

A Literature Survey / First Year Postgraduate Studies

Super-resolution imaging of protein structures

Student	Rachel Alcraft
Supervisor	Mark A. Williams
Date	15 th July 2022

Birkbeck, University of London

(8,678 words)

Table of Contents

Figures and Tables	ii
Glossary	iii
1) Introduction	1
2) A brief overview of refinement	2
2.1) Structure factors.....	2
2.2) Refinement models	4
3) A brief overview of resolution.....	5
4) The highest resolution structures	6
5) A review of the claims.....	10
6) A summary of the claims.....	24
6.1) Unusual hydrogen bonds	24
6.2) Non-planar peptide bonds.....	25
6.3) Backbone angles	25
6.4) C:O and C:N+1	26
6.4) Protonation states.....	26
6.5) Electron and proton transfer	27
6.6) Topology analysis.....	27
7) An analysis of the claims.....	28
7.1) Statistical quality	28
7.2) Model bias	30
7.3) Ring artefacts.....	33
7.4) Atom Positions	34
7.5) Hydrogen Bonds and geometry	34
8) Next Steps.....	36
8.1) Statistical quality and model bias	36

8.2) Electron density information from averaging.....	36
8.3) Hydrogen bonds and geometry.....	37
8.4) Atom position.....	37
8.5) Topology and bonds.....	37
9) Conclusion.....	38
References.....	39

Figures and Tables

*Figure 2.1-Diffraction patterns (Crick & Kendrew, 1957),.....	2
*Equation 2.1-Structure factor equations (Afonine, 2010).....	3
*Figure 2.2-Hexachlorobenzene, hand drawing by Kathleen Lonsdale (1931).....	3
*Equation 2.2-The multipole model formula (Afonine, 2010).....	4
Figure 3.1-Resolution over time.....	5
Table 4.1-The highest resolution structures.....	8
Figure 4.1-Citations between the high-res papers.....	9
*Figure 5.1-Interpreting difference density as hydrogen (Moroz et al., 2021 – Figure 5.....	11
*Figure 5.2-Averaged experimental difference density (Jelsch et al, 2000 – Figure 2).	12
Figure 5.3-Visualised electron density of 1yk4.....	14
*Figure 5.4- Histidine protonation states (Liebschner et al, 2013 – Figure 4).....	17
*Figure 5.5-Electron density in the C:O bond (Ahmed 2012).....	19
*Figure 5.6-Correlations between CA:C:N+1 and C:O (Zarychta et al, 2015).....	20
Figure 5.7-Correlation between C:O and CA:C:N+1.....	20
Figure 5.8 Unbalanced rings and hydrogens.....	21
Table 6.1-Hydrogen bond claims in papers.....	24
Table 6.2-Non-planar peptide bond claims in papers.....	25
Figure 7.1-Synthetic IAM density at different b-factors.....	29
*Figure 7.2-B-factor restricted electron density (Ahmed, 2012 – Figure 6.11).....	29
*Figure 7.3-Deformation density pictures of crambin (Jelsch et al, 2000 – Figure 3).....	31
*Figure 7.4 D-orbitals visualised in Nature (Zuo et al, 1999).....	31
*Figure 7.5-A 3d visualisation of deformation density (Pröpper et al, 2013).....	32

*Where asterisked the Figure is directly copied from the referenced paper.

Glossary

- **Atomic resolution** – 1.2Å or better.
- **Bfactors** – or temperature factors, a measure of the disorder of an atom.
- **Bond electrons** – electrons involved in chemical bonds, shared between atoms.
- **Deformation density** – the difference between the experimental electron density and a spherical non-bonded density. This may show non-spherical regions, covalent electrons, and hydrogens.
- **Histidine protonation states** – around pH 7 there are 3 possible protonation states of histidine, with either or both (or neither, not protonated) of ND1 or NE2 protonated (covalently bonded to a hydrogen).
- **Hydrogen bond** – a predominantly electrostatic attraction between a covalently bonded hydrogen atom attracted to a lone pair in another atom. In protein backbones they play an important role in the secondary structure formation, e.g., in a helix the carbonyl oxygen hydrogen is bonded to the backbone HN+4.
- **IAM** – the Independent Atom Model models electron density as spherical atoms using an exponential decay model with scattering factors.
- **Isotropic** – identical in all directions (opposite is anisotropic).
- **Lone pair** – a pair of valence electrons not participating in a covalent bond. Their presence can influence the geometry of a molecule.
- **Low barrier hydrogen bond** – evenly shared strong, short, hydrogen bond.
- **Residual density maps**, or difference density – the difference between the observed and the calculated maps.
- **Riding hydrogens** – hydrogen position implied by their bonded atom
- **Ultrahigh resolution** – 0.8Å or better.
- **Valence electrons** – electrons in the outermost shell of an atom that can form bonds.
- **Temperature factors** – or bfactors, a measure of the disorder of an atom.
- **Laplacian** – the sum of the 2nd partial derivatives or eigenvalues of a numerical matrix, e.g., of the electron density.

1) Introduction

Recent developments in applications of machine learning and artificial intelligence have brought to popular attention the importance of protein structure determination, which can “facilitate a mechanistic understanding of their function” (Jumper et al, 2021). These AI techniques rely on having training data, in this case obtained by accurate structure determination by other methods. The Protein Databank (PDB, Dutta et al, 2010) is a vast, free resource in which all protein structures and, where possible, the experimental evidence used to determine them, are stored - portions of which were used for the training data of AlphaFold (Jumper et al, 2021). By whichever method these structures are determined, a process of refinement is undertaken in which various decisions must be made to juggle a multidimensional puzzle into something that fits with chemical expectations, resulting in a structure - aka a list of 3d coordinates for every atom.

The weighting of these decisions depends on the reason and purpose of the structure determination. It is unlikely to be the case that the purpose of structure determination is an accurate location of every single atom within a protein. When Engh and Huber reviewed the future of geometric parameter standards they were pessimistic, stating that “protein structures are generally solved not to build a statistically optimized protein database, but to discover biophysical functional mechanisms” (Engh & Huber, 2006).

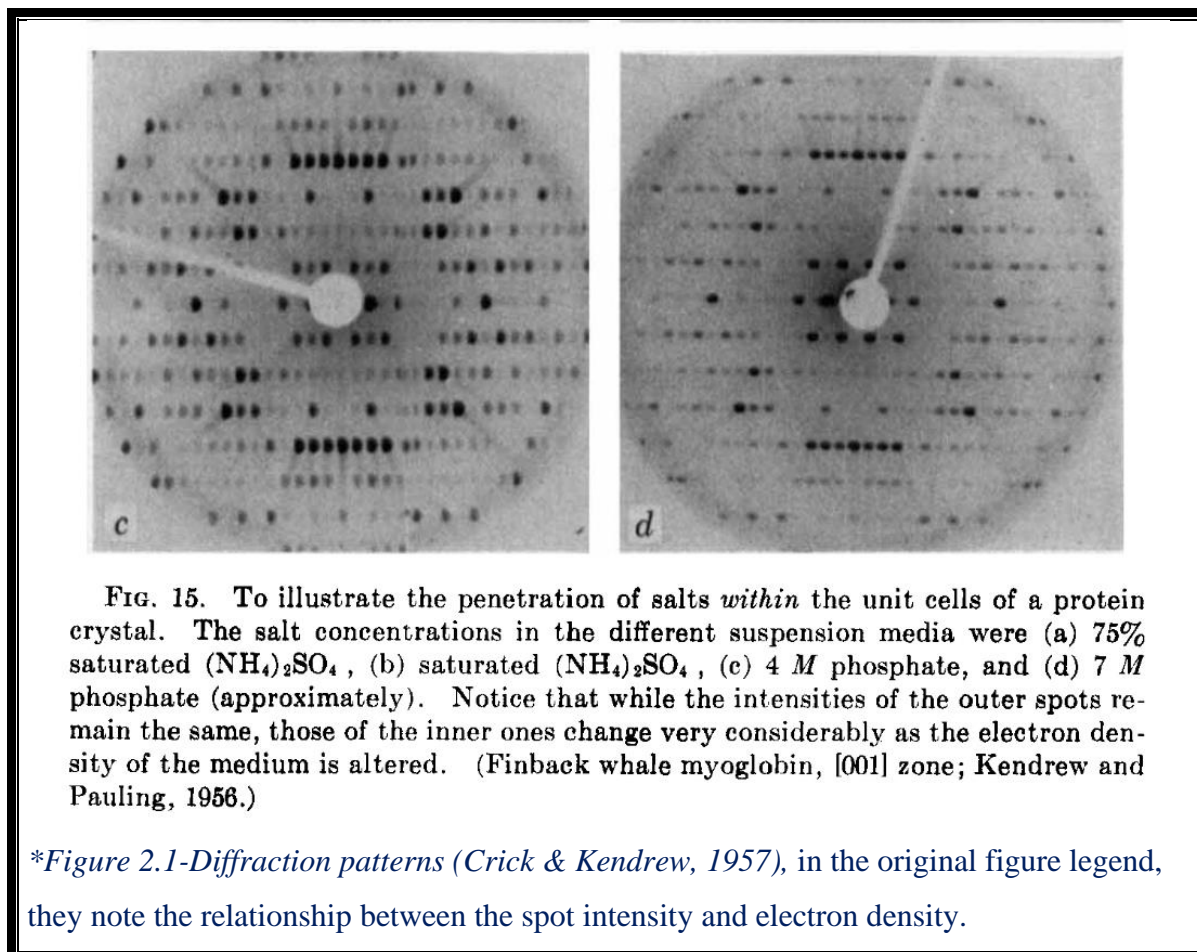
Atom position is not the only information there is to determine function, the experimental results of x-ray crystallography yield the likely distribution of electrons in the structure. As the resolution of x-ray crystallography improves these distributions become clearer until at atomic resolution ($\approx <1.2\text{\AA}$) individual atoms can be resolved, and their electron distribution can be seen to varying degrees of statistical accuracy.

In the tradition of a literature review, this study seeks to be exhaustive on a topic and so sets out to examine the methods, results and claims made by every protein structure solved at an ultrahigh subatomic resolution of $<0.8\text{\AA}$. Initially this study examines all the structures in the PDB that are solved at this level, and then goes on to examine some of the supporting theory and methods, finally identifying issues open for future study.

2) A brief overview of refinement

2.1) Structure factors


The experimental evidence of an x-ray crystallography experiment is a diffraction pattern. Figure 2.1 shows an early diffraction pattern from Crick & Kendrew (1957).



The reflections are created by constructive interference from x-rays diffracting from different planes. The features of each reflection's position and intensity represent a wave with magnitude and relative phase which is represented mathematically as a structure factor. The missing information measurable from the experiment is the phase (Cowtan, 2003) – the reason the solution of x-ray structures is a complicated mathematical puzzle. Since the diffraction is caused by electrons, the experimental evidence relates to the density of electrons encountered. The electron density and structure factors relate by Fourier transforms, and these equations underly the structural solutions in x-ray crystallography. The relationship between these functions is usefully given by Pavel Afonine on the Phenix website (Afonine, 2010) for the

spherical atom model, which is also referred to as the Independent Atom Model (IAM), see Equation 2.1.

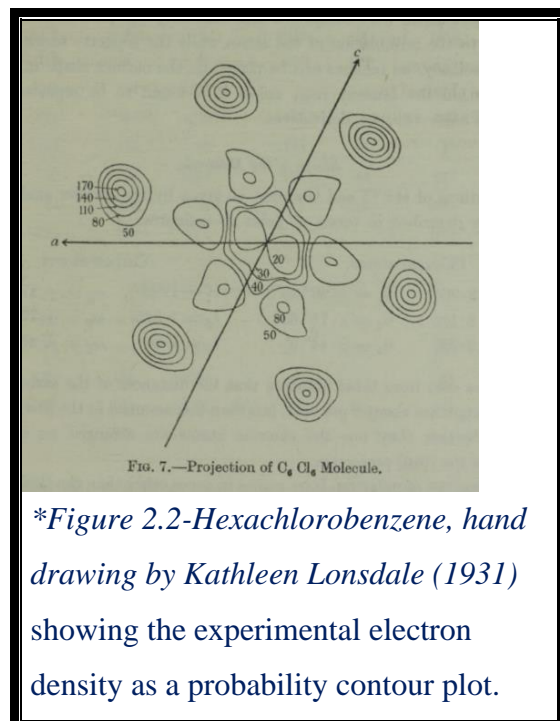
$$\rho_{crystal}(\mathbf{r}) = \sum_{i=1}^{Natoms} q_i \sum_{k=1}^5 a_k \left(\frac{4\pi}{b_k + B_i} \right)^{3/2} \exp\left(-\frac{4\pi^2 |\mathbf{r} - \mathbf{r}_{0,i}|^2}{b_k + B_i} \right)$$


 Fourier transformation

Set of structure factors $\{\mathbf{F}(\mathbf{s})\}$, where each one is:

$$\mathbf{F}(\mathbf{s}) = \sum_{i=1}^{Natoms} q_i \sum_{k=1}^5 a_k \exp\left(-\frac{b_k s^2}{4} \right) \exp(-B_i s^2 / 4) \exp(2i\pi \mathbf{r}_i \mathbf{s})$$

**Equation 2.1-Structure factor equations (Afonine, 2010) where $\rho(\mathbf{r})$ is electron density, q_i is occupancy, a_k and b_k are atomic scattering factors, \mathbf{r} and \mathbf{s} are position vectors.*



The first structure solved by the Fourier Method was in 1931 by Kathleen Lonsdale (Lonsdale, 1931), which she referred to as “Structure Determination by Trial and Error”. Her trial-and-error produced the structure of Hexachlorobenzene, Figure 2.2. This demonstrates how the electron density images were given nearly 100 years ago, as a projection with the probability contours shown.

2.2) Refinement models

An iterative process takes place in which phases are estimated and structure factors computed, and then iteratively improved. Most commonly, the initial phases are implied from the atomic positions of what is believed to be a similar protein, with structure factors calculated from the chosen mathematical models. The initial phases can also be estimated by direct methods (Woolfson, 1987). Two important and common models for the density are described below.

2.2.1) The Independent Atom Model

The Independent Atom Model (IAM) is a model of all the atoms in a structure as if they are not interacting, a promolecule, spherical, using a gaussian estimation of the decay. The formula is given in Equation 2.1. This is the simplest model used to iterate with during refinement and may also be used to compare against the observed density.

2.2.2) Aspherical model - multipole model

At ultrahigh resolutions there is evidence that this model breaks down and the more sophisticated multipole model can be used. This was initially suggested by Hansen and Coppens (1978) and considers interactions between atoms. The formula is replicated from Afonine (2010) in Equation 2.2.

$$\rho_{atom}(\mathbf{r}) = \rho_{core}(\mathbf{r}) + P_{val}\kappa^3\rho_{val}(\kappa\mathbf{r}) + \sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa'\mathbf{r}) \cdot \sum_{m=-l}^l P_{lm} y_{lm}(\theta, \phi)$$

$\rho_{ATOM} = \text{core electrons}$
 $+ \text{valence electrons}$
 $+ \text{non-spherical part of the valence electron distribution}$

**Equation 2.2-The multipole model formula (Afonine, 2010) - this model incorporates valence electrons.*

3) A brief overview of resolution

Towards the end of the 1990s, the first ultrahigh resolution structures of proteins were published, made possible by technological advancements including: synchrotron sources; detectors; cryocooling; and compute power. Traditional x-ray crystallographic refinement uses a spherical atomic density model, geometry parameters, and isotropic b-factors without hydrogens: finding a best fit structure but losing the possible genuine deviation from standards in geometry, losing protonation states and losing non-spherical features of chemical bonds. As the resolution increases to less than the atomic bond distance, it becomes possible to see hydrogen peaks and deformed density from bonds or lone pairs and charge distribution (Hakanpää et al., 2006), potentially elucidating function and interactions. But with the extra information comes a more elaborate problem to be solved in refinement - more complex models with more degrees of freedom are needed to model the non-sphericity such as the multipole model (Hansen and Coppens, 1978), the bond electron model (Afonine et al., 2002) or the invariom model (Dittrich et al., 2004).

Figure 3.1(a) shows the number of protein structures deposited per year, and (b) those that were strictly less than 1.2\AA . For both, there is a blip in 2020 but the atomic resolution depositions appear to continue an upwards trend - the global pandemic has impacted protein structure determination like everything else, and the number of structures solved at atomic resolution is distorted over the 2019-2022 period.

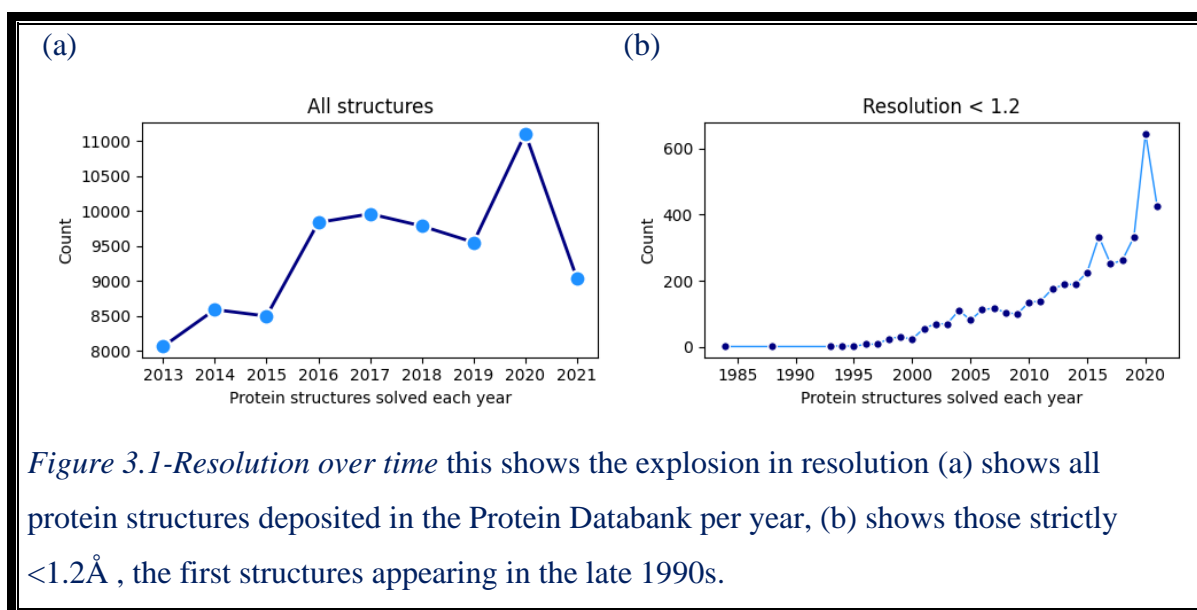


Figure 3.1-Resolution over time this shows the explosion in resolution (a) shows all protein structures deposited in the Protein Databank per year, (b) shows those strictly $<1.2\text{\AA}$, the first structures appearing in the late 1990s.

4) The highest resolution structures

The highest resolution protein structures in the PDB are summarised in Table 4.1. The search was made using the following criteria on 1st March 2022:

- Experimental Method = X-RAY DIFFRACTION
- Refinement Resolution < 0.8Å
- Polymer Entity Type = Protein

This resulted in 54 structures, but some were without deposition papers and removed from scope, leaving 37 deposition papers covering 43 structures. The “Notes” column highlights features or techniques in the paper of interest in this study. The notes in brackets are out of scope in this study.

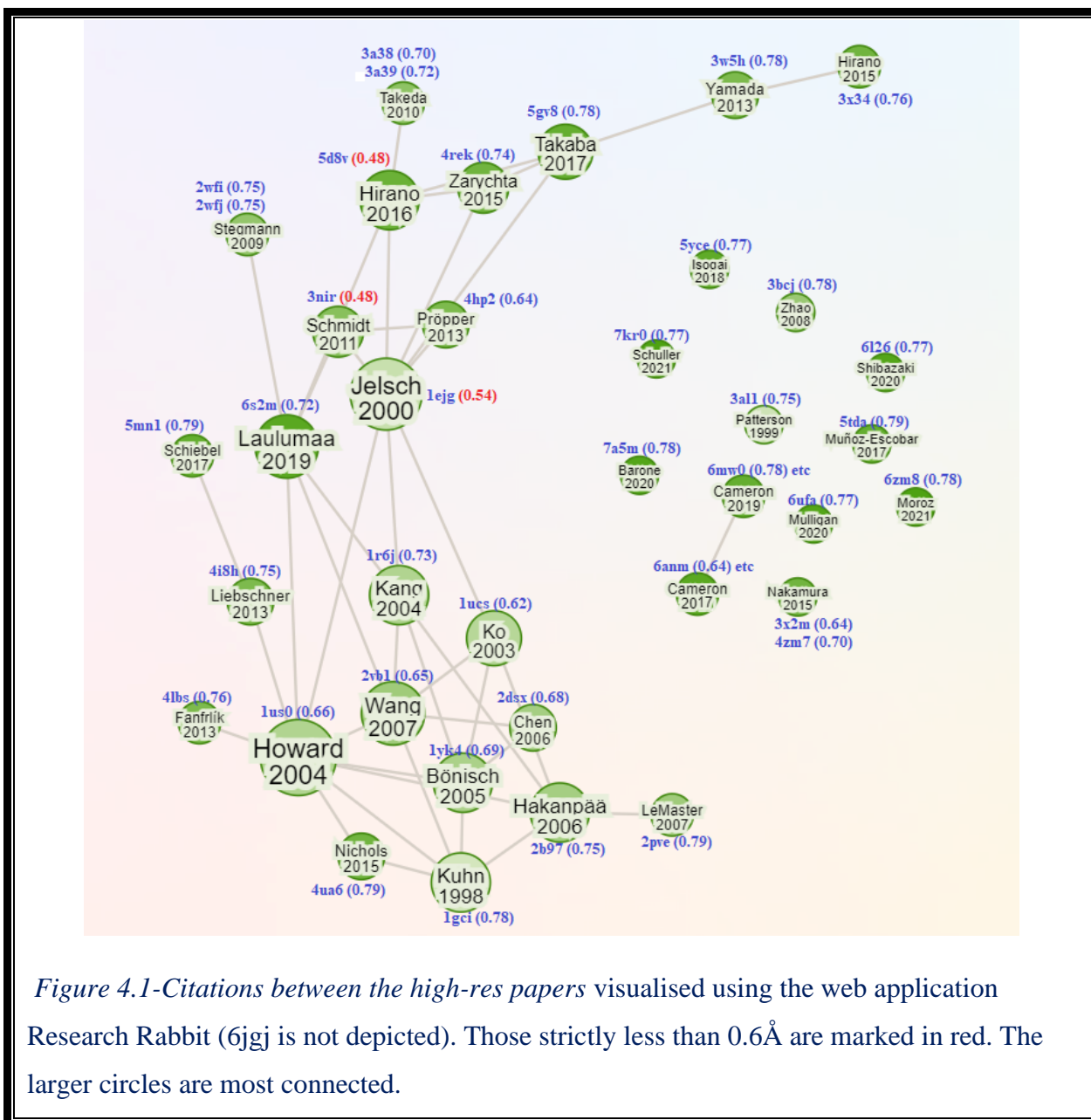
Entry ID	Release Date	Res	Reference	Title	Notes
1ejg	05/04/2000	0.54	(Jelsch et al., 2000)	Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin.	Multipolar Residual maps Deformation maps Transferable library Charge density analysis Average maps Valence electrons
1gci	21/10/1998	0.78	(Kuhn et al., 1998)	The 0.78 Å structure of a serine protease: <i>Bacillus lentus</i> subtilisin.	Hydrogens Hydrogen bonds Difference maps No electron density
1r6j	04/05/2004	0.73	(Kang et al., 2004)	The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between small molecule and macromolecular crystal chemistry	Non-planar omega Hydrogens Hydrogen bonds History and methods
1ucs	06/05/2003	0.62	(Ko et al., 2003)	The refined crystal structure of an eel pout type III antifreeze protein RD1 at 0.62-Å resolution reveals structural microheterogeneity of protein and solvation.	Hydrogens Hydrogen bonds Residual maps Valence electrons Non-planar omega
1us0	07/05/2004	0.66	(Howard et al., 2004)	Ultrahigh Resolution Drug Design I: Details of Interactions in Human Aldose Reductase-Inhibitor Complex at 0.66 Å.	Hydrogens Hydrogen bonds Bonds electrons Geometry deviation Non-planar omega Protonation
1yk4	17/01/2006	0.69	(Bönisch et al., 2005)	Ultrahigh-resolution study on <i>Pyrococcus abyssi</i> rubredoxin. I. 0.69 Å X-ray structure of mutant W4L/R5S.	Direct methods Iron-Sulfur Hydrogen bonds Electron transfer Non-spherical density Fourier truncation
2b97	28/03/2006	0.75	(Hakanpää et al., 2006)	Hydrophobin HFBII in detail: ultrahigh-resolution structure at 0.75 Å.	History and justification Hydrogen bonds Unrestrained Omit maps
2dsx	10/10/2006	0.68	(Chen et al., 2006)	Crystal structure of rubredoxin from <i>Desulfovibrio gigas</i> to ultra-high 0.68Å resolution	Electron transfer Direct methods

					Hydrogens Iron-Sulfur π - π interactions no electron density
2pve	18/12/2007	0.79	(LeMaster et al., 2007)	NMR and X-ray analysis of structural additivity in metal binding site-swapped hybrids of rubredoxin.	Iron-sulfur (Stability) (Mostly NMR)
2vb1	18/09/2007	0.65	(Wang et al., 2007)	Triclinic Lysozyme at 0.65 Å Resolution.	Anisotropic b-factors Unrestrained Conformations Geometry Hydrogen bonds Non-planar omega
2wfi 2wfj	16/06/2009	0.75 0.75	(Stegmann et al., 2009)	The Thermodynamic Influence of Trapped Water Molecules on a Protein-Ligand Interaction	(Free energy) (Mutation)
3a38 3a39	26/01/2010	0.70 0.72	(Takeda et al., 2010)	Detailed assessment of X-ray induced structural perturbation in a crystalline state protein.	Iron-sulfur Hydrogen Electron transfer Omit maps (Radiation damage)
3a1	04/11/1998	0.75	(Patterson et al., 1999)	Centrosymmetric bilayers in the 0.75 Å resolution structure of a designed alpha-helical peptide, D,L-Alpha-1.	(Designed peptide)
3bcj	08/04/2008	0.78	(Zhao et al., 2008)	Unusual Binding Mode of the 2S4R Stereoisomer of the Potent Aldose Reductase Cyclic Imide Inhibitor Fidarestat (2S4S) in the 15 K Crystal Structure of the Ternary Complex Refined at 0.78 Å Resolution: Implications for the Inhibition Mechanism	Hydrogens Difference density Hydrogen bonds Protonation
3nir	18/05/2011	0.48	(Schmidt et al., 2011)	Crystal structure of small protein crambin at 0.48 Å resolution	Multipole Conformations Bond electrons Function not known
3w5h	17/07/2013	0.78	(Yamada et al., 2013)	Elucidations of the catalytic cycle of NADH-cytochrome b5 reductase by X-ray crystallography: new insights into regulation of efficient electron transfer	Electron transfer Hydrogens Omit map Hydrogen bonds
3x2m 4zm7	14/10/2015	0.64 0.70	(Nakamura et al., 2015)	"Newton's cradle" proton relay with amide-imidic acid tautomerization in inverting cellulase visualized by neutron crystallography.	(Neutron crystal'phy) Proton transfer
3x34	15/07/2015	0.76	(Hirano et al., 2015)	High-resolution crystal structures of the solubilized domain of porcine cytochrome b5.	Haem Unrestrained Hydrogens Protonation
4hp2	02/10/2013	0.64	(Pröpper et al., 2013)	Invariom refinement of a new monoclinic solvate of thioestrepton at 0.64 angstrom resolution.	Invariom method Deformation density Omit map Hydrogen bonds Conformations (Electrostatic potential)
4i8h	19/12/2012	0.75	(Liebschner et al., 2013)	On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin.	Geometry Hydrogens Histidine Protonation states Fo-Fc C:O bond lengths
4lbs	30/04/2014	0.76	(Fanfrlík et al., 2013)	Modulation of aldose reductase inhibition by halogen bond tuning.	(Halogen bond) (Free energy) Difference map Bond densities
4rek	15/04/2015	0.74	(Zarychta et al., 2015)	Cholesterol oxidase: ultrahigh-resolution crystal structure and multipolar atom model-based analysis.	Large protein Average maps Topology analysis

					Hydrogen bonds
4ua6	24/06/2015	0.79	(Nichols et al., 2015)	Ligand-Induced Proton Transfer and Low-Barrier Hydrogen Bond Revealed by X-ray Crystallography.	Low-barrier HB Proton transfer Protonation (MD Simulations)
5d8v	25/05/2016	0.48	(Hirano et al., 2016)	Charge-density analysis of an iron-sulfur protein at an ultra-high resolution of 0.48 angstrom	Iron-sulfur Valence electrons Multipole Electron transfer Bader AIM
5gv8	05/04/2017	0.78	(Takaba et al., 2017)	Distribution of valence electrons of the flavin cofactor in NADH-cytochrome b5 reductase.	Topology analysis Hydrogen bonds Non-planar omega Omit map
5mn1	24/05/2017	0.79	(Schiebel et al., 2017)	Charges Shift Protonation: Neutron Diffraction Reveals that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin.	Protonation (Neutron diffraction) (Quantum mechanics)
5tda	22/03/2017	0.79	(Muñoz-Escobar et al., 2017)	Bound Waters Mediate Binding of Diverse Substrates to a Ubiquitin Ligase.	Bound waters (N-degrons) 2Fo-Fc
5yce	19/09/2018	0.77	(Isogai et al., 2018)	Tracing whale myoglobin evolution by resurrecting ancient proteins.	(Evolution) (Engineered) (Free energy)
6anm 6ann	15/11/2017	0.64 0.76	(Cameron et al., 2017)	Crystal and NMR Structures of a Peptidomimetic beta-Turn That Provides Facile Synthesis of 13-Membered Cyclic Tetrapeptides.	(Engineered) (Cyclic) (NMR)
6jgj	17/04/2019	0.78	(Takaba et al., 2019)	Subatomic resolution X-ray structures of green fluorescent protein.	Protonation Hydrogen bonds Topology analysis
6l27	01/04/2020	0.77	(Shibazaki et al., 2020)	Direct Observation of the Protonation States in the Mutant Green Fluorescent Protein.	Hydrogen bonding Protonation Proton transfer (Neutron diffraction)
6mw0 6mw1 6mw2	11/09/2019	0.78 0.77 0.77	(Cameron et al., 2019)	Investigations of the key macrolactamisation step in the synthesis of cyclic tetrapeptide pseudoxylallemycin A.	(Engineered) (Cyclic)
6s2m	28/08/2019	0.72	(Laulumaa & Kursula, 2019)	Sub-Atomic Resolution Crystal Structures Reveal Conserved Geometric Outliers at Functional Sites.	Peptide bonds Bent rings Geometric outliers Non-planar omega
6ufa	02/12/2020	0.77	(Mulligan et al., 2020)	Computational design of mixed chirality peptide macrocycles with internal symmetry.	(Engineered)
6zm8	14/07/2021	0.78	(Moroz et al., 2021)	Fungal GH25 muramidases: New family members with applications in animal nutrition and a crystal structure at 0.78 angstrom resolution.	Protonated Short hydrogen bond Difference map
7a5m	07/10/2020	0.78	(Barone et al., 2020)	Designed nanomolar small-molecule inhibitors of Ena/VASP EVH1 interaction impair invasion and extravasation of breast cancer cells.	(Engineered)
7kr0	09/12/2020	0.77	(Schuller et al., 2021)	Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking.	(Covid) (Fragment docking) Water networks Conformations Electron density

Table 4.1-The highest resolution structures in the protein databank. The “Notes” column highlights analyses or methods in the papers that are interesting in this study. In brackets are interesting aspects of the paper not reviewed further here.

The relationships between the deposition papers for each of these structures were examined using the online tool Research Rabbit (www.researchrabbit.ai). Figure 4.1 depicts the relationships between these papers in terms of mutual citations. Evidently, there is a distinct group of structures with mutual citations, and these are the papers and structures that are primarily in scope for this study. The standalone papers tend to use high resolution for unique purposes – one group are engineering cyclic structures for example (Cameron 2017,2019).



5) A review of the claims

Below is a summary of some of the interesting claims made for these ultrahigh resolution structures, and their purported evidence. I have concentrated on selected claims that outline and advance techniques and evidence that could be examined further in this study. The structures are introduced in chronological order and where I refer for the first time to a new technique or observation, the text is bold.

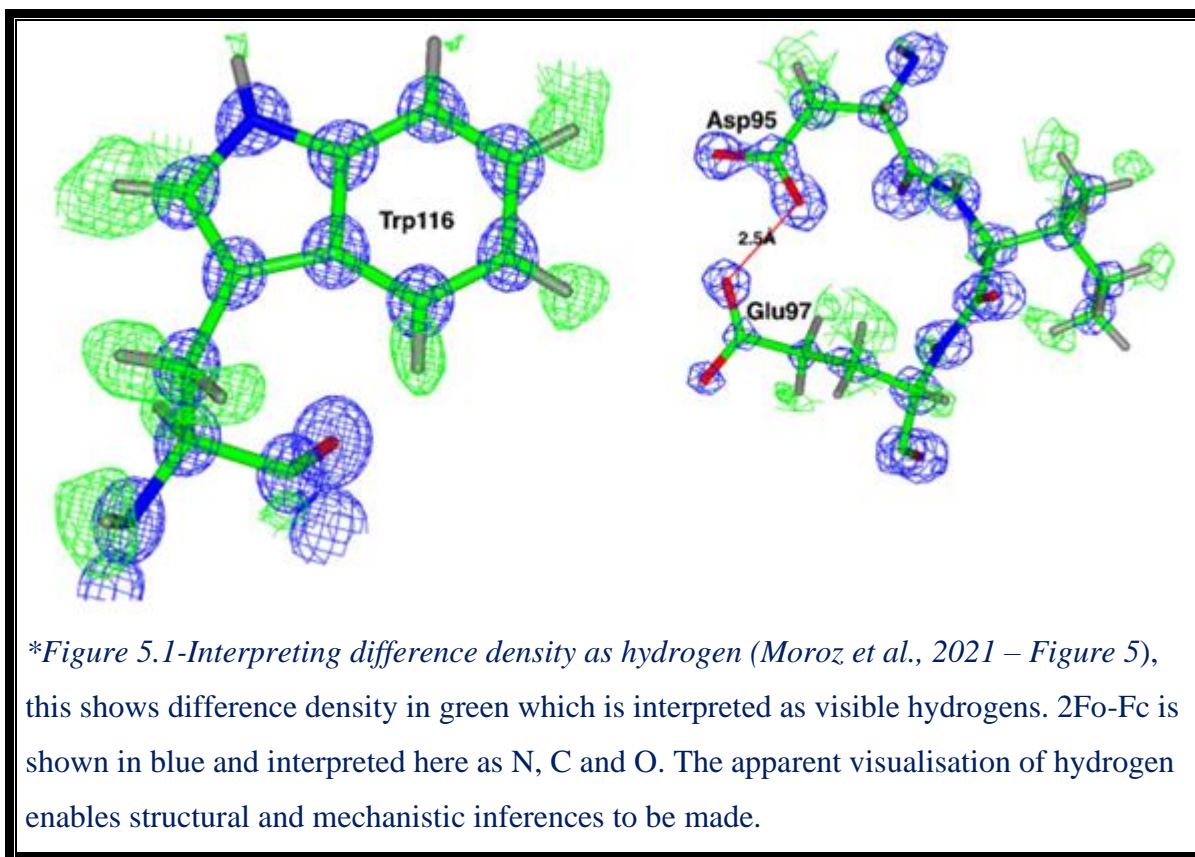
The claims made are often not new discoveries found in ultrahigh resolution structures, rather they provide additional experimental evidence for claims already made either in lower resolution structures or theoretically.

One of the most important tools for the analysis of structures, which is even more accurate at ultrahigh resolution, is the ability to **see hydrogens in the difference maps**. There are different kinds of difference maps which can be very briefly summarised as:

- Difference density - the difference between observed density and the model density, which may show unmodelled features such as hydrogen.
- Hydrogen omit maps – specifically omits hydrogen so that hydrogen can be inferred in the difference density.
- Deformation density – compares the observed density with a spherical model (not necessarily the one used in refinement) and may show experimental evidence of non-spherical features.

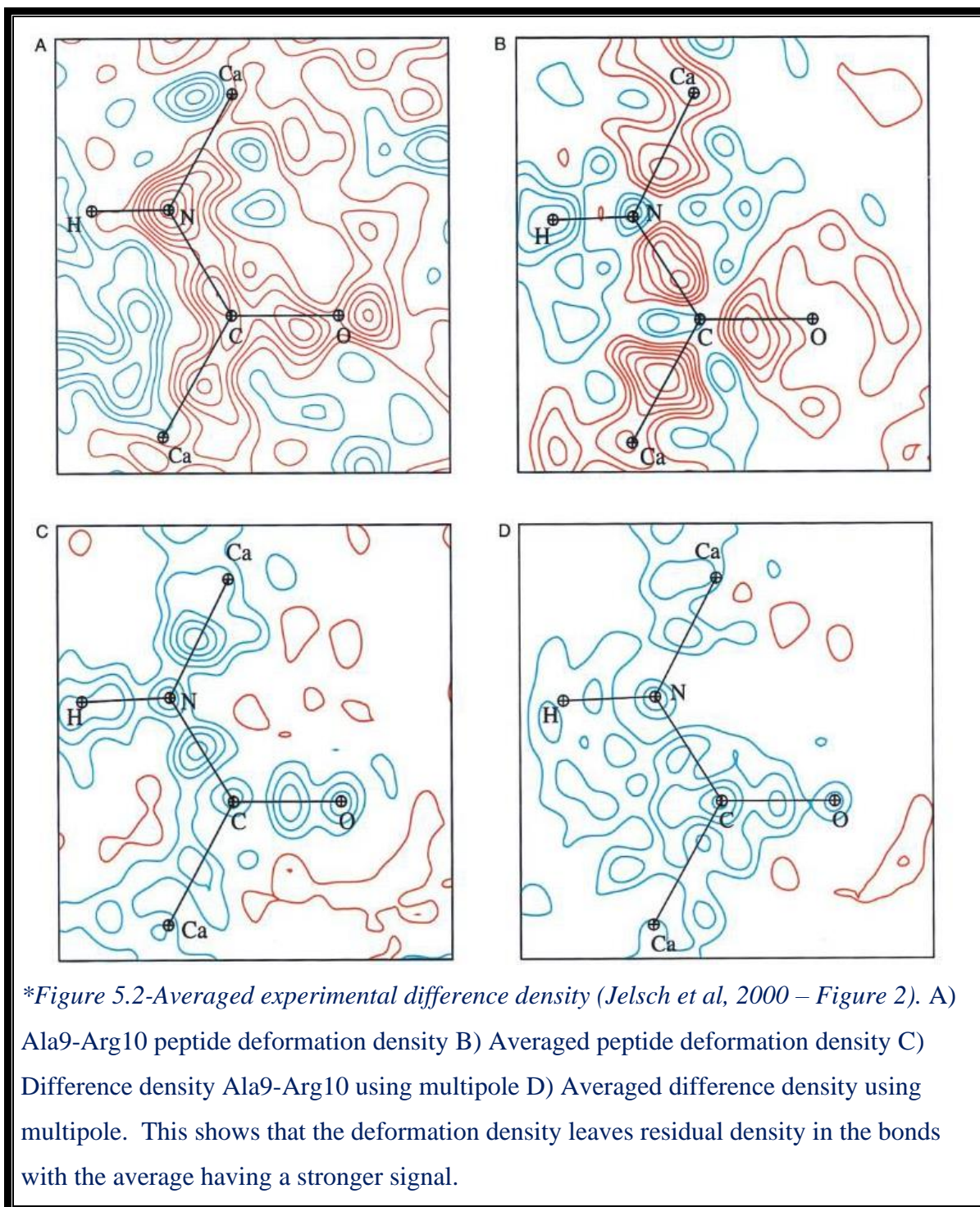
Since 2008, it has been a requirement to deposit electron density with the structure in the PDB as $2F_o-F_c$ and F_o-F_c where F_o is the observed density (experimental) and F_c is the calculated density (model). At the end of the iterative process to solve the phases there are two sets of structure factors: for the observed density and for the model. The difference density is created by converting the difference in these structure factors to electron density.

In Figure 5.1 the difference density in green is superposed over the electron density in blue. The green difference is positive which means that there exists observed density that wasn't modelled. Since the hydrogens weren't modelled that means, by the conclusion drawn by these methods, that the positive density blobs are hydrogens, which is the basis of much analysis, including hydrogen bonds, protonation states, active sites, and transfer mechanisms.



The first $<0.8\text{\AA}$ ultrahigh resolution structure was deposited in 1998, **1gci**, at 0.78\AA (Kuhn et al, 1998). It uses the difference technique to find an **unusually short hydrogen bond** of 2.62\AA between Asp32 OD2 and His64 ND1 having observed a hydrogen 1.2\AA from ND1 and 1.5\AA from OD2. This short hydrogen bond is part of the functional catalytic triad, and the hydrogen had previously been identified by NMR experiments. It was possible to view this hydrogen by a repeated refinement adding more hydrogens as they became visible, iteratively, which gradually reduced the noise in the map and enabled ever more hydrogens to become clear.

The ultrahigh resolution structures discussed in this study fulfil the requirements that are needed to refine a structure using a more sophisticated model than IAM. Those requirements suggest that the structure needs to have a resolution of $\leq 0.85\text{\AA}$, good electron density with sharp peaks (disorder, flexibility and multiple occupancy will limit the use) and that when a spherical refinement model is used there should be deformation density present in bonds (Pröpper et al, 2013).



In 2000 one of the very highest resolution structures was deposited, **1ejg** at 0.54Å (Jelsch et al, 2000) which has proved to be an important structure in the field. This used an aspherical model for refinement – the multipole model (Hansen and Coppens, 1978). The spherical model for this structure breaks down as valence electron density can be seen in chemical bonds (Jelsch et al, 2000). The multipole model considers the aspherical nature of electron density and the

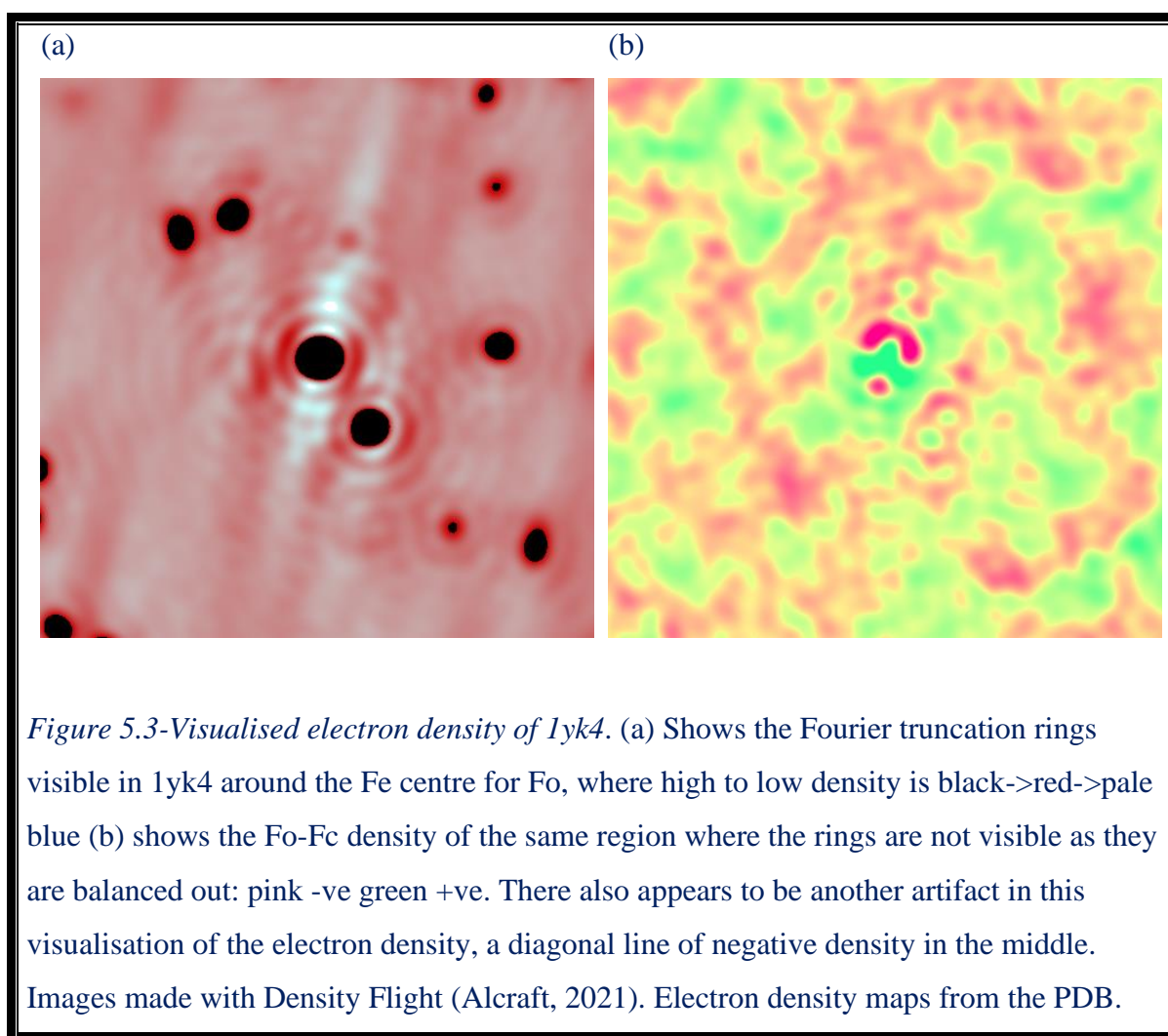
valence electrons in atomic bonds to refine against a more accurate model. To see the valence electrons in difference density deformation density is used - comparing the observed density with a spherical model. For 1ejg, Jelsch et al additionally use a technique of averaging the deformation density over all the ordered peptide bonds to observe the finest features they can as the averaging reduces the statistical noise. A striking image is shown in the paper of deformation electron density featuring a carbonyl oxygen lone pair in a static map (Jelsch et al, 2000 - Figure 3, replicated here in Figure 7.3) with the average deformation image of the peptide bonds looking quite similar (Figure 5.2).

Structure **1ucs** uses a spherical model for refinement but still achieves visualisation of valence electrons (Ko et al, 2003), residual density is seen in most bonds and near the sulphurs of methionine. Using the observable hydrogens, they are able to identify 2 unusual hydrogen bonds - 2 intra-residue hydrogen bonds with acidic sidechains (GLU35/OE2 and ASP58/OD1) both bend back **to hydrogen bond with their own backbone N**.

High-resolution structure determination allows relaxation of restraints in refinement and greater confidence in geometric outliers from experimental evidence (Laulumaa & Kursula 2019). One important feature from this is the identification of **non-planar peptide bonds** and functional analyses that may result. This is the case for **1r6j** (Kang et al, 2004) with a lowest omega (ω , CA:C:N+1:C+1) value of 162.2° for SER261 with electron density evidence. This structure also uses observable hydrogens to identify hydrogen bonds - from this they can identify weak hydrogen bonds of the **unusual type CH-O** stabilising a beta-sheet with a hydrogen bond distance longer than usual at around 3.16Å, and 3.17Å with the corresponding amide group and the C α .

In structure **1us0** (Howard et al, 2004) they counted hydrogen as observable if it had a difference density greater than 1σ , and they observed a linear correlation between observed hydrogens and atom temperature factor. They report 27 cases of deviation from peptide planar of more than 10° - for some of these they attribute the deviation to sp³ hybridisation, and in most of the cases there is evidence of a strong hydrogen bond formed by the N, for example SER76-LYS77 with omega of 167.4° which is stabilised by a hydrogen bond between the SER76 OG and LYS77 N of an unusually short distance of just 2.68Å. Structure 1us0 observes a departure from spherical density in its residual electron density found around atoms near the active site. They use the presence of density in bonds to infer the bond type, for example between residues 44/45 there is residual bond density in both C:O and C:N+1, but in 45/46 the

bond density is mostly in C:O. From this they infer that the 45/46 peptide bond has a **C:O double bond and a C:N+1 single bond**, which is also suggested by the bond geometry. For 44/45 there are distances of C:O=1.241Å and C:N+1=1.330Å and for 45/46 C:O=1.220Å and C:N+1=1.355Å. 1us0 clearly shows the orientation and **protonation state** for His110 in the electron density via visible residual density for the NE2 atom. These protonation states are depicted in Figure 5.4.



1yk4 (Bönisch et al, 2005) was able to identify unambiguously all atom types by the density volume, except for the disordered Lys7 sidechain. The hydrogen bond network between iron and sulfur is of interest, and the experimental data was able to confirm 6 previously suggested hydrogen bonds and refute 3. They report 8 peptide bonds that significantly deviate from planarity, and an additional insight - they note a correlation between the **backbone angle tau** (N:CA:C) and backbone dihedral psi (N:CA:C:N+1), along with a bimodal distribution of tau with means of 109.2° and 113.5° (in line with 2 different psi populations). They analysed the

distances between **C:O and C:N+1 and found a negative correlation**, noting that at resolutions $<0.8\text{\AA}$ the assumption of spherical atoms no longer holds. The structure 1yk4 has evident bond electron density in the difference maps, refined with a spherical model, citing Afonine (Afonine et al, 2004) as evidence that well-ordered atoms at ultrahigh resolution can show bond electron density. Also citing Afonine, they point out the risk of **Fourier truncation effects** in what they describe as unbalanced density maps (Bönisch et al, 2005, Afonine et al, 2004). This is something I can observe directly in the electron density, see Figure 5.3 using my own visualisation tools from Density Flight (2021). I note that the advice (Afonine et al, 2004) that you can have confidence in the Fo-Fc map as the rings “balance” could be unsatisfactory in terms of method precision and it additionally suggests that we cannot have confidence in 2Fo-Fc.

Conventional refinement of ultrahigh **2b97** (Hakanpää et al, 2006) left residual density in: most of the carbonyl oxygens; around the main chain peptide bonds; at the sulfur positions in disulfide bridges; and around the manganese ion – only the manganese residual density was present in the 1.0\AA high-resolution structure on which the solution was based. In 2b97 520/990 H atoms could be directly observed in the omit map. The combination of observable hydrogens and accurate bond lengths allows for identification of protonation states, but it does need the restraints to be relaxed in refinement for confidence in the experimental evidence of the bond lengths. In this structure this was only partially successful as restraints were put on some of the weaker density side chains, and with nearly half the hydrogens not visible (an absent hydrogen does not exclude its protonation). A technique was employed in 2b97 to measure the **true error of the geometry** in the unrestrained structure: they used the variability in the $C\alpha$ - $C\beta$ which would not be expected to change for functional or structural reasons, measuring 5 more than 0.05\AA from the small molecule standards. This structure had an extreme unrestrained ω -outlier at 159.1° between Ile31 and Val32. Unsatisfied hydrogen bonds were found, so they searched for unusual weak hydrogen bonds formed between carbons or π -rings at $\leq 3.8\text{\AA}$. 47 were found, and on inspection 41 of these had non-zero density where a hydrogen would be expected, making them considered to be directly observed. Short hydrogen bonds were also found in the structure – 2.67\AA was the shortest reliable hydrogen bond found between His42O and Ser45OG.

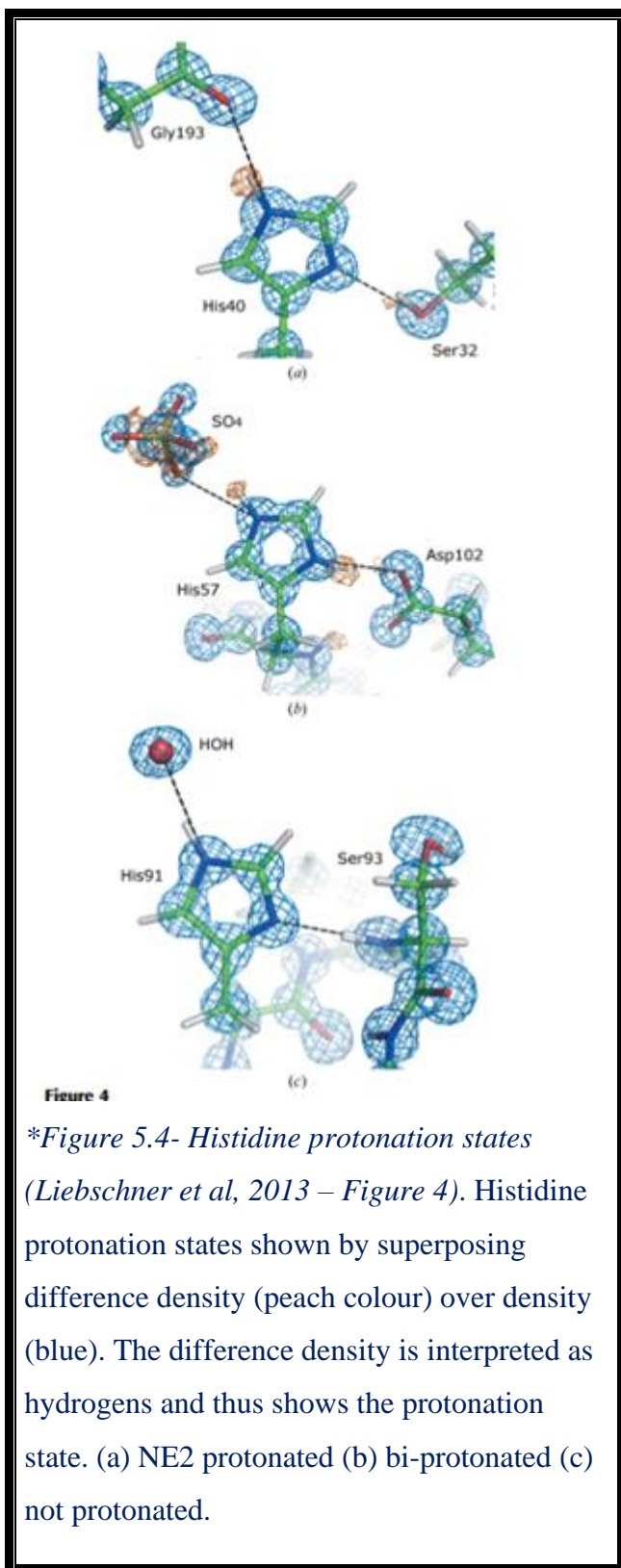
The presence of **alternate conformations** can be seen with accuracy in some ultrahigh-resolution structures, and this can give evidence for functional features. The structure **2dsx** (Chen et al, 2006) hypothesises the structural features that enable **electron transfer** – via the

iron centre and through aromatic rings. They suggest that the multiple conformations of the iron-sulfur cluster demonstrate this, along with the evidence of hydrogen bonds in this area and the binding of water to peptides. They also observe an asymmetric shifting of density around the 4 sulfurs, as they say “Thus, SG9 and SG42 densities move toward the Fe atom whereas SG6 and SG39 densities shift away from the Fe atom”. They suggest this could be the reason for the asymmetrical bond lengths and that it could imply some dynamism: “The **direction of density movement** might imply the dynamic range of the FeA4Cys centre, which is related to the electron transfer mechanism”. No other paper in this study mentions direction of density movement as relating to function. This sort of density asymmetry could be accounted for by occupancy and by bfactor, both perhaps a dynamism, and we see in other papers that bond distances are discussed in relation to function. Unfortunately, they did not deposit the electron density for this structure so the claims cannot be assessed.

Structure **2vb1** (Wang et al, 2007) also shows multiple conformations along the backbone and side chain. Some of these conformations are hinged by a hydrogen bond which is evident at high resolution, such as the atoms around His15, Asp87, Asp90 and Thr89. There is also evidence in the difference map for this structure having 2 hydrogen positions around the hydroxyl of Tyr53, allowing 2 possible alternative hydrogen bonds in the vicinity – these fine geometric features could suggest functional features. The tau angle is again mentioned in this structure, where they note that the unrestrained distribution is wide and might correlate with the local conformation, citing Esposito et al (2000) who detect a clear dependence of tau on conformation. This structure also shows outliers in ω -planarity, with the largest deviation being residues 62-63 with an angle of 152.7°.

In structure **3bcj** (Zhao et al, 2008) the existence of hydrogen densities near the amide ring gives evidence of protonation state and hydrogen bonding that led to a determination of the inhibitor mechanism.

Structures **3a38** and **3a39** (Takeda et al, 2010) show clear features of non-spherical density around the Fe-S bonds, indicated by the presence of density in the deformation map for the structures with less radiation damage, which they interpret as visible bonding electrons. In this iron-sulfur electron transfer cluster, they state that the position of the hydrogens in the bound waters define their orientation rather than just their position as is found in medium resolution structures. In this structure they found some water molecules with 2 hydrogen bumps which is a feature not yet understood.



It is not always possible to identify function even at the very highest resolution. The function of crambin remains elusive despite the highest resolution 0.48Å structure **3nir** (Schmidt et al, 2011). Although pushing the refinement to what they consider to be the absolute limit of current practicalities and interpretation of electron density, further understanding of the function was not made.

Structure **4hp2** uses IAM refinement to establish that valence electron density can be seen, which they say is a prerequisite for a further aspherical model (Pröpper et al, 2013). They then go on to use the **invariom method** (Dittrich et al, 2004) for further refinement which is an alternative method to the multipole model (Hansen & Coppens, 1978). They show that residual density that is present in bonds after IAM refinement disappears after invariom refinement, thus suggesting the model has better taken the bonding electrons into account.

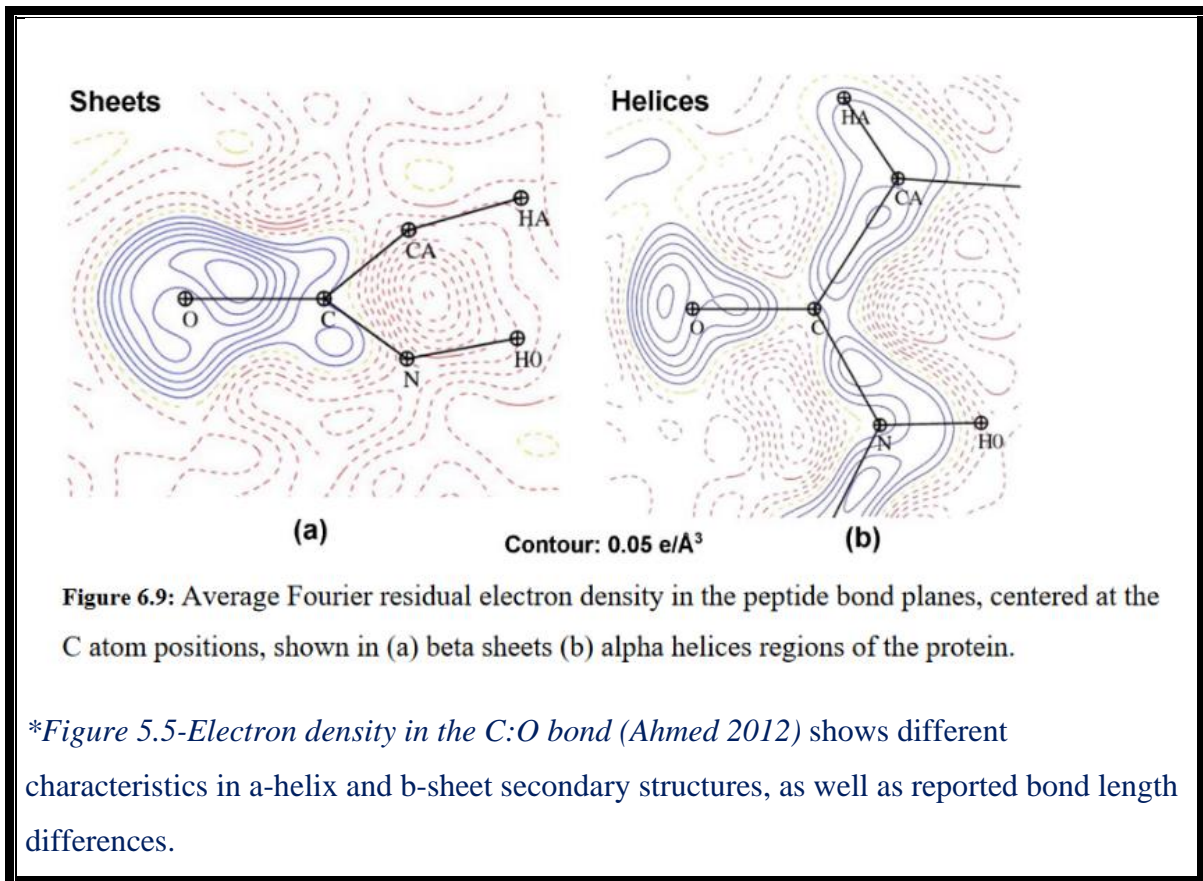
A study of significance in the examination of accurate bond positions and geometric parameters is 4i8h which compares 5 high resolution trypsin

structures (Liebschner et al, 2013). One of the structures is in scope for this study at 0.75Å although the other 4 are all below 0.9Å. It looks at the comparative calcium position wrt to the 6 surrounding oxygens in the 5 structures to examine variability. This study shows great agreement between these coordinate positions in all but one water molecule that is not bound

to the protein, which suggests it is less constrained than the others and the difference may be observed rather than an uncertainty of measurement. **Disulfide bonds** are an important feature in protein structure and this study compared the 6 disulfide bonds. The standard length is 2.04Å and they took the bond to be broken above 2.20Å. Here they found that the highest resolution structure was the one with the most broken disulfide bonds, with only 3 fully intact: and it also shows more cysteine conformations. This is consistent with known impacts of radiation damage at high resolution. In some cases of multiple occupancy, it is suggested that multiple occupancy shows evidence of function, but in this case, it is suggested that there is a breakdown in structural stability because of the effects of the experiment. The five-structure comparative study (Liebschner et al, 2013) looks at the 3 histidines in all 5 structures and infers the protonation states based on deformation density, where visible peaks for hydrogens are seen in the difference maps, shown in Figure 5.4 which is representative of figures of this type that depict protonation states.

Structure **4rek** (Zarychta et al, 2015) applies **Bader's Atoms in Molecules theory** (AIM, Bader, 1994) for a topology analysis of the charge density. In this theory, a saddle point and bond path between 2 atoms indicates an interaction – using this theory they tabulate all the atoms in the protein that interact with the FAD cofactor. An analysis of this type, using the second derivative, can only be done with extremely high-quality electron density. An average density map is also plotted for 4rek comparing the impact different b-factors have on the results. The bonding density is clear, and the clarity decreases as the b-factors increase. Structure 4rek undertakes a stereochemical analysis of the main chain hydrogen bonds in the large structure (>500 residues), looking at both distance and directionality using a mixture of observable hydrogens and known stereochemistry. With the observable hydrogens, they can measure the N-H-O angles and the N-H and O-H distances, adding to the understanding of this geometry – with nothing much surprising, but the interesting observation that *i*+3 α -helices have bifurcated hydrogen bonds with D(onar)-H-A(cceptor) angles mostly in the 120-140° range and long O:H distances > 2.5Å. The C:O bond lengths were examined in the structure 4rek (Zarychta et al, 2015) and compared with the C:N+1 distances for different secondary structures – as previously discussed in structures 1us0 and 1yk4. Here, they find that the mean *i*+3 C=O bond length is longer than the other secondary structures, which is of interest in combination with their observations (reported in this study) on the *i*+3 hydrogen bond geometry being distinct. They also suggest a relationship between the shorter peptide bond distances in the structure in α -helices which correspond to the longest C:O distances. A related observation has been made

by the same group where they averaged the electron density in α -helices and β -sheets separately and found a distinct difference in the character of the electron density and distances (Ahmed, 2012), shown as Figure 5.5. In the 4rek paper they additionally look at the correlation between C:O and the backbone angle $\tau+1$ (CA:C:N+1) because they observe that this angle is lowest in β -sheets. They find no meaningful correlation, although a distinct grouping for different secondary structures – see Figure 5.6.



Independently, my own analysis on a large population of high-resolution structures has concurred that there is no evident correlation between these distances using the coordinates in the pdb, but if instead we look at the correlation between electron density maxima near the atom coordinates, then there is a very strong correlation (Figure 5.7).

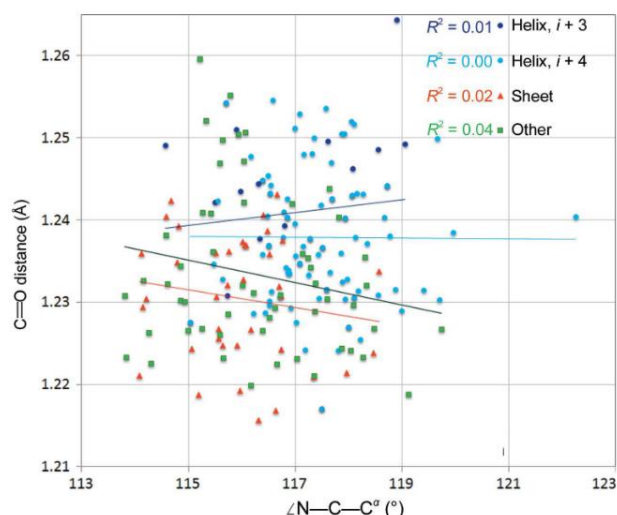


Figure 18
Scatterplot of the C=O distances *versus* the $\angle C^\alpha - C - N$ angles. The linear fits are shown to highlight the average values of distances (despite correlation coefficients of close to zero between the two variables).

**Figure 5.6-Correlations between CA:C:N+1 and C:O*
(Zarychta et al, 2015) this shows weak correlations between the geometric parameters and weak clustering for different secondary structures.

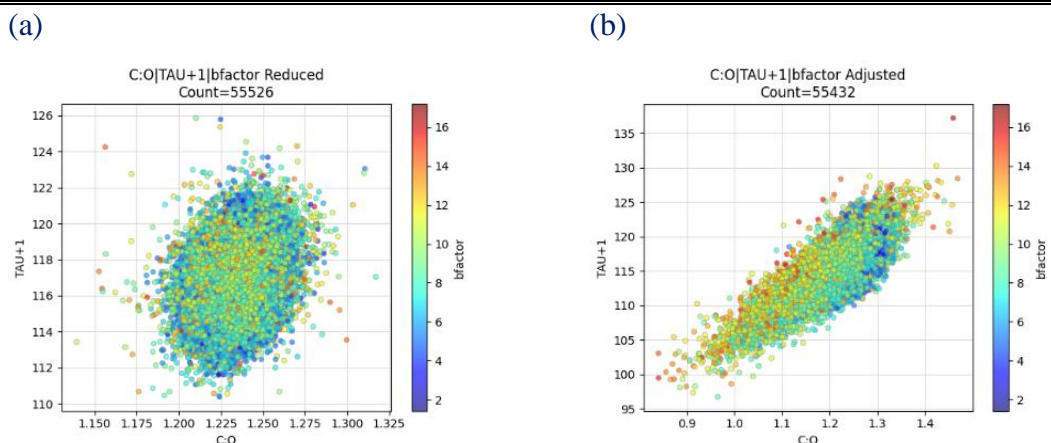
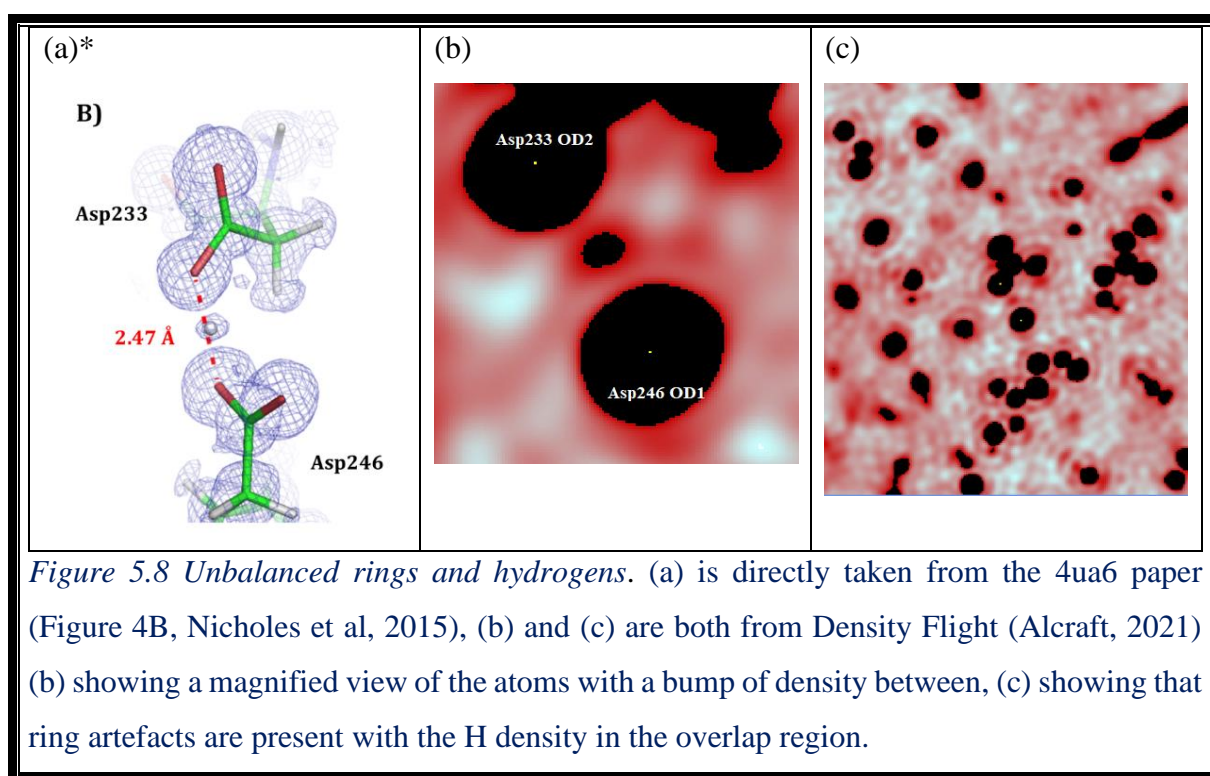


Figure 5.7-Correlation between C:O and CA:C:N+1 (tau+1) where a) is calculated from the PDB coordinates for high resolution structures and b) is the nearest electron density maxima to the atom coordinates for the same structures (some fail to converge) (LeucipPy, Alcraft, 2021). Why there is a linear relationship in the maxima positions might warrant further investigation – and is there are relationship with bfactor?

The side chain protonation states are implicated in enzyme activity and a 3-structure study including **4ua6** at 0.79Å (Nichols et al, 2015) shows an unusually **short low barrier hydrogen bond (LBHB)** of 2.53Å between Ser70 and Lys73 with a verified hydrogen in the difference map. These 3 structures provide experimental evidence into the mechanism of this proton transfer and evidence for LBHB in proteins that are the subject of debate. This same structure also finds a LBHB that seems to play a structural role in the protein between Asp233 and Asp246, also with observable hydrogen. For this they use the 2Fo-Fc map rather than the difference map to observe a hydrogen peak and possibly the unbalanced map with ring artefacts could be problematic. See Figure 5.8 for their Figure (a) and the 2Fo-Fc density from Density Flight (b,c) (Alcraft, 2021). In the more distant view 5.8(c) ring artefacts can be seen, and the hydrogen position is possibly where they overlap making it hard to distinguish physical evidence from the artefact.



One of the structures, **3x34** (Hirano et al, 2015) is analysed along with 4 structures, the others just out of scope, in oxidised and reduced forms 1 and 2 that examine the haem group (unrestrained refinement) of cytochrome b₅. They compare the C:O distances in the propionate

rings of these structures and use the distances as evidence that the reduced haem indicates C:O double bonds (1.21Å) but in the oxidised state the equal distances indicate a charged state.

The ultrahigh resolution structures **3x2m** and **4zm7** have enabled an understanding of a proton relay, in conjunction with **neutron diffraction** which contrasts experimental evidence of neutron (atom) position. The position of H between Asn92 and Asp114 suggest the presence of a proton relay through water molecules.

The joint highest resolution structure in the pdb, **5d8v** (Hirano et al, 2016) examined the electron distribution around the iron-sulfur clusters. They identify and visualise S *3p* and Fe *3d* electron density and perform a topology analysis (Bader, 1994) in which they analyse Laplacian and gradient maps in the iron-sulfur region. An important feature of this topological analysis is the **bond critical point** (BCP) that is marked by the crossing point of the boundary between the atoms and the path between them and BCPs were calculated for all the Fe-S interactions, finding slightly distorted paths asymmetry and unequal strengths at the BCP. They find an inconsistency in the AIM analysis of the charge and the multipole parameters which they could not interpret easily.

Topology analysis continues in 2017 with the structure **5gv8** (Takaba et al, 2017), finding the BCP for the isoalloxazine ring and a topological analysis of the hydrogen bonds. They find the hydrogen bond paths to be slightly curved, in contrast to the straighter covalent bonds and they find some bond paths in non-classical hydrogen bonds for FAD with C α of Tyr65 and C α of Ile81. There are other unexpected features identified in topology analysis, for example single bonds where double ones are expected and a surprising electron distribution in FAD around N5 via hydrogen analysis.

Protonation is the main interest behind the high-resolution solution of structure **5mn1** (Schiebel et al, 2017), and they combine neutron diffraction with x-ray diffraction to investigate binding of trypsin with aniline and 2-aminopyridine with experimental evidence for the protonated forms given by the combination of electron density and neutron density maps.

Further topological analysis is undertaken for structure **6jgj**, in which they observe that the usual method of understanding protein interactions is through geometry, but that the ultrahigh resolution structure allows for a topology analysis that detects intramolecular interactions as features in the electron density (Takaba et al, 2019). They find that their electron density is of good quality as most of the hydrogens are observable. They use a method to find non-covalent interactions by Johnson et al (2010) with the gradient, Laplacian and Hessian eigenvalues used

to identify features such as nuclei positions with local maxima and all negative eigenvalues. Where there is 1 positive and 2 negative eigenvalues this identifies bonding regions, with the sign of λ_2 important in the bonding type. In this analysis they found conventionally understood interactions plus a new one between C2 and Thr62O which they regard as a lone pair π interaction.

The structure **6s2m** (Laulumaa & Kursula 2019) has geometric outliers which the group kept during refinement after inspecting the electron density for evidence. These include non-planar Arg guanidinium groups; non-planar Phe16 and Trp97 rings; peptide bond distortions on both sides of Thr80. They further suggest that geometric outliers can be functional, citing Flocco & Mowbray (1994) who detail the statistically high occurrence of planar stacked arginines over the centre of aromatic sidechains - the proposal being that the structure is stable, yet a hydrogen bond has not formed with the three N atoms of guanidinium group, leaving them available for other functional purposes. More examples can be found of unusual hydrogen bond positions that suggest functional features in the structure 6s2m - there is evidence through hydrogen bond geometry that there is a break in a beta sheet at LEU10 which is a known portal region.

The paper for structure **6l27** (Shibazaki et al, 2020) uses the hydrogen bonding network, in association with alternate conformations to make claims about the protonation state and proton transfer in these conformations around the residue Glu222 and the different hydrogen bonds seen with OE2 and OG2 in the 2 conformations.

The most recent structure, **6zm8** (Moroz et al, 2021), suggests that Asp95 donates a proton to the substrate, the short hydrogen bond distance of 2.5Å between Asp95 and Glu97 provides evidence for this claim.

6) A summary of the claims

Some of the above claims are summarised by category for ease of reference.

6.1) Unusual hydrogen bonds

Many examples are given of unusual hydrogen bonds throughout the papers, a few are detailed in Table 6.1 with numerical examples.

(a) Unusually short hydrogen bonds.				
Structure	Residues	Distance	Comment	Evidence
1gci	Asp32 OD2-His64 ND1	2.62Å	Short HB in catalytic triad	2Fo-Fc
1ucs	Asp58 OD2-Glu25 OE2	2.46Å	Crystal contact short HB	Distance
1us0	Ser76 OG-Lys77 N	2.68Å	Short HB	Distance
2b97	His42 O-Ser45 OG	2.67Å	Short hydrogen bond	Omit map
4ua6	Ser70 OG-Lys73 NZ	2.53Å	LBHB – active site	2Fo-Fc
4ua6	Asp233 OD2-Asp246 OD1	2.47Å	LBHB – buried residue	2Fo-Fc
(b) Unusually long hydrogen bonds.				
1r6j	Ile269 CA-Asn237 O	3.16Å	Weak CH-O HB in β -strand*	Fo-Fc
1r6j	Ile269 N-Asn237 O	3.17Å	Weak β -strand amide HB	Fo-Fc
(c) Non classical partners				
1r6j	Ile269 CA-Asn237 O	3.16Å	Weak CH-O HB in β -strand*	Fo-Fc
5gv8	Tyr65 CA - FAD	-	CH...N/O non classical	Topology
(d) Functional				
2vb1	Tyr53 OH-Asp66 O	2.7Å	2 conformations could suggest function	Omit map
	Tyr53 OH-1054 O	2.9Å		

Table 6.1-Hydrogen bond claims in papers, this summarises the hydrogen bonds that are specifically mentioned with residues and distances with (a) unusually short, (b) unusually long and (c) non classical partners and (d) functional.

(* Entered in 2 categories: long and non-classical.)

6.2) Non-planar peptide bonds

The peptide bond has been considered planar since Linus Pauling “because of the resonance of the double bond between the carbon-oxygen and carbon-nitrogen positions” (Pauling et al, 1951). In their paper on deviations from planarity for omega, MacArthur & Thornton (1996) analysed small molecules in the Cambridge Structural Database (CSD) and structures in the PDB and find that for trans-peptide bonds with standard amino acids there is a mean omega of 179.7° with a standard deviation of 5.9° and 179.6° and 4.7° respectively. A selection of non-planar peptide bonds are given in Table 6.2, where numerical examples have been given in the literature. A confident identification of non-planar omega has the potential to identify important sites in the protein.

Structure	Residues	Dihedral	Comment
1r6j	Ser261-Gly262	162.2°	Overall mean omega slightly low
1ucs	Met21-Met22	165°	Met21 HB with Asn8
1ucs	Gln44-Val45	168°	Gln44 HB with Lys61
1us0	Ser76-Lys77	167.4°	Associated with a short HB
2vb1	Trp62-Trp63	152.7°	Multiple conformations
6s2m	Thr80-Lys81	162.3°	Also has non-planar rings

Table 6.2-Non-planar peptide bond claims in papers, this shows some of the non-planar omegas angles that are specifically mentioned with residues and distances.

6.3) Backbone angles

Discussions of the backbone angle tau (N:CA:C) are less common than those around hydrogen bonds and omega, however mentions include:

- 1yk4 suggests tau correlates with psi and has a bimodal distribution of 109.2° and 113.5° associating with different psi values.
- 2vb1 suggests a wide distribution correlating with local conformation (which means with phi and psi).
- 4rek looks at the next backbone angle tau+1 (CA:C:N+1) and its correlation with C:O

6.4) C:O and C:N+1

Less commonly mentioned is a relationship between C:O and the peptide bond C:N+1, and the discussion around the electron sharing in the peptide bond.

- 1us0 finds that residual density is either in C:O or C:N+1 and they infer different bond types with the sharing of electrons.
- 4rek mentions C:O in relation to C:N+1 and they find that shorter C:O correlates with longer C:N+1 in secondary structures.
- Not in this study, Ahmed's PhD (2012) potentially shows a different character in the C:O bond depending on secondary structure (though it could be a statistical artefact or due to bfactor).

The very latest ultrahigh resolution structure 7vos (which was released after the search criteria for this study on 1st June 2022) is another iron-sulfur cluster containing structure from Japan (Hanazono et al, 2022) and they find a -0.56 correlation between C:O and C:N+1.

6.4) Protonation states

Protonation states are inferred from the presence of hydrogens in the difference density – mostly for histidine but also for aspartate.

- 1us0 shows the protonation state of His110.
- 3bcj has evidence of protonation state near the amide ring.
- 4i8h shows the protonation state of His57 (Figure 5.4).
- 4ua6 shows evidence of side chain protonation.
- 5mn1 shows protonation using neutron diffraction to compare.
- 6l27 shows protonation for Glu97.

6.5) Electron and proton transfer

By analysing multiple conformations and/or hydrogen bonding networks and protonation states some papers infer mechanistic function.

- 2dsx suggests electron transfer via the iron centre and through the aromatic rings.
- 2vb1 has multiple hydrogen bond conformations which suggest functional features.
- 3a38 has an iron-sulfur electron transfer cluster.
- 4ua6 shows evidence for proton transfer, possibly linked to the LBHB.
- 6l27 claims proton transfer around Glu222 from the multiple conformation hydrogen bond network.

6.6) Topology analysis

Methods of numerically analysing the electron density topology for derivatives based on gradient and Laplacian maps to find features such as atom boundaries and bond types were made by: 4rek, 5gv8, 6jgj (Bader, 1994; Johnson et al, 2010). These methods have potential to make future discoveries about interactions.

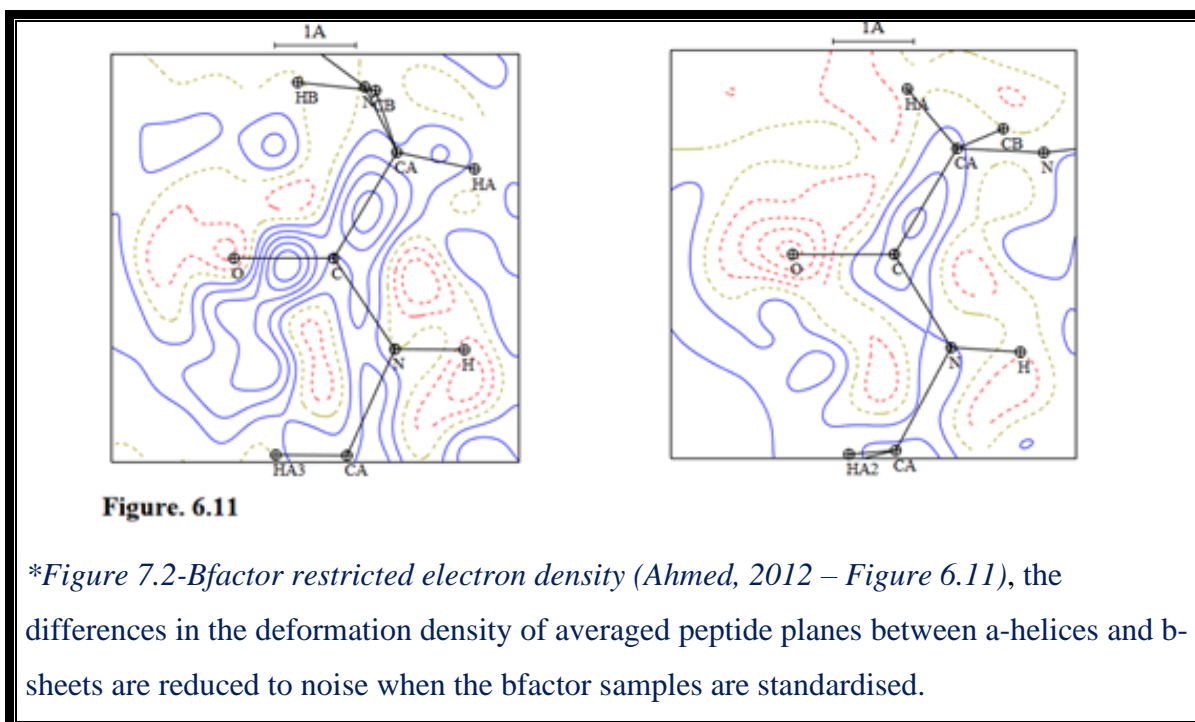
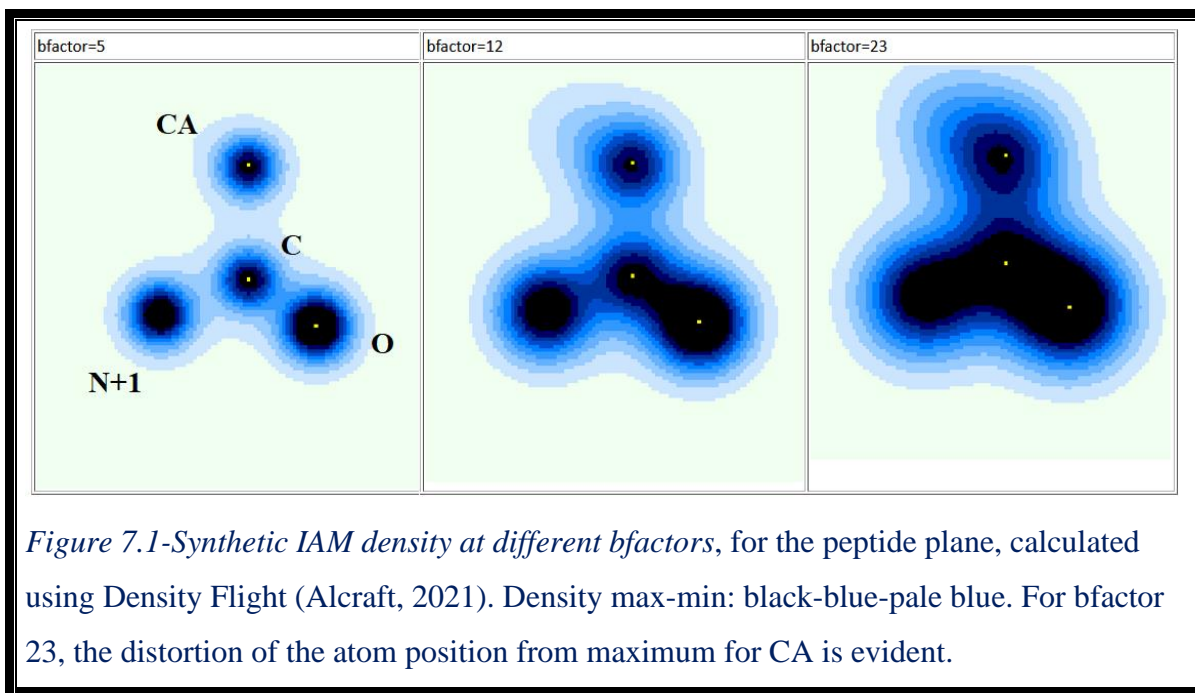
7) An analysis of the claims

In the 2010 textbook *Crystal Structure Analysis*, Glusker and Trueblood described the techniques of charge density analysis and deformation density, particularly in conjunction with neutron diffraction, saying “The future of this area is bright” (Glusker & Trueblood, 2010). In 2000 the structure 1ejg (Jelsch et al, 2000) was ground-breaking in its techniques, yielding 180 citations, but 15 years later, the same group published the 4rek structure, a much larger protein, (Zarychta et al, 2015) with the addition of topology analysis, but otherwise not building on the techniques for interpretation of the electron density. They use a technique of averaging electron density of equivalent chemical groups that no other group appears to have used, we have seen 1ejg and 4rek from the same group in this study, and additionally Housset et al (1999) – associated with the same group. An analysis follows in which I have picked out some points to consider based on the claims made in all these papers and a wider survey of opinion to understand the progress of research in this area.

7.1) Statistical quality

The underlying basis of all these claims relies on the quality of the electron density. In some of these papers some doubt is cast on the quality of the electron density which limits their analysis. For structure 1ejg “the signal to noise ratio of the crambin residual map was increased by averaging” (Jelsch et al, 2000). A later PhD student in this same group used the averaging method in his thesis (Ahmed, 2012) however despite the compelling images that this produces for a comparison of an averaged peptide bond for α -helices and β -sheets (Figure 5.5), he concludes that these results are not meaningful due to different b-factors for the two subsets of peptide groups. His concern about the different b-factor populations can be demonstrated theoretically to be justifiable - when calculating the electron density from IAM for a purely hypothetical structure, and comparing b-factors, it is certainly the case that the b-factor dramatically distorts the C:O bond - Figure 7.1 shows this for a synthetic peptide plane.

Ahmed then restricts his samples to the same b-factor range in which case the evidence is inconclusive as the difference between the deformation densities is indistinguishable from noise. He suggests that the ability to detect the difference in electron density between bonds for different secondary structures is at the limit of what can be observed, and his own initial results could be an artefact (Ahmed, 2012 – see chapter 6), reproduced here as Figure 7.2.



It is difficult to ascertain more information from Ahmed's samples - when he restricts the populations of secondary structure samples to the same bfactor he is using non-representative samples. It is not the lack of evidence in this comparison that suggests that it is hard to know the difference, but the difficulty we have in comparing samples with different b-factors as the bfactor distorts the evidence so we cannot compare the true populations.

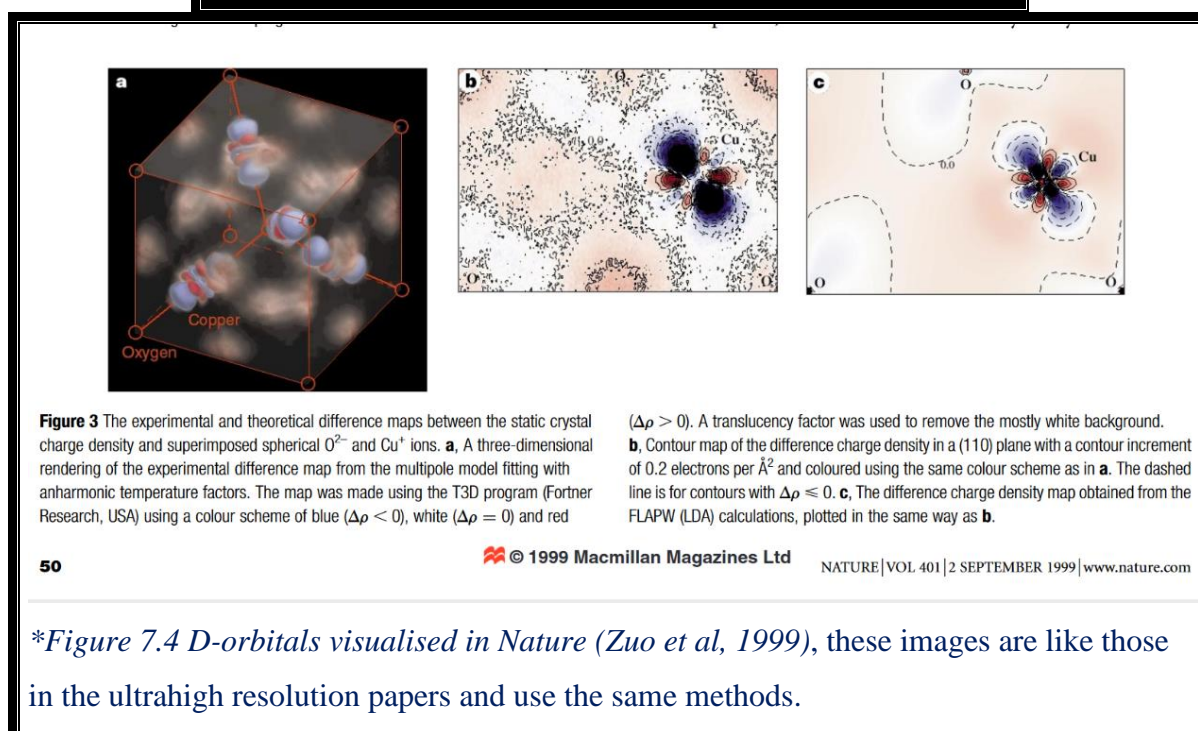
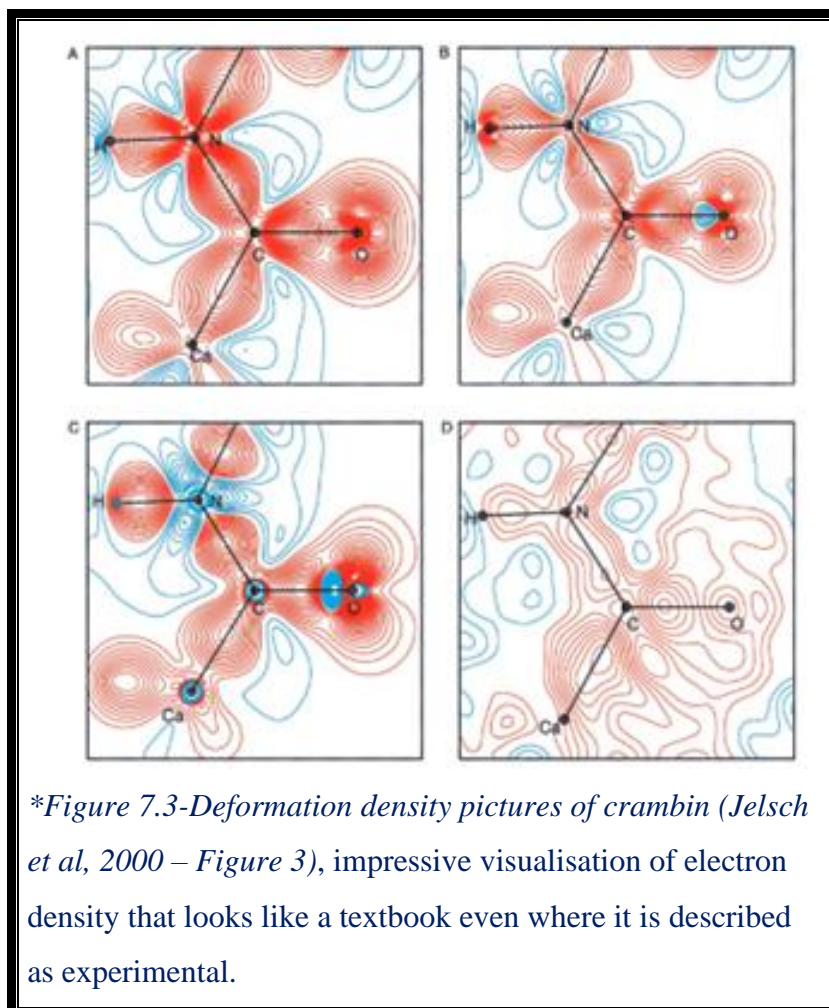
The statistical quality of electron density is analysed by Ian Tickle (2012) in terms of accuracy or precision: one using the difference density for accuracy and one only the observed density for precision in signal-to-noise ratio. These calculations are implemented in EDSTATS which is used in the solution of some ultrahigh resolution structures (e.g. Ramos et al, 2021) but not by any in this particular group of $<0.8\text{\AA}$ papers. Metrics for both accuracy and precision would be of interest in these structures. Outside the world of protein biology, there is discussion on the validity of analysing differences between probability distributions at all. A team working on machine learning applications to statistical data analysis (Sugiyama et al, 2012) discuss how a two-stage approach to estimating density differences has the problem of errors magnifying in the second stage, they propose a single-stage difference estimation – a least squares density-difference. This is dealing with a different format of data as here the difference data is calculated from structure factors – nevertheless it is effectively a difference between 2 density distributions and their insights could apply.

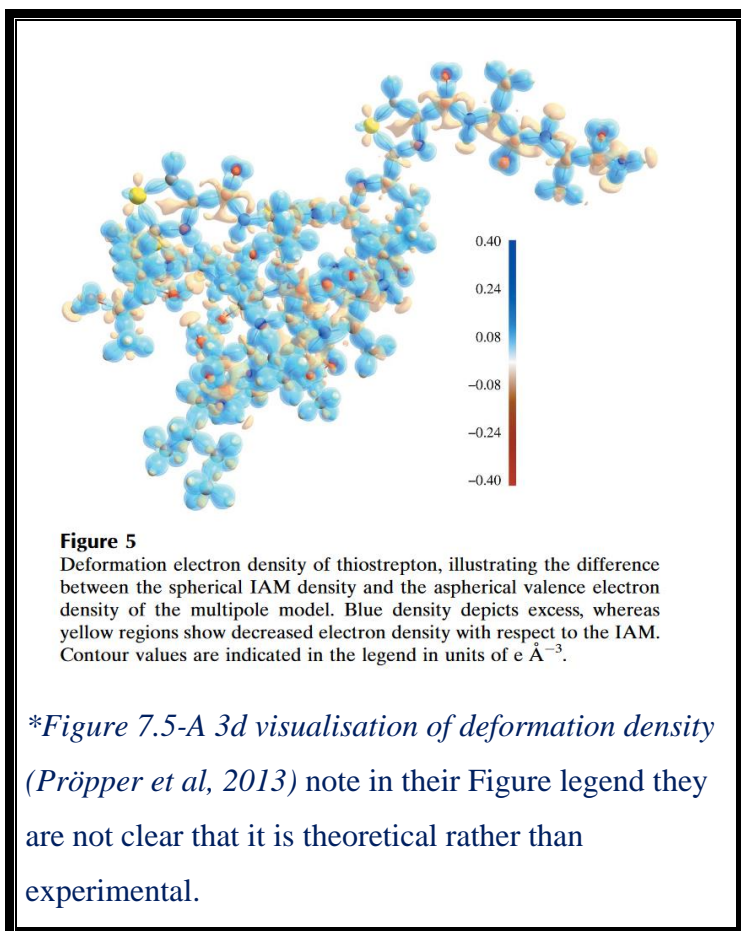
If we could quantify the signal-to-noise ratio, the ability to infer more from the electron density maps by averaging is appealing. We would want to establish that the averaging had made that improvement quantifiably. The suggestion by Ahmed (2012) that the limits of interpretability had been reached may have been off putting to this method, but this comment has wider implications – what can we tell at a noisier individual sample level if we can't infer anything above the noise at an average? Can we really see individual valence electrons and hydrogens or are they indistinguishable from noise?

7.2) Model bias

In the above we talk about difference density, and some of the maps and average maps in the papers show some quite striking images of difference density that look almost like text-book pictures of orbitals. Figure 7.3 shows the deformation pictures of crambin, both theoretical and experimental. The experimental static images look like a textbook.

Orbital claims have not been made in any of these papers but a small molecule paper that uses a similar technique of multipole method and difference density published in Nature came close with the title “Direct observation of d-orbital holes and Cu-Cu bonding in Cu₂O” (Zuo et al, 1999) which I have replicated in Figure 7.4.





This caused a storm, with Scerri (2000) writing a letter that said “Text-book orbitals will never be observed”, with the original authors responding “We entirely agree with Scerri that orbital wave functions are unobservable and that all the indistinguishable electrons in a crystal define a single quantum state” continuing with “it would be perverse not to mention the undoubted resemblance of our final result to the simple one-electron model of a d-orbital charge-density distribution” (Zuo et al, 2001). This is important to highlight,

because it would be odd not to wonder at it – it is not mentioned in any of the high-resolution papers in this study yet many of them show images similar to the d-orbital image: Jeslch et al 2000; Zarrychta et al 2015; Takaba et al, 2015; Pröpper et al, 2013, Hirano et al 2016. It can be difficult to discern what is being shown, often they are a “static deformation map” which is explained as being “static” because the anisotropic displacement parameters are not considered (Pröpper et al, 2013). It can be unclear when looking at these images whether you are looking at a model or at experimental data; one example is in Pröpper et al (2013) (see Figure 7.5). It is a striking image of the entire 3d structure with visible carbonyl lone pairs. The figure description says it is deformation density but, in the text, they advise that a Bader-AIM analysis was not appropriate because the electron density was derived from theoretical calculations and so it would not provide additional information. So: purely a model, but quite obscurely explained.

The process of x-ray crystallography refinement is one in which a model is developed and then refined iteratively until the model matches the experimental data near enough. This leaves an inevitable problem of model-bias in the results which can end up imprinted in what should be the experimental evidence. Richard Henderson says, “one must not underestimate the ingenuity

of humans to invent new ways to deceive themselves” (Henderson, 2013) referring to an image of Einstein being constructed out of noise in cryo-EM data. When being concerned about model bias I refer to the well-known model bias in crystallography from using a model to calculate phases which leads to features of the model being present in the resulting density map even if they are not there (Terwilliger et al, 2008). But I also intend it to mean any other model biases that may result from methods used to iterate towards a solution in the manner of the Einstein result. I am not aware of a metric to measure model bias of either kind.

7.3) Ring artefacts

An important issue raised is that from the structure 1yk4 in which they mention the visible evidence of rings in the electron density caused by Fourier truncation (Figure 5.3). This is discussed in a paper by Afonine et al (2004) in which they suggest that when the observed and calculated density balance out the problem is resolved – i.e. in the Fo-Fc map.

However, 2Fo-Fc density is used in much analysis, including topology analysis. Truncation artefacts are perhaps problematic for the topology analyses performed by 4rek, 5gv8 (which finds an inconsistency in comparison with multipole) and 6jgj (which finds single bonds where double are expected) all of which I have verified to contain ring artefacts. Further exploration is needed to understand how these topology analyses use the electron density, but where the rings problem exists the unbalanced density contains artefacts, so that must be corrected for topology analysis to be successful. Methods must exist for this - the rings problem is known – however given all these high-resolution structures present with ring artefacts the present methods are not sufficient, or not being used. Either way, a systematic correction of the ring artefacts is necessary for electron density analysis.

A very different method of refinement to any so far discussed makes use of the rings. It was introduced in 2015 and is based on image definition evaluation functions. It uses the Fourier truncation rings, to create a solution for the phases through machine learning (Li et al, 2015). They refer to it as a kind of “focus” because the rings come into focus when the phases are correct, the method is in its infancy but has an appeal as it is a model free method and could make an interesting comparison if used as an alternative refinement method on any of these structures.

7.4) Atom Positions

There have been systematic analyses of the geometry of atomic resolution structures, for example a study that looked at whether any change should be made to the **target refinement parameters** based on ultrahigh resolution structures (Jaskolski et al, 2007) and suggestions that there is an improvement in, for example, the chi values at higher resolution (Morris et al, 1992) - although I observe that care must be taken over the **rarity effect** – the possibility that a convergence of values to an ideal is only due to rarer sampling (Density Flight, Alcraft, 2020). Confidence in these analyses relies on the higher confidence of the atomic position from electron density evidence and the ability to relax geometric restraints in refinement. The atom position is generally believed to be at the maxima of the electron density (Bader, 1994; Johnson et al, 2010) – although not for hydrogens due to the asymmetry of the distribution of the single shared electron. I have shown that the atomic positions are not always at the electron density maxima in my analysis (see the adjusted tau / C:O plot in Figure 5.7) and that b-factors distort electron density in theoretical IAM models (Figure 7.1). Atom position is not the direct experimental result of x-ray crystallography experiments, but rather electron density. Some further understanding of this is warranted: electron density is considered the final arbiter of atom evidence (Wlodawer, 2017) but how far can the electron density maximum stray from the atom position and under what circumstances? How do these circumstances influence the atom position and the geometry and to what extent and accuracy? Movement, occupancy, signal-to-noise, experimental quality: many factors will influence whether there is a true observation of omega at 152.7° (2vb1) including whether the atom positions are at the density maxima.

7.5) Hydrogen Bonds and geometry

The presence of so many short hydrogen bonds and non-standard hydrogen bonds suggests a systematic and statistical analysis of ultrahigh-resolution structures would be interesting. Each solved structure is an individual story, an anecdote, but all together they may provide a wider picture of stereochemical and functional interest.

A widespread analysis of the peptide plane would be interesting, along with any links to function and other features as suggested in these papers. Other geometric features mentioned by several authors are those concerning correlations between C:O, C:N+1 and tau, suggesting

further analysis of these could be warranted. The finding of a bimodal distribution of tau with psi in 1yk4 (Bönisch et al, 2005) is interesting with regards to literature on psi and whether it is bimodal (Jaskolski et al, 2007) and whether it is correlated with local conformation (Balasco et al, 2017). We have seen that the correlation between C:O and tau+1 changes substantially when atom position is replaced with electron density maxima (Figure 5.7), so it could be interesting to review claims about outlier geometrical positions from this perspective.

8) Next Steps

Points of interest that could be further analysed that stand out from the literature include:

8.1) Statistical quality and model bias

There are questions about the statistical quality of the maps, both from the perspective of metrics used to identify accuracy (Tickle, 2012) and from questions raised about individual peptide bond samples (Jelsch et al, 2010) and the possibility of averaging maps including those of different secondary structures (Ahmed, 2012). We have also seen that there are impacts of model bias that are not understood and not measurable. There are methods to reduce model bias such as the maximum entropy method (Smaalen & Netzel, 2009) and iterative omit maps (Terwilliger et al, 2008); there are methods that infer phases directly although mostly only applicable to small molecules due to compute time (Woolfson, 1987) and there is also potentially a method to refine phases directly using machine learning and the rings artefact (Li et al, 2015).

Action: For deeper analysis of electron density including comparative analysis across structures a re-refinement of density would be needed, including metrics for statistical quality and model bias.

8.2) Electron density information from averaging

Averaging maps has the potential to be a powerful tool. With the newly refined electron densities (proposed above) it may be possible to gather enough samples by using multiple structures to gain bonding insights into new chemical groups other than the peptide bond (which can be averaged using a single structure due to the large number of repeats and its near planarity). Uncertainty about the impact of bfactor on the electron density would need to be addressed. There are no standard methods for averaging maps with only one group using the approach in this study

Action: Investigate methods for averaging electron density. Investigate averaging over multiple structures, with an understanding of bfactor influence on both effect and sampling.

8.3) Hydrogen bonds and geometry

The hydrogen bonds that have been mentioned are fascinating and would warrant a systematic analysis of ultrahigh resolution structures to find more examples and investigate the environment in which they are found.

Action: Widescale analysis of extreme geometry and the environment. This analysis could extend to a systematic direct analysis of hydrogens and hydrogen bonds in difference density and density to find the extent of experimental evidence.

8.4) Atom position

The geometry is tied to atom position and an understanding of the placing of atom position wrt the experimental evidence is needed – post refinement how do we justify an atom position?

Action: Widescale analysis of atom position and experimental evidence in terms of the electron density, looking at statistical quality and position uncertainty, density volumes, intensities, and electron density topology. A quantifiable justification of the atom position would inform the extreme geometry analysis.

8.5) Topology and ring artefacts

Hydrogen bonds and other interactions have also been identified by topology analysis of electron density from first and second derivatives. The problem with ring artefacts would seem to make this unreliable, so the removal of ring artefacts would be necessary for this analysis. Current methods of density modification could be implemented, or some new method.

Action: Remove ring artefacts and analyse electron density – are the same hydrogen bonds and other interactions identified as through geometry?

9) Conclusion

This analysis of the top $<0.8\text{\AA}$ structures papers has proved unexpectedly fascinating as the claims and features of the structures have unravelled chronologically. The individual stories that are pursued in each paper purport to provide experimental evidence for claims that give insight and support for observations that may have been made theoretically or in lower resolution structures without strong evidence. These ideas together provide a broader picture of outstanding questions in ultrahigh-resolution macromolecular structures. Implementing the actions suggested could facilitate insight into some important questions.

References

- Afonine, P.V., Facts About Maps. (2010). Available online at http://phenix-online.org/phenixwebsite_static/mainsite/files/presentations/latest/pavel_maps.pdf
- Afonine, P., Pichon-Pesme, V., Muzet, N., Jelsch, C., Lecomte, C., & Urzhumtsev, A. (2002). *Modelling of bond electron density by Gaussian scatters at subatomic resolution*. 10. Available online at http://legacy.ccp4.ac.uk/newsletters/newsletter41/07_elect_scatoct02/afonurzh.pdf
- Afonine, P. V., Lunin, V. Y., Muzet, N., & Urzhumtsev, A. (2004). On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps. *Acta Crystallographica Section D Biological Crystallography*, 60(2), 260–274. <https://doi.org/10.1107/S0907444903026209>
- Ahmed, M. (2012). Ultra high resolution crystallography of small molecules and proteins. *Phd Thesis, Université de Lorraine*. <https://hal.univ-lorraine.fr/tel-01749192>
- Alcraft, R.S. (2020). *What new things can we learn from ultrahigh-resolution protein crystal structures?* MSc Thesis, Birkbeck College, University of London.
- Alcraft, R.S. (2021). Density Flight (software application).
- Alcraft, R.S. (2021). LeucipPy (python library).
- Bader, R. F. W. (1994). *Atoms in molecules: A quantum theory*. Clarendon Press ; Oxford University Press.
- Balasco, N., Esposito, L., & Vitagliano, L. (2017). Factors affecting the amplitude of the τ angle in proteins: A revisitation. *Acta Crystallographica Section D Structural Biology*, 73(7), 618–625. <https://doi.org/10.1107/S2059798317007793>
- Barone, M., Müller, M., Chiha, S., Ren, J., Albat, D., Soicke, A., Dohmen, S., Klein, M., Bruns, J., van Dinther, M., Opitz, R., Lindemann, P., Beerbaum, M., Motzny, K., Roske, Y.,

- Schmieder, P., Volkmer, R., Nazaré, M., Heinemann, U., ... Kühne, R. (2020). Designed nanomolar small-molecule inhibitors of Ena/VASP EVH1 interaction impair invasion and extravasation of breast cancer cells. *Proceedings of the National Academy of Sciences*, *117*(47), 29684–29690. <https://doi.org/10.1073/pnas.2007213117>
- Bönisch, H., Schmidt, C. L., Bianco, P., & Ladenstein, R. (2005). Ultrahigh-resolution study on *Pyrococcus abyssi* rubredoxin. I. 0.69 Å X-ray structure of mutant W4L/R5S. *Acta Crystallographica Section D Biological Crystallography*, *61*(7), 990–1004. <https://doi.org/10.1107/S0907444490501293X>
- Cameron, A. J., Squire, C. J., Edwards, P. J. B., Harjes, E., & Sarojini, V. (2017). Crystal and NMR Structures of a Peptidomimetic β -Turn That Provides Facile Synthesis of 13-Membered Cyclic Tetrapeptides. *Chemistry - An Asian Journal*, *12*(24), 3195–3202. <https://doi.org/10.1002/asia.201701422>
- Cameron, A. J., Squire, C. J., Gérenton, A., Stubbing, L. A., Harris, P. W. R., & Brimble, M. A. (2019). Investigations of the key macrolactamisation step in the synthesis of cyclic tetrapeptide pseudoxyllallemycin A. *Organic & Biomolecular Chemistry*, *17*(16), 3902–3913. <https://doi.org/10.1039/C9OB00227H>
- Chen, C.-J., Lin, Y.-H., Huang, Y.-C., & Liu, M.-Y. (2006). Crystal structure of rubredoxin from *Desulfovibrio gigas* to ultra-high 0.68Å resolution. *Biochemical and Biophysical Research Communications*, *349*(1), 79–90. <https://doi.org/10.1016/j.bbrc.2006.07.205>
- Crick, F. H. C., & Kendrew, J. C. (1957). X-Ray Analysis and Protein Structure. In *Advances in Protein Chemistry* (Vol. 12, pp. 133–214). Elsevier. [https://doi.org/10.1016/S0065-3233\(08\)60116-3](https://doi.org/10.1016/S0065-3233(08)60116-3)
- Cowtan, K. (2003). Phase Problem in X-ray Crystallography, and Its Solution. In John Wiley & Sons, Ltd (Ed.), *ELS* (1st ed.). Wiley. <https://doi.org/10.1038/npg.els.0002722>

- Dittrich, B., Koritsánszky, T., & Luger, P. (2004). A Simple Approach to Nonspherical Electron Densities by Using Invariants. *Angewandte Chemie International Edition*, 43(20), 2718–2721. <https://doi.org/10.1002/anie.200353596>
- Dutta, S., Zardecki, C., Goodsell, D. S., & Berman, H. M. (2010). Promoting a structural view of biology for varied audiences: An overview of RCSB PDB resources and experiences. *Journal of Applied Crystallography*, 43(5), 1224–1229. <https://doi.org/10.1107/S002188981002371X>
- Engh, R. A., & Huber, R. (2006). 18.3. Structure quality and target parameters. *International Tables for Crystallography, F*, 382–392. <http://dx.doi.org/10.1107/97809553602060000695>
- Esposito, L., Vitagliano, L., Sica, F., Sorrentino, G., Zagari, A., & Mazzarella, L. (2000). The ultrahigh resolution crystal structure of ribonuclease A containing an isoaspartyl residue: Hydration and stereochemical analysis. *Journal of Molecular Biology*, 297(3), 713–732. <https://doi.org/10.1006/jmbi.2000.3597>
- Fanfrlík, J., Kolář, M., Kamlar, M., Hurný, D., Ruiz, F. X., Cousido-Siah, A., Mitschler, A., Řezáč, J., Munusamy, E., Lepšík, M., Matějček, P., Veselý, J., Podjarný, A., & Hobza, P. (2013). Modulation of Aldose Reductase Inhibition by Halogen Bond Tuning. *ACS Chemical Biology*, 8(11), 2484–2492. <https://doi.org/10.1021/cb400526n>
- Flocco, Maria M., and Sherry L. Mowbray. ‘Planar Stacking Interactions of Arginine and Aromatic Side-Chains in Proteins’. *Journal of Molecular Biology* 235, no. 2 (January 1994): 709–17. <https://doi.org/10.1006/jmbi.1994.1022>.
- Glusker, J. P., & Trueblood, K. N. (2010). *Crystal Structure Analysis: A Primer* (3rd ed.). Oxford University Press. <https://doi.org/10.1093/oso/9780199576340.001.0001>

- Hakanpää, J., Linder, M., Popov, A., Schmidt, A., & Rouvinen, J. (2006). Hydrophobin HFBII in detail: Ultrahigh-resolution structure at 0.75 Å. *Acta Crystallographica Section D Biological Crystallography*, 62(4), 356–367. <https://doi.org/10.1107/S0907444906000862>
- Hanazono, Y., Hirano, Y., Takeda, K., Kusaka, K., Tamada, T., & Miki, K. (2022). Revisiting the concept of peptide bond planarity in an iron-sulfur protein by neutron structure analysis. *Science Advances*, 8(20), eabn2276. <https://doi.org/10.1126/sciadv.abn2276>
- Hansen, N. K., & Coppens, P. (1978). Testing aspherical atom refinements on small-molecule data sets. *Acta Crystallographica Section A*, 34(6), 909–921. <https://doi.org/10.1107/S0567739478001886>
- Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45), 18037–18041. <https://doi.org/10.1073/pnas.1314449110>
- Hirano, Y., Kimura, S., & Tamada, T. (2015). High-resolution crystal structures of the solubilized domain of porcine cytochrome *b*₅. *Acta Crystallographica Section D Biological Crystallography*, 71(7), 1572–1581. <https://doi.org/10.1107/S1399004715009438>
- Hirano, Y., Takeda, K., & Miki, K. (2016). Charge-density analysis of an iron–sulfur protein at an ultra-high resolution of 0.48 Å. *Nature*, 534(7606), 281–284. <https://doi.org/10.1038/nature18001>
- Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A., & Podjarny, A. (2004). Ultrahigh resolution drug design I: Details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. *Proteins: Structure, Function, and Bioinformatics*, 55(4), 792–804. <https://doi.org/10.1002/prot.20015>

- Housset, D., Pichon, V., Jelsch, C., Maierhofer, A., Fontecilla, J. C., & Lecomte, C. (2000). *Towards the charge-density study of proteins: A room-temperature scorpion-toxin structure at 0.96 Å resolution as a first test case*. 10.
- Isogai, Y., Imamura, H., Nakae, S., Sumi, T., Takahashi, K., Nakagawa, T., Tsuneshige, A., & Shirai, T. (2018). Tracing whale myoglobin evolution by resurrecting ancient proteins. *Scientific Reports*, 8(1), 16883. <https://doi.org/10.1038/s41598-018-34984-6>
- Jaskolski, M., Gilski, M., Dauter, Z., & Wlodawer, A. (2007). Stereochemical restraints revisited: How accurate are refinement targets and how much should protein structures be allowed to deviate from them? *Acta Crystallographica Section D Biological Crystallography*, 63(5), 611–620. <https://doi.org/10.1107/S090744490700978X>
- Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H., & Lecomte, C. (2000). Accurate protein crystallography at ultra-high resolution: Valence electron distribution in crambin. *Proceedings of the National Academy of Sciences*, 97(7), 3171–3176. <https://doi.org/10.1073/pnas.97.7.3171>
- Johnson, E. R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A. J., & Yang, W. (2010). Revealing Noncovalent Interactions. *Journal of the American Chemical Society*, 132(18), 6498–6506. <https://doi.org/10.1021/ja100936w>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

- Kang, B. S., Devedjiev, Y., Derewenda, U., & Derewenda, Z. S. (2004). The PDZ2 Domain of Syntenin at Ultra-high Resolution: Bridging the Gap Between Macromolecular and Small Molecule Crystallography. *Journal of Molecular Biology*, 338(3), 483–493.
<https://doi.org/10.1016/j.jmb.2004.02.057>
- Ko, T.-P., Robinson, H., Gao, Y.-G., Cheng, C.-H. C., DeVries, A. L., & Wang, A. H.-J. (2003). The Refined Crystal Structure of an Eel Pout Type III Antifreeze Protein RD1 at 0.62-Å Resolution Reveals Structural Microheterogeneity of Protein and Solvation. *Biophysical Journal*, 84(2), 1228–1237. [https://doi.org/10.1016/S0006-3495\(03\)74938-8](https://doi.org/10.1016/S0006-3495(03)74938-8)
- Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M., & Bott, R. (1998). The 0.78 Å Structure of a Serine Protease: *Bacillus lentus* Subtilisin . *Biochemistry*, 37(39), 13446–13452. <https://doi.org/10.1021/bi9813983>
- Laulumaa, S., & Kursula, P. (2019). Sub-Atomic Resolution Crystal Structures Reveal Conserved Geometric Outliers at Functional Sites. *Molecules*, 24(17), 3044.
<https://doi.org/10.3390/molecules24173044>
- LeMaster, D. M., Anderson, J. S., Wang, L., Guo, Y., Li, H., & Hernández, G. (2007). NMR and X-ray analysis of structural additivity in metal binding site-swapped hybrids of rubredoxin. *BMC Structural Biology*, 7(1), 81. <https://doi.org/10.1186/1472-6807-7-81>
- Li, H., He, M., & Zhang, Z. (2015). Image definition evaluation functions for X-ray crystallography: A new perspective on the phase problem. *Acta Crystallographica Section A Foundations and Advances*, 71(5), 526–533. <https://doi.org/10.1107/S2053273315012103>
- Liebschner, D., Dauter, M., Brzuszkiewicz, A., & Dauter, Z. (2013). On the reproducibility of protein crystal structures: Five atomic resolution structures of trypsin. *Acta Crystallographica Section D Biological Crystallography*, 69(8), 1447–1462.
<https://doi.org/10.1107/S0907444913009050>

- Lonsdale, K. (1931). *An X-ray Analysis of the Structure of Hexachlorobenzene, using the Fourier Method*. Royal Society. <https://royalsocietypublishing.org/doi/pdf/10.1098/rspa.1931.0166>
- MacArthur, M. W., & Thornton, J. M. (1996). Deviations from Planarity of the Peptide Bond in Peptides and Proteins. *Journal of Molecular Biology*, 264(5), 1180–1195. <https://doi.org/10.1006/jmbi.1996.0705>
- Malinska, M., Dauter, M., Kowiel, M., Jaskolski, M., & Dauter, Z. (2015). Protonation and geometry of histidine rings. *Acta Crystallographica Section D Biological Crystallography*, 71(7), 1444–1454. <https://doi.org/10.1107/S1399004715007816>
- Minichino, A., Habash, J., Raftery, J., & Helliwell, J. R. (2003). The properties of $(2F_o - F_c)$ and $(F_o - F_c)$ electron-density maps at medium-to-high resolutions. *Acta Crystallographica Section D Biological Crystallography*, 59(5), 843–849. <https://doi.org/10.1107/S0907444903004219>
- Moroz, O. V., Blagova, E., Taylor, E., Turkenburg, J. P., Skov, L. K., Gippert, G. P., Schnorr, K. M., Ming, L., Ye, L., Klausen, M., Cohn, M. T., Schmidt, E. G. W., Nymand-Grarup, S., Davies, G. J., & Wilson, K. S. (2021). Fungal GH25 muramidases: New family members with applications in animal nutrition and a crystal structure at 0.78Å resolution. *PLOS ONE*, 16(3), e0248190. <https://doi.org/10.1371/journal.pone.0248190>
- Morris, A. L., MacArthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins: Structure, Function, and Genetics*, 12(4), 345–364. <https://doi.org/10.1002/prot.340120407>
- Mulligan, V. K., Kang, C. S., Sawaya, M. R., Rettie, S., Li, X., Antselovich, I., Craven, T. W., Watkins, A. M., Labonte, J. W., DiMaio, F., Yeates, T. O., & Baker, D. (2020). Computational design of mixed chirality peptide macrocycles with internal symmetry. *Protein Science*, 29(12), 2433–2445. <https://doi.org/10.1002/pro.3974>

- Muñoz-Escobar, J., Matta-Camacho, E., Cho, C., Kozlov, G., & Gehring, K. (2017). Bound Waters Mediate Binding of Diverse Substrates to a Ubiquitin Ligase. *Structure*, 25(5), 719-729.e3. <https://doi.org/10.1016/j.str.2017.03.004>
- Nakamura, A., Ishida, T., Kusaka, K., Yamada, T., Fushinobu, S., Tanaka, I., Kaneko, S., Ohta, K., Tanaka, H., Inaka, K., Higuchi, Y., Niimura, N., Samejima, M., & Igarashi, K. (2015). “Newton’s cradle” proton relay with amide–imidic acid tautomerization in inverting cellulase visualized by neutron crystallography. *Science Advances*, 1(7), e1500263. <https://doi.org/10.1126/sciadv.1500263>
- Nichols, D. A., Hargis, J. C., Sanishvili, R., Jaishankar, P., Defrees, K., Smith, E. W., Wang, K. K., Prati, F., Renslo, A. R., Woodcock, H. L., & Chen, Y. (2015). Ligand-Induced Proton Transfer and Low-Barrier Hydrogen Bond Revealed by X-ray Crystallography. *Journal of the American Chemical Society*, 137(25), 8086–8095. <https://doi.org/10.1021/jacs.5b00749>
- Patterson, W. R., Anderson, D. H., Degrado, W. F., Cascio, D., & Eisenberg, D. (1999). Centrosymmetric bilayers in the 0.75 Å resolution structure of a designed alpha-helical peptide, D, L-Alpha-1. *Protein Science*, 8(7), 1410–1422. <https://doi.org/10.1110/ps.8.7.1410>
- Pauling, L., Corey, R.B., and Branson, H.R. (1951). The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. USA* 37, 205–211
- Pröpper, K., Holstein, J. J., Hübschle, C. B., Bond, C. S., & Dittrich, B. (2013). Invariom refinement of a new monoclinic solvate of thiostrepton at 0.64 Å resolution. *Acta Crystallographica Section D Biological Crystallography*, 69(8), 1530–1539. <https://doi.org/10.1107/S0907444913010664>
- Ramos, J., Laux, V., Haertlein, M., Boeri Erba, E., McAuley, K. E., Forsyth, V. T., Mossou, E., Larsen, S., & Langkilde, A. E. (2021). Structural insights into protein folding, stability and

activity using *in vivo* perdeuteration of hen egg-white lysozyme. *IUCrJ*, 8(3), 372–386.

<https://doi.org/10.1107/S2052252521001299>

Scerri, E. R. (2000). Have Orbitals Really Been Observed? *Journal of Chemical Education*, 77, 3.

Schiebel, J., Gaspari, R., Sandner, A., Ngo, K., Gerber, H.-D., Cavalli, A., Ostermann, A., Heine,

A., & Klebe, G. (2017). Charges Shift Protonation: Neutron Diffraction Reveals that Aniline and 2-Aminopyridine Become Protonated Upon Binding to Trypsin. *Angewandte Chemie International Edition*, 56(17), 4887–4890. <https://doi.org/10.1002/anie.201701038>

<https://doi.org/10.1002/anie.201701038>

Schmidt, A., Teeter, M., Weckert, E., & Lamzin, V. S. (2011). Crystal structure of small protein

crambin at 0.48 Å resolution. *Acta Crystallographica Section F Structural Biology and Crystallization Communications*, 67(4), 424–428.

<https://doi.org/10.1107/S1744309110052607>

Schuller, M., Correy, G. J., Gahbauer, S., Fearon, D., Wu, T., Díaz, R. E., Young, I. D., Carvalho

Martins, L., Smith, D. H., Schulze-Gahmen, U., Owens, T. W., Deshpande, I., Merz, G. E.,

Thwin, A. C., Biel, J. T., Peters, J. K., Moritz, M., Herrera, N., Kratochvil, H. T., ... Ahel, I.

(2021). Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking. *Science Advances*, 7(16), eabf8711.

<https://doi.org/10.1126/sciadv.abf8711>

Shatsky, M., Hall, R. J., Brenner, S. E., & Glaeser, R. M. (2009). A method for the alignment of

heterogeneous macromolecules from electron microscopy. *Journal of Structural Biology*,

166(1), 67–78. <https://doi.org/10.1016/j.jsb.2008.12.008>

Shibazaki, C., Shimizu, R., Kagotani, Y., Ostermann, A., Schrader, T. E., & Adachi, M. (2020).

Direct Observation of the Protonation States in the Mutant Green Fluorescent Protein. *The*

Journal of Physical Chemistry Letters, 11(2), 492–496.

<https://doi.org/10.1021/acs.jpcllett.9b03252>

- Smaalen, S. van, & Netzel, J. (2009). The maximum entropy method in accurate charge-density studies. *Physica Scripta*, 79(4), 048304. <https://doi.org/10.1088/0031-8949/79/04/048304>
- Stegmann, C. M., Seeliger, D., Sheldrick, G. M., de Groot, B. L., & Wahl, M. C. (2009). The Thermodynamic Influence of Trapped Water Molecules on a Protein-Ligand Interaction. *Angewandte Chemie International Edition*, 48(28), 5207–5210. <https://doi.org/10.1002/anie.200900481>
- Sugiyama, M., Kanamori, T., Suzuki, T., Plessis, M. C. du, Liu, S., & Takeuchi, I. (2012). *Density-Difference Estimation* (arXiv:1207.0099). arXiv. <http://arxiv.org/abs/1207.0099>
- Takaba, K., Tai, Y., Eki, H., Dao, H.-A., Hanazono, Y., Hasegawa, K., Miki, K., & Takeda, K. (2019). Subatomic resolution X-ray structures of green fluorescent protein. *IUCrJ*, 6(3), 387–400. <https://doi.org/10.1107/S205225251900246X>
- Takaba, K., Takeda, K., Kosugi, M., Tamada, T., & Miki, K. (2017). Distribution of valence electrons of the flavin cofactor in NADH-cytochrome b5 reductase. *Scientific Reports*, 7(1), 43162. <https://doi.org/10.1038/srep43162>
- Takeda, K., Kusumoto, K., Hirano, Y., & Miki, K. (2010). Detailed assessment of X-ray induced structural perturbation in a crystalline state protein. *Journal of Structural Biology*, 169(2), 135–144. <https://doi.org/10.1016/j.jsb.2009.09.012>
- Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Adams, P. D., Read, R. J., Zwart, P. H., & Hung, L.-W. (2008). Iterative-build OMIT maps: Map improvement by iterative model building and refinement without model bias. *Acta Crystallographica Section D Biological Crystallography*, 64(5), 515–524. <https://doi.org/10.1107/S0907444908004319>

- Tickle, I. J. (2012). Statistical quality indicators for electron-density maps. *Acta Crystallographica Section D Biological Crystallography*, 68(4), 454–467.
<https://doi.org/10.1107/S0907444911035918>
- Wang, J., Dauter, M., Alkire, R., Joachimiak, A., & Dauter, Z. (2007). Triclinic lysozyme at 0.65 Å resolution. *Acta Crystallographica Section D Biological Crystallography*, 63(12), 1254–1268. <https://doi.org/10.1107/S0907444907054224>
- Wlodawer, A., Dauter, Z., & Jaskolski, M. (Eds.). (2017). *Protein Crystallography: Methods and Protocols* (Vol. 1607). Springer New York. <https://doi.org/10.1007/978-1-4939-7000-1>
- Woolfson, M. M. (1987). Direct methods – from birth to maturity. *Acta Crystallographica Section A Foundations of Crystallography*, 43(5), 593–612.
<https://doi.org/10.1107/S0108767387098854>
- Yamada, M., Tamada, T., Takeda, K., Matsumoto, F., Ohno, H., Kosugi, M., Takaba, K., Shoyama, Y., Kimura, S., Kuroki, R., & Miki, K. (2013). Elucidations of the Catalytic Cycle of NADH-Cytochrome b5 Reductase by X-ray Crystallography: New Insights into Regulation of Efficient Electron Transfer. *Journal of Molecular Biology*, 425(22), 4295–4306. <https://doi.org/10.1016/j.jmb.2013.06.010>
- Zarychta, B., Lyubimov, A., Ahmed, M., Munshi, P., Guillot, B., Vrielink, A., & Jelsch, C. (2015). Cholesterol oxidase: Ultrahigh-resolution crystal structure and multipolar atom model-based analysis. *Acta Crystallographica Section D Biological Crystallography*, 71(4), 954–968. <https://doi.org/10.1107/S1399004715002382>
- Zhao, H.-T., Hazemann, I., Mitschler, A., Carbone, V., Joachimiak, A., Ginell, S., Podjarny, A., & El-Kabbani, O. (2008). Unusual Binding Mode of the 2 S 4 R Stereoisomer of the Potent Aldose Reductase Cyclic Imide Inhibitor Fidarestat (2 S 4 S) in the 15 K Crystal Structure of the Ternary Complex Refined at 0.78 Å Resolution: Implications for the Inhibition

Mechanism. *Journal of Medicinal Chemistry*, 51(5), 1478–1481.

<https://doi.org/10.1021/jm701514k>

Zuo, J. M., Kim, M., O’Keeffe, M., & Spence, J. C. H. (1999). Direct observation of d-orbital holes and Cu–Cu bonding in Cu₂O. *Nature*, 401(6748), 49–52. <https://doi.org/10.1038/43403>

Zuo, J. M., O’Keefe, M., & Spence, J. C. H. (2001). Have Orbitals Really Been Observed? *Journal of Chemical Education*, 78(7), 877. <https://doi.org/10.1021/ed078p877.2>