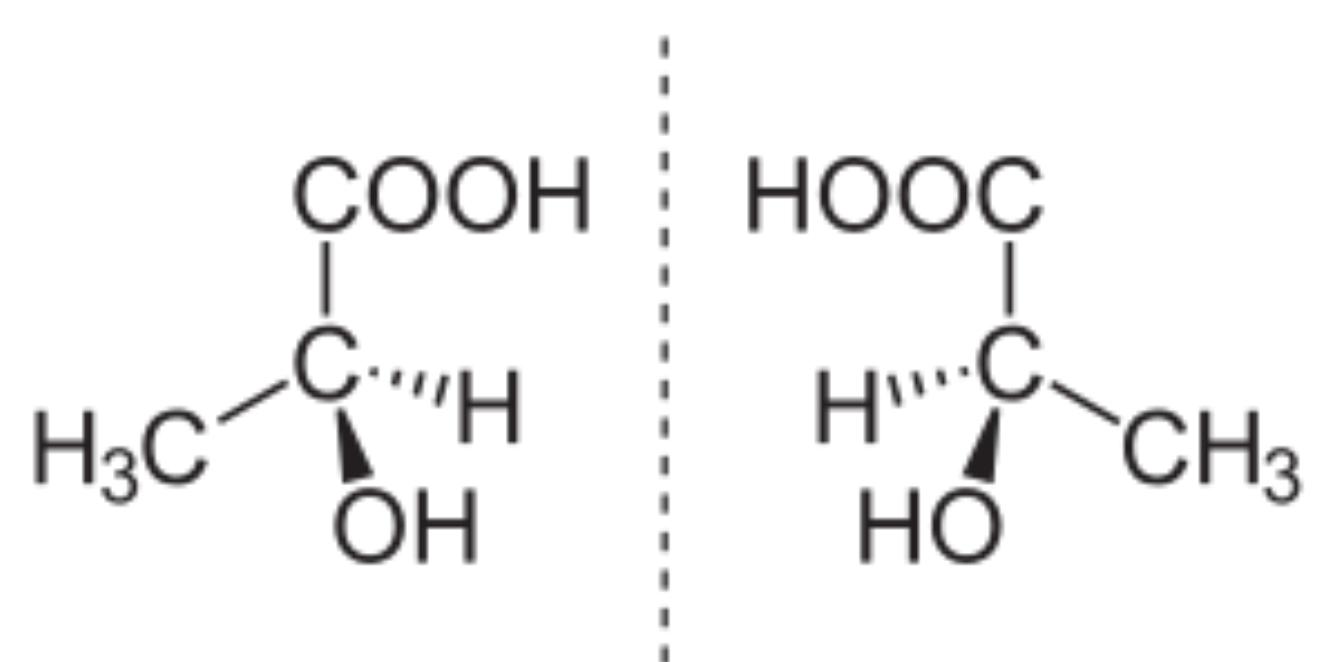
pKa for ligands is also important!



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



Choosing the right enantiomer



**Next: How to get ligand(s) into a protein and run a virtual screening after coffee**

