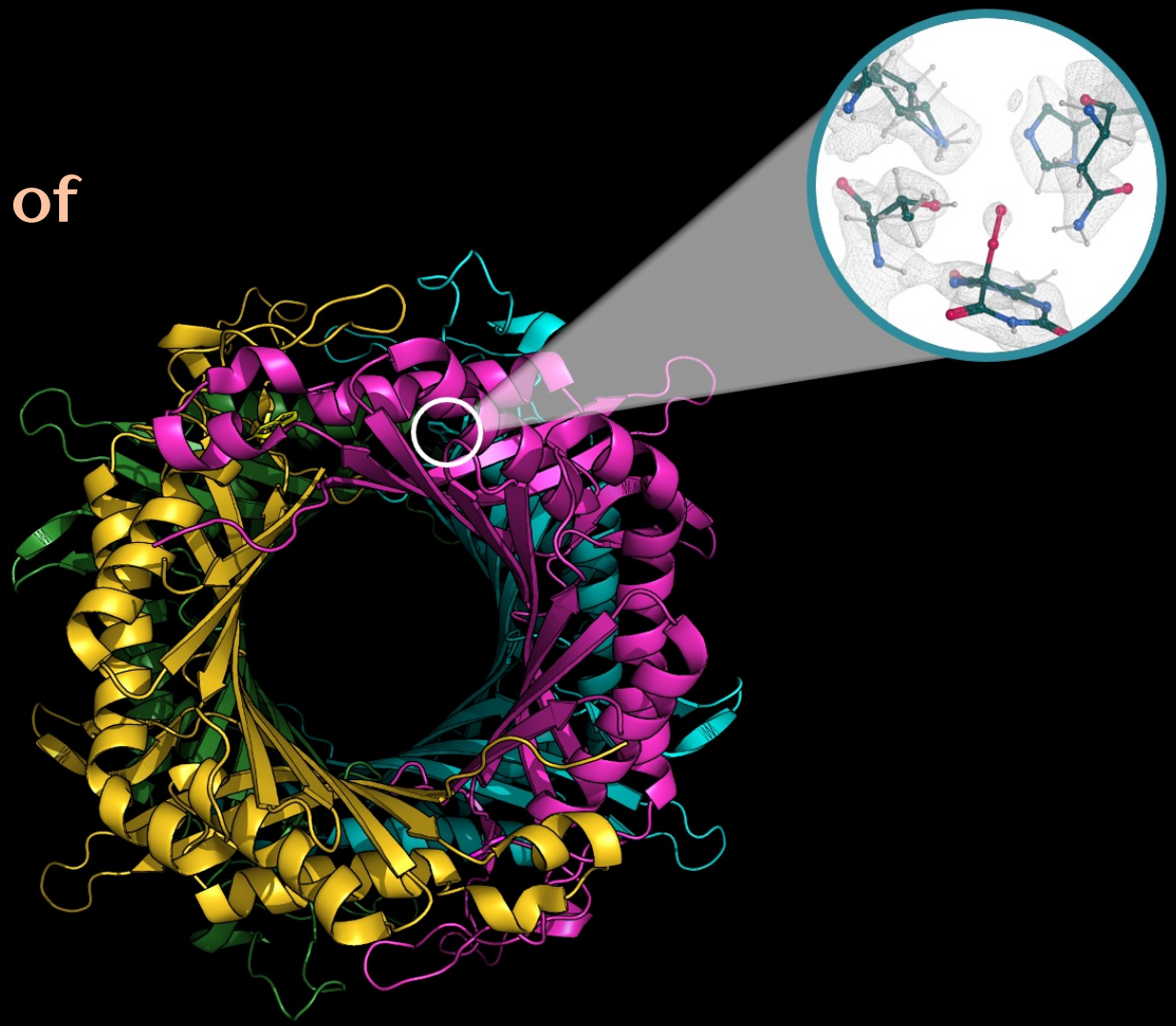
Enhancing Structural Refinement of Macromolecules obtained from Neutron Crystallography



Lucrezia Catapano

ECM34 Satellite-Early Science on the NMX Macromolecular Diffractometer at the European Spallation Source

26th August 2024, Padova, Italy

Computational Structural Biology Group

Software for macromolecular structure analysis

Lucrezia Catapano[1,2], Paul Emsley[1], Fei Long[1], Robert A. Nicholls[1], Rangana Warshamanage[1], Keitaro Yamashita[1], Garib N. Murshudov[1]

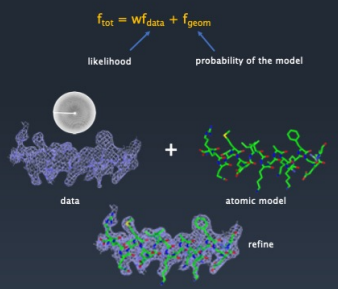
[1]MRC Laboratory of Molecular Biology, Cambridge, UK, [2]Randall Centre for Cell and Molecular Biophysics, King's College London, UK

Research in the Computational Structural Biology Group focuses on the development of mathematical, statistical, and computational techniques to derive reliable atomic models from noisy and incomplete experimental data. We use Bayesian statistics machinery to allow the joint use of data from several sources, such as stereochemical and structural prior knowledge, and data derived from crystallography and cryo-EM. Our group has developed such software tools as *REFMAC5* and *Servalcat* for atomic model refinement and updated map generation; *AceDRG*, *ProSMART* and *LIBG* for the generation and organisation of restraints; *BALBES*, an automatic molecular replacement pipeline; *EMDA* for map manipulation and Maximum Likelihood map averaging; *Coot* for model building, validation, visualisation, and real-space refinement of atomic models of biological macromolecules. Our computational tools are distributed as a part of the CCP4 and CCP-EM suites.

Macromolecular refinement

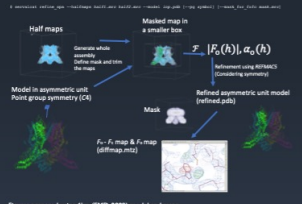
REFMAC5

REFinement of MACromolecular Structures. It uses the Maximum Likelihood method and some elements of Bayesian Statistics to perform full model refinement and map calculation. Originally designed for use with data from crystallography, REFMAC5 has been adapted and extended to support data from other sources including cryo-EM (Murshudov *et al.*, 2011)



Servalcat

Structure refinement and validation for crystallography and single particle analysis. It is a Python package and standalone program for the refinement and map calculation of cryo-EM SPA structures. It implements a refinement pipeline using REFMAC5 (Yamashita *et al.*, 2021)

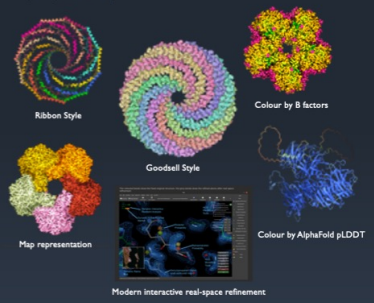


Figures prepared using 6qp (EMD-0882) model and maps

Model building & validation

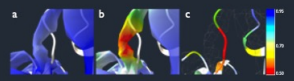
Coot

Crystallographic Object-Oriented Toolkit for macromolecular model building, model completion and validation. Particularly suitable for protein modelling using MX and cryo-EM data (Emsley *et al.*, 2010)



EMDA

Electron Microscopy Data Analytical toolkit. It is a Python library module for Electron Microscopy map and model manipulations (Warshamanage *et al.*, 2022)

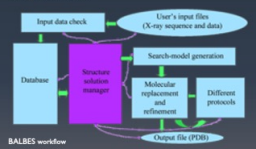


Identifying model-map discrepancies by local correlation. a) EMD-5623 primary map density near residues Lys52-Val54 of chain U of the 3JYI model coloured by, $CC_{map,model}$. b) the same density coloured by $CC_{map,model}$. c) corresponding atomic model coloured by $CC_{map,model}$ showing low correlation residues Lys52-Val54 of chain U.

Molecular replacement

BALBES

Automatic molecular replacement pipeline. A system for solving protein structures using X-ray crystallographic data, which aims to integrate all components necessary for finding a solution structure by molecular replacement. (Long *et al.*, 2008)

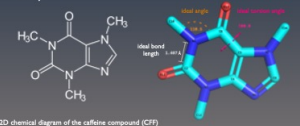


BALBES workflow

Restraints generation

AceDRG

Stereo-chemical description generator for monomers/ligands and links. Encapsulates information about local chemical and topological environments derived from a small molecule database (the Crystallography Open Database) (Long et *al.*, 2017)

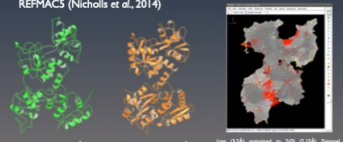


2D chemical diagram of the caffeine compound (CFF)

ideal bond lengths, angles & torsion angles for CFF

ProSMART

Procrustes Structural Matching Alignment and Restraints Tool. It has two main purposes: conformation-independent comparison of protein structures, and the generation of interatomic distance restraints for subsequent use in macromolecular crystallographic refinement by REFMAC5 (Nicholls *et al.*, 2014)



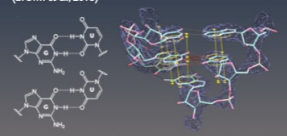
Low resolution 3.5Å structure (1ryx)

High resolution 2.15Å homologue structure (2d3i)

1ryx (3.5Å) restrained to 2d3i (2.15Å). External restraints generated by ProSMART are visualised in Coot. Grey lines indicate similar interatomic distances between the low resolution structure and the homologue. Red lines imply the interatomic distances of the low resolution structure are longer and blue that they are shorter.

LIBG

LIBG generates restraints to help stabilise refinement of DNA/RNA models. Restraints are generated for base-pairs, stacking planes, sugar puckers, and other torsion angles (Brown *et al.*, 2015)



G:U base pair shown in (top) wobble and (bottom) reverse wobble configuration

LIBG restraints in Coot. Grey: G:U wobble restraints. red: G:U reverse wobble configuration restraints. yellow: parallel-plane restraints

Acknowledgement

CCP4 CCP-EM

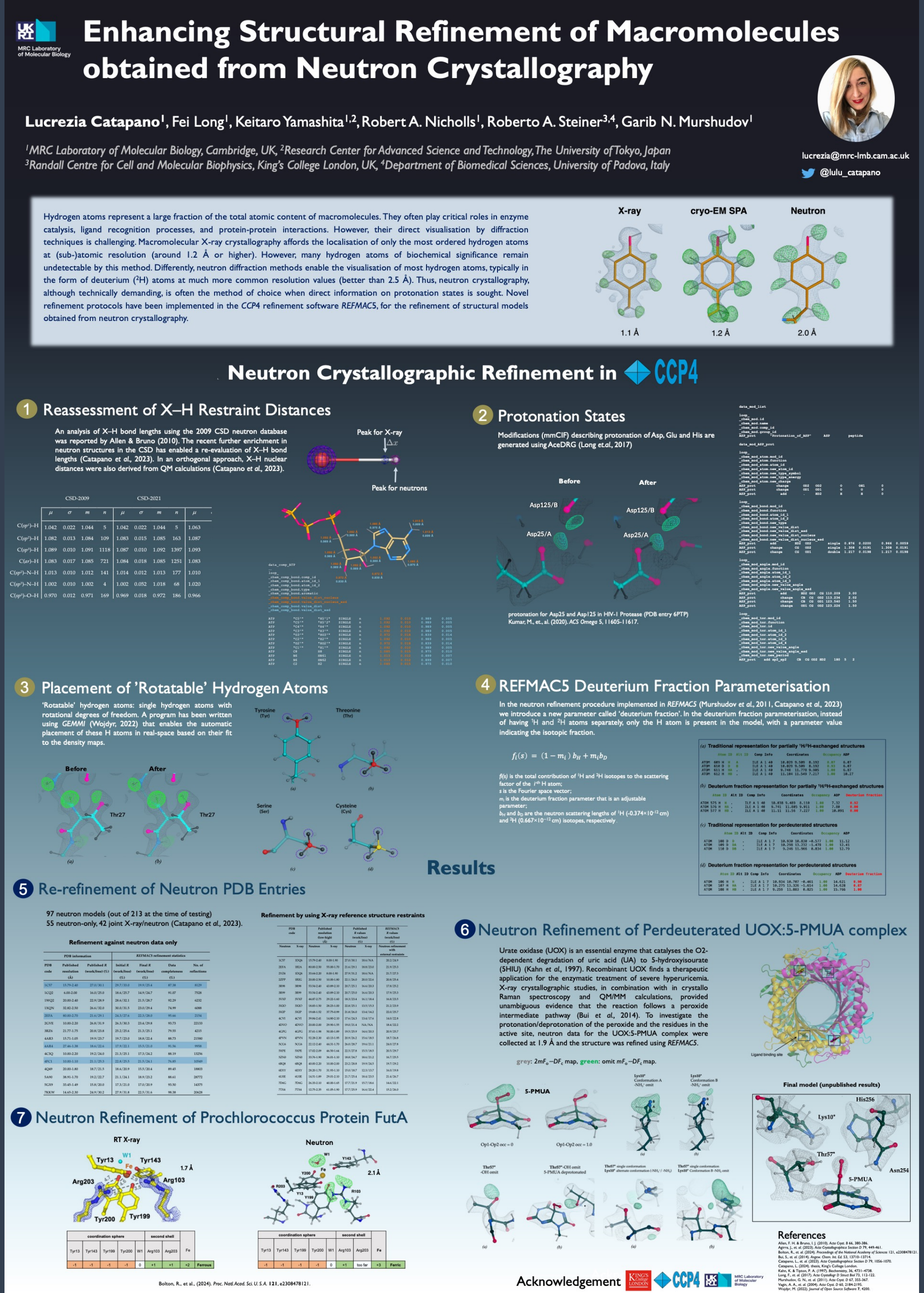
REFERENCES

Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P. & Murshudov, G. (2015). Acta Cryst. D 71, 136–153.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D 66, 486–501.
Long, F., Nicholls, R. A., Emsley, P., Graeulis, S., Merkys, A., Vaitkus, A. & Murshudov, G. N. (2017). Acta Cryst. D 73, 112–122.
Long, F., Vagin, A.A., Young, P. & Murshudov, G. N. (2008). Acta Cryst. D 64, 125–132.
Nicholls, R. A., Fischer, F. T., Nicholls, E. & Murshudov, G. N. (2014). Acta Cryst. D 70, 2487–2499.
Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T. & Bricogne, G. (2011). Acta Cryst. D 67, 293–302.
Warshamanage, R., Yamashita, K. & Murshudov, G. N. (2021). J. Struct. Biol. 214, 107826.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). Acta Cryst. D 67, 355–367.

https://www2.mrc-lmb.cam.ac.uk/groups/murshudov/

MRC-Laboratory of Molecular Biology

Cambridge, UK



Wednesday 28th of August

Poster ID: 613

First floor, panel nr. 28

Outline

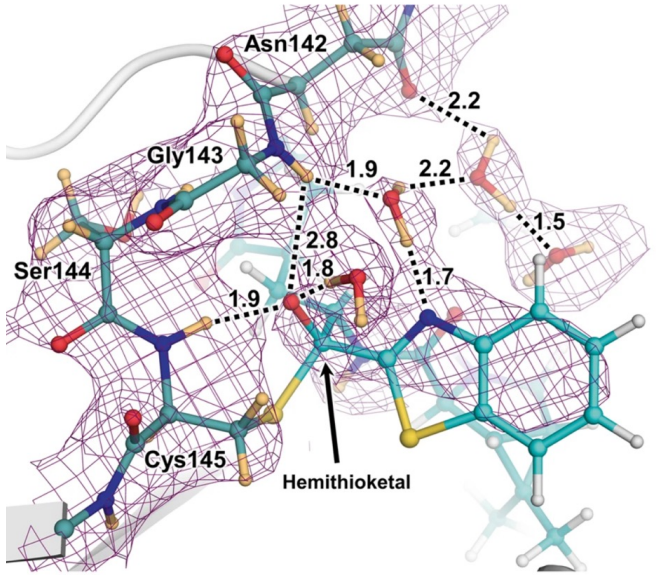
- Hydrogen atoms in Structural Biology
- Basics of Neutron Crystallography
- Neutron refinement within REFMAC5
- Re-refinement of neutron PDB entries
- Neutron studies of iron binding protein FutA

H atoms account for
~50% of protein atoms
~35% of nucleic acid atoms

Key roles in **enzyme chemistry**

^{1}H ydrogen atoms matter!

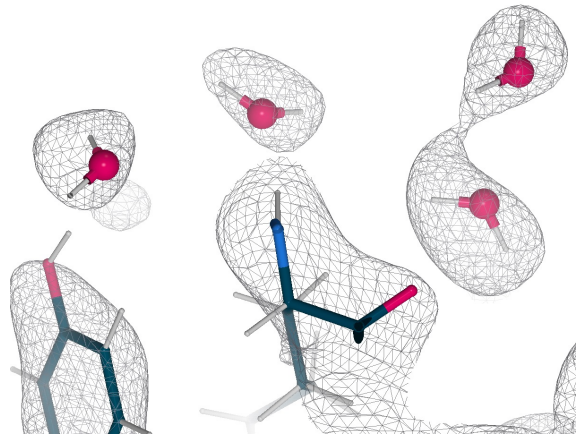
Ligand binding preferences



BBH-1 binding in M^{pro} active site

Kneller, D. W., et al. (2022). *Nature Communications* 13, 2268.

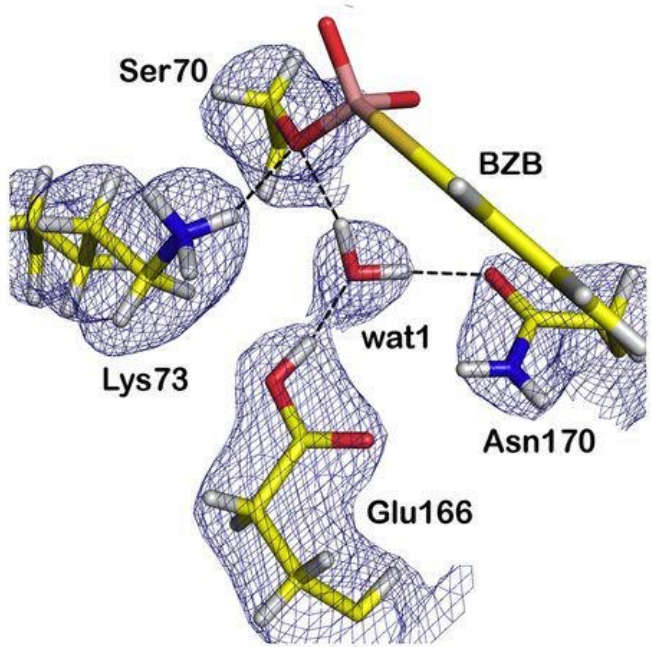
Water molecules organisation



Water network in concanavalin A (PDB code:1XQN)

Blakeley, M. P., et al. (2004). *PNAS* 101, 16405-16410.

Protonation & orientation

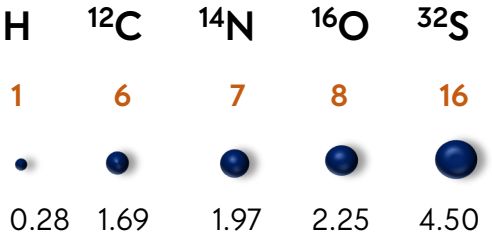


Protonation states of the active site of Toho-1 β-lactamase in complex with BZB

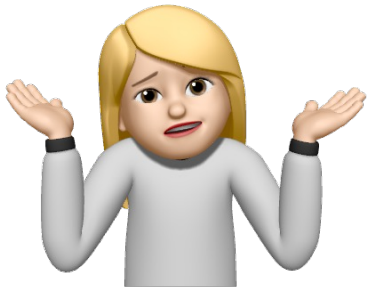
Tomanicek, S. J., et al. (2013). *J. Biol. Chem.* 288, 4715-4722.

Hydrogen atoms 'invisible' in X-ray structures

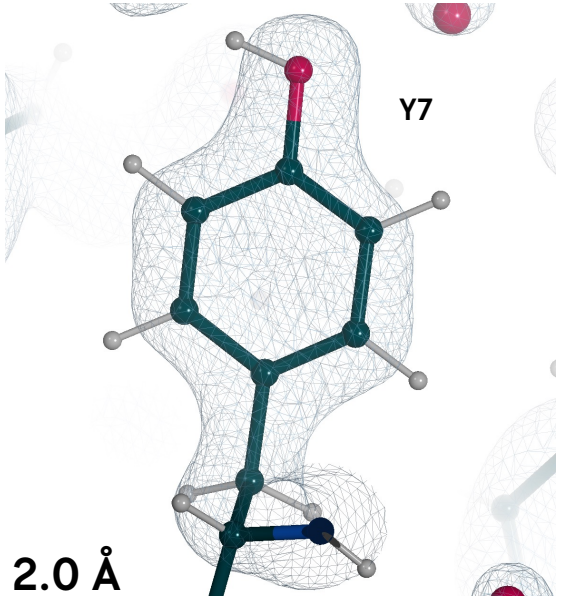
X-rays interact with **electron cloud**



X-ray scattering (10^{-12} cm) proportional to Z

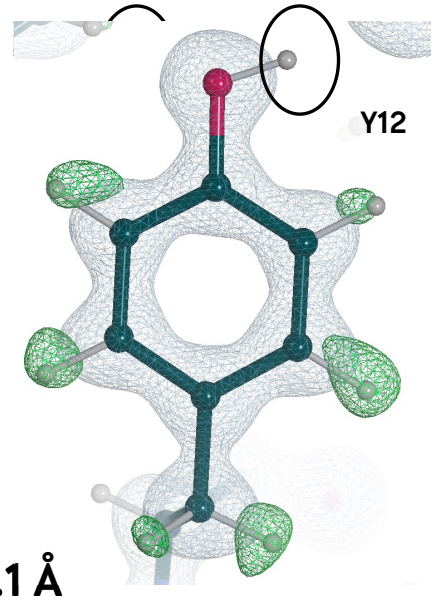


H atoms are **invisible** in electron density maps at typical resolutions **(2.0 Å)**



PDB code: 1TBT
grey: $2mF_o-DF_c$ electron density map
Fisher, Z., et al., (2005). *Biochemistry* 44, 1097-1105.

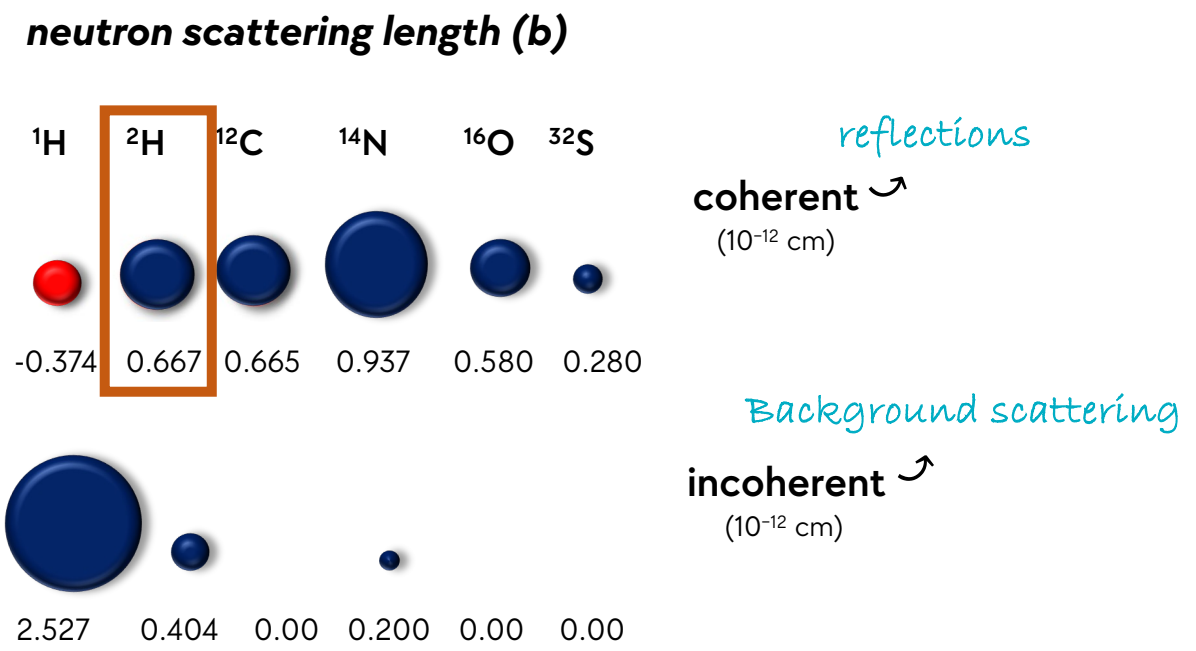
Some H atoms are often **undetected** at (sub-)atomic resolution **(<1.2 Å)**



PDB code: 3KYU
grey: $2mF_o-DF_c$ electron density map
green: hydrogen-omitted $mF_o - DF_c$ map
Gardberg, A. S., et al., (2010). *Acta Cryst.* D66, 558–567.

Neutrons reveal H atom positions

- Neutrons interact with atomic **nuclei**
- Scattering permits a detailed discrimination of **isotopes**
- Non-destructive probe (**room-temperature data collection**)



Sears, V. F. (1992). *Neutron News* 3, 26-37.

Neutrons reveal H atom positions

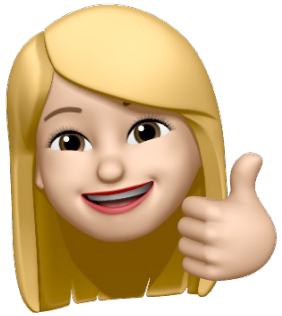
$^{1}H/^{2}H$ exchange:

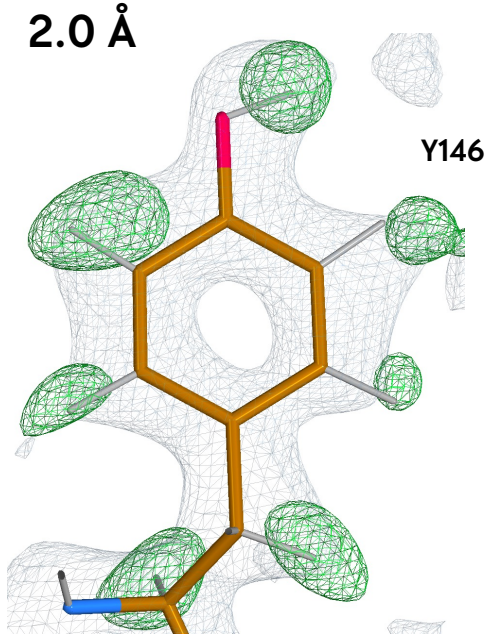
Improve the **signal-to-noise** ratio

Reduce the **background scattering**

Clearer visualization of **neutron scattering density maps**

No cancellation effects (**perdeuterated** structures)



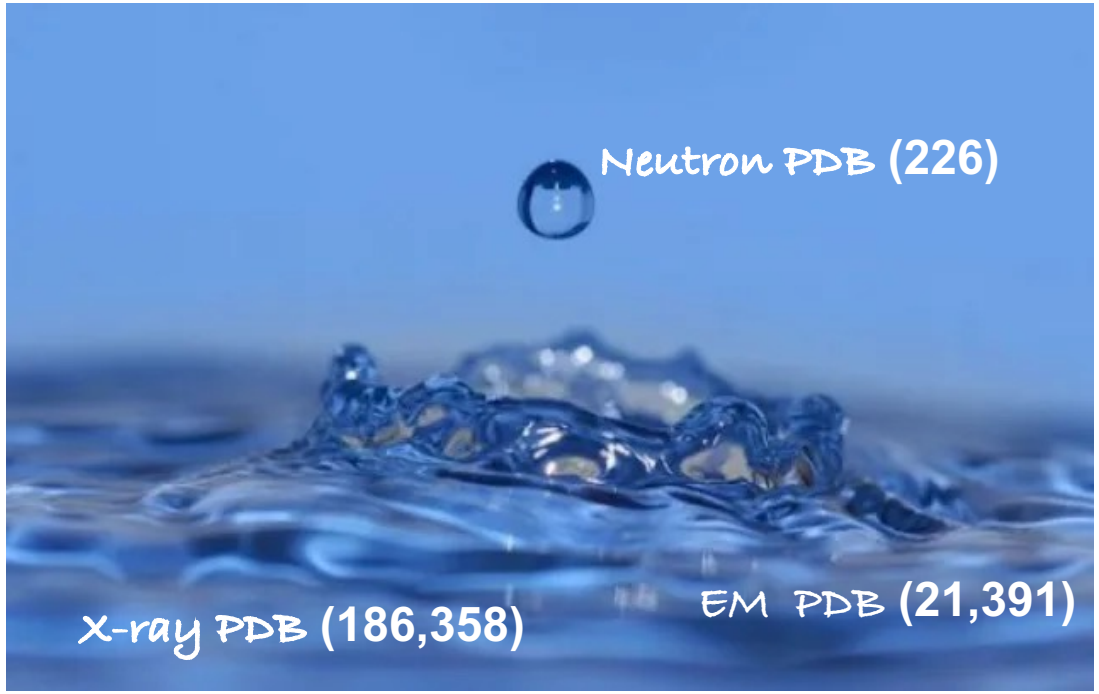


PDB code: 1CQ2
grey: $2mF_o-DF_c$ neutron scattering length density map
green: deuterium-omitted $mF_o - DF_c$ map
Shu, F., Ramakrishnan, V., & Schoenborn, B. P. (2000). *PNAS* *97*(8), 3872–3877.

Neutron macromolecular crystallography

Limitations

- **Low flux** of neutron beams
- Long data collection time (**several days or weeks**)
- Large crystals (**>0.1 mm^3**)
- Small number of neutron beamlines worldwide



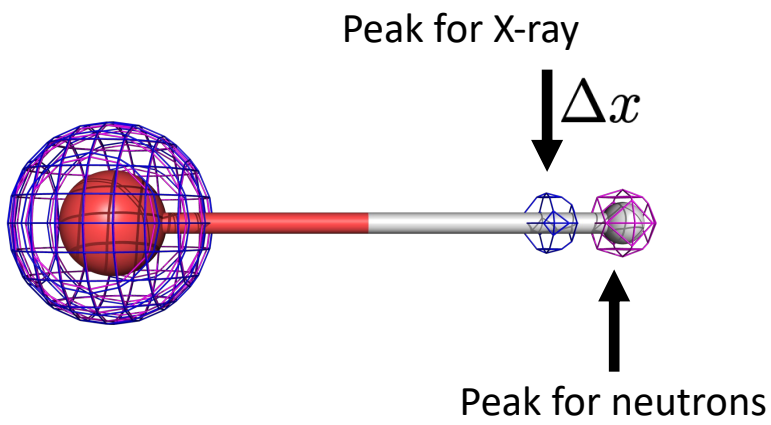
Source PDB: (https://www.rcsb.org/) Date: 13th July 2024

Hydrogen density in EM

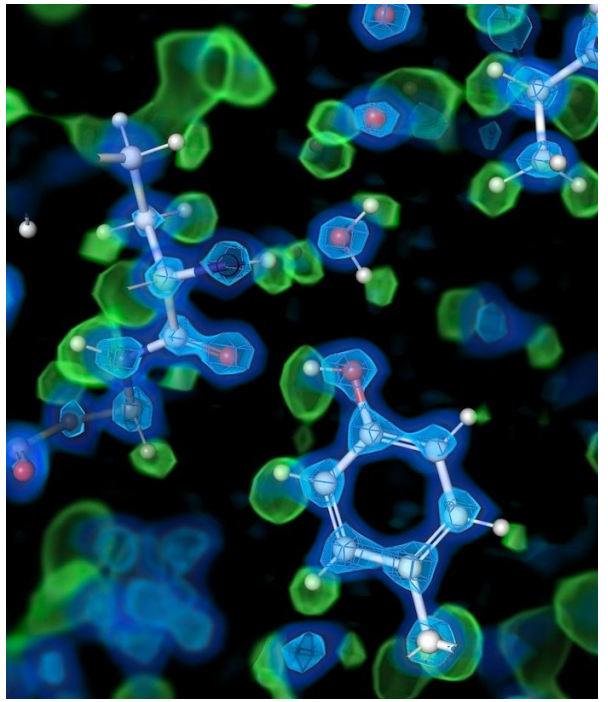
Electron scattering by H atom

$$f_e(s) = \frac{me^2}{8\pi h^2 \varepsilon_0} \frac{Ze^{-2\pi i s^T \Delta x} - f_X(s)}{s^2}$$

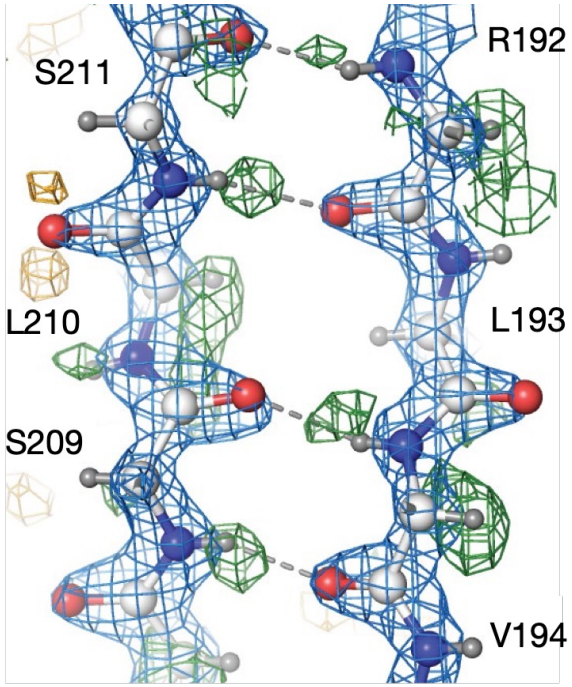
Mott-Bethe formula



green: hydrogen-omitted F_o - DF_c maps



Apoferritin at 1.22 Å



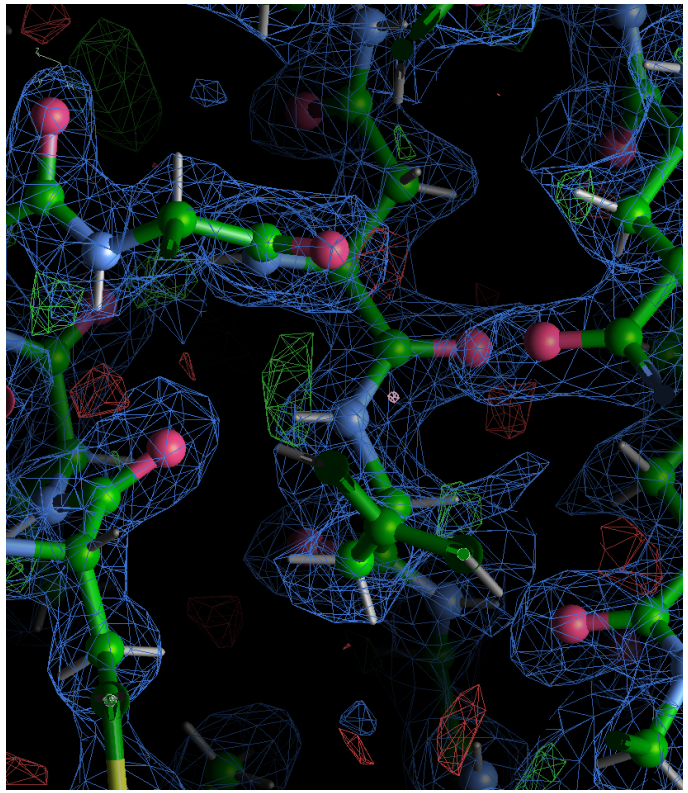
GABA$_A$R at 1.7 Å

Nakane, T., et al. (2020). *Nature* 587, 152-156.

Keitaro Yamashita

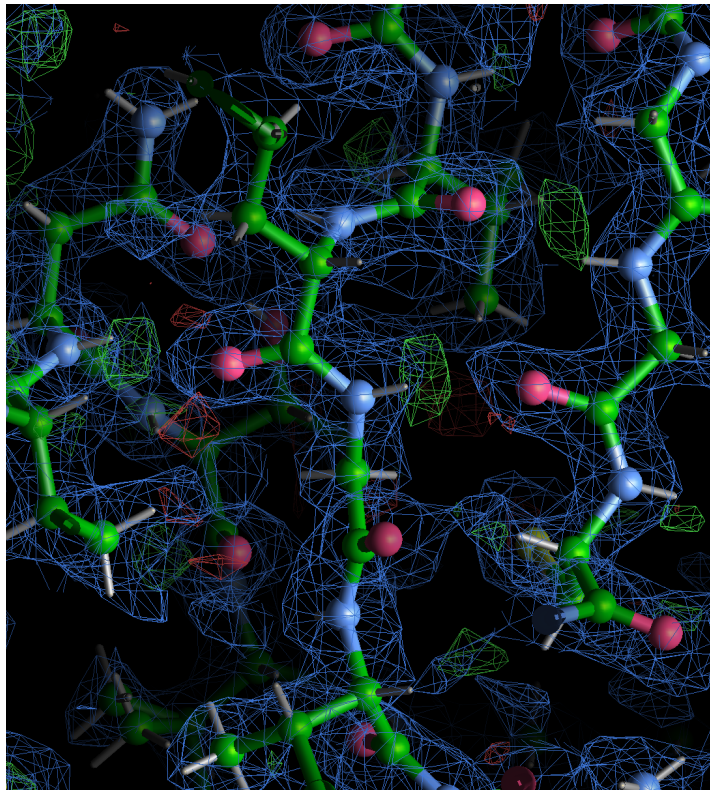
Hydrogen density in EM

green: hydrogen omit weighted F_o-DF_c maps



β-galactosidase from *E. coli* at 1.9 Å
PDB: 6cvm / EMD-7770

Bartesaghi, A.,et al. (2018). *Structure* 26, 848-856.e843.



β-galactosidase from *T. maritima* at 2.0 Å
PDB: 6s6z / EMD-10109

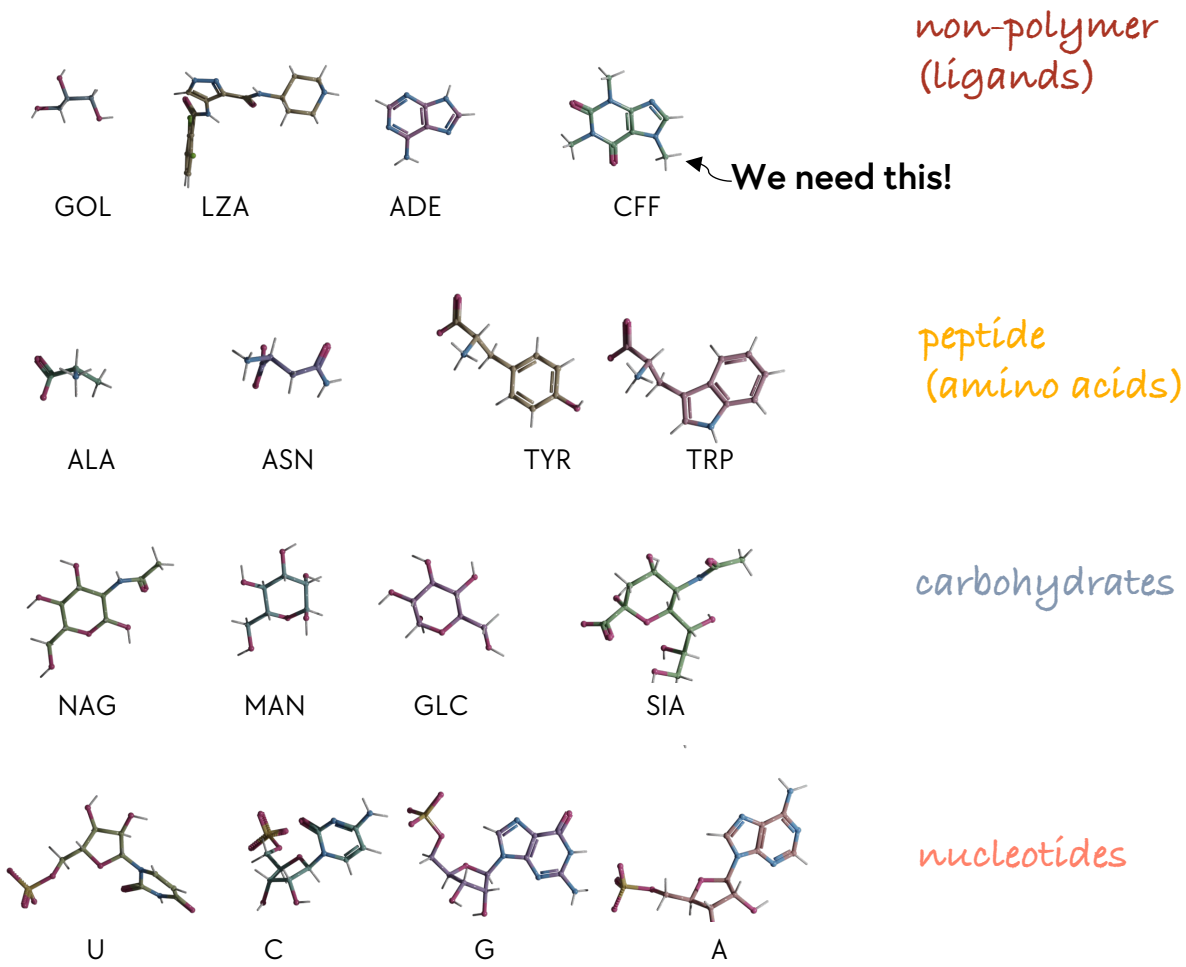
Míguez Amil, S., et al. (2020). *ACS Chem. Biol.* 15, 179-188.

New developments in REFMAC5 for neutron refinement

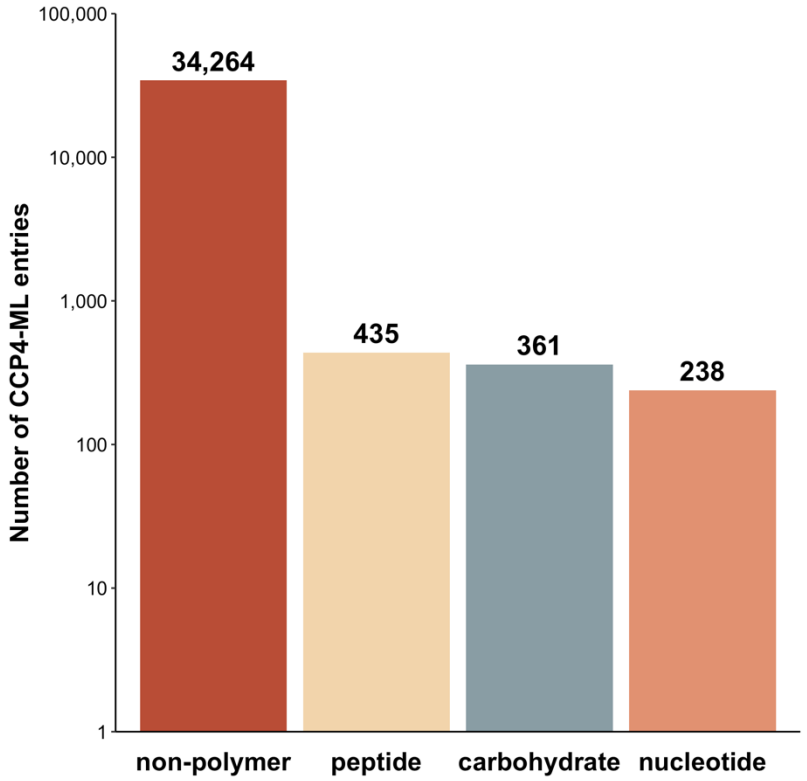
Restraints for Macromolecular Refinement

Restraints help to ensure that the model is **chemically sensible**

They are organised in the **CCP4 Monomer Library**

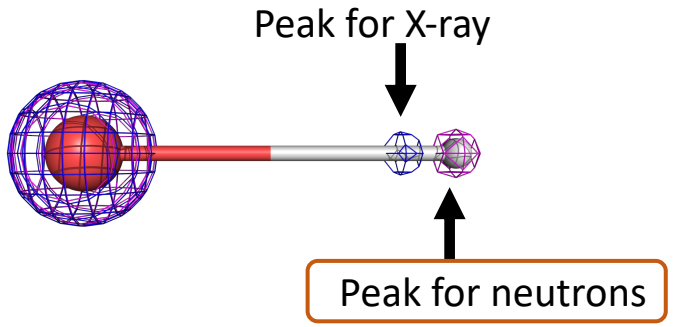


35,298 monomers
>100 of **modifications** and **links**



Vagin, A. A. et al., (2004). *Acta Cryst.* D60, 2184-2195
Nicholls, R. A. et al. (2021). *Acta Cryst.* D77, 712-726

Restraints for Neutron crystallography and EM



research papers

Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Bond lengths in organic and metal-organic compounds revisited: *X*—H bond lengths from neutron diffraction data[1]

Frank H. Allen* and Ian J. Bruno

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England

The number of structures in the Cambridge Structural Database (CSD) has increased by an order of magnitude since the preparation of two major compilations of standard bond lengths in mid-1985. It is now of interest to examine whether this huge increase in data availability has implications

Received 26 February 2010
Accepted 30 March 2010

Neutron studies in the Cambridge Structural Database (CSD)

	CSD-2009	CSD-2021	ratio 2021:2009
All	1213	2362	1.95
Organic	811	1452	1.79
With coordinates, $R \leq 0.10$	664	1220	1.84
With coordinates, $R \leq 0.10$*	561	1068	1.90
With coordinates, $R \leq 0.075$*	461	894	1.94
With coordinates, $R \leq 0.05$*	302	604	2.00
Metal-organic	402	910	2.26
Powder studies	217	707	3.26

*Structures counted have no disorder and no residual coordinate errors

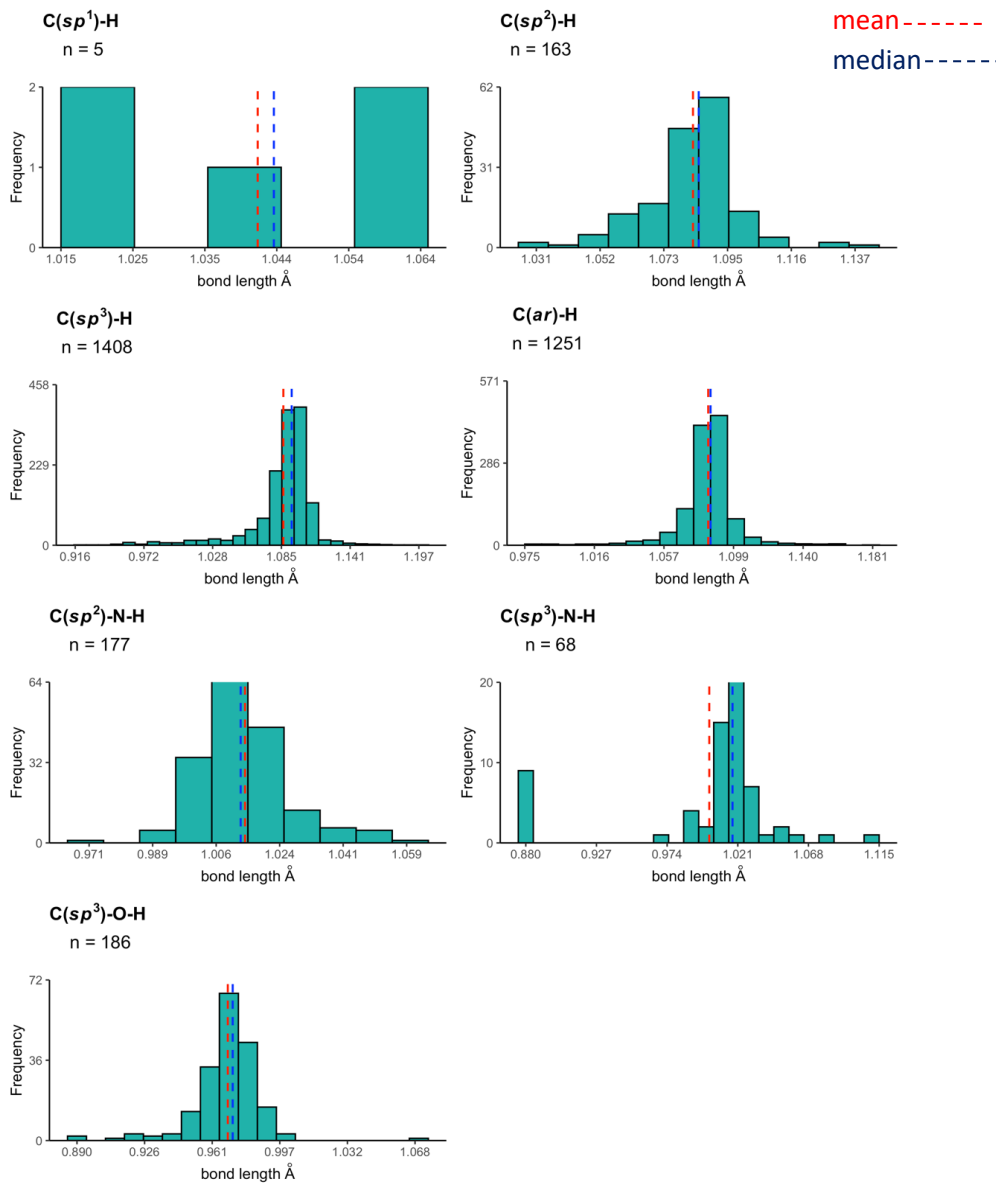
Restraints for Neutron crystallography and EM

CSD-2021 neutron dataset analysis

647 organic compounds :

- non-polymeric
- without disorder
- with R-factors ≤ 0.075

Tools: ***ConQuest, Mercury***



Restraints for Neutron crystallography and EM

Fei Long

Quantum Mechanics (QM) calculations

- **2652** small molecules from DrugBank
- DFT calculations at B3LYP/6-311++G** level of theory
- Tools: **AceDRG, GAMESS-US**

	CSD-2009				CSD-2021				QM			
	μ	σ	m	n	μ	σ	m	n	μ	σ	m	n
C(sp^1)–H	1.042	0.022	1.044	5	1.042	0.022	1.044	5	1.063	-	1.063	9
C(sp^2)–H	1.082	0.013	1.084	109	1.083	0.015	1.085	163	1.087	-	1.085	538
C(sp^3)–H	1.089	0.010	1.091	1118	1.087	0.010	1.092	1397	1.093	-	1.093	12985
C(ar)–H	1.083	0.017	1.085	721	1.084	0.018	1.085	1251	1.083	-	1.083	3906
C(sp^2)–N–H	1.013	0.010	1.012	141	1.014	0.012	1.013	177	1.010	-	1.009	1055
C(sp^3)–N–H	1.002	0.010	1.002	4	1.002	0.052	1.018	68	1.020	-	1.019	1172
C(sp^3)–O–H	0.970	0.012	0.971	169	0.969	0.018	0.972	186	0.966	-	0.964	1229
S–H	1.338	-	1.338	1	1.338	-	1.338	1	1.345	-	1.345	83

The letters μ, σ, m and n represent the mean, standard deviation, median, and number of observations

Catapano, L., et al., (2023). *Acta Crystallographica Section D* 79, 1056-1070.

Restraints for Neutron crystallography and EM

Inclusion of hydrogen nucleus distances

in the CCP4 Monomer Library

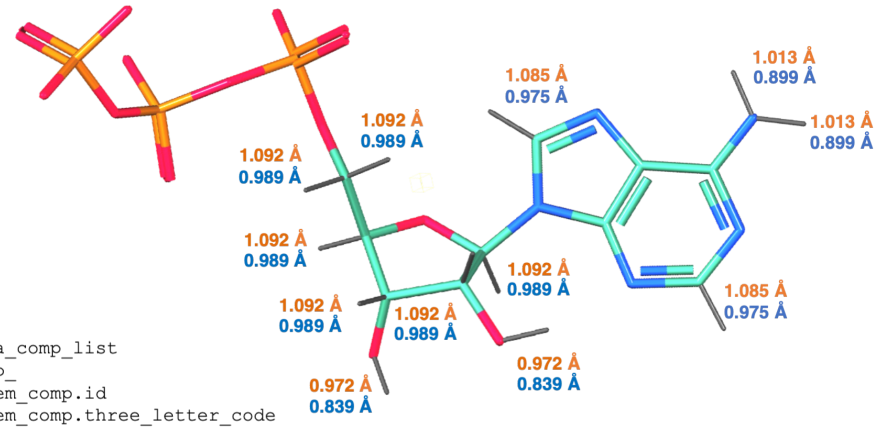
Acta Cryst D STRUCTURAL BIOLOGY
ISSN 2059-7983

AceDRG: a stereochemical description generator for ligands

Fei Long,[a] Robert A. Nicholls,[a] Paul Emsley,[a] Saulius Gražulis,[b] Andrius Merkys,[b] Antanas Vaitkus[b] and Garib N. Murshudov[a]*

[a]Structural Studies, MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, England, and [b]Institute of Biotechnology, Sauletekio al. 7, LT-10257 Vilnius, Lithuania. *Correspondence e-mail: garib@mrc-lmb.cam.ac.uk

Received 29 September 2016
Accepted 3 January 2017




```
#
data_comp_list
loop_
_chem_comp.id
_chem_comp.three_letter_code
_chem_comp.name
_chem_comp.group
_chem_comp.number_atoms_all
_chem_comp.number_atoms_nh
_chem_comp.desc_level
ATP     ATP      "ADENOSINE-5'-TRIPHOSPHATE"     NON-POLYMER     43      31      .
#
data_comp_ATP
...
loop_
_chem_comp_bond.comp_id
_chem_comp_bond.atom_id_1
_chem_comp_bond.atom_id_2
_chem_comp_bond.type
_chem_comp_bond.aromatic
_chem_comp_bond.value_dist_nucleus
_chem_comp_bond.value_dist_nucleus_esd
_chem_comp_bond.value_dist
_chem_comp_bond.value_dist_esd

ATP         "C5'"       "H5'1"     SINGLE     n          1.092       0.010       0.989       0.005
ATP         "C5'"       "H5'2"     SINGLE     n          1.092       0.010       0.989       0.005
ATP         "C4'"       "H4'"      SINGLE     n          1.092       0.010       0.989       0.005
ATP         "C3'"       "H3'"      SINGLE     n          1.092       0.010       0.989       0.005
ATP         "O3'"       "HO3'"     SINGLE     n          0.972       0.018       0.839       0.014
ATP         "C2'"       "H2'"      SINGLE     n          1.092       0.010       0.989       0.005
ATP         "O2'"       "HO2'"     SINGLE     n          0.972       0.018       0.839       0.014
ATP         "C1'"       "H1'"      SINGLE     n          1.092       0.010       0.989       0.005
ATP         C8          H8         SINGLE     n          1.085       0.015       0.975       0.010
ATP         N6          HN61       SINGLE     n          1.013       0.012       0.899       0.007
ATP         N6          HN62       SINGLE     n          1.013       0.012       0.899       0.007
ATP         C2          H2         SINGLE     n          1.085       0.015       0.975       0.010
```

Catapano, L., et al., (2023). *Acta Crystallographica Section D* 79, 1056-1070.

Treatment of hydrogen atoms

The riding hydrogen model

- The concept of riding hydrogen model ('riding H') relies on the fact that the coordinates of most (but not all) H atoms in proteins can be unambiguously expressed through the coordinates of their covalently bound non-H atoms, known as 'parent' atoms.

- For routine refinement, the riding model is very efficient, for example, for amide H atoms in the protein backbone, for those bound to C^{α} atoms, for those attached to aromatic carbon atoms. These H atoms are often referred as **'fixed'** H atoms.

- REFMAC5 generates riding H positions using the information from the CCP4-ML restraints. A H atom must be away from its parent at a distance equal to the 'ideal' distance and it must obey chirality and planarity rules.

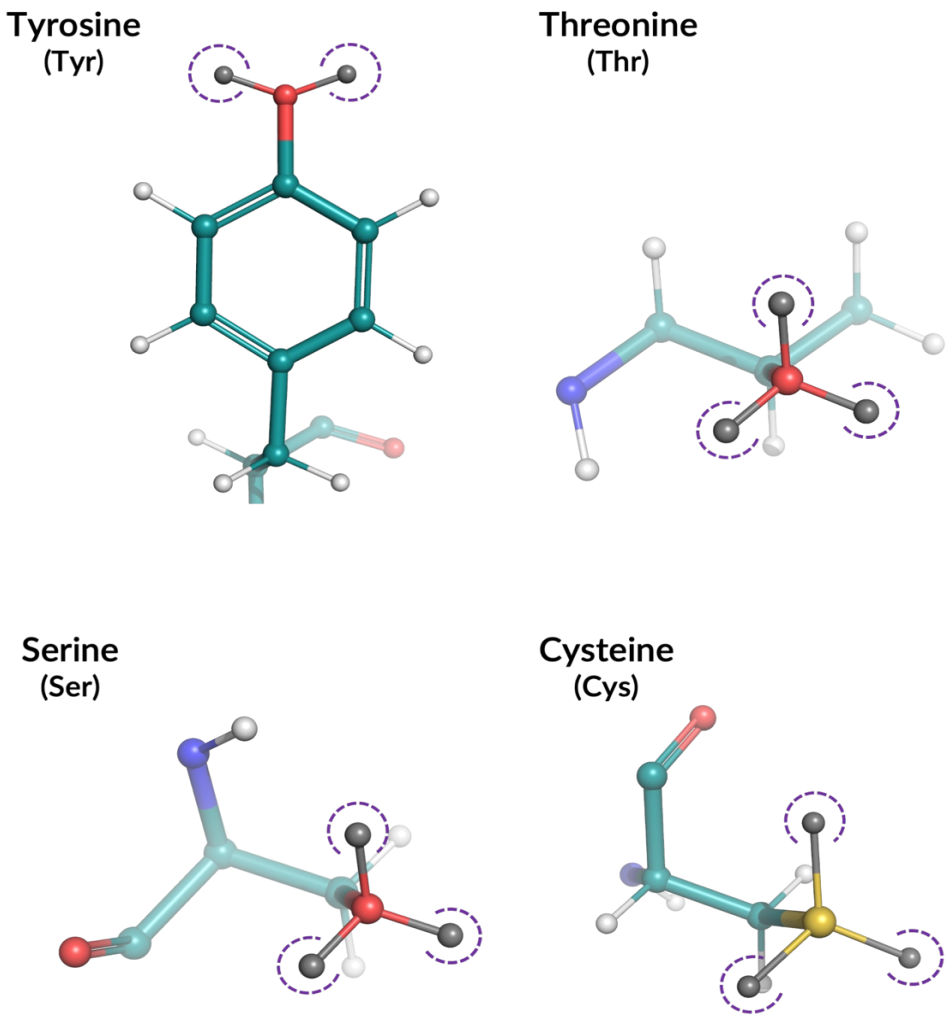
Treatment of hydrogen atoms

Placing 'rotatable' hydrogen atoms

- **'Rotatable'** hydrogen atoms: single hydrogen atoms with rotational degrees of freedom

- H atoms on ***hydroxyl (–OH)*** groups of Tyr, Thr, Ser and ***thiol (–SH)*** groups of Cys

- The accurate position of a 'rotatable' H atom can only be determined based on ***experimental data***



Treatment of hydrogen atoms

Placing 'rotatable' hydrogen atoms

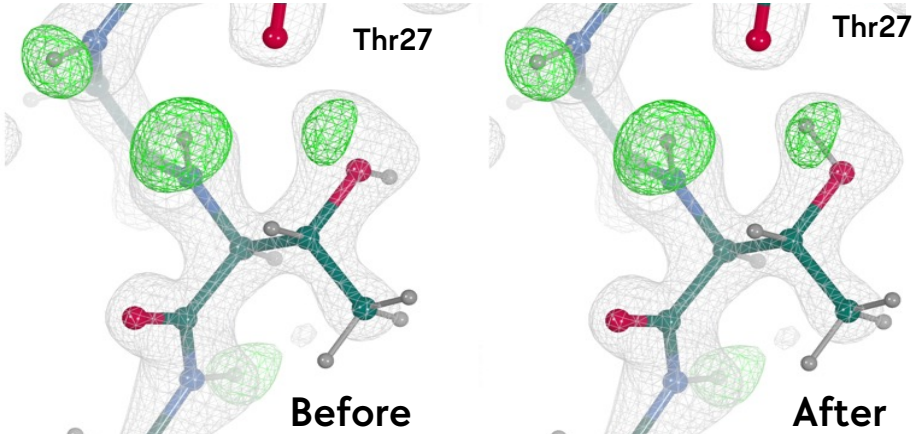
a program has been written using ***GEMMI*** that enables the automatic placement of these H atoms in ***real-space*** based on their fit to the density maps:

1. calculation of an H-omit ***mFo–DFc*** map by removing from the model the rotatable H atoms for Tyr, Ser, Thr and Cys residues.

2. calculation of favourable H positions (two for Tyr; three for Ser, Thr and Cys) based on ***parent atom positions and torsion angles.***

3. optimisation of H positions by ***best fit to the mFo–DFc*** neutron scattering length density maps.

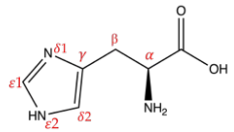
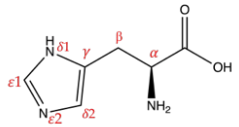
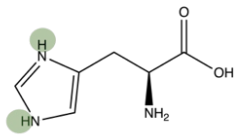
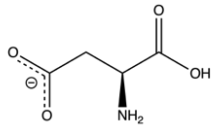
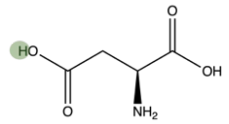
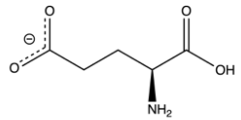
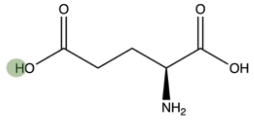
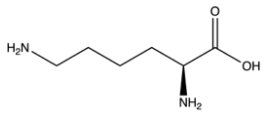
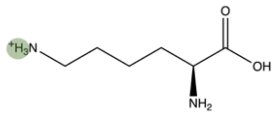
grey: $2mF_o–DF_c$ neutron scattering length density map
green: H-omitted $mF_o - DF_c$ map

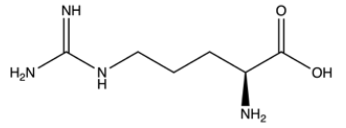
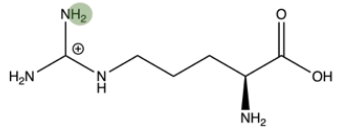
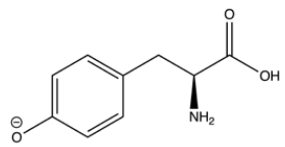
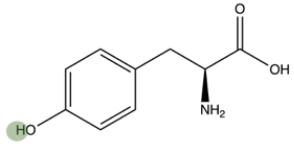
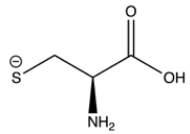
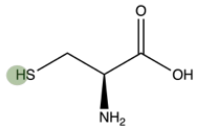
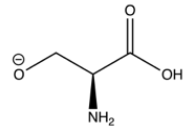
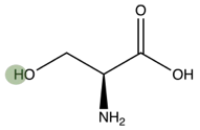
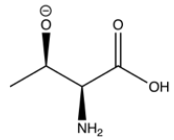
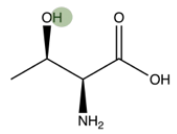


Rubredoxin 1.5 A, PDB code: 1VCX
Kurihara, K., et., al. (2004). *PNAS.* 101, 11215-11220

Treatment of hydrogen atoms

Protonation states

Amino acid p*Ka*	Deprotonated species	Protonated species
Histidine (His) 6.0	 $N^{\varepsilon 2}$ tautomer  $N^{\delta 1}$ tautomer	
Aspartate (Asp) 3.9		
Glutamate (Glu) 4.1		
Lysine (Lys) 10.5		

Arginine (Arg) 12.5		
Tyrosine (Tyr) 10.5		
Cysteine (Cys) 8.3		
Serine (Ser) 13		
Threonine (Thr) 13		

Treatment of hydrogen atoms

Protonation states

- ***Asp and Glu*** side chains are predominantly deprotonated at physiological conditions, thus the CCP4-ML provides their component dictionary entries in the deprotonated forms.

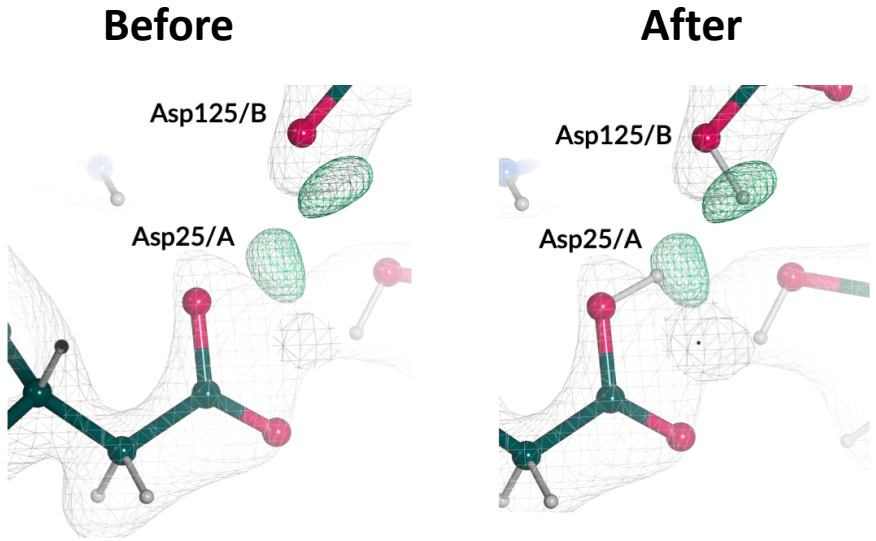
- the imidazole of ***His*** is typically found in a singly protonated state at physiological pH, with the residue that can exist as either ***Nε2*** protonated tautomer or ***Nδ1*** protonated tautomer. The CCP4-ML provides the His component dictionary entry as doubly protonated.

- Within CCP4, the determination of protonation (or deprotonation) forms is achieved by applying ***modifications*** to the current monomer to ensure proper geometry is maintained during refinement

Treatment of hydrogen atoms

Protonation states

- **Modifications** (mmCIF) describing protonation of Asp, Glu and His are generated using **AceDRG**



protonation for Asp25 and Asp125 in HIV-1 Protease (PDB entry 6PTP)

Kumar, M., et., al. (2020). *ACS Omega* 5, 11605-11617.

```
data_mod_list

loop_
_chem_mod.id
_chem_mod.name
_chem_mod.comp_id
_chem_mod.group_id
ASP_prot       "Protonation_of_ASP"      ASP         peptide

data_mod_ASP_prot

loop_
_chem_mod_atom.mod_id
_chem_mod_atom.function
_chem_mod_atom.atom_id
_chem_mod_atom.new_atom_id
_chem_mod_atom.new_type_symbol
_chem_mod_atom.new_type_energy
_chem_mod_atom.new_charge
ASP_prot          change       OD2   OD2         O         OH1        0
ASP_prot          change       OD1   OD1         O          O         0
ASP_prot            add        .     HD2         H          H         0

loop_
_chem_mod_bond.mod_id
_chem_mod_bond.function
_chem_mod_bond.atom_id_1
_chem_mod_bond.atom_id_2
_chem_mod_bond.new_type
_chem_mod_bond.new_value_dist
_chem_mod_bond.new_value_dist_esd
_chem_mod_bond.new_value_dist_nucleus
_chem_mod_bond.new_value_dist_nucleus_esd
ASP_prot       add          HD2   OD2         single  0.876  0.0200     0.966  0.0059
ASP_prot       change       CG    OD2         single  1.308  0.0191     1.308  0.0191
ASP_prot       change       CG    OD1         double  1.217  0.0198     1.217  0.0198

loop_
_chem_mod_angle.mod_id
_chem_mod_angle.function
_chem_mod_angle.atom_id_1
_chem_mod_angle.atom_id_2
_chem_mod_angle.atom_id_3
_chem_mod_angle.new_value_angle
_chem_mod_angle.new_value_angle_esd
ASP_prot            add        HD2 OD2  CG 110.209    3.00
ASP_prot            change     CB  CG   OD2 113.234    2.02
ASP_prot            change     CB  CG   OD1 123.540    1.50
ASP_prot            change     OD1 CG   OD2 123.226    1.50

loop_
_chem_mod_tor.mod_id
_chem_mod_tor.function
_chem_mod_tor.id
_chem_mod_tor.atom_id_1
_chem_mod_tor.atom_id_2
_chem_mod_tor.atom_id_3
_chem_mod_tor.atom_id_4
_chem_mod_tor.new_value_angle
_chem_mod_tor.new_value_angle_esd
_chem_mod_tor.new_period
ASP_prot    add sp2_sp2     CB  CG OD2 HD2    180  5   2
```

Neutron refinement with REFMAC5/refmacat

Deuterium fraction parameterisation

$$f_i(s) = (1 - m_i)\, b_H + m_i b_D$$

fi(s) is the total contribution of protium ^{1}H and ^{2}H isotopes to the scattering factor of the ***i***th H atom

s is the Fourier space vector

m_i is the deuterium fraction parameter that is an adjustable parameter

b_H and b_D are the neutron scattering lengths of ^{1}H (-0.374×10^{-12} cm) and ^{2}H (0.667×10^{-12} cm) isotopes, respectively

In the model, use ***_atom_site.ccp4_deuterium_fraction*** instead of having ^{1}H and ^{2}H atoms separately

Catapano, L., et al., (2023). *Acta Crystallographica Section D* 79, 1056-1070.

Deuterium fraction parameter

```
ATOM   352 N  N    . SER A 1 24  12.597 -4.159  14.144 1.00 6.79   24  SER A N    1
ATOM   353 C  CA   . SER A 1 24  13.849 -4.499  14.798 1.00 7.22   24  SER A CA   1
ATOM   354 C  C    . SER A 1 24  15.037 -4.410  13.835 1.00 6.56   24  SER A C    1
ATOM   355 O  O    . SER A 1 24  14.886 -4.586  12.629 1.00 6.13   24  SER A O    1
ATOM   356 C  CB   . SER A 1 24  13.775 -5.904  15.419 1.00 7.52   24  SER A CB   1
ATOM   357 O  OG   . SER A 1 24  13.373 -6.885  14.483 1.00 9.28   24  SER A OG   1
ATOM   358 H  H    A SER A 1 24  12.000 -4.855  13.866 0.14 6.79   24  SER A H    1
ATOM   359 D  D    B SER A 1 24  12.000 -4.855  13.866 0.86 6.79   24  SER A D    1
ATOM   360 H  HA   . SER A 1 24  14.000 -3.798  15.603 1.00 6.88   24  SER A HA   1
ATOM   361 H  HB2  . SER A 1 24  14.756 -6.188  15.792 1.00 6.52   24  SER A HB2  1
ATOM   362 H  HB3  . SER A 1 24  13.082 -5.908  16.242 1.00 9.87   24  SER A HB3  1
ATOM   363 D  DG   . SER A 1 24  14.122 -7.106  13.939 1.00 12.20  24  SER A DG   1
```

_atom_site.ccp4_deuterium_fraction

```
ATOM   331 N  N    . SER Ap 1 24  12.59  -4.173  14.126 1 7.939  24  A 1 0
ATOM   332 C  CA   . SER Ap 1 24  13.842 -4.479  14.801 1 8.287  24  A 1 0
ATOM   333 C  C    . SER Ap 1 24  15.019 -4.371  13.833 1 7.903  24  A 1 0
ATOM   334 O  O    . SER Ap 1 24  14.865 -4.592  12.642 1 7.724  24  A 1 0
ATOM   335 C  CB   . SER Ap 1 24  13.805 -5.842  15.429 1 8.732  24  A 1 0
ATOM   336 O  OG   . SER Ap 1 24  13.394 -6.837  14.505 1 9.213  24  A 1 0
ATOM   337 H  H    . SER Ap 1 24  11.996 -4.949  13.784 1 8.035  24  A 1 0.918506
ATOM   338 H  HA   . SER Ap 1 24  13.98  -3.744  15.591 1 8.239  24  A 1 0
ATOM   339 H  HB2  . SER Ap 1 24  14.796 -6.089  15.808 1 8.741  24  A 1 0
ATOM   340 H  HB3  . SER Ap 1 24  13.117 -5.828  16.273 1 8.741  24  A 1 0
ATOM   341 H  HG   . SER Ap 1 24  14.157 -7.146  14.029 1 9.924  24  A 1 0.886209
```

Catapano, L., et al., (2023). *Acta Crystallographica Section D* **79**, 1056-1070.

Deuterium fraction parameter

(a) **Traditional representation for partially $^{1}H/^{2}H$-exchanged structures**

	Atom ID		Alt ID	Comp Info	Coordinates			Occupancy	ADP
ATOM	609	H	A	ILE A 1 40	10.029	9.509	8.192	0.07	6.07
ATOM	610	D	B	ILE A 1 40	10.029	9.509	8.192	0.93	6.07
ATOM	611	H	.	ILE A 1 40	9.748	11.778	9.905	1.00	6.87
ATOM	612	H	.	ILE A 1 40	11.104	11.549	7.217	1.00	10.27

(b) **Deuterium fraction representation for partially $^{1}H/^{2}H$-exchanged structures**

	Atom ID		Alt ID	Comp Info	Coordinates			Occupancy	ADP	Deuterium fraction
ATOM	575	H	.	ILE A 1 40	10.038	9.489	8.110	1.00	7.32	0.92
ATOM	576	H	.	ILE A 1 40	9.741	11.805	9.911	1.00	7.80	0.00
ATOM	577	H	.	ILE A 1 40	11.11	11.56	7.227	1.00	10.091	0.00

(c) **Traditional representation for perdeuterated structures**

	Atom ID		Alt ID	Comp Info	Coordinates			Occupancy	ADP
ATOM	108	D	D	. ILE A 1 7	10.930	10.830	−0.577	1.00	11.12
ATOM	109	D	DA	. ILE A 1 7	10.298	13.232	−1.478	1.00	12.41
ATOM	110	D	DB	. ILE A 1 7	9.246	11.966	0.834	1.00	12.79

(d) **Deuterium fraction representation for perdeuterated structures**

	Atom ID		Alt ID	Comp Info	Coordinates			Occupancy	ADP	Deuterium fraction
ATOM	106	H	H	. ILE A 1 7	10.934	10.707	−0.461	1.00	14.621	0.90
ATOM	107	H	HA	. ILE A 1 7	10.275	13.326	−1.614	1.00	14.628	0.87
ATOM	108	H	HB	. ILE A 1 7	9.259	11.883	0.825	1.00	15.766	1.00

Catapano, L., et al., (2023). *Acta Crystallographica Section D* 79, 1056-1070.

Neutron refinement with REFMAC5/refmacat

X-ray reference structure restraints

- **neutron** models are often **poor** in geometric quality
- the geometry of all **non-H atoms** is more accurately determined by **X-ray**.
- **X-ray** models are almost **always available** before a corresponding neutron structure.
- **Interatomic distances** between non-H atoms are **restrained** to the reference X-ray structure

$$\rho(x,\sigma,\alpha) = \begin{cases} \frac{1}{2}\left(\frac{x}{\sigma}\right)^{\alpha}, & if\ \alpha = 2 \\ \log\left(\frac{1}{2}\left(\frac{x}{\sigma}\right)^{2} + 1\right), & if\ \alpha = 0 \\ 1 - e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^{2}} & if\ \alpha = -\infty \\ \frac{|\alpha-2|}{\alpha}\left(\left(\frac{\left(\frac{x}{\sigma}\right)^{2}}{|\alpha-2|} + 1\right)^{\frac{\alpha}{2}} - 1\right) & otherwise \end{cases}$$

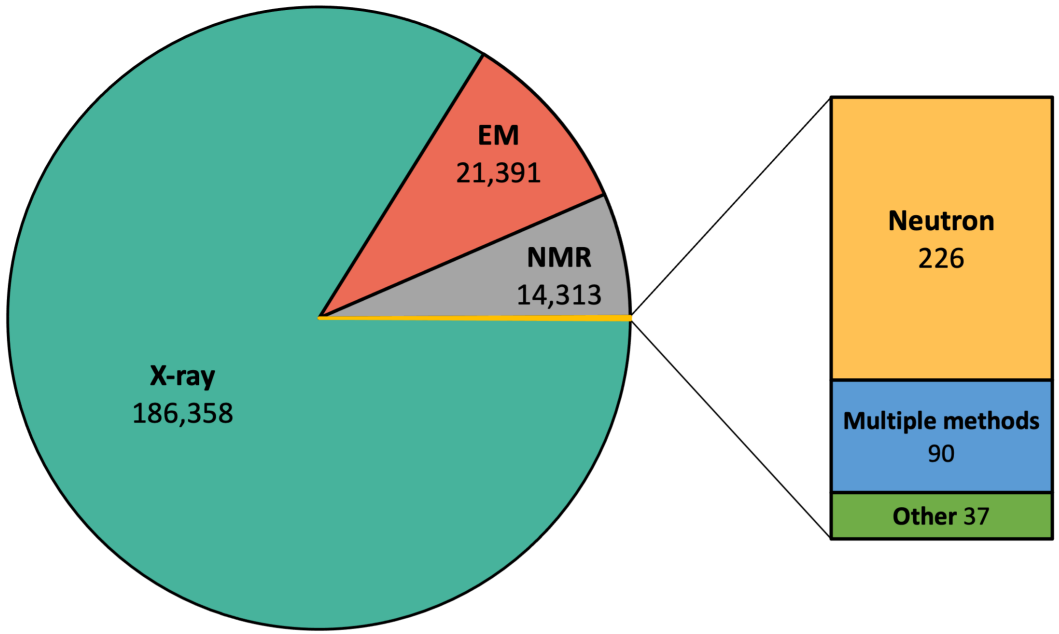
$\boldsymbol{x = d_{model} - d_{ref}}$, is the difference between model and reference structure interatomic distances

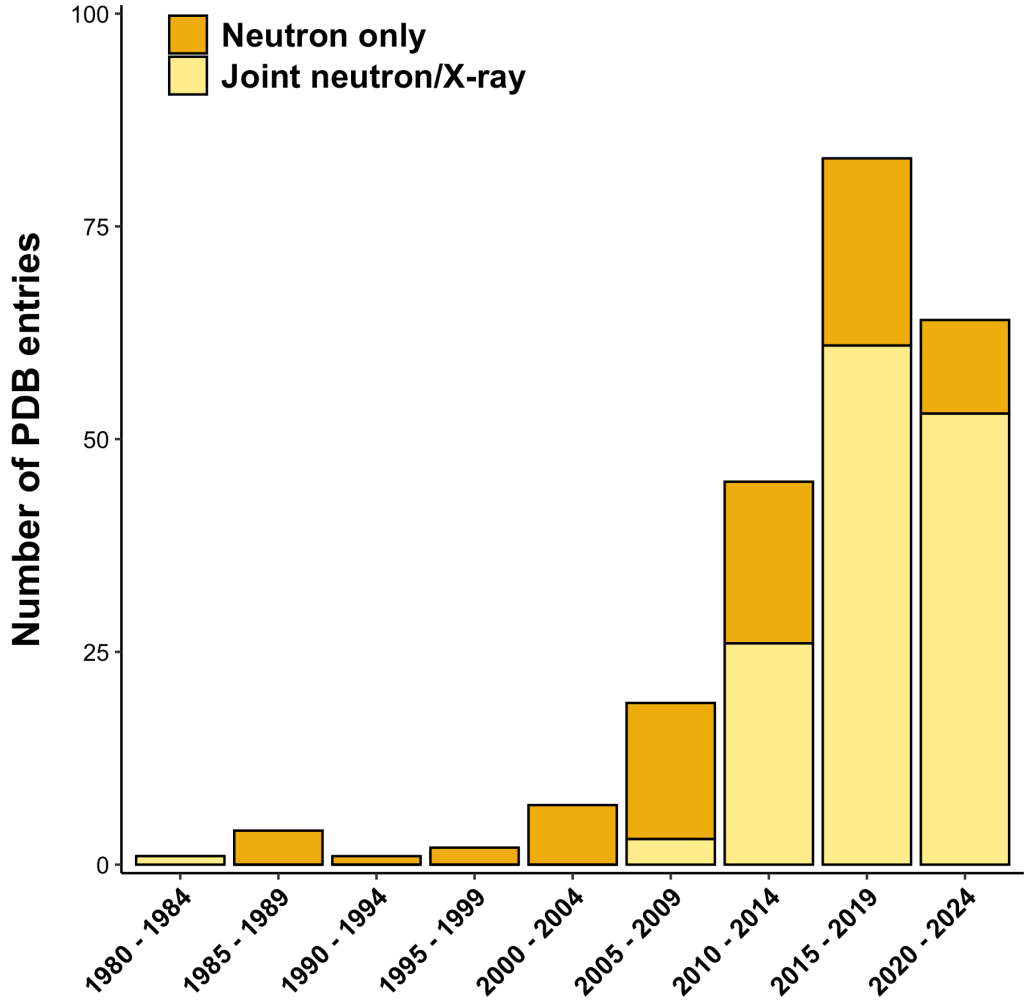
small or negative values of $\boldsymbol{\alpha}$ increase robustness

J. T. Barron, *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4326-4334

Re-refinement of neutron PDB entries

Neutron entries in the PDB

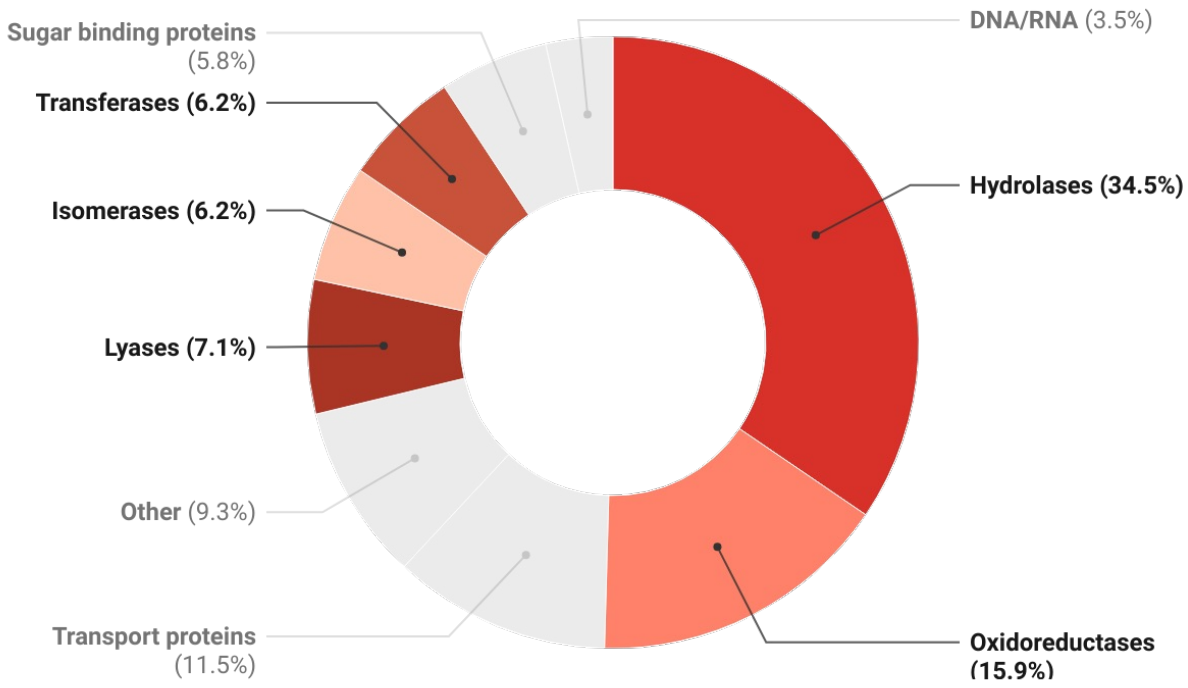




Source PDB: (https://www.rcsb.org/) Date: 13th July 2024

Neutron entries in the PDB

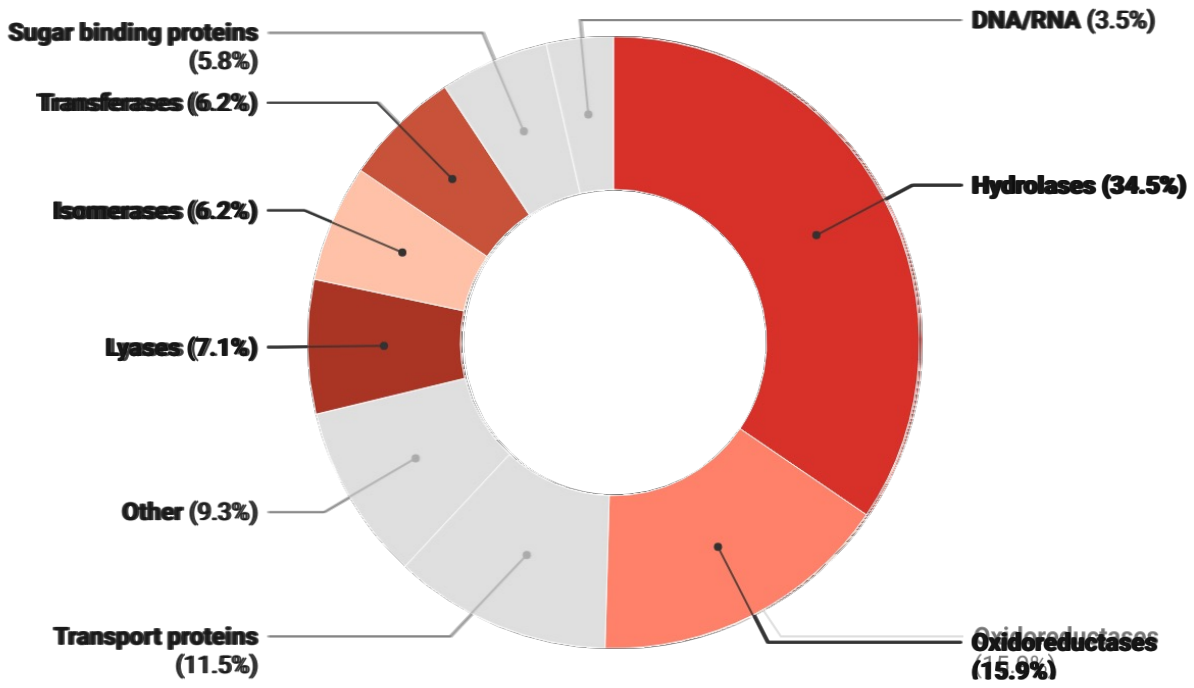
~70% of all neutron PDB entries are ***enzymes***



Source PDB: (https://www.rcsb.org/) Date: 13th July 2024

Neutron entries in the PDB

~70% of all neutron PDB entries are ***enzymes***



Hydrolases (EC 3): e.g., trypsin, HIV-1 protease, β-lactamase

Oxidoreductases (EC 1): e.g., heme peroxidases

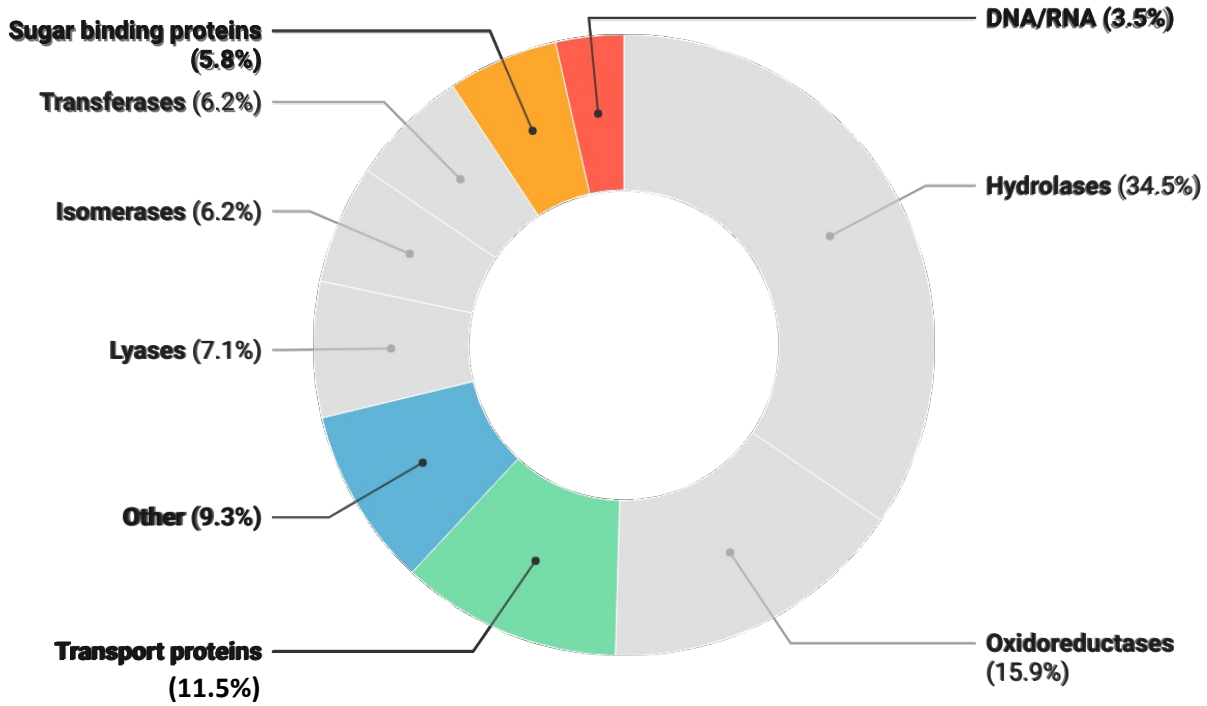
Lyases (EC 4): e.g., carbonic anhydrase II

Isomerases (EC 5): e.g., xylose isomerases

Transferases (EC 2): e.g., aminoglycoside acetyltransferase

Source PDB: (https://www.rcsb.org/) Date: 13th July 2024

Neutron entries in the PDB



Transport proteins: e.g., myoglobin, hemoglobin

Sugar binding proteins: e.g., concanavalin A

DNA/RNA: e.g., B and Z-forms of DNA, Sarcin-Ricin loop RNA

Others: e.g., signaling proteins, fluorescent proteins, etc.

Source PDB: (https://www.rcsb.org/) Date: 13th July 2024

Re-refinement of neutron PDB entries – deuterium fraction

97 neutron models (out of 213 at the time of testing)

55 neutron-only, **42 joint X-ray/neutron**

Strategy:

- Refinement against neutron data only
- Refinement of ***deuterium fraction parameter*** for only polar H atoms in case of $^{1}H/^{2}H$-exchanged samples or all H atoms in case of perdeuterated samples

Re-refinement of selected neutron-only models from the PDB

PDB information			*REFMAC*5 refinement statistics			
PDB code	Published resolution (Å)	Published *R* (work/free) (%)	Initial *R* (work/free) (%)	Final *R* (work/free) (%)	Data completeness (%)	No. of reflections
1C57	15.79-2.40	27.0/30.1	29.7/33.0	19.9/25.4	87.38	8129
1CQ2	6.00-2.00	16.0/25.0	18.6/25.7	14.9/24.7	91.07	7528
1WQ2	20.00-2.40	22.9/28.9	28.6/32.1	21.5/28.7	92.29	6232
1XQN	32.82-2.50	26.6/32.0	30.0/31.5	23.0/29.4	74.99	6088
2EFA	80.00-2.70	21.6/29.1	24.3/27.6	22.3/28.0	95.66	2154
2GVE	10.00-2.20	26.8/31.9	26.3/30.3	23.4/29.8	93.73	22133
3RZ6	21.77-1.75	20.8/23.8	25.2/25.6	21.3/25.1	79.55	4215
4AR3	15.71-1.05	19.9/23.7	19.7/23.0	18.8/22.4	88.73	21580
4AR4	27.46-1.38	18.6/22.6	17.9/22.1	15.5/21.0	91.56	9958
4C3Q	10.00-2.20	19.2/24.0	21.3/25.1	17.3/24.2	88.19	13256
4FC1	10.00-1.10	21.1/25.3	22.8/25.5	21.5/24.1	76.85	10549
4Q49	20.00-1.80	18.7/21.5	18.6/20.9	15.5/20.4	89.45	18803
5A90	38.91-1.70	19.2/22.7	21.1/24.1	18.9/23.2	88.61	28772
5GX9	33.45-1.49	15.8/20.0	17.3/21.0	17.0/20.9	93.50	14375
7KKW	14.65-2.30	24.9/30.2	27.9/31.8	22.5/31.6	98.38	20628

Catapano, L., et al., (2023*). Acta Crystallographica Section D* 79, 1056-1070.

Re-refinement of neutron PDB entries – deuterium fraction

97 neutron models (out of 213 at the time of testing)

55 neutron-only, **42 joint X-ray/neutron**

Strategy:

- Refinement against neutron data only
- Refinement of ***deuterium fraction parameter*** for only polar H atoms in case of $^{1}H/^{2}H$-exchanged samples or all H atoms in case of perdeuterated samples

Re-refinement of selected joint X-ray/neutron models from the PDB

PDB information			*REFMAC*5 refinement statistics			
PDB code	Published resolution (Å)	Published *R* (work/free) (%)	Initial *R* (work/free) (%)	Final *R* (work/free) (%)	Data completeness (%)	No. of reflections
3R98	53.54-2.40	20.7/25.1	20.8/26.1	18.4/25.7	74.99	11610
3R99	53.54-2.40	20.7/25.0	20.8/26.1	18.5/25.6	74.99	11610
3VXF	44.87-2.75	18.3/23.4	17.7/23.6	17.0/24.3	73.72	7670
3X2O	18.83-1.50	22.8/25.1	23.6/25.6	21.1/24.0	93.49	23109
4CVI	39.84-2.41	17.6/24.3	17.6/23.8	14.8/22.9	73.74	11426
4NY6	26.77-1.85	17.6/22.4	19.6/22.3	15.2/21.1	89.97	4654
4QDW	20.00-1.80	16.6/17.9	19.2/18.2	16.2/17.5	72.86	31459
4XPV	20.00-2.00	26.4/30.4	27.3/30.0	24.8/29.9	80.58	11251
5CG5	53.57-2.40	18.6/22.9	19.3/22.9	17.5/22.8	98.45	17458
5MOO	22.09-1.43	17.0/18.5	19.9/20.9	18.0/20.3	93.83	36397
5MOQ	25.43-1.50	15.0/16.7	17.1/18.5	15.3/18.5	89.39	30123
5MOR	19.67-1.49	19.6/20.7	22.8/23.8	18.9/22.0	93.38	32293
5MOS	22.15-1.50	16.6/18.0	18.7/19.5	16.9/19.8	88.77	30027
6EXY	28.20-1.70	15.0/18.7	18.6/19.5	15.8/19.5	95.37	14401
6U0F	13.79-2.00	21.9/24.1	25.2/26.5	20.6/25.9	86.6	12087

Catapano, L., et al., (2023). *Acta Crystallographica Section D* 79, 1056-1070.

Re-refinement of neutron PDB entries

Reference structure restraints

Strategy:

- for the ***'neutron-only'*** entries the corresponding X-ray reference structures were chosen based on their high structural similarity to the neutron refined structures using ***'Find Similar Assemblies'*** option from the PDB.

- For the ***joint X-ray/neutron structures*** these models were subjected to refinement against their corresponding X-ray data using REFMAC5 (10 refinement cycles). The output model obtained from this refinement process was subsequently employed as a reference model for the generation of external restraints using ***ProSMART.***

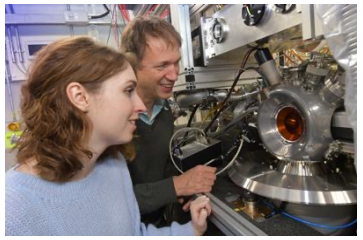
- Refinement of ***deuterium fraction parameter***

improve both the Rwork and Rfree values for low resolution neutron structures by ~2–3 percentage points in certain cases

PDB code		Published resolution (low-high) (Å)		Published *R* values (work/free) (%)		*REFMAC5* *R* values (work/free) (%)
Neutron	X-ray	Neutron	X-ray	Neutron	X-ray	Neutron refinement with external restraints
1C57	1DQ6	15.79-2.40	8.00-1.90	27.0/30.1	18.6/NA	20.2/24.9
2EFA	1B2A	80.00-2.50	55.00-1.70	21.6/29.1	18.8/23.0	21.9/25.3
2YZ4	1DQ6	33.64-2.20	8.00-1.90	27.9/31.2	18.6/NA	21.7/27.3
2ZPP	1B2G	20.00-2.50	10.00-1.80	22.1/26.0	20.0/22.6	20.8/25.4
3R98	3R98	53.54-2.40	43.89-2.10	20.7/25.1	16.6/20.3	17.8/25.2
3R99	3R99	53.54-2.40	43.89-2.10	20.7/25.0	16.6/20.3	17.9/25.3
3VXF	3VXF	44.87-2.75	29.22-1.60	18.3/23.4	16.1/18.4	16.8/23.5
3X2O	3X2O	18.83-1.50	28.23-1.00	22.8/25.1	13.5/15.3	21.2/23.9
3X2P	3X2P	19.68-1.52	37.75-0.99	21.8/26.0	13.4/14.2	22.0/25.7
4CVI	4CVI	39.84-2.41	14.80-2.10	17.6/24.3	13.4/17.6	14.0/22.8
4DVO	4DVO	20.00-2.00	29.90-1.55	19.0/21.4	NA/NA	18.4/22.2
4GPG	4GPG	37.61-1.98	50.00-1.89	19.5/25.9	14.6/20.3	20.9/25.7
4PVN	4PVN	52.28-2.30	43.13-1.95	20.9/26.2	15.6/18.5	18.7/26.8
5CG6	5CG6	22.12-2.40	44.31-1.70	26.0/28.7	19.6/21.1	24.0/27.8
5XPE	5XPE	17.02-2.09	46.50-1.64	22.5/27.8	15.5/18.5	20.5/29.7
5ZN0	5ZN0	33.76-1.90	36.01-1.10	18.8/24.7	18.6/21.2	16.7/25.5
6BQ8	6BQ8	40.00-2.20	10.00-2.00	23.2/28.8	19.9/24.5	19.7/29.2
6EXY	6EXY	28.20-1.70	31.93-1.10	15.0/18.7	12.3/13.7	16.0/19.8
6U0E	6U0E	14.51-1.89	29.01-2.10	21.7/25.4	18.4/23.5	21.4/26.7
7D6G	7D6G	26.33-2.10	40.00-1.65	17.7/21.9	15.7/18.6	14.6/22.1
7TX4	7TX4	12.75-2.35	61.05-1.90	17.7/25.9	16.6/22.4	15.2/26.0

Catapano, L., et al., (2023). *Acta Crystallographica Section D* 79, 1056-1070.

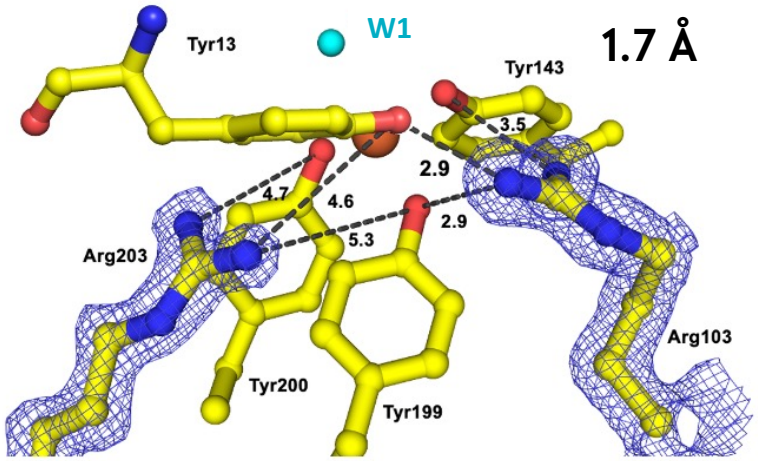
Neutron studies of iron binding protein FutA



Rachel Bolton & Ivo Tews

Cyanobacterial **iron binding** protein

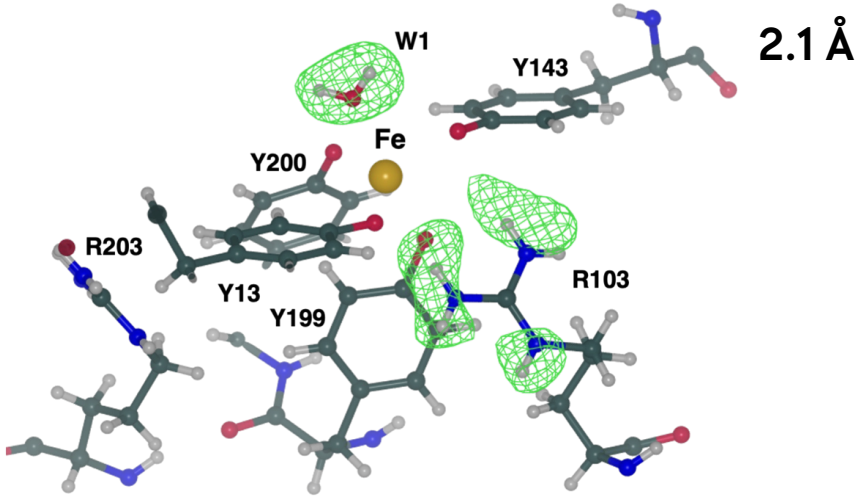
RT X-ray



X-ray induce **photoreduction** of **ferric (Fe^{3+})** FutA

coordination sphere					second shell			
Tyr13	Tyr143	Tyr199	Tyr200	W1	Arg103	Arg203	Fe	
-1	-1	-1	-1	0	+1	+1	+2	**Ferrous**

Neutron structure reveals the **protonation states** of iron coordinating residues

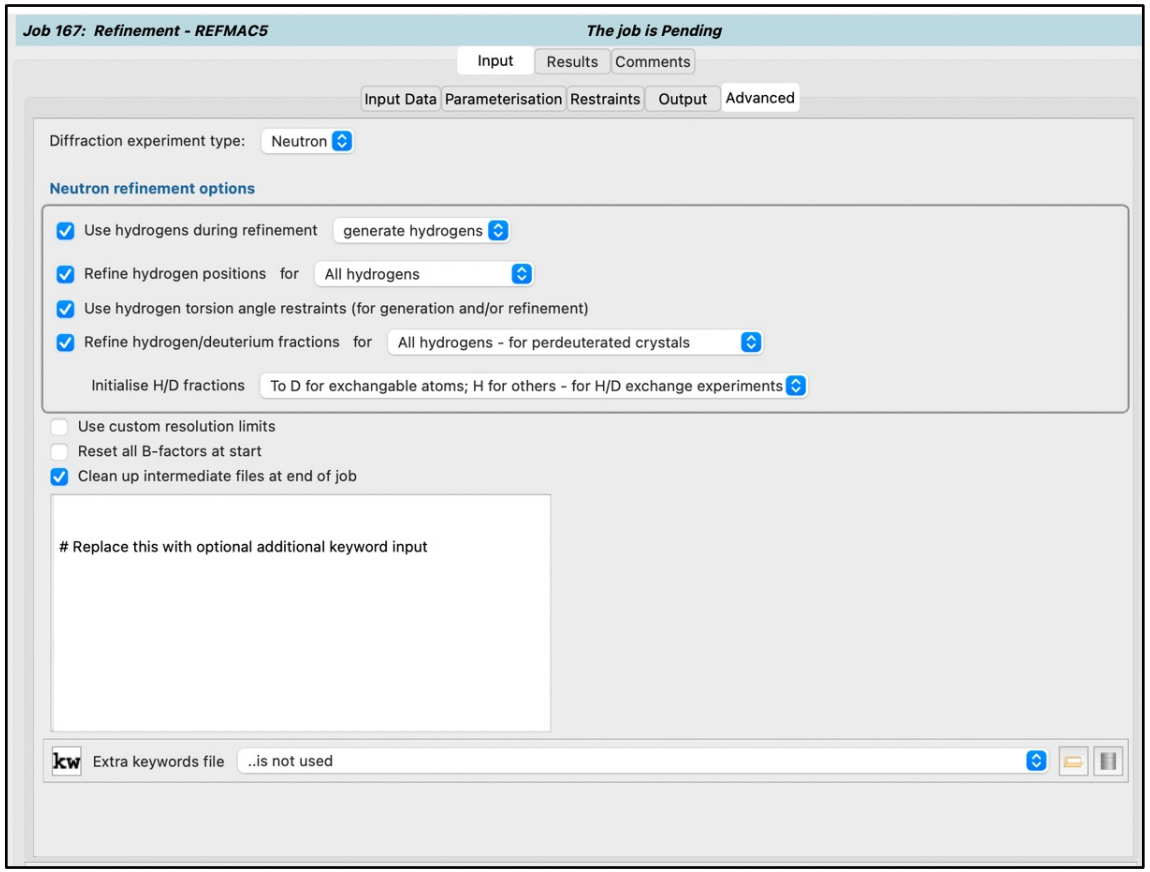


coordination sphere					second shell			
Tyr13	Tyr143	Tyr199	Tyr200	W1	Arg103	Arg203	Fe	
-1	-1	-1	-1	0	+1	too far	+3	**Ferric**

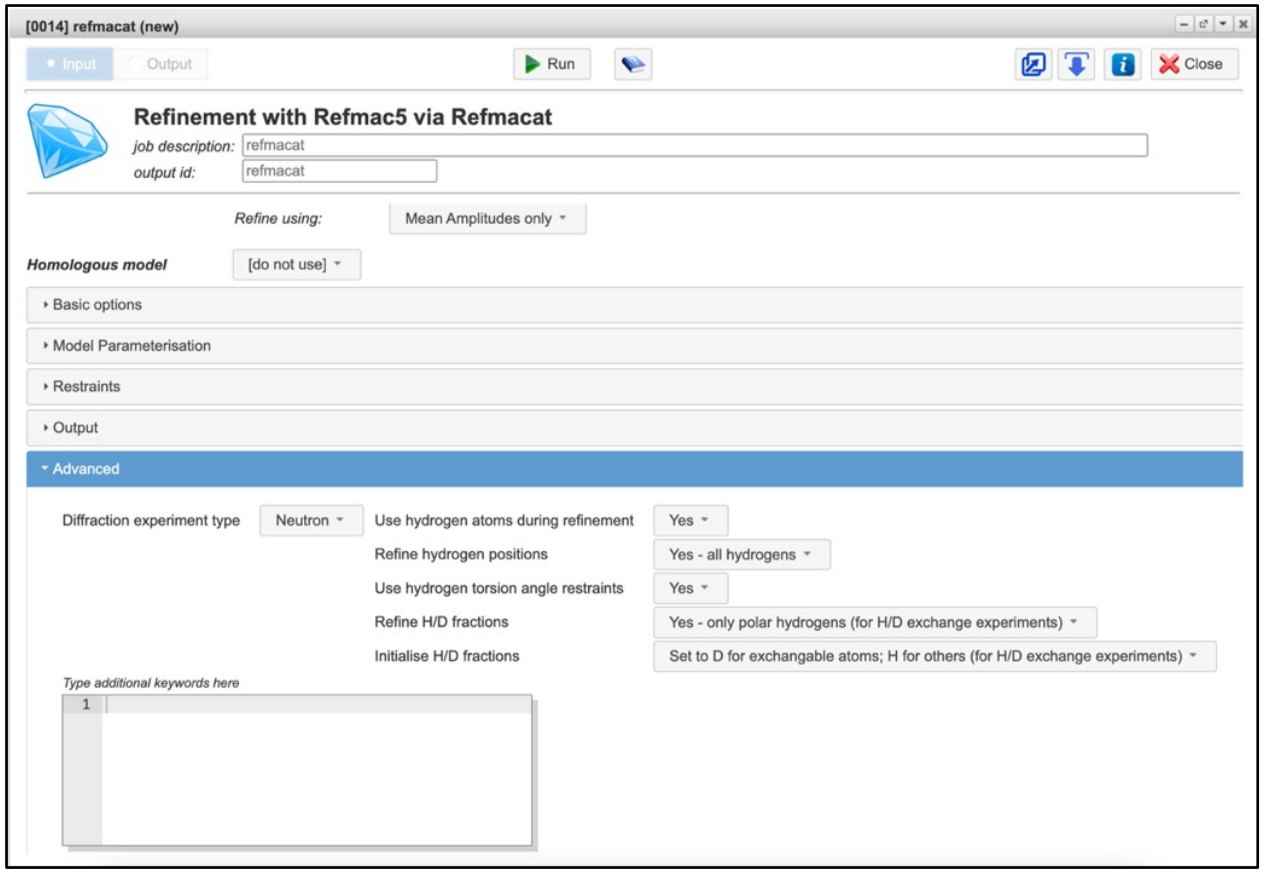
Bolton, R., et al., (2024). *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2308478121.

Availability

CCP4i2



CCP4 Cloud



Documentation in progress...

Future work

- Development of algorithms for the precise arrangement of ***water molecules***. Specifically, they will focus on refining the orientation of hydrogen positions within these molecules, utilising information derived from density maps.

- In addition to considerations driven solely by map-based criteria, these algorithms must incorporate information on ***hydrogen bonding interactions***. This inclusion is essential for obtaining a more comprehensive understanding of the molecular landscape.

- Further experiments with UOX in complex with the natural substrate ***UA*** are important. These, however, will require the use of ***cryo-trapping*** as the ***5-PIU*** intermediate is short lived and thus unstable during the long neutron diffraction experiments

Acknowledgement

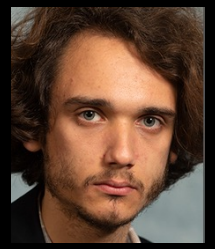
MRC-LMB Computational Structural Biology Group

Paul Emsley

Garib N. Murshudov

Fei Long

Alexander Shtyrov



King's College London

Roberto A. Steiner

Soi Bui

Lindsay McGregor

Global Phasing

Marcin Wojdyr

University of Southampton

Rachel Bolton

Ivo Tews

MRC-LMB Computing

Jake Grimmett

Toby Darling

Ivan Clayson

lucrezia@mrc-lmb.cam.ac.uk

Twitter: lulu_catapano

KING'S College LONDON

CCP4

UKRI MRC Laboratory of Molecular Biology