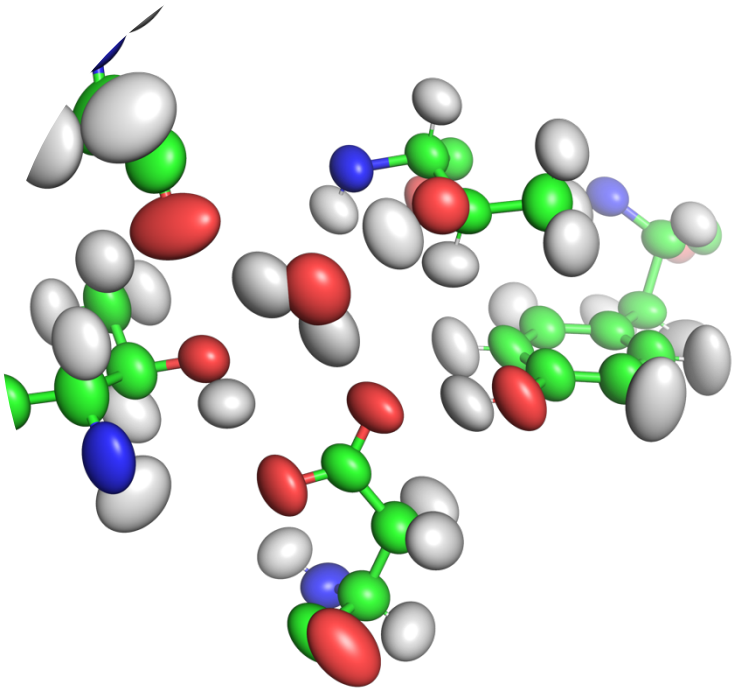
UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES



PhD Thesis

João Carlos Moreno Ramos

Protein and water dynamics at the atomic level

Supervisor: Annette Eva Langkilde

Co-supervisors: Estelle Mossou, V. Trevor Forsyth, Sine Larsen

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen on the 14th of April 2021

Protein and water dynamics at the atomic level

PhD thesis

April 2021

João Carlos Moreno Ramos

Life Sciences group, Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.

Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Jagtvej 162, 2100 København Ø, Denmark.

Email: ramosj@ill.fr; khv884@sund.ku.dk

Supervisors:

Associate Professor Annette Eva Langkilde (primary supervisor),

Department of Drug Design and Pharmacology, University of Copenhagen, Denmark.

Doctor Estelle Mossou (co-supervisor),

Institut Laue-Langevin, Grenoble, France; Keele University, Staffordshire, United Kingdom.

Professor V. Trevor Forsyth (co-supervisor),

Institut Laue-Langevin, Grenoble, France; Keele University, Staffordshire, United Kingdom.

Professor *emerita* Sine Larsen (co-supervisor),

Department of Chemistry, University of Copenhagen, Denmark.

Assessment Committee:

Professor Jette Sandholm Jensen Kastrup,

Department of Drug Design and Pharmacology, University of Copenhagen, Denmark.

Professor Poul Nissen,

Aarhus University, Denmark.

Professor *emeritus* John Helliwell,

University of Manchester, United Kingdom.

Submitted:

14th of April 2021

Cover image:

Ellipsoidal representation of anisotropic atomic displacement parameters in a region of perdeuterated hen egg-white lysozyme structure from neutron diffraction data. Carbon atoms colored in green, Oxygen in red, Nitrogen in blue and Deuterium in white.

Preface

This thesis dissertation was produced to comply with the requirements for obtaining a PhD degree from the Faculty of Health and Medical Sciences, University of Copenhagen, Denmark. The work developed is the result of a joint project between the Department of Drug Design and Pharmacology of the University of Copenhagen (Denmark) and the Diffraction and Life Sciences groups of the Institut Laue-Langevin, Grenoble (France). Associate Professor Annette Eva Langkilde was the primary supervisor and co-supervision was provided by Dr. Estelle Mossou, Professor V. Trevor Forsyth, and Professor *emerita* Sine Larsen.

This 3.5-year project was conducted primarily at the Institut Laue-Langevin, where most of the experimental work was carried out. Short stays at the University of Copenhagen occurred during the timeframe of the project, aimed at data processing and analysis, and at completing mandatory University courses required for obtaining the PhD degree. The PhD studentship was funded by the Institut Laue-Langevin.

The chief objective of this project was the determination of atomic displacement parameters of perdeuterated hen egg-white lysozyme from complete atomic resolution neutron diffraction data. These results, together with the parameters obtained from equivalent X-ray diffraction data, would allow comparative analysis aimed at exploring the potential bias of atomic motion derived from X-ray data. The work developed provided the means to obtain high quality neutron diffraction data, at room temperature and 100 K, and equivalent X-ray datasets. The findings described in this thesis provide novel and unique insight into protein structure and dynamics from neutron diffraction, and into the effects of deuteration and *in vitro* refolding in protein biophysical and structural properties. Preliminary analysis indicates differences between atomic motion obtained from the neutron and X-ray data.

The thesis dissertation is divided in 7 chapters. Chapter 1 serves as an introduction to the project, emphasizing the relevance of exploring atomic motion in protein crystals, and providing the appropriate background on neutron macromolecular crystallography, protein perdeuteration and the case study, hen egg-white lysozyme. Chapter 2 focuses on describing important aspects of the methodologies employed throughout the PhD work, giving some details on the directions taken experimentally. Chapter 3 to 6 describes manuscripts produced as a result of the work developed, reporting our main findings and giving an appropriate outline of the whole project. A short synopsis is provided before each manuscript to offer context in regard to the PhD project and describe the main motivations behind the work produced. Shortcomings are also presented for the analysis reported in the manuscripts. Finally, Chapter 7 describes the main conclusions from the results presented in the manuscripts and offers future perspectives with a focus on the analysis of atomic thermal motion in protein crystals from X-ray and neutron data.

Manuscripts included in this thesis

1. **Ramos, J.**, Laux, V., Haertlein, M., Boeri Erba, E., McAuley, K. E., Forsyth, V. T., Mossou, E., Larsen, S. & Langkilde, A. E. *Structural insights into protein folding, stability and activity using in vivo perdeuteration of hen egg-white lysozyme.* (2021). IUCrJ 8.
2. **Ramos, J.**, Laux, V., Haertlein, M., Forsyth, V. T., Mossou, E., Larsen, S. & Langkilde, A. E. *The impact of deuteration and in vitro refolding on the structural and biophysical properties of hen egg-white lysozyme*. (Manuscript in preparation).
3. **Ramos, J.**, Mossou, E., Laux, V., Mason, S., Lemee, M., Diederichs, K., Haertlein, M., Forsyth, V. T., Larsen, S. & Langkilde, A. E. *A complete picture of hen egg-white lysozyme active site at room temperature from 1 Å resolution neutron diffraction data*. (Manuscript in preparation).
4. **Ramos, J.**, Laux, V., Lemee, M., Bowler, M., Diederichs, K., Haertlein, M., Forsyth, V. T., Mossou, E., Larsen, S. & Langkilde, A. E. *Protein dynamics at the atomic level: comparing atomic displacement parameters from X-ray and neutron diffraction*. (Manuscript draft).

Acknowledgements

I would like to dedicate this work to my family and to Mathilde Folacci. Thank you for the unconditional and continued support. You are a great part of who I am and of my motivation to keep pushing to be better.

I'm grateful for the many friends and colleagues that I've met along the years. I've spent three and a half amazing years in Grenoble because I had the chance to interact with each one of you. I sincerely hope to stay in touch with many of you in the future and all I can do is to wish you all the best. I cannot also forget my friends from Portugal, which have always been there to welcome me back and to make me feel like I've never left.

Thank you to the supervisors and colleagues that were involved in my work from my bachelor's and master's in Biochemistry. It was then that I've found my passion for protein crystallography and that I learned a great deal about how to work in a laboratory and to conduct scientific research. A special thank you to Marino Santos, who was a great colleague and advised me during my application to the PhD position that I obtained. I'm also grateful to Francisco Leisico, who I've met in the research group in Portugal and coincidently came to Grenoble as well, being always a great friend and someone with whom I could always discuss crystallography.

Finally, I'm sincerely grateful to Dr. Annette Langkilde, Dr. Estelle Mossou, Prof. Trevor Forsyth and Prof. *emerita* Sine Larsen, for giving me the amazing opportunity of working in this PhD project. This opportunity was life changing and I believe that I became a better researcher and also a better person because of it. It was a privilege to work with you and I think that I couldn't have asked for better supervisors. I have to extend my gratitude to Valerie Laux and Dr. Michael Haertlein, who were essential to the success of this PhD project and a pleasure to work with. Also, I would like to thank all the collaborators of this project, which have contributed immensely to its success. Finally, I'm grateful to the Institut Laue-Langevin and the University of Copenhagen for the PhD studentship.

Abstract

Protein biological function is intimately related to structure as well as dynamics at the molecular and atomic scales. X-ray diffraction has contributed immensely to the understanding of protein function mainly through the description of structural properties. However, the study of atomic motion has been somewhat overlooked due to the limitations inherent to X-ray models and to the challenges faced in obtaining neutron diffraction data. For many years, in small-molecule crystallography, the use of multipole models to describe X-ray data has been employed to accurately model the electron density and characterize the nature of chemical bonding in molecules. At present, the accuracy of X-ray diffraction data from protein crystals is not at the level required to enable this type of refinements. In contrast, small-molecule studies have shown the capabilities of neutron diffraction in terms of providing accurate information on atomic and molecular motion. It is the aim of the research presented in this thesis to extend the potential of neutron diffraction to the study of proteins to get information of their structure and dynamics not biased by the limitations of the X-ray models.

The main aim of the present work is to push the boundaries of neutron macromolecular crystallography to obtain anisotropic atomic displacement parameters (ADPs) for perdeuterated hen egg-white lysozyme (D-HEWL), providing a basis for comparisons with X-ray derived ADPs.

For the success of this study, a strategy for the production of significant quantities of D-HEWL was developed, consisting in its overexpression in *E. coli* inclusion bodies, followed by protein purification and *in vitro* refolding. The refolded D-HEWL, and its hydrogenated variant, were characterized in terms of biophysical and structural properties. Both variants were shown to be thermally stable and active, and with an identical overall structural fold compared to the native unlabelled protein, from *Gallus gallus*. Minor variations in protein structure, were however observed, mainly in the Lys97-Gly104 region, related to the Asn103 peptide-plane flip. These changes were linked to slight decreases in protein thermal stability and enzymatic function.

Complete atomic resolution neutron diffraction data was successfully measured at room temperature (RT) and 100 K, for D-HEWL crystals grown at the pH where the enzyme is fully active. The neutron data were complemented by equivalent X-ray datasets obtained in the same conditions. The RT D-HEWL neutron structure elucidated the detailed configuration and dynamics of protein residues and water molecules in the enzyme's active site, in an active conformation. This dataset also provided insight into the potential of complete atomic resolution neutron data for a protein crystal, which has not been reported earlier. The D-HEWL X-ray and neutron datasets at RT and 100 K are the foundation for the present study of atomic motion. Preliminary results, from the D-HEWL RT data, suggest differences in the ADPs derived from the X-ray and neutron models. The X-ray derived ADPs appear to be larger and more isotropic that those obtained from neutrons. Additionally, the neutron model seems to describe more accurately structural disorder, affecting less the modelling of the ADPs. The neutron ADPs seem to contain biologically relevant information that is more physically realistic

than can be derived from the X-ray model. Finally, an outline is given of additional work in order to extract reliable ADPs from the 100 K neutron dataset, which could corroborate the observations made at RT.

Resumé (in Danish)

Proteiners biologiske funktion er tæt forbundet med deres struktur og den molekylære og atomare dynamik. Røntgendiffraktion har især gennem beskrivelse af proteiners strukturelle egenskaber bidraget enormt til vores forståelse af biologiske molekylers funktion. Undersøgelser af de atomare bevægelser er dog noget overset på grund af begrænsninger i de røntgenbaserede modeller og de udfordringer der er i indsamling af neutrondiffraktionsdata. Multipol-modeller er i en årrække blevet anvendt i små-molekyle Røntgenstudier til en nøjagtig modellering af elektrontætheden og efterfølgende karakterisering af kemiske bindinger i molekylerne. Men denne metode er dog ikke anvendelig til undersøgelser af proteiner, da det har vist sig umuligt at opnå tilstrækkeligt nøjagtige diffraktionsdata fra proteinkrystaller. Neutron diffraktion har til gengæld vist sig egnet til nøjagtig bestemmelse af atomare og molekylære bevægelser for mindre molekyler, giver mulighed for at få et uforstyrret billede af atomkernernes position og bevægelse.

Målet med dette projekt er at skubbe grænserne for makromolekylær neutronkrystallografi og bestemme anisotrope atomare forskydningsparametre (ADP'er) for perdeutereret lysozym fra hønse æggehvide (D-HEWL), og derved opnå basis for sammenligning med ADP'er bestemt via Røntgendiffraktion.

For at opnå dette, blev der først udviklet en strategi til fremstilling at større mængder D-HEWL, bestående af overekspression i *E. Coli* i såkaldte inclusion bodies, efterfulgt af oprensning og *in vitro* refoldning. De biofysiske og strukturelle egenskaber af refoldet D-HEWL og den tilsvarende hydrogenerede variant blev karakteriseret. Begge varianter viste sig termisk stabile og aktive, og med overordnet struktur svarende til den oprindelige form fra *Gallus gallus*. Der kunne dog observeres små variationer i strukturerne, især i regionen omkring Lys97-Gly104 relateret til peptid-flip af Asn103. Ændringerne her blev koblet til en lille svækkelse af termisk stabilitet og enzymatisk funktion.

Komplette neutrondiffraktionsdata blev samlet til atomar opløsning ved stuetemperatur (RT) og 100 K på krystaller af D-HEWL dyrket under betingelser hvor proteinet er aktivt,, og tilsvarende komplementære røntgen datasæt blev også målt ved disse temperaturer. Med strukturen fra neutron D-HEWL RT eksperimentet blev detaljer i konfigurationen og dynamikken af både proteinrester og vandmolekyler i proteinets aktive site, i dets aktive form, belyst. Dette datasæt gav desuden , en hidtil uset, indsigt i det potentiale der er i komplette neutron data med atomar opløsning. Datasættene på D-HEWL fra både Røntgen og neutroner ved stuetemperatur og 100 K er fundamentet i dette studie af atomare forskydninger. De foreløbige resultater, baseret på D-HEWL RT data, peger på forskelle i de ADP'er der opnås fra Røntgen og neutron baserede modeller. ADP'er fra Røntgen diffraktion forekommer større og mere isotrope end de tilsvarende parametre fra neutron eksperimentet.

Umiddelbart er beskrivelsen af strukturel uorden desuden mere nøjagtig i neutronmodellen og derved påvirkes ADP'erne i mindre grad. Alt i alt, reflekterer de neutronbaserede ADP'er formentlig mere biologisk og fysisk relevant information end de tilsvarende ADP'er baseret på Røntgen modeller. 100 K neutron datasættet kan muligvis bekræftet observationerne gjort ved stuetemperatur, men yderligere arbejde er som slutteligt beskrevet nødvendigt for at bestemme pålidelige ADP'er fra 100 K neutron datasættet.

Abbreviations

ADP	Atomic displacement parameter
CC	Correlation coefficient
Cryo-EM	Cryogenic electron microscopy
D-HEWL	Perdeuterated hen egg-white lysozyme
D-HEWL$_{EC}$	Perdeuterated hen egg-white lysozyme expressed in *Escherichia coli*
D-HEWL$_{PP}$	Perdeuterated hen egg-white lysozyme expressed in *Pichia pastoris*
DSC	Differential scanning calorimetry
DSF	Differential scanning fluorimetry
ESRF	European Synchrotron Radiation Facility
ESI	Electrospray ionization
HEWL	Hen egg-white lysozyme
HEWL$_{EC}$	Hen egg-white lysozyme expressed in *Escherichia coli*
H-HEWL	Hydrogenated hen egg-white lysozyme
H-HEWL$_{EC}$	Hydrogenated hen egg-white lysozyme expressed in *Escherichia coli*
ILL	Institut Laue-Langevin
MALDI-TOF	Matrix assisted laser desorption ionization – time-of-flight
MS	Mass spectrometry
MW	Molecular weight
NAG	N-acetyl-glucosamine
NAM	N-acetylmuramic acid
NMR	Nuclear magnetic resonance
NMX	Neutron macromolecular crystallography
NR	Neutron reflectometry
PDB	Protein Data Bank
PEG	Polyethylene glycol
pI	Isoelectric point
R.m.s.d.	Root mean-square deviation
RT	Room temperature
SAXS	Small-angle X-ray scattering
SANS	Small-angle neutron scattering
SEC	Size-exclusion chromatography
TLS	Translation-libration-screw
T_m	Melting temperature
XaDHL	Haloalkane dehalogenase from *Xanthobacter autotrophicus*
XFEL	X-ray free-electron laser

Table of contents

Chapter 1

Introduction

1.1. The relationship between protein structure and function

The central dogma of structural biology is that biological function can be deduced from structural data, supported by complementary biochemical and biophysical information. Prime examples of this assumption are the plethora of X-ray crystallographic studies aimed at understanding protein function through the determination of protein structure in different conformations (Strynadka & James, 1991; Van Tilbeurgh *et al.*, 1993; Van Den Bedem *et al.*, 2013; Pedersen *et al.*, 2007; Jin *et al.*, 2020; Kneller *et al.*, 2020). In many cases, this work entails the resolution of both native and ligand-bound complex structures. Nevertheless, what connects the different dots of structural information is protein dynamics, both at the molecular and atomic scales. It has been shown that proteins in their native state explore a multitude of conformations, through thermodynamical fluctuations (Ramanathan *et al.*, 2014). This high degree of flexibility in proteins is often crucial for their biological functions, driving *e.g.*, the diffusion or transport of ions or small molecules, the interaction with enzymatic substrates and products, or the formation of multimeric complexes.

Solution-state nuclear magnetic resonance (NMR) allows the detection in real-time, and in solution, of changes in protein conformation, *e.g.*, induced by the interaction with ligands [for a review, (Kovermann *et al.*, 2016)]. Protein dynamics can be detected not only from studies in different conditions but also by the averaging of multiple conformations that obey to the experimentally determined geometric restraints, in the same experiment. Although protein function and ligand-binding have been major focuses in NMR (Kumari *et al.*, 2021; Chen *et al.*, 2021; Liu *et al.*, 2021), relevant work has also been developed regarding the dynamics of protein folding (Miranker *et al.*, 1991, 1993; Zhuravleva & Korzhnev, 2017). More recently, cryogenic electron microscopy (cryo-EM) has become one of the methods of choice to study protein structure and dynamics (Jain *et al.*, 2019; Pillon *et al.*, 2021; Zhang *et al.*, 2021; Matsumoto *et al.*, 2021). This method has unique advantages that make it suitable for the observation of large macromolecular complexes and the detection of conformational variants often trapped in the same sample. Additionally, small-angle X-ray/neutron scattering can be valuable techniques for the search of conformational changes of proteins in solution (Rosenblum *et al.*, 2007; Vijayakrishnan *et al.*, 2010; Cuypers, Trubitsyna *et al.*, 2013). Finally, with the advances being made in the field of X-ray free-electron lasers (XFELs), the experimental observation of protein dynamics at the near-atomic level through serial femtosecond crystallography (SFX) is becoming a reality (Zhou *et al.*, 2016; Gisriel *et al.*, 2019; Mehrabi *et al.*, 2019; Pandey *et al.*, 2020).

While conformational changes are essential in unravelling protein interactions, and subsequently biological function, so far, the dynamics at the atomic level in macromolecules have been somewhat overlooked. Considering the aforementioned methods, only X-ray crystallography is currently able to provide experimental information regarding atomic thermal motion. Extensive knowledge has been gathered by small-molecule crystallography in terms of the atomic motion in crystals, through the combination of X-ray and neutron diffraction studies (Koritsanszky & Coppens, 2001; Madsen, 2006;

Munshi *et al.*, 2008). In contrast, macromolecular crystallography remains limited by the information available from X-ray diffraction, since this type of studies using neutrons is currently extremely challenging. The experimental data obtained for small molecules show that often the assessment of atomic thermal motion from X-rays is substantially different than that from neutron diffraction (Blessing, 1995). It is, however, accepted that neutron diffraction is able to provide an unbiased depiction of atomic motion, since the results obtained from X-rays do not permit the deconvolution between atomic motion and non-spherical deformation of valence electron density distribution (Blessing, 1995; Koritsanszky & Coppens, 2001).

The experimental observation of realistic and unbiased atomic thermal motion is crucial for a better fundamental understanding of the dynamics in a crystal structure. Small-molecule studies, including the refinement of multipole models, have provided valuable knowledge on the nature of chemical bonding in crystals through the analysis of X-ray charge-density, based on Bader's quantum theory of atoms in molecules (QTAM) (Bader, 1994). This type of empirical efforts has currently no parallel in macromolecules, with the exception of a few attempts at refining multipole models from ultra-high resolution X-ray diffraction data (Held & Van Smaalen, 2014). Hence, the main objective of this thesis is to push the boundaries of macromolecular neutron crystallography (NMX) towards the description of atomic thermal motion in protein crystals.

To achieve the project's main goal, it is vital to collect complete atomic resolution neutron diffraction data, which, to our knowledge, has never been achieved before for protein crystals. Additionally, the neutron data need to be collected at different temperatures, and equivalent X-ray diffraction data has to be acquired for subsequent comparative analysis. This work poses a series of technical challenges that need to be addressed to yield the desired experimental data:

i) Production of significant quantities of perdeuterated protein.
ii) Structural and biophysical characterization of recombinantly produced perdeuterated protein, and subsequent quantification of isotope effects.
iii) Optimization of protein crystallization in deuterated conditions to produce high quality diffracting crystals and crystal growth to several mm^3 in volume for neutron diffraction.
iv) Complete atomic resolution neutron and X-ray data collection at both room temperature and 100 K.
v) Neutron diffraction data reduction.
vi) Neutron model refinements.
vii) Comparative analysis of atomic thermal motion obtained from X-ray and neutron diffraction data.

If successful, this work has the potential to clarify the hypothetical bias inherent to X-ray diffraction in the estimation of atomic motion. The comparison between X-ray and neutron derived atomic motion can provide a new basis for the interpretation of crystal structures obtained from X-ray diffraction.

Ultimately, this work can pave the way for the investment and development of NMX by unlocking the full potential of the technique and showcasing its complementarity to X-ray crystallography. Additional information should also be obtained concerning the effects of perdeuteration in protein structure and dynamics.

1.2. Neutron macromolecular crystallography

Neutron diffraction stands as a strongly complementary technique to X-ray diffraction. A striking example of this relationship can be found in macromolecular crystallography. While X-ray crystallography has been the prominent method to decipher the structure of biomolecules, as easily demonstrated by the number of structures deposited in the Protein Data Bank (PDB) (Berman *et al.*, 2000) from the most used methods (Fig. 1), neutron diffraction has contributed immensely to the understanding of protein function and interactions through the determination of Hydrogen (H) positions. In fact, the observation of protonation states and hydrogen-bond (H-bond) networks has been the main driving force behind the interest in NMX, since this type of information is precluded in X-ray diffraction studies.

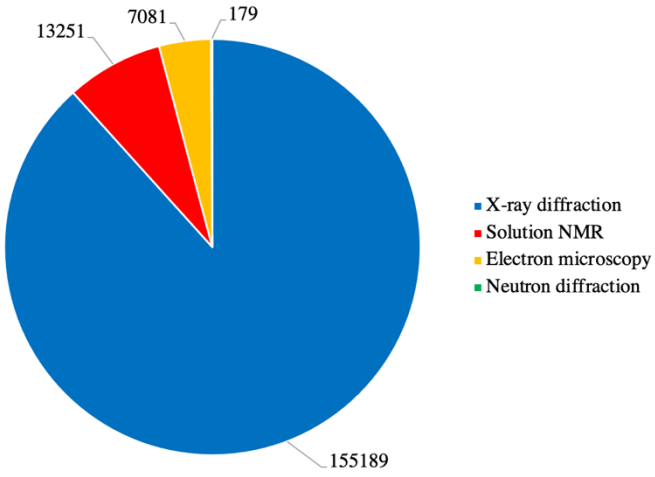


Figure 1

Comparison of the number of structures deposited in the PDB (on the 28th of March of 2021) using or including the use of the most common methods: X-ray diffraction, solution-state nuclear magnetic resonance, electron microscopy and neutron diffraction. This representation emphasizes the overwhelming contribution to the deciphering of macromolecular structures from X-ray diffraction compared to the remaining techniques.

The difference between the detail available from neutron and X-ray diffraction is directly linked to the nature of the interaction between both radiation types and the atoms in the crystal. X-rays are scattered by the electrons of the atoms, meaning that the atomic scattering length is directly proportional to the atomic number (Fig. 2). On the other hand, neutrons scatter from the atomic nuclei, resulting in a non-linear relationship between neutron coherent scattering length and atomic number (Fig. 2). Interestingly, while H atoms have very weak contributions to the scattering of X-rays, the coherent neutron scattering length of its isotope Deuterium (D) is of similar magnitude as other common atoms in biomolecules, *e.g.*, C, N, O (Sears, 1992). Therefore, while H positions are elusive in X-ray electron density maps, in neutron maps, they are readily available at resolutions better than ~2.5 Å (Chen *et al.*, 2012). Concomitantly, NMX studies routinely involve the partial or full deuteration of samples, through

solvent exchanges to D_2O, *e.g.*, using soaking or vapor diffusion, and recombinant expression of selectively deuterated biomolecules. Further insight into these techniques is given in subsequent sections of this thesis.

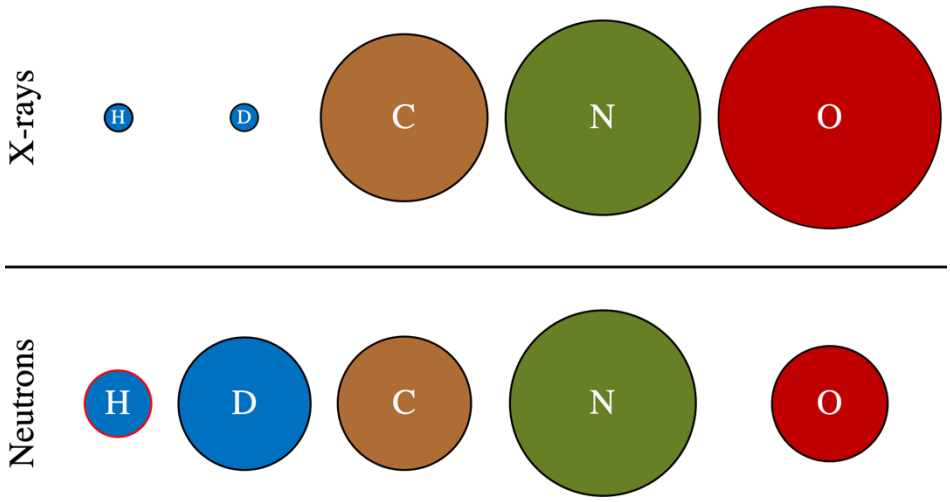


Figure 2

Representation of the relative atomic coherent scattering lengths of Hydrogen, Deuterium, Carbon, Nitrogen, and Oxygen, for X-rays and neutrons. While for X-rays the scattering length is proportional to the atomic number, for neutrons this relationship is non-linear. Note that in the case of H, for neutrons, the coherent scattering length is negative, represented by the red contour.

Additionally, neutron diffraction and its resulting structural model may provide a better description of both structural disorder and atomic thermal motion, when compared to X-ray diffraction. In a molecule, atoms vibrate away from their equilibrium position, in a temperature-dependent manner. The higher the temperature, the greater will the atomic motion be, and vice-versa. Nevertheless, the atomic movement is affected by the chemical environment, by covalent and non-covalent interactions, resulting in characteristic behaviors which have different magnitudes in distinct directions. Debye (1914) and Waller (1923) have conducted early studies on the effects of temperature on the relationship between atomic motion and the Bragg reflections measured in X-ray diffraction experiments. As a result, the Debye-Waller factor (W_i) was discovered and implemented in the equation that describes electron density from the Fourier sum of the experimentally determined reflection intensities, converted into structure factors [F(h,k,l)] (eq. 1) (Koetzle & McIntyre, 2012).

$$\rho(x,y,z) = \sum F(h,k,l)\exp\{-2\pi i(hx+ky+lz)\} \quad (1)$$

Subsequently, the structure factors' function (eq. 2) (Koetzle & McIntyre, 2012) depends on the Fourier sum of the atomic scattering factor (b_i), the atomic site occupancy (a_i), and the Debye-Waller factor (W_i), over the Miller indices (*hkl*) and atomic coordinates (*xyz*).

$$F(h,k,l) = \sum a_i b_i \exp(-W_i) \exp\{2\pi i(hx + ky + lz)\} \quad (2)$$

The scattering factor, or as mentioned previously, the scattering length, is characteristic for each atom type, describing the magnitude of their interaction with the incident radiation. The site occupancy quantifies the fraction of scatterers which can be found in a particular plane of atoms. Meanwhile, the Debye-Waller factor, as described earlier, accounts for the atomic thermal motion (eq. 3) (Koetzle & McIntyre, 2012).

$$W_i = \exp\left\{\frac{-U_i Q^2}{2}\right\} = \exp\left\{-8\pi^2 U_i \left(\frac{sin^2\theta}{\lambda^2}\right)\right\} \quad (3)$$

Equation 3 represents the isotropic approximation to W_i, where U_i is the isotropic atomic displacement parameter (ADP), θ is the half of the scattering angle, λ is the incident radiation wavelength, and Q is the scattering vector. However, the assumption that the atomic thermal motion is isotropic (*i.e.*, of similar magnitude in all directions) is physically unrealistic. This approximation is often made when diffraction data resolution is worse than 1.2 Å (*i.e.*, atomic resolution) and the data-to-parameter ratio in model refinement is low. A more accurate description of atomic thermal motion is given by the anisotropic model (eq. 4) (Koetzle & McIntyre, 2012), which introduces second-order tensors to the ADPs for the three reciprocal lattice axes (a, b, c).

$$W_i = \exp\begin{Bmatrix} -2\pi^2 U_{11} a^2 h^2 + U_{22} b^2 k^2 + U_{33} c^2 l^2 \\ + 2U_{12} a \cdot bhk + 2U_{13} a \cdot chl + 2U_{23} b \cdot ckl \end{Bmatrix} \quad (4)$$

Importantly, the relationships described in the aforementioned equations hold true for neutron diffraction. An interesting difference between the atomic interaction with X-rays and neutrons is that the scattering amplitude for neutrons is constant, independent of momentum transfer, since the atomic radius is ~10^4 times smaller than the neutron wavelengths typically used in diffraction experiments (Koetzle & McIntyre, 2012). Meanwhile, for X-rays the scattering intensity decays exponentially with Q, due to the destructive interference between waves that scatter from different regions of the electron density cloud. As a consequence, in the case of neutrons, there is a minimal correlation between the Debye-Waller factor and the site occupancy, while for X-rays the decay of intensity with Q arising from the atomic scattering factor and from the Debye-Waller factor can be very similar (Koetzle & McIntyre, 2012). Concomitantly, the fact that X-rays scatter from electrons renders more difficult the accurate determination of the atomic coordinates, since electron density clouds are often distorted due to covalent and non-covalent chemical bonding between atoms with different electrophilic potentials. On the other

hand, neutrons scatter from the atomic nuclei, meaning that the maximum of the nuclear density in neutron maps corresponds to the atomic equilibrium position. Additionally, neutron maps provide information on the position of H/D atoms, which constituted approximately 50% of all protein atoms, whereas in X-ray electron density maps only a few H positions can be identified when ultra-high resolution data is available (Wang *et al.*, 2007; Schmidt *et al.*, 2011). Finally, when collecting X-ray diffraction data to high resolutions, in synchrotrons, radiation damage is a major concern and if present can lead to misleading structural data (Garman, 2003; Russi *et al.*, 2017). Meanwhile, this issue is not a reality in neutron diffraction experiments, because, at the energies used, neutrons are non-ionizing radiation and therefore room temperature (RT) measurements are routine in NMX studies (Blakeley, 2009). The combination of the aforementioned properties makes neutron diffraction the technique of choice for the study of atomic thermal motion in a crystal, as it is already the case for small molecules (Koritsanszky & Coppens, 2001; Madsen, 2006; Munshi *et al.*, 2008; Koetzle & McIntyre, 2012). Nevertheless, NMX poses a series of challenges to the measurement of complete atomic resolution diffraction data, which is vital for the refinement of anisotropic ADPs. These hurdles stem from the relatively low neutron flux available at current reactor and spallation sources in comparison with the photon flux possible at synchrotron facilities. This difference can be of ~10^9 orders of magnitude, which implicates the use of samples with significantly larger volumes (in the order of mm^3 as opposed to μm^3) and extended data collection times (several days rather than seconds) in NMX (Blakeley, 2009). For proteins, growing crystals of several mm^3 requires a significant amount of precious material and is sometimes impossible since it is an unfavorable thermodynamical process. Additionally, growing protein crystals can lead to undesired crystal defects, such as twinning and multiple crystals. Another concern arises in low-temperature studies, since cryo-cooling crystals typically results in increased mosaicity and can lead to loss of diffraction quality, especially in the case of large protein crystals for which the cryo-cooling step is hindered by a slower temperature gradient (Kriminski *et al.*, 2003; Chinte *et al.*, 2005). These challenges of NMX are discussed further in the Methods section.

Illustrating the difficulties of working in NMX is the considerably low number of neutron structures deposited in the PDB in comparison with X-ray crystallography (Fig.1). As of the 28th of March of 2021, there were only 70 structures obtained from neutron diffraction data alone, while an additional 109 models were the result of joint X-ray and neutron data refinements. Over the last decade, the joint X-ray and neutron refinement strategy has contributed considerably to the number of neutron structures released by the PDB (Fig. 3). These structures benefit from an increase in the data-to-parameter ratio provided by the addition of X-ray data to the refinements, usually accompanied by improvements in resolution limits. In fact, according to a recent survey of the PDB (Liebschner *et al.*, 2018), the average completeness and resolution limit of neutron diffraction data is approximately 80% and 2 Å, respectively. Once again, these limitations in both completeness and resolution of neutron data are consequences of the limited neutron flux in dedicated instruments. To date, there is only one protein structure that has been determined from neutron diffraction data up to atomic resolution, being that of

crambin to 1.1 Å (PDB entry 4fc1; Chen *et al.*, 2012). Despite of its small molecular weight (MW), 4.7 kDa (46 residues), and small unit cell (a = 22.79 Å, b = 18.83 Å, c = 41.04 Å, α = 90°, β = 90.9°, γ = 90°), the neutron dataset is only 78.8% complete, 65.8% in the outer resolution shell. Therefore, the low data-to-parameter ratio available rendered impossible the refinement of anisotropic ADPs for all protein atoms, including H/D. Moreover, the sample used was of hydrogenated crambin solvent exchanged to D_2O, meaning that ~75% of all H positions were still occupied by H instead of D. Since H has a negative coherent neutron scattering length (Sears, 1992), these will be absent from the positive neutron density maps and thus cannot be reliably be refined in terms of anisotropic motion.

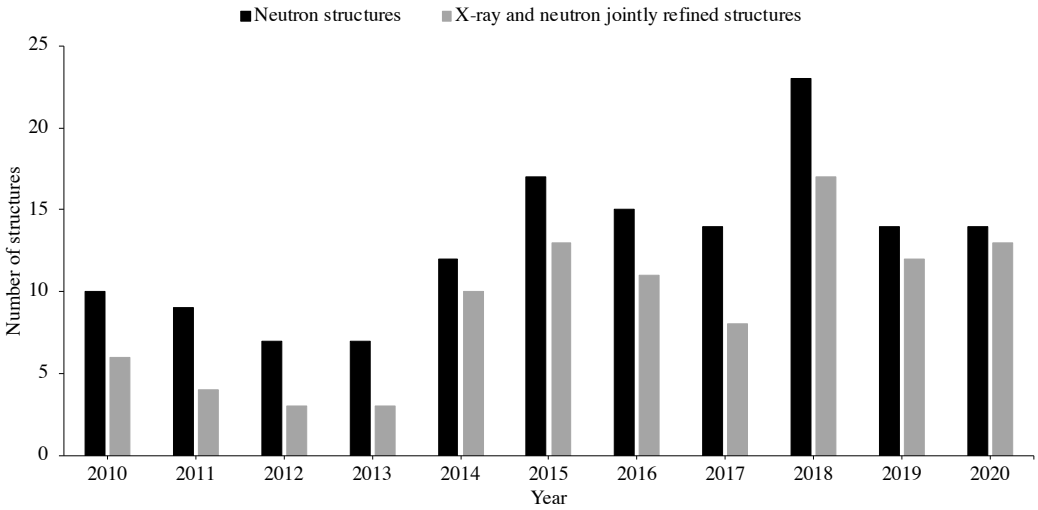


Figure 3

Number of neutron structures released by the PDB between 2010 and 2020, highlighting the significant contribution of the X-ray and neutron joint refinement strategy to the overall number of models obtained from neutron data.

1.3. Protein perdeuteration

As previously mentioned, H atoms comprise approximately 50% of the atoms in a protein. Despite of being elusive to X-ray crystallography, even at atomic resolutions, they play a pivotal role in protein biological function (Niimura & Bau, 2008; Blakeley, 2009; Oksanen *et al.*, 2017). Perdeuteration is a technique in which all H atoms in a biomolecule are replaced by the isotope, D. This substitution occurs both in non-exchangeable positions (*i.e.*, bound to C) which represent about 75% of all H positions, and in exchangeable sites (*i.e.*, bound to N, O, S) equivalent to the remaining 25% of H positions. Biomolecular deuteration is an essential tool when employing techniques such as NMX, small-angle neutron scattering (SANS), neutron reflectometry (NR), neutron spectroscopy and NMR (Haertlein *et al.*, 2016; Varga *et al.*, 2007). These methods take advantage of the different physical properties of H and D, namely their neutron coherent and incoherent scattering lengths and cross sections, and, in the case of NMR, their nuclear spin. In neutron studies, the fact that D has a positive coherent scattering length comparable to ^{12}C, ^{14}N and ^{16}O, while for H this value is negative (Table 1) (Sears, 1992), is exploited not only to improve the quality of the neutron data collected, but also to deconvolute different scattering contributions.

In NMX, the use of perdeuterated protein can produce several benefits due to the elimination of H from the sample, which has a large incoherent scattering cross section (Table 1) (Sears, 1992). By replacing H for D, an increase in signal-to-noise ratio can be obtained which may lead to the use of crystals with smaller volumes, by one order of magnitude (Hazemann *et al.*, 2005), and to improvements in diffraction data resolution (Blakeley, 2009). Moreover, the subtraction of H from a sample helps to avoid cancelation effects in the Fourier maps (Fisher *et al.*, 2014), arising from the negative coherent scattering length of H. For SANS, deuteration is used in contrast matching techniques (Dunne *et al.*, 2017; Laux *et al.*, 2008; Haertlein *et al.*, 2016) to enable the masking of specific components of protein-protein (Vijayakrishnan *et al.*, 2010), protein-nucleic acid (Cuypers, Trubitsyna *et al.*, 2013), or protein-lipid complexes (Breyton *et al.*, 2013). This is achieved through selective deuteration of a complex and by matching the scattering length density of a molecule to that of a D_2O/H_2O solvent mixture. Contrast matching has also been applied to the study of membrane proteins (Josts *et al.*, 2018; Nitsche *et al.*, 2018; Kehlenbeck *et al.*, 2019) with the help of stealth-deuterated nanodiscs (Maric *et al.*, 2014, 2015). In similar fashion, selective deuteration and contrast matching have been widely employed in NR studies (Grage *et al.*, 2011; Moulin *et al.*, 2018; Waldie *et al.*, 2018, 2019, 2020). In neutron spectroscopy studies, deuterium labelling and reverse labelling have been used to study the dynamics of biological macromolecules (Foglia *et al.*, 2016) and their hydration water (Wood *et al.*, 2013). In the case of NMR, partial deuteration is applied to allow the study of high MW proteins or protein complexes, taking advantage of the integer nuclear spin of D. This approach simplifies the NMR spectra from the remaining ^{1}H nuclei and increases the signal-to-noise ratio due to the effects on the relaxation of bonded or adjacent ^{1}H, ^{13}C and ^{15}N atoms (Sattler & Fesik, 1996).

Table 1

Atomic coherent and incoherent neutron scattering lengths and cross sections of the most common atoms in a protein (Sears, 1992). Z – atomic number, A – mass number, b_c – bound coherent scattering length, b_i – bound incoherent scattering length, σ_c – bound coherent cross section, σ_i – bound incoherent cross section.

element	Z	A	b_c (fm)	b_i (fm)	σ_c (barn)	σ_i (barn)
H	1	1	-3.74	25.27	1.75	80.27
D	1	2	6.67	4.04	5.59	2.05
C	6	12	6.65	0	5.55	0
N	7	14	9.37	2.02	11.03	0.51
O	8	16	5.80	0	4.23	0

Several studies on protein perdeuteration have described the isotope effects, from solvent and macromolecular H/D substitution, on protein biophysical and structural properties. The biophysical properties of a protein, mainly its thermal stability and dynamics, appear to be altered by its partial or full deuteration. In a number of cases, the perdeuterated variant of a protein seems to be less thermally stable than its hydrogenated counterpart (Berns, 1963; Hattori *et al.*, 1965; Brockwell *et al.*, 2001; Meilleur *et al.*, 2004; Koruza *et al.*, 2018; Nichols *et al.*, 2020). Meanwhile, both variants display in general higher thermal stability in D_2O solvent than in H_2O (Hattori *et al.*, 1965; Harrington & von Hippel, 1961; Makhatadze *et al.*, 1995; Dong *et al.*, 1997; Kuhlman & Raleigh, 1998; Sasisanker *et al.*, 2004; Efimova *et al.*, 2006). The macromolecular isotope effect is thought to impact protein stability by weakening the amino acids side-chain hydrophobic interactions, as described by Hattori *et al.*, 1965, due to the lower vibrational amplitudes of the C-D bond, compared to the C-H bond. This change leads to lower steric requirements for the hydrophobic interactions between residues side-chains. On the other hand, the higher thermal stability observed for proteins in D_2O solvent seems to be a consequence of stronger protein-protein and protein-water H-bonds, promoted by H/D substitution in exchangeable positions, and of differences in solvation and dynamics between heavy water (D_2O) and H_2O. A stronger hydrophobic effect has been observed in D_2O leading to more compact protein structures and reduced solvent accessible surfaces (Svergun *et al.*, 1998; Sasisanker *et al.*, 2004; Efimova *et al.*, 2006; Jasnin *et al.*, 2008). Moreover, a decrease in protein flexibility in D_2O has been noted in a direct comparison of hydrogenated and perdeuterated rat γE-crystallin in H_2O and D_2O (Artero *et al.*, 2005), where the solvent isotope effect is the most noticeable factor impacting the residues B-factors.

Conversely, X-ray and neutron crystallographic studies have investigated the impact of perdeuteration in protein structure. In general, identical crystal structures have been observed for a hydrogenated protein and its perdeuterated variant (Gamble *et al.*, 1994; Meilleur *et al.*, 2004; Artero *et al.*, 2005; Liu *et al.*, 2007; Koruza *et al.*, 2019). Nevertheless, the perdeuteration of haloalkane

dehalogenase from *Xanthobacter autotrophicus* (XaDHL) stands out as an example where minor changes in protein structure can result in dramatic shifts in the interpretation of biological function. In the work of Liu *et al.* (2007), it was observed that in the perdeuterated variant of XaDHL the hydrolytic water molecule, essential for enzymatic catalysis, was absent from the active site, leading to the displacement of the residues Asp124 and His289 and subsequent formation of an H-bond. This structural variation resulted in the inactivation of the perdeuterated form of XaDHL. It is important to note that both hydrogenated and perdeuterated variants of XaDHL were crystallized in H_2O, and due to the lower solubility of the perdeuterated protein, the pH of its crystallization condition was 4.6, instead of 6.0 used for the hydrogenated form. The more acidic pH used to crystallize the perdeuterated protein may play a role in the protonation of Asp124 and His289 and subsequent side-chain displacement and exclusion of the hydrolytic water from the active site. This example shows that validation of the structural information obtained from a perdeuterated variant with a direct comparison to its hydrogenated counterpart is crucial. More important seems to be the use of identical crystallization conditions for both hydrogenated and perdeuterated proteins, even though this can be challenging due to variations in solubility linked to protein perdeuteration.

An interesting observation was made by Koruza *et al.* (2019) regarding both solvent and macromolecular H/D isotope effects on protein structure. The authors studied the structures of three forms of human carbonic anhydrase IX, with six surface amino acid mutations in the catalytic domain: i) hydrogenated protein in H_2O; ii) hydrogenated protein vapor buffer-exchanged to D_2O; iii) and perdeuterated protein vapor buffer-exchanged to D_2O. Although the crystal structures were very similar, minor changes in the active sites' arrangement of water molecules were noted. Increased water disorder is visible in the hydrogenated protein in D_2O compared to H_2O. This disorder is especially noticable in the perdeuterated protein in D_2O. Koruza and collaborators discuss that this observation seems to be linked to either protein deuteration effects or the presence of a formate ion, although this ion is also present in one of the hydrogenated structures, in H_2O. This example suggests, once again, that the H/D isotope effect can result in minor variations on protein and water structure which have always to be considered in regard to the structural data from the hydrogenated protein.

In the present thesis, protein perdeuteration was an essential step, since the objective was to eliminate, to the maximum possible extent, the presence of H atoms in the crystals submitted to neutron diffraction. In this way, neither the cancellation effects in the Fourier maps or the increase in diffraction background noise, arising from the scattering of H atoms, would degrade the quality of the neutron diffraction data collected. Therefore, perdeuterated protein was expressed recombinantly in fully deuterated culture medium and later crystallized in 100% D_2O solution, using crystallization conditions similar to those employed for the hydrogenated variant. Subsequently, extensive comparisons of protein biophysical and structural properties were performed between the perdeuterated and hydrogenated recombinant variants, and also with the native, commercially available, hydrogenated protein.

1.4. Hen egg-white lysozyme

Hen egg-white lysozyme (HEWL) was the first enzyme structure to be solved by X-ray crystallography (Blake *et al.*, 1965). Later it became a widely used model in structural biology due to several factors: i) it can be obtained at low cost; ii) it can be crystallized in various space groups and under well-known conditions; iii) and it is a remarkably stable protein. Therefore, HEWL has been the subject of numerous studies on protein folding (Miranker *et al.*, 1991, 1993; Radford *et al.*, 1992; Wildegger & Kiefhaber, 1997), protein crystallization (Durbin & Feher, 1986; McPherson & Delucas, 2015; Darmanin *et al.*, 2016), and on neutron and X-ray diffraction (Mason *et al.*, 1984; Bon *et al.*, 1999; Russi *et al.*, 2017; Mehrabi *et al.*, 2019).

The highest resolution crystal structure of HEWL from X-ray diffraction data is that reported by Wang *et al.* (2007), at 0.65 Å resolution (PDB entry 2vb1). The crystal form used in their work is the triclinic (P1), obtained in sodium nitrate and sodium acetate pH 4.7, by sitting-drop vapor-diffusion. The crystal solvent content was 27%, which is significantly lower than that of other HEWL crystal systems, such as monoclinic (31%, *e.g.*, in PDB entry 3wl2), tetragonal (40%, *e.g.*, in PDB entry 4b4e), and orthorhombic (44%, *e.g.*, in PDB entry code 6f1o). The difference in solvent content observed between the various crystal forms explains why the P1 space group yields the highest resolution X-ray diffraction data, since the disordered bulk solvent in a crystal has a destructive interference to the Bragg reflections and provides higher background noise to the recorded diffraction data (Matthews, 1968; Kantardjieff & Rupp, 2003; Weichenberger *et al.*, 2015). While high resolution NMX studies have been recently conducted on other proteins, such as crambin (Chen *et al.*, 2012) and rubredoxin (Cuypers, Mason *et al.*, 2013), only four neutron structures of HEWL have been reported so far (Mason *et al.*, 1984; Bon *et al.*, 1999; Niimura *et al.*, 1997; Kita & Morimoto, 2020), none of these being perdeuterated. The combination of the aforementioned motives led to the choice of triclinic HEWL as the ideal target for the present study.

Lysozyme was first described in 1922 by Alexander Fleming (Fleming, 1922), whom had observed the presence of a bacteriolytic element in several tissues and secretions from humans, animals and plants. One of the sources of this element was egg-white. Later it was found that HEWL is a hydrolase that cleaves the 1,4-β-linkages between N-acetylmuramic acid (NAM) and N-acetyl-glucosamine (NAG) residues in bacterial cell wall peptidoglycan (Fig. 4). The first indications about the natural substrate of HEWL came from early experiments (Meyer *et al.*, 1936; Epstein & Chain, 1940; Meyer *et al.*, 1946), which revealed that the enzyme releases N-acetyl-amino sugars from gram-positive *Micrococcus lysodeikticus*. Later, Berger & Weiser (1957) showed that lysozyme was also able of degrading chitin, the linear polymer of NAG. These results led to the conclusion that HEWL possesses a β-(1-4) glucosaminidase activity. Subsequently, Salton & Ghuysen (1959) reported the isolation of a tetrasaccharide product from *M. lysodeikticus* cell wall degradation by lysozyme, which comprised equimolar amounts of NAM and NAG residues. Wenzel *et al.* (1962) described that the monomer NAG

is a competitive inhibitor of lysozyme activity, which elucidates its inactivity towards chitosan (poly-glucosamine) and cellulose and highlights the importance of the N-acetyl group in the substrate molecule.

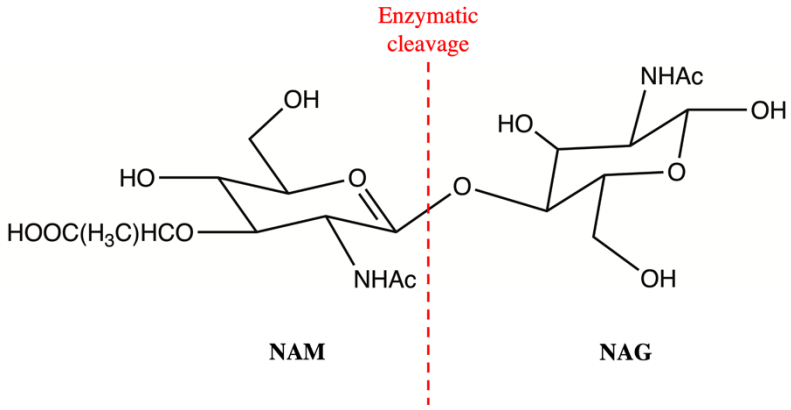


Figure 4

Representation of the HEWL enzymatic cleavage of the 1,4-β-linkage between N-acetylmuramic acid (NAM) and N-acetyl-glucosamine (NAG) residues.

Early crystallographic studies of inhibitor binding, at low resolution (6 Å), allowed the identification of the lysozyme binding cleft (Johnson & Phillips, 1965). The authors described six positions on the surface of the enzyme able to bind individual sugar residues (termed A through F). Further work at 2 Å resolution demonstrated the specific interactions between the lysozyme residues at the binding cleft and the polysaccharide inhibitor residues (Fig. 5) (Phillips, 1967; Blake *et al.*, 1967). Noticeably, the main-chain CO and NH groups of Ala107 and Asn59, respectively, and the NH groups of the side-chains of Trp62 and Trp63 were found to stabilize the binding of the A, B and C residues of tri-NAG through H-bonds. Model building led to the extrapolation of the remaining D, E and F sites and corresponding interactions with the binding cleft residues.

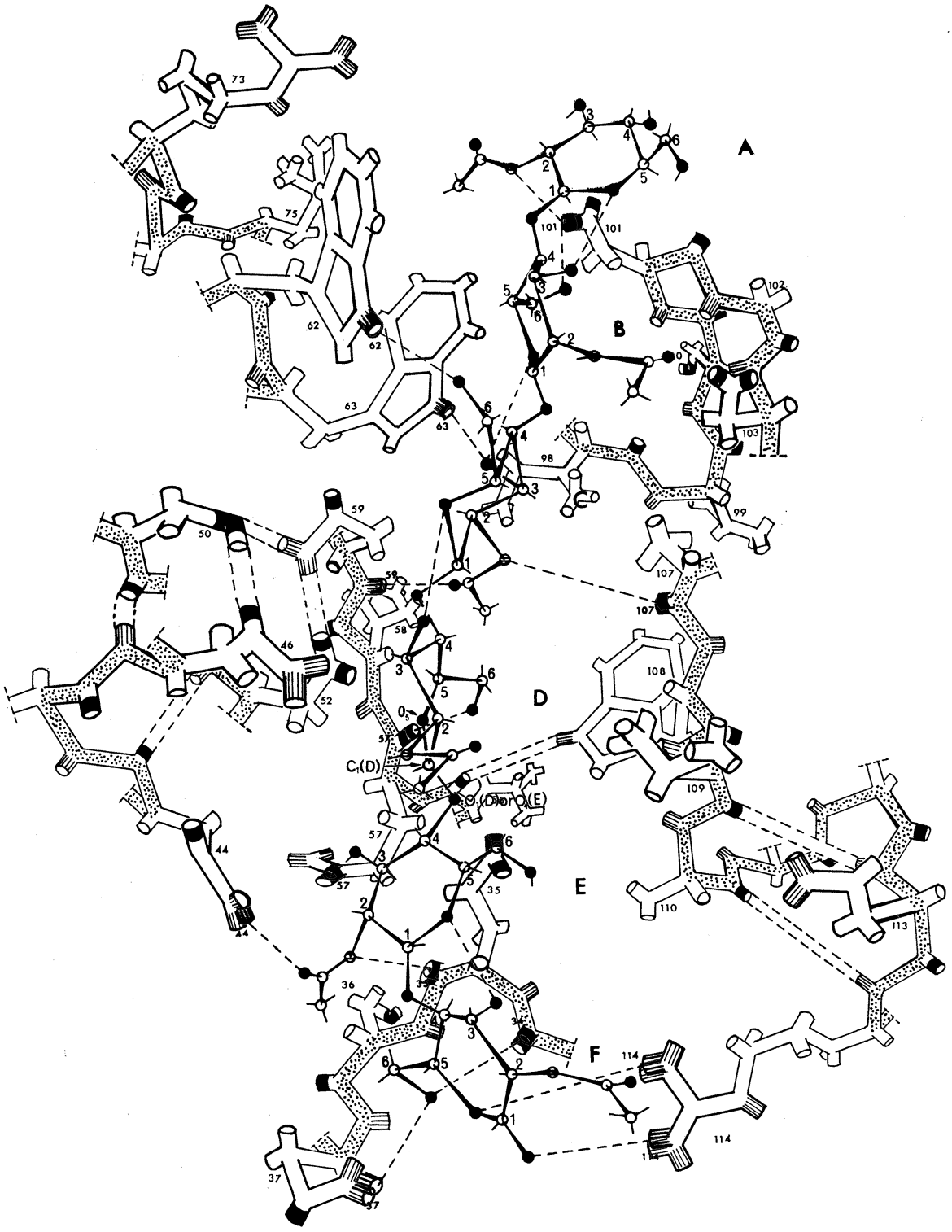


Figure 5

Arrangement of HEWL active site and cleft where substrate molecules bind, as reported by Blake *et al.* (1967). The protein main-chain is shown speckled and residues side-chains are represented by empty thick sticks. The overlayed hexasaccharide substrate molecule is displayed by thin filled sticks, and its monomers are labelled from A through F. The extrapolated protein-ligand H-bond interactions are depicted as dashed lines.

According to the proposed model, the two catalytic residues, Glu35 and Asp52, would be located on each side of the 1,4-β-linkage of the substrate between residues D and E, indicating their pivotal role in the enzyme's mechanism of action. Moreover, Blake *et al.* (1967) discussed a potential distortion of the hexasaccharide residue in site D due to apparent steric clashes with the enzyme binding cleft amino acids. This hypothesis was later validated by crystallographic studies of lysozyme-substrate complexes at 1.5 Å resolution (Strynadka & James, 1991). The authors described a distortion from a chair to sofa conformation of the NAM residue D (*i.e.*, where the CH_2OH group moves from an equatorial to a quasi-axial position, promoting the coplanarity of the ring atoms). Additionally, Strynadka & James (1991) clarified the roles of the Asn59-Cys64, Val99-Thr118, and Arg68-Cys78 regions in the accommodation of the substrate molecule, by comparing the atomic coordinates and variations in B-factors of the substrate-bound and unbound structures of lysozyme. Two mechanisms of action were proposed for lysozyme by Phillips (1967) and Koshland Jr. (1953). The fundamental difference between both pathways concerns the stabilization of a long-lived carboxylate-oxocarbonium ion pair between Asp52 and the anomeric carbon atom of NAM residue D (Fig. 6) (Phillips, 1967), or the formation of a short-lived oxocarbonium transition state that would be converted into a tetrahedral covalent intermediate (Koshland Jr., 1953). The work of Strynadka & James (1991) contributed with evidence supporting the existence of a long-lived carboxylate-oxocarbonium ion pair as initially proposed by Phillips (1967).

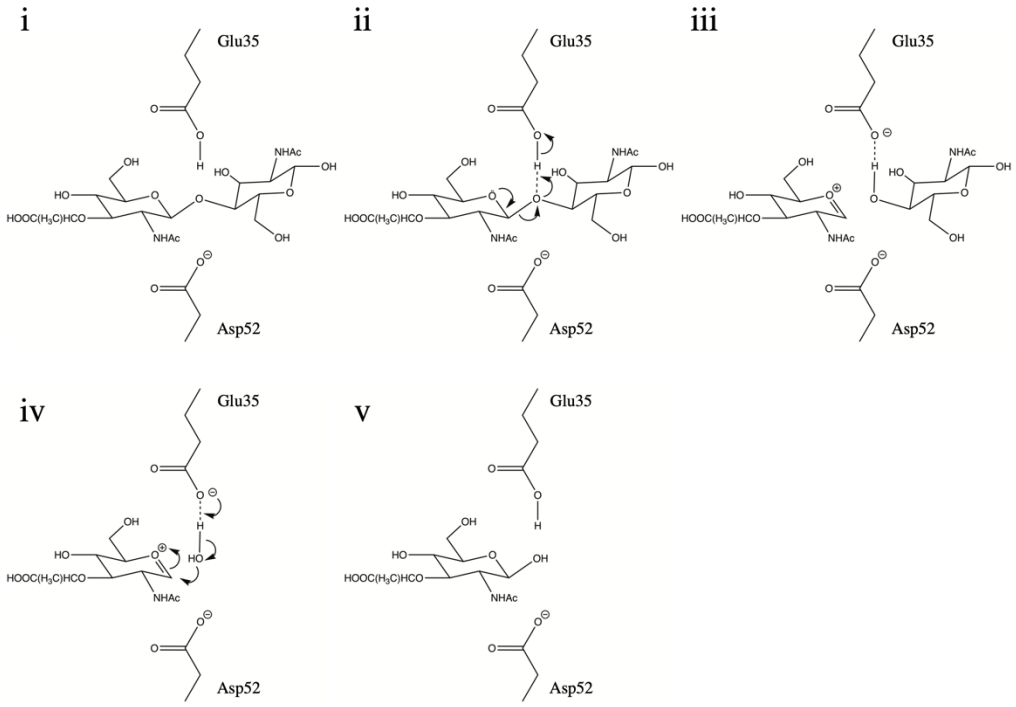


Figure 6

Representation of the HEWL enzymatic mechanism as proposed by Phillips (1967), where the positively charged NAM residue, after the cleavage of the glycosidic bond, is stabilized by the negatively charged Asp52 carboxylate in a long-lived ion par, in step iii.

Chapter 2

Methods

2.1. *In vivo* recombinant protein expression

Methodologies and applications of biomolecular perdeuteration in structural biology were described in detail by Haertlein *et al.* (2016). Perdeuterated recombinant protein expression can be achieved through chemical synthesis (*in vitro*) or biosynthesis (*in vivo*). An *in vitro* approach is only suited to the production of short peptides and, through cell-free systems, of specific proteins which may present, *e.g.*, toxicity issues to cells. In general, *in vivo* expression of perdeuterated macromolecules, using bacteria, yeast, insect, and mammalian cells, is preferred. Generally, bacterial and yeast systems can be readily adapted to grow in media with high levels of deuterium, while higher organisms have lower tolerance to deuteration and also require the supplementation of deuterated precursors, such as essential amino acids [for a review see Katz & Crespi (1966)]. The expression system selection depends primarily on the recombinant protein in study and its co- and post-translational modifications (*e.g.*, disulphide bond formation, glycosylation, and phosphorylation). The gram-negative bacteria *Escherichia coli* is widely used in recombinant protein expression; however, it is not always capable of performing all the aforementioned modifications, leading in some cases to limited protein solubility and to the formation of inclusion bodies. Subsequently, the overexpressed aggregated protein needs to be refolded to a native-like state before further biophysical and structural studies. Another limitation of the *E. coli* system is its double-membraned envelope, hindering in general the secretion of proteins to the growth medium, contrary to gram-positive systems, such as *Bacillus* or *Brevibacillus*, or to the yeast system, *Pichia pastoris*. This characteristic of the bacterial system can pose difficulties to the purification of the recombinant protein, which needs then to be separated from the extract of lysed cells. On the other hand, the *P. pastoris* expression system allows the formation of disulphide bonds, resulting in the expression of correctly folded and soluble proteins. Moreover, the addition of the α-factor secretion signal, from *Saccharomyces cerevisiae*, to the recombinant protein sequence permits the secretion of the target protein to the growth medium, generally simplifying its purification process.

The most frequently used *E. coli* strain for the expression of recombinant perdeuterated protein is the BL21 strain, which easily adapts to grow in D_2O through a procedure that takes several days. The adaptation process was reported by Paliy *et al.* (2003) on *E. coli* JM109, MRE600 and MRE600Rif strains. A similar stepwise adaptation protocol to deuterated minimal medium is used by the Life Sciences Group of the Institute Laue Langevin (ILL) (Fig. 7) (Haertlein *et al.*, 2016). After adaptation to deuterated minimal media conditions, a prototrophic *E. coli* strain, such as BL21(DE3), can be grown in fermenters by high cell density cultivation, where d_8-glycerol, deuterated acetate or succinate is added and used as carbon source. Recombinant protein expression using *E. coli* BL21(DE3) involves the cloning of the gene of interest under the control of bacteriophage T7 transcription and the supply of T7 RNA polymerase. Overexpression can be induced by adding isopropyl β-D-1- thiogalactopyranoside (IPTG), if a *lacUV5* promoter is present to regulate the expression of the T7 RNA polymerase gene (Studier *et al.*, 1990).

The *P. pastoris* system possesses the advantages of other eukaryotic systems for heterologous protein expression (*e.g.*, disulphide bond formation), while being able to grow in deuterated minimal media. The adaptation to deuterated conditions can be achieved through a similar process as described for the *E. coli* system (Fig. 1), although the composition of the growth medium is different. Additionally, *P. pastoris* can also use d_8-glycerol as carbon source. Recombinant protein expression is methanol-induced by using the alcohol oxidase (AOX1) promoter, which has been included in a series of commercially available *P. pastoris* expression vectors.

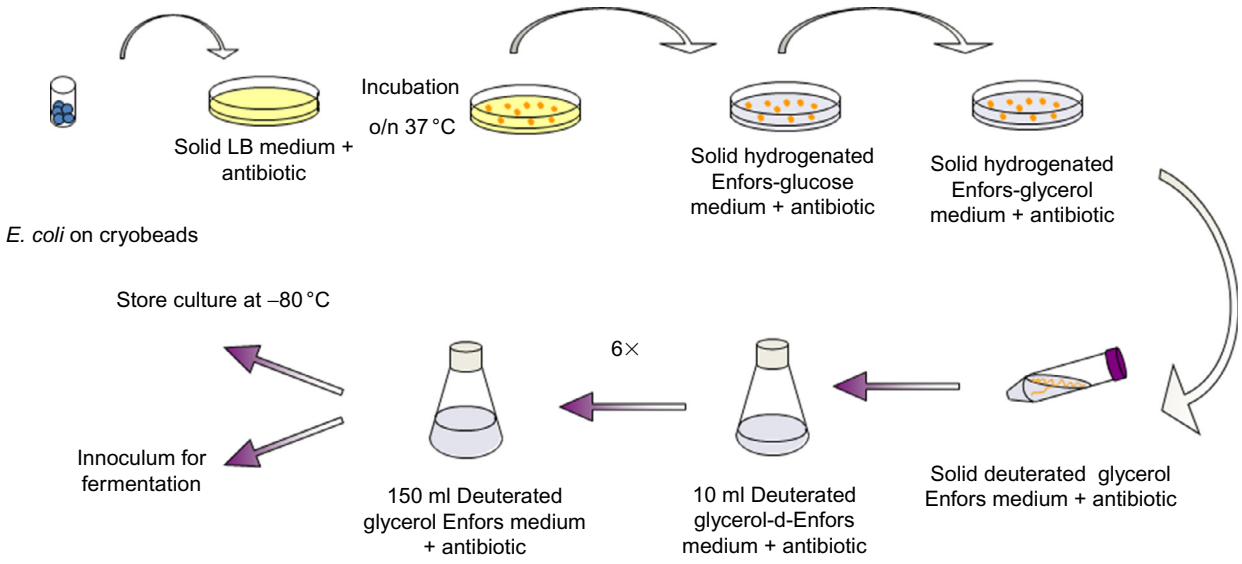


Figure 7

Schematic representation of the adaptation protocol to deuterated minimal medium for *E. coli* cells, developed by the Life Sciences Group at ILL (Haertlein *et al.*, 2016).

In this study, the *P. pastoris* expression system was initially used for the expression of perdeuterated HEWL (D-HEWL). After several unsuccessful attempts at matching the previously reported protein yields of approximately 11 mg per liter of culture, alternative strategies for recombinant protein expression were considered. The *E. coli* system was selected, knowing that HEWL would be produced in inclusion bodies, due to lack of disulphide-bond formation. Tests were first conducted in hydrogenated conditions, to ensure that the expression system yielded significant amounts of recombinant protein. Later, a protein purification approach had to be developed consisting, first, of the removal of contaminants from the inclusion body suspension, and then of the extraction of considerably pure protein for subsequent *in vitro* refolding. The works of Batas *et al.* (1999) and Palmer & Wingfield (2004) were extremely useful in the design of the purification protocol used for D-HEWL. This protocol is described in detail in Manuscript 1 of this thesis.

2.2. *In vitro* protein refolding

The *thermodynamic hypothesis*, formulated by Anfinsen. C. B. (1973), states that the three-dimensional structure of a native protein in its physiological conditions is governed by a search for its lowest possible Gibbs free energy state. This postulation suggests that the only information required to achieve a native protein fold is embedded in its amino acid sequence, when in its physiological environment. The hypothesis was validated and demonstrated by extensive work of Anfinsen and others on bovine pancreatic ribonuclease and *staphylococcal* nuclease (Anfinsen *et al.*, 1962; Haber & Anfinsen, 1962; Givol *et al.*, 1964; Epstein *et al.*, 1971). These findings have paved the way for the application of *in vitro* refolding where reduced solubility of recombinant proteins led to aggregation in the form of inclusion bodies, and perhaps more importantly enlightened the scientific efforts to predict *in silico* the three-dimensional structures of challenging proteins which have been elusive to structural experimental methods.

There are essentially two methods that have been used to refold proteins *in vitro*, namely dilution and dialysis. Both techniques aim at reducing the concentration of denaturing agents to levels which no longer inhibit protein-protein interactions, allowing the formation of secondary and tertiary structures. An obstacle to the refolding process is usually high concentration of protein, which leads to nonspecific interactions, aggregation and subsequent precipitation. To avoid the occurrence of this phenomenon, which hinders protein refolding yields, very low concentrations of protein are usually employed in refolding experiments. Additionally, the choice of refolding buffer is essential to allow correct protein refolding and it is often the experimental parameter most difficult to optimize, due to the numerous variables at play (*e.g.*, pH, ionic charges, salt concentration, oxidizing/reducing agents, detergents, ligands). A plethora of studies have been conducted over the years in an attempt to rationalize and facilitate the discovery of appropriate protein refolding conditions (Armstrong *et al.*, 1999; Vincentelli *et al.*, 2004; Willis *et al.*, 2005; Cowieson *et al.*, 2006; Biter *et al.*, 2016). An interesting approach to protein refolding *in vitro* is the high-throughput method reported by Biter *et al.* (2016). Taking the concept that the native protein structure should correspond to the lowest Gibbs free energy state, the authors have employed differential scanning fluorimetry (DSF), which is a technique that measures protein thermal stability, to screen for the conditions most likely to yield a correctly refolded protein. By using this approach, they were able to rapidly screen numerous refolding buffer solutions for several model proteins (*e.g.*, glucose isomerase, lysozyme and carbonic anhydrase) and challenging cases, from bacterial inclusion bodies, such as irisin and fibroblast growth factors 19 and 21. Methodologies of protein refolding from solubilized inclusion bodies were described in detail by Burgess (2009).

Batas & Chaudhuri (1996) have taken a different approach to the dilution method of protein refolding by showing that it was possible to refold HEWL and bovine carbonic anhydrase in-column, through size-exclusion chromatography (SEC). This method has the advantage of combining the

dilution/desalting of the unfolded protein from the denaturing buffer into the refolding solution, with the separation of monomeric and aggregated fractions of protein. Furthermore, the authors demonstrated that active protein could be recovered by performing the experiment even at very high protein concentrations (*i.e.*, up to 80 mg/mL). While in the batch dilution method, the refolding yield decreases dramatically with the increase in protein concentration, in the procedure described by Batas & Chaudhuri (1996), the yields seem to be less affected. Further studies showed that this approach could also be applied to refold recombinantly expressed HEWL, from *E. coli* inclusion bodies (Batas *et al.*, 1999). Their work also included a centrifugation-based process for inclusion body washing and unfolded protein purification through SEC in denaturing conditions, which paved the way for our efforts in producing refolded D-HEWL. The refolding approach used for D-HEWL was also adapted from Batas *et al.* (1999), although important variations were made regarding the type of SEC column used and elution flowrate. Interestingly, it was found that this method resulted in larger protein yields for D-HEWL, compared to the hydrogenated variant. It was also noted that the refolding in-column was less effective in D_2O than in H_2O. This observation could be explained by differences in folding dynamics, since D has twice the mass of H, and H/D comprise about 50% of all atoms in a protein. Hence, the macromolecular isotope effect can have an important role in the protein refolding process. Furthermore, the solvent isotope effect can also be a contributing factor to differences in refolding, due to the higher viscosity of D_2O (1.25 cP at 20° C) compared to H_2O (1.00 cP at 20° C) (Millero *et al.*, 1971) and to variations in solvation. Since larger protein yields were achieved in H_2O, this approach was implemented throughout this study, although it was likely to result in the caging of a few H atoms in exchangeable positions which upon refolding would become inaccessible to solvent. The refolding protocol used in this study for both perdeuterated and hydrogenated variants of recombinant HEWL is reported in Manuscript 1. Additionally, a brief comparison of the refolding SEC elution profiles for both proteins is presented in Manuscript 2.

2.3. Biophysical characterization

Proteins possess certain biophysical features which are intimately related with their structure and biological function. Properties such as thermal stability and enzymatic activity can provide relevant information regarding protein folding. As discussed previously, a native protein structure is usually associated with its global Gibbs free energy minimum, meaning that techniques able to probe protein thermal stability can inform about different protein folding states. Moreover, macromolecular and solvent deuteration are known to impact the thermal stability of a protein by promoting changes in structure and dynamics. Similar insights can also be extracted from enzymatic activity assays, due to the relationship between protein structure, dynamics and function.

2.3.1. Mass spectrometry

Mass spectrometry (MS) is a versatile biophysical technique that has been employed in studies of protein folding (Miranker *et al.*, 1993), protein-protein interactions (Smits & Vermeulen, 2016), and protein labelling (Yee *et al.*, 2016; Kita & Morimoto, 2016). This method allows the separation of several species from a sample according to their molecular masses and the charges gained during the ionization process. The methodology employed depends strongly on the components of the analyzed sample and on the scientific question posed.

Kita & Morimoto (2016) have employed matrix assisted laser desorption ionization – time-of-flight (MALDI-TOF) MS to assess the H/D exchange in buried regions of HEWL structure. By dissolving HEWL powder in D_2O buffer solution and promoting unfolding and refolding of the protein structure through addition of strong base/acid and/or through heating and cooling, the authors demonstrated that H positions otherwise inaccessible to solvent were exchanged by D. In the case where the HEWL solution was subjected to acidification, heated to 80° C, cooled down and later exposed to a basic pH, an increase in mass of 20 Da was recorded, suggesting the H/D exchange in 20 positions.

Electrospray ionization (ESI) MS can be routinely used to assess the labelling effectiveness of recombinant production of perdeuterated molecules (Hazemann *et al.*, 2005; Manzoni *et al.*, 2016; Gajdos *et al.*, 2020). A straightforward procedure to assess the deuterium-labelling of a recombinantly expressed protein, which was employed in this project, is to allow full back-exchange in H_2O solution through dialysis. In this way, partial H/D back-exchange is avoided during the MS experiment and the resulting masses account exclusively for D in non-exchangeable positions (*i.e.*, bound to C). Partial back-exchange would occur if H_2O were present in the preparations used in the MS analyzes. Nevertheless, if the target protein is expressed in perdeuterated conditions, all exchangeable H positions (*i.e.*, bound to N, O, S) will be occupied by D. Therefore, if the MS analyzes shows full deuteration in non-exchangeable positions, the protein can be considered as successfully perdeuterated.

2.3.2. Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a method of measuring the thermal stability of a protein by following its denaturation and/or refolding process through an extrinsic or intrinsic fluorescence signal. This signal can come from a dye which binds hydrophobic amino acids (extrinsic) or from the intrinsic fluorescence of tryptophan residues. These amino acids due to their hydrophobic nature are usually buried in the hydrophobic core of a protein, meaning that their fluorescence signal is readily quenched by neighboring residues. However, upon denaturation, these residues become exposed to the solvent molecules and no longer protected by the protein fold. Therefore, the denatured protein will have a stronger fluorescence signal. Melting temperature (T_m) is the parameter used to compare the thermal stability of different proteins or of the same protein in different conditions (*i.e.*, buffer compositions). This value corresponds to the inflection point of the protein denaturing or renaturing curves, where 50% of the molecules in solution are in an unfolded state. A simple and widely used way of visualizing protein melting curves and extracting the T_m values is to calculate the first derivative of the fluorescence signal over time – the T_m will correspond to the maximum of the curve. The thermal stability of a protein depends strongly on the solution in which it is dissolved. Variables such as pH, salt concentration, and the presence of additives can cause significant changes to protein stability.

An interesting application of DSF is its use in macromolecular crystallography, where the choice of protein buffer is of pivotal importance in the efforts of protein crystallization (Geders *et al.*, 2012; Reinhard *et al.*, 2013; Deller *et al.*, 2016; Gao *et al.*, 2020). Additionally, DSF can also provide valuable information regarding protein-ligand interactions, since these are often associated with variations in protein thermal stability (Forneris *et al.*, 2009; Holdgate *et al.*, 2010; Bai *et al.*, 2019; Ramos *et al.*, 2019). This method and, as alternative, differential scanning calorimetry (DSC) have also been widely applied to the study of protein deuteration (Berns, 1963; Hattori *et al.*, 1965; Brockwell *et al.*, 2001; Koruza *et al.*, 2018; Nichols *et al.*, 2020), where it may be possible to deconvolute the contributions of the macromolecular and the solvent deuteration effects.

In this study, DSF was employed in the biophysical characterization of different HEWL variants. In Manuscript 1, perdeuterated HEWL obtained from the *E. coli* and *P. pastoris* systems were compared with hydrogenated HEWL (H-HEWL) from *Gallus gallus* to assess the effects of refolding and perdeuteration on protein thermal stability. Moreover, in Manuscript 2, a direct comparison between D-HEWL and H-HEWL, recombinantly expressed in *E. coli* and posteriorly refolded, was performed. This work has allowed the deconvolution of the effects on protein thermal stability of *in vitro* refolding, perdeuteration, and solvent H/D substitution.

2.3.3. Enzymatic activity assays

A simple method to measure HEWL activity was developed by Shugar (1952) with the purpose of studying the ultra-violet inactivation of this enzyme. This procedure entails the use of a *M. lysodeikticus* cell suspension in M/15 phosphate buffer pH 7.1, to which a solution of HEWL is added. The enzymatic reaction is followed by measuring the decrease in optical density at 450 nm with a spectrophotometer. Hence, one activity unit corresponds to a decrease in absorbance of 0.001 per minute, in a specific condition. The author discussed the effects of two critical variables on assay reproducibility, namely pH and temperature. The method was designed for neutral and basic solutions, where HEWL activity is lower and thus allow monitoring over a few minutes. Additionally, it was noted that the results were highly variable with minor changes in temperature – the author estimated a temperature coefficient (Q_{10}) greater than 3 (*i.e.*, a temperature variation of 0.2° C could lead to measurement errors of 3% or more).

Alternative methods have been described to measure HEWL activity, particularly through fluorescence assays, which provide higher sensitivity when probing solutions with low concentrations of enzyme. An example of such protocols is the EnzCheck® lysozyme assay kit (Molecular Probes). In this assay, a *M. lysodeikticus* cell suspension is used where the bacteria cell wall was highly labelled with fluorescein, to a degree which causes quenching of the fluorescent signal. Upon HEWL degradation of the cell wall, fluorescence is released at 518 nm, from excitation at 494 nm. This approach seems valuable for situations where high sensitivities are required and also where a direct measurement of the activity is advantageous (*e.g.*, in conditions that can be problematic for cell viability). An obvious application for this method is the probing of HEWL activity in D_2O, since heavy water can pose difficulties to cell viability and interfere with activity measurements from cell suspension turbidity.

In the present study, the fluorescence-based assay was initially used to assess the activity of different HEWL variants. However, after several measurements, the lack of reproducibility was problematic, leading to significant errors in the estimated protein activities. This issue is likely related to the high sensitivity of the method and to the fact that it is better suited for impure samples with minor concentrations of HEWL, rather than pure samples at high HEWL concentrations. Therefore, the turbidity-based method was later employed, yielding reproducible results, as reported in Manuscripts 1 & 2. Unfortunately, meaningful activity measurements could only be obtained in H_2O solution, while in D_2O the decrease in turbidity was extremely rapid. These observations are probably linked with the lower cell viability in D_2O, which did not allow direct probing of HEWL activity.

2.4. Macromolecular crystallography

Macromolecular crystallography has been the main contributor to the understanding of protein structure, namely through X-ray diffraction. The disparity between the number of structures reported from each method is due to several factors, such as maturity of the method, technological advancements, limitations inherent to the technique. For instances, crystallography was initially applied to the structure determination of small molecules, in 1913 (Bragg, 1913), and later yielded the first structure of an enzyme, HEWL (Blake *et al.*, 1965). Meanwhile, the first protein structure obtained by NMR was described in 1985 (Williamson *et al.*, 1985), and the first virus structures by cryo-EM were reported in 1984 (Adrian *et al.*, 1984). In recent years, the cryo-EM revolution has been taking place due to earlier advancements in detector technology (Faruqi *et al.*, 2003) and sample preparation (Dubochet & McDowall, 1981), which allowed the transition from low resolution data to high resolution structures, and most recently to atomic resolution (Nakane *et al.*, 2020). On the other hand, NMR was considered as a promising alternative to X-ray crystallography, since high MW structures became more accessible, because of isotope labelling and multidimensional NMR techniques (Clore & Gronenborn, 1994; Sattler & Fesik, 1996). However, the MW limitation of NMR proved to be a systematic hurdle to the widespread use of the technique.

As mentioned in the Introduction section, neutron crystallography possesses unique capabilities to probe not only protein structure but also dynamics, by estimation of ADPs. Nevertheless, this method entails a series of challenges in order to retrieve the aforementioned information. Protein perdeuteration is the first hurdle of neutron crystallography, since the expression of recombinant perdeuterated material is expensive and sometimes associated with low yields (Haertlein *et al.*, 2016). As described previously, isotope substitution can lead to variations in protein biophysical properties, and in some cases structural changes. A consequence of the macromolecular and solvent isotope effects on protein stability is often the need to optimize the protein crystallization conditions (Hazemann *et al.*, 2005; Liu *et al.*, 2007; Petit-Haertlein *et al.*, 2009; Koruza *et al.*, 2018), leading occasionally to significant changes in the crystallization buffer used. Another challenge in neutron crystallography is growing relatively large crystals compared to those used in X-ray crystallography, to compensate for the reduced neutron flux available. Neutron diffraction data collection is also distant from the current practices in X-ray crystallography, especially when using synchrotron radiation. While X-ray diffraction data collection at a synchrotron is usually achieved within minutes, with fluxes of ~10^{16} photons/cm^2/s , neutron data is routinely recorded during several days, using fluxes of ~10^6 to 10^8 neutrons/cm^2/s (Blakeley, 2009). Finally, recent developments of crystallographic software have resolved previous limitations and facilitated model refinement from neutron data (Afonine *et al.*, 2010; Gruene *et al.*, 2014).

2.4.1. Protein crystallization

The triclinic crystal form (P1) of HEWL was shown to produce the higher quality diffraction from X-rays (*e.g.*, Wang *et al.*, 2007), rendering this crystal system has a good candidate for the present atomic resolution NMX study. Triclinic HEWL can be obtained by using sodium nitrate as precipitant and by performing a precipitation step at 4° C followed by storage at room temperature. This methodology was introduced by Vidal *et al.* (1999) in a HEWL crystallization study where it was noted that the monoclinic crystal form prevails in the absence of the precipitation step. The observation of monoclinic HEWL crystals in this condition without temperature variation was made by Steinrauf (1959) and Sieker (1988). Later, Legrand *et al.* (2002) explained this phenomenon of crystal system selectivity through solubility differences, and their temperature dependences, observed between the monoclinic and the triclinic forms (Fig. 8). While at 4° C the HEWL crystal systems present similar solubilities, at temperatures between 18° and 25° C, the monoclinic form shows a higher solubility compared to that of the triclinic, meaning that the crystals are less stable and tend to dissolve. These results support the hypothesis that both monoclinic and triclinic crystalline nuclei are formed in the presence of sodium nitrate during the precipitation step at 4° C. Afterwards, when the temperature is raised above 18° C, the monoclinic nuclei tend to dissolve, while the triclinic form prevails (Fig. 9d).

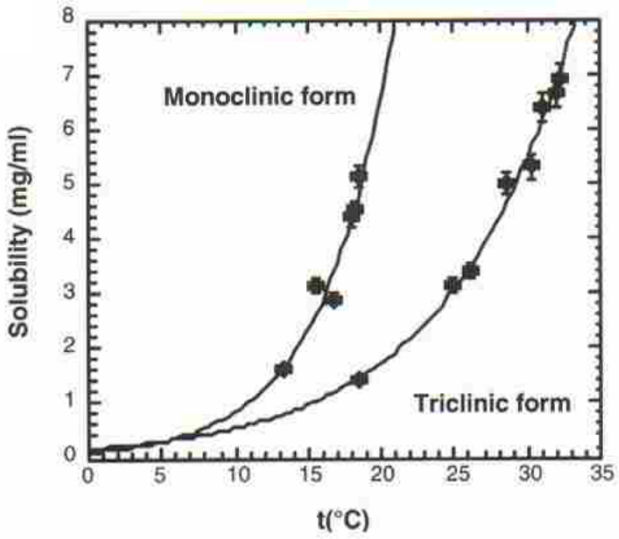


Figure 8

HEWL solubility as a function of temperature for 0.2 M sodium nitrate, as reported by Legrand *et al.* (2002). The differences in the solubility curves of the monoclinic and triclinic HEWL forms explain the selective phenomenon behind the temperature variation method. Figure reproduced with permission of the International Union of Crystallography.

Macromolecular and solvent isotope effects can influence significantly protein crystallization conditions. This observation is directly linked with changes in protein stability and solubility upon perdeuteration and solvent H/D substitution. Therefore, the optimization of crystallization conditions is

often an additional challenge in neutron crystallography, where studying the perdeuterated form of a protein is advantageous. For example, in the work of Hazemann *et al.* (2005), the crystallization conditions used for perdeuterated human aldose reductase were closely similar to those reported for its hydrogenated variant. However, a decrease in solubility of the perdeuterated form was noted under these conditions, leading to the formation of amorphous precipitate. The authors surpassed this hurdle by employing microseeding techniques to yield high quality crystals of 0.15 mm^3. Similarly, Petit-Haertlein *et al.* (2009) performed a grid search for the optimal crystallization conditions of perdeuterated anti-freeze protein in D_2O, while using the conditions for the hydrogenated form as starting point. The optimized conditions varied relative to that for the hydrogenated variant by decreasing the temperature (from 295 to 285 K), changing the pH from 4.5 to 4.8 (pD 5.2), and decreasing the concentrations of both sodium acetate buffer (from 50 to 20 mM) and ammonium sulfate precipitant (from 1.5 to 1.1 M). These variations suggest a decrease in solubility of the perdeuterated form of anti-freeze protein, however, the different temperatures and pH used hinder the drawing of definitive conclusions. On the other hand, the study performed by Koruza *et al.* (2018), on different hydrogenated and partially deuterated variants of human carbonic anhydrase, showed that identical hydrogenated crystallization conditions can yield crystals of both hydrogenated and deuterated proteins. Nevertheless, differences in crystallization behavior were evident due to the macromolecular isotope effect. The deuterated crystals obtained were usually smaller than those of their hydrogenated counterparts. Only in one case, the crystallization condition tested yielded good quality crystals of deuterated protein compared to amorphous precipitate of hydrogenated protein.

More often than not crystallization experiments present problems in yielding high quality crystals, which are associated with the plethora of variables involved in protein crystallization (*e.g.*, protein intrinsic disorder, protein stability, choice of crystallization technique, temperature, precipitant, pH, protein and precipitant concentrations). In some cases, experimental conditions can lead to the precipitation region of the phase diagram, resulting in amorphous protein aggregation or crystalline precipitate. On the other hand, crystalline nucleation can occur too rapidly causing the formation of microcrystals or twinned crystals with poor diffraction qualities. The inverse scenario is also possible, when nucleation kinetics are slow, and crystals take weeks or months to appear. Additionally, in certain situations, as in many neutron crystallographic studies, the desired crystal volumes are significantly larger than the ones obtained in initial conditions. Attempting to solve the aforementioned difficulties in protein crystallization, various seeding approaches have been developed over the years (Stura & Wilson, 1990; Bergfors, 2003; D'Arcy *et al.*, 2003, 2014). In general, seeding comprises the provision of nucleation sites which can support crystal growth in supersaturated conditions and help avoid the rate-limiting step of spontaneous nucleation (Stura & Wilson, 1990; Bergfors, 2003).

Microseeding can be employed in neutron crystallography to aid the formation of crystals of perdeuterated proteins, overcoming the commonly observed changes in protein stability and solubility. In this study, initially crystals of H-HEWL in deuterated buffer were crushed into microscopic seeds

which were later used to induce the growth of perdeuterated crystals. Using this approach, crystal forms can be selectivity crystallized in conditions otherwise yielding polymorphic crystals, as in the case of HEWL nitrate. This technique is denoted cross-seeding and it has been often implemented in mutant variants, heavy atom derivatives, and complexes with ligands (Sanishvili *et al.*, 1994; Thaller *et al.*, 1981; Hassell *et al.*, 2007). A screen of precipitant and protein concentrations is, usually, required in order to probe the phase diagram and to identify the undersaturated and supersaturated zones, and subsequently the metastable region, where crystal growth is possible (Stura & Wilson, 1990; Bergfors, 2003). Optimization of the seed stock used is also recommended since a lower dilution stock might result in crystal showers, while an excessive dilution may not contain sufficient seeds (Stura & Wilson, 1990; Bergfors, 2003; D'Arcy *et al.*, 2014). In Manuscript 1, a detailed description of the crystallization process used to obtain triclinic D-HEWL is available. Identical crystallization conditions were also used to produce triclinic crystals of recombinant refolded H-HEWL, as noted in Manuscript 2. As expected, D-HEWL could not be crystallized in the triclinic form using the temperature variation step, reported by Vidal *et al.* (1999) (Fig. 9). Additionally, the recombinant refolded H-HEWL was also unable to crystallize in similar conditions. Surprisingly, only through microsseding in microbatch under oil, high quality crystals could be produced (Fig. 10a).

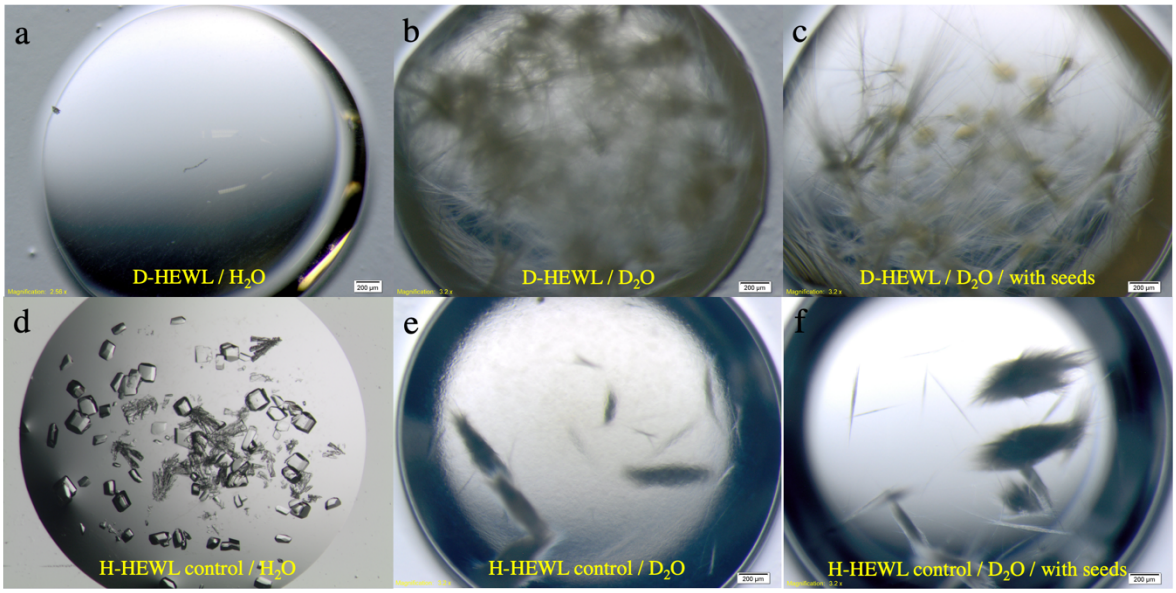


Figure 9

General crystallization behavior of D-HEWL in sodium nitrate and sodium acetate pH 4.5 (a-c) compared to commercially available H-HEWL (d-f), as a control. In the triclinic condition, the temperature variation step (Vidal *et al.*, 1999) was used in batch-like conditions. H_2O and D_2O solutions were tried, and H-HEWL microseeds were also added in deuterated conditions.

On the other hand, macroseeding is a routine practice in neutron crystallographic studies, since relatively large crystal volumes are required, compared to current habit in X-ray crystallography. Different approaches can be exploited to grow large crystals and their success seems to depend strongly on the system studied. Traditionally, macroseeding would comprise the transfer of a good quality crystal to a new crystallization drop with fresh protein and precipitant solution. This drop would then undergo equilibration and the protein in solution would be consumed to grow the existent crystal. However, when several "feeding cycles" are necessary to significantly increase crystal volume, transferring crystals to new drops can be detrimental. These successive manipulations can potentially lead to crystal defects due to cracking and also produce secondary nucleation. Alternative solutions can be implemented, where crystal handling is reduced by replacing only the surrounding crystallization solution with fresh protein and precipitant mixture, or even just by adding fresh protein to the equilibrated crystallization drop. In the present work, crystals were grown to mm^3 volumes using macroseeding (Fig. 10) and by replacing the crystallization solution with a fresh protein and precipitant mixture, without manipulating the crystal. Occasionally, when excessive secondary nucleation occurred or smaller crystals were found growing attached to the main crystal in the drop, the crystal of interest would be carefully manipulated with tools and transferred to a fresh crystallization drop.

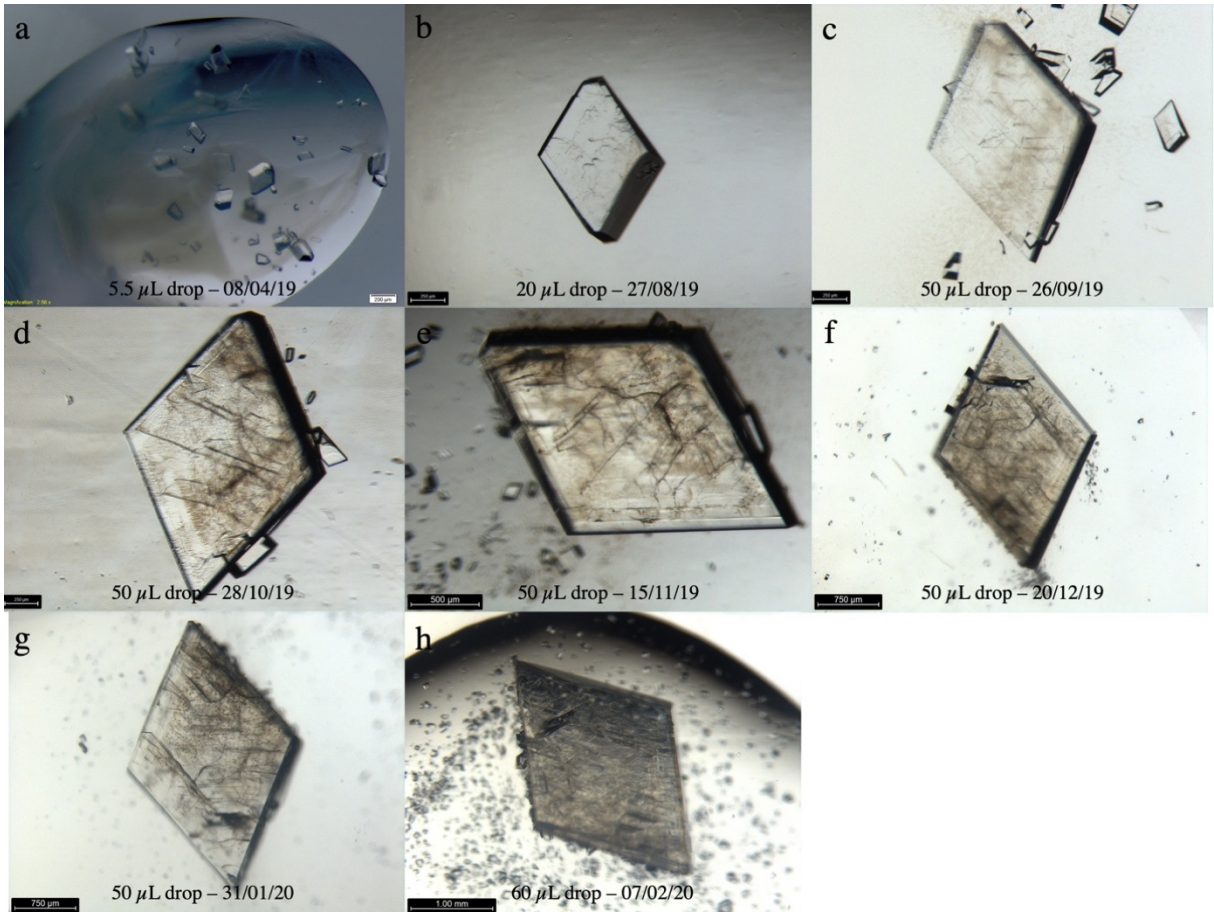


Figure 10

Crystal growth of a triclinic D-HEWL crystal using macroseeding in batch-like conditions, at 18 ° C. Typical crystallization drop of triclinic D-HEWL in microbatch under oil, after microseeding and equilibration for about 1 week (a). Growth progress of the D-HEWL crystal used for RT neutron diffraction data collection (b-h). The crystal in (b) was first macroseeded 1 month prior, from the initial 5.5 μL crystallization drop to a 20 μL solution. The crystallization solution, since (b) to (h) was generally replaced by fresh protein and precipitant mixture after 1-2 weeks of equilibration.

2.4.2. X-ray diffraction

To obtain the atomic resolution and complete X-ray diffraction datasets necessary for the success of the present work, high quality crystals and synchrotron radiation are essential. From previous studies, it was known that triclinic HEWL crystals possess the desired diffracting quality (Walsh *et al.*, 1998; Wang *et al.*, 2007). Additionally, synchrotron beamlines have proven for many years to be capable of providing the means (*i.e.*, X-ray beam energies, instrument geometry, detector technology) to collect complete atomic and ultra-high resolution X-ray diffraction data (Dauter *et al.*, 2010). In the case of the triclinic P1 system, it is evident that achieving high data completeness is challenging. This is a consequence of the absence of symmetry in the P1 unit cell, leading to the existence of a significant blind region in reciprocal space if the data collection is performed only by rotating the crystal over one axis (commonly named φ, phi). Therefore, an additional axis rotation must be employed to cover this region of reciprocal space. Currently, there are many macromolecular crystallography synchrotron beamlines equipped to allow the rotation of the crystal over a different axis, usually denoted as κ (kappa). By performing a 180° φ scan in two significantly different κ orientations, it is usually possible to collect complete X-ray diffraction data on a single triclinic HEWL crystal. Other hurdles of working with this crystal system are achieving high diffraction data redundancy and avoiding radiation damage, since the data collection strategy consists of two φ scans. The latter is particularly relevant in room temperature synchrotron X-ray diffraction experiments. Collecting data from multiple isomorphous crystals, grown in identical conditions, is a suitable strategy to overcome the aforementioned hurdles.

In cryocrystallography experiments, the choice of cryoprotectant solution and soaking strategy can dictate the quality of the diffraction data collected (Senda *et al.*, 2016). A cryoprotectant is an additive to the solution in which a crystal is flash-cooled helping to avoid nucleation of crystalline ice. Protein crystals comprise on average approximately 50% of solvent (Matthews, 1968), mostly in disordered solvent channels, which means that the formation of ice can disrupt crystalline order, and lead to loss of diffraction quality (Pflugrath, 2015). Moreover, the presence of ice around a sample submitted to X-rays results in their contribution to diffraction, showing characteristic powder-like diffraction (commonly named ice rings) (Pflugrath, 2015). These ice rings introduce errors in the estimate of neighboring reflections intensities and may lead to their complete exclusion from the diffraction dataset. The use of a cryoprotectant compound, such as glycerol, ethylene glycol, PEG 400, 2-methyl-2,4-pentanediol, typically at concentrations of 25-50% (v/v) (w/v for PEG) allows the vitrification of the sample and prevents crystalline ice formation (Garman, 2003; Pflugrath, 2015). Nevertheless, soaking a protein crystal in a solution with these cryoprotecting agents can be detrimental to its diffraction quality, since these molecules enter the crystal solvent channels and may disturb the crystal packing. An often-observed consequence of these perturbations is the increase in crystal mosaicity (Harp *et al.*, 1998). Despite the hurdles of cryocrystallography, this technique has been common practice in X-ray crystallography because of its benefits related to the prevention of radiation

damage. It has allowed the exploitation of the high X-ray energies produced by synchrotron rings in improving the quality of the diffraction data collected and in facilitating the study of smaller and poorly diffracting samples.

On the other hand, room temperature X-ray crystallography has become a rare practice at synchrotron beamlines. Several factors contribute to this trend, such as the difficulties in transporting crystals in solution from laboratories to synchrotron facilities, the time-consuming and cumbersome sample mounting at beamlines, and, most importantly, the loss of crystal quality caused by radiation damage (Fischer, 2021). However, room temperature X-ray data collection can be advantageous, since the experimental conditions are closer to the protein physiological environment and it has been shown that additional biologically relevant information can be retrieved, compared to cryogenic studies (Fraser *et al.*, 2011; Russi *et al.*, 2017; Fischer, 2021). Fischer (2021) presented some interesting solutions on how to overcome the aforementioned hurdles of room temperature X-ray crystallography, from sample transport to mounting, data collection and processing. While many synchrotron beamlines possess humidity controllers to avoid crystal dehydration during data collection at room temperature, other techniques can be employed to the same effect. Before cryocrystallography became routine, crystals were being measured inside sealed quartz or borosilicate glass capillaries which contained crystallization solution to maintain the humidity levels around the sample. Nowadays, it can be challenging to mount microcrystals in capillaries and also to accommodate them in current beamline specifications, which are optimized for short-length magnetic pins. Mac Sweeney & D'Arcy (2003) demonstrated that standard nylon loops can be used to mount a crystal at room temperature by covering it with a short capillary and sealing it with plasticine. At the end of the capillary, crystallization solution is added to prevent crystal dehydration. The disadvantage of using quartz or glass capillaries is their scattering and absorption of X-rays, contributing to the diffraction experiment background noise (Fischer, 2021). As an alternative, MiTiGen has produced a polyester tubing which can perform the same function as the capillaries without significant X-ray absorption (Kalinin *et al.*, 2005). Another method is the use of oils, such as silicon oil or paraffin oil, to cover the crystals and impede dehydration (Pflugrath, 2015), although these oils can also produce undesired diffraction background noise.

2.4.3. Neutron diffraction

For the success of the present work, obtaining complete atomic resolution neutron diffraction data was paramount. D19, at ILL, is an instrument with unique capabilities, suitable for the collection of the desired neutron data. It is a thermal neutron diffractometer which can operate at monochromatic neutron wavelengths between 0.8 and 2.4 Å. The instrument monochromator is positioned close to the reactor core, which results in the availability of a relatively high neutron flux (up to ~10^8 neutrons/cm^2/s). D19 possesses a large position-sensitive detector, working at high gas pressure (5 atm ^{3}He + 1 atm CF_4). This position-sensitive detector covers 120° horizontally and 30° vertically, allowing an extensive coverage of the reciprocal space, essential to the collection of complete data. The sample is held in a C-shaped Eulerian cradle and can be rotated relative to the incident neutron beam over three different axes (phi φ, chi χ, and omega ω). These various possible sample orientations facilitate the collection of complete data by minimizing the extent of blind regions, even for triclinic systems. The characteristics of D19 enable the study of large chemical systems and small proteins, data collection at very high resolution, and allow an extensive coverage of the reciprocal space. Nevertheless, collecting neutron diffraction data at D19 is significantly different from measuring X-ray diffraction at synchrotron beamlines. The neutron flux available at D19 is approximately 8 orders of magnitude lower than that of common synchrotron beamlines. This discrepancy has to be compensated by using significantly larger crystal volumes, sample perdeuteration, and longer exposure times. While collecting X-ray data takes in most cases less than a minute, to collect high quality neutron diffraction data, several days are required. This problem is further emphasized when studying a crystal system such as the triclinic P1. Acquiring a complete neutron diffraction dataset from a triclinic protein crystal in D19 can take up to 12 days. Such a large time period for one data collection poses several problems in terms of maintaining a consistent experimental setup and sample environment. Typically, collecting neutron data from a triclinic protein crystal at D19 entails the acquisition of 25 individual ω scans, in static stepwise frames. The design of these scans is dictated by geometric constraints inherent to the instrument setup, where χ can be moved from 90° to 180°, and the ω range is from -25° to 39°, at χ 90°-105°, and -30° to 49°, at χ 130°-180°. The axis of rotation parallel to the sample holder is named φ, and can rotate 360°.

Room temperature neutron experiments at D19 are facilitated by the use of capillaries to mount protein crystals (Fig. 11). The capillaries are usually made of quartz or borosilicate glass which have negligible neutron absorption. At the end of the capillary, crystallization solution is added to maintain the humidity levels and avoid crystal dehydration. The crystal is usually placed in-between quartz wool with the intent of fixing its position throughout the data collection. Finally, the other end of the capillary is sealed with epoxy glue to avoid evaporation. The capillary is then secured onto a goniometer head with plasticine. Immediately before neutron data acquisition, the goniometer head with the mounted crystal is attached to the Eulerian cradle and the crystal is aligned with the neutron beam. Typically,

several crystals are prepared and mounted to be tested on the neutron beam, in order to find the best diffracting sample. The possibility of testing different crystals is a distinctive advantage of room temperature neutron diffraction experiments, since the samples are stable inside the capillaries and no radiation damage arises from the neutron beam.

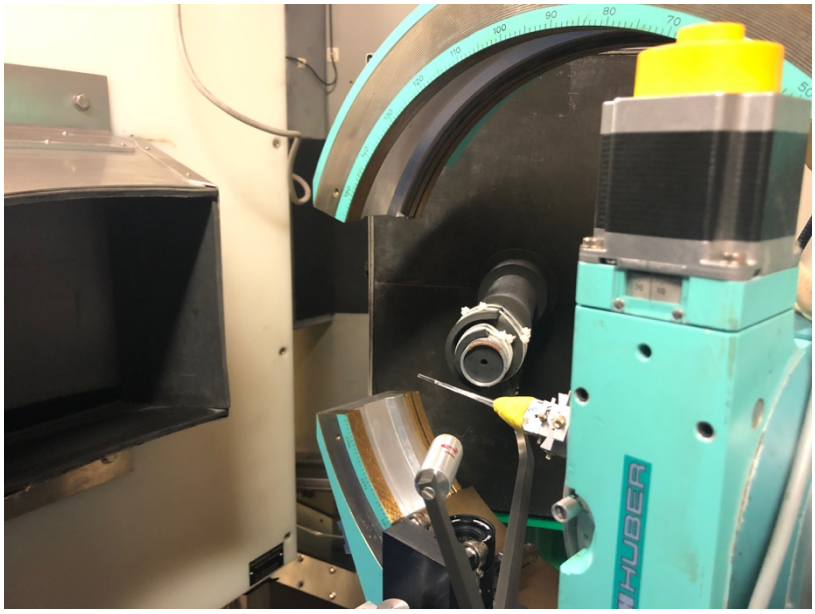


Figure 11

Picture of the D-HEWL crystal mounted in capillary and attached to the C-shaped Eulerian cradle at D19, which was used in the neutron data collection at room temperature (Manuscripts 3 & 4).

On the other hand, cryogenic experiments with protein crystals in D19 can be extremely challenging. Indeed, neutron cryocrystallography is an uncommon practice, since neutrons do not cause visible radiation damage because they are non-ionizing radiation contrary to X-rays. Therefore, neutron crystallographic studies, in general, are conducted at room temperature avoiding the hurdles associated with crystal cooling (*e.g.*, ice formation, degrading crystal quality). Kriminski *et al.* (2003) have investigated the impacts of flash-cooling protein crystals. The authors have listed the factors involved in successfully cooling crystals from most to least important as follows: 1) crystal solvent content and solvent composition; 2) crystal size and shape; 3) amount of residual liquid around the crystal; 4) cooling method (liquid plunge versus gas stream); 5) choice of gas/liquid; 6) relative speed between cooling fluid and crystal. Later, Chinte *et al.* (2005) demonstrated that sample size plays a crucial role in successful sample vitrification. These observations illustrate the complexity of flash-cooling crystals for neutron crystallographic studies, where samples can measure several mm^3. Previous work, related to this project, showed that soaking a large HEWL crystal during several minutes in increasingly more concentrated glycerol solutions was very effective. Moreover, it was found that the required final concentration of cryoprotectant used in flash-cooling was lower for large crystals compared to that for

crystals of sizes typically used in X-ray experiments. This seems to be related to the fact that the majority of the solvent when cooling small crystals is surrounding the crystal, while in large crystals it is located inside the crystal, in solvent channels. An interesting observation has been made by Weik *et al.* (2001) relative to the behavior of solvent in protein crystals at cryogenic temperatures. The authors reported that solvent in large channels behaves like disordered bulk water, while in small channels ice nucleation is prevented by protein-water interactions. Additionally, the crystals used in neutron studies are usually too large and heavy to be mounted in conventional nylon loops, used routinely in X-ray crystallography. To overcome this, Romoli *et al.* (2014) have developed a SPINE-compatible 'carboloop' capable of sustaining large protein crystals for both neutron and X-ray diffraction experiments. These 'carboloops' are made from vitreous carbon and shaped by laser cutting. Posteriorly, the 'carboloops' are glued onto 1.5 mm aluminum tubes which are fixed to standard SPINE magnetic bases.

To obtain 100 K neutron diffraction data, for the present study, carbon loops were manufactured using a similar methodology to that described by Romoli *et al.* (2014) (Fig. 12a). These loops were utilized to mount the large D-HEWL crystals onto D19 goniometer, after soaking in fully deuterated cryoprotectant solution. The setup at D19 was covered by a plastic sheet meant to preserve the humidity levels around the sample and cryogenic system (Fig. 12b). Additionally, a dish with dry silica beads was placed next to the Eulerian cradle, and periodically changed, to absorb humidity. The major concern during data collection was the formation of ice around the sample and on the carbon loop (Fig. 12c). Due to the long duration of the experiment (~12 days), significant amounts of ice would usually form in a matter of hours. For this reason, the standard 25 individual ω scans, normally measured for this crystal system at D19, had to be shortened. Half-scans were thus performed and, whenever suitable, ice would be manually removed in-between scans. Certain sample orientations would promote the formation of ice whereas others would result virtually in no ice being formed, depending on the goniometer orientation relative to the cryo-stream. The latter cases were usually at χ 130°-180° and negative ω values, while the orientations where significant ice would build-up were at χ 90°-105° and positive ω values. After observing these tendencies, a data collection strategy was conceived which ensured the following aspects:

i) Consecutive scans, without ice removal, could be performed for χ 130°-180° and negative ω values.

ii) Scans at χ 90°-105° and positive ω values had to be monitored, and if needed paused for the removal of ice.

iii) Different orientations of φ had to used for consecutive scans, and if possible opposite orientations (180° rotations).

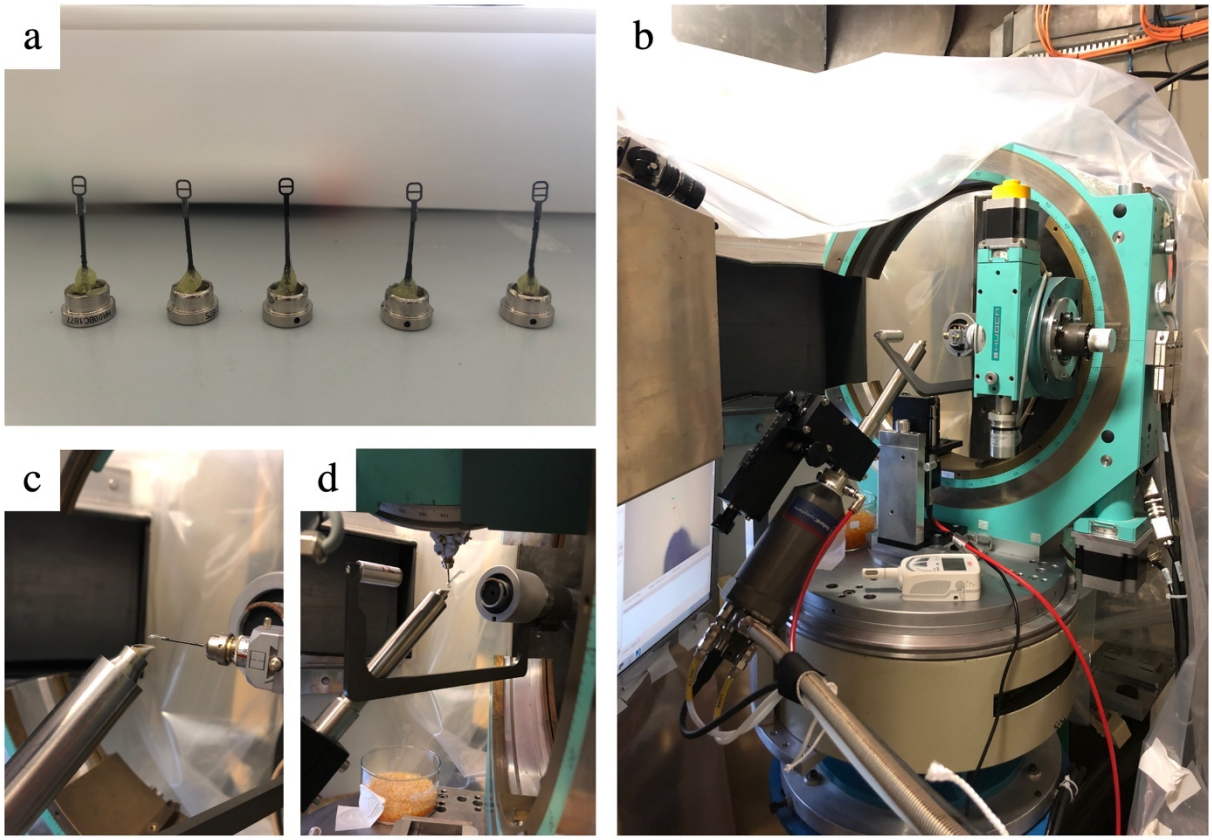


Figure 12

Pictures of critical elements used in the 100 K neutron diffraction experiment at D19, ILL (Manuscript 4). Carbon loops were manufactured to hold the large crystals (a). The instrument was covered with a plastic sheet and had dry silica beads placed nearby to control humidity (b). The carbon loop was aligned to the neutron beam prior to flash-cooling the crystal (c). Ice formation around the sample and on the carbon loop was the main hurdle during data collection (d).

2.4.3.1. Neutron data processing

Neutron diffraction data collected at D19, ILL, is usually processed with a program called RETREAT. This program performs data reduction by integrating the Bragg reflections in such a way as to minimize the ratio between σ(I) and I. The recorded reflections are treated as three-dimensional ellipsoids and weak reflections are assumed to have similar shapes to those of the nearest strong reflections. Previous work related to this project, on D_2O soaked H-HEWL crystals, showed that the integration of weak reflections using RETREAT was problematic, hindering the access to high resolution data. As a consequence, D19 beam time, in July 2019, was dedicated to measurements of long-exposure diffraction from a D-HEWL crystal at room temperature, to help build a library of reflection shapes at high-angle to be used by RETREAT in the reduction of subsequent high resolution data. However, challenges remained in data reduction as the data collected was still only a small fraction of a complete dataset and at a specific temperature. Another, and more successful approach, was done in collaboration with Prof. Kay Diederichs, from the University of Konstanz (Germany), enabled the adaptation of the XDS (Kabsch, 2010) program to D19 neutron data reduction. Processing using XDS (Kabsch, 2010) resolves the previously mentioned issues of integrating weak reflections and its use was of pivotal importance to the present study.

XDS (Kabsch, 2010) was employed in the reduction of both neutron and X-ray data in this project. Regarding the processing of neutron data, it was noted that XDS has difficulties to correctly index short ω scans of D19 data. This issue is emphasized by the lack of symmetry of the P1 space group, which does not provide enough, if any, symmetry-equivalent reflections in short scans and, thus, imped the calculation of important metrics such as R-factors and correlation-coefficients (CC). The D19 data is usually processed separately for each recorded ω scan, and then scaled using XSCALE (Kabsch, 2010). It was found that using the XDSCC12 program (Assmann *et al.*, 2020) to assess the CC for the final dataset with and without each scan provides valuable insight into issues in data reduction and can help to identify scans recorded upon changes in the experimental setup (*e.g.*, crystal movement, or presence of ice rings). After excluding scans which worsen the quality of the dataset, if present, the final reflection file can be converted to SHELX or CCP4 format using XDSCONV (Kabsch, 2010), and flags can be generated for the calculation of R_{free}. Where direct comparisons of models from refinements of different neutron or X-ray diffraction datasets were relevant, the R_{free} flags were generated for one dataset and then copied onto the remaining reflection files, using a python script written by Dr. Annette Langkilde.

2.4.3.2. Model refinements

Numerous structures of HEWL are available in the PDB and, thus, molecular replacement is commonly used to solve the crystal structure of this protein from diffraction data. In principle, this process implies the removal of alternate residue conformations, water molecules, ions, and ligands, if present, to avoid biasing the subsequent electron density or neutron density maps.

PHENIX (Liebschner *et al.*, 2019) is a commonly used software for model refinement against X-ray and neutron diffraction data. Recent developments have allowed better modelling of neutron structures, by enabling the refinement of H and D positions in both protein residues and water molecules, and also by allowing joint X-ray and neutron refinements (Afonine *et al.*, 2010; Liebschner *et al.*, 2018, 2019). In the case of a perdeuterated structure soaked in D_2O solvent, all H/D positions should be occupied by D atoms, and these will be visible in the neutron maps at resolutions better than ~2.5 Å (Chen *et al.*, 2012). On the other hand, if the crystal is of a hydrogenated protein soaked in D_2O solvent, all non-exchangeable positions will be occupied by H atoms, while the exchangeable positions can either have H or D atoms. For this purpose, PHENIX (Liebschner *et al.*, 2019) allows the modelling of H and D in the same position and the refinement of their occupancies, being correlated and their sum equal to 1. However, it is currently not possible to refine both H/D in protein residues modelled with alternate conformations, in the case of a perdeuterated structure. This problem is especially evident in high resolution neutron structures and leads to inaccurate models. In this study, atomic resolution neutron data is sought and, thus, the refinement of D atoms in disordered residues would not be possible in PHENIX (Liebschner *et al.*, 2019). Furthermore, since the refolding of D-HEWL was performed in H_2O solvent, caged H atoms would be expected in certain exchangeable positions, increasing the complexity of structure modelling.

Meanwhile, SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) is uniquely suited for model refinement against high resolution neutron and X-ray diffraction data, being routinely used in small-molecule crystallography. The work developed by Gruene *et al.* (2014) provided the basis for neutron refinements of protein structures, by reporting a series of restraints and constrains for deuterated amino acids and water molecules. SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) allows a high degree of control over model refinement, since the user is able to add, edit, or remove a plethora of parameters according to the type of refinement being performed, and the applied restraints are easily accessible. Additionally, fewer limitations are imposed to the format of the model being refined, compared to PHENIX (Liebschner *et al.*, 2019) which uses the standard *pdb* format. This flexibility enables *e.g.*, the modelling of H and/or D atoms in any H position of any particular residue.

Different model refinement programs apply distinct methodologies to obtain the best model possible. Variations in *e.g.*, minimization function, bulk solvent correction, and even in the rejection criteria applied to select the reflections used in refinement, have an impact in the accuracy of the refined model and hinder direct comparisons between models obtained from different programs. For this reason, the models from X-ray and neutron data used in this study for ADP analysis were all refined

using the same program, SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015). Details on the refinement strategies are described in Manuscripts 3 & 4.

Chapter 3

Manuscript 1

Synopsis

The first challenge faced in this study was the production of significant quantities of perdeuterated lysozyme to be used in the growth of large protein crystals for neutron diffraction experiments. While the project started with the production of material using the *P. pastoris* expression system, as described elsewhere (Campbell *et al.*, 2018), the transition to the *E. coli* system was necessary to increase the protein yield. An eventual improvement in protein production, and proportional reduction of the associated costs, was possible, even with the subsequent difficulties related to protein purification from inclusion bodies and protein *in vitro* refolding. The following published manuscript (Manuscript 1; Ramos *et al.*, 2021) describes in detail the methodology used to produce recombinant refolded perdeuterated HEWL (D-$HEWL_{EC}$) and the benefits of using the *E. coli* system compared to that of *P. pastoris*. Biophysical characterization of both perdeuterated variants (D-$HEWL_{EC}$ and D-$HEWL_{PP}$) is reported, focusing on protein thermal stability and enzymatic activity, including the comparison with unlabelled hydrogenated HEWL from *G. gallus* (H-HEWL). Moreover, the crystal structures of the three HEWL variants are determined from X-ray diffraction data at atomic resolution, enabling the detailed description of protein perdeuteration and refolding effects. Variations in protein structure are observed and can be linked to the changes in protein thermal stability and activity.

The work presented in this chapter proves that stable and active perdeuterated protein can be produced using the *E. coli* expression system and subsequent protein *in vitro* refolding. Additionally, it was shown that this new strategy for protein perdeuteration results in significant gains in protein yield and in the reduction of protein production costs. Altogether, these observations validated our new approach to obtain perdeuterated HEWL, forming the basis for our subsequent crystallographic studies. It is, however, important to note that data on the recombinant refolded hydrogenated HEWL (H-$HEWL_{EC}$) are still lacking, which could be valuable in separating the effects of deuteration and *in vitro* refolding in protein structural and biophysical properties.

IUCrJ
ISSN 2052-2525
BIOLOGY | MEDICINE

Structural insights into protein folding, stability and activity using *in vivo* perdeuteration of hen egg-white lysozyme

Joao Ramos,[a,b,c] Valerie Laux,[a,b] Michael Haertlein,[a,b] Elisabetta Boeri Erba,[b,d] Katherine E. McAuley,[e,f] V. Trevor Forsyth,[a,b,g] Estelle Mossou,[a,b,g] Sine Larsen[h] and Annette E. Langkilde[c]*

[a]Life Sciences Group, Institut Laue–Langevin, 71 Avenue des Martyrs, 38000 Grenoble, France, [b]Partnership for Structural Biology (PSB), 71 Avenue des Martyrs, 38000 Grenoble, France, [c]Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark, [d]Institut de Biologie Structurale, Université de Grenoble Alpes, CEA, CNRS, 71 Avenue des Martyrs, 38000 Grenoble, France, [e]Diamond Light Source, Didcot OX11 0DE, United Kingdom, [f]Paul Scherrer Institute, Forschungsstrasse 111, 5232 Villigen, Switzerland, [g]Faculty of Natural Sciences, Keele University, Newcastle-under-Lyme ST5 5BG, United Kingdom, and [h]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark. *Correspondence e-mail: annette.langkilde@sund.ku.dk

Received 30 November 2020
Accepted 3 February 2021

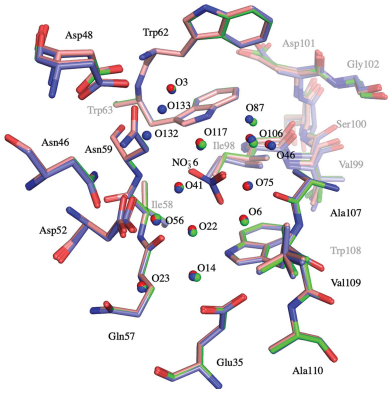
Edited by J. L. Smith, University of Michigan, USA

Keywords: perdeuterated lysozyme; HEWL; isotope effect; protein refolding; biophysical characterization; X-ray crystallography; protein structure; structural biology.

PDB references: hen egg-white lysozyme, hydrogenated, 7avf; perdeuterated, expressed in *E. coli* and refolded, 7ave; perdeuterated, produced in *Pichia pastoris*, 7avg

Supporting information: this article has supporting information at www.iucrj.org

This structural and biophysical study exploited a method of perdeuterating hen egg-white lysozyme based on the expression of insoluble protein in *Escherichia coli* followed by in-column chemical refolding. This allowed detailed comparisons with perdeuterated lysozyme produced in the yeast *Pichia pastoris*, as well as with unlabelled lysozyme. Both perdeuterated variants exhibit reduced thermal stability and enzymatic activity in comparison with hydrogenated lysozyme. The thermal stability of refolded perdeuterated lysozyme is 4.9°C lower than that of the perdeuterated variant expressed and secreted in yeast and 6.8°C lower than that of the hydrogenated *Gallus gallus* protein. However, both perdeuterated variants exhibit a comparable activity. Atomic resolution X-ray crystallographic analyses show that the differences in thermal stability and enzymatic function are correlated with refolding and deuteration effects. The hydrogen/deuterium isotope effect causes a decrease in the stability and activity of the perdeuterated analogues; this is believed to occur through a combination of changes to hydrophobicity and protein dynamics. The lower level of thermal stability of the refolded perdeuterated lysozyme is caused by the unrestrained Asn103 peptide-plane flip during the unfolded state, leading to a significant increase in disorder of the Lys97–Gly104 region following subsequent refolding. An ancillary outcome of this study has been the development of an efficient and financially viable protocol that allows stable and active perdeuterated lysozyme to be more easily available for scientific applications.



1. Introduction

Biomolecular deuteration is widely used in structural biology, where it plays a crucial role in techniques such as neutron macromolecular crystallography (NMX), small-angle neutron scattering (SANS), neutron reflectometry (NR), neutron spectroscopy and nuclear magnetic resonance (NMR) (Haertlein *et al.*, 2016; Varga *et al.*, 2007). For neutron studies, the fact that deuterium (D), an isotope of hydrogen (H), possesses a positive coherent neutron scattering length and a small incoherent neutron scattering cross section, and also an integer nuclear spin, is of central importance. In NMX, perdeuteration may be used to eliminate the incoherent scattering arising from the two spin states of the H atom (incoherent scattering cross section of 80.27 barns; Sears,

1992); this allows the use of samples that are approximately one order of magnitude smaller by volume (Hazemann *et al.*, 2005) and may result in improved resolution (Blakeley, 2009). Perdeuteration enhances the visibility of the coherent signal (Bragg reflections), fully exploiting the higher coherent scattering length of deuterium (6.67 fm) in comparison with that of hydrogen (−3.74 fm). Furthermore, perdeuteration helps to avoid cancellation effects (arising from the negative scattering length of hydrogen) that may occur for neutron Fourier maps based on data with intermediate resolution (for example $d > 1.6$ Å). In SANS, the use of deuterated samples allows contrast-matching techniques (Dunne *et al.*, 2017; Laux *et al.*, 2008; Haertlein *et al.*, 2016) to provide unique information on protein–protein (Vijayakrishnan *et al.*, 2010), protein–nucleic acid (Cuypers, Trubitsyna *et al.*, 2013) or protein–lipid (Breyton *et al.*, 2013) interactions. Sophisticated technologies have also been developed to allow the production of stealth-deuterated nanodiscs (Maric *et al.*, 2014, 2015) for the study of membrane proteins (Josts *et al.*, 2018; Nitsche *et al.*, 2018; Kehlenbeck *et al.*, 2019). In the case of NR, a wide range of research now routinely exploits the contrast enabled through the use of selective deuteration (Grage *et al.*, 2011; Moulin *et al.*, 2018; Waldie *et al.*, 2018, 2019, 2020). Deuterium labelling and reverse labelling have also been used for neutron scattering studies of the dynamics of biological macromolecules (Foglia *et al.*, 2016), in particular when coupled with hydration water dynamics (Wood *et al.*, 2013). In the case of solution-state NMR, deuterium labelling is essential for multidimensional heteronuclear NMR studies of proteins, especially high-molecular-weight proteins and macromolecular complexes. Partial deuteration simplifies the NMR spectra from the remaining ^{1}H nuclei and also contributes to spectra with a higher signal-to-noise ratio owing to the effects on the relaxation of bonded or adjacent ^{1}H, ^{13}C and ^{15}N atoms (Sattler & Fesik, 1996). While major developments in *in vivo* deuteration technologies have occurred in the last 15 years, the expression of deuterated protein is often complex and expensive and may be associated with low yields. The way in which it is carried out depends on the downstream application and on the labelling regime needed to answer the scientific questions posed (Haertlein *et al.*, 2016). In the case of neutron crystallographic applications, the goal is invariably to perdeuterate the sample so that the incoherent scattering from hydrogen is removed from the recorded data to the maximum possible extent.

Hen egg-white lysozyme (HEWL) was the first enzyme structure to be solved by X-ray crystallography (Blake *et al.*, 1965), and has subsequently become a widely used model in structural biology in a variety of contexts including protein-folding studies (Miranker *et al.*, 1991, 1993; Radford *et al.*, 1992; Wildegger & Kiefhaber, 1997) and crystallization (Durbin & Feher, 1986; McPherson & DeLucas, 2015; Darmanin *et al.*, 2016). It is a small (129 residues, 14.3 kDa) and stable protein that in its hydrogenated form can be acquired at low cost and crystallized in numerous space groups under well known conditions. HEWL is a hydrolase from *Gallus gallus* that cleaves the 1,4-β-linkages between *N*-acetyl-muramic acid and *N*-acetyl-D-glucosamine residues in peptidoglycan. The recombinant production of HEWL in *Escherichia coli* is challenged by the reductive environment of the bacterial cytosol, which prevents the correct formation of its four disulfide bridges, resulting in the formation of inclusion bodies. The use of the yeast *Pichia pastoris* has been investigated as an expression system for the production of recombinant HEWL (Liu *et al.*, 2003; Li *et al.*, 2012; Mine *et al.*, 1999; Campbell *et al.*, 2018). While this approach results in the production of high-quality protein, the low yield is problematic for neutron crystallographic and spectroscopic applications. For this reason, we developed an approach whereby large quantities of insoluble protein were produced as inclusion bodies in *E. coli*, followed by an optimized refolding process, significantly improving the yield.

Using this strategy, large amounts of correctly folded perdeuterated HEWL (D-HEWL) can be obtained at a financially viable level. Of particular interest is the fact that the refolded perdeuterated lysozyme from *E. coli* (D-HEWL$_{EC}$) provides important insights into the structural and biophysical properties of HEWL when compared with those of the perdeuterated analogue produced in *P. pastoris* (D-HEWL$_{PP}$) and those of the commercially available non-recombinant hydrogenated protein (H-HEWL). These variants are identical in primary structure, with the exception of an additional glycine at the N-terminus of D-HEWL$_{EC}$.

Atomic resolution X-ray structures have been determined for all three variants using triclinic crystals obtained in closely comparable conditions. The effect of deuteration on reduced thermal stability and activity is noted. The structural analyses highlight subtle but important differences that are related to the decrease in the thermal stability of D-HEWL$_{EC}$; these differences are of significance for protein folding (Biter *et al.*, 2016), enzymatic activity (Lea & Simeonov, 2012), crystallization (Geders *et al.*, 2012; Reinhard *et al.*, 2013) and protein–ligand interactions (Bai *et al.*, 2019; Forneris *et al.*, 2009; Holdgate *et al.*, 2010; Ramos *et al.*, 2019). The improved yield (by a factor of more than three compared with that found using *P. pastoris*) paves the way for a wide range of studies that can exploit H/D isotopic substitution in this protein.

2. Methods

2.1. Expression of D-HEWL$_{EC}$

Recombinant D-HEWL was overexpressed in *E. coli* BL21 (DE3) cells grown in a Labfors 2.3 l computer-controlled fermenter (Infors, France) using a high cell-density culture (HCDC) strategy. Transformation of chemically competent cells with the vector pET-28a(+) (GenScript) containing codon-optimized cDNA for HEWL expression (Supplementary Fig. S1) was performed by heat shock. Using a lysogeny broth (LB) solid medium supplemented with 40 µg ml^{-1} kanamycin (catalogue No. 60615; Sigma–Aldrich), transformed cells were selected. The cells containing the vector were then adapted to hydrogenated Enfors minimal medium containing hydrogenated glycerol and kanamycin. The cells

were further adapted to fully deuterated Enfors medium supplemented with d_8-glycerol (catalogue no. DLM-558-PK; Eurisotop) and antibiotic. 100 ml precultures were prepared to inoculate 1.4 l minimal medium in the fermenter. During batch and fed-batch phases, the pD (pD = $pH_{measured}$ + 0.4; Glasoe & Long, 1960) was maintained at 6.4 by adding NaOD. The gas-flow rate of sterile-filtered air was 0.5 l min^{-1}. Stirring was adjusted to ensure a dissolved oxygen level of 30%. The initial glycerol supply was consumed during the batch phase. The cells were then fed continuously with fresh feeding solution containing 12% d_8-glycerol in an exponential manner (fed-batch phase). When the cell density reached an OD of 10, recombinant protein expression was induced by adding IPTG to a final concentration of 1 m*M*. The cells were harvested after 24 h of induction. The final volume of cell culture extracted from the fermenter was approximately 1.8 l.

2.2. Inclusion-body separation of D-HEWL$_{EC}$

The inclusion bodies were purified in a centrifugation-based approach, with several washing steps to remove nucleic acids, lipids and other contaminants. After pelleting the *E. coli* cells, lysis was promoted by sonication with a Vibra-Cell ultrasonic liquid processor (VCX-750-220, Sonics & Materials), performing three cycles of 30 s at amplitude 0.8 in a buffer consisting of 0.1 *M* Tris–HCl pH 8.45, 150 m*M* NaCl, 25 m*M* EDTA, 25 m*M* DTT, 0.5% Triton X-100. The suspension was centrifuged at 10 080*g* for 1 h at 4°C to separate the soluble and insoluble fractions. The supernatant was discarded and the pellet was solubilized in 0.1 *M* Tris–HCl pH 8.45, 150 m*M* NaCl, 25 m*M* EDTA, 25 m*M* DTT, 1 *M* guanidine–HCl, 1% Triton X-100 using a homogenizer (D1000, Benchmark). The suspension was sonicated three times for 10 s at amplitude 0.4 and was then centrifuged at 22 680*g* for 30 min at 4°C. This washing step was performed six times, with Triton X-100 excluded from the buffer in the last two cycles.

2.3. Purification of D-HEWL$_{EC}$

The inclusion bodies were solubilized in 0.1 *M* Tris–HCl pH 8.45, 150 m*M* NaCl, 25 m*M* EDTA, 25 m*M* DTT, 6 *M* guanidine–HCl using a homogenizer. The suspension was sonicated three times for 10 s at amplitude 0.4 and then centrifuged at 22 680*g* for 1 h at 4°C. The soluble fraction was collected and filtered through 0.4 µm filters. Purification of unfolded protein was performed by gel filtration on a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) equilibrated in the same buffer. Protein fractions of 5 ml were diluted to avoid saturation of the UV detector of the HPLC and were injected into the column, running an isocratic flow at 1.0 ml min^{-1}. Pure protein eluted at 0.6–0.7 column volumes (CV). The fractions of pure protein collected were frozen at −80°C until the refolding procedure.

2.4. Refolding of D-HEWL$_{EC}$

Denatured protein was refolded at room temperature in a size-exclusion chromatography (SEC) setup using a HiLoad 16/600 Superdex 200 pg column equilibrated with 0.1 *M* Tris–HCl pH 8.45, 2 *M* urea, 1 m*M* EDTA, 3 m*M* reduced glutathione, 0.3 m*M* oxidized glutathione as described by Batas & Chaudhuri (1996). 5 ml injections of pure unfolded HEWL at concentrations of 1–2 mg ml^{-1} were performed in each run; the isocratic flow was set to 0.1 ml min^{-1}, resulting in monomeric HEWL fractions being collected at 0.9 CV.

The protein buffer was exchanged to 50 m*M* sodium acetate pD 4.5 in D_2O by desalting using two coupled HiTrap 5 ml desalting columns (GE Healthcare). Injections of 2.5 ml of protein at 0.6 mg ml^{-1} were performed. The protein was subsequently concentrated to 20 mg ml^{-1} for crystallization experiments.

2.5. Expression and purification of D-HEWL$_{PP}$

The expression of D-HEWL$_{PP}$ was achieved as described by Campbell *et al.* (2018). Since the protein was secreted into the extracellular medium, the supernatant was recovered upon cell pelleting. The supernatant was diluted by the addition of 50 m*M* Tris–HCl pH 7.8 buffer to achieve a solution conductivity of below 10 mS cm^{-1}. Pure protein was obtained by ion-exchange chromatography (IEC) using an SP-Sepharose column (GE Healthcare) and elution with a 30 ml NaCl gradient from 0 to 1 *M* in 50 m*M* Tris–HCl pH 7.8 buffer. Following the same approach as the final buffer exchange of D-HEWL$_{EC}$, the D-HEWL$_{PP}$ buffer was exchanged to 50 m*M* sodium acetate pD 4.5 in D_2O by desalting. The protein was concentrated to 30 mg ml^{-1} for crystallization experiments.

2.6. Mass spectrometry (MS)

MS under denaturing conditions was utilized to assess the mass of the intact deuterated proteins and their degree of labelling. Specifically, liquid-chromatography/electrospray ionization mass spectrometry (LC/ESI-MS) on a 6210 TOF mass spectrometer coupled to an HPLC system (1100 series, Agilent Technologies) was performed. Data acquisition was carried out in positive-ion mode, and mass spectra were recorded in the 300–3200 m/z range. The following experimental settings were utilized: the ESI source temperature was set to 300°C, N_2 was used as a drying gas (with a flow rate of 7 l min^{-1}) and as a nebulizer gas (using a pressure of 69 kPa) and the capillary needle voltage was 4 kV. Voltages in the first part of the instrument were set as follows: the voltage of the fragmentor was 250 V and that of the skimmer was 60 V. The acquisition rate was one spectrum per second. Instrument pressure values were typically 2.33 Torr (rough vacuum) and 4.6×10^{-7} Torr (TOF vacuum). The mass spectrometer was calibrated with tuning mix (ESI-L, Agilent Technologies). The HPLC mobile phases were prepared with HPLC-grade solvents. The mobile phase *A* composition was 95% H_2O, 5% acetonitrile (ACN), 0.03% trifluoroacetic acid (TFA). The mobile phase *B* composition was 95% ACN, 5% H_2O, 0.03% TFA.

As partial D-to-H back-exchange would be possible during the experiment, both samples were dialyzed against 50 m*M* sodium acetate pH 4.5 buffer in H_2O prior to the MS experiment to ensure full back-exchange and thus allow the

evaluation of the number of D atoms in all non-exchangeable positions.

Just before the analysis, the samples were diluted in 0.03% TFA to obtain a concentration of 5 µ*M* and a volume of 20 µl. The samples were loaded into glass vials, which were placed on a sample loader refrigerated at 10°C. 4 µl of each sample (*i.e.* ∼20 pmol of protein) was injected into the HPLC system directly connected to the mass spectrometer. The injected sample was first trapped and desalted on an RP-C8 cartridge for 3 min at a flow rate of 50 µl min^{-1} using 100% mobile phase *A*. Afterwards, the proteins were separated on an RP-C8 column using a linear gradient from 5 to 95% mobile phase *B* for 15 min and subjected to ESI prior to the TOF detection of their *m*/*z* signals. The software *MassHunter BioConfirm* (version B.07.00; Agilent Technologies) was used to calculate masses from *m*/*z* values obtained during the MS experiments.

2.7. Differential scanning fluorimetry (DSF)

DSF measurements were performed using a Prometheus instrument (NanoTemper). The setup included a temperature ramp from 20 to 95°C with increments of 1.0°C min^{-1}, following unfolding by the intrinsic fluorescent signal from the tryptophan residues (six tryptophans in HEWL). Lyophilized H-HEWL powder was dissolved in 50 m*M* sodium acetate pD 4.5 in D_2O to match the conditions of D-HEWL$_{EC}$ and D-HEWL$_{PP}$. The experiment was repeated in the hydrogenated buffer of the activity assay, where the samples were diluted in 0.1 *M* sodium phosphate pH 7.5, 0.1 *M* NaCl, 2 m*M* NaN_3 in H_2O in a ratio of at least 1:50. The results presented correspond to samples at concentrations of 0.3 mg ml^{-1} with a 40% excitation power and were obtained for at least two HEWL preparations as duplicate or triplicate measurements for every condition.

2.8. HEWL activity assays

The activity assays were performed based on the work of Shugar (1952). The activity is followed by the absorbance at 450 nm at 25°C, with measurements every minute for 20 min. Nunc 96-well flat-bottom plates (Thermo Fisher Scientific) were used with each sample in triplicate, including negative controls without protein. The 100 µl samples used for these experiments comprised 50 µl protein sample at 0.2 mg ml^{-1} and 50 µl *Micrococcus lysodeikticus* cell suspension in H_2O with 0.1 *M* sodium phosphate pH 7.5, 0.1 *M* NaCl, 2 m*M* NaN_3. After averaging triplicates of each experiment, the activity curves were plotted against time, and the linear phase ($R^2 > 0.91$) corresponding to the first 8 min of reaction was considered to retrieve the initial velocities. Standard deviations were derived from three separate experiments and a *t*-test was performed for each pair of results to assess the significance of the homoscedastic hypothesis, meaning the probability of the pairs of measured values being equal.

2.9. Protein crystallization

H-HEWL (catalogue No. L6876; Sigma–Aldrich) was crystallized in the triclinic form in batch-like conditions using a precipitation step as described by Vidal *et al.* (1999). 5 µl drops were prepared consisting of 2.5 µl H-HEWL at 20 mg ml^{-1} dissolved in deionized water and 2.5 µl 0.4 *M* $NaNO_3$, 50 m*M* sodium acetate pH 4.5. Under these conditions, monoclinic crystals readily formed at room temperature. To obtain the triclinic crystal form, the crystallization plate was stored at 4°C overnight and then subsequently kept at 18°C. During the cold storage, crystals of both the triclinic and monoclinic forms nucleate. When the temperature is raised, the less stable monoclinic form dissolves, leaving almost exclusively nuclei of the triclinic form (Legrand *et al.*, 2002). Triclinic crystals appeared after three days.

Triclinic crystals of D-HEWL were obtained by initial microseeding using triclinic H-HEWL seeds from a crystal in 100% D_2O buffer. The seed solution was made by crushing the crystal in 0.3 *M* $NaNO_3$, 50 m*M* sodium acetate pD 4.5. Subsequently, the solution was transferred to an Eppendorf tube containing a zirconium silicate ceramic seed bead (Hampton Research) and vortexed to produce microseeds. Seed stocks of 1:100 and 1:1000 dilutions were used in the crystallization experiments. Sitting drops of 5.5 µl were prepared by microbatch under oil and stored at 18°C. The drops consisted of 2.5 µl D-HEWL$_{EC}$ at 20 mg ml^{-1} or D-HEWL$_{PP}$ at 30 mg ml^{-1}, 2.5 µl 0.3 *M* $NaNO_3$, 50 m*M* sodium acetate pD 4.5 and 0.5 µl of the H-HEWL seed solution. Triclinic crystals of D-HEWL of up to 0.1 mm^3 were obtained within one week.

2.10. X-ray data collection, processing and model refinement

Synchrotron X-ray diffraction data were collected at 100 K from crystals of H-HEWL, D-HEWL$_{EC}$ and D-HEWL$_{PP}$. The data collections were performed on beamline I03 at Diamond Light Source (DLS), UK and on BioMAX at MAX IV, Sweden (Table 1). Crystals of approximately 0.1 mm^3 were cooled in cryoprotectant solutions of 25–35%(*v*/*v*) glycerol or d_8-glycerol with 0.3 *M* $NaNO_3$ and 50 m*M* sodium acetate pH/pD 4.5 in H_2O for H-HEWL or D_2O for both D-HEWL forms. Due to the low triclinic crystal symmetry, the data sets were measured in two different κ orientations to improve the completeness of the data. The data were reduced, merged and scaled using *XDS* (Kabsch, 2010). Initial phases were estimated by molecular replacement in *Phenix* (Liebschner *et al.*, 2019), using the structure deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000) as entry 4yeo (Shabalin *et al.*, 2015), stripped of ligands and water molecules, as a starting model. Model refinement was performed using *Phenix* (Liebschner *et al.*, 2019), with the same set of reflections flagged for the R_{free} calculation. Model building was achieved using *Coot* (Emsley *et al.*, 2010). The D-HEWL$_{EC}$ model from a late stage of refinement was used for the initial refinement of D-HEWL$_{PP}$ and H-HEWL to maintain the labelling of residue disorder as well as of water molecules and ions. H/D atoms were added to the models as riding atoms in ideal positions. The occupancy of water molecules and ions was refined for atoms with *B* factors above 20 Å^2 and was otherwise fixed to 1. Water molecules which displayed a density lower than 1.5 σ in

Table 1
X-ray diffraction data-collection and model-refinement statistics for H-HEWL, D-HEWL$_{EC}$ and D-HEWL$_{PP}$

Values in parentheses are for the outer resolution shell.

	H-HEWL	D-HEWL$_{EC}$	D-HEWL$_{PP}$
Cryoprotectant	25%(*v*/*v*) glycerol	35%(*v*/*v*) d_8-glycerol	30%(*v*/*v*) d_8-glycerol
Strategy	2 κ orientations, 180° scans	2 κ orientations, 180° scans	2 κ orientations, 360° scans
Beamline and source	I03, DLS	I03, DLS	BioMAX, MAX IV
Detector	EIGER2 XE 16M	EIGER2 XE 16M	EIGER hybrid-pixel 16M
Wavelength (Å)	0.7293	0.7293	0.7999
Resolution range (Å)	31.99–1.00 (1.036–1.000)	32.01–0.98 (1.015–0.980)	32.01–1.00 (1.036–1.000)
Space group	*P*1	*P*1	*P*1
a, *b*, *c* (Å)	26.76, 31.07, 33.77	26.67, 30.97, 33.74	26.67, 30.97, 33.74
α, β, γ (°)	89.211, 72.459, 67.863	89.439, 72.818, 67.503	89.439, 72.818, 67.503
Total reflections	178195 (17670)	278571 (24674)	337122 (32144)
Unique reflections	50297 (4908)	52966 (5018)	49991 (4880)
Multiplicity	3.5 (3.6)	5.3 (4.9)	6.7 (6.6)
Completeness (%)	97.47 (94.95)	97.44 (92.67)	97.71 (95.50)
Mean $I/\sigma(I)$	9.25 (2.81)	18.09 (4.21)	15.07 (6.92)
Wilson *B* factor (Å^2)	8.43	7.12	8.81
R_{merge}	0.0727 (0.386)	0.0424 (0.273)	0.0803 (0.262)
R_{meas}	0.0856 (0.453)	0.0470 (0.306)	0.0876 (0.284)
$R_{p.i.m.}$	0.0449 (0.236)	0.0201 (0.136)	0.0344 (0.109)
$CC_{1/2}$	0.996 (0.884)	0.999 (0.935)	0.992 (0.979)
CC*	0.999 (0.969)	1.00 (0.983)	0.998 (0.995)
Reflections used in refinement	50286 (4906)	52964 (5018)	49982 (4878)
Reflections used for R_{free}	2398 (218)	2510 (223)	2400 (222)
R_{work}	0.1172 (0.1577)	0.1049 (0.1323)	0.1205 (0.1206)
R_{free}	0.1319 (0.1740)	0.1145 (0.1423)	0.1341 (0.1326)
CC_{work}	0.976 (0.956)	0.977 (0.969)	0.965 (0.977)
CC_{free}	0.976 (0.945)	0.972 (0.963)	0.941 (0.970)
No. of non-H/D atoms			
Total	1502	1467	1474
Macromolecule	1308	1291	1299
Ligands	40	40	40
Solvent	154	136	135
Protein residues	129	130	129
R.m.s.d., bond lengths (Å)	0.011	0.008	0.013
R.m.s.d., angles (°)	1.46	1.42	1.62
Ramachandran favoured (%)	96.85	97.66	97.64
Ramachandran allowed (%)	3.15	2.34	2.36
Ramachandran outliers (%)	0	0	0
Rotamer outliers (%)	0.7	1.44	0.71
Clashscore	4.18	2.32	4.6
Average *B* factor (Å^2)			
Overall	11.06	9.96	11.84
Macromolecule	10.27	9.19	11.33
Ligands	16.50	16.71	16.55
Solvent	16.37	15.25	15.35

the $2F_o - F_c$ electron-density map were removed from the models.

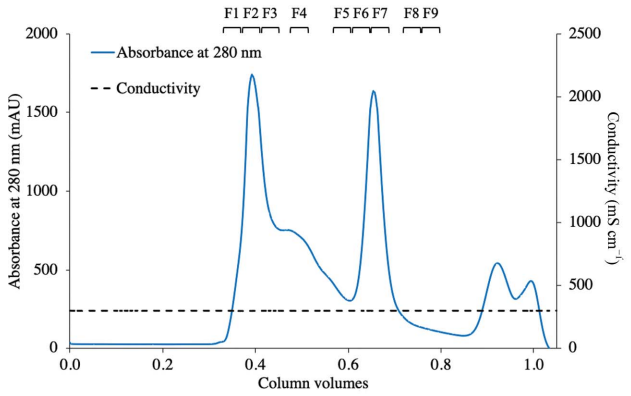
2.11. X-ray structure analysis and comparison

Structural alignment of the entire protein chains was achieved with the *CEALIGN* plugin (Shindyalov & Bourne, 1998) using the C^α atoms from 128 residues, while alignment of the Lys97–Gly104 region (using all atoms) was performed with the *SUPER* function of *PyMOL* (version 2.0; Schrödinger). *EDSTATS* (Tickle, 2012) from the *CCP*4 suite (Winn *et al.*, 2011) was employed to evaluate the quality of the models according to the data, allowing the identification of residues that may not have been reliably modelled for further analysis. The combination of cutoffs considered was 90% for the RSCC, 1σ for the sample RSZO and −3σ and +3σ for RSZO− and RSZO+, respectively. Hydrogen-bond analysis was performed using *HBPLUS* (McDonald & Thornton, 1994). Results that included intra-residue interactions and residues that were not reliably modelled according to the metrics from *EDSTATS* (Tickle, 2012) were not considered for the comparison between models, with the exception of Thr89 from D-HEWL$_{PP}$, which participates in an extensive hydrogen-bond network involving His15, Asp87 and Asn93. The graphical representations presented here were made in *PyMOL*.

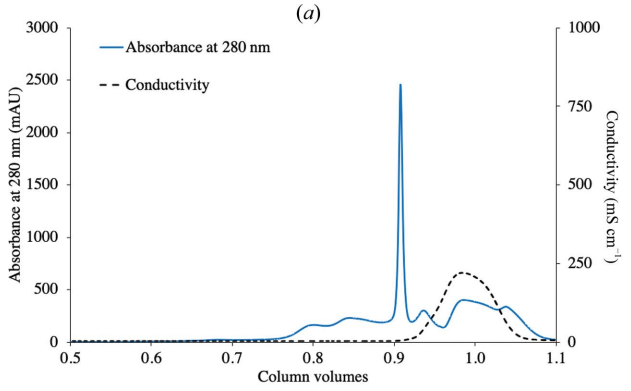
3. Results

3.1. Increased yield by refolding from inclusion bodies

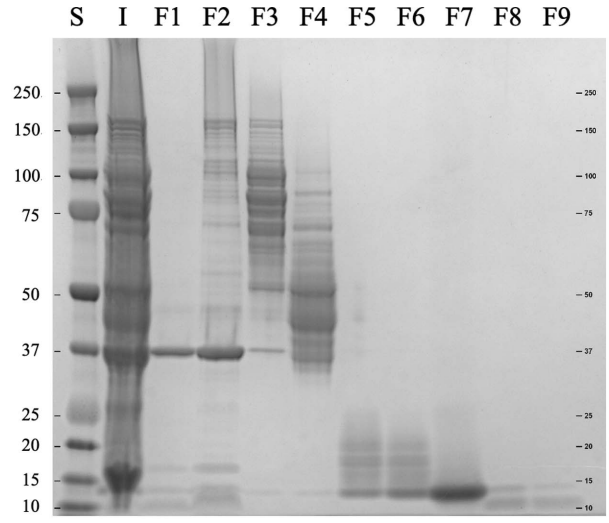
Inclusion bodies from D-HEWL$_{EC}$ expression were separated from insoluble contaminants, as shown by SDS–PAGE of the supernatants from the washing steps (Supplementary Fig. S2). The untagged D-HEWL$_{EC}$ was further purified by gel



(*a*)



(*b*)



(*b*)

Expression system	*E. coli*	*P. pastoris*
Cell paste	27 g l^{-1}	50 g l^{-1}
Pure protein	186 mg l^{-1}	11 mg l^{-1}
Refolded protein*	37 mg l^{-1}	—

(*d*)

Figure 1
The expression of insoluble D-HEWL in *E. coli* followed by refolding increases the yield of protein production by more than threefold. (*a*) Chromatogram from the denaturing SEC, yielding pure unfolded D-HEWL$_{EC}$, which eluted at 0.6–0.7 CV. (*b*) Fractions from the denaturing SEC on a 12% SDS–PAGE gel. Lane S, Precision Plus Protein Dual Xtra Standards (Bio-Rad); lane I, injected sample of unpurified D-HEWL$_{EC}$ from the inclusion-body washing steps; lanes F1–F9, collected fractions from the denaturing SEC as indicated at the top of (*a*). Fractions F5–F7 were used in subsequent refolding experiments. (*c*) Refolding SEC chromatogram, where monomeric D-HEWL elutes at 0.9 CV. The fractions eluting before and after the monomeric refolded D-HEWL$_{EC}$ are likely to be misfolded or oligomeric and partially unfolded forms of D-HEWL$_{EC}$, respectively. This is followed by the elution of the guanidine–HCl and the DTT from the denaturing buffer, as shown by the increase in conductivity. (*d*) Comparison of the D-HEWL expression yields between the two systems, *E. coli* and *P. pastoris*. *Considering an average refolding yield of 20%, the final yield of D-HEWL$_{EC}$ production is 37 mg l^{-1} without further denaturing and refolding of the misfolded, oligomeric and partially unfolded fractions.

filtration, eluting as a single peak around 0.6–0.7 CV, with fractions F5–F7 being collected for the refolding step [Figs. 1(*a*) and 1(*b*)].

D-HEWL$_{EC}$ was refolded in-column using a low flow rate of 0.1 ml min^{-1}, which allowed desalting of the unfolded protein and separation of the monomeric and oligomeric, misfolded and partially unfolded fractions [Fig. 1(*c*)]. The fractions of refolded D-HEWL in refolding buffer and in deuterated protein buffer are shown on a 12% SDS–PAGE gel in Supplementary Fig. S3. The refolding yields were impacted by the fact that the molecular weight of HEWL is close to the lower exclusion limit of the gel-filtration column (M_r = 10 kDa), which hindered optimal separation of the monomeric protein fraction from the denaturing buffer. Injections of 6.5 mg unfolded protein resulted in average refolding yields of 20%. The expression, purification and refolding strategy yielded 186 mg of pure protein per litre of culture on average, from which, considering a consistent refolding yield of 20%, 37 mg was recovered in a native-like state. These results represent more than a threefold increase in D-HEWL production compared with the *P. pastoris* system [Fig. 1(*d*)].

3.2. All non-exchangeable H positions are fully deuterated in both D-HEWL$_{EC}$ and D-HEWL$_{PP}$

The deuteration level of D-HEWL$_{EC}$ and D-HEWL$_{PP}$ was assessed by LC/ESI-MS. Prior to the MS experiments, both samples were dialyzed against 50 m*M* sodium acetate pH 4.5 buffer in H_2O to avoid partial back-exchange during the experiment. Therefore, the expected masses included D in all

Table 2
Expected and observed masses for D-HEWL$_{EC}$ and D-HEWL$_{PP}$ in the MS experiments.

Sample	MW of hydrogenated oxidized form (Da)	No. of non-exchangeable H positions	No. of exchangeable H positions	Expected mass of perdeuterated variant in H_2O (Da)	Observed mass in H_2O (Da)
D-HEWL$_{EC}$	14362	698	256	15064	15060
D-HEWL$_{PP}$	14305	696	255	15005	15005

non-exchangeable positions (*i.e.* bound to C) and, with full back-exchange, H in all labile positions (Table 2).

D-HEWL$_{EC}$ has 130 residues, with one additional glycine at the N-terminus compared with the other HEWL variants studied (Supplementary Fig. S1), resulting in differences in the expected masses. The masses observed by MS of 15 060 and 15 005 Da (Supplementary Fig. S4) for D-HEWL$_{EC}$ and D-HEWL$_{PP}$, respectively, closely match the expected values (Table 2) and verify the successful replacement of H atoms by D atoms in non-exchangeable positions. D-HEWL$_{EC}$ shows a minor difference of 4 Da between the expected and the observed masses, which shows that 99.4% of all non-exchangeable positions are occupied by D. The D-HEWL$_{PP}$ observed mass exactly matched the expected value of the fully deuterated form.

3.3. Perdeuterated variants of lysozyme are stable and active

DSF assays were performed to retrieve information on the folding and stability of D-HEWL$_{EC}$ and D-HEWL$_{PP}$ using H-HEWL as a reference. Results were obtained using the same deuterated buffer (50 m*M* sodium acetate pD 4.5 in D_2O) and showed that both variants of D-HEWL are thermally less stable than H-HEWL [Figs. 2(*a*) and 2(*b*)]. The refolded D-HEWL$_{EC}$ is less thermally stable than D-HEWL$_{PP}$, with a difference in melting temperature of 4.9°C. Moreover, compared with H-HEWL, the refolded D-HEWL$_{EC}$ shows a decrease in thermal stability of 6.8°C. If the D-HEWL$_{EC}$ was not completely separated from denaturing salts upon refolding, a small population of misfolded protein could potentially be present in the sample. To test this, D-HEWL$_{EC}$ crystals were washed and dissolved in protein buffer (from now on referred to as D-HEWL$_{EC}$ after crystallization) and analyzed by DSF [Figs. 2(*a*) and 2(*b*)]. With differences of less than 1°C observed between D-HEWL$_{EC}$ before and after crystallization, it was concluded that the lower thermal stability was not attributable to the presence of misfolded protein in the D-HEWL$_{EC}$ sample.

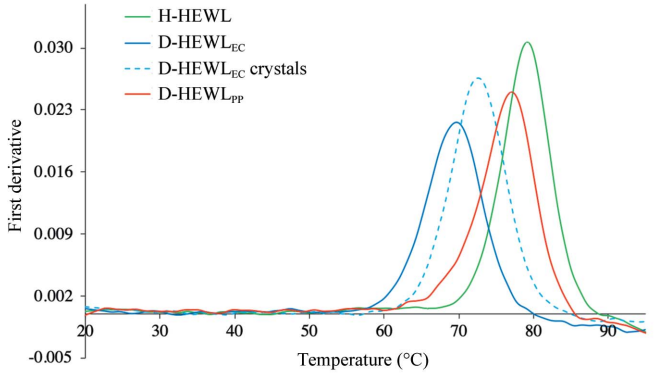
The enzymatic activities of the D-HEWL variants were also assessed. As part of this, DSF measurements were performed in activity-assay buffer (0.1 *M* sodium phosphate buffer pH 7.5 in H_2O with 0.1 *M* NaCl and 2 m*M* NaN_3). A systematic decrease in stability of all of the samples was observed in this buffer [Fig. 2(*a*)]. The D-HEWL variants are less active than H-HEWL (D-HEWL$_{EC}$ and D-HEWL$_{PP}$ exhibited 51% and 67% of the activity of H-HEWL, respectively; both were significantly different, with *t*-test *p* values of <0.05) [Fig. 2(*c*)]. Conversely, the activity difference between D-HEWL$_{EC}$ and D-HEWL$_{PP}$ is not significant (p = 0.19). As for the thermal stability, no significant differences were observed between D-HEWL$_{EC}$ before and after crystallization.

3.4. Structural similarities and differences

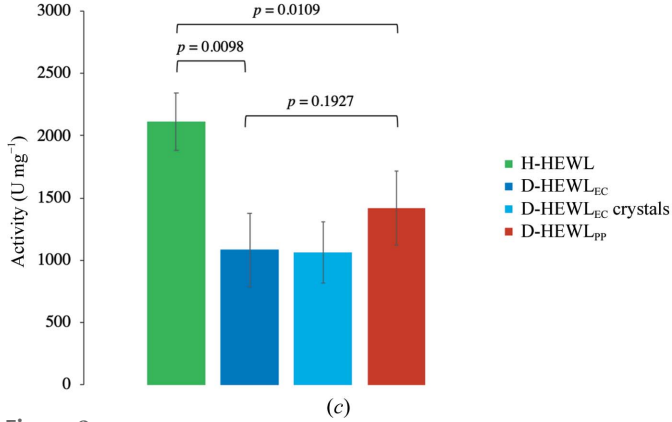
Atomic resolution X-ray diffraction data were collected for all three variants: H-HEWL, D-HEWL$_{EC}$ and D-HEWL$_{PP}$. The data sets all extended to at least 1.00 Å resolution, although, as evident from the merging statistics (Table 1), the resolution cutoff was limited by the experimental geometry (detector distance and coverage) and not by the diffraction power of the crystals. Given the low symmetry of the *P*1 space group, a data-collection strategy with sweeps collected in two distinct crystal orientations (different κ angles) was implemented. An overall completeness of greater than 90% was obtained to a resolution of 1.00 Å. Refinement of the three variants of lysozyme provided the basis for comparison of the features and differences between the structures. The quality and resolution of the diffraction data allowed the visualization of elusive structural detail, including side-chain and main-

Sample	Protein buffer in D_2O	Activity-assay buffer in H_2O
	Unfolding T_m (°C)	Unfolding T_m (°C)
H-HEWL	78.8 ± 0.2	73.0 ± 0.1
D-HEWL$_{EC}$	72.0 ± 0.1	66.8 ± 0.1
D-HEWL$_{EC}$ crystals	72.5 ± 0.0	67.6 ± 0.1
D-HEWL$_{PP}$	76.9 ± 0.1	70.7 ± 0.2

(*a*)



(*b*)



(*c*)

Figure 2
Stability and activity of the HEWL variants. (*a*) T_m values for H-HEWL, D-HEWL$_{EC}$ (before and after crystallization) and D-HEWL$_{PP}$ in deuterated protein buffer (50 m*M* sodium acetate pD 4.5) and in hydrogenated activity-assay buffer (0.1 *M* sodium phosphate pH 7.5, 0.1 *M* sodium chloride, 2 m*M* sodium azide). (*b*) Thermal unfolding curves (first derivative against temperature) of H-HEWL (green), D-HEWL$_{EC}$ before crystallization (continuous dark blue line) and after crystallization (dashed light blue line) and D-HEWL$_{PP}$ (red) in deuterated protein buffer. (*c*) Enzymatic activity of H-HEWL (green), D-HEWL$_{EC}$ (dark blue, before crystallization; light blue, after crystallization) and D-HEWL$_{PP}$ (red) in the hydrogenated activity-assay buffer. The *p*-values represent the significance of the homoscedastic hypothesis, meaning the probability of the pairs of measured values being equal.

chain disorder, and the interpretation of complex hydrogen-bonding patterns and their underlying structural dynamics.

3.4.1. Secondary and tertiary structures are retained. Numerous structures of HEWL are available in the PDB, representing a multitude of crystallization conditions, different space groups, ligands, humidity levels, mutations *etc.*, but a benchmark in this large pool of structures is the *P*1 structure refined to 0.65 Å reesolution by Wang *et al.* (2007) (PDB entry 2vb1). With a r.m.s.d. of 0.23 Å between the C^{α} atoms, the three-dimensional structure of H-HEWL obtained in our study closely matches this model. There are some differences between the disorder modelled in the two structures, which may reflect the difference in resolution of the corresponding data sets.

In order to perform a comparison of the structure of H-HEWL with the structures of the two D-HEWL variants, the model of H-HEWL was obtained from similar crystallization conditions, data-collection and refinement parameters and resolution limits to those for the D-HEWL structures.

The structural alignment based on C^{α} atoms between the three HEWL variants showed a high degree of similarity, with an r.m.s.d. of 0.11 Å for both D-HEWL$_{EC}$ and D-HEWL$_{PP}$ in comparison with H-HEWL [Fig. 3(*a*)]. The conserved tertiary and secondary structures indicate that perdeuteration did not have a significant impact on the overall protein fold, as has been demonstrated in many neutron crystallographic studies of other proteins (Artero *et al.*, 2005; Haupt *et al.*, 2014; Langan *et al.*, 2014; Cuypers, Mason *et al.*, 2013; Yee *et al.*, 2019; Liu *et al.*, 2007; Koruza *et al.*, 2019). The r.m.s.d. between the two D-HEWL variants was 0.13 Å, suggesting that the refolding process had little effect on the global protein fold.

3.4.2. Alternate conformations and hydrogen-bond patterns. The atomic resolution X-ray data enabled a detailed description of backbone and side-chain disorder. Alternate conformations were modelled for approximately 30% of the protein residues. Overall, the structures exhibited similar disorder patterns; the only exceptions were residues Glu7, Asn19, Ser24, Gln41, Thr89 and Gln121 (Supplementary Fig. S5). Given the high resolution of the X-ray data, the structural analysis includes a comparison of hydrogen bonds in the three structures, as this is of central interest for an understanding of differences in thermal stability.

Even at this high resolution, the electron-density maps in specific regions do not allow unambiguous modelling. Thus, the results from *HBPLUS* (McDonald & Thornton, 1994) were filtered considering the RSCC and RSZO metrics from *EDSTATS* (Tickle, 2012; details are shown in Supplementary Figs. S6–S8) to ensure the reliability of the subsequent analysis. The following residues did not comply with the applied cutoffs in one or more of the structures: Gly0, Ala9,

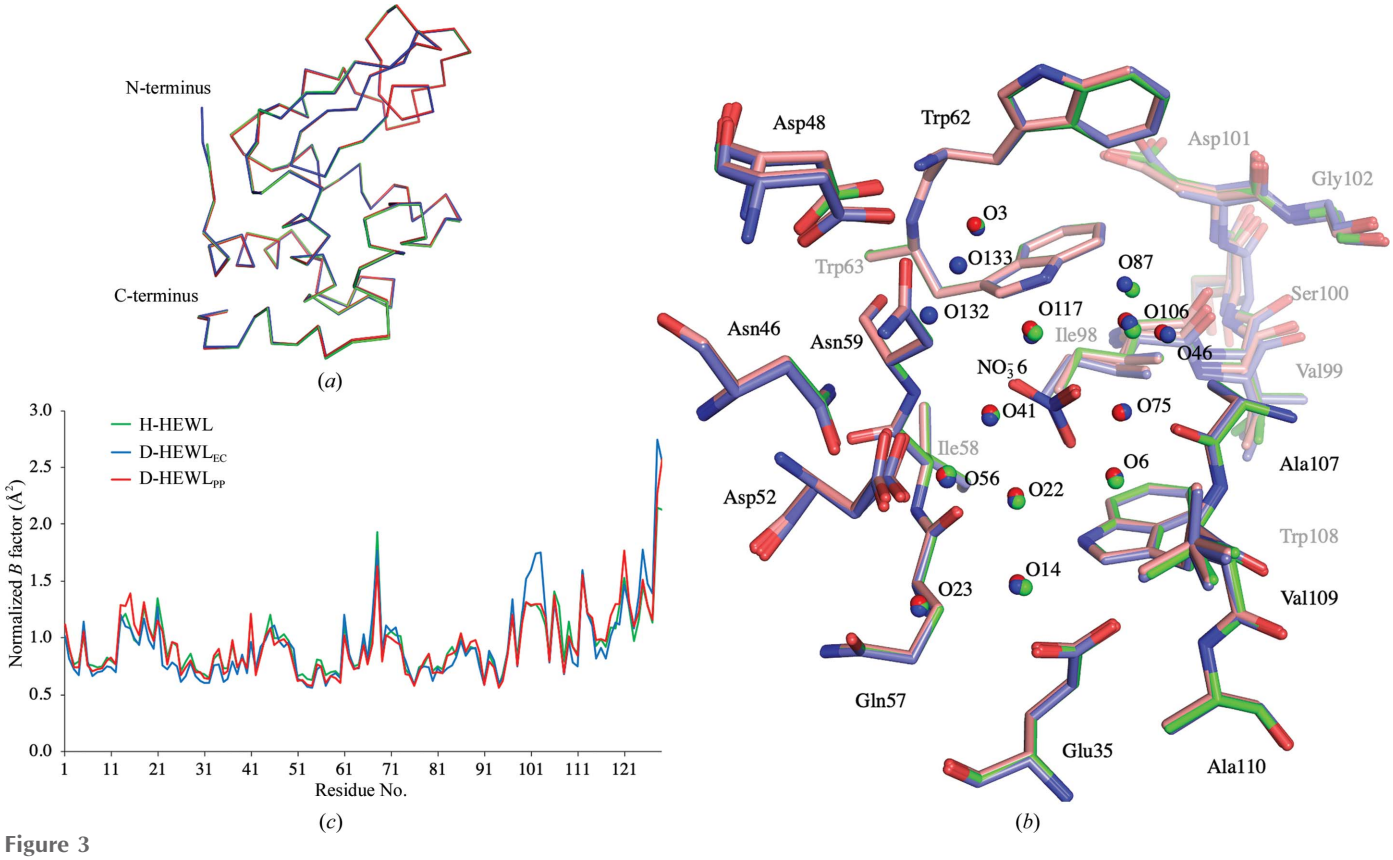


Figure 3
The overall structures and normalized *B* factors of the three HEWL variants. (*a*) Ribbon representation of the structurally aligned models of H-HEWL (green), D-HEWL$_{EC}$ (blue) and D-HEWL$_{PP}$ (red). (*b*) The active site and the polysaccharide-binding cleft shown for all three molecules: H-HEWL (green), D-HEWL$_{EC}$ (blue) and D-HEWL$_{PP}$ (red). (*c*) Plot of the normalized residue-averaged *B* factors from the H-HEWL (green), D-HEWL$_{EC}$ (blue) and D-HEWL$_{PP}$ (red) models.

Gly26, Asn27, Ala32, Phe38, Tyr53, Leu56, Ile58, Trp62, Cys64, Thr89, Ser91, Val92, Asp101, Gly102, Asn103, Gly104, Met105, Asn106, Ala107, Cys127 and Leu129. Discrepancies in hydrogen-bond lengths larger than 0.1 Å between all three structures were considered and inspected individually.

A comparison of the active site, with the catalytic residues Asp35 and Glu52, and the polysaccharide-binding cleft (Phillips, 1967) initially showed only minor differences in residue positions and conformations [Fig. 3(*b*)]. However, the residues Lys97–Gly104 display a high level of disorder, reflected by increased *B* values, most noticeably in the structure of D-HEWL$_{EC}$ [Fig. 3(*c*)]. As also reported by Wang *et al.* (2007), this region contains main-chain disorder due to a partial peptide-plane flip of Asn103, which causes strain on the backbone of residues Lys97–Gly104 (Fig. 4), propagating through hydrogen-bond interactions. The occupancy of the loop conformation associated with the flipped Asn103 (conformation *B*) was 46% in D-HEWL$_{EC}$, 38% in D-HEWL$_{PP}$ and 33% in H-HEWL. Structural alignment of this region (Lys97–Gly104, using all atoms) showed that in comparison with H-HEWL, D-HEWL$_{EC}$ and D-HEWL$_{PP}$ deviate by 0.27 and 0.16 Å, respectively. Meanwhile, the r.m.s.d. between D-HEWL$_{EC}$ and D-HEWL$_{PP}$ was 0.21 Å.

Main-chain disorder was also observed in the Lys13–Gly16 region, which is part of the first α-helix, with variations in the Gly16 N–Lys13 O hydrogen bond (Supplementary Fig. S9). However, this relates to the disorder of the His15 side chain, together with the interaction of Lys13 with the C-terminal residue Leu129 and of Gly16 with the disordered Arg114 via crystal contacts. The alternate conformations of His15, *A* and *B*, appear to be stabilized by water-mediated hydrogen bonds to Asn93 and by a hydrogen-bond to nitrate ion 9, respectively (Fig. 5).

The most evident differences between the structures in this region are the disorder of Thr89 in H-HEWL, and more profoundly in D-HEWL$_{PP}$, and the absence of water 81 in D-HEWL$_{PP}$. For His15*A*, water 81 seems to be important in restraining Thr89 in H-HEWL and D-HEWL$_{EC}$, contrary to the observation in D-HEWL$_{PP}$. In the absence of water 81 in D-HEWL$_{PP}$, a significant displacement of Thr89 occurs, stabilizing the His15 side chain. Furthermore, a steric clash with Thr89 appears to force flipping of the Asp87 side chain. The visualization of this extended hydrogen-bond network is supported by the similar refined occupancies of His15*A*, water 57, Thr89*B* and Asp87*B* of 47%, 34%, 39% and 41%, respectively. Meanwhile, in the H-HEWL and D-HEWL$_{EC}$ structures, His15*A* interacts with Asn93 and Asp87 through hydrogen bonds mediated by waters 57 and 81, as shown by their refined occupancies (51% for His15*A*, 65% for water 57 and 55% for water 81 in H-HEWL; 58% for His15*A*, 70% for water 57 and 55% for water 81 in D-HEWL$_{EC}$).

On the other hand, His15*B* in all three HEWL structures forms a hydrogen bond to nitrate ion 9, which is further stabilized by hydrogen bonds to Ile88 N and water 72. This interaction network is supported by the refined occupancies of His15*B* and nitrate ion 9 (42% and 49% in H-HEWL, 49% and 56% in D-HEWL$_{EC}$ and 66% and 47% in D-HEWL$_{PP}$, respectively). The low *B* factor refined for the O2 atom of this nitrate ion revealed the presence of a water molecule when the nitrate is not occupying the space (the occupancy of nitrate 9 O2 is 1, while the nitrate occupancy is refined based on N, O1 and O3). Additionally, in D-HEWL$_{PP}$ the His15*B* side chain forms a hydrogen bond to the nitrate ion, which replaces its interaction with Thr89 and promotes the interaction of Thr89*A* with Asp87*A*, as shown by their matching occupancies of 61% and 59%, respectively.

The presence of Gly0 at the N-terminus of D-HEWL$_{EC}$ influences the hydrogen-bond pattern in this region. Specifically, Gly0 cancels the Lys1 N–Thr40 OG1 interaction, instead favouring a Thr40 OG1–Lys1 O hydrogen bond (Fig. 6). Additionally, Gly0 does not interact with other protein residues and increases the disorder of the N-terminus of D-HEWL$_{EC}$. In H-HEWL and D-HEWL$_{PP}$, water molecule 138 occupies the position of Gly0 and enables water-mediated hydrogen bonds between Lys1 N and Ser86*B* OG.

In addition, several minor differences between the three structures were noted, where D-HEWL$_{EC}$ in particular stands out. In D-HEWL$_{EC}$ Asn19 was observed in a single confor-

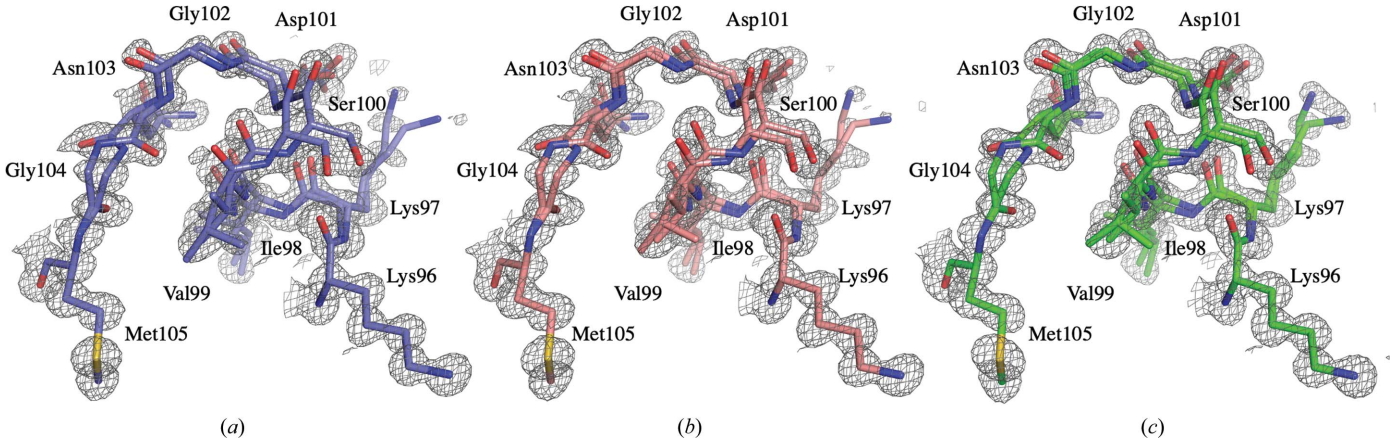


Figure 4
Increased disorder in the Lys97–Gly104 region of D-HEWL$_{EC}$ compared with both D-HEWL$_{PP}$ and H-HEWL. Representation of the backbone disorder resulting from the strain induced by the Asn103 partial peptide flip in D-HEWL$_{EC}$ (*a*), D-HEWL$_{PP}$ (*b*) and H-HEWL (*c*). The $2F_o - F_c$ electron-density maps represented are contoured at 1σ.

mation, allowing a stable Gly22 N–Asn19 O hydrogen bond of 2.96 Å, while in D-HEWL$_{PP}$ and H-HEWL disorder was observed, with the major conformation (occupancies of 60% and 69%, respectively) resulting in a significantly longer Gly22 N–Asn19 O hydrogen bond (Supplementary Fig. S10). This variation is correlated with the alternate conformations of the Asn19 side chain in H-HEWL and D-HEWL$_{PP}$, where the minor conformation of Asn19 participates in crystal contacts with Ser81 O, while the major conformation is involved in crystal contacts with the disordered Gln41 OE1. In D-HEWL$_{EC}$ only the latter conformation is present, as Gln41 is ordered, resulting in a single conformation of Asn19 with the shorter intramolecular Gly22 N–Asn19 O hydrogen bond. This shorter interaction suggests a more stable 3_{10}-helix between Tyr20 and Gly22 in D-HEWL$_{EC}$, although this may be a consequence of the stable crystal contact between the side chains of Asn19 and Gln41, thus not influencing stability in solution.

In all three structures Ser81 adopts two distinct conformations, giving rise to different Leu84 N–Ser81 O hydrogen-bond lengths (Supplementary Fig. S11), where the major conformation corresponds to the shorter of the two interactions. However, the lower occupancy of this major conformation in D-HEWL$_{EC}$ (66% compared with 82% in both D-HEWL$_{PP}$ and H-HEWL), together with the larger difference in the hydrogen-bond lengths of the two conformations, indicates that the Leu84 N–Ser81 O interaction is potentially weaker in D-HEWL$_{EC}$, destabilizing its 3_{10}-helix.

Furthermore, in another 3_{10}-helix (Val120–Arg125), minor variations were observed in the Arg125 NH2–Asp119 OD2 and Arg125 NH2–Gln121*B* OE1 hydrogen bonds, with the shorter Arg125 NH2–Asp119 OD2 interactions found in H-

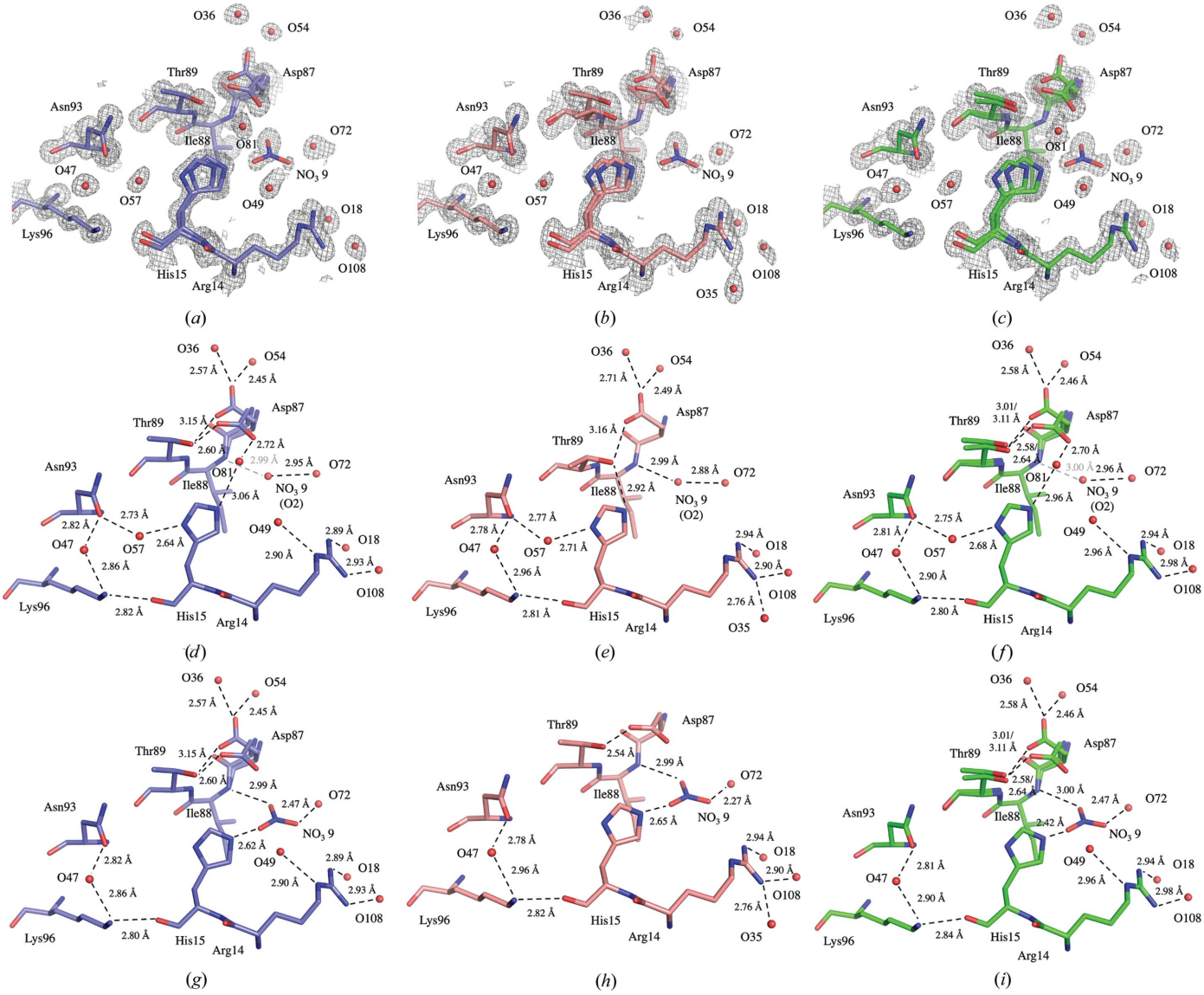


Figure 5
Disorder and hydrogen-bond patterns surrounding the His15 side chain. A representation is shown of the overall environment around His15 in D-HEWL$_{EC}$ (*a*), D-HEWL$_{PP}$ (*b*) and H-HEWL (*c*). The $2F_o - F_c$ electron-density maps represented are contoured at 1σ. Highlighted hydrogen-bond interactions correlated with His15 side-chain disorder are shown for conformation *A* of D-HEWL$_{EC}$ (*d*), D-HEWL$_{PP}$ (*e*) and H-HEWL (*f*) and for conformation *B* of D-HEWL$_{EC}$ (*g*), D-HEWL$_{PP}$ (*h*) and H-HEWL (*i*).

HEWL (Supplementary Fig. S12). Additionally, in D-HEWL$_{EC}$ Arg5 forms longer side chain–main chain hydrogen bonds to Trp123 O and Arg125 O, respectively, representing a minor destabilization of the tertiary structure in comparison to D-HEWL$_{PP}$ and H-HEWL.

4. Discussion

By using an *E. coli* expression system in parallel with in-column protein refolding, it is possible to obtain a more than threefold gain in the production of D-HEWL in comparison with yields for the *P. pastoris* system. The increase in yield is proportional to the financial cost reduction of protein production, since the approximate cost per litre of *E. coli* and *P. pastoris* cultures is similar. The cost is dominated by the deuterated materials, which for D-HEWL production using the *E. coli* system is roughly 140 euros per milligram of protein, in comparison to approximately 450 euros per milligram using *P. pastoris*. Although non-optimal in separating monomeric lysozyme from denaturing salts, the SEC column used in refolding provided the highest yields when compared with analytical columns. This observation is related to difficulties in removing such high concentrations of salt and the need to separate oligomeric from monomeric fractions while injecting milligram amounts of sample. Furthermore, the yield of the protocol can be further increased by dialyzing the oligomeric, misfolded and partially unfolded fractions of D-HEWL$_{EC}$ from refolding against denaturing buffer and reinjecting them into a refolding SEC. Complete perdeuteration of non-exchangeable sites in both D-HEWL variants was demonstrated by mass spectrometry. A similar refolding approach has been applied for the production of a perdeuterated antifreeze protein (Petit-Haertlein *et al.*, 2009), with the difference that refolding was carried out in a deuterated buffer. The refolding of perdeuterated lysozyme reported here is, to our knowledge, the first example of a perdeuterated protein exceeding 7 kDa and with multiple disulfide bonds. Refolding in D_2O was also attempted; however, it led to a decrease in the refolding yield (data not shown) owing to reduced separation of the monomeric protein fraction and denaturing salts. This observation is likely to be due to the slower dynamics in heavy water, resulting in a delay in the elution of the folded monomeric lysozyme fraction. Additionally, using D_2O would not be cost-effective, given the numerous refolding SEC runs that are required to obtain several milligrams of refolded protein. As the refolding of D-HEWL$_{EC}$ was performed in H_2O buffer, it may result in the caging of H atoms in exchangeable positions, *i.e.* exchanged during the unfolded state and then trapped upon refolding. The protein fold may keep specific regions protected from any interaction with solvent molecules; hence, to exchange these H atoms to D atoms the protein must be at least partially unfolded in D_2O buffer. To unambiguously identify the positions occupied by caged H atoms in the protein structure, neutron crystallography or NMR experiments are required. An indication of relevant positions is found in a reverse setup, where 20 H atoms were exchanged to D using unfolding and refolding processes of H-HEWL in D_2O (Kita & Morimoto, 2016), as observed in the neutron structure deposited in the PDB (PDB entry 6k8g; Kita & Morimoto, 2020).

Biophysical characterization of both D-HEWL variants and commercially available unlabelled HEWL shows that both D-HEWL molecules are stable and active. The perdeuterated variants showed lower thermal stability relative to the hydrogenated protein both in D_2O and H_2O buffers, in line with what has been reported in several biophysical studies on protein deuteration (Berns, 1963; Hattori *et al.*, 1965; Brockwell *et al.*, 2001; Meilleur *et al.*, 2004; Koruza *et al.*, 2018; Nichols *et al.*, 2020). Additionally, it seems that both hydrogenated and perdeuterated forms of HEWL have an increased transition temperature in D_2O compared with H_2O, as described in previous studies (Makhatadze *et al.*, 1995; Efimova *et al.*, 2007). However, the data presented here are not sufficient to draw definitive conclusions on this solvent-isotope effect, since the D_2O and H_2O buffers used have significantly different compositions (aimed at crystallization and activity measurements, respectively). Additionally, the presence of residual H atoms in H-HEWL, due to the limited

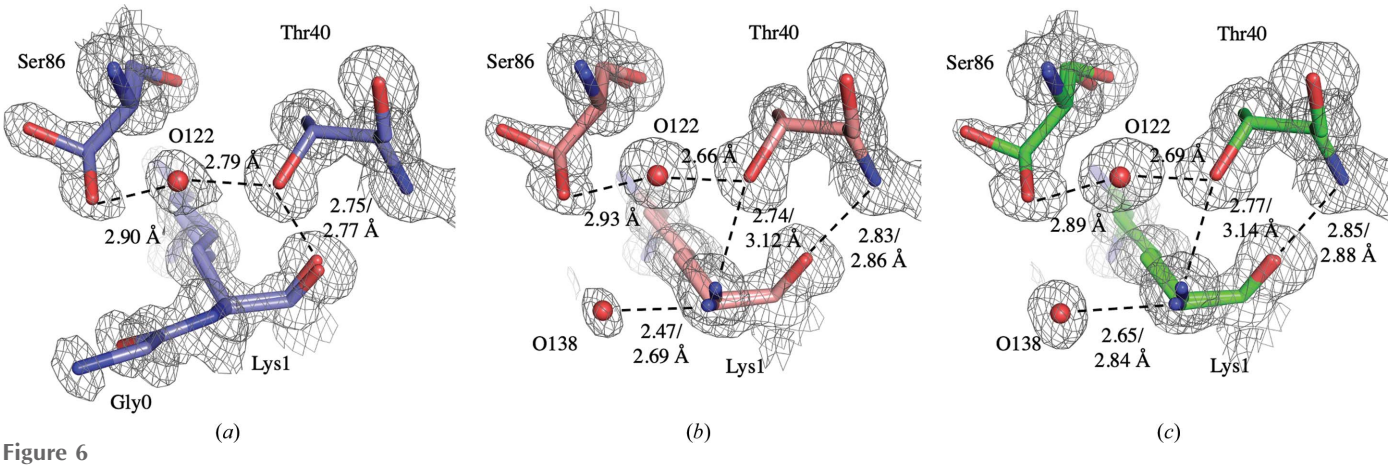


Figure 6
Representation of the differences in the hydrogen-bond patterns involving Lys1 and Thr40 in D-HEWL$_{EC}$ with the additional Gly0 residue (*a*), D-HEWL$_{PP}$ (*b*) and H-HEWL (*c*). The $2F_o - F_c$ electron-density maps represented are contoured at 1σ.

time for H/D exchange and the limited solvent accessibility of specific protein regions to the D_2O solvent, cannot be ruled out. The differences in protein thermal stability can be correlated with the enzymatic activities. Perdeuteration of the protein was expected to affect its dynamics and consequently its stability and activity, and in this study a decrease in stability as well as in relative activity compared with H-HEWL is observed. The differences between D-HEWL$_{EC}$ and D-HEWL$_{PP}$ are likely to be a consequence of the refolding procedure. The additional N-terminal glycine residue in D-HEWL$_{EC}$ may also cause a slight destabilization of the protein. However, the activity results do not allow a conclusive correlation of the effect of refolding on activity, since under the conditions used the difference in activity between the two perdeuterated variants is not statistically significant. This further emphasizes the similarity between the D-HEWL variants and validates the refolding approach to obtain stable and active D-HEWL.

While a large number of HEWL crystal structures have been published, the detailed comparisons needed for this study of perdeuterated and hydrogenated HEWL required the growth of crystals under closely comparable conditions, with only minor variations relating to the seeding procedure and precipitant concentrations. The atomic resolution X-ray data for both perdeuterated samples, as well as for the reference unlabelled sample, have been analyzed in detail, revealing structural features that can be related to the observations on stability and activity. The crystal packing and overall structures were, as expected, found to be essentially identical, with negligible differences in the unit-cell parameters. Moreover, the nitrate and acetate ions that are essential to crystallization were located and refined in identical positions in the three models, with similar *B* factors (Table 1). However, despite the close similarity between the three structures, there are some clear variations in hydrogen-bond distances, which appear to be related to the differences in protein stability.

An important factor contributing to the reduced thermal stability of the D-HEWL structures is the effect of H/D substitution on hydrophobic interactions. As described by Hattori *et al.* (1965), deuterium-substituted nonpolar amino-acid side chains have a reduced steric requirement due to the smaller amplitudes of vibration of the C—D bond compared with C—H, leading to weaker hydrophobic interactions between the residue side chains; this has also been noted in mass-spectrometric studies (Yee *et al.*, 2016). Additionally, D_2O has a stronger hydrophobic effect than H_2O, leading to changes in solvation, more compact structures and a decrease in protein flexibility (Svergun *et al.*, 1998; Sasisanker *et al.*, 2004; Efimova *et al.*, 2007; Jasnin *et al.*, 2008). This is observed in the crystal structures, where a larger number of structural water molecules were identified in H-HEWL compared with both D-HEWL variants. Moreover, the molecular surface and solvent-accessible surface areas of D-HEWL$_{PP}$ were 15 555 and 8 200 Å^2, respectively, whereas those for H-HEWL were 15 725 and 8 274 Å^2. The corresponding values for D-HEWL$_{EC}$ are not directly comparable due to the presence of the additional Gly0 residue. Finally, protein dynamics are expected to be influenced by deuteration since D is twice as heavy as H, which in the case of HEWL corresponds to a mass increase of at least ∼700 Da. All of these factors play a role in the interaction with substrate molecules, since the enzymatic activity is strongly dependent on protein dynamics and the displacement of water molecules to accommodate the substrate, consistent with the decreased activity observed in the perdeuterated variants.

The disorder observed in the structures is evidently linked to the intricate networks of hydrogen bonds. However, only a few regions of the models show distinct disorder due to variations in the hydrogen-bond patterns. These are the cases of the Thr40 N–Lys1 O and Gly22 N–Asn19 O hydrogen bonds and the His15 side chain. The differences observed in the Thr40 N–Lys1 O interaction are due to the presence of Gly0 at the N-terminus of D-HEWL$_{EC}$, leading to increased disorder. In the case of the Gly22 N–Asn19 O hydrogen bond, the alternate conformation of Asn19 is favoured by the side-chain disorder of Gln41 that is present in D-HEWL$_{PP}$ and H-HEWL, resulting in a weaker Gly22 N–Asn19 O interaction. Finally, the His15 side-chain disorder, with differentiation between the two conformational networks *A* and *B*, appears to be linked to partial occupancies of waters 57 and 81 and of nitrate ion 9, and potentially to variations in protonation states. In D-HEWL$_{PP}$, the absence of water 81 seems to promote the disorder of Thr89 and subsequently the flipping of Asp87 to stabilize His15*A*. The protonation states are not evident, even in the 0.65 Å resolution structure (Wang *et al.*, 2007), and obtaining an unambiguous picture of the protonation of lysozyme will require high-quality and high-resolution neutron diffraction data. In conclusion, these minor variations in the protein structure alone are not likely to explain the decrease in stability observed in the D-HEWL structures.

The main difference in the crystal structures that can be correlated with variations in protein stability is the disorder of the Lys97–Gly104 region due to the partial peptide-plane flip of Asn103. Peptide-plane flipping occurs in the early stages of protein folding, particularly when glycine is in the $i + 1$ position, since the structure is not yet restrained by hydrogen bonds between protein residues (Hayward, 2001). Although not frequent due to its energetically unfavored conformation, peptide flipping remains underrepresented in the PDB (Berman *et al.*, 2000). This was found to be correlated, among other factors, with the resolution of the X-ray data available to determine the crystal structures (Stewart *et al.*, 1990; Weiss *et al.*, 1998). Peptide flipping can be responsible for amyloid formation (Milner-White *et al.*, 2006; Yang *et al.*, 2006) or can confer structural flexibility that is essential for protein function (Weiss *et al.*, 1998; Ludwig *et al.*, 1997; Keedy *et al.*, 2015). As described by Wang *et al.* (2007), the backbone disorder in this region is a consequence of the Asn103 peptide-plane flip. In their H-HEWL crystal structure determined from X-ray data at 0.65 Å resolution, the flipped conformation has a refined occupancy of 35%, which is consistent with our H-HEWL model in which the flipped conformation of Asn103 was refined with an occupancy of 33%. This observation suggests that the likelihood of Asn103 peptide flipping in

native H-HEWL is constant. Conversely, in the D-HEWL models the refined occupancies for the flipped conformation are greater: 46% in D-HEWL$_{EC}$ and 38% in D-HEWL$_{PP}$. While D-HEWL$_{EC}$ was chemically unfolded and then refolded by slowly changing its buffer from 6 *M* guanidine–HCl to a 2 *M* urea H_2O solution, D-HEWL$_{PP}$ was folded in deuterated conditions during expression. Thus, both D-HEWL variants were subjected to different folding environments compared with H-HEWL, which are associated with slower solvent dynamics and the H/D-isotope effect, which could favour the peptide-plane flip of Asn103. Interestingly, when the protein is completely unfolded, as is the case for D-HEWL$_{EC}$, it appears that the probability of the peptide flip occurring or not is identical, suggesting a high degree of freedom between the two conformations. In the case of D-HEWL$_{PP}$, the solvent-isotope effect may be responsible for this by slowing down the folding dynamics and increasing the likelihood of peptide flipping. This destabilized region is not only part of the enzyme active site, and therefore relevant to substrate binding, as reported by Strynadka & James (1991), but also protects a hydrophobic pocket containing Trp28, Trp62, Trp63 and Trp108. The increase in disorder of this loop region may therefore be correlated with the decrease in protein thermal stability measured for D-HEWL$_{EC}$ when compared with D-HEWL$_{PP}$.

The results presented here support the widespread understanding that perdeuteration has no significant effect on secondary and tertiary protein structures. Nevertheless, the hydrophobic effect and the slower dynamics caused by perdeuteration have an impact on protein stability and activity. Ultimately, this study emphasizes the capability to use *E. coli* for the expression of recombinant insoluble protein and subsequent refolding for the production of large amounts of perdeuterated material, enabling a wide range of new science in the future. In addition, this work highlights the fact that studies of deuterated proteins can reveal crucial and highly specific aspects of protein conformation related to variations in protein thermal stability.

5. Data accessibility

The X-ray diffraction data and models have been deposited in the PDB with accession codes 7ave (D-HEWL$_{EC}$), 7avf (H-HEWL) and 7avg (D-HEWL$_{PP}$).

Acknowledgements

JR thanks the Institut Laue–Langevin (ILL) for a PhD studentship. All authors thank the ILL for access to its Deuteration Laboratory within the Partnership for Structural Biology (PSB; https://www.psb-grenoble.eu/). We acknowledge the MS platform of the Grenoble Instruct–ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the PSB. Platform access was supported by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). We thank Diamond Light Source for beamtime at I03 and MAX IV for beamtime at BioMAX, and also the beamline staff at both facilities for their support.

Funding information

We thank DANSCATT for travel support to MAX IV, made possible through the funding from the Danish Agency for Science, Technology and Innovation. AEL acknowledges funding from the Lundbeck Foundation Initiative BRAIN-STRUC. VTF acknowledges support from the UK Engineering and Physical Sciences Research Council (EPSRC), which funded the Deuteration Laboratory (D-Lab) in ILL's Life Sciences Group under grants GR/R99393/01 and EP/C015452/1.

References

Artero, J.-B., Härtlein, M., McSweeney, S. & Timmins, P. (2005). *Acta Cryst.* D**61**, 1541–1549.
Bai, N., Roder, H., Dickson, A. & Karanicolas, J. (2019). *Sci. Rep.* **9**, 2650.
Batas, B. & Chaudhuri, J. B. (1996). *Biotechnol. Bioeng.* **50**, 16–23.
Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.
Berns, D. S. (1963). *Biochemistry*, **2**, 1377–1380.
Biter, A. B., de la Peña, A. H., Thapar, R., Lin, J. Z. & Phillips, K. J. (2016). *Sci. Rep.* **6**, 18906.
Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature*, **206**, 757–761.
Blakeley, M. P. (2009). *Crystallogr. Rev.* **15**, 157–218.
Breyton, C., Gabel, F., Lethier, M., Flayhan, A., Durand, G., Jault, J., Juillan-Binard, C., Imbert, L., Moulin, M., Ravaud, S., Härtlein, M. & Ebel, C. (2013). *Eur. Phys. J. E*, **36**, 71.
Brockwell, D., Yu, L., Cooper, S., McCleland, S., Cooper, A., Attwood, D., Gaskell, S. J. & Barber, J. (2001). *Protein Sci.* **10**, 572–580.
Campbell, R. A., Tummino, A., Varga, I., Milyaeva, O. Y., Krycki, M. M., Lin, S. Y., Laux, V., Haertlein, M., Forsyth, V. T. & Noskov, B. A. (2018). *Langmuir*, **34**, 5020–5029.
Cuypers, M. G., Mason, S. A., Blakeley, M. P., Mitchell, E. P., Haertlein, M. & Forsyth, V. T. (2013). *Angew. Chem. Int. Ed.* **52**, 1022–1025.
Cuypers, M. G., Trubitsyna, M., Callow, P., Forsyth, V. T. & Richardson, J. M. (2013). *Nucleic Acids Res.* **41**, 2020–2033.
Darmanin, C., Strachan, J., Adda, C. G., Ve, T., Kobe, B. & Abbey, B. (2016). *Sci. Rep.* **6**, 25345.
Dunne, O., Weidenhaupt, M., Callow, P., Martel, A., Moulin, M., Perkins, S. J., Haertlein, M. & Forsyth, V. T. (2017). *Eur. Biophys. J.* **46**, 425–432.
Durbin, S. D. & Feher, G. (1986). *J. Cryst. Growth*, **76**, 583–592.
Efimova, Y. M., Haemers, S., Wierczinski, B., Norde, W. & Well, A. A. (2007). *Biopolymers*, **85**, 264–273.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Foglia, F., Hazael, R., Simeoni, G. G., Appavou, M., Moulin, M., Haertlein, M., Forsyth, V. T., Seydel, T., Daniel, I., Meersman, F. & McMillan, P. F. (2016). *Sci. Rep.* **6**, 18862.
Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. (2009). *FEBS J.* **276**, 2833–2840.
Geders, T. W., Gustafson, K. & Finzel, B. C. (2012). *Acta Cryst.* F**68**, 596–600.
Glasoe, P. K. & Long, F. A. (1960). *J. Phys. Chem.* **64**, 188–190.
Grage, S. L., Keleshian, A. M., Turdzeladze, T., Battle, A. R., Tay, W. C., May, R. P., Holt, S. A., Contera, S. A., Haertlein, M., Moulin,

M., Pal, P., Rohde, P. R., Forsyth, V. T., Watts, A., Huang, K. C., Ulrich, A. S. & Martinac, B. (2011). *Biophys. J.* **100**, 1252–1260.
Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O. & Forsyth, V. T. (2016). *Methods Enzymol.* **566**, 113–157.
Hattori, A., Crespi, H. L. & Katz, J. J. (1965). *Biochemistry*, **4**, 1213–1225.
Haupt, M., Blakeley, M. P., Fisher, S. J., Mason, S. A., Cooper, J. B., Mitchell, E. P. & Forsyth, V. T. (2014). *IUCrJ*, **1**, 429–438.
Hayward, S. (2001). *Protein Sci.* **10**, 2219–2227.
Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Cryst.* D**61**, 1413–1417.
Holdgate, G. A., Anderson, M., Edfeldt, F. & Geschwindner, S. (2010). *J. Struct. Biol.* **172**, 142–157.
Jasnin, M., Tehei, M., Moulin, M., Haertlein, M. & Zaccai, G. (2008). *Eur. Biophys. J.* **37**, 613–617.
Josts, I., Nitsche, J., Maric, S., Mertens, H. D., Moulin, M., Haertlein, M., Prevost, S., Svergun, D. I., Busch, S., Forsyth, V. T. & Tidow, H. (2018). *Structure*, **26**, 1072–1079.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 133–144.
Keedy, D. A., Fraser, J. S. & van den Bedem, H. (2015). *PLoS Comput. Biol.* **11**, e1004507.
Kehlenbeck, D.-M., Josts, I., Nitsche, J., Busch, S., Forsyth, V. T. & Tidow, H. (2019). *Biol. Chem.* **400**, 1509–1518.
Kita, A. & Morimoto, Y. (2016). *Mol. Biotechnol.* **58**, 130–136.
Kita, A. & Morimoto, Y. (2020). *J. Appl. Cryst.* **53**, 837–840.
Koruza, K., Lafumat, B., Nyblom, M., Mahon, B. P., Knecht, W., McKenna, R. & Fisher, S. Z. (2019). *Acta Cryst.* D**75**, 895–903.
Koruza, K., Lafumat, B., Végvári, Á., Knecht, W. & Fisher, S. Z. (2018). *Arch. Biochem. Biophys.* **645**, 26–33.
Langan, P., Sangha, A. K., Wymore, T., Parks, J. M., Yang, Z. K., Hanson, B. L., Fisher, Z., Mason, S. A., Blakeley, M. P., Forsyth, V. T., Glusker, J. P., Carrell, H. L., Smith, J. C., Keen, D. A., Graham, D. E. & Kovalevsky, A. (2014). *Structure*, **22**, 1287–1300.
Laux, V., Callow, P., Svergun, D. I., Timmins, P. A., Forsyth, V. T. & Haertlein, M. (2008). *Eur. Biophys. J.* **37**, 815–822.
Lea, W. A. & Simeonov, A. (2012). *PLoS One*, **7**, e36219.
Legrand, L., Riès-Kautt, M. & Robert, M.-C. (2002). *Acta Cryst.* D**58**, 1564–1567.
Li, J., Wang, Y. J. K., Wang, N., Li, G., Sun, N. & Liu, D. (2012). *Afr. J. Biotechnol.* **11**, 11887–11893.
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.
Liu, S. T., Saito, A., Azakami, H. & Kato, A. (2003). *Protein Expr. Purif.* **27**, 304–312.
Liu, X., Hanson, B. L., Langan, P. & Viola, R. E. (2007). *Acta Cryst.* D**63**, 1000–1008.
Ludwig, M. L., Pattridge, K. A., Metzger, A. L., Dixon, M. M., Eren, M., Feng, Y. & Swenson, R. P. (1997). *Biochemistry*, **36**, 1259–1280.
Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. (1995). *Nat. Struct. Mol. Biol.* **2**, 852–855.
Maric, S., Skar-Gislinge, N., Midtgaard, S., Thygesen, M. B., Schiller, J., Frielinghaus, H., Moulin, M., Haertlein, M., Forsyth, V. T., Pomorski, T. G. & Arleth, L. (2014). *Acta Cryst.* D**70**, 317–328.
Maric, S., Thygesen, M. B., Schiller, J., Marek, M., Moulin, M., Haertlein, M., Forsyth, V. T., Bogdanov, M., Dowhan, W., Arleth, L. & Pomorski, T. G. (2015). *Appl. Microbiol. Biotechnol.* **99**, 241–254.
McDonald, I. K. & Thornton, J. M. (1994). *J. Mol. Biol.* **238**, 777–793.
McPherson, A. & DeLucas, L. J. (2015). *NPJ Microgravity*, **1**, 15010.
Meilleur, F., Contzen, J., Myles, D. A. A. & Jung, C. (2004). *Biochemistry*, **43**, 8744–8753.
Milner-White, E. J., Watson, J. D., Qi, G. & Hayward, S. (2006). *Structure*, **14**, 1369–1376.
Mine, S., Ueda, T., Hashimoto, Y., Tanaka, Y. & Imoto, T. (1999). *FEBS Lett.* **448**, 33–37.
Miranker, A., Radford, S. E., Karplus, M. & Dobson, C. M. (1991). *Nature*, **349**, 633–636.
Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. (1993). *Science*, **262**, 896–900.
Moulin, M., Strohmeier, G. A., Hirz, M., Thompson, K. C., Rennie, A. R., Campbell, R. A., Pichler, H., Maric, S., Forsyth, V. T. & Haertlein, M. (2018). *Chem. Phys. Lipids*, **212**, 80–87.
Nichols, P. J., Falconer, I., Griffin, A., Mant, C., Hodges, R., McKnight, C. J., Vögeli, B. & Vugmeyster, L. (2020). *Protein Sci.* **29**, 1641–1654.
Nitsche, J., Josts, I., Heidemann, J., Mertens, H. D., Maric, S., Moulin, M., Haertlein, M., Busch, S., Forsyth, V. T., Svergun, D. I., Uetrecht, C. & Tidow, H. (2018). *Commun. Biol.* **1**, 206.
Petit-Haertlein, I., Blakeley, M. P., Howard, E., Hazemann, I., Mitschler, A., Haertlein, M. & Podjarny, A. (2009). *Acta Cryst.* F**65**, 406–409.
Phillips, D. C. (1967). *Proc. Natl Acad. Sci. USA*, **57**, 483–495.
Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). *Nature*, **358**, 302–307.
Ramos, J., Muthukumaran, J., Freire, F., Paquete-Ferreira, J., Otrelo-Cardoso, A., Svergun, D., Panjkovich, A. & Santos-Silva, T. (2019). *Int. J. Mol. Sci.* **20**, 860.
Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. (2013). *Acta Cryst.* F**69**, 209–214.
Sasisanker, P., Oleinikova, A., Weingärtner, H., Ravindra, R. & Winter, R. (2004). *Phys. Chem. Chem. Phys.* **6**, 1899–1905.
Sattler, M. & Fesik, S. W. (1996). *Structure*, **4**, 1245–1249.
Sears, V. F. (1992). *Neutron News*, **3**(3), 26–37.
Shabalin, I., Dauter, Z., Jaskolski, M., Minor, W. & Wlodawer, A. (2015). *Acta Cryst.* D**71**, 1965–1979.
Shindyalov, I. N. & Bourne, P. E. (1998). *Protein Eng. Des. Sel.* **11**, 739–747.
Shugar, D. (1952). *Biochim. Biophys. Acta*, **8**, 302–309.
Stewart, D. E., Sarkar, A. & Wampler, J. E. (1990). *J. Mol. Biol.* **214**, 253–260.
Strynadka, N. C. J. & James, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.
Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 2267–2272.
Tickle, I. J. (2012). *Acta Cryst.* D**68**, 454–467.
Varga, K., Aslimovska, L., Parrot, I., Dauvergne, M. T., Haertlein, M., Forsyth, V. T. & Watts, A. (2007). *Biochim. Biophys. Acta*, **1768**, 3029–3035.
Vidal, O., Robert, M.-C., Arnoux, B. & Capelle, B. (1999). *J. Cryst. Growth*, **196**, 559–571.
Vijayakrishnan, S., Kelly, S. M., Gilbert, R. J. C., Callow, P., Bhella, D., Forsyth, T., Lindsay, J. G. & Byron, O. (2010). *J. Mol. Biol.* **399**, 71–93.
Waldie, S., Lind, T. K., Browning, K., Moulin, M., Haertlein, M., Forsyth, V. T., Luchini, A., Strohmeier, G. A., Pichler, H., Maric, S. & Cárdenas, M. (2018). *Langmuir*, **34**, 472–479.
Waldie, S., Moulin, M., Porcar, L., Pichler, H., Strohmeier, G. A., Skoda, M., Forsyth, V. T., Haertlein, M., Maric, S. & Cárdenas, M. (2019). *Sci. Rep.* **9**, 5118.
Waldie, S., Sebastiani, F., Browning, K., Maric, S., Lind, T. K., Yepuri, N., Darwish, T. A., Moulin, M., Strohmeier, G., Pichler, H., Skoda, M. W. A., Maestro, A., Haertlein, M., Forsyth, V. T., Bengtsson, E., Malmsten, M. & Cárdenas, M. (2020). *Biochim. Biophys. Acta*, **1865**, 158769.
Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Cryst.* D**63**, 1254–1268.
Weiss, M. S., Jabs, A. & Hilgenfeld, R. (1998). *Nat. Struct. Mol. Biol.* **5**, 676.
Wildegger, G. & Kiefhaber, T. (1997). *J. Mol. Biol.* **270**, 294–304.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst.* D**67**, 235–242.

Wood, K., Gallat, F. X., Otten, R., van Heel, A. J., Lethier, M., van Eijck, L., Moulin, M., Haertlein, M., Weik, M. & Mulder, F. A. A. (2013). *Angew. Chem. Int. Ed.* **52**, 665–668.

Yang, M., Lei, M., Yordanov, B. & Huo, S. (2006). *J. Phys. Chem. B*, **110**, 5829–5833.

Yee, A. W., Aldeghi, M., Blakeley, M. P., Ostermann, A., Mas, P. J., Moulin, M., de Sanctis, D., Bowler, M. W., Mueller-Dieckmann, C., Mitchell, E. P., Haertlein, M., de Groot, B. L., Boeri Erba, E. & Forsyth, V. T. (2019). *Nat. Commun.* **10**, 925.

Yee, A. W., Moulin, M., Breteau, N., Haertlein, M., Mitchell, E. P., Cooper, J. B., Boeri Erba, E. & Forsyth, V. T. (2016). *Angew. Chem. Int. Ed.* **55**, 9292–9296.

IUCrJ

Volume 8 (2021)

Supporting information for article:

Structural insights into protein folding, stability and activity using *in vivo* perdeuteration of hen egg-white lysozyme

Joao Ramos, Valerie Laux, Michael Haertlein, Elisabetta Erba Boeri, Katherine E. McAuley, V. Trevor Forsyth, Estelle Mossou, Sine Larsen and Annette E. Langkilde

a

```
atg ggt aag gtt ttt ggt cgt tgc gaa ctg gcg gcg gcg atg aag cgt cac ggt ctg gac
(M)  G   K   V   F   G   R   C   E   L   A   A   A   M   K   R   H   G   L   D

aat tat cgt ggt tat agc ctg ggt aac tgg gtg tgc gcg gcg aag ttc gag agc aac ttt
 N   Y   R   G   Y   S   L   G   N   W   V   C   A   A   K   F   E   S   N   F

aac acc cag gcg acc aac cgt aac acc gac ggt agc acc gat tac ggc atc ctg caa att
 N   T   Q   A   T   N   R   N   T   D   G   S   T   D   Y   G   I   L   Q   I

aac agc cgt tgg tgg tgc aac gat ggt cgt acc ccg ggc agc cgt aac ctg tgc aac atc
 N   S   R   W   W   C   N   D   G   R   T   P   G   S   R   N   L   C   N   I

ccg tgc agc gcg ctg ctg agc agc gac att acc gcg agc gtg aac tgc gcg aag aaa atc
 P   C   S   A   L   L   S   S   D   I   T   A   S   V   N   C   A   K   K   I

gtt agc gat ggt aac ggc atg aac gcg tgg gtt gcg tgg cgt aac cgt tgc aaa ggc acc
 V   S   D   G   N   G   M   N   A   W   V   A   W   R   N   R   C   K   G   T

gat gtt cag gcg tgg att cgt ggc tgc cgt ctg taa
 D   V   Q   A   W   I   R   G   C   R   L   -
```

b

```
aaa gtg ttt ggc cgt tgc gaa ctg gcg gcg gcg atg aaa cgt cat ggc ctg gat aac acc
 K   V   F   G   R   C   E   L   A   A   A   M   K   R   H   G   L   D   N   T

cag gcg acc aac cgt aac acc gat ggc agc acc gat tat ggc att ctg cag att aac tat
 Q   A   T   N   R   N   T   D   G   S   T   D   Y   G   I   L   Q   I   N   Y

cgt ggc tat agc ctg ggc aac tgg gtg tgt gcc gcc aaa ttt gaa agc aac ttc aac agc
 R   G   Y   S   L   G   N   W   V   C   A   A   K   F   E   S   N   F   N   S

cgt tgg tgg tgt aac gat ggc cgt acc ccg ggc agc cgt aac ctg tgt aac att ccg tgt
 R   W   W   C   N   D   G   R   T   P   G   S   R   N   L   C   N   I   P   C

agc gcc ctg ctg tct agc gat att acc gcc agc gtg aac tgt gcc aaa aaa att gtg agc
 S   A   L   L   S   S   D   I   T   A   S   V   N   C   A   K   K   I   V   S

gat ggc aac ggc atg aac gcc tgg gtg gcg tgg cgt aac cgt tgt aaa ggc acc gat gtt
 D   G   N   G   M   N   A   W   V   A   W   R   N   R   C   K   G   T   D   V

cag gcc tgg att cgt ggc tgc cgt ctg taa
 Q   A   W   I   R   G   C   R   L   -
```

Figure S1

DNA coding and protein sequences for the recombinantly expressed proteins, D-HEWL$_{EC}$ (a) and D-HEWL$_{PP}$ (b).

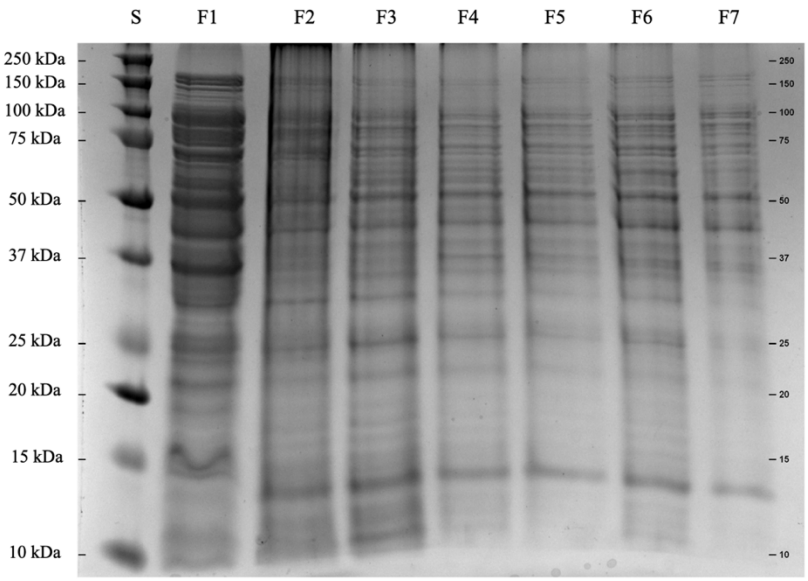


Figure S2

Samples from the inclusion body washing steps in a 12% SDS-PAGE gel – lanes: S. Precision Plus Protein™ Dual Xtra Standards (Bio-Rad); F1. Soluble fraction from cell lysis; F2. Soluble fraction from 1st washing cycle; F3. Soluble fraction from 2nd washing cycle; F4. Soluble fraction from 3rd washing cycle; F5. Soluble fraction from 4th washing cycle; F6. Soluble fraction from 5th washing cycle; F7. Soluble fraction from 6th washing cycle.

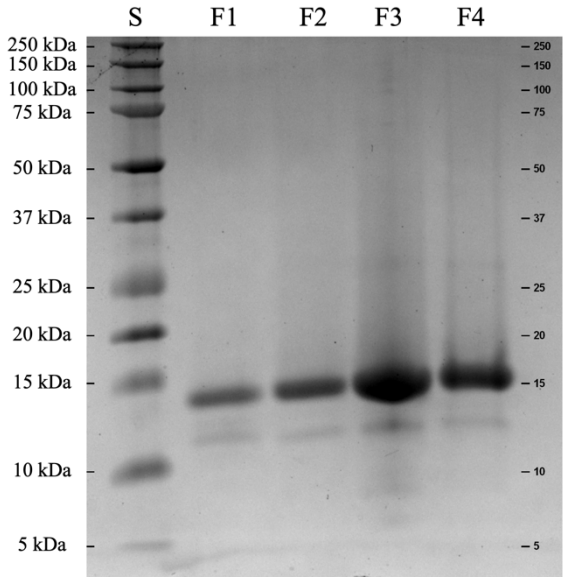


Figure S3

Samples from D-HEWL refolding SEC and desalting to deuterated protein buffer in a 12% SDS-PAGE gel – lanes S. Precision Plus Protein™ Dual Xtra Standards (Bio-Rad); F1. Fractions from refolding SEC before 0.9 CV; F2. Refolded monomeric D-HEWL in refolding buffer; F3. Refolded monomeric D-HEWL in protein buffer in D_2O; F4. Diluted sample of F3.

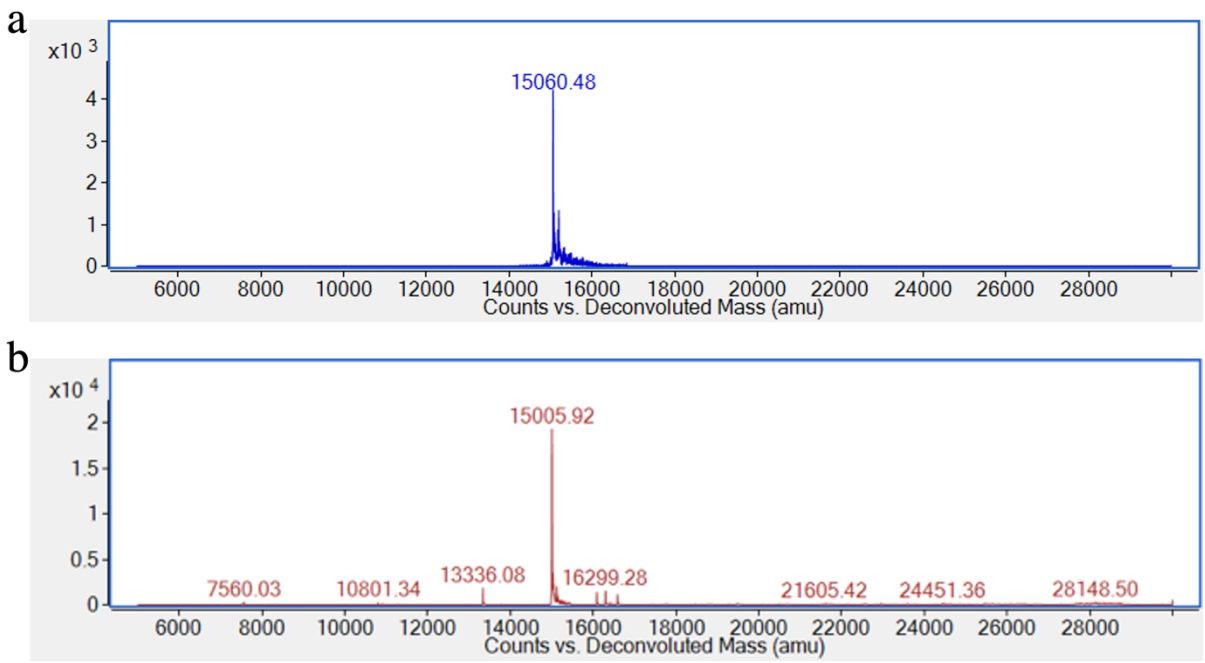


Figure S4

Mass spectra of D-HEWL$_{EC}$ (a) and D-HEWL$_{PP}$ (b) from LC/ESI-MS analysis. The most abundant masses observed were of 15060 and 15005 Da, for D-HEWL$_{EC}$ and D-HEWL$_{PP}$, respectively.

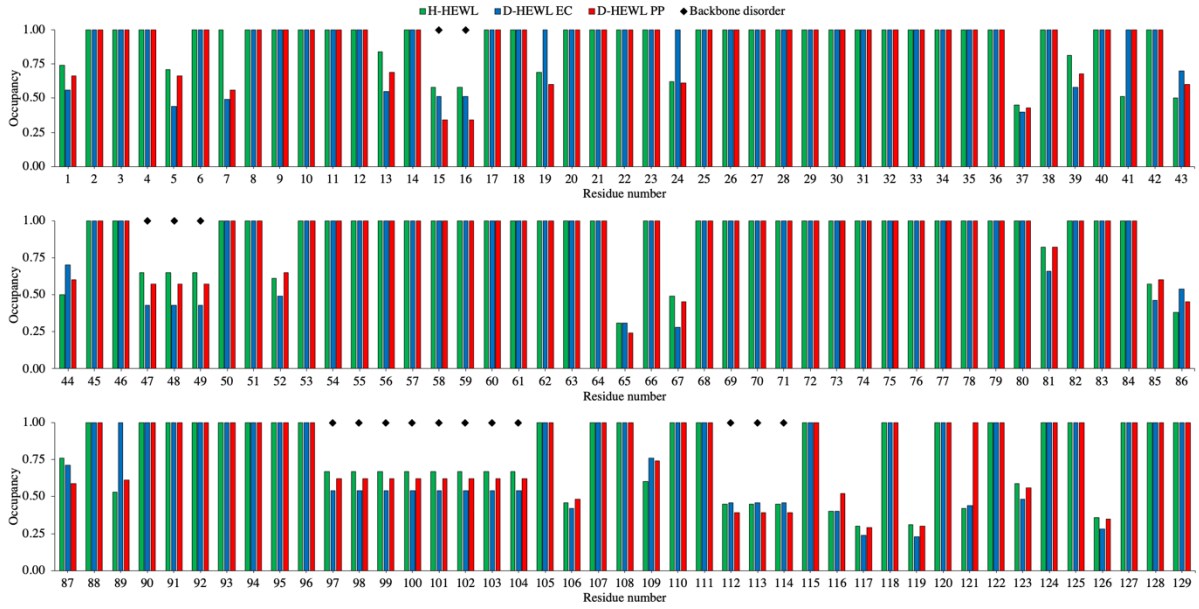


Figure S5

Residue occupancies of conformation A by residue number in H-HEWL (green), D-HEWL$_{EC}$ (blue) and D-HEWL$_{PP}$ (red). An occupancy of 1 represents single confirmations, while occupancies <1 indicate alternate residue conformations (modelled with a total residue occupancy of 1). Regions containing backbone disorder are highlighted (♦).

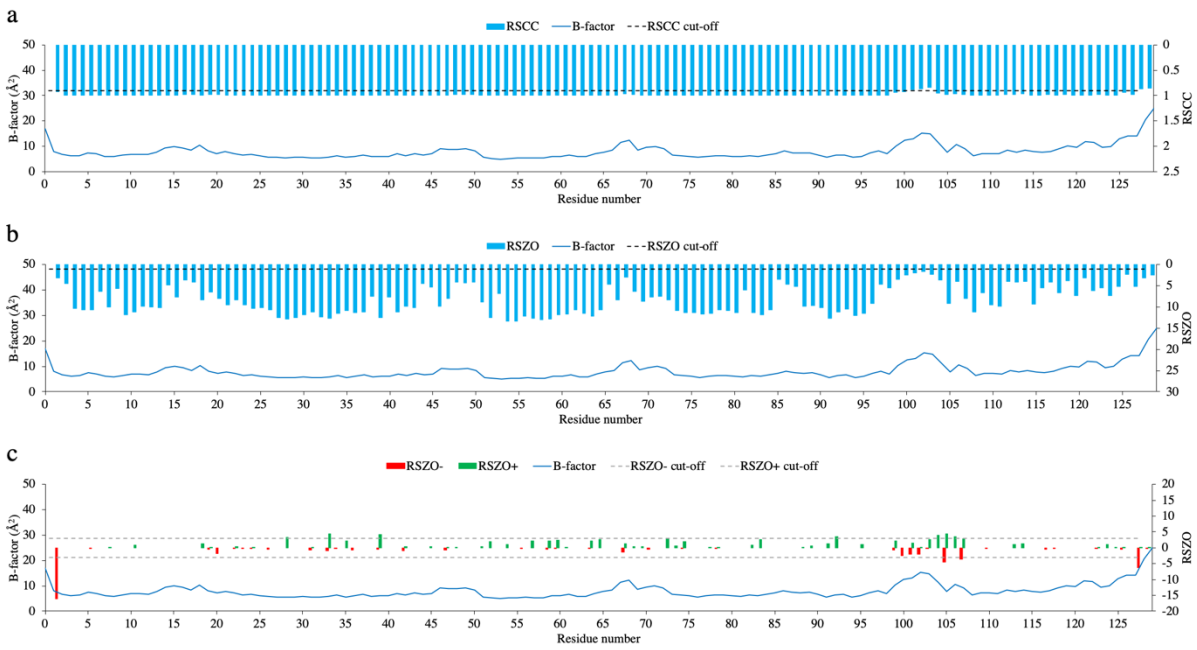


Figure S6

Validation of the quality of the D-HEWL$_{EC}$ model against the X-ray data, according to the RSCC (a), RSZD (b), RSZD- and RSZD+ (c), metrics from EDSTATS. The cut-offs applied were: 90% for RSCC, 1 σ for RSZD, -3 σ for RSZD- and +3 σ for RSZD+. In the D-HEWL$_{EC}$ model, the residues excluded from the detailed H-bond analysis were Gly0, Asn27, Ala32, Phe38, Val92, Asp101, Gly102, Asn103, Gly104, Met105, Asn106, Ala107, Arg128 and Leu129.

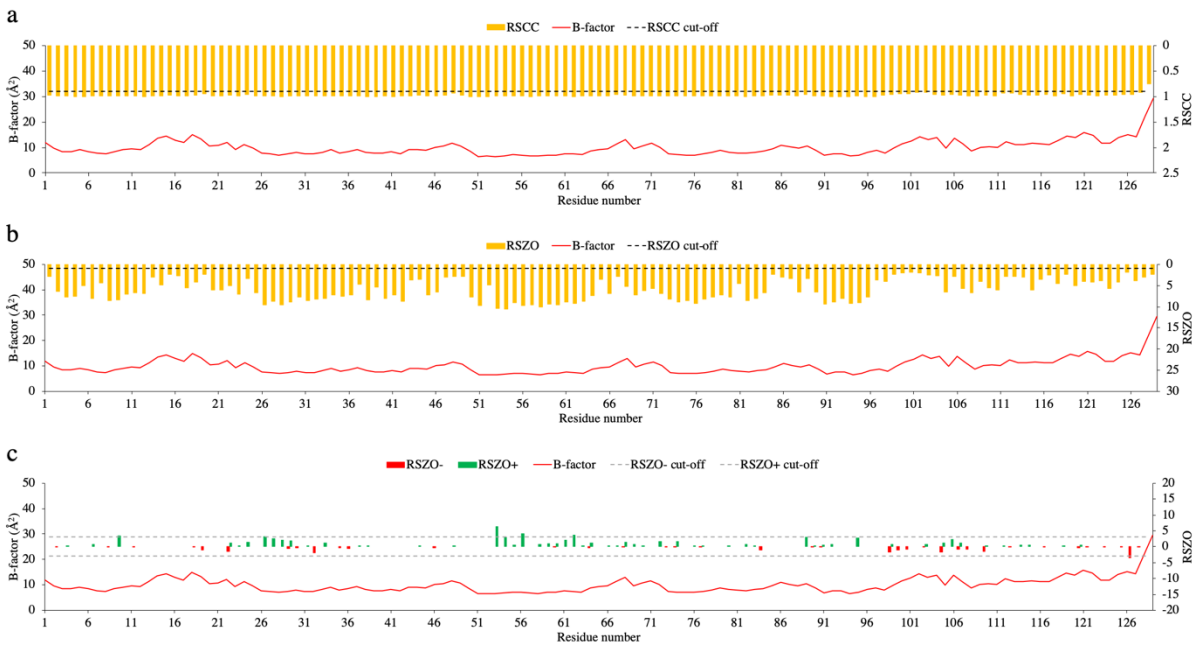


Figure S7

Validation of the quality of the D-HEWL$_{PP}$ model against the X-ray data, according to the RSCC (a), RSZD (b), RSZD- and RSZD+ (c), metrics from EDSTATS. The cut-offs applied were: 90% for RSCC, 1 σ for RSZD, -3 σ for RSZD- and +3 σ for RSZD+. In the D-HEWL$_{PP}$ model, the residues excluded from the detailed. H-bond analysis were Ala9, Gly26, Tyr53, Leu56, Trp62, Thr89, Cys127 and Leu129.

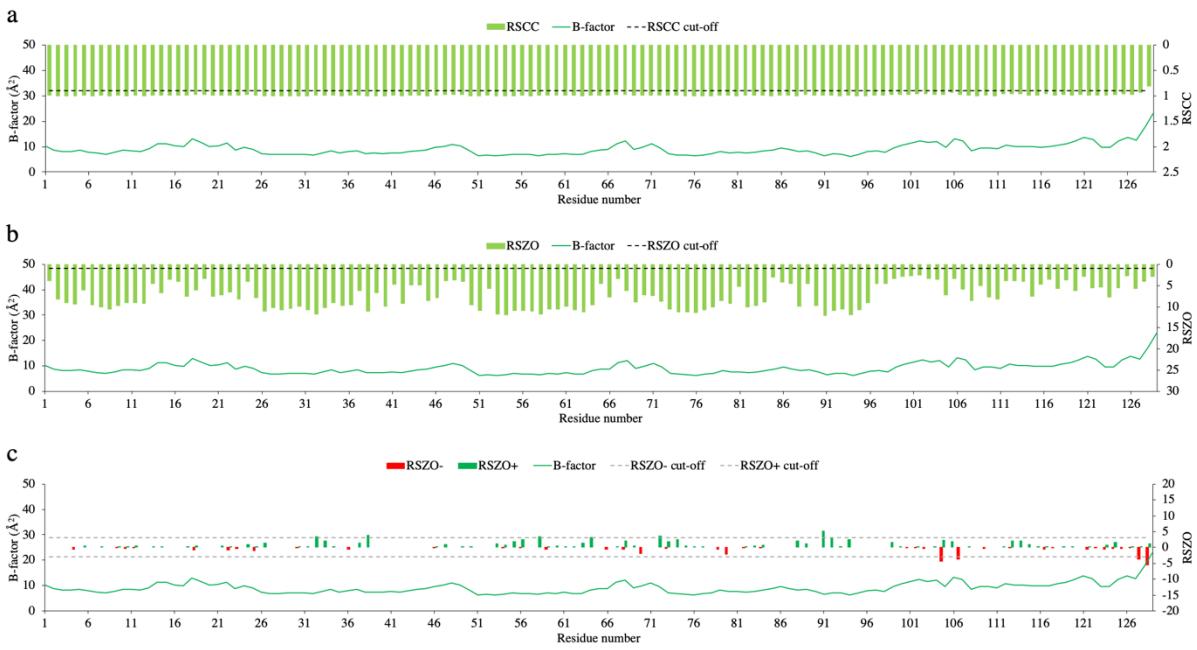


Figure S8

Validation of the quality of the H-HEWL model against the X-ray data, according to the RSCC (a), RSZD (b), RSZD- and RSZD+ (c), metrics from EDSTATS. The cut-offs applied were: 90% for RSCC, 1 σ for RSZD, -3 σ for RSZD- and +3 σ for RSZD+. In the H-HEWL model, the residues excluded from the detailed H-bond analysis were Ala32, Phe38, Ile58, Cys64, Ser72, Ser91, Met105, Ala107, Arg128 and Leu129.

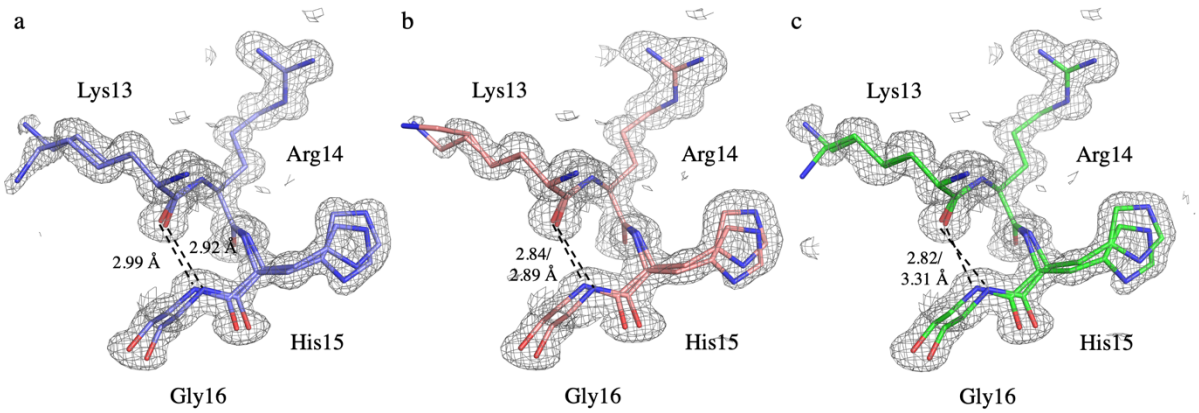


Figure S9

Representation of the Lys13-Gly16 backbone disorder and variation in the Gly16(N)-Lys13(O) H-bond between D-HEWL$_{EC}$ (a), D-HEWL$_{PP}$ (b) and H-HEWL (c). The $2F_o$-F_c electron density map represented is contoured at 1 σ.

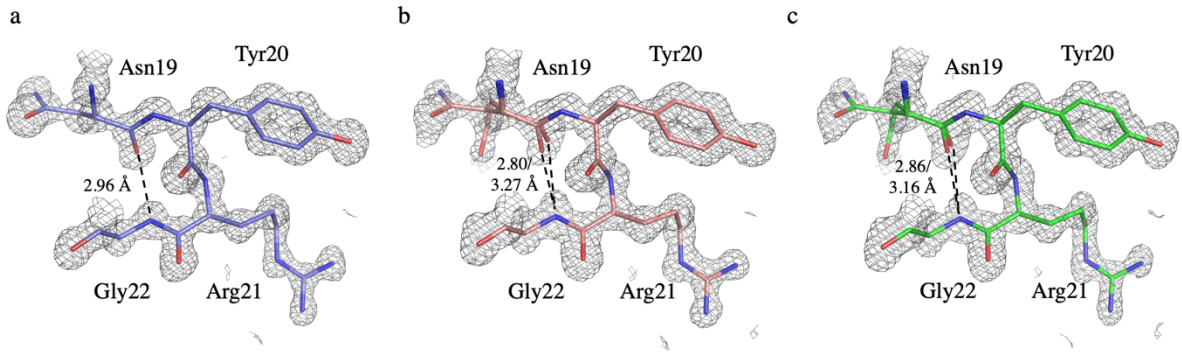


Figure S10

Representation of the differences in Gly229(N)-Asn19(O) H-bond between D-HEWL$_{EC}$ (a), D-HEWL$_{PP}$ (b) and H-HEWL (c). The $2F_o$-F_c electron density map represented is contoured at 1 σ.

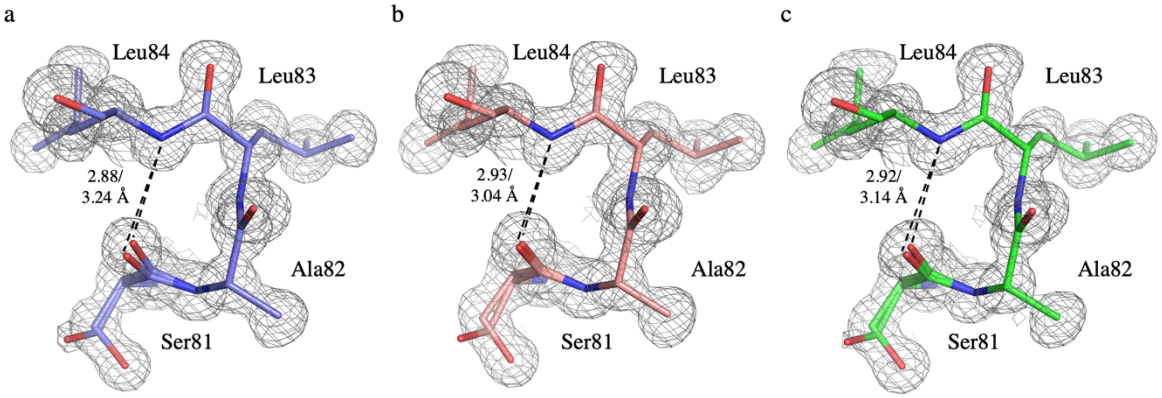


Figure S11

Representation of the differences in the Ser81(N)-Leu84(O) H-bond between D-HEWL$_{EC}$ (a), D-HEWL$_{PP}$ (b) and H-HEWL (c). The $2F_o$-F_c electron density map represented is contoured at 1 σ.

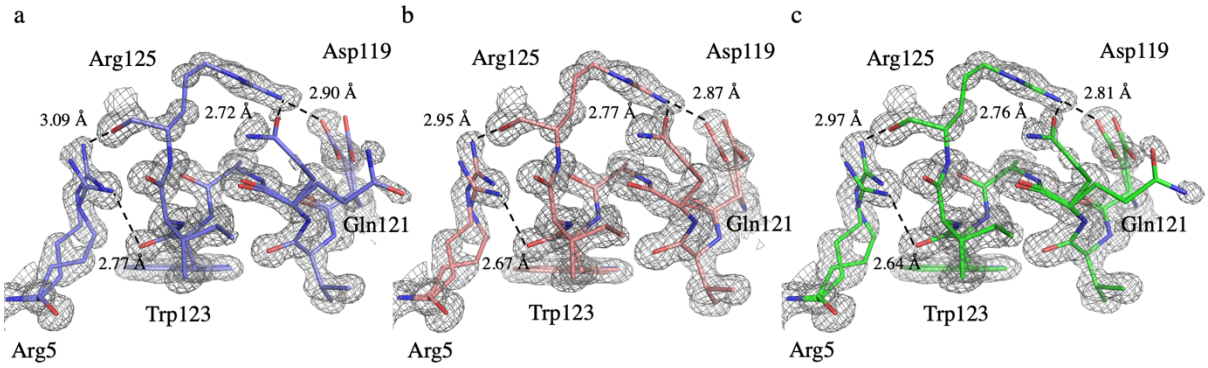


Figure S12

Representation of the differences in the H-bond pattern, involving Arg5(NH1 and NH2), Trp123(O), Arg125(O and NH2), Asp 119(OD2) and Gln121(OE1), between D-HEWL$_{EC}$ (a), D-HEWL$_{PP}$ (b) and H-HEWL (c). The $2F_o$-F_c electron density map represented is contoured at 1 σ.

Chapter 4

Manuscript 2

Synopsis

As previously mentioned, protein perdeuteration provides considerable benefits to the quality of neutron diffraction data, thus, becoming a routine practice in NMX studies. Nevertheless, changes in protein biophysical and structural properties have been reported in several cases. Hence, the comprehensive analysis of deuteration effects on a specific target is essential for the interpretation of subsequent experimental data. Additionally, it is common to observe changes *e.g.*, in protein thermal stability and enzymatic activity, upon recombinant expression and also *in vitro* refolding of a protein, in comparison to its native variant. However, not always these variations are explained through structural data, even less from atomic resolution X-ray diffraction. In this chapter, a manuscript in preparation is presented (Manuscript 2), which unveils relevant details of the impact of deuteration and *in vitro* refolding on the biophysical and structural properties of HEWL expressed in *E. coli*. In contrast with Manuscript 1, this work enables the separation of the macromolecular and solvent isotope effects, and the effects of protein *in vitro* refolding.

The impact of deuteration and *in vitro* refolding on the structural and biophysical properties of hen egg-white lysozyme

Joao Ramos[1,2,3], Valerie Laux[1,2], Michael Haertlein[1,2], Trevor Forsyth[1,2,4], Estelle Mossou[1,2,4], Sine Larsen[5], Annette Langkilde[3*]

[1]*Institut Laue-Langevin, 71 avenue de Martyrs, 38000 Grenoble, France.*

[2]*Partnership for Structural Biology (PSB), 71 avenue de Martyrs, 38000 Grenoble, France.*

[3]*Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.*

[4]*Faculty of Natural Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.*

[5]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.*

*corresponding author, annette.langkilde@sund.ku.dk

Abstract

Protein perdeuteration can provide higher quality data and novel insights into protein structure and dynamics in techniques such as neutron crystallography and nuclear magnetic resonance. *In vitro* refolding is often used to recover recombinant protein overexpressed in the form of inclusion bodies and may be part of the strategy to obtain *e.g.*, perdeuterated protein. In this study, the influence of *in vitro* refolding, protein perdeuteration, and solvent isotope effects on protein thermal stability were quantified through systematic comparisons. Additionally, the structure of recombinant refolded hydrogenated hen egg-white lysozyme (H-HEWL$_{EC}$) was obtained through X-ray diffraction data measured to 0.89 Å resolution and compared to previously published structures of refolded perdeuterated HEWL and unlabelled native HEWL. *In vitro* refolding was found to have a stronger effect on protein structure and thermal stability than protein perdeuteration. An unexpected full peptide-plane flip of Asn103 was observed in H-HEWL$_{EC}$ and this peptide flip was shown to affect protein activity by altering the configuration of the enzyme's binding cleft.

Keywords

Isotope effect, protein perdeuteration, *in vitro* refolding, biophysical characterization, X-ray diffraction, peptide-plane flip.

Introduction

The Hydrogen (H) – Deuterium (D) isotope effects on protein structure, stability and activity have been investigated with the aim to provide a consensus on whether deuteration impacts the interpretation of structural and biophysical data. H atoms comprise approximately 50% of all atoms in a protein and they are essential to biological function, either by playing a role in protein folding, protein interactions, or by participating in the catalysis of enzymatic reactions. However, the substitution of H for its heavier isotope D, only in exchangeable positions, *i.e.*, bound to N, O (~25%), or in all positions, *i.e.*, perdeuteration, is expected to impact protein dynamics. Biomolecular deuteration is routinely employed in neutron macromolecular crystallography (NMX), small-angle neutron scattering (SANS), neutron reflectometry, neutron spectroscopy, and nuclear magnetic resonance (NMR), due to the different physical properties of H and D (Haertlein *et al.*, 2016; Varga *et al.*, 2007). A standout example is the contrast matching technique used in SANS (Dunne *et al.*, 2017; Laux *et al.*, 2008; Haertlein *et al.*, 2016), where different concentrations of a mixture of D_2O and H_2O can mask the scattering contributions of certain components of a macromolecular complex (Vijayakrishnan *et al.*, 2010; Cuypers *et al.*, 2013; Breyton *et al.*, 2013). Another technique that explores extensively the H_2O to D_2O solvent exchange and biomolecular deuteration is NMX. Similarly to the case of SANS, this is due to the large incoherent neutron scattering cross section of H (80.27 barn), in contrast with its isotope D (2.05 barn), which also has a coherent neutron scattering length comparable to ^{12}C, ^{14}N and ^{16}O (Sears, 1992). In NMX, soaking hydrogenated protein crystals in D_2O, allowing H/D exchange in labile positions and removing H from the solvent, can lead to a gain in neutron scattering signal or to a decrease in the crystal volume requirements of 2.5 to 3-fold (Shu *et al.*, 2000; Hazemann *et al.*, 2005). In the case of perdeuteration the impact in the crystal volume requirements for neutron diffraction experiments is more drastic, representing a reduction by 1 order of magnitude (Hazemann *et al.*, 2005). Additionally, protein perdeuteration enables the mitigation of cancellation effects in the Fourier maps, which arise from the negative coherent neutron scattering contribution of H atoms (Chen *et al.*, 2012; Fisher *et al.*, 2014).

Several studies on the macromolecular H/D isotope effect have reported decreases in thermal stability of a perdeuterated protein compared to its hydrogenated variant (Berns, 1963; Hattori *et al.*, 1965; Brockwell *et al.*, 2001; Meilleur *et al.*, 2004; Koruza *et al.*, 2018; Nichols *et al.*, 2020). Concomitantly, observations have been made regarding the solvent H/D isotope effect which suggest that both perdeuterated and hydrogenated proteins are thermally more stable in D_2O than in H_2O (Hattori *et al.*, 1965; Harrington & von Hippel, 1961; Makhatadze *et al.*, 1995; Dong *et al.*, 1997; Kuhlman & Raleigh, 1998; Sasisanker *et al.*, 2004; Efimova *et al.*, 2006). While the macromolecular isotope effect seems to be caused by changes to the strength of hydrophobic interactions between residues side-chains, as described by Hattori *et al.* (1965), the solvent isotope effect appears to be a consequence of variations in protein-protein and protein-water H-bonds, due to the H/D substitution in

exchangeable positions, and of different solvation potentials of D_2O and H_2O (Svergun *et al.*, 1998; Sasisanker *et al.*, 2004; Efimova *et al.*, 2006; Artero *et al.*, 2005; Jasnin *et al.*, 2008).

Conversely, NMX studies have reported only minor structural differences between the hydrogenated and the perdeuterated variants of a protein, which don't impact the protein fold (Gamble *et al.*, 1994; Meilleur *et al.*, 2004; Artero *et al.*, 2005; Liu *et al.*, 2007; Koruza *et al.*, 2019; Ramos *et al.*, 2021). However, changes in protein solubility are apparent and often discussed due to the extensive efforts made to adapt crystallization conditions of hydrogenated protein to the deuterated variant (Budayova-Spano *et al.*, 2000; Koruza *et al.*, 2018; Liu *et al.*, 2007). Often the crystallization conditions used for the perdeuterated variant comprise reductions in precipitant and/or protein concentration, which reflect decreases in protein solubility compared to the hydrogenated variant (Hazemann *et al.*, 2005; Petit-Haertlein *et al.*, 2009). Nevertheless, crystallographic studies have been performed in which the crystallization conditions for both proteins are substantially different (Liu *et al.*, 2007). These variations need to be considered when discussing the structural changes observed between the two variants and when attributing those to macromolecular isotope effects.

Batas & Chaudhuri (1996) have reported a refolding approach for hydrogenated hen egg-white lysozyme (H-HEWL), which yields significant amounts of active protein by performing *in vitro* refolding in-column using high concentrations of denatured sample. This method was later applied to recombinant HEWL, expressed in *E. coli* inclusion bodies (Batas *et al.*, 1999). However, the structural characterization of both refolded H-HEWL variants was missing, raising questions about which impacts the refolding process has in the observed reduction of protein activity. Recently, we showed that a similar approach could be used for the production of perdeuterated HEWL (D-$HEWL_{EC}$) (Ramos *et al.*, 2021).

In our previous study the refolded perdeuterated variant was compared to the unlabeled hydrogenated variant (H-HEWL), however the influences of both macromolecular and solvent isotope substitution (H/D) could not be completely separated from the refolding effects. In this study, we have further characterized and systematically compared the three variants (H-HEWL, H-$HEWL_{EC}$ and, D-$HEWL_{EC}$), including the refolded hydrogenated variant to separate the different effects. The crystal structure of H-$HEWL_{EC}$ was obtained from X-ray diffraction data at 0.89 Å resolution for comparison with the previous structures (PDB entries 7ave and 7avf; Ramos *et al.*, 2021). The observed changes in enzymatic activity could be correlated to structural differences. Additionally, the thermal stabilities of the three variants were investigated by differential scanning fluorimetry (DSF) in hydrogenated as well as deuterated solutions. The *in vitro* refolding, the macromolecular and the solvent isotope effects could then be addressed by systematic comparison. It is relevant to note that, in terms of primary structure, both $HEWL_{EC}$ variants differ from H-HEWL by an additional Gly residue in the N-terminus. This Gly residue has been shown to disrupt hydrogen-bond (H-bond) interactions at the protein's N-terminus, rendering this region more disordered (Ramos *et al.*, 2021). Nevertheless, it is unlikely that this residue would perturb protein folding or function.

Methods

Protein materials

Commercially available H-HEWL (catalogue No. L6876; Sigma–Aldrich) was used without additional processing. D-HEWL$_{EC}$ was recombinantly expressed in *E. coli* BL21(DE3) cells in the form of inclusion bodies, and posteriorly purified and refolded in-column, as described by Ramos *et al.* (2021). Using a similarly approach, H-HEWL$_{EC}$ was expressed in *E. coli* BL21(DE3) cells, purified and refolded. Buffer compositions and procedures employed in inclusion body washing, protein purification by gel filtration, and refolding in-column were identical for both proteins. The main differences between the protein production methods applied to the two HEWL$_{EC}$ variants were the culture media used for protein overexpression and the final protein buffer exchange, performed after refolding. D-HEWL$_{EC}$ was expressed in *E. coli* cultures adapted to fully deuterated minimal medium (Ramos *et al.*, 2021), while H-HEWL$_{EC}$ was produced in *E. coli* cultures grown in H_2O minimal medium. Additionally, D-HEWL$_{EC}$ was buffer exchanged to a D_2O solution of 50 mM sodium acetate pD 4.5 (pD = pH + 0.4) (Glasoe & Long, 1960). Meanwhile, H-HEWL$_{EC}$ was buffer exchanged to a H_2O solution of 50 mM sodium acetate pH 4.5. The proteins were concentrated in their final buffer solutions to 20 mg/mL for further experiments.

Differential Scanning Fluorimetry (DSF)

Protein stability was assessed by DSF using a Prometheus instrument from NanoTemper. Protein unfolding is measured through the intrinsic fluorescent signal of Trp residues (six Trp in HEWL), which is quenched when the protein is folded and is enhanced when these residues are exposed to the solvent in the protein unfolded state. H-HEWL, D-HEWL$_{EC}$, and H-HEWL$_{EC}$ were tested each in a total of four different buffer solutions: 1) 50 mM sodium acetate pH 4.5 in H_2O; 2) 50 mM sodium acetate pD 4.5 in D_2O; 3) 0.1 M sodium phosphate pH 7.5, 0.1 M NaCl, 2 mM NaN_3 in H_2O; 4) 0.1 M sodium phosphate pD 7.5, 0.1 M NaCl, 2 mM NaN_3 in D_2O. The samples were diluted from stock solutions at 20 mg/mL in each buffer solution to a final protein concentration of approximately 0.2 mg/mL. The fluorescence was measured at 90% instrument excitation power, in temperature ramps from 20° to 95° C, with 0.1° C/min increments. The results presented were obtained from three different experiments with duplicate or triplicate measurements for each condition.

Activity assays

The enzymatic activity was measured using the method reported by Shugar (1952), which estimates the activity rates by following the decrease in absorbance at 450 nm when HEWL is added to a cell suspension of *Micrococcus lysodeikticus*. The conditions used are described elsewhere (Ramos *et al.*, 2021). The results presented in Fig. S2 were obtained from three HEWL$_{EC}$ preparations, containing triplicate measurements for each condition, including a control without protein. The triplicates from each experiment were averaged and plotted against time to obtain the activity curves.

The initial velocities were retrieved from the linear phase ($R^2 > 0.95$) corresponding to the first 5 minutes of reaction. The final activity rates and corresponding standard deviations were obtained from averaging the results from the three experiments.

Protein crystallization

H-HEWL$_{EC}$ triclinic crystals were obtained by microseeding triclinic H-HEWL, in microbatch under oil at 18° C. The crystallization drop contained 2.5 µL of 20 mg/mL refolded H-HEWL$_{EC}$, 2.5 µL of 0.3 M $NaNO_3$ and 50 mM sodium acetate pH 4.5 in H_2O, and 0.5 µL of H-HEWL seeds. Crystals of approximately 0.1 mm^3 appeared within 1 week.

X-ray diffraction data collection and processing

A triclinic crystal of H-HEWL$_{EC}$ was flash-cooled in liquid N_2 after soaking in a solution containing 30% (*v/v*) glycerol, 0.4 M $NaNO_3$ and 50 mM sodium acetate pH 4.5 in H_2O. X-ray diffraction data at 100 K was measured from the H-HEWL$_{EC}$ crystal in the BioMAX beamline, at MAX IV, Sweden. Data to 0.89 resolution Å were recorded in two different κ orientations, with 180° scans, to maximize reciprocal space coverage and data completeness. Data reduction was performed using XDS (Kabsch, 2010) and the two 180° scans were scaled using XSCALE (Kabsch, 2010). The data was converted to SHELX format using XDSCONV (Kabsch, 2010) and 5% of the reflections were randomly flagged for R_{free} calculations.

Model refinements

PDB entry 7ave (Ramos *et al.*, 2021) was used as initial model for the refinement of H-HEWL$_{EC}$ in SHELXL (Sheldrick, 2015). Alternate conformations, water molecules and ions were removed from the initial model and the anisotropic ADPs were converted to isotropic. After the refinement of residue disorder, hydrogen atoms were added in idealized positions and fixed as riding atoms. Anisotropic ADPs were refined for all non-hydrogen atoms, including waters and ions. Additionally, occupancies were refined for water molecules and ions with B-factors larger than 30 $Å^2$. SIMU (0.1) and XPND (0.001) restraints were removed in the last refinement rounds.

Analysis of HEWL structures

The H-HEWL and D-HEWL$_{EC}$ models (PDB entries 7avf and 7ave; Ramos *et al.*, 2021), were used in the comparisons with the H-HEWL$_{EC}$ structure. The structural alignments between the HEWL structures were performed using GESAMT, from the CCP4 suite (Winn *et al.*, 2011). The alignments of the Lys97-Gly104 and binding cleft regions were done in PyMOL (version 2.0, Schrödinger), using the ALIGN function with 0 refinement cycles. All representations of protein structures were made using PyMOL.

Results

In vitro refolding has a stronger impact in lysozyme's thermal stability than deuteration

The results from the DSF experiments show clear trends regarding the effects in protein thermal stability of *in vitro* refolding, protein perdeuteration, H/D solvent substitution and buffer solution pH (Fig. 1).

It is evident that the refolding process impacted the thermal stability of H-HEWL$_{EC}$. In H_2O, H-HEWL$_{EC}$ displays a decrease in melting temperature (T_m) of 4.7° C and 3.5° C, at pH 4.5 and 7.5 respectively, in comparison with H-HEWL. Meanwhile, in D_2O, H-HEWL$_{EC}$ is less thermally stable than H-HEWL by 4.5° C and 3.7° C, at pD 4.5 and 7.5 respectively. These reductions in T_m seem to be constant regardless of the solvent isotope substitution, with minor variations likely due to the differences between pH and pD (pD = pH + 0.4) (Glasoe & Long, 1960).

HEWL$_{EC}$ perdeuteration appears to reduce protein thermal stability by 1.6° C and 1.3° C, at pH 4.5 and pH 7.5, respectively. Interestingly, these variations are smaller than those observed between H-HEWL and H-HEWL$_{EC}$, highlighting that the refolding process has a stronger impact in protein thermal stability than perdeuteration. Additionally, the differences in T_m appear similar in H_2O and D_2O solutions, thus reflecting effects of protein perdeuteration and not solvent isotope substitution.

Comparing the individual variants in H_2O and D_2O respectively, the solvent isotope effect has similar magnitudes across the three HEWL variants, varying only according to the buffer solution pH/pD. In the case of H-HEWL, replacing H_2O for D_2O increased its thermal stability by 1.9° C and 2.8° C, at pH/pD 4.5 and 7.5, respectively. Meanwhile, both H-HEWL$_{EC}$ and D-HEWL$_{EC}$ were also found to be more thermally stable in D_2O than in H_2O, with identical variations in T_m of 2.1° C at pH/pD 4.5, and of 2.6° C at pH/pD 7.5.

It may seem evident that pH as an impact in protein thermal stability. However, since the buffer solutions at pH/pD 4.5 and 7.5 vary significantly in composition, it is not appropriate to perform a direct comparison of the pH/pD effects. Nevertheless, it is clear that substantial reductions in thermal stability occur in the three HEWL variants when shifting from pH/pD 4.5 to pH/pD 7.5 buffer solution.

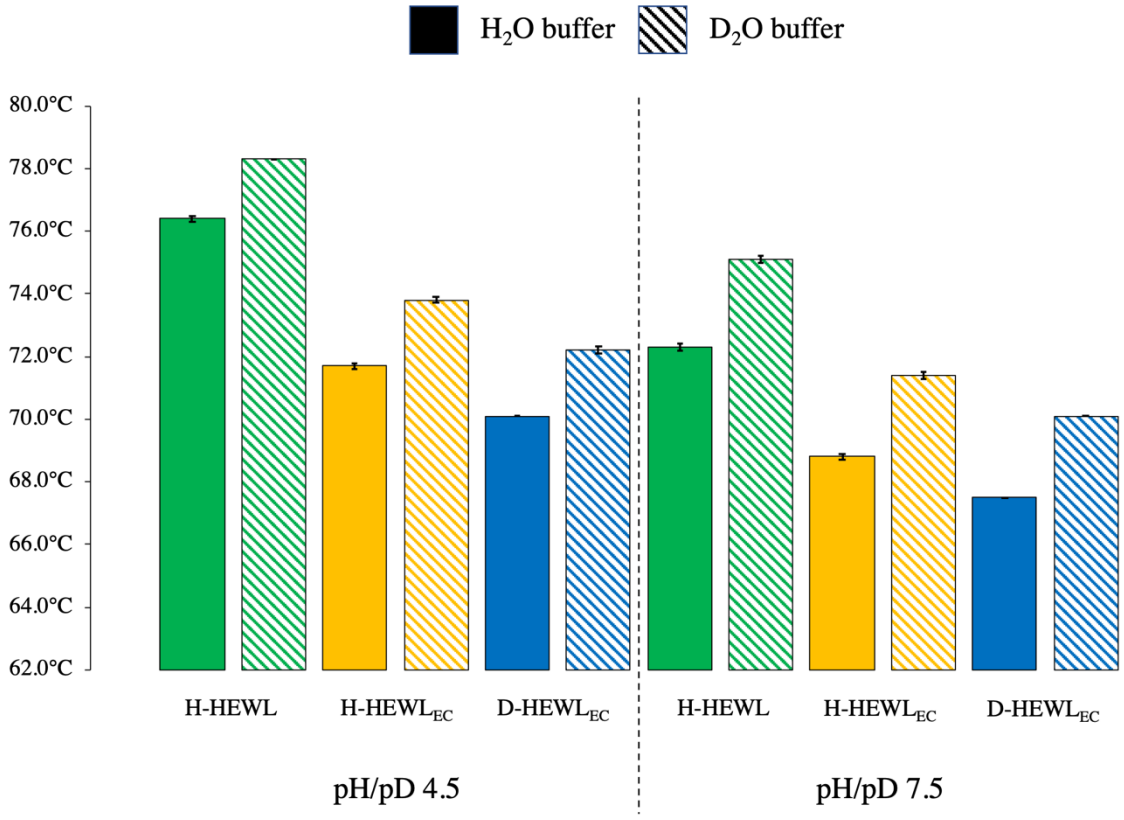


Figure 1

Melting temperatures measured in the DSF experiments for H-HEWL, H-HEWL$_{EC}$ and D-HEWL$_{EC}$, in buffer solutions at pH 4.5 and 7.5 in H_2O, and at pD 4.5 and 7.5 in D_2O. The error bars correspond to the standard error deviations of the measurements performed in three DSF experiments.

In vitro refolding leads to loss of enzymatic activity

Previous studies have reported reductions in enzymatic activity of refolded HEWL, compared to the native variant (Batas & Chaudhuri, 1996; Batas *et al.*, 1999). However, these observations were not explained through structural data. In this study, activity assays were performed for all three HEWL variants in similar conditions to those described elsewhere (Ramos *et al.*, 2021). The results indicate that H-HEWL$_{EC}$ and D-HEWL$_{EC}$ retain 66% and 69% of H-HEWL activity, respectively (Supplementary Fig. S1). Although our previous work had reported an activity of 51% for D-HEWL$_{EC}$ (Ramos *et al.*, 2021), the analysis performed considered the first 8 minutes of reaction for the extrapolation of initial velocities, with $R^2 > 0.91$, while in the present study the first 5 minutes were used, resulting in $R^2 > 0.95$.

Protein fold is retained despite in vitro refolding and perdeuteration

The H-HEWL$_{EC}$ triclinic crystal was obtained in close to identical conditions to those reported for H-HEWL and D-HEWL$_{EC}$ (PDB entries 7avf and 7ave; Ramos *et al.*, 2021) and comparable X-ray

diffraction data extending to 0.89 Å resolution has been collected at 100 K. The data merging and model refinement statistics are presented in Table 1. The HEWL crystal structure reported by Wang *et al.* (2007), obtained from X-ray diffraction data to 0.65 Å resolution, stands out as a reference among the plethora of HEWL structures deposited in the PDB (Berman *et al.*, 2000). However, a detailed comparison between the structures of several HEWL variants is better performed when the experimental conditions during protein crystallization, X-ray data collection, processing and model refinement are highly similar.

Structural alignments were performed for the following pairs of crystal structures: H-HEWL – H-HEWL$_{EC}$ (highlighting refolding effects); D-HEWL$_{EC}$ – H-HEWL$_{EC}$ (highlighting deuteration effects); and H-HEWL – D-HEWL$_{EC}$ (combined effects as previously observed for comparison). Overall, considering H-HEWL as a reference, the protein fold is retained in the recombinant variants, despite the *in vitro* refolding performed to obtain H-HEWL$_{EC}$ and both the refolding and perdeuteration of D-HEWL$_{EC}$ (Fig. 2). This evidence is supported by the fact that both HEWL$_{EC}$ were successfully crystallized in closely comparable conditions, in the same crystal space group and with similar unit cell constants. Interestingly, the H-HEWL and D-HEWL$_{EC}$ are highly similar and the main deviations in protein structure are observed in relation to H-HEWL$_{EC}$. While D-HEWL$_{EC}$ differs only by 0.17 Å from the H-HEWL structure, H-HEWL$_{EC}$ shows a r.m.s.d. of 0.57 Å. Additionally, H-HEWL$_{EC}$ was found to have a r.m.s.d of 0.56 Å when compared to D-HEWL$_{EC}$. The main differences observed between H-HEWL$_{EC}$ and the other two HEWL variants reside in the Thr47-Gly49 and Lys97-Gly104 regions, which are at the protein surface and involve crystal contacts.

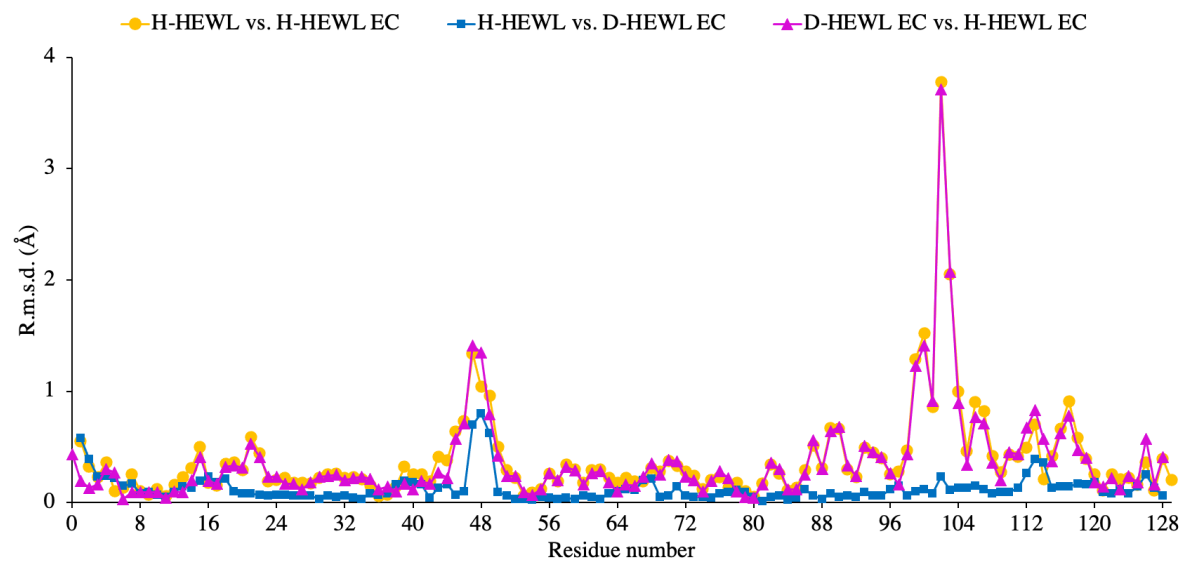


Figure 2

The three HEWL variants share a similar overall fold, however, the structure of H-HEWL$_{EC}$ displays significant variations in the Thr47-Gly49 and Lys97-Gly104 regions. The structures were structurally aligned using GESAMT, from the CCP4 suite (Winn *et al.*, 2011), and the r.m.s.d. between each pair of structures are plotted for each protein residue.

Table 1

X-ray diffraction data merging and model refinement statistics for H-HEWL$_{EC}$. The values in parenthesis correspond to the outer resolution shell.

	H-HEWL$_{EC}$
Temperature (K)	100
Source	BioMAX, MAX IV, SE
Detector	EIGER hybrid-pixel 16M
Wavelength (Å)	0.700
Resolution range (Å)	22.46 - 0.89 (0.92 - 0.89)
Space group	P1
Unit cell **a, b, c (Å)** **α, β, γ (°)**	 26.12, 30.70, 33.57 88.968, 72.768, 69.505
Total reflections	234437 (23523)
Unique reflections	68446 (6717)
Multiplicity	3.4 (3.5)
Completeness (%)	96.2 (94.3)
Mean I/σ(I)	7.4 (2.2)
R_{merge}	0.083 (0.647)
R_{meas}	0.098 (0.762)
R_{pim}	0.052 (0.398)
$CC_{1/2}$	99.6 (79.9)
Refl. used in refinement with *Fo*/σ(*Fo*) > 4 / all refl.	50427 / 65023
Refl. used for R_{free} with *Fo*/σ(*Fo*) > 4 / all refl.	2635 / 3423
R_{work} for refl. with *Fo*/σ(*Fo*) > 4 / all refl.	10.89 / 13.55
R_{free} for refl. with *Fo*/σ(*Fo*) > 4 / all refl.	12.24 / 14.99
Number of non-hydrogen atoms (overall)	1517
macromolecules	1304
ligands	40
solvent	164
Protein residues	130
R.m.s.d., bond lengths (Å)	0.0235
R.m.s.d., angles (°)	3.40
Ramachandran favored (%)	96.88
Ramachandran allowed (%)	3.12
Ramachandran outliers (%)	0
Rotamer outliers (%)	2.17
Clashscore	11.48
Average B-factor (overall) (Å^2)	12.25
macromolecules	11.12
ligands	14.08
solvent	21.56

The Asn103 peptide-plane flip is key in HEWL refolding

In our previous study (Ramos *et al.*, 2021), the Asn103 peptide-plane flip was shown to play a central role in the *in vitro* refolding of D-HEWL$_{EC}$, as well as in the production of D-HEWL in *Pichia pastoris*. While a partial Asn103 peptide-plane flip had been described in H-HEWL (PDB entries 2vb1, Wang *et al.*, 2007; and 7avf, Ramos *et al.*, 2021) with a refined occupancy of 33%, in D-HEWL$_{EC}$ (PDB entry 7ave), it was found that this alternate conformation was present in 46% of the protein molecules in the crystal. This peptide-plane flip is thought to occur in the initial stages of protein folding and it was shown to be responsible for the backbone disorder in the Lys97-Gly104 region by Wang *et al.* (2007) and Ramos *et al.* (2021).

Interestingly, in the H-HEWL$_{EC}$ structure, the Lys97-Gly104 region deviates further away from the conformations present both in H-HEWL and D-HEWL$_{EC}$ (Fig. 3). The Lys97-Gly104 region in H-HEWL$_{EC}$ differs 2.37 Å and 2.77 Å from those of H-HEWL and D-HEWL$_{EC}$, respectively. Meanwhile, between D-HEWL$_{EC}$ and H-HEWL there is a r.m.s.d. of only 0.62 Å. This greater disparity in conformation shown by H-HEWL$_{EC}$ is linked to the fact that, according to the 2*Fo-Fc* electron density map, both A and B conformations include a peptide-plane flip of Asn103 (Supplementary Fig. S2). As a result, a significant shift in this region's backbone structure occurs, predominantly from Ser100 to Asn103. Variations in side-chains conformations can also be observed, namely for Asp101 and Asn103. In the case of Asn103 side-chain, significant disorder is evident by the lack of density in the 2*Fo-Fc* electron density map contoured at 1 σ. Meanwhile, for Asp101, the electron density suggests a different position of the residue side-chain compared to the models of the other HEWL variants (Ramos *et al.*, 2021), which leads to the disruption of a crystal contact with Glu7, through an H-bond.

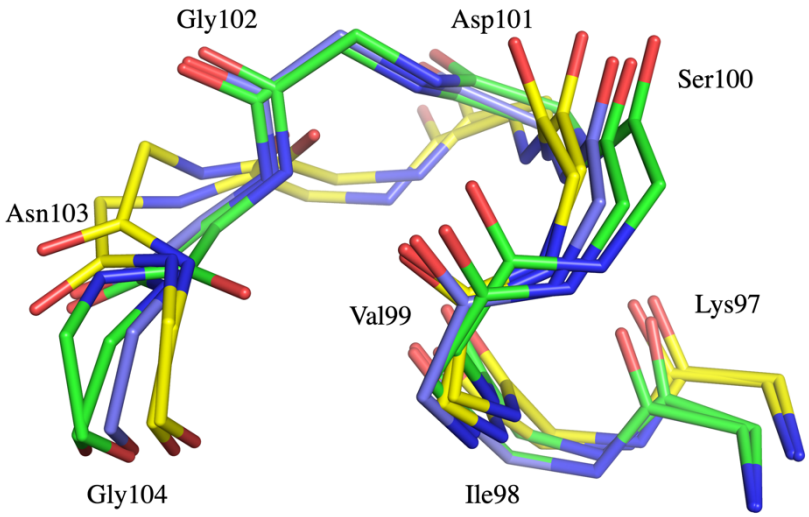


Figure 3

The Asn103 peptide-plane flip in both conformation of the Lys97-Gly104 regions in H-HEWL$_{EC}$ results in considerable shift of the backbone structure. Backbone conformations of the Lys97-Gly104 regions of D-HEWL$_{EC}$ (blue; PDB entry 7ave; Ramos *et al.*, 2021) and H-HEWL$_{EC}$ (yellow) superimposed on the H-HEWL model (green; PDB entry 7avf; Ramos *et al.*, 2021).

In D-HEWL$_{EC}$, an alternate conformation of Glu7, B with refined occupancy of 51%, was found to be stabilized by H-bonds to Lys1 Nζ, Gly4 N and Val2 O, with the latter being mediated by a water molecule. Positive density in the *Fo-Fc* electron density map for the carboxylate oxygens in this conformation suggests the presence of two water molecules when Glu7 occupies the conformation A. The refined occupancies of both Glu7 and Lys97-Gly104 conformations B are similar (51% and 46%, respectively), and Glu7 conformation A and Asp101 are involved in crystal contact interactions indicating a correlation between the Asn103 peptide-plane flip and the displacement of the Glu7 side-chain. A contributing factor for this association is the observation that Asp101/B side-chain, linked to the Asn103 peptide flip conformation, is moved further away from Glu7/A in D-HEWL$_{EC}$, hindering the possibility of interaction via H-bond. Interestingly, in the case of H-HEWL$_{EC}$, where the Asn103 peptide-plane flip appears to be present in both conformations of the Lys97-Gly104 region and where the Glu7 Oε1 – Asp101 Oδ2 interaction is completely disrupted, the Glu7 side-chain is found in a single conformation involved in intra-molecular H-bonds to Lys1 Nζ, Gly4 N and Val2 O (Fig. 4).

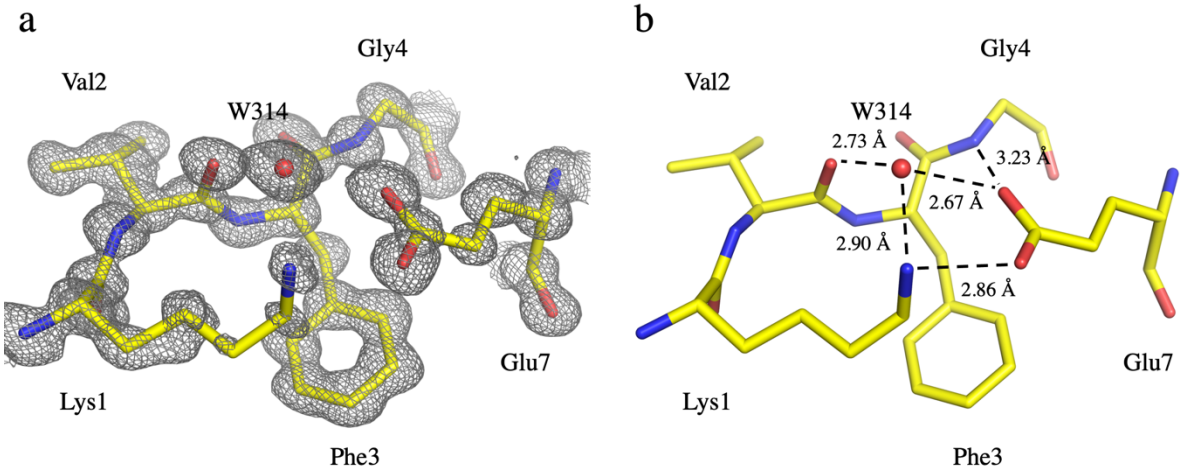


Figure 4

Single Glu7 conformation present in H-HEWL$_{EC}$ and respective H-bond interactions with Lys1 Nζ, Gly4 N and Val2 O. W314 mediates the interaction between Glu7 and Val2 through H-bonds. The 2*Fo-Fc* electron density map presented was contoured at 1 σ.

The Glu7 conformation in H-HEWL$_{EC}$ appears to be further linked to the displacement of the Thr47-Gly49 region via crystal contacts. Although the disorder in the Thr47-Gly49 region is present in the three HEWL variants, the interaction patterns vary in the case of H-HEWL$_{EC}$ compared to both H-HEWL and D-HEWL$_{EC}$. In H-HEWL and D-HEWL$_{EC}$, Thr47/A is involved in water (W34) mediated H-bonds with Leu75 O and Glu7(/A) Oε2 via crystal contacts, while Thr47/B appears to be able of establishing also direct H-bonds with Leu75 O and Asn74 O. On the other hand, in H-HEWL$_{EC}$, the fact that Glu7 is in a single conformation, identical to Glu7/B in D-HEWL$_{EC}$, eliminates the potential water mediated H-bonds with Glu7(/A) Oε2 and Leu75 O via crystal contacts found in both H-HEWL and D-HEWL$_{EC}$. Moreover, this water molecule seems to be absent in the H-HEWL$_{EC}$ model. Therefore, while Thr47/A seems to be stabilized by weak H-bond interactions with Asp48, mediated by W367, Thr47/B Oγ1 is able of participating in H-bonds with W365, W314 and Glu7 Oε2, via crystal contacts with distances of 2.71 Å, 3.17 Å and 3.26 Å, respectively (Supplementary Fig. S3).

In vitro refolding perturbs the configuration of HEWL binding cleft

We reported only minor differences in HEWL binding cleft when comparing D-HEWL$_{EC}$ with H-HEWL, with the exception of increased disorder in the Lys97-Gly104 region (Ramos *et al.*, 2021). This small variation was promoted by a greater occurrence of the partial Asn103 peptide-plane flip in D-HEWL$_{EC}$ (46% as opposed to 33%) and the corresponding displacement of Lys97-Gly104 conformation B.

In the case of H-HEWL$_{EC}$, a significant shift of several binding cleft residues can be observed (Fig. 5). Both Asn46 and Asp48 appear to be displaced in comparison with H-HEWL and D-HEWL$_{EC}$, which can be explained by the crystal contacts previously described with Glu7, affecting the Thr47-Gly49 region. Additionally, the Lys97-Gly104 region, associated with the Asn103 peptide-plane flip, is considerably dislocated, most noticeably the residues Asp101, Gly102 and Asn103. This variation in protein structure is directly linked with the fact that the Asn103 peptide-plane flip occurs in both Lys97-Gly104 conformations in H-HEWL$_{EC}$, as opposed to what is observed in the other HEWL variants. Additionally, this shift in the protein backbone seems to be propagated until Val109, due to the Asn106 N – Asn103 O and Ala107 N – Gly104 O H-bond interactions. Concomitantly, a displacement of the water molecules arrangement can be observed in HEWL active site and binding cleft. It is interesting to note also the absence of the nitrate ion in the H-HEWL$_{EC}$ binding cleft, which is present in the other HEWL variants, suggesting a link to the displacement of the protein backbone and water structure.

An alternate conformation of residue Asn44 is present in H-HEWL$_{EC}$ (Supplementary Fig. S4), which hasn't been observed in H-HEWL and D-HEWL$_{EC}$ (Ramos *et al.*, 2021), nor in the 0.65 Å resolution H-HEWL structure by Wang *et al.* (2007). This Asn44 conformation has a refined occupancy of 55% and establishes a 2.68 Å H-bond to Gln57 Oε1. Additionally, the minor conformation, which can be found in the other HEWL variants, appears to be significantly disordered with clear electron density, in the 2*Fo-Fc* electron density map at 1 σ, only for the side-chain Cγ. Concomitantly, a fully occupied water molecule (W387), absent in the other HEWL structures, is present, forming H-bonds to Asn46 Oδ1 and Asp52 Oδ2 (Supplementary Fig. S5).

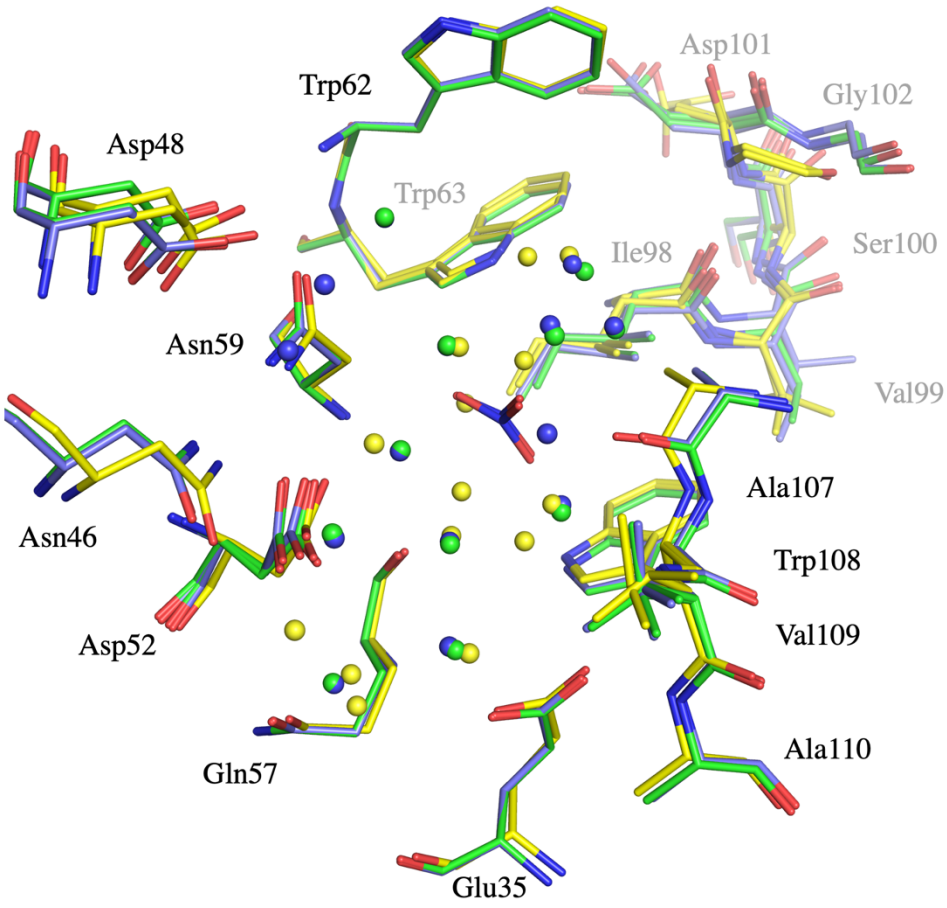


Figure 5

Configuration of H-HEWL$_{EC}$ (yellow) and D-HEWL$_{EC}$ (blue) binding clefts superposed to that of H-HEWL (green). Water molecules are represented as spheres colored according to the color of the respective model and they are not labelled for clarity. The nitrate ions placed at the center of the figure belong to the D-HEWL$_{EC}$ and H-HEWL models.

Discussion

In this study, the contributions of *in vitro* refolding, protein perdeuteration and solvent isotope substitution towards the thermal stability of HEWL were quantified by DSF. Surprisingly, the observations suggest that *in vitro* refolding has a stronger impact in HEWL thermal stability than protein perdeuteration, with respective decreases in T_m greater than 3.5° C compared to variations smaller than 1.6° C. These changes in protein thermal stability are consistent throughout different buffer compositions and also in H_2O and D_2O solvent, where the minor deviations found can be explained by the aforementioned difference between pH and pD (Glasoe & Long, 1960). Moreover, the reduction in protein thermal stability observed due to protein perdeuteration, is in agreement with several previous studies (Berns, 1963; Hattori *et al.*, 1965; Brockwell *et al.*, 2001; Meilleur *et al.*, 2004; Koruza *et al.*, 2018; Nichols *et al.*, 2020).

The effects of solvent isotope substitution were measurable and found to be of similar magnitude in all three HEWL variants. Protein thermal stability was increased when replacing the H_2O solvent for D_2O, with differences in T_m between 1.9° C and 2.8° C. The small variations observed between H-HEWL and both $HEWL_{EC}$ variants might be related to distinct solvation structures which could be a consequence of the protein refolding process, or to the addition of a Gly residue at the N-terminus of $HEWL_{EC}$. Nevertheless, the measurement of higher protein thermal stability in D_2O when compared to H_2O is coherent with previous reports (Hattori *et al.*, 1965; Harrington & von Hippel, 1961; Makhatadze *et al.*, 1995; Dong *et al.*, 1997; Kuhlman & Raleigh, 1998; Sasisanker *et al.*, 2004; Efimova *et al.*, 2006). It is interesting to note that the solvent isotope effect appears to be significantly stronger than the macromolecular isotope effect regarding protein thermal stability. This observation suggests that weakening the hydrophobic interactions of the protein residues through perdeuteration perturbs protein structure and stability to a lower degree in comparison to the changes in protein solvation caused by solvent isotope substitution for D_2O.

Throughout the protein thermal stability measurements, it is evident that the HEWL variants are less thermally stable in buffer solutions at pH/pD 7.5 than at pH/pD 4.5. A direct comparison between the results obtained at different pH/pD was not possible, since the buffer solutions have very different compositions. Nevertheless, since HEWL pI is 11, it was expected that the different variants would be less thermally stable at pH/pD 7.5 compared to a more acidic buffer solution. This observation is also consistent with the optimal pH for the enzyme's activity being around 5.

The H-$HEWL_{EC}$ crystal structure, obtained from X-ray diffraction data to 0.89 Å resolution, provided valuable insight regarding the effects of both *in vitro* refolding and perdeuteration on protein structure, which can be linked to differences in thermal stability and enzymatic activity. Although the protein fold was shown to be conserved in H-$HEWL_{EC}$ compared to H-HEWL, significant variations were observed which, similarly to the case of D-$HEWL_{EC}$, stem from the peptide-plane flip of Asn103. In D-$HEWL_{EC}$, the Asn103 partial peptide-plane flip seemed to be favored as a consequence of protein

perdeuteration and *in vitro* refolding (Ramos *et al.*, 2021). However, it was not possible to deconvolute both perdeuteration and refolding effects since the structure of H-HEWL$_{EC}$ was unknown. Furthermore, similar observations were made for the perdeuterated HEWL variant expressed in *Pichia pastoris*, which was produced in fully deuterated medium. Thus, the slower folding dynamics caused by a combination of more viscous solvents and protein perdeuteration seemed likely to contribute to the increased probability of the Asn103 peptide-plane flip to occur. Unexpectedly, here, the structure of H-HEWL$_{EC}$ revealed that this peptide-plane flip is favored by the conditions of *in vitro* refolding, since both conformations of the Lys97-Gly104 region are associated with a peptide-plane flip of Asn103. As a consequence, a significant shift in the Lys97-Gly104 conformation is observed which has ramifications in terms of changes to the configuration of HEWL binding cleft and to the usual crystal contacts of triclinic HEWL. Interestingly, the elution profiles of H-HEWL$_{EC}$ and D-HEWL$_{EC}$ during refolding show variations (Supplementary Fig. S6), which could be explained by the macromolecular isotope effect. In the case of H-HEWL$_{EC}$ the quantities of aggregated and unfolded or misfolded protein appear to be greater, and as a consequence the refolding yield is reduced, compared to D-HEWL$_{EC}$. For H-HEWL$_{EC}$, a combination of the reduction in MW (of approximately 1 kDa compared to D-HEWL$_{EC}$), which can hinder the optimal separation from denaturing salts during gel filtration, and the faster protein dynamics, which can promote a larger conformational energy landscape, could play a role in the observed changes to protein structure.

In terms of crystals contacts, the full Asn103 peptide-plane flip disrupts the H-bond interaction between Asp101 and Glu7 through the distortion of the Lys97-Gly104 region. Additionally, the peptide-plane flip also favors the interaction of Glu7 with Thr47, shifting the region comprising Thr47-Gly49. The aforementioned regions are thus the main differences between structures of H-HEWL$_{EC}$ and both D-HEWL$_{EC}$ and H-HEWL. The perturbation of the crystal contact interactions explains why H-HEWL$_{EC}$ was only successfully crystallized in the triclinic form when using microseeding of triclinic H-HEWL.

While crystal contacts seem to be responsible for significant differences in the crystal structure of H-HEWL$_{EC}$, in comparison with other HEWL variants, they are not a reflection of protein structure in solution and of its biophysical properties. However, the fact that the Asn103 peptide-plane flip occurs in both conformations of the Lys97-Gly104 regions has implications in the configuration of HEWL binding cleft. More specifically, the increased backbone disorder, compared to other HEWL variants, is propagated to additional residues of the binding cleft through H-bond interactions. A drastic shift of the protein backbone can be observed between residues Ser100 and Val109. This region is responsible for the interaction with substrate molecules A, B, and C (Blake *et al.*, 1967; Phillips, 1967). Several X-ray crystal structures of HEWL bound to inhibitor carbohydrates proved the importance of that region for enzymatic activity (Strynadka & James, 1991; Cheetham *et al.*, 1992; Maenaka *et al.*, 1995; Tanaka *et al.*, 2021). Although, the further clarification would require a structure of the H-HEWL$_{EC}$-inhibitor

complex, one could interpret the observed changes in the binding cleft of the apo-form as indications that its ability to interact the substrate molecule is hindered.

The observations made concerning the structure of H-HEWL$_{EC}$ serve as an explanation for the reduced enzymatic activity reported for both non-recombinant and recombinant H-HEWL in previous studies (Batas & Chaudhuri, 1996; Batas *et al.*, 1999). In the present work, H-HEWL$_{EC}$ and D-HEWL$_{EC}$ were refolded in identical conditions and subsequent enzymatic assays were performed for all HEWL variants, including native H-HEWL. Despite of showing similar levels of function, the reasons for the reduction in activity compared to H-HEWL, for H-HEWL$_{EC}$ and D-HEWL$_{EC}$, seem to be differ, according to their crystal structures. While H-HEWL$_{EC}$ function appears to be hindered by the differences in configuration of the enzyme's binding cleft, D-HEWL$_{EC}$ activity might be primarily affected by the macromolecular isotope effect, which alters protein dynamics.

Ultimately, this study proved that *in vitro* refolding may lead to significant changes in protein structure which hinder its thermal stability and activity. Nevertheless, the expression of recombinant protein in inclusion bodies and posterior refolding remains a potentially attractive approach to obtain active and thermally stable proteins, as in the case of D-HEWL$_{EC}$ (Ramos *et al.*, 2021). Unexpectedly, the perdeuterated protein was found to be closer in structure to H-HEWL, than the hydrogenated recombinant variant, due to different folding dynamics associated with the macromolecular isotope effect. It is, however, striking how the macromolecular isotope effect combined with *in vitro* refolding can lead to such variations in protein structure and our study underlines the need of carefully characterizing the different variants whenever possible to increase our understanding of the multimodal effects.

References

Artero, J., Härtlein, M., McSweeney, S. & Timmins, P. (2005). *Acta Crystallogr.* **D61**, 1541–1549.

Batas, B. & Chaudhuri, J. B. (1996). *Biotechnol. Bioeng.* **50**, 16–23.

Batas, B., Schiraldi, C. & Chaudhuri, J. B. (1999). *J. Biotechnol.* **68**, 149–158.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.

Berns, D. S. (1963). *Biochemistry*. **2**, 1377–1380.

Blake, C. C., Johnson, L. N., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1967). *Proc. R. Soc. London. Ser. B. Biol. Sci.* **167**, 378–388.

Breyton, C., Gabel, F., Lethier, M., Flayhan, A., Durand, G., Jault, J. M., Juillan-Binard, C., Imbert, L., Moulin, M., Ravaud, S., Härtlein, M. & Ebel, C. (2013). *Eur. Phys. J. E*. **36**,.

Brockwell, D., Yu, L., Cooper, S., McCleland, S., Cooper, A., Attwood, D., Gaskell, S. J. & Barber, J. (2001). *Protein Sci.* **10**, 572–580.

Budayova-Spano, M., Lafont, S., Astier, J.-P., Ebel, C. & Veesler, S. (2000). *J. Cryst. Growth*. **217**, 311–319.

Cheetham, J. C., Artymiuk, P. J. & Phillips, D. C. (1992). *J. Mol. Biol.* **224**, 613–628.

Chen, J. C. H., Hanson, B. L., Fisher, S. Z., Langan, P. & Kovalevsky, A. Y. (2012). *PNAS*. **109**, 15301–15306.

Cuypers, M. G., Trubitsyna, M., Callow, P., Forsyth, V. T. & Richardson, J. M. (2013). *Nucleic Acids Res.* **41**, 2020–2033.

Dong, A., Kendrick, B., Kreilgård, L., Matsuura, J., Manning, M. C. & Carpenter, J. F. (1997). *Arch. Biochem. Biophys.* **347**, 213–220.

Dunne, O., Weidenhaupt, M., Callow, P., Martel, A., Moulin, M., Perkins, S. J., Haertlein, M. & Forsyth, V. T. (2017). *Eur. Biophys. J.* **46**, 425–432.

Efimova, Y. M., Haemers, S., Wierczinski, B., Norde, W. & van Well, A. A. (2006). *Biopolymers*. **85**, 264–273.

Fisher, S. J., Blakeley, M. P., Howard, E. I., Petit-Haertlein, I., Haertlein, M., Mitschler, A., Cousido-Siah, A., Salvay, A. G., Popov, A., Muller-Dieckmann, C., Petrova, T. & Podjarny, A. (2014). *Acta Crystallogr.* **D70**, 3266–3272.

Gamble, T. R., Clauser, K. R. & Kossiakoff, A. A. (1994). *Biophys. Chem.* **53**, 15–26.

Glasoe, P. K. & Long, F. A. (1960). *J. Phys. Chem.* **64**, 188–190.

Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O. & Forsyth, V. T. (2016). *Methods in Enzymology*, Vol. *566*, pp. 113–157.

Harrington, W. F. & von Hippel, P. H. (1961). *Arch. Biochem. Biophys.* **92**, 100–113.

Hattori, A., Crespi, H. L. & Katz, J. J. (1965). *Biochemistry*. **4**, 1213–1225.

Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Crystallogr.* **D61**, 1413–1417.

Jasnin, M., Tehei, M., Moulin, M., Haertlein, M. & Zaccai, G. (2008). *Eur. Biophys. J.* **37**, 613–617.

Kabsch, W. (2010). *Acta Crystallogr.* **D66**, 133–144.

Koruza, K., Lafumat, B., Nyblom, M., Mahon, B. P., Knecht, W., McKenna, R. & Fisher, S. Z. (2019). *Acta Crystallogr.* **D75**, 895–903.

Koruza, K., Lafumat, B., Végvári, Knecht, W. & Fisher, S. Z. (2018). *Arch. Biochem. Biophys.* **645**, 26–33.

Kuhlman, B. & Raleigh, D. P. (1998). *Protein Sci.* **7**, 2405–2412.

Laux, V., Callow, P., Svergun, D. I., Timmins, P. A., Forsyth, V. T. & Haertlein, M. (2008). *Eur. Biophys. J.* **37**, 815–822.

Liu, X., Hanson, B. L., Langan, P. & Viola, R. E. (2007). *Acta Crystallogr.* **D63**, 1000–1008.

Maenaka, K., Matsushima, M., Song, H., Sunada, F., Watanabe, K. & Kumagai, I. (1995). *J. Mol. Biol.* **247**, 281–293.

Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. (1995). *Nat. Struct. Biol.* **2**, 852–855.

Meilleur, F., Contzen, J., Myles, D. A. A. & Jung, C. (2004). *Biochemistry*. **43**, 8744–8753.

Nichols, P. J., Falconer, I., Griffin, A., Mant, C., Hodges, R., McKnight, C. J., Vögeli, B. & Vugmeyster, L. (2020). *Protein Sci.* **29**, 1641–1654.

Petit-Haertlein, I., Blakeley, M. P., Howard, E., Hazemann, I., Mitschler, A., Haertlein, M. & Podjarny, A. (2009). *Acta Crystallogr.* **F65**, 406–409.

Phillips, D. C. (1967). *PNAS*. **57**, 484–495.

Ramos, J., Laux, V., Haertlein, M., Boeri Erba, E., McAuley, K. E., Forsyth, V. T., Mossou, E., Larsen, S. & Langkilde, A. E. (2021). *IUCrJ*. **8**,.

Sasisanker, P., Oleinikova, A., Weingärtner, H., Ravindra, R. & Winter, R. (2004). *Phys. Chem. Chem. Phys.* **6**, 1899–1905.

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

Sheldrick, G. M. (2015). *Acta Crystallogr.* **C71**, 3–8.

Shu, F., Ramakrishnan, V. & Schoenborn, B. P. (2000). *PNAS*. **97**, 3872–3877.

Shugar, D. (1952). *Biochim. Biophys. Acta*. **8**, 302–309.

Strynadka, N. C. J. & James, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.

Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *PNAS*. **95**, 2267–2272.

Tanaka, I., Nishinomiya, R., Goto, R., Shimazakic, S. & Chataked, T. (2021). *Acta Crystallogr.* **D77**, 288–292.

Varga, K., Aslimovska, L., Parrot, I., Dauvergne, M. T., Haertlein, M., Forsyth, V. T. & Watts, A. (2007). *Biochim. Biophys. Acta - Biomembr.* **1768**, 3029–3035.

Vijayakrishnan, S., Kelly, S. M., Gilbert, R. J. C., Callow, P., Bhella, D., Forsyth, T., Lindsay, J. G. & Byron, O. (2010). *J. Mol. Biol.* **399**, 71–93.

Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Crystallogr.* **D63**, 1254–1268.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Crystallogr.* **D67**, 235–242.

Supporting Information

The impact of deuteration and *in vitro* refolding on the structural and biophysical properties of hen egg-white lysozyme

Joao Ramos[1,2,3], Valerie Laux[1,2], Michael Haertlein[1,2], Trevor Forsyth[1,2,4], Estelle Mossou[1,2,4], Sine Larsen[5], Annette Langkilde[3*]

[1]*Institut Laue-Langevin, 71 avenue de Martyrs, 38000 Grenoble, France.*

[2]*Partnership for Structural Biology (PSB), 71 avenue de Martyrs, 38000 Grenoble, France.*

[3]*Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.*

[4]*Faculty of Natural Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.*

[5]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.*

*corresponding author, annette.langkilde@sund.ku.dk

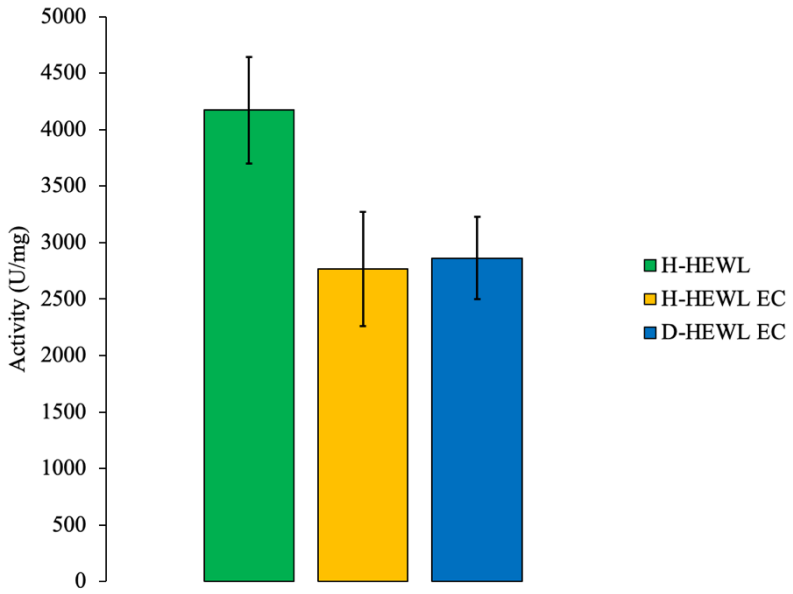


Figure S1

Measured activity rates of H-HEWL, H-HEWL$_{EC}$ and D-HEWL$_{EC}$, at 25° C, at pH 7.5 in H_2O buffer solutions.

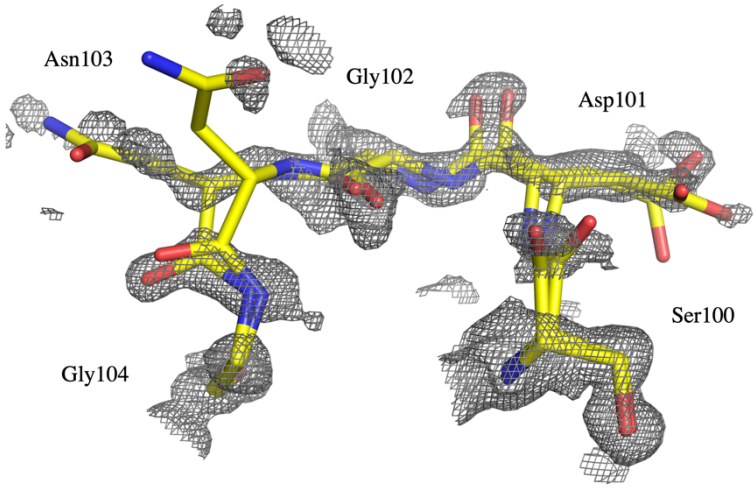


Figure S2

Full Asn103 peptide-plane flip and subsequent disorder of the Lys97-Gly104 region in H-HEWL$_{EC}$. The 2*Fo-Fc* electron density map is contoured at 1 σ.

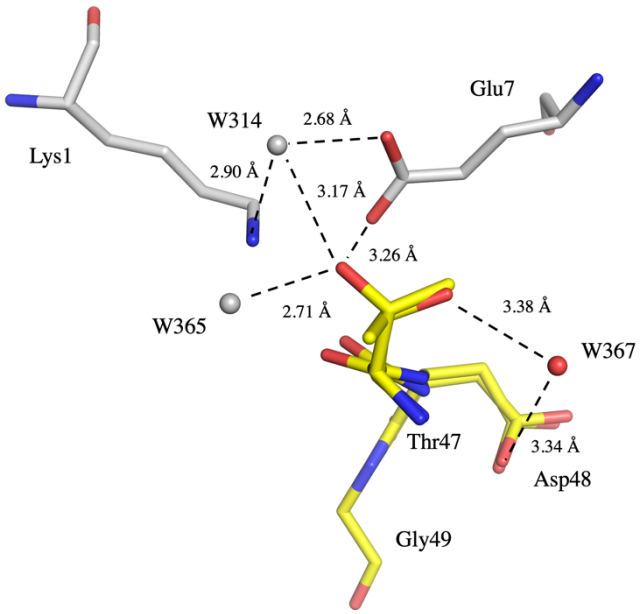


Figure S3

Representation of the H-bond interactions of Thr47 conformations A and B via crystal contacts, which promote the backbone disorder in the Thr47-Gly49 region in H-HEWL$_{EC}$. The residues and waters belonging to adjacent asymmetric units are colored in grey.

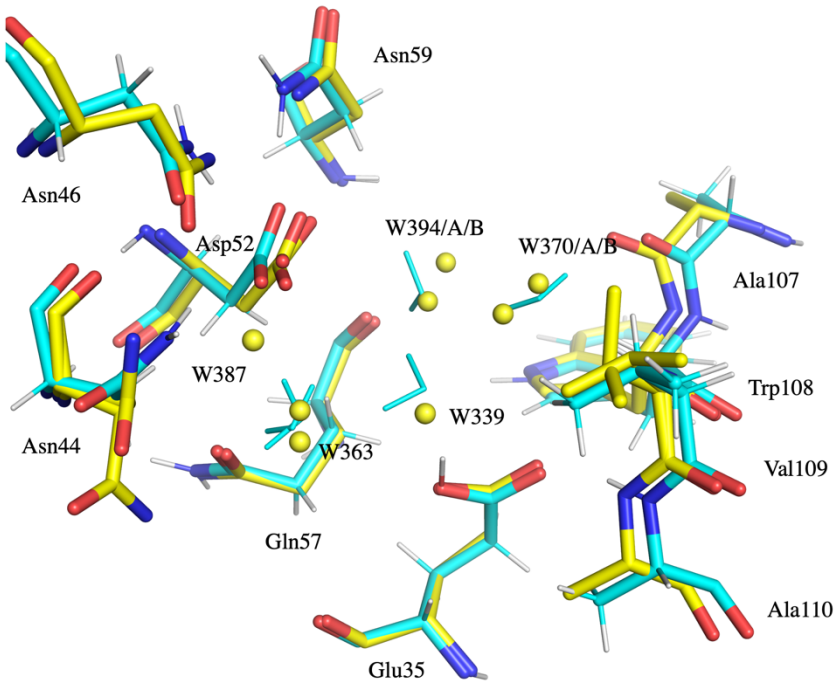


Figure S4

Comparison of the active site structure between the 100 K X-ray structure of H-HEWL$_{EC}$ (yellow) and the 298 K neutron structure of D-HEWL$_{EC}$ (cyan). The respective water molecules are colored according to each model's color. For clarity only the water molecules of H-HEWL$_{EC}$ are labelled.

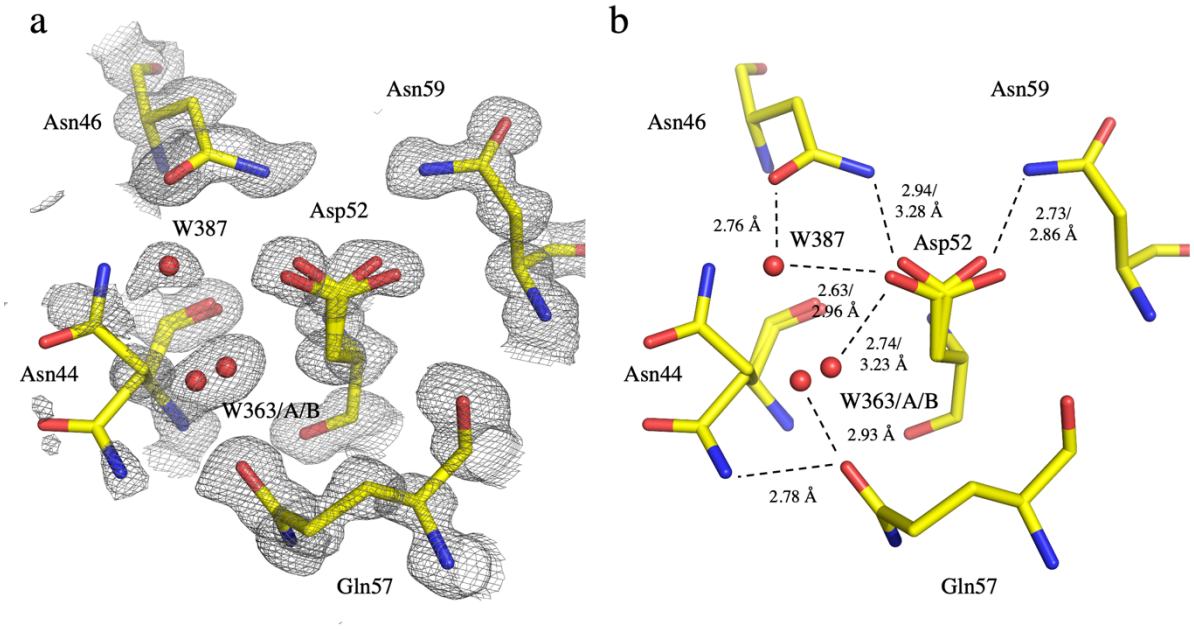


Figure S5

Alternate conformations of Asn44 and W387 found in the H-HEWL$_{EC}$ structure (a), with corresponding active site H-bond interactions (b).

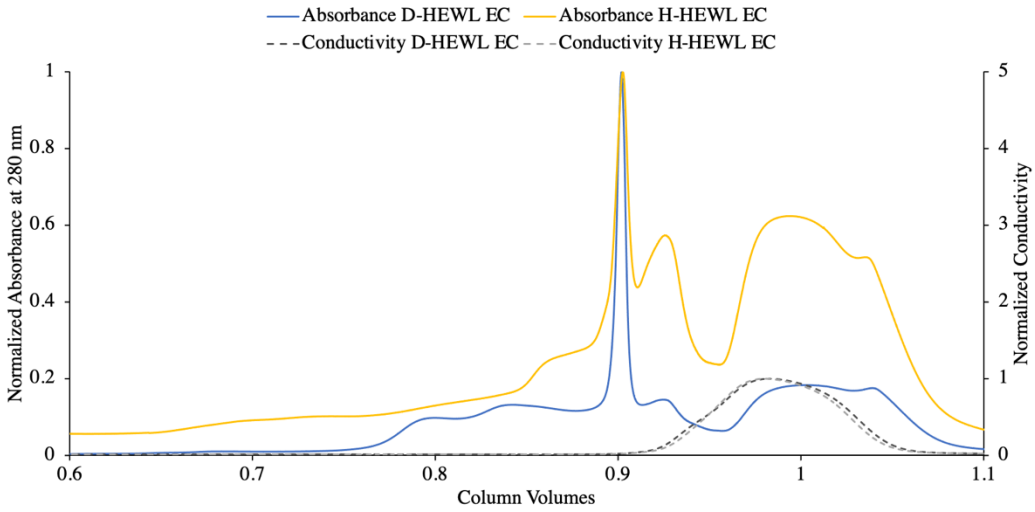


Figure S6

Superposed chromatograms of the refolding gel filtrations of D-HEWL$_{EC}$ and H-HEWL$_{EC}$.

Chapter 5

Manuscript 3

Synopsis

Complete atomic resolution neutron diffraction data was successful obtained from a crystal of perdeuterated HEWL at room temperature and close to the enzyme's optimal pH. This dataset provided not only the basis for the chief study presented in this thesis, but also unique insight into the structure and dynamics of the enzyme in its active state. The unambiguous assignment of protein residue and water molecule conformations through neutron diffraction helps to clarify the enzyme's mechanism of action. In the following manuscript in preparation (Manuscript 3), the conformations of important active site residues and water molecules are evident. This work provides a clear picture of the protein in the active state, which is precluded in X-ray diffraction studies, even at atomic resolution, due to the inability of assigning H positions. Furthermore, important detail is given regarding the benefits of protein perdeuteration in NMX studies, highlighting the existence of cancelation effects at atomic resolution, which were not previously considered (Chen *et al.*, 2012). This report constitutes, to our knowledge, the first time complete atomic resolution neutron diffraction data was obtained from a protein crystal. Finally, the impact of neutron data quality (*e.g.*, resolution, completeness) is discussed, also in light of the common data quality of neutron structures deposited in the PDB. It is important to note that the completeness test presented in the manuscript does not reflect common neutron dataset quality, which corresponds to 80% completeness and 2 Å resolution. The reported results reflect instead incomplete neutron data at atomic resolution. Hence, further analysis is required to truly compare the information available from complete atomic resolution and standard neutron diffraction datasets.

The analysis of refined anisotropic ADPs is omitted in the following study since measurements at different temperatures are required to separate static and dynamic disorder, which interfere in the estimation of atomic motion. Additionally, equivalent X-ray diffraction data are necessary to answer the questions related to the potential bias of ADPs from X-rays.

A complete picture of hen egg-white lysozyme active site at room temperature from 1 Å resolution neutron diffraction data

Joao Ramos[1,2,3], Estelle Mossou[1,4], Valerie Laux[1,2], Sax A. Mason[1], Marie-Helene Lemee[1], Kay Diederichs[5], Michael Haertlein[1,2], V. Trevor Forsyth[1,2,4], Sine Larsen[6], Annette E. Langkilde[3]*

[1]*Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.*
[2]*Partnership for Structural Biology (PSB), 71 avenue des Martyrs, 38000 Grenoble, France.*
[3]*Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.*
[4]*Faculty of Natural Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.*
[5]*Department of Biology, University of Konstanz, Box 647, D-78457 Konstanz, Germany.*
[6]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.*
*corresponding author, annette.langkilde@sund.ku.dk

Abstract

Complete neutron diffraction data to 1 Å resolution have been collected on crystals of perdeuterated hen egg-white lysozyme as well as of D_2O buffer exchanged hydrogenated variant. Despite being one of the most studied enzymes in structural biology, there are still questions and ambiguities surrounding its mechanism and active site. The unambiguous assignment of Hydrogen (H)/Deuterium (D) positions in protein residues and water molecules facilitates the interpretation of complex hydrogen-bond networks and provides an explicit and complete picture of the active site, including protonation of the catalytic Glu35 and interactions of the non-protonated Asp52. In addition, due to protein refolding in H_2O solution, caged H atoms were found in several backbone amide positions of specific secondary structure motifs, revealing a high degree of protection against solvent substitution, overall consistent with previous observations. While neutron scattering density cancellation effects, arising from H atoms, are generally thought to be only visible at resolutions better than 1.5 Å, our data suggest that the presence of H atoms in the partially deuterated (D_2O soaked) variant leads to the preclusion of protein residue disorder in neutron maps at atomic resolution, emphasizing further the importance of perdeuteration in neutron diffraction studies.

Keywords

Neutron crystallography, atomic resolution, perdeuterated lysozyme, HEWL active site, neutron scattering cancellation effects, protein folding, caged hydrogens, H/D exchange.

Introduction

Neutron macromolecular crystallography (NMX) is regarded as a complementary method to macromolecular X-ray crystallography. This consideration is well justified not only by the supplementary information available from neutron diffraction data, but also by the limitations inherent to instrumentation and sample requirements of NMX. These boundaries stem from the relatively low neutron flux at current reactor and spallation sources compared to the photon flux offered by synchrotrons. The neutron flux can be lower by 9 orders of magnitude and this hinders the intensity of the neutron data recorded, which has to be compensated by the use of larger crystal volumes (mm^3 instead of μm^3), sample perdeuteration, and extended data collection time (several days as opposed to seconds). In neutron scattering studies, isotopic substitution of hydrogen (H) for deuterium (D) has significant benefits to data quality. While X-rays interact with the electrons of atoms and, so, their scattering lengths increase with the atomic number, neutron scattering is from atomic nuclei, which possess a similar coherent scattering length in the case of D, ^{12}C, ^{14}N, ^{16}O (Sears, 1992). On the other hand, H has a negative coherent scattering length (-3.74 fm) and, also, a large incoherent scattering cross section (80.27 barn) (Sears, 1992). Therefore, the presence of H atoms has two undesired effects: 1) cancellation of the coherent neutron scattering signal from other atom types, leading to loss of diffraction information (Chen *et al.*, 2012; Fisher *et al.*, 2014); 2) significant increase of the background noise in diffraction images, resulting in the weakening of Bragg reflections at higher resolutions (Hazemann *et al.*, 2005; Cuypers *et al.*, 2013). The necessity of producing relatively large crystals together with the requirement of sample deuteration, and also the challenge of obtaining complete and high resolution diffraction data, are factors that hinder the widespread interest and investment in studying protein structure through NMX. Routinely in NMX, large hydrogenated protein crystals soaked in D_2O solutions are studied (Liebschner *et al.*, 2018), since the production of perdeuterated protein can be cumbersome and expensive (Haertlein *et al.*, 2016). In the case of perdeuteration, crystal volume requirements for neutron diffraction studies can be reduced by 1 order of magnitude, since all incoherent scattering from H atoms is eliminated from the sample (Hazemann *et al.*, 2005). It is estimated that the signal-to-noise ratio decreases by a factor of three if D_2O soaking of hydrogenated protein crystal is used instead of producing perdeuterated sample (Shu *et al.*, 2000; Hazemann *et al.*, 2005).

Until the 1st of February 2021, 177 crystal structures had been deposited in the Protein Data Bank (PDB) (Berman *et al.*, 2000) based on neutron diffraction data, compared to the 153676 structures derived from X-ray data. Furthermore, out of these 177 models, 107 (~60%) were refined against both neutron and X-ray data, due to the low completeness and limited resolution of the neutron data alone. To date, there are no reports of complete neutron diffraction data on proteins extending to atomic resolution (*i.e.*, < 1.2 Å). Nevertheless, NMX work conducted in recent years has contributed with biologically relevant insight to enzymatic function (Blakeley *et al.*, 2008; Oksanen *et al.*, 2009; Cuypers *et al.*, 2013; Howard *et al.*, 2016; Schiebel *et al.*, 2018), protein and water dynamics (Niimura & Bau,

2008; Chen *et al.*, 2012; Fisher *et al.*, 2014), and drug design (Fisher *et al.*, 2012; Weber *et al.*, 2013; Aggarwal *et al.*, 2016; Gerlits *et al.*, 2016). These advancements were possible because neutron diffraction data at resolutions better than 2.5 Å allow the location of D atoms (Chen *et al.*, 2012). Thus, unambiguous assignment of protonation states, hydrogen-bond (H-bond) networks, and solvent interactions become possible from high quality neutron diffraction data. While H atoms are fundamental for protein biological function and comprise approximately 50% of all protein atoms, they remain elusive to X-ray diffraction, even at atomic resolutions.

In the case of the hen egg-white lysozyme (HEWL) catalytic mechanism, the main unanswered question concerns the role of Asp52 in the stabilization of electron deficient substrate molecule after the breakage of the glycosidic bond. Two hypotheses were proposed, which suggest the formation of a short-lived covalent bond between Asp52 and the substrate (Koshland Jr., 1953), or the existence of a positively charged carbonium ion that is compensated by the negatively charged Asp52 (Phillips, 1967). However, the unambiguous orientation of certain active site residues (*e.g.*, Asn44, Asn46, Gln57) and water molecules, and their subsequent H-bond interactions remain unresolved from X-ray data, even at ultra-high resolution (Wang *et al.*, 2007). Additionally, the protonation states of both catalytic residues, Glu35 and Asp52, have only been observed in two neutron studies, where data was collected at room temperature and at a pH close to the optimal pH for enzymatic activity (Mason *et al.*, 1984; Bon *et al.*, 1999). Despite of the valuable insight provided by these NMX studies regarding the protonation of HEWL catalytic residues, a clear and complete picture of the active site interactions is still lacking.

Liebschner *et al.* (2018) had surveyed all neutron structures deposited in the PDB and concluded that the average neutron dataset completeness was ~80%, while the average resolution extended to ~2 Å. Currently, the highest resolution neutron structure reported is of crambin (46 residues, MW = 4.7 kDa), at 1.1 Å (PDB entry 4fc1) (Chen *et al.*, 2012). Despite the small orthorhombic unit cell of crambin (a = 22.79 Å, b = 18.82 Å, c = 41.04 Å), the dataset obtained is only 78.8% complete, 65.8% in its outer resolution shell. The study of crambin is based on a crystal of hydrogenated protein, which was buffer exchanged to D_2O solution. This detail reveals that neutron scattering cancellation effects took place upon neutron data collection, due to the negative scattering contribution of H atoms. Thus, it is reasonable to assume that this may have been detrimental to the quality of the diffraction data collected and limited the information available in the observed neutron scattering length density maps (from now on referred to as neutron maps), as shown also in other studies (Shu *et al.*, 2000; Fisher *et al.*, 2014).

The clear and complete picture of the HEWL active site is obtained from neutron diffraction data to 1 Å resolution collected on perdeuterated HEWL (D-HEWL) and D_2O buffer exchanged hydrogenated HEWL (H-HEWL). The resulting neutron structures are particularly relevant for the study of enzymatic function, since they reflect the state of HEWL at room temperature and close to the enzyme's optimal pH, depicting its biologically active conformation. To our knowledge, these datasets constitute the first complete neutron diffraction data at atomic resolution obtained from protein crystals.

The unambiguous assignment of H/D positions in both neutron models enabled the clarification of the orientations of protein residues and water molecules, unravelling the intricate H-bond networks present in HEWL active site. Additionally, the importance of protein perdeuteration in neutron diffraction data quality is discussed, through a direct comparison between D-HEWL and H-HEWL neutron datasets. The effects of data resolution and completeness are explored, showcasing the type of information available from neutron crystallographic data at different levels.

Results and Discussion

1 Å resolution neutron structure of perdeuterated lysozyme

Room temperature neutron diffraction data was collected from a triclinic (P1) ~4 mm^3 D-HEWL crystal, at D19 (ILL). The resolution extends to 1 Å, being 97.1% complete (91.8% in the outer resolution shell). Obtaining this high quality neutron dataset was possible due to several factors which played a role in overcoming the current limitations of NMX: 1) protein perdeuteration; 2) production of large well-diffracting crystals (triclinic with ~30% solvent content); 3) availability of thermal neutrons and of a large position-sensitive curved detector at D19 (ILL), covering 120° (NN segments) horizontally and 30° vertically; 4) the use of XDS (Kabsch, 2010) in D19 neutron data reduction. Additionally, room temperature neutron diffraction data to 1.07 Å has also been collected at D19 (ILL) from a ~6 mm^3 triclinic H-HEWL crystal, soaked in D_2O based buffer for solvent exchange and H/D exchange in labile positions. The data merging and model refinement statistics for both H-HEWL and D-HEWL are presented in Table 1.

The D19 instrument is ideal for high resolution NMX studies (Cuypers *et al.*, 2013; Langan *et al.*, 2014). Its tunable monochromatic beam of thermal neutrons and large detector provide an extensive coverage of the reciprocal space with reduced reflection overlap. However, previous work (data not shown) suggested that the built-in data reduction software was not suitable for the correct integration of weaker intensity reflections. In addition, the lack of profile fitting and the inability to resolve overlapping reflections hampered access to atomic resolution diffraction data for large unit cells. To overcome this issue, XDS (Kabsch, 2010) was implemented in D19 neutron data reduction, also applicable for future NMX studies at D19 with the aim of investigating structural and dynamical protein details only available through high resolution neutron data.

SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) was employed in the model refinement against neutron diffraction data due to its high degree of flexibility, *e.g.*, with possible modelling of partially occupied H/D positions, also in residues with alternate conformations. The model refinement strategy used is summarized in Table 2, alongside with the corresponding variations in R-factors, GooF and data-to-parameter ratio. Residues with alternate conformations, water molecules and ions were modelled to a great extent while maintaining isotropic ADPs, due to the limited data-to-parameter ratio of 7.1 at 1.13 Å (R_{work} = 18.09%, R_{free} = 22.95%). This resolution cut-off was initially used to maintain conservative data merging statistics: I/σ (I) = 3.1; $CC_{1/2}$ = 87.2%; R_{meas} = 34.3% in the outer resolution shell (1.19 - 1.13 Å). By extending the data to 1 Å resolution, the gap between the R-factors was reduced from 4.86% to 4.53%, and the GooF decreased from 3.19 to 2.69, showing that the inclusion of diffraction data to 1 Å improved the accuracy of the model (Fig. S1), as suggested by Karplus & Diederichs (2012) and Maly *et al.* (2020). Confirming this increase in data, and thereby in data-to-parameter ratio, subsequent refinement rounds included the refinement of anisotropic ADPs for 89% of all atoms in the asymmetric unit.

While the refinement of isotropic ADPs corresponds to the addition of 1 parameter/atom, refining anisotropic ADPs increases the number of parameters to 6 per atom. Additionally, since H/D atoms comprise ~50% of all protein atoms and neutron diffraction data allows their determination, the number of parameters is already doubled. In the case of D-HEWL, the number of parameters in model refinement when assuming isotropic ADPs is approximately 10000. This number increases to ~15000 when refining anisotropic ADPs for non-H/D atoms, and further to ~20000 when D atoms are included in the refinement of anisotropic ADPs. This observation emphasizes the importance of obtaining complete neutron data in order to be able to reliably refine anisotropic ADPs for all atoms.

The refinement of anisotropic ADPs (*cf.* Methods) yielded at each stage an improvement in the R-factors and the difference between R_{free} and R_{work} did not surpass 5.3% (Table 2 & Supplementary Table S1), justifying this approach and suggesting no significant model overfitting. The final D-HEWL neutron model (R_{work} = 11.86%, R_{free} = 17.15%, GooF = 2.10) includes 24 alternate conformations of protein residues, 58 structural water molecules (42 of which modelled as D_2O), and 6 nitrate ions.

Table 1

Data merging and model refinement statistics for the H-HEWL and D-HEWL structures determined by neutron diffraction. The values in parenthesis correspond to the respective outer resolution shell.

	H-HEWL	**D-HEWL**
Temperature (K)	298	298
Source	D19, ILL, FR	D19, ILL, FR
Detector	^{3}He + CF_4	^{3}He + CF_4
Wavelength (Å)	1.458	1.455
Resolution range (Å)	21.67 – 1.07 (1.12 – 1.07)	21.44 – 1.00 (1.04 – 1.00)
Space group	P 1	P 1
Unit cell **a, b, c (Å)** **α, β, γ (°)**	 27.31, 31.9, 34.43 88.611, 71.673, 68.251	 27.1, 31.59, 34.1 89.156, 72.027, 67.751
Total reflections	96927 (8360)	130852 (8899)
Unique reflections	43105 (5069)	51887 (4927)
Multiplicity	2.2 (1.7)	2.5 (1.8)
Completeness (%)	95.8 (90.2)	97.1 (91.8)
Mean I/ σ(I)	5.7 (1.4)	7.1 (1.5)
R_{merge}	0.100 (0.656)	0.078 (0.544)
R_{meas}	0.121 (0.861)	0.092 (0.707)
R_{pim}	0.066 (0.551)	0.047 (0.446)
$CC_{1/2}$	0.990 (0.470)	0.995 (0.543)
Refl. used in refinement with *Fo*/σ(*Fo*) > 4 / all refl.	25735 / 40949	34066 / 58718
Refl. used for R-free with *Fo*/σ(*Fo*) > 4 / all refl.	1363 / 2156	1789 / 3090
R_{work} for refl. with *Fo*/σ(*Fo*) > 4 / all refl.	14.13 / 20.73	11.86 / 17.05
R_{free} for refl. with *Fo*/σ(*Fo*) > 4 / all refl.	18.82 / 25.52	17.15 / 22.30
Number of non-H/D atoms (overall)	1179	1288
macromolecules	1088	1203
ligands	24	24
solvent	67	58
Protein residues	129	130
R.m.s.d., bond lengths (Å)	0.0108	0.0118
R.m.s.d., angles (°)	2.35	2.55
Ramachandran favored (%)	99.21	99.22
Ramachandran allowed (%)	0.79	0.78
Ramachandran outliers (%)	0	0
Rotamer outliers (%)	3.51	2.40
Clashscore	7.38	10.82
Average B-factor (overall) (Å^2)	20.75	19.50
macromolecule	20.14	19.09
ligands	27.81	30.01
solvent	24.43	23.90

Table 2

D-HEWL model refinement strategy used in SHELXL against neutron diffraction data. The changes in R-factors for reflections with I/σ (I) > 4, goodness-of-fit (GooF), and the data-to-parameter ratio are presented for each step. Corresponding table for H-HEWL is available as Table S1.

Round	Refinement step	Resolution (Å)	R_{work} (%)	R_{free} (%)	GooF	Data-to-parameter ratio
1	Rigid body refinement	22.84 – 2.0	41.87	44.58	10.0	10.3
2	Coordinates and isotropic ADPs (non-H/D)	22.84 – 1.5	35.50	42.41	7.65	4.7
3	Adding D atoms in idealized positions	22.84 – 1.5	33.55	39.20	7.62	8.9
4	Adding D_2O and ions	22.84 – 1.13	26.68	31.18	4.32	12.5
5	Replacing caged H atoms	22.84 – 1.13	21.39	24.45	3.49	12.5
6	Modelling protein residue disorder	22.84 – 1.13	20.20	23.77	3.35	12.0
7	Refining D atom positions	22.84 – 1.13	18.58	23.31	3.26	7.2
8	Increasing data resolution	22.84 – 1.00	19.01	23.54	2.69	8.8
9	Anisotropic ADPs for non-H/D atoms	22.84 – 1.00	14.84	20.56	2.34	6.2
10	Anisotropic ADPs for protein D atoms	22.84 – 1.00	13.14	18.57	2.29	5.2
11	Anisotropic ADPs for well-ordered waters	22.84 – 1.00	12.29	17.94	2.17	5.1
12	Minor corrections	22.84 – 1.00	11.86	17.15	2.10	5.1

Water molecules and nitrate ions

The neutron densities for acetate and nitrate ions, normally found in triclinic HEWL, were not well defined, likely due to disorder and partial occupancy. Distinguishing between acetate and nitrate ions is more difficult in the neutron maps because of the similar neutron scattering lengths of the atoms. Using the corresponding D-HEWL X-ray structure as a reference (PDB entry 7ave; Ramos *et al.* 2021), it was possible to identify neutron densities at six different locations likely from nitrate ions. Water molecules at the surface of the protein were hydrogen-bonded to the main-chain peptide groups either as donors to O or acceptors of N-H groups. Water molecules participating in only one interaction with protein residues were usually (partially) disordered, resulting in neutron density only being evident for one or two atoms (D, O or D-O), while the remaining atom(s) could not be modelled unambiguously. It was possible to model a total of 42 well-ordered water molecules as D_2O, while 16 disordered waters were modelled as single O atoms. Evidently, the number of water molecules identified in the neutron model is significantly lower compared to the 136 observed in the X-ray structure of D-HEWL (PDB entry 7ave; Ramos *et al.* 2021). Additional factors that contribute to the preclusion of water molecules in the neutron maps are the significantly lower intensity of the neutron data compared to the reported X-ray data at high resolution and the difference in data collection temperatures. Moreover, since the coherent neutron scattering length of D atoms is comparable to that of O, disorder of the water molecules results in diffuse scattering length densities. This is in contrast to the results from X-ray diffraction data, where the electron density of water molecules is dominated by the scattering contribution of O atoms.

Water molecules are an essential part of a protein structure, either by mediating H-bonds which stabilize protein conformation or by allowing a degree of protein flexibility, which enables protein function. While H-bond interactions typically can be deduced by geometry and distance between atoms in X-ray structures, neutron data can provide a way of identifying unambiguously the interacting atoms. Besides the network in the active site discussed below, the H-bond network connecting residues Gly49, Thr51 and Asp66 is an example of the role of water meditated H-bonds (Fig. 1). The 1 Å resolution neutron map allows the identification of the donor and acceptor atom pairs and, interestingly, shows that one of the O atoms from the Asp66 carboxylate group acts as a bifurcated acceptor for hydrogen atoms from both Tyr53 and W302.

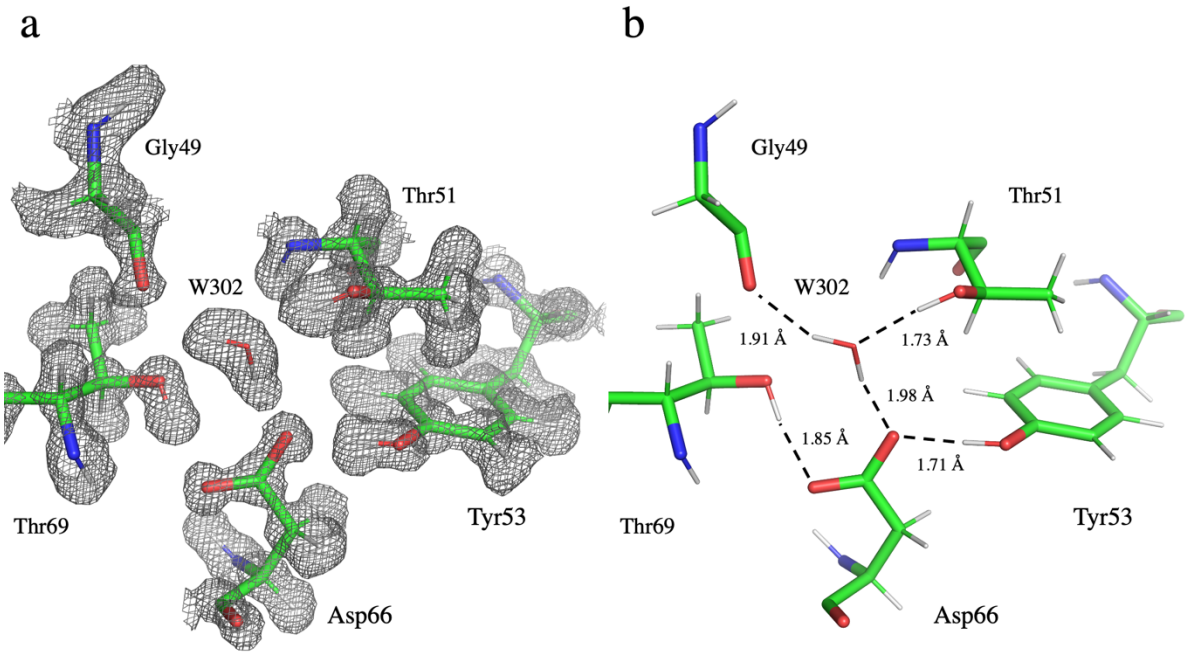


Figure 1

The importance of water mediated H-bond interactions in HEWL structure and the ability to unambiguously determine the H-bond pairs from neutron data at 1 Å resolution. The H-bond network enabled by the presence of W302, with 2*Fo-Fc* neutron map contoured at 1 σ (a), and corresponding H-bond pairs and distances (b).

Disorder and Asn103 peptide-plane flip

The observed disorder of protein residues was similar to that previously reported in the structure modelled based on X-ray data to 0.98 Å resolution (Supplementary Fig. S2) (PDB entry 7ave; Ramos *et al.* 2021). There are some exceptions, where alternate conformations were visible in the electron density maps but not in the neutron maps, which might be associated with the weaker intensity of the neutron data compared to that of the X-ray data. On the other hand, for residues Asn19, Trp28, Trp63, Arg68, Ile124, alternate conformations were evident in the neutron maps but not in the electron density maps. Interestingly, the alternate conformation for Asn19 was not observed in the D-HEWL X-ray structure, but it was visible in both the unlabeled H-HEWL and the D-HEWL expressed in *Pichia*

pastoris (Ramos *et al.*, 2021). In the case of Trp28, Trp63, Arg68, Ile124, the modelled alternate conformations were not present in the X-ray electron density maps of either H-HEWL and D-HEWL. Furthermore, excluding the Arg68, these residues display single conformations in the 0.65 Å resolution X-ray structure reported by Wang *et al.* (2007). Although the intensity of the neutron diffraction data collected is weaker than that of the aforementioned X-ray data, it seems that the additional coherent scattering signal from the D atoms contributes to the observation of disorder, especially in residues with larger side-chains. Backbone disorder is a hallmark of atomic and ultra-high resolution X-ray diffraction data, however, to our knowledge, this level of structural detail has not been described before from neutron data. The neutron structure of D-HEWL at 1 Å resolution displays main-chain disorder in three different regions, namely His15-Gly16, Lys97-Gly104, and Lys116-Gly117. This disorder has been observed in previous X-ray crystallographic studies of D-HEWL (Ramos *et al.*, 2021). Noticeably, the partial peptide-plane flip of Asn103 and the associated backbone disorder (Lys97-Gly104), otherwise evident only in atomic and ultra-high resolution X-ray structures of HEWL (Walsh *et al.*, 1998; Wang *et al.*, 2007; Ramos *et al.*, 2021), are also visible in the neutron maps of D-HEWL. The refined occupancy for the Lys97-Gly104 conformation associated with the peptide-plane flip of Asn103 is identical in both the 100 K X-ray and the room temperature neutron D-HEWL structures (46%).

The active site – the full and detailed picture

Three HEWL structures solved using neutron diffraction are currently deposited in the PDB (Niimura *et al.*, 1997; Bon *et al.*, 1999; Kita & Morimoto, 2020). The most recent, PDB entry 6k8g (Kita & Morimoto, 2020), using joint X-ray and neutron refinement of the structure in the tetragonal crystal form focused on H/D exchange. However, the protonation of Glu35 is not observed even though it has been crystallized at pH 4.6. With a higher pH of 7 during crystallization, the HEWL structure from Niimura *et al.* (1997) (PDB entry 1io5) is far away from the enzyme's most active form and also no protonation of Glu35 is observed. Protonation of Glu35 is thus only observed in a single deposited structure (PDB entry 1lzn, Bon *et al.*, 1999) and additionally in an earlier study by Mason *et al.* (1984), while in both cases the full details of the H-bond patterns and orientation of neighboring water molecules are unclear.

With room temperature neutron data collection and crystallization conditions in pH close to the optimal pH for enzymatic activity, our structure represents the active form of HEWL, as also confirmed during characterization of the perdeuterated variants (Ramos *et al.*, 2021). The neutron scattering length densities obtained in this complete atomic resolution study show clearly the protonation of Glu35 and allow the unambiguous assignment of the orientations of both Asn46 and surrounding water molecules, thus providing a full picture of the H-bond patterns around both catalytic residues (Fig. 2). The configuration of the active site was found to be identical between D-HEWL and H-HEWL, with a minor exception regarding the orientation of W332 (Supplementary Fig. S3).

Glu35 Oε1 is undoubtedly deuterated forming a H-bond to W332/A, while the side-chain conformation is stabilized by the Ala110 N-D ··· Glu35 Oε1 H-bond. The carboxylate of Asp52 is closely interacting with three neighboring Asn residues. A short H-bond is present from Asn59 Nδ2-Dδ2b ··· Asp52 Oδ1, whereas Asp 52 Oδ2 is a bifurcated acceptor with longer H-bonds from Asn46 Nδ2-Dδ2a and either Asn44 Nδ2-Dδ2b or W306. While the short 1.65 Å H-bond between Asn59 and Asp52 appears to be constraining the mobility of Asp52 Oδ1, Oδ2 seems to be more loosely restrained due to the weaker 2.47 Å H-bond from Asn46 and to the apparent dynamics of Asn44 and W306. Despite of the assignment of Asn44 side-chain orientation being clear in the neutron map, the position of both Dδ2a and Dδ2b is not well defined. This observation is supported by the higher B-factors of both deuterium atoms (~ 30 $Å^2$) when compared to those of Asn46 and Asn59, ~15 $Å^2$ and ~10 $Å^2$, respectively. Concomitantly, the disorder of W306 is evident, fluctuating between a short H-bond to Asp52 Oδ2 or to Gln57 Oε1. These observations are consistent with a looser association of Asp 52 Oδ2, which permits the stabilization of the electron deficient substrate molecule after the glycosidic bond breakage.

W332/A occupies the position of the catalytic water, responsible for reverting the protonation state of Glu35 side-chain and for the nucleophilic attack to the substrate. Despite of its importance in the enzymatic catalysis and of its short 1.83 Å H-bond to Glu35 Dε1, this water molecule is transient. Its dynamics is shown by its refined occupancy of 41% and by the presence of an alternate conformation interacting with Gln57 O. The disorder displayed by the water molecules in the active site demonstrates the high degree of dynamics present in the enzyme binding cleft. While water orientations and subsequent interactions are evident, none of the water molecules is retained in a single position. Both the catalytic water and the water stabilizing Asp52 display alternate conformations related with the displacements required for enzymatic activity. Additionally, W330 was refined in a single conformation but its refined ADPs (31 $Å^2$ for O) suggest disorder, even though its interactions with Ala107 O and Val109 N-D are clear.

Interestingly, despite the similar crystallization conditions and outstanding X-ray data resolution of 0.65 Å (Wang *et al.*, 2007) with observable electron density for several H atoms, protonation of Glu35 is not observed or included, neither in the original model (Wang *et al.*, 2007) nor in the subsequent bonding analysis using multipole parameters (Held & Van Smaalen, 2014). However, as with most of the X-ray data, this experiment was performed at 100 K and whether influenced by cryoprotecting agents or temperature, the data were not collected from a crystal representing an active state. Additionally, radiation damage is always a concern when collecting high resolution diffraction data with synchrotron X-rays. On the other hand, neutrons are non-ionizing radiation and so measurements even at room temperature do not alter protein structure. This reality emphasizes the importance of neutron crystallography in deciphering enzymatic mechanisms (Kossiakoff & Spencer, 1980; Blakeley *et al.*, 2008; Oksanen *et al.*, 2009; Cuypers *et al.*, 2013; Gerlits *et al.*, 2016; Vandavasi

et al., 2016). To fully resolve the questions concerning the HEWL mechanism, additional information is needed. Previous and recent substrate bound structures (Blake *et al.*, 1967; Phillips, 1967; Strynadka & James, 1991; Cheetham *et al.*, 1992; Maenaka *et al.*, 1995; Ogata *et al.*, 2013; Tanaka *et al.*, 2021) have provided significant contributions, but while using X-ray data we note that particular efforts are needed to ensure correct interpretation of H-bond patterns, and careful consideration of the side-chain conformation of Asn46 and Asn44, as well as protonation states.

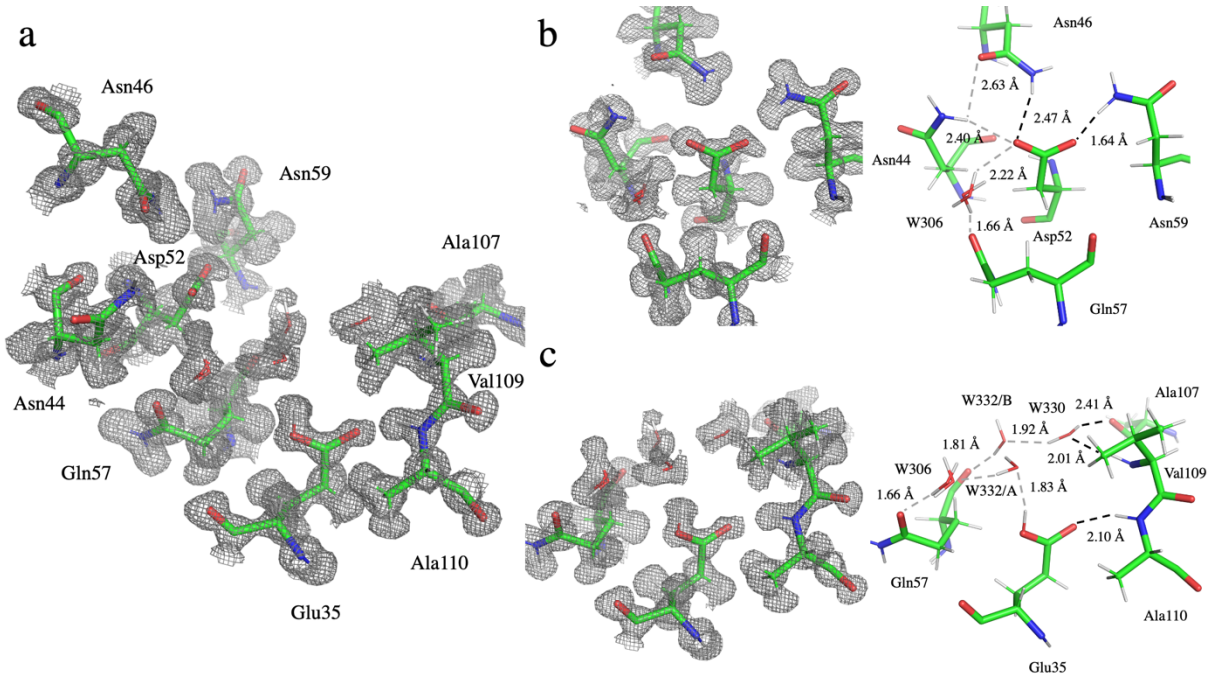


Figure 2

A complete picture of perdeuterated HEWL active site, resolving residue and water orientations and subsequent H-bond networks. The overall configuration of the active site according to the neutron maps, including the catalytic residues Glu35 and Asp52, and respective H-bond interactions (a). H-bond networks stabilizing Asp52 (b) and Glu35 (c), where the interactions connected by black dashed lines appear to be ordered while the ones with grey dashed lines seem to be dynamic (b). The 2*Fo-Fc* neutron maps presented were contoured at 1 σ.

Caged Hs and protein folding

Perdeuteration of HEWL, in this study, involved recombinant expression in *E. coli* inclusion bodies, in fully deuterated conditions (Ramos *et al.*, 2021). Subsequently, pure protein in the unfolded state was subjected to *in vitro* refolding in-column (Ramos *et al.*, 2021). Since the refolding process occurred in H_2O buffer at pH 8, D to H exchange occurred in labile positions (*i.e.*, bound to N, O). Upon (re)folding, some of these positions become inaccessible to solvent since they are buried in the protein structure, in hydrophobic regions or forming strong H-bonds in secondary structure motifs. The refolded

protein was exchanged to a fully deuterated buffer at pD 4.5 (pD = pH + 0.4) (Glasoe & Long, 1960) and crystallized, leaving only H atoms in buried labile positions. Due to the negative neutron scattering length of H atoms and the atomic resolution of the neutron data obtained, these particular sites with caged H atoms are clearly observed (Fig. 3). In total 35 H atoms in backbone N-H groups were identified having full H occupancy, while 21 positions were refined as partially occupied by H and D atoms. 38 caged H atoms participate in main-chain N-H···O H-bonds within α-helices or β-strands. An additional 12 H were found in residues immediately before or after these secondary structure motifs. These results are coherent with the assumptions that secondary structure formation occurs in early stages of protein folding and that these H-bonds are well protected from exchange with solvent molecules (Ikeguchi *et al.*, 1986; Kuwajima, 1989; Miranker *et al.*, 1991; Wildegger & Kiefhaber, 1997). Similar observations were made using NMR by Miranker *et al.* (1991), where it was shown that folding and unfolding transitions of HEWL can be explained by a cooperative two-state model. HEWL was shown to have two domains (α-domain and β-domain) with distinct folding kinetics, which are separated by the enzyme substrate binding cleft (Miranker *et al.*, 1991; Radford *et al.*, 1992). Their observations of protected amide groups agree with the neutron crystallographic data shown.

A comparison between the caged H found in the D-HEWL and H-HEWL neutron structures shows that an additional 10 positions are occupied by H in backbone amides of H-HEWL, six of these H positions appeared to be fully occupied, while the remaining four sites were partially exchanged (Supplementary Table S2). While D-HEWL was exchanged to D_2O solution after refolding, H-HEWL was crystallized in H_2O and only soaked in D_2O solution at the final stage before the neutron diffraction data collection. In the case of H-HEWL, ~50% of all main-chain amide positions are fully or partially occupied by caged H, and the additional caged H found are likely the result of the shorter exchange time (days/weeks compared to months). Thus, sites with reduced flexibility and solvent accessibility resulting in slower exchange were not fully exchanged in our H-HEWL variant. It is, however, interesting to note that a few H atoms were not exchanged in the intrinsically disordered N- and C-terminus regions, namely in residues Phe3, Gly4 and Arg125, indicating slower exchange at these sites. The protection of both Phe3 and Gly4 was not observed in the studies of Miranker *et al.* (1991) and Radford *et al.* (1992), confirming that faster exchange occurs in solution but not in the crystal form.

The previously reported partially deuterated HEWL structures (Mason *et al.*, 1984; Niimura *et al.*, 1997; Bon *et al.*, 1999; Kita & Morimoto, 2020) represent different approaches and timeframes with respect to H/D exchange, being *e.g.*, before crystallization or through slow vapor diffusion after crystallization. Likewise, the refinement approaches differs and the Niimura *et al.* (1997) model includes fully occupied D atoms in all labile positions. Comparing the H/D exchange in the three remaining models with our D-HEWL model, only minor differences are observed (Supplementary Table S2). These discrepancies may be ascribed to differences in data quality and refinement

approaches. Nevertheless, the majority of the protected backbone positions appear consistent with about 30 fully protected positions located mainly in the major α-helices and in the triple stranded β-sheet.

Our neutron data also confirms a caged H in the side-chain amine group of Trp28, displaying full H occupancy, as observed in some of the previous neutron experiments (Mason *et al.*, 1984; Kita & Morimoto, 2020). Trp28 is buried in the hydrophobic core of the HEWL structure and participates in hydrophobic interactions with Leu17, Tyr20, Tyr23, Leu56, Ala95, Val99, Trp108. The disorder of Trp28 side-chain appears to be associated with the displacement of Val99, which belongs to the disordered Lys97-Gly104 region influenced by the partial peptide-plane flip of Asn103. The Trp28 side-chain amine group does not appear to be involved in strong H-bond interactions, since the distance between the Trp28 Hε1 and Tyr23 O is equal or longer than 2.09 Å, showing that the dominant interactions leading to Trp28 side-chain disorder are hydrophobic. Radford *et al.* (1992) had reported similar protection behaviors at the millisecond time scale for Trp28 and Trp111, both incorporating the HEWL α-domain. However, despite of the strong H-bond between Trp111 Nε1 and Asn27 Oδ1 (1.88 Å), the Trp111 side-chain H is fully exchanged by D, according to the neutron maps. Trp111 is located near the protein surface, highly exposed to the solvent. On the other hand, the presence of a caged H in Trp28 side-chain emphasizes that this residue is at the epicenter of the HEWL hydrophobic collapse during folding.

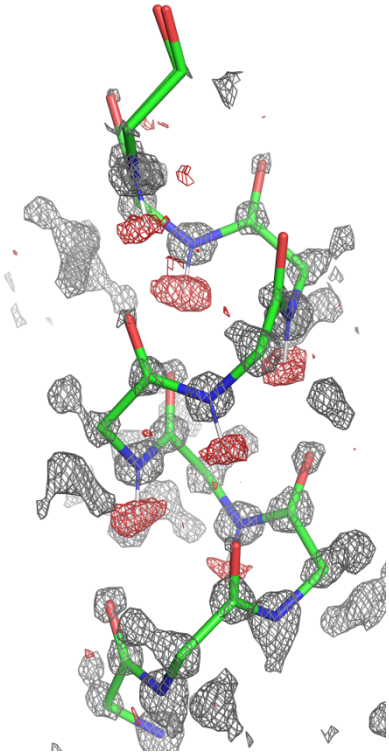


Figure 3

H atoms identified by their negative neutron scattering length densities. Caged H atoms in the main-chain N-H groups of residues in the first α-helix of refolded D-HEWL with 2*Fo-Fc* contoured at 3.5 σ (grey) and at -2 σ (red).

Perdeuteration increases the quality of neutron data

Chen *et al.* (2012) demonstrated that the non-exchangeable H atoms negative coherent neutron scattering results in cancellation effects in neutron maps at resolutions worse than 1.5 Å. The authors noted that at these resolutions, neutron density was lost for carbon atoms in aliphatic side-chains. Fisher *et al.* (2014) also reported that cancellation effects arising from selectively protonated hydrophobic residues were observed in the neutron densities of nearby water molecules. Nevertheless, to our knowledge there is no previous reports of NMX studies at resolutions superior to 1.5 Å, where both hydrogenated (D_2O soaked) and perdeuterated protein variants were compared. The comparison between our neutron structures of H-HEWL and D-HEWL provides further insight to cancellation effects arising from the neutron scattering contribution of H atoms.

It is clear by comparing the $I/\sigma(I)$ of both neutron datasets (Table 1) that protein perdeuteration has a significant impact on data quality. In the experimental conditions used, complete neutron diffraction data to 1 Å resolution was only possible in the case of the perdeuterated crystal, even though the H-HEWL crystal was larger (6 mm^3 compared to 4 mm^3 of the D-HEWL crystal). By eliminating the negative coherent scattering of H atoms, the coherent signal from other types of nuclei is enhanced. Moreover, by subtracting the incoherent scattering contribution of H atoms, the diffraction background noise is reduced allowing the detection of weaker Bragg reflections. The observed improvement in diffraction data quality emphasizes the importance of protein perdeuteration in NMX studies, as shown in previous work (Shu *et al.*, 2000; Hazemann *et al.*, 2005; Cuypers *et al.*, 2013; Fisher *et al.*, 2014).

Cancellation effects are evident also at atomic resolution

The model refinement of H-HEWL was performed in SHELXL using a similar strategy has for D-HEWL, with the exception that H atoms were added to non-exchangeable positions (*i.e.*, bound to C) instead of D, and that these H atoms were fixed in idealized positions and their ADPs refined as isotropic (Supplementary Table S1). The final H-HEWL neutron model (R_{work} = 14.13%, R_{free} = 18.82%, GooF = 2.07) includes 9 alternate conformations of protein residues, 67 structural waters molecules (including 53 D_2O molecules), and 6 nitrate ions. 36 structural D_2O molecules and 4 nitrate ions are conserved in the H-HEWL when compared with the D-HEWL neutron model. While the identification of neutron densities belonging to nitrate ions was aided by the previously reported H-HEWL X-ray structure (PDB entry 7avf, Ramos *et al*. 2021), the modelling of water molecules was performed at the strongest positive peaks in the *Fo-Fc* neutron map, similar to what was performed for the D-HEWL neutron model. Hence, these 36 conserved water molecules seem to be crucial for protein structure, stabilizing its conformation, regardless of the macromolecular isotope effect. While perdeuteration appears to have no significant impact in the ability to interpret the location of well-ordered structural water molecules and also ions in the neutron maps, there is a significant impact in the observation of protein residue disorder (Fig. 4 & Supplementary Fig. S4). The few alternate conformations modelled in the neutron H-HEWL model find correspondence in the 0.65 Å HEWL

structure determined by X-rays (Wang *et al.*, 2007). Moreover, with the exception of the residues Arg45 and Trp111, similar disorder had been reported in previous work (Ramos *et al.*, 2021). Interestingly, the typical cancellation effects produced by the neutron scattering contribution of non-exchangeable H atoms in neutron maps, by "removing" density for atoms in their vicinity is not evident at atomic resolution, as also shown by Chen *et al.* (2012). Instead, the absence of neutron density for protein residue disorder in many cases is observed. Nevertheless, the lack of well-defined negative density for H atoms is indicative that residue disorder may be present. This observation further emphasizes the relevance of perdeuterating samples for NMX studies at atomic resolutions, in order to avoid the preclusion of significant macromolecular structural details.

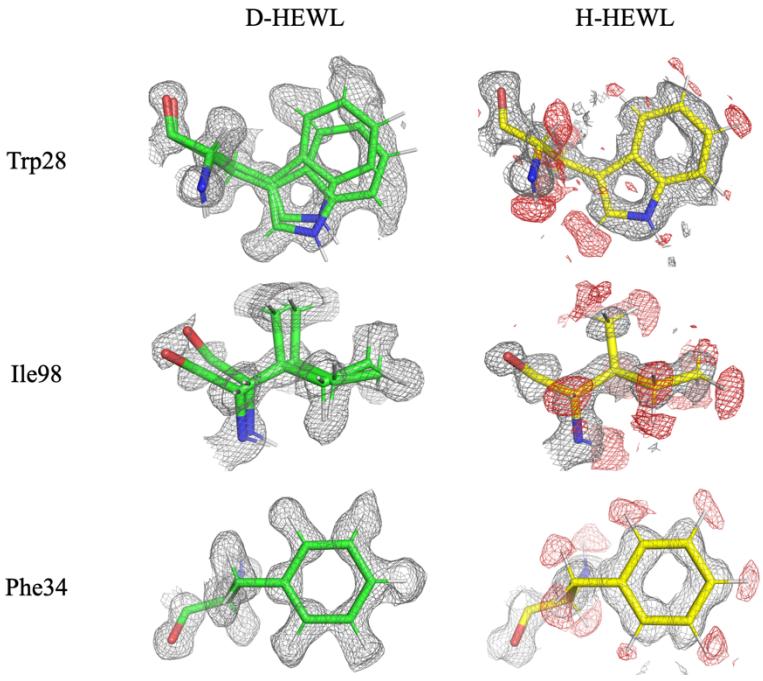


Figure 4

Impact of perdeuteration on the quality of neutron density maps. Protein residue disorder becomes evident in D-HEWL neutron maps, while these details are precluded in H-HEWL due to the negative neutron scattering contribution of non-exchangeable H atoms, as seen for Trp28 and Ile98. Phe34 is an example of an ordered residue in both models. The positive 2*Fo-Fc* neutron scattering length density were contoured at 1 σ (grey) and the negative peaks are depicted with an additional contour at -1.5 σ (red).

Effects of data resolution and completeness on neutron maps

At resolutions better than ~2.5 Å, D atoms become readily visible in neutron maps. This resolution threshold is considered in efforts to study protein structure through NMX (Sørensen *et al.*, 2018), since the main advantage of the technique is the ability of locating H (and D) positions and subsequently resolve protonation states, H-bond networks and solvent interactions in macromolecules. Moreover, according to Liebschner *et al.* (2018), the average diffraction data resolution for neutron structures deposited in the PDB is also around 2 Å. Chen *et al.* (2012) have demonstrated the effects of resolution, from 2.5 Å to 1.1 Å, in the visibility of atoms in neutron maps. However, in this case observations were made from incomplete neutron data to 1.1 Å resolution, from a D_2O buffer exchanged crystal of hydrogenated crambin. Therefore, both the lack of data completeness and the cancellation effects arising from the neutron scattering of H atoms may have an impact in the neutron maps calculated. The complete neutron datasets obtained for D-HEWL and H-HEWL allow a direct comparison between neutron maps of perdeuterated and hydrogenated samples up to 1 Å resolution (Fig. 5). It is clear that resolutions better than 2.5 Å are required to locate D atoms in well-ordered regions, such as the case of Tyr53. At atomic resolution, neutron density for individual atoms is observed, reflecting anisotropic thermal motion, for both perdeuterated and hydrogenated molecules.

In the case of the D-HEWL neutron maps, the atomic anisotropic behavior can also be observed for D atoms. Additionally, it seems that the negative neutron scattering of neighboring H atoms has no impact in the neutron densities of other atoms in well-ordered residues, as previously noted by Chen *et al.* (2012). However, as mentioned above, cancellation effects might impact neutron maps at high resolution by precluding protein residue disorder. Also, at high resolutions (*i.e.*, < 2 Å), H atoms become visible in the negative 2*Fo-Fc* neutron map, providing further insight into potential disorder, not displayed in the positive neutron map.

Liebschner *et al.* (2018) showed that the average neutron diffraction data completeness for structures available in the PDB is approximately 80%. This is a consequence of the relatively low neutron flux used in NMX experiments, which lead to considerably longer exposure times per diffraction frame and often hinder the complete coverage of reciprocal space (Blakeley, 2009). For this reason, a strategy of model refinement has been increasingly adopted in NMX studies, where a joint refinement against isomorphous neutron and X-ray data is performed (Adams *et al.*, 2009; Afonine *et al.*, 2010). In this way, the data-to-parameter ratio is significantly increased and inaccuracies arising from a refinement against incomplete data are mitigated. Nevertheless, there are currently 53 neutron structures in the PDB (out of a total 177) obtained exclusively from incomplete neutron diffraction data (*i.e.*, overall completeness < 90%). To investigate the effects of model refinement against incomplete data, the amount of neutron data for D-HEWL was reduced by including only the first 14 ω scans recorded, from the total 29 scans (resulting in 84% of overall completeness), simulating an experiment with less allocated time. The same reflections were marked for the R_{free} calculations and the last refinement round in SHELXL (Table 2) was performed. The corresponding data merging and refinement statistics are presented in Supplementary Table S3.

Surprisingly, no obvious differences were found in an inspection of the 2*Fo-Fc* and *Fo-Fc* neutron maps for both complete and partial datasets. Neutron densities continued to show residue disorder and the most intense peaks in the *Fo-Fc* map did not suggest over or under modelling. Minor variations in the refined occupancies of disordered residues and partial caged H were observed (< 4%), indicating that the lack of data completeness results in a decrease of model accuracy. However, the r.m.s.d. from ideal values for bond lengths, angles and other restraints, did not show significant variations. The data-to-parameter ratio decreased by 0.4, to 5.1, thus remaining acceptable to perform the refinement of anisotropic ADPs. However, a comparison of the refined isotropic ADPs for the Cα of ordered protein residues (*i.e.*, single conformation residues, excluding also N-terminus Gly0 and C-terminus Arg128 and Leu129) suggests that these are slightly overestimated when using the incomplete dataset, with an average difference of 0.2 $Å^2$ to the refined ADPs for D-HEWL (Supplementary Figure S5). This discrepancy extends to the full model as reflected by the overall average B-factors of 19.62 $Å^2$ and 19.50 $Å^2$ for the incomplete and complete datasets respectively. These results emphasize the importance of obtaining complete diffraction data in order to optimize model accuracy. In particular,

anisotropic ADP refinement is not recommended based on incomplete data despite the reasonable data-to-parameter ratio. Nevertheless, these results also highlight the fact that incomplete data (~80%) can still provide other important insights to protein structures.

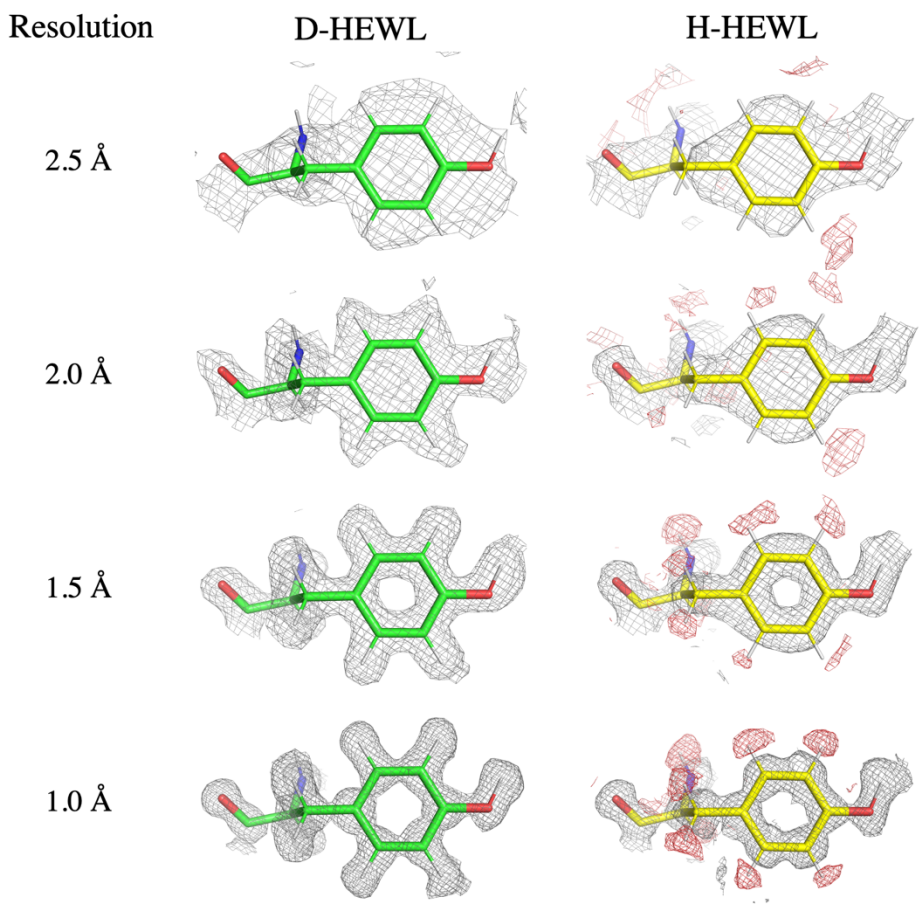


Figure 5

The effects of neutron diffraction data resolution in neutron density maps, using Tyr53 as example. The 2*Fo-Fc* neutron density maps presented were calculated in SHELXL using the final phases and models of D-HEWL and H-HEWL at 1.00 Å and 1.07 Å resolution, respectively. The 2*Fo-Fc* neutron density maps were contoured at 1.0 σ (grey), and -1.5 σ in the case of the negative density map for H-HEWL (red).

Conclusions

To our knowledge, this work constitutes the first report of complete atomic resolution neutron diffraction data for a protein crystal. The high quality neutron data enabled the unambiguous assignment of H/D positions which provided valuable insight and clarity of the HEWL active site. This work should contribute to more accurate HEWL models obtained from X-ray diffraction data and facilitate the interpretation of structural data focusing on the enzymatic mechanism of HEWL.

Partial deuteration through soaking in D_2O remains the most accessible approach to NMX and, if high quality data is obtained, it permits the determination of protonation states and details of H-bond networks involving the most well-ordered water molecules. However, the neutron data recorded is still influenced by the negative neutron scattering from the non-exchangeable H atoms and cancellation effects will mask *e.g.,* disorder even at atomic resolution. The detailed comparison between neutron data obtained from hydrogenated D_2O soaked and perdeuterated HEWL crystals proved that perdeuteration was crucial to the recording of complete neutron data to 1 Å resolution, even with a smaller crystal volume of the perdeuterated sample. Furthermore, unique structural and dynamical detail was observed for D-HEWL, while absent for H-HEWL, due to the complete substitution of H for D in non-exchangeable positions.

Finally, the effects of resolution and completeness of neutron diffraction data in the resulting neutron maps were investigated, completing the picture of the information available to NMX studies. Once again, perdeuteration proved to be advantageous by enabling the refinement of coordinates and anisotropic ADPs for D atoms at atomic resolutions. Completeness was shown to play a role in the accuracy of the refined neutron models, with discrepancies observed in the refined residue and caged H occupancies, and anisotropic ADPs, between complete and incomplete datasets.

Methods

Protein materials

D-HEWL was recombinantly produced as previously described (Ramos *et al.*, 2021). Briefly, the protein was expressed in *E. coli* BL21(DE3) cells in the form of inclusion bodies using fully deuterated medium. D-HEWL was purified and subsequently refolded *in vitro*. H-HEWL was purchased as lyophilized powder (Sigma-Aldrich, L6876).

Protein crystallization

H-HEWL crystals in the P1 form were initially obtained by a cooling step as previously described (Vidal *et al.*, 1999). Crystallization was initially performed in hanging drops, in batch-like conditions with 2 % $NaNO_3$, 0.1 M Na acetate, pH 4.5 and 10 mg/mL H-HEWL. Crystals were harvested after a week, and single crystals were transferred to 9-well glass plates, kept in closed containers with surrounding matching buffer solution. $NaNO_3$ was lowered to 1.8% to reduce further nucleation. To ensure continued crystal growth, the solution in the drops was replaced with fresh protein solution every other week for several months. Secondary nucleation was removed to the extent possible if present. During the week before neutron data collection, the surrounding solution was replaced by a fully D_2O based version of the buffer and the protein solution was replaced by buffer with gradually increased D_2O content. Finally, the crystal was equilibrated against 100% D_2O buffer in the capillary. Triclinic D-HEWL crystals were obtained by microseeding, using seeds from D_2O buffer exchanged H-HEWL. These D-HEWL crystals were then used to produce seeds for new batches of crystals in 5 µL drops, including 2.5 µL of protein at 20 mg/mL and 2.5 µL of 0.28 M sodium nitrate and 0.05 M Na acetate pD 4.5, in microbatch under oil at 18° C (Ramos *et al.*, 2021). In order to grow the crystals to volumes appropriate for neutron diffraction experiments, these were macroseeded in 20-60 µL drops, in batch-like conditions at 18° C, using 9-well glass plates and sandwich box setup from Hampton Research. After every 1-2 weeks in which the drops were allowed to equilibrate, the solution around the crystals was replaced by new solutions of protein and precipitant.

Neutron diffraction data collection and processing

Neutron diffraction data was collected at room temperature at the D19 instrument, at Institute Laue-Langevin (Grenoble, France), from triclinic D-HEWL and H-HEWL crystals of approximately 4 and 6 mm^3, respectively (Supplementary Fig. S6). The crystals were mounted in quartz capillaries containing crystallization solution in D_2O to prevent dehydration. Quartz wool was used to fix the crystals' position inside the capillaries, and the capillaries were sealed with epoxy glue. Data was collected from the D-HEWL crystal in 29 ω scans, in static frames of 0.1° steps and exposure times of ~60 s per frame. For the H-HEWL crystal, data was also collected in 29 ω scans, in static frames of 0.07° steps and exposure of ~40-60 s. The data were reduced with XDS (Kabsch, 2010) and scaled

using XSCALE (Kabsch, 2010). The program XDSCC12 (Assmann *et al.*, 2020), designed to optimize data processing in multi-dataset crystallography, was employed to determine the correlation coefficients of the overall data with and without each ω scans. All 29 ω scans were used for model refinement. The data was converted to hkl format, suitable for SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) and 5% of all reflections were randomly flagged for the R_{free} calculations using XDSCONV (Kabsch, 2010). The resolution cutoff was initially evaluated based on completeness, R-factors, I/σ(I), and $CC_{1/2}$, but ultimately fixed based on paired refinement as suggested by Karplus & Diederichs (2012) and Maly *et al.* (2020). This was done by following the changes in R-factors, more specifically in the difference between R_{free} and R_{work}, and in the refinement goodness-of-fit (GooF) as a function of increasing data resolution (Supplementary Fig. S1).

Model refinement

Model refinement was performed using SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) starting from an initial model based on PDB entry 7ave (Ramos *et al.*, 2021), after removing alternate residue conformations, water molecules and ions. Geometric restraints (*i.e.*, bond lengths, bond angles, and chiral volumes) for deuterated amino acids were applied as described by Gruene *et al.* (2014). Since H/D atoms represent ~50% of all atoms in the protein and the contribution of D atoms to the neutron coherent scattering is of similar magnitude as C, N, O, these were included in the initial stages of model refinement. The refinement strategy (Table 2 and Supplementary Table S1) generally consisted of gradual increases in the number of parameters, associated with the inclusion of more data. After each round of refinement, the neutron maps were inspected visually using COOT (Emsley *et al.*, 2010). The majority of model building and refinement actions were performed using data to 1.13 Å resolution. After the refinements converged, paired refinements were conducted using progressively higher resolution limits to evaluate if the additional weaker diffraction data could improve the accuracy of the refined models (Supplementary Fig. S1). Occupancies for protein residues, water molecules and ions were initially fixed at 1, and when disorder was modelled the sum of alternate conformations was set to 1. Idealized positions were applied to H and initially also to D atoms. Later D atoms positions, isotropic and anisotropic ADPs were refined, while maintaining H in idealized positions and refining isotropic ADPs. SIMU (0.1) and XNPD (0.001) restraints on isotropic and anisotropic ADPs were kept throughout the refinements.

In the neutron models of D-HEWL and H-HEWL, anisotropic ADPs were refined for all protein atoms, except for H atoms and for the side-chains of residues without visible neutron density at 1 σ. These exceptions were Gly0 in D-HEWL, Lys13 in H-HEWL, and Asn106, Arg112, Asp119, Gln121, Arg128, and Leu129 in both models. D_2O molecules were also refined with anisotropic ADPs in both models, if the isotropic B-factor for each of the three atoms was inferior to 30 $Å^2$. This corresponded to

17 and 19 D_2O molecules in D-HEWL and H-HEWL, respectively. The data merging and model refinement statistics are presented in Table 1.

To evaluate the impact of diffraction data resolution in neutron maps, the final models for D-HEWL and H-HEWL were used, and Fourier maps were recalculated at each resolution limit presented by performing a 0-cycle refinement in SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015). In the test of diffraction data completeness, the first 14 scans of neutron data for D-HEWL were used and the flags for R_{free} calculations were copied from the complete dataset. The final refinement in SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) was rerun against the incomplete data. The data merging and model refinement statistics are available in Supplementary Table S2.

References

Adams, P. D., Mustyakimov, M., Afonine, P. V. & Langan, P. (2009). *Acta Crystallogr.* **D65**, 567–573.

Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P. & Adams, P. D. (2010). *Acta Crystallogr.* **D66**, 1153–1163.

Aggarwal, M., Kovalevsky, A. Y., Velazquez, H., Fisher, S. Z., Smith, J. C. & McKenna, R. (2016). *IUCrJ*. **3**, 319–325.

Assmann, G. M., Wang, M. & Diederichs, K. (2020). *Acta Crystallogr.* **D76**, 636–652.

Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.

Blakeley, M. P. (2009). *Crystallography Reviews*, Vol. pp. 157–218.

Blakeley, M. P., Langan, P., Niimura, N. & Podjarny, A. (2008). *Curr Opin Struct Biol*. **18**, 593–600.

Chen, J. C. H., Hanson, B. L., Fisher, S. Z., Langan, P. & Kovalevsky, A. Y. (2012). *PNAS*. **109**, 15301–15306.

Cuypers, M. G., Mason, S. A., Blakeley, M. P., Mitchell, E. P., Haertlein, M. & Forsyth, V. T. (2013). *Angew. Chemie - Int. Ed.* **52**, 1022–1025.

Cuypers, M. G., Mason, S. A., Mossou, E., Haertlein, M., Forsyth, V. T. & Mitchell, E. P. (2016). *Sci. Rep.* **6:31487**,.

Dauter, Z., Jaskolski, M. & Wlodawer, A. (2010). *J. Synchrotron Radiat.* **17**, 433–444.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr.* **D66**, 486–501.

Fisher, S. J., Blakeley, M. P., Howard, E. I., Petit-Haertlein, I., Haertlein, M., Mitschler, A., Cousido-Siah, A., Salvay, A. G., Popov, A., Muller-Dieckmann, C., Petrova, T. & Podjarny, A. (2014). *Acta Crystallogr.* **D70**, 3266–3272.

Fisher, S. Z., Aggarwal, M., Kovalevsky, A. Y., Silverman, D. N. & McKenna, R. (2012). *J. Am. Chem. Soc.* **134**, 14726–14729.

Gerlits, O., Wymore, T., Das, A., Shen, C.-H., Parks, J. M., Smith, J. C., Weiss, K. L., Keen, D. A., Blakeley, M. P., Louis, J. M., Langan, P., Weber, I. T. & Kovalevsky, A. (2016). *Angew. Chemie*. **128**, 5008–5011.

Glasoe, P. K. & Long, F. A. (1960). *J. Phys. Chem.* **64**, 188–190.

Gruene, T., Hahn, H. W., Luebben, A. V., Meilleur, F. & Sheldrick, G. M. (2014). *J. Appl. Crystallogr.* **47**, 462–466.

Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O. & Forsyth, V. T. (2016). *Methods in Enzymology*, Vol. *566*, pp. 113–157.

Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Crystallogr.* **D61**, 1413–1417.

Howard, E. I., Guillot, B., Blakeley, M. P., Haertlein, M., Moulin, M., Mitschler, A., Fadel, F., Valsecchi, W. M., Tomizaki, T., Petrova, T., Claudot, J. & Podjarny, A. (2016). *IUCrJ*. **3**, 115–126.

Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986). *Biochemistry*. **25**, 6965–6972.

Kabsch, W. (2010). *Acta Crystallogr.* **D66**, 133–144.

Karplus, P. A. & Diederichs, K. (2012). *Science (80-.).* **336**, 1030–1033.

Koetzle, T. F. & McIntyre, G. J. (2002). *Characterization of Materials*, Vol. pp. 2192–2205.

Kuwajima, K. (1989). *Proteins Struct. Funct. Bioinforma.* **6**, 87–103.

Langan, P., Sangha, A. K., Wymore, T., Parks, J. M., Yang, Z. K., Hanson, B. L., Fisher, Z., Mason, S. A., Blakeley, M. P., Forsyth, V. T., Glusker, J. P., Carrell, H. L., Smith, J. C., Keen, D. A., Graham, D. E. & Kovalevsky, A. (2014). *Structure*. **22**, 1287–1300.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Langan, P. & Adams, P. D. (2018). *Acta Crystallogr.* **D74**, 800–813.

Maly, M., Diederichs, K., Dohnalek, J. & Kolenko, P. (2020). *IUCrJ*. **7**, 681–692.

Miranker, A., Radford, S. E., Karplus, M. & Dobson, C. M. (1991). *Nature*. **349**, 633–636.

Munshi, P., Madsen, A. Ø., Spackman, M. A., Larsen, S. & Destro, R. (2008). *Acta Crystallogr.* **A64**, 465–475.

Myles, D. A. A., Dauvergne, F., Blakeley, M. P. & Meilleur, F. (2012). *J. Appl. Crystallogr.* **45**, 686–692.

Niimura, N. & Bau, R. (2008). *Acta Crystallogr.* **A64**, 12–22.

Oksanen, E., Blakeley, M. P., Bonneté, F., Dauvergne, M. T., Dauvergne, F. & Budayova-Spano, M. (2009). *J. R. Soc.*

Interface. **6**, S599--S610.

Oksanen, E., Dauvergne, F., Goldman, A. & Budayova-Spano, M. (2010). *J. Appl. Crystallogr.* **43**, 1113–1120.

Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). *Nature*. **358**, 302–307.

Romoli, F., Mossou, E., Cuypers, M., Van Der Linden, P., Carpentier, P., Mason, S. A., Forsyth, V. T. & McSweeney, S. (2014). *Acta Crystallogr.* **F70**, 681–684.

Schiebel, J., Gaspari, R., Wulsdorf, T., Ngo, K., Sohn, C., Schrader, T. E., Cavalli, A., Ostermann, A., Heine, A. & Klebe, G. (2018). *Nat. Commun.* **9:3559**,.

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

Sheldrick, G. M. (2015). *Acta Crystallogr.* **C71**, 3–8.

Shu, F., Ramakrishnan, V. & Schoenborn, B. P. (2000). *PNAS*. **97**, 3872–3877.

Sørensen, T. L.-M., Hjorth-Jensen, S. J., Oksanen, E., Andersen, J. L., Olesen, C., Møllerd, J. V. & Nissen, P. (2018). *Acta Crystallogr.* **D74**, 1208–1218.

Vidal, O., Robert, M. C., Arnoux, B. & Capelle, B. (1999). *J. Cryst. Growth*. **196**, 559–571.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Crystallogr.* **D54**, 522–546.

Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Crystallogr.* **D63**, 1254–1268.

Weber, I. T., Waltman, M. J., Mustyakimov, M., Blakeley, M. P., Keen, D. A., Ghosh, A. K., Langan, P. & Kovalevsky, A. Y. (2013). *J. Med. Chem.* **56**, 5631–5635.

Wildegger, G. & Kiefhaber, T. (1997). *J. Mol. Biol.* **270**, 294–304.

Supporting Information

A complete picture of hen egg-white lysozyme active site at room temperature from 1 Å resolution neutron diffraction data

Joao Ramos[1,2,3], Estelle Mossou[1,4], Valerie Laux[1,2], Sax A. Mason[1], Marie-Helene Lemee[1], Kay Diederichs[5], Michael Haertlein[1,2], V. Trevor Forsyth[1,2,4], Sine Larsen[6], Annette E. Langkilde[3]*

[1]Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.

[2]Partnership for Structural Biology (PSB), 71 avenue des Martyrs, 38000 Grenoble, France.

[3]Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.

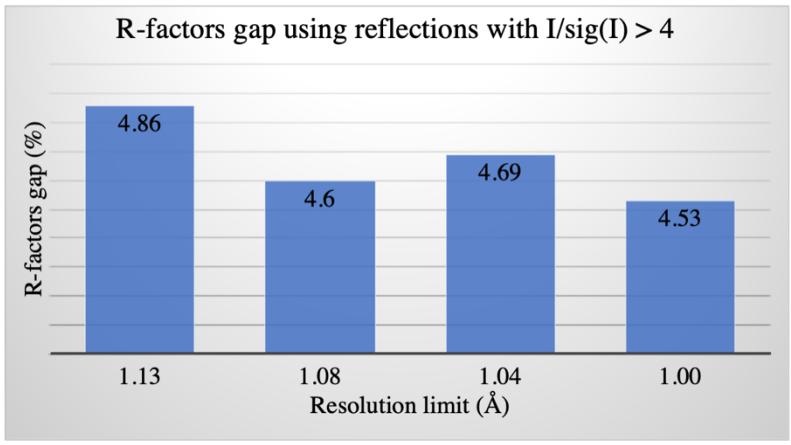
[4]Faculty of Natural Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.

[5]Department of Biology, University of Konstanz, Box 647, D-78457 Konstanz, Germany.

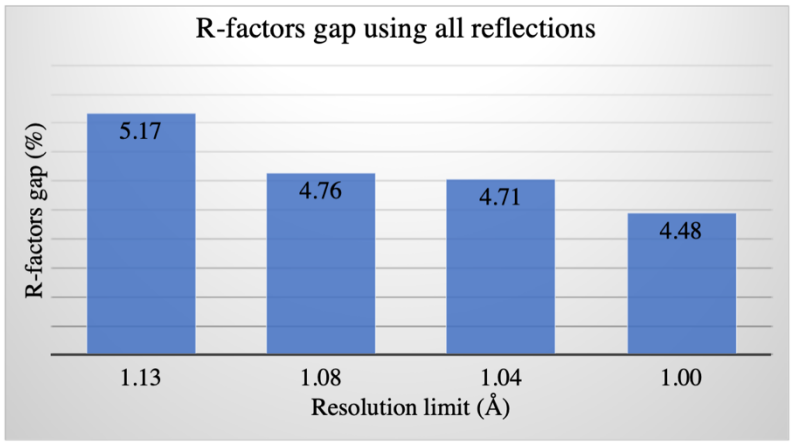
[6]Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.

*corresponding author, annette.langkilde@sund.ku.dk

a



b



c

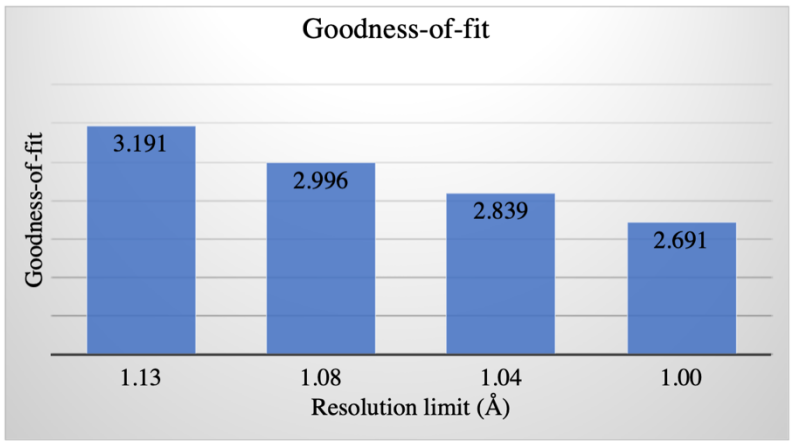


Figure S1

Testing the validity of adding weaker diffraction data to the model refinement. (a) Difference observed between the R_{free} and R_{work} for the stronger reflections, I/σ(I) > 4, with different data resolution limits. (b) Changes in the difference between R_{free} and R_{work} for all dataset reflections, through the use of different data resolution cut-offs. (c) Variations in the SHELXL model refinement score, goodness-of-fit, by applying different data resolution limits.

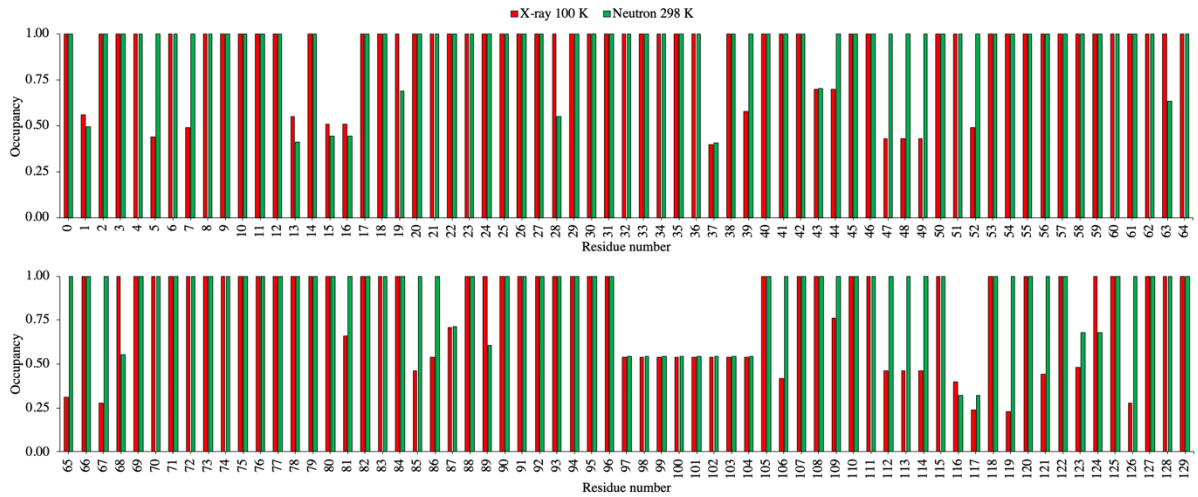


Figure S2

Protein residue disorder is visible in neutron maps of D-HEWL. Differences in the modelled disorder for D-HEWL, trough variations in the refined occupancies of residues conformation A, from X-ray data at 100 K (PDB entry 7ave) (red bars) and neutron data at 298 K (green bars).

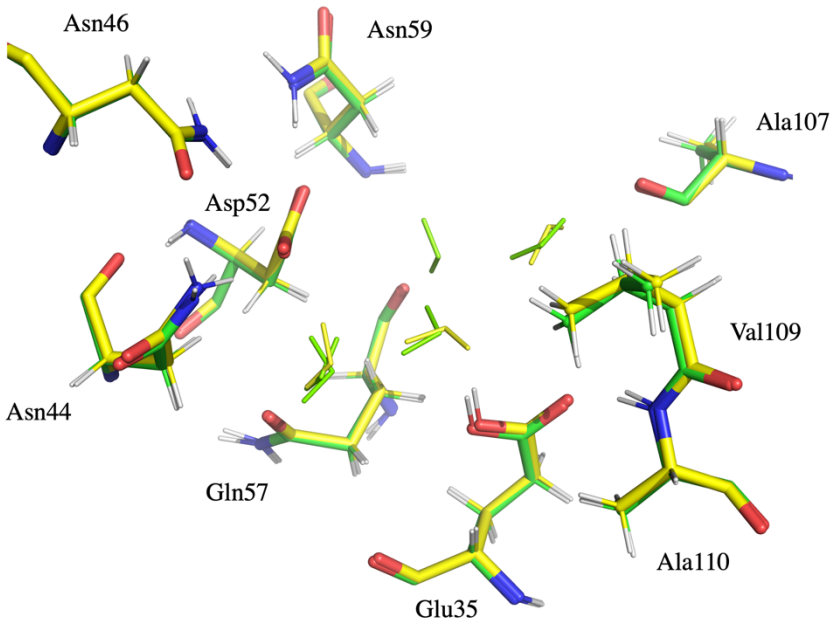


Figure S3

The configuration of the active sites of D-HEWL and H-HEWL is identical, with the only exception being the orientation of the catalytic water hydrogen-bonded to Glu35. The carbon atoms and water molecules of the D-HEWL model are colored in green, while in the H-HEWL model they are in yellow.

Table S1

Model refinement strategy used in SHELXL against neutron diffraction data for H-HEWL. The changes in R-factors for reflections with I/σ(I) > 4, goodness-of-fit (GooF), and the data-to-parameter ratio are presented for each step. *The last round of refinement yielded higher R-factors compared to round 11, since several instances of model overfitting were corrected, thus reducing the gap between R_{free} and R_{work} for all reflections, from 5.0 to 4.8.

Round	Refinement step	Resolution (Å)	R_{work} (%)	R_{free} (%)	GooF	Data-to-parameter ratio
1	Rigid body	22.84 – 2.0	45.61	46.94	8.84	10.3
2	Coordinates and isotropic ADPs	22.84 – 1.5	32.07	37.57	5.82	4.7
3	Adding H atoms in idealized positions	22.84 – 1.5	31.06	35.06	5.88	4.9
4	Adding D atoms in exchangeable positions	22.84 – 1.5	33.00	36.90	6.20	5.4
5	Adding D_2O and ions	22.84 – 1.13	28.91	32.06	3.80	9.1
6	Adding caged H atoms	22.84 – 1.13	21.35	23.93	2.91	9.1
7	Modelling protein residue disorder	22.84 – 1.13	19.59	21.98	2.71	8.7
8	Refining D atoms positions	22.84 – 1.13	18.46	20.69	2.59	7.5
9	Anisotropic ADPs for non-H/D atoms	22.84 – 1.07	14.92	19.02	2.15	5.4
10	Anisotropic ADPs for D atoms	22.84 – 1.07	14.68	19.08	2.16	5.1
11	Anisotropic ADPs for well-ordered waters	22.84 – 1.07	13.88	18.51	2.06	5.0
12*	Minor corrections	22.84 – 1.07	14.13	18.82	2.07	5.1

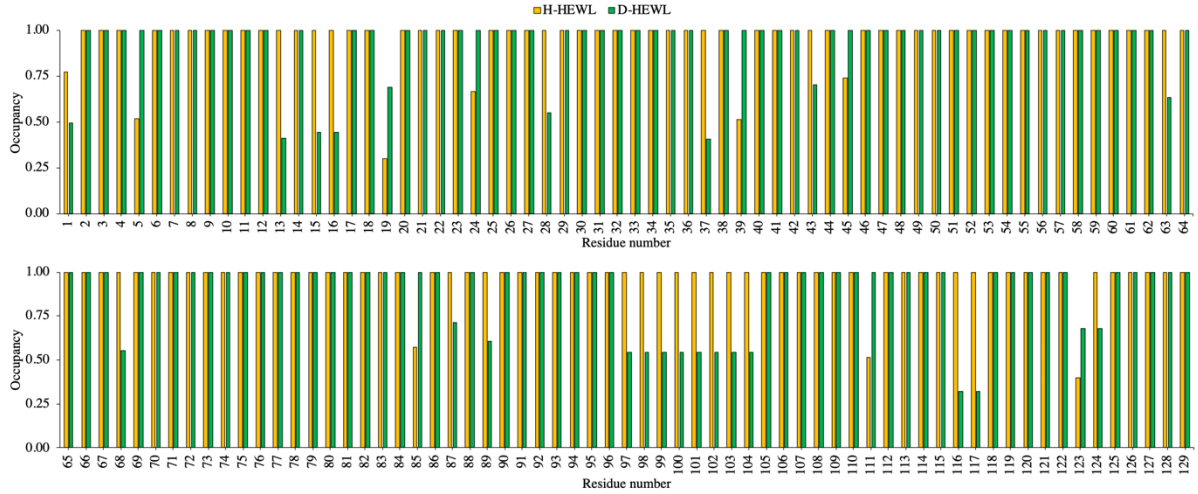


Figure S4

Protein perdeuteration has a significant impact in the observation of protein residue disorder, even at atomic resolutions. Refined occupancies of alternate conformation A by residue number in the 298 K neutron models of H-HEWL (yellow bars) and D-HEWL (green bars).

Table S2

Refined D occupancies in backbone N-H groups compared to all HEWL neutron structures published to date.

Residue number	1lzn	6k8g	D-HEWL	Mason *et al.* (1984)
	D_{occ}= 1 or D_{occ}= 0	$D_{occ}+H_{occ}$= 1	$D_{occ}+H_{occ}$= 1	D_{occ} > 0.5 based on Fig. 1
2	1	0.74	1	x
3	0	0.94	1	
4	1	0.92	1	
5	1	0.84	1	x
6	1	0.75	1	x
7	1	1	1	x
8	0	0.43	0	
9	0	0.52	0	
10	0	0	0	
11	0	0.64	0	
12	0	0	0	
13	0	0	0	
14	1	1	1	x
15	1	0.86	1	x
16	1	0.97	1	x
17	1	0.82	1	x
18	1	0.56	1	x
19	1	0.65	1	x
20	1	0.72	1	x
21	1	0.31	1	x
22	1	0.91	1	x
23	0	1	0.40	
24	1	0.56	1	x
25	1	1	1	x
26	1	0.79	1	x
27	1	0.35	0.20	
28	0	0	0	
29	0	0.36	0	
30	0	0	0	
31	0	0	0	
32	0	0	0	
33	0	0	0.33	
34	0	0.17	0	
35	1	0.67	1	x
36	1	0.99	1	x
37	1	1	0	x
38	0	0.09	0	
39	0	0.17	0	
40	1	0.72	0.49	
41	1	1	1	x
42	0	1	0.44	
43	1	0.86	1	x
44	0	1	0	
45	1	0.74	1	x
46	1	0.77	1	x
47	1	0.62	1	x
48	1	0.62	1	x
49	1	0.76	1	x
50	1	0.87	1	x
51	1	0.53	0.82	x
52	0	0	0	
53	0	0.1	0.12	
54	0	0	0	
55	1	0.83	0.74	x
56	1	0.94	0.54	x
57	0	0.21	0.07	
58	0	0	0	
59	1	0.6	0.23	x
60	0	0.11	0	
61	0	0.71	0.23	
62	1	1	1	x
63	0	1	0	
64	0	0.07	0	
65	0	0.4	0	
66	1	0.87	1	x
67	1	0.65	1	x
68	1	0.55	1	x

69	1	0.42	1	x
71	1	1	1	x
72	1	0.72	1	x
73	1	0.82	1	x
74	1	0.6	1	x
75	0	1	0.30	
76	0	0.74	0	
77	1	1	1	x
78	1	1	0.06	
80	1	1	0.26	x
81	1	0.85	1	x
82	1	0.92	0.50	
83	0	0.12	0	
84	0	0.65	0.44	
85	1	0.99	1	x
86	1	0.59	1	x
87	1	0.84	1	x
88	1	0.97	1	x
89	1	0.7	1	x
90	1	0.44	1	x
91	1	0.98	1	x
92	0	0	0	
93	1	0.87	0.36	x
94	1	0.86	0.42	
95	0	0.06	0	
96	0	0	0	
97	0	0.58	0	
98	0	0	0	
99	0	0.18	0	
100	1	0.7	1	
101	1	0.64	1	
102	1	0.53	1	
103	1	1	1	
104	1	0.56	1	
105	1	0.66	1	x
106	1	0.85	1	x
107	1	0.95	1	x
108	1	0.79	0.47	x
109	1	0.69	1	x
110	1	0.89	1	x
111	1	1	0.33	x
112	1	0.94	0	x
113	1	0.83	1	
114	0	1	1	
115	1	1	1	x
116	1	0.71	0	x
117	1	0.74	1	x
118	1	0.6	1	
119	1	0.9	1	x
120	1	0.58	1	
121	1	0.69	1	x
122	1	0.93	1	
123	1	1	0	
124	1	0.99	0	
125	1	0.58	1	x
126	1	0.86	1	x
127	1	1	1	x
128	1	1	1	x
129	1	0.6	1	x

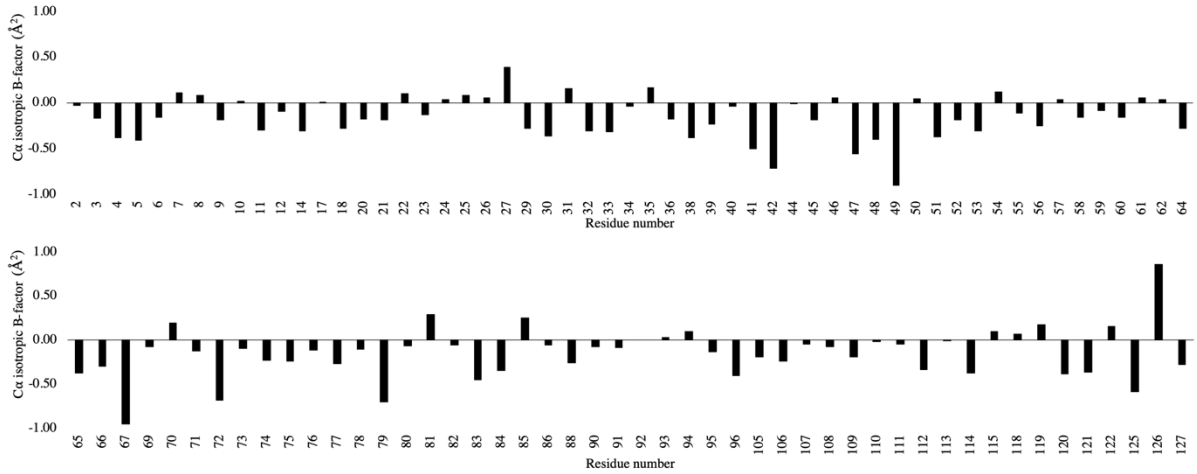


Figure S5

Lack of neutron diffraction data completeness hinders model accuracy. Comparison of refined C_α isotropic ADPs for ordered protein residues of D-HEWL using incomplete and complete neutron datasets. The bars correspond to the difference between the refined ADPs of complete and incomplete datasets ($\Delta U = U_{complete} - U_{incomplete}$).

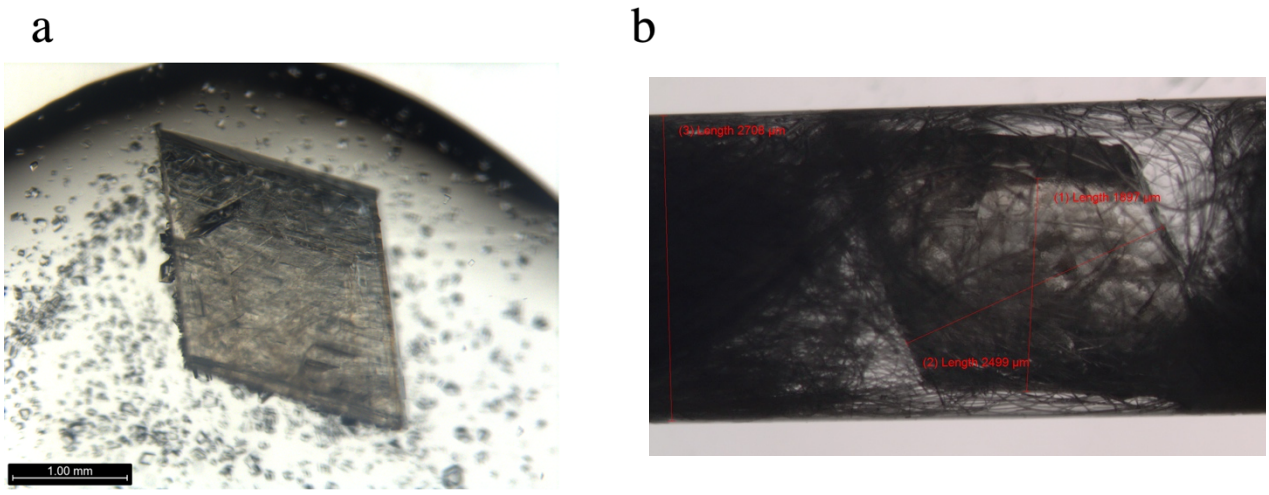


Figure S6

Crystals of D-HEWL (a) and H-HEWL (b) measured in the room temperature neutron diffraction experiments in D19, at ILL.

Table S3

Data merging and model refinement statistics from incomplete D-HEWL neutron data. The values in parenthesis correspond to the outer resolution shell.

	D-HEWL
Temperature (K)	298
Source	D19, ILL, FR
Detector	$^3He + CF_4$
Wavelength (Å)	1.455
Resolution range (Å)	21.44 – 1.00 (1.04 – 1.00)
Space group	P 1
Unit cell **a, b, c (Å)** **α, β, γ (°)**	27.1, 31.59, 34.1 89.156, 72.027, 67.751
Total reflections	89077 (6104)
Unique reflections	44895 (3938)
Multiplicity	2.0 (1.6)
Completeness (%)	84.0 (73.4)
Mean I/σ(I)	6.1 (1.2)
R_{merge}	0.069 (0.528)
R_{meas}	0.086 (0.707)
R_{pim}	0.049 (0.466)
$CC_{1/2}$	0.995 (0.517)
Refl. used in refinement with *Fo*/σ(*Fo*) > 4 / all refl.	27318 / 42662
Refl. used for R-free with *Fo*/σ(*Fo*) > 4 / all refl.	1687 / 2245
R_{work} **for refl. with *Fo*/σ(*Fo*) > 4 / all refl.**	11.75 / 17.09
R_{free} **for refl. with *Fo*/σ(*Fo*) > 4 / all refl.**	16.87 / 22.28
Number of non-hydrogen atoms (overall)	1288
macromolecules	1203
ligands	24
solvent	58
Protein residues	130
R.m.s.d., bond lengths (Å)	0.0112
R.m.s.d., angles (°)	2.46
Ramachandran favored (%)	99.22
Ramachandran allowed (%)	0.78
Ramachandran outliers (%)	0
Rotamer outliers (%)	2.40
Clashscore	9.98
Average B-factor (overall) (Å^2)	19.62
macromolecule	19.21
ligands	30.1
solvent	23.91

Chapter 6

Manuscript 4

Synopsis

In small-molecule crystallography, discrepancies are often observed between ADPs derived from X-ray and neutron data, when atomic models are used. However, for protein crystals, this type of studies is still lacking due to the challenges inherent to the acquisition of complete atomic resolution neutron diffraction data. Additionally, measurements performed at different temperatures are required to separate atomic thermal motion from static and dynamic disorder in a crystal. In this chapter, a manuscript draft is presented (Manuscript 4), which attempts to address the current limited knowledge in regard to atomic thermal motion in protein crystals, and the potential bias of ADPs in X-ray models.

Unfortunately, due to time constraints, the work presented in this manuscript remains unfinished. The main pending task concerns the analysis of the 100 K neutron diffraction data, which seems to be affected by the presence of ice rings. Ice ring diffraction patterns could be observed during data collection at D19, and these might be biasing the refinement of the model and its ADPs. From the differences in the model R-factors (> 5%, between R_{work} and R_{free}), it is apparent that the model has bias, and especially from the discrepancy between the R-factors for the strong reflections, $Fo/\sigma(Fo) > 4$, and for all data, it seems that the lower angle reflections are problematic. Additionally, the refined mean ADPs for the 100 K neutron model are closer to the ones observed for both the X-ray and neutron models at room temperature. It would be expected that the mean ADPs at 100 K would be lower and comparable to those obtained for the X-ray data at 100 K. Several errors can impact the estimation of ADPs and introduce bias during model refinement:

i) Inaccurate estimation of reflections' intensities (*e.g.*, from the presence of ice rings, or from crystal twinning).
ii) Inappropriate absorption correction.
iii) Unsuitable bulk solvent correction.
iv) Errors in the modelling of disorder.

As previously mentioned, we suspect that the presence of ice rings is leading to the overestimation of the intensities of lower angle reflections during data reduction. The absorption correction is usually non-problematic from data processing with XDS (Kabsch, 2010), and since the crystal was perdeuterated, this minimizes the effects of incoherent neutron scattering from H atoms. The bulk solvent correction performed by SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) is less sophisticated than the one used *e.g.*, in PHENIX (Liebschner *et al.*, 2019), however, several neutron and X-ray datasets at both room temperature and 100 K have been refined in SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) and similar observations regarding the R-factors have not been made. Finally, the errors in modelling have been addressed by performing refinements from the initial model used for all different X-ray and neutron datasets, without alternate conformations, water molecules or ions. Similar disorder to that modelled previously has been observed in the resulting 2*Fo*-*Fc* neutron maps and

identical differences between the R-factors have also been identified. A test was conducted where the Wilson-plot outliers (*i.e.*, reflections with abnormally large I/σ(I) compared to other reflections in the same resolution shell) identified during data reduction in XDS were removed from the final dataset. A few refinement rounds were performed and the resulting gaps between R_{work} and R_{free} were significantly lower, of ~ 5%. Importantly, the discrepancy between the R-factors for strong reflections and all data was significantly reduced. Although these results are indicative that the problem resides in the overestimation of a set of reflections (likely due to the presence of ice rings), further analysis of the neutron maps and model quality are required. For this reason, the 100 K neutron model is omitted in the following manuscript and the analysis is focused on the neutron and X-ray data at room temperature.

Protein dynamics at the atomic level: comparing atomic displacement parameters from X-ray and neutron diffraction

Joao Ramos[1,2,3], Valerie Laux[1,2], Marie-Helene Lemee[4], Matthew W. Bowler[5], Kay Diederichs[6], Michael Haertlein[1,2], V. Trevor Forsyth[1,2,7], Estelle Mossou[1,2,7], Sine Larsen[8], Annette E. Langkilde[3]*

[1]*Life Sciences Group, Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.*
[2]*Partnership for Structural Biology (PSB), 71 avenue des Martyrs, 38000 Grenoble, France.*
[3]*Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.*
[4]*Diffraction Group, Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.*
[5]*European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, 38000 Grenoble, France.*
[6]*Department of Biology, University of Konstanz, Box 647, D-78457 Konstanz, Germany.*
[7]*Faculty of Natural Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.*
[8]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.*
*corresponding author, annette.langkilde@sund.ku.dk

Abstract

Protein biological function is intimately related to dynamics at both the molecular and the atomic scales. The estimation of anisotropic atomic displacement parameters (ADPs) is commonly performed for atomic resolution X-ray diffraction data, aiming at describing atomic motion in a crystal structure. In this study, the analysis of anisotropic ADPs for protein crystals is performed through complete atomic resolution neutron and X-ray data from perdeuterated hen egg-white lysozyme, at both room temperature and 100 K. It is shown that the ADPs derived from neutron data are systematically smaller than those from X-ray data. The ADPs obtained from X-rays appear to be more isotropic than those refined from neutron data. Neutron crystallographic studies seem to reduce the impact of static and dynamic disorder in the refinement of ADPs, by allowing a better interpretation of rigid-body dynamics using TLS models. Additionally, a comparison between neutron and X-ray derived ADPs for the enzymatic catalytic residues, Glu35 and Asp52, provides unique insight into their protonation states.

Keywords

Neutron crystallography, X-ray diffraction, atomic resolution, perdeuterated lysozyme, atomic displacement parameters, atomic thermal motion, protein dynamics.

Introduction

In the analysis of single-crystal diffraction data, the Debye-Waller factor (W_i) is used in the calculation of the structure factors, to correct for the loss in intensity of the Bragg reflections caused by atomic thermal motion (Debye, 1914; Waller, 1923). The fact that atoms vibrate away from their equilibrium positions leads to destructive interreference of the scattered waves, particularly at larger scattering angles (*i.e.*, high resolution). This motion is described by the atomic displacement parameters (ADPs, *U*) which estimate how much an atom deviates from its equilibrium position in all directions and can, thus, provide insight into local and global dynamics of a molecule in a crystal. The atomic thermal motion is usually dominated by macromolecular bulk movements, being reflected in the refined model ADPs. TLS models allow the description of rigid-body motion in a crystal structure by grouping protein residues according to their dynamical behavior in translation-libration-screw movements (Cruickshank, 1956; Schomaker & Trueblood, 1968). While atomic displacement has different magnitudes in different directions, its estimation for proteins from diffraction data is usually performed by assuming the motion is similar in all directions, *i.e.*, isotropic. This approximation represents the addition of only 1 parameter/atom to the model refinement and is appropriate when diffraction data resolution is worse than 1.2 Å, and the data-to-parameter ratio is low. However, if atomic resolution data is available, the more realistic description of atomic motion is anisotropic, requiring 6 different parameters per atom. Due to the significant increase in the number of parameters used in refinement, complete data is required to avoid model overfitting, by maintaining a data-to-parameter ratio of at least 5.

The physical meaning of anisotropic ADPs derived from atomic resolution X-ray data is hindered by limitations inherent to the method, unless multipole models are employed (Blessing, 1995; Koritsanszky & Coppens, 2001; Oddershede & Larsen, 2004; Koetzle & McIntyre, 2012). Since X-rays interact with electrons, the center of an atom is not directly probed leading to errors related with the deformation of valence electron density distributions. Additionally, while Hydrogen (H) atoms comprise approximately 50% of all atoms in a protein, their determination remains elusive using X-ray data, even at ultra-high resolutions (Wang *et al.*, 2007; Schmidt *et al.*, 2011). Moreover, despite facilitating the measurement of high resolution X-ray data, synchrotrons have also potentiated the effects of radiation damage in crystallographic studies (Dauter *et al.*, 2010). On the other hand, neutron macromolecular crystallography (NMX) poses as the technique of choice to probe local and global dynamics of a molecule in a crystal (Koritsanszky & Coppens, 2001; Munshi *et al.*, 2008; Koetzle & McIntyre, 2012). Contrary to X-rays, neutrons are non-ionizing radiation, allowing routine measurements at room temperature (Blakeley, 2009). Neutron maps facilitate the identification of H and Deuterium (D) positions at resolutions better than 2.5 Å (Chen *et al.*, 2012; Ramos *et al.* in prep, Manuscript 3). Finally, neutrons interact with the nuclei of atoms, meaning that the maximum of the neutron density corresponds to the atomic equilibrium position. The determination of anisotropic ADPs using neutron diffraction is routine for small molecules (Koritsanszky & Coppens, 2001; Munshi *et al.*,

2008). However, the measurement of atomic resolution and complete neutron data from protein crystals remains extremely challenging.

The major limitation of NMX is the relatively low neutron flux, at reactor and spallation sources, when compared to the photon flux currently available at synchrotrons. When considering monochromatic diffractometers, the difference between neutron and X-ray fluxes can be of approximately 9 orders of magnitude. Therefore, the lower neutron flux has to be compensated by the use of larger crystal volumes, sample perdeuteration, and extended data collection times (Blakeley, 2009). Protein perdeuteration has been shown to have a significant impact in improving neutron diffraction data quality and in reducing sample volume requirements (Shu *et al.*, 2000; Hazemann *et al.*, 2005; Blakeley, 2009; Fisher *et al.*, 2014; Ramos *et al.* in prep, Manuscript 3). These benefits are associated with the distinct neutron scattering properties of H and D. While H possesses a negative neutron coherent scattering length (-3.74 fm) and a large incoherent scattering cross section (80.27 barn), D has a positive coherent scattering length (6.67 fm) comparable to ^{12}C, ^{14}N, ^{16}O (Sears, 1992).

Chen *et al.* (2012) reported the first atomic resolution neutron structure of a protein – crambin at 1.1 Å (PDB entry 4fc1). The fact that the dataset obtained was incomplete (78.8% overall completeness, 65.8% in the outer resolution shell), hindered the refinement of anisotropic ADPs against neutron diffraction data alone. Thus, the authors chose to treat the protein model as a rigid body while refining anisotropic ADPs for all non-hydrogen protein atoms. Posteriorly, anisotropic ADPs were refined for D atoms in 20 of the 46 residues of crambin. Recently, we demonstrated the potential of atomic resolution and complete neutron diffraction data by refining anisotropic ADPs for protein atoms and several water molecules, in our study of perdeuterated lysozyme (Ramos *et al.* in prep, Manuscript 3). However, to study protein dynamics through the estimation of anisotropic ADPs, and to evaluate the refined parameters from X-ray and neutron data, we performed experiments at different temperatures. This work can also provide important insight into protein static and dynamical disorder.

The crystal structures of perdeuterated hen egg-white lysozyme (D-HEWL), with refined anisotropic ADPs from complete atomic resolution neutron diffraction datasets at both room temperature (RT) and 100 K, are reported. Equivalent atomic resolution X-ray diffraction datasets are also presented. The analysis of static and dynamical disorder was performed using the four D-HEWL crystal structures, allowing the identification of well-ordered atoms, for which the refined ADPs are likely to reflect solely atomic motion. X-ray and neutron derived ADPs are analyzed and compared, highlighting their differences in reflecting structural and dynamical information. To our knowledge, this study represents the first time atomic thermal motion is described in protein crystals using both X-ray and neutron diffraction data. Additionally, unique detail from neutron data is provided concerning the atomic motion of D atoms.

Results and Discussion

D-HEWL disorder in X-ray and neutron models

Neutron diffraction data to 1 Å resolution was collected at 100 K, in D19 (ILL) from a triclinic D-HEWL crystal of approximately 5 mm^3. The corresponding 1 Å resolution RT neutron dataset from an identical crystal has been reported elsewhere (Ramos *et al.* in prep, Manuscript 3). The 100 K neutron model of D-HEWL was refined using SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015), from the final RT model. The data merging and refinement statistics are presented in Table 1.

The RT neutron model includes 24 protein residues with alternate conformations, 58 structural water molecules (of which 42 were modelled as D_2O), and 6 nitrate ions. On the other hand, the 100 K neutron model contains 23 protein residues with alternate conformations, 87 structural water molecules (of which 45 were modelled as D_2O), and also 6 nitrate ions. All nitrate ions were modelled in identical positions at both temperatures. Regarding the structural waters found in both models, more specifically the most well-ordered which were modelled as D_2O, 34 seem to be conserved at RT and 100 K. In Supplementary Fig. S1, the refined occupancies for the conformation A of all protein residues in the two models are compared. 77% of the protein residues present single conformations at both temperatures. Nevertheless, a close inspection of the single conformation residues' isotropic B-factors (B_{eq}), in the RT model (mean B_{eq} of 17.4 $Å^2$), shows that 27 of these are likely disordered or highly flexible, although alternate conformations were not visible in the 2*Fo-Fc* neutron maps contoured at 1σ (Fig. 1a).

X-ray diffraction data to 1.07 Å resolution has been collected from triclinic D-HEWL at RT in the MASSIF-1 beamline (ESRF). An equivalent X-ray dataset at 100 K has been reported in previous work (Ramos *et al.*, 2021), however, the model refinement for this study was performed with SHELXL (Sheldrick, 2015), instead of PHENIX (PDB entry 7ave). Similarly, the RT model has been refined in SHELXL (Sheldrick, 2015), allowing a direct comparison between the ADPs refined from data at both temperatures and across the methods. The data merging and refinement statistics for the two X-ray D-HEWL models are presented in Table 1. Additionally, the model refinement strategy for the RT structure is summarized in Table S1, while the 100 K model was refined using the final X-ray RT structure as model, as was the case for the neutron data.

The RT X-ray model of D-HEWL includes 27 alternate protein residue conformations, 81 structural water molecules, and 10 ions (8 nitrates and 2 acetates). Meanwhile, in the 100 K X-ray structure, 33 protein residue alternate conformations, 159 water molecules, and also 10 ions (8 nitrates and 2 acetates) were modelled. 9 of the acetate and nitrate ions were identified in identical positions at both temperatures, whereas one nitrate ion was found near Asn113 and Arg114 at RT, and another nitrate in the enzyme's binding cleft at 100 K. As expected, a significantly larger number of water molecules were found in D-HEWL at low temperature, however, it was found that 68 of these are conserved in both models, emphasizing their importance in protein folding and hydration. The residue disorder in both structures was compared and, in this instance, 75% of all residues were found in single

conformations at both temperatures (Supplementary Fig. S2). Considering only the residues in single conformations, the mean residue B_{eq} is 17.7 Å^2 at RT, and 10.8 Å^2 at 100 K. Assuming that the residues with B_{eq} greater than average are likely to suffer from disorder, biasing the refined ADPs, a total of 65 residues were, thus, considered as well-ordered with reliable ADPs (Fig. 1b).

Since the comparison of neutron and X-ray derived ADPs is of great interest, it is important to establish which protein residues can be considered well-ordered. A total of 62 protein residues appear to be well-ordered in the X-ray and neutron models at both temperatures (Supplementary Table S2).

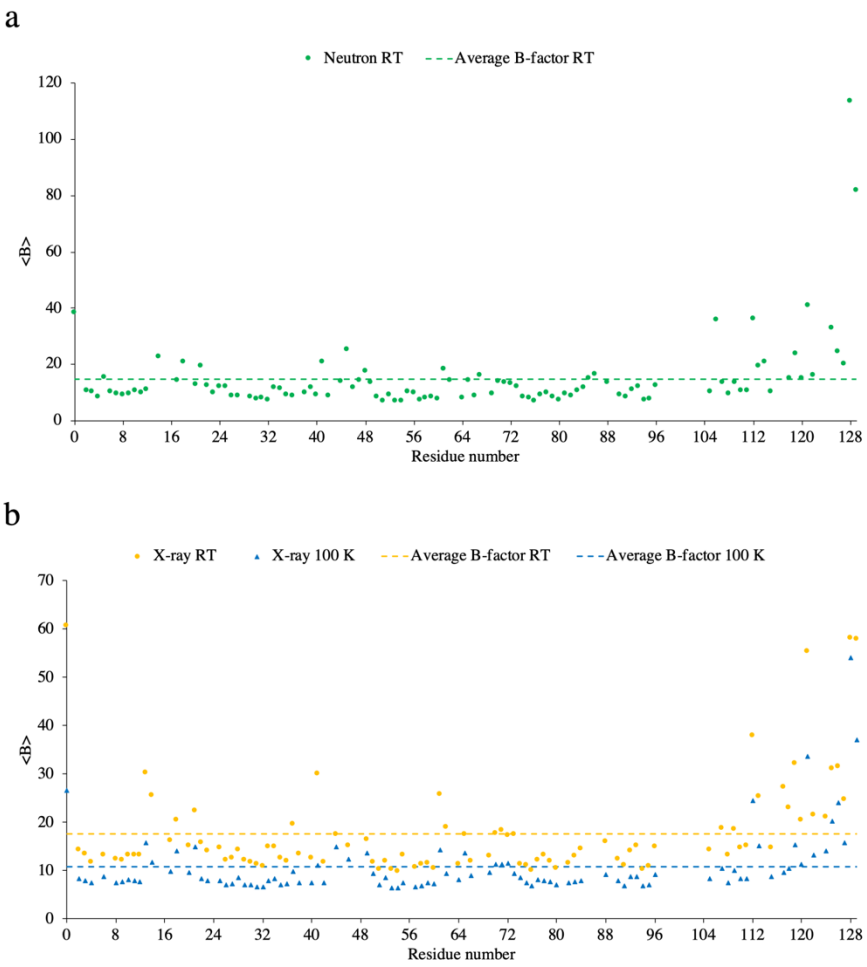


Figure 1

Residues with B_{eq} higher than average are likely to suffer from disorder, biasing the refinement of anisotropic ADPs. These disordered residues were identified in the RT neutron model (a) and in the X-ray models at RT and 100 K (b), being excluded from the subsequent ADP analysis.

Table 1

Merging data and model refinement statistics corresponding to the D-HEWL X-ray datasets at room temperature and 100 K, and neutron data at 100 K. The values in parenthesis correspond to the respective outer resolution shell.

	D-HEWL		
Radiation	Neutron	X-rays	
Temperature (K)	100	100	298
Source	D19, ILL, FR	I03, Diamond, UK	MASSIF-1, ESRF, FR
Detector	$^3He + CF_4$	EIGER2 XE 16M	PILATUS3 2M
Wavelength (Å)	1.45616	0.7293	0.965459
Resolution range (Å)	28.87 - 1.00 (1.04 - 1.00)	32.01 - 0.98 (1.015 - 0.98)	21.53 - 1.07 (1.11 - 1.07)
Space group	P1	P1	P1
Unit cell **a, b, c (Å)** **α, β, γ (°)**	26.84, 31.28, 33.8 88.944, 72.561, 68.098	26.67, 30.97, 33.74 89.439, 72.818, 67.503	27.15, 31.76, 34.07 88.746, 71.715, 67.985
Total reflections	130484 (8586)	278571 (24674)	549037 (27052)
Unique reflections	50814 (4841)	52966 (5018)	43570 (4049)
Multiplicity	2.6 (1.8)	5.3 (4.9)	12.6 (6.7)
Completeness (%)	97.1 (92.4)	97.44 (92.67)	99.2 (92.9)
Mean I/σ(I)	5.3 (0.8)	18.09 (4.21)	33.4 (7.5)
R_{merge}	0.105 (0.661)	0.0424 (0.273)	0.042 (0.0225)
R_{meas}	0.126 (0.867)	0.0470 (0.306)	0.043 (0.245)
R_{pim}	0.067 (0.555)	0.0201 (0.136)	0.012 (0.09)
$CC_{1/2}$	0.990 (0.408)	0.999 (0.935)	0.999 (0.965)
Refl. used in refinement with *Fo*/σ (*Fo*) > 4 / all data	23566 / 48200	46095 / 50313	39049 / 41391
Reflections used for R_{free} with *Fo*/σ (*Fo*) > 4 / all data	1254 / 2535	2434 / 2649	2047 / 2179
R_{work} for *Fo*/σ (*Fo*) > 4 / all data	14.67 / 24.86	9.62 / 10.04	9.62 / 9.93
R_{free} for *Fo*/σ (*Fo*) > 4 / all data	22.42 / 31.30	12.34 / 12.84	12.70 / 13.2
Number of non-H/D atoms (overall)	1308	1477	1340
macromolecules	1192	1275	1219
ligands	24	40	40
solvent	87	159	81
Protein residues	130	130	130
R.m.s.d., bond lengths (Å)	0.011	0.02	0.017
R.m.s.d., angles (°)	2.69	2.94	2.68
Ramachandran favored (%)	94.53	98.44	99.22
Ramachandran allowed (%)	3.91	1.56	0
Ramachandran outliers (%)	1.56	0	0.78
Rotamer outliers (%)	3.23	3.7	2.34
Clashscore	12.18	11.34	6.55
Average B-factor (overall) ($Å^2$)	18.07	13.86	20.51
macromolecules	17.91	12.23	19.26
ligands	20.3	26.22	38.74
solvent	18.66	25.17	32.27

The atomic thermal motion estimated by X-rays is larger than that from neutrons

Comparing the RT D-HEWL structures obtained from neutron and X-ray data can provide insight into the potential bias of X-rays in the estimation of anisotropic ADPs. Considering only the 62 well-ordered protein residues identified in the analysis of model disorder, their B_{eq} were compared in Fig. 2. It appears that the ADPs refined from X-ray data (B_{eq}^{X}) are systematically larger than those obtained from neutron data (B_{eq}^{N}). The mean difference between B_{eq}^{N} and B_{eq}^{X} (ΔB_{eq}^{N-X}) for the protein residues is 3.3 Å^2, with a standard deviation of 0.5 Å^2. Since the atomic motion in the residue side-chains is expected to be larger than in the protein backbone, similar comparisons were performed isolating these groups of atoms (Supplementary Fig. S3). Interestingly, similar variations in ΔB_{eq}^{N-X} were observed. For the main-chain atoms the mean ΔB_{eq}^{N-X} is 3.3 Å^2, while for the side-chains it is 3.4 Å^2. As expected, the standard deviation is greater for side-chains (0.8 Å^2), compared to that found between the main-chain atoms (0.4 Å^2).

Extensive work has been performed in small molecules towards the description of atomic thermal motion, using both X-ray and neutron diffraction (Blessing, 1995; Koritsanszky & Coppens, 2001; Oddershede & Larsen, 2004; Madsen, 2006; Munshi *et al.*, 2008). The observation of larger anisotropic ADPs refined from X-ray data in comparison with those obtained from neutron data, has been made previously in small-molecule crystallography. Blessing (1995) has described several corrections that can be performed to compensate for the experimentally observed ΔB_{eq}^{N-X}, which can be caused by isotropic or anisotropic errors in data collection or processing. Only variations in data collection temperature may result in an isotropic discrepancy between B_{eq}^{N} and B_{eq}^{X}. Thus, if the temperatures are identical, the ΔB_{eq}^{N-X} should only be explained by differences in accuracy of the modelling used to fit X-ray and neutron data. Typically, atomic models are used to fit diffraction data, however, it has been shown that this approach affects the quality of the refined ADPs (Koritsanszky & Coppens, 2001; Oddershede & Larsen, 2004; Munshi *et al.*, 2008). On the other hand, multipole models can describe more accurately atomic motion from X-ray diffraction data. For example, Oddershede & Larsen (2004) proved that atomic thermal motion could be deconvoluted from the electron density in multitemperature X-ray studies of naphthalene. Held & Van Smaalen (2014) have attempted to describe HEWL X-ray structure using a multipole model, however, the improvements made in comparison to the atomic model were not significant. The authors have justified this observation with the fact that the refined ADPs for proteins are significantly larger, compared to those obtained for small molecules.

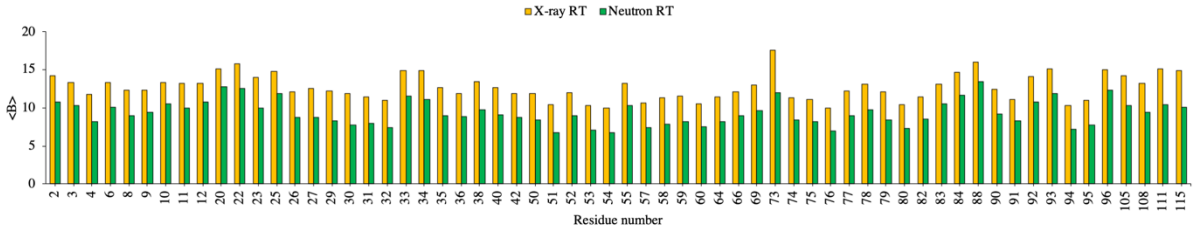


Figure 2

The refined B_{eq} from RT X-ray data appear to be systematically overestimated, when compared to those obtained from RT neutron data. The mean B_{eq} for each well-ordered protein residue are presented for the RT D-HEWL X-ray and neutron models.

ADPs refined from X-ray data have decreased anisotropy

The error associated with anisotropic ADP refinement from X-ray diffraction data is a consequence of the inaccuracy of the atomic model in describing atomic motion, not accounting for the non-spherical deformation of the valence electron density distribution caused by chemical bonding. Therefore, when model refinement is performed only for ultra-high resolution X-ray data (*i.e.*, > 0.8 Å) or when multipolar valence electron density distributions are parameterized, the $\Delta B_{eq}^{N\text{-}X}$ is usually reduced (Blessing, 1995). The anisotropy of atomic motion is determined by the ratio of the minimum and maximum eigenvalues of the 3 × 3 matrix of anisotropic ADPs (Trueblood *et al.*, 1996). A more isotropic behavior is reflected by an anisotropy close to 1, while values approaching 0 represent higher anisotropy. Studies have shown that the mean anisotropy (<A>) in protein crystals, from atomic and near-atomic resolution structures deposited in the Protein Data Bank (PDB) (Berman *et al.*, 2000), fluctuates between 0.4 and 0.5 (Merritt, 1999; Zucker *et al.*, 2010).

A comparison of the results obtained for the D-HEWL RT X-ray and neutron structures (Table 2) shows that the X-ray model reflects reduced ADP anisotropy, in contrast with the neutron structure. The <A> is 0.48 and 0.43 for the X-ray and neutron models, respectively, with associated errors [σ(A)] of 0.17 and 0.16. This discrepancy indicates bias of the X-ray model, which does not reliably describe atomic motion, due to the convolution of this phenomenon with the electron density deformation promoted by chemical bonding. Significant deviations in <A> can be found for the most electrophilic atoms, O and S, which, interestingly, would be the atoms with the most deformed electron densities when bonding to atoms such as C and H, as opposed to what occurs, for instances, in C-C or C-H bonds.

If only the 62 D-HEWL residues considered as well-ordered are analyzed in terms of ADP anisotropy, their behavior appears to be more isotropic than that reflected by the entire models. While for the X-ray structure <A> is 0.59 (σ(A) = 0.12), for the neutron model <A> is 0.51 (σ(A) = 0.13). Although the discrepancy in ADP anisotropy between X-rays and neutrons is maintained, the overestimation of <A> found for the overall protein structures suggests that static and dynamic disorder can affect significantly the description of anisotropy. This observation has also implications in the observed <A>, between 0.4 and 0.5, reported for several atomic and near-atomic resolution X-ray

structures deposited in the PDB (Merritt, 1999; Zucker *et al.*, 2010), since disorder was not accounted for in these studies.

Table 2

Mean anisotropy (<A>) of the ADPs refined for D-HEWL RT and 100 K from X-ray and neutron data, using the PARVATI web server (Zucker *et al.*, 2010). All protein non-H/D atoms were considered for the calculations.

	X-ray RT		**X-ray 100 K**		**Neutron RT**	
	<A>	**σ(A)**	**<A>**	**σ(A)**	**<A>**	**σ(A)**
Protein	0.484	0.168	0.471	0.185	0.426	0.159
N	0.453	0.183	0.446	0.204	0.426	0.174
C	0.498	0.165	0.484	0.178	0.442	0.151
O	0.464	0.157	0.449	0.182	0.371	0.158
S	0.642	0.135	0.623	0.151	0.451	0.107

Hinge bending mechanism identified in neutron model

TLS models can be used in model refinement strategies, since the number of additional parameters is minimized while providing a good approximation of the atomic B-factors. TLS refinements have, thus, proved to be useful in improving the accuracy of crystallographic models, especially when atomic resolution is not available, and the data-to-parameter ratio is low (Painter & Merritt, 2006*a,b*; Zucker *et al.*, 2010). Additionally, TLS analysis can be performed on refined crystal structures that include anisotropic ADPs, as a way of extracting information in terms of protein dynamics at the molecular scale. This type of analysis is achieved by programs such as TLSMD (Painter & Merritt, 2006*a*,*b*), and TLS models have provided relevant biological information regarding protein function, *e.g.*, by describing inter-domain hinge motion (Wilson & Brunger, 2000; Papiz *et al.*, 2003; Chaudhry *et al.*, 2004; Bernett *et al.*, 2004).

A hinge-bending mechanism has also been described for HEWL, in which two domains are separated by the enzyme's active site (McCammon *et al.*, 1976; Bruccoleri *et al.*, 1986). These domains correspond to the two folding domains (α- and β-domains) described by Miranker *et al.* (1991). The TLS analysis performed for the D-HEWL RT X-ray and neutron structures showed different rigid-body dynamics, as a consequence of different fitting of the refined B_{eq} (Fig. 3). For the neutron model, two main domains (residues 0-45; 46-104) were modelled around the catalytic residues Glu35 and Asp52, as expected, with additional segmentations of the very flexible C-terminal region (residues 105-112; 113-124; 125-129). On the other hand, the TLS model for the X-ray structure does not separate the two domains at the catalytic site. While significant disorder was observed in both N- and C-terminal regions, leading to their segmentation (residues 0-15; 121-129), the first main domain (residues 16-60) includes both Glu35 and Asp52. Moreover, distinct rigid-body dynamics were modelled for the Arg61-Arg68 region, and the second main domain (residues 69-120) is significantly shifted in the amino acid sequence, when compared to that of the neutron model (residues 46-104). A similar TLS model was also obtained for the 0.65 Å resolution HEWL X-ray structure by Wang *et al.* (2007) (Supplementary

Fig. S4). Another indication that the TLS model finds better agreement with the refined B_{eq} from neutron data, in comparison with those from X-ray data, is the significant decrease in least squares residual upon model segmentation, from 2 to 20. Moreover, the steep reduction in residuals for the first 2 to 5 segmentations of the neutron model is a clear sign that these parts of the structure present relevant rigid-body movement. The work conducted by Bruccoleri *et al.* (1986) suggests that the catalytic residues Glu35 and Asp52 belong to separate domains with distinct rigid-body dynamics, which upon ligand binding would move these residues closer together facilitating enzymatic activity. The TLS model that displayed a structure segmentation more coherent with the enzyme's biological function was the one fitted to the neutron data. These results suggest that the B_{eq} from neutron data possess important biophysical meaning, while those obtained from X-ray data appear to be somewhat affected, partly losing biological relevance. An explanation for the worse fit to the X-ray B_{eq} could be the influence of side-chain disorder, leading to the overestimation of the residues B_{eq}. As reported by Ramos *et al.* in prep, Manuscript 3, neutron data, particularly from perdeuterated variants, can provide additional insight into protein disorder, compared to X-ray data, since the neutron maps include information on the positions of H/D atoms, facilitating the unambiguous modelling of residue side-chains, such as Asn and Gln, and also in the case of large residues like Arg and Trp.

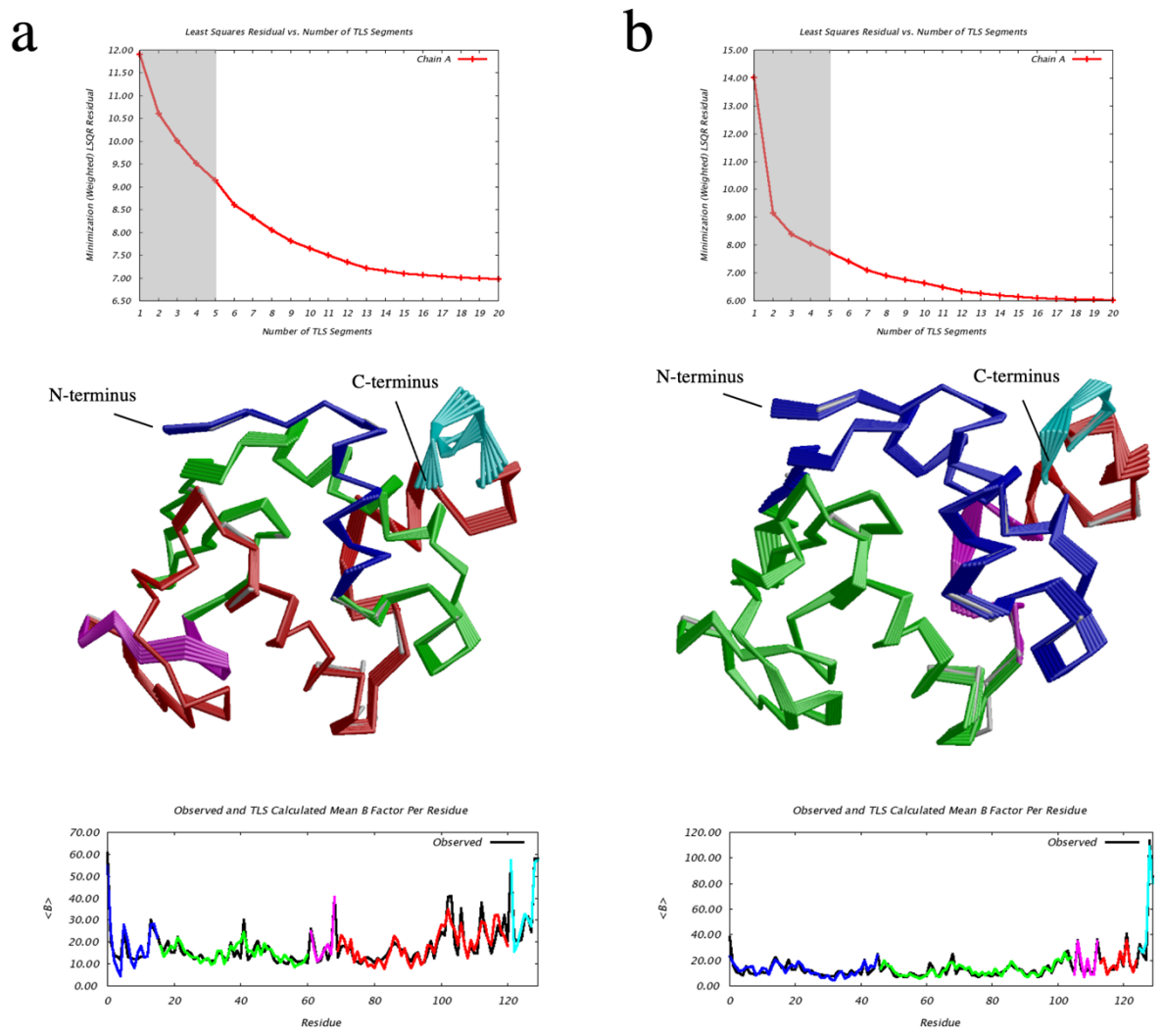


Figure 3

The TLS models fit better the residues B_{eq} obtained from neutron data, compared to those from X-ray data. The results from the TLS analysis on both X-ray (a) and neutron (b) structures are presented, showing the quality of the TLS model fit using 1 to 20 segments (top), the segmentation of the three-dimensional structures assuming 5 rigid bodies (middle), and fitted B_{eq} of the corresponding segments (colored) with the experimental values (black) (bottom).

Protein dynamics at the atomic level

The aforementioned differences between the ADPs refined based on X-ray diffraction data and those from neutron data were found consistent over 62 well-ordered protein residues. This coherence is well illustrated by the small standard deviations for the $\Delta B_{eq}^{N\text{-}X}$ of all atoms and also of main-chain atoms. However, a closer inspection of the refined ADPs for some of these well-ordered residues provides a clearer idea of the information available from both models. Additionally, the neutron data presented allows unprecedented detail regarding the dynamics of D atoms.

In Fig. 4, ellipsoidal representations of the neutron and X-ray refined B_{eq} are shown for the catalytic residues Glu35 and Asp52, as well as for the hydrophobic residues Phe34 and Val92. In every case, the B_{eq} refined from X-ray data are significantly larger than those from neutron data. In Glu35, the

difference in B_{eq} between Oε1 and Oε2 is approximately the same in both X-ray and neutron models. Oε1 displays a lower B_{eq} in the two models, which could be indicative of the protonation state of Glu35 side-chain. Clearly the neutron model shows that Oε1 is protonated and that the anisotropic behavior of Dε1 is similar to the atom to which it is bonded, as expected. Regarding Aps52, it is interesting to find that the X-ray B_{eq} is significantly larger for Oδ2, than for Oδ1. Meanwhile, in the neutron model, both atoms have similar B_{eq}. This discrepancy could be explained by the delocalization of the negative charge of Asp52 carboxylate group towards Oδ2, which plays a role in the stabilization of the electron deficient substrate molecule after breakage of its glycosidic bond. Asp52 is, thus, a prime example of the bias of X-ray models in the estimation of ADPs, not just reflecting atomic motion but also the deformation of the valence electron density distribution. In Phe34, significant $\Delta B_{eq}^{N\text{-}X}$, of up to 4.9 Å^2, can be observed for the C atoms of the aromatic ring of the side-chain. In this case, the X-ray B_{eq} are not only biased by the deformation of the valence electron densities due to the C-H bonds, but also by the delocalized electron density of the aromatic ring. This latter phenomenon is somewhat analogous to what is observed for Asp52 Oδ2. Finally, in Val92, $\Delta B_{eq}^{N\text{-}X}$ of 3.9 Å^2 and 5.4 Å^2 are present for Cγ1 and Cγ2, respectively. These atoms are part of methyl groups (CH_3) which possess a high degree of rotational freedom. Therefore, it would be expected that their refined ADPs would reflect significant motions. Once more, the X-ray ADPs seem to exacerbate the motion of the C atoms because of the distortion of their valence electron density by the covalent bonds to the respective 3 H atoms.

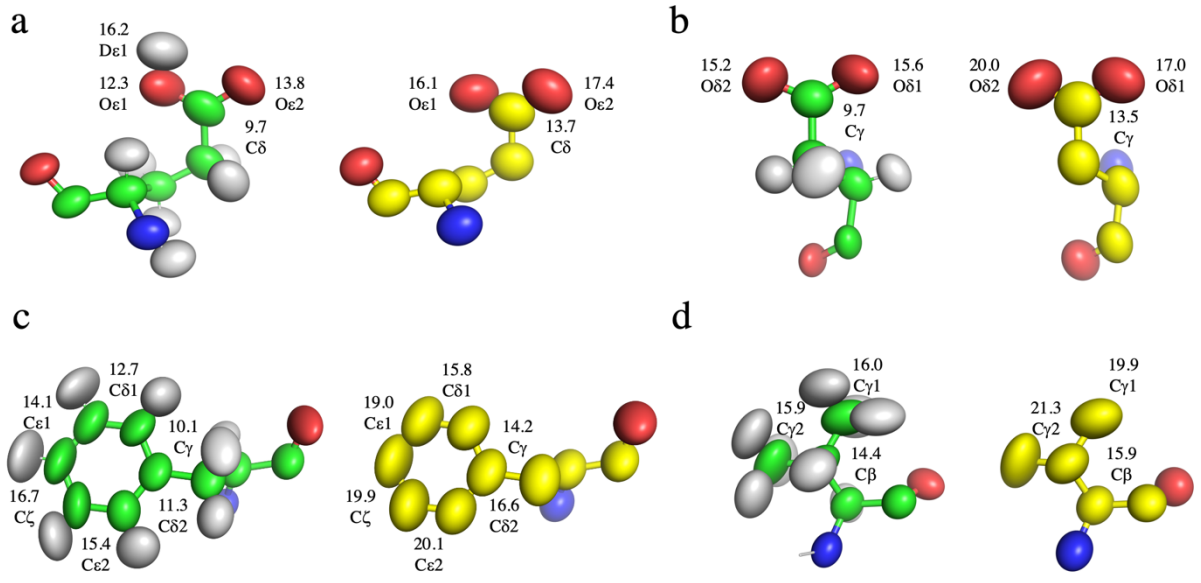


Figure 4

Protein dynamics at the atomic level, using Glu35 (a), Asp52 (b), Phe34 (c) and Val92 (d) as examples. The refined atomic B_{eq} are represented by ellipsoids and the respective values for some of the side-chain atoms are shown. The D-HEWL RT neutron and X-ray models are colored in green and yellow, respectively.

Conclusions

This study demonstrated the discrepancies in the refinement of anisotropic ADPs between X-ray and neutron diffraction data for protein crystals. Static and dynamic disorder were separated through structural analysis of data measured at different temperatures. The results presented for RT D-HEWL suggest that atomic thermal motion estimated from X-ray data is larger than that from neutron data. Moreover, X-ray ADPs are described as more isotropic than those obtained in the neutron model.

It seems that protein static and dynamic disorder have significant impacts in the estimation of ADP anisotropy and also in their analysis using TLS models. These effects can lead to misinterpretation of protein dynamics and function through X-ray crystallographic data. NMX seems, thus, to be advantageous in the description of both protein disorder and atomic motion, when complete diffraction datasets are available.

Interesting details were found in the refined B_{eq} in both X-ray and neutron models, especially for HEWL catalytic residues, which can be useful for the interpretation of atomic B_{eq} in future X-ray crystallographic studies.

Methods

Protein materials

D-HEWL was expressed in *E. coli* BL21(DE3) cells in the form of inclusion bodies, purified, and subsequently refolded *in vitro*, as described previously (Ramos *et al.*, 2021). The protein buffer was exchanged to 50 mM sodium acetate pD 4.5 (pD = pH + 0.4) (Glasoe & Long, 1960) in D_2O. The protein was concentrated to 20 mg/mL for the crystallization experiments.

Protein crystallization

Triclinic crystals of D-HEWL were initially obtained in D_2O, using microbatch under oil and microseeding of triclinic hydrogenated HEWL at 18 ° C, as reported previously (Ramos *et al.*, 2021). Later, new seeds of perdeuterated crystals were prepared and triclinic D-HEWL crystals were obtained in similar conditions. The drops contained 2.5 μL of 20 mg/mL D-HEWL, 2.5 μL of 0.3 M $NaNO_3$ and 50 mM sodium acetate pD 4.5 solution in D_2O, and 0.5 μL of seed stock. Crystals of approximately 0.1 mm^3 appeared within 1 week.

To grow crystals suitable for neutron diffraction experiments, macroseeding in batch-like conditions was employed, also at 18 ° C. The 0.1 mm^3 D-HEWL crystals were transferred to new drops of 20 μL containing 10 mg/mL of D-HEWL, 0.15 M $NaNO_3$ and 50 mM sodium acetate pD 4.5 in D_2O. After drop equilibration for 1-2 weeks, the crystallization solution was replaced with fresh protein and precipitant mixture. After a few weeks in 20 μL drops, the crystals were moved to 9-well glass plates in sandwich boxes from Hampton Research, in order to allow the use of larger crystallization drops (40 – 60 μL).

Neutron diffraction data collection and processing

RT neutron diffraction data was collected from a ~4 mm^3 triclinic D-HEWL crystal at D19 (ILL) as described previously (Ramos *et al.* in prep, Manuscript 3). The 100 K dataset was recorded from a ~5 mm^3 triclinic D-HEWL crystal also at D19 (ILL). Prior to cryocooling, the crystal was soaked in 0.4 M $NaNO_3$ and 50 mM sodium acetate pD 4.5 solutions in D_2O, containing incremental concentrations of d_8-glycerol. After 78 minutes in 5% (v/v) d_8-glycerol, the solution was replaced by a mixture with 10% (v/v) d_8-glycerol for 10 minutes and later by a solution with 15% (v/v) d_8-glycerol for another 10 minutes. Finally, the crystal was soaked for 10 minutes in 20% (v/v) d_8-glycerol before being mounted and flash-cooled in the N_2 stream. Neutron data was recorded in 29 ω scans in static frames of 0.1° steps, similar to what was performed in the RT data collection (Ramos *et al.* in prep, Manuscript 3). However, each ω scan was split into two in order to reduce the amount of ice formed in the loop, at the interfaces between the cryo-stream, the dry-air stream and the surrounding air at the instrument. The exposure per frame was of approximately 52 s. 3 of the total 58 ω scans were repeated at lower exposure times (20 s) to improve the measurement accuracy of the low angle reflections.

The data was processed with XDS (Kabsch, 2010), similarly to what was reported for the RT dataset (Ramos *et al.* in prep, Manuscript 3). While scaling with XSCALE (Kabsch, 2010), the shorter exposure scans were cut at 1.53 Å resolution, to avoid adding noise to the averaging of weaker reflections at higher resolutions from the 52 s scans. XDSCC12 (Assmann *et al.*, 2020) was employed to assess the quality of each scan compared to the remaining dataset, leading to the inclusion of all recorded scans in the final dataset. The data was then converted to SHELX format and the flags for R_{free} calculations were copied from the RT neutron dataset, accounting for about 5% of the data.

X-ray diffraction data collection and processing

For the 100 K X-ray experiment, triclinic D-HEWL crystals of ~0.1 mm^3 were cryocooled after soaking in 35 % (v/v) d_8-glycerol, 0.3 M $NaNO_3$ and 50 mM sodium acetate pD 4.5, in D_2O (Ramos *et al.*, 2021). X-ray diffraction data was collected and processed as described by Ramos *et al.* (2021).

In the RT X-ray experiment, triclinic D-HEWL crystals of ~0.1 mm^3 were recovered from drops where large D-HEWL crystals were grown for neutron experiments, in batch-like conditions at 18° C. The crystals were fished out with standard nylon loops which were covered by 3 cm quartz capillaries (Charlessupper Company) and sealed with plasticine, similarly to what was described by Mac Sweeney & D'Arcy (2003). At the end of the capillary crystallization solution was added to avoid crystal dehydration. X-ray diffraction data at RT was collected in beamline MASSIF-1 (ESRF). A total of 7 D-HEWL crystals were measured in 360° sweeps with steps of 0.1°. Data reduction was performed with XDS (Kabsch, 2010) and signs of radiation damage were investigated by merging and scaling each dataset with Aimless (Evans & Murshudov, 2013). Data from 4 of the crystals were excluded after 200° of rotation and from other 2 crystals after 250° and 300°. For the remaining crystal, the variation of R_{merge} over number of frames suggested that the crystal could have moved at the start of the data collection, and thus the data included comprised only the last 210°. Afterwards, the filtered data from the 7 crystals were scaled using XSCALE (Kabsch, 2010).

The X-ray datasets of both RT and 100 K experiments were converted to SHELX format using XDSCONV (Kabsch, 2010) and approximately 5% of the reflections were flagged for the R_{free} calculation by copying the flags from the neutron RT dataset.

Model refinement

Model refinements were performed in SHELXL (Gruene *et al.*, 2014; Sheldrick, 2015) due to its flexibility and suitability for refinement against high resolution X-ray and neutron data. Both X-ray and neutron RT models were refined first and used as starting models for the refinements against the 100 K datasets. The initial model for the RT X-ray and neutron refinements was the D-HEWL crystal structure – PDB entry 7ave (Ramos *et al.*, 2021) – stripped of protein residue alternate conformations, water molecules and ions, and anisotropic B-factors were converted to isotropic. Later, the 100 K models were remodeled in COOT (Emsley *et al.*, 2010) according to the strongest peaks in the

respective *Fo-Fc* electron density and neutron maps and corrections were made after inspection of the density fit analysis and B-factors variance.

The neutron RT model was refined following a strategy reported by Ramos *et al.* in prep, Manuscript 3. The restraints for deuterated amino acids were retrieved from the work of Gruene *et al.* (2014). The refined anisotropic ADPs were restrained by the SIMU (0.1) and XPND (0.001) commands from SHELXL (Sheldrick, 2015). On the other hand, the X-ray RT model refinement strategy is summarized in Supplementary Table S1. Regarding the X-ray models, the most significant change was the inclusion of H atoms in idealized positions with isotropic B-factors constrained to 1.2 times the B-factor of the atom to which they are bonded, and 1.5 times when bonded to side-chain O atoms (*e.g.*, Ser, Thr, Tyr), or to CH_3 and NH_3 groups. Occupancies of water molecules and ions were refined if their B-factors were greater than 30 $Å^2$. However, whenever the refined occupancy was superior to 0.95, its value was fixed at 1. Contrary to the neutron model refinements, the SIMU (0.1) and XPND (0.001) restraints to ADPs were removed in late stages of the X-ray structures refinements.

Structural comparisons

The refined B_{eq} from the RT X-ray and neutron models were analyzed only for non-H/D atoms. The residue, main-chain, and side-chain B_{eq} were calculate using the structural comparison tool from PHENIX (Liebschner *et al.*, 2019).

ADP anisotropies were compared between the D-HEWL RT X-ray and neutron models using the PARVATI web server (Zucker *et al.*, 2010). Water molecules, ions and H/D atoms were removed from the submitted *pdb* files and the default settings were used.

TLS analysis was performed using TLSMD (Painter & Merritt, 2006*a*,*b*) for the RT D-HEWL X-ray and neutron structures, excluding H/D atoms. The default settings were employed. An identical approach was used for the 0.65 Å resolution HEWL X-ray structure reported by Wang *et al.* (2007) (PDB entry 2vb1).

References

Assmann, G. M., Wang, M. & Diederichs, K. (2020). *Acta Crystallogr.* **D76**, 636–652.

Berman, H., Westbrook, J., Feng, Z., Gilliand, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bernett, M. J., Somasundaram, T. & Blaber, M. (2004). *Proteins Struct. Funct. Genet.* **57**, 626–634.

Blakeley, M. P. (2009). *Crystallography Reviews*, Vol. pp. 157–218.

Blessing, R. H. (1995). *Acta Crystallogr.* **B51**, 816–823.

Bruccoleri, R. E., Karplus, M. & McCammon, J. A. (1986). *Biopolymers*. **25**, 1767–1802.

Chaudhry, C., Horwich, A. L., Brunger, A. T. & Adams, P. D. (2004). *J. Mol. Biol.* **342**, 229–245.

Chen, J. C. H., Hanson, B. L., Fisher, S. Z., Langan, P. & Kovalevsky, A. Y. (2012). *PNAS.* **109**, 15301–15306.

Cruickshank, D. W. J. (1956). *Acta Crystallogr.* **9**, 754–756.

Dauter, Z., Jaskolski, M. & Wlodawer, A. (2010). *J. Synchrotron Radiat.* **17**, 433–444.

Debye, P. (1914). *Ann. Phys.* **43**, 49.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr.* **D66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr.* **D69**, 1204–1214.

Fisher, S. J., Blakeley, M. P., Howard, E. I., Petit-Haertlein, I., Haertlein, M., Mitschler, A., Cousido-Siah, A., Salvay, A. G., Popov, A., Muller-Dieckmann, C., Petrova, T. & Podjarny, A. (2014). *Acta Crystallogr.* **D70**, 3266–3272.

Glasoe, P. K. & Long, F. A. (1960). *J. Phys. Chem.* **64**, 188–190.

Gruene, T., Hahn, H. W., Luebben, A. V., Meilleur, F. & Sheldrick, G. M. (2014). *J. Appl. Crystallogr.* **47**, 462–466.

Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Crystallogr.* **D61**, 1413–1417.

Held, J. & Van Smaalen, S. (2014). *Acta Crystallogr.* **D70**, 1136–1146.

Kabsch, W. (2010). *Acta Crystallogr.* **D66**, 133–144.

Koetzle, T. F. & McIntyre, G. J. (2012). *Characterization of Materials*, Vol. pp. 2192–2205.

Koritsanszky, T. S. & Coppens, P. (2001). *Chem Rev*. **101**, 1583–1627.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Crystallogr.* **D75**, 861–877.

Madsen, A. Ø. (2006). *J. Appl. Crystallogr.* **39**, 757–758.

McCammon, J. A., Gelin, B. R., Karplus, M. & Wolynes, P. G. (1976). *Nature*. **262**, 325–326.

Merritt, E. A. (1999). *Acta Crystallogr.* **D55**, 1109–1117.

Miranker, A., Radford, S. E., Karplus, M. & Dobson, C. M. (1991). *Nature*. **349**, 633–636.

Munshi, P., Madsen, A. Ø., Spackman, M. A., Larsen, S. & Destro, R. (2008). *Acta Crystallogr.* **A64**, 465–475.

Oddershede, J. & Larsen, S. (2004). *J. Phys. Chem. A*. **108**, 1057–1063.

Painter, J. & Merritt, E. A. (2006*a*). *Acta Crystallogr.* **D62**, 439–450.

Painter, J. & Merritt, E. A. (2006*b*). *J. Appl. Crystallogr.* **39**, 109–111.

Papiz, M. Z., Prince, S. M., Howard, T., Cogdell, R. J. & Isaacs, N. W. (2003). *J. Mol. Biol.* **326**, 1523–1538.

Ramos, J., Laux, V., Haertlein, M., Boeri Erba, E., McAuley, K. E., Forsyth, V. T., Mossou, E., Larsen, S. & Langkilde, A. E. (2021). *IUCrJ*. **8**,.

Ramos, J., Mossou, E., Laux, V., Mason, S., Lemee, M., Diederichs, K., Haertlein, M., Forsyth, V. T., Larsen, S. & Langkilde, A. E. (2021). A complete picture of hen egg-white lysozyme active site at room temperature from 1 Å resolution neutron diffraction. Manuscript in preparation.

Schmidt, A., Teeter, M., Weckert, E. & Lamzin, V. S. (2011). *Acta Crystallogr.* **F67**, 424–428.

Schomaker, V. & Trueblood, K. N. (1968). *Acta Crystallogr.* **B24**, 63–76.

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

Sheldrick, G. M. (2015). *Acta Crystallogr.* **C71**, 3–8.

Shu, F., Ramakrishnan, V. & Schoenborn, B. P. (2000). *PNAS*. **97**, 3872–3877.

Mac Sweeney, A. & D'Arcy, A. (2003). *J. Appl. Crystallogr.* **36**, 165–166.

Trueblood, K. N., Bürgi, H. B., Burzlaff, H., Dunitz, J. D., Gramaccioli, C. M., Schulz, H. H., Shmueli, U. & Abrahams, S. C. (1996). *Acta Crystallogr.* **A52**, 770–781.

Waller, I. (1923). *Zeitschrift Für Phys.* **17**, 398–408.

Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Crystallogr.* **D63**, 1254–1268.

Wilson, M. A. & Brunger, A. T. (2000). *J. Mol. Biol.* **301**, 1237–1256.

Zucker, F., Champ, P. C. & Merritt, E. A. (2010). *Acta Crystallogr.* **D66**, 889–900.

Supporting Information

Protein dynamics at the atomic level: comparing atomic displacement parameters from X-ray and neutron diffraction

Joao Ramos[1,2,3], Valerie Laux[1,2], Marie-Helene Lemee[4], Matthew W. Bowler[5], Kay Diederichs[6], Michael Haertlein[1,2], V. Trevor Forsyth[1,2,7], Estelle Mossou[1,2,7], Sine Larsen[8], Annette E. Langkilde[3]*

[1]*Life Sciences Group, Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.*
[2]*Partnership for Structural Biology (PSB), 71 avenue des Martyrs, 38000 Grenoble, France.*
[3]*Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark.*
[4]*Diffraction Group, Institut Laue-Langevin, 71 avenue des Martyrs, 38000 Grenoble, France.*
[5]*European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, 38000 Grenoble, France.*
[6]*Department of Biology, University of Konstanz, Box 647, D-78457 Konstanz, Germany.*
[7]*Faculty of Natural Sciences, Keele University, Staffordshire, ST5 5BG, United Kingdom.*
[8]*Department of Chemistry, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark.*
*corresponding author, annette.langkilde@sund.ku.dk

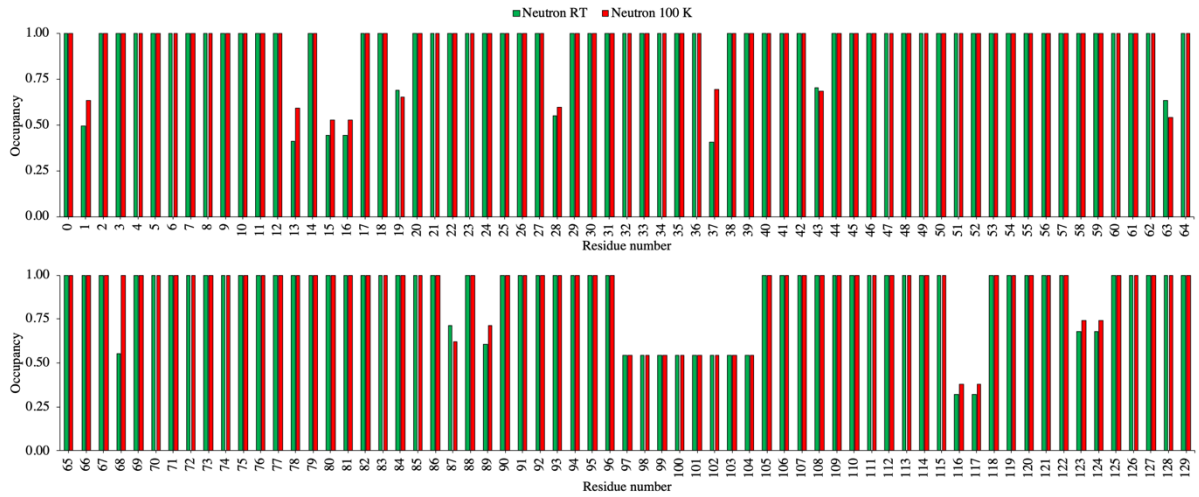


Figure S1

Refined occupancies for residues conformations A in the neutron RT and 100 K D-HEWL structures.

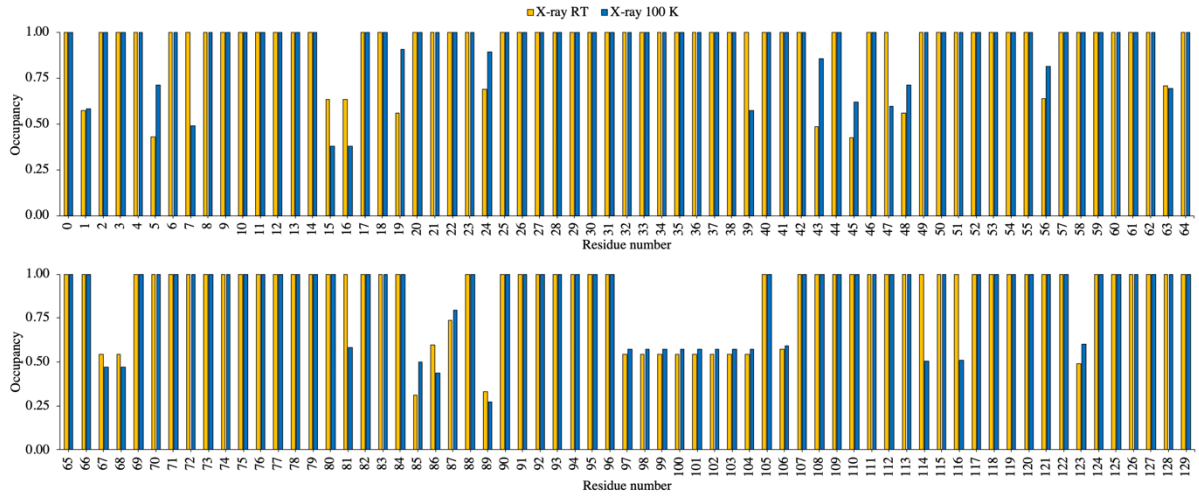


Figure S2

Refined occupancies for residues conformations A in the X-ray RT and 100 K D-HEWL structures.

Table S1

Summary of the refinement strategy used in SHELXL for the X-ray RT data, highlighting the variations in R-factors (for reflections with $Fo/\sigma(Fo) > 4$), goodness-of-fit, and respective data-to-parameter ratio.

Round	Refinement step	Resolution range (Å)	Rwork (%)	Rfree (%)	GooF	Data-to-parameter ratio
1	Rigid body refinement	22.84-2.0	33.17	32.25	10.7	10.4
2	Coordinate and isotropic B-factor refinement	22.84-1.5	23.06	25.87	7.1	4.8
3	Adding waters and ions	22.84-1.5	18.33	21.65	5.9	4.5
4	Resolution cut	22.84-1.07	18.61	20.90	4.9	10.6
5	Modelling disorder	22.84-1.07	17.71	19.81	4.8	9.4
6	Adding Hydrogens	22.84-1.07	16.24	19.39	4.5	9.1
7	Anisotropic ADP refinement of non-H protein atoms (except for residues with very high B-factors)	22.84-1.07	11.78	14.97	2.9	6.2
8	Refining occupancies for waters and ions	22.84-1.07	11.39	14.78	2.8	6.2
9	Anisotropic ADP refinement of all non-H protein atoms	22.84-1.07	10.54	13.98	2.7	6.0
10	Removing SIMU restraints	22.84-1.07	10.36	13.66	2.6	5.0
11	Removing XPND restraints	22.84-1.07	10.32	13.59	2.6	5.0
12	Anisotropic ADP refinement of ions	22.84-1.07	10.19	13.50	2.6	4.9
13	Anisotropic ADP refinement of waters	22.84-1.07	9.66	12.86	2.4	4.7
14	Minor corrections	22.84-1.07	9.61	12.68	2.4	4.8

Table S2

List of well-ordered residues common to all X-ray and neutron structures at both room temperature and 100 K, and respective mean B_{eq}.

Residue	<B_{eq}> (Å^2)			Residue	<B_{eq}> (Å^2)		
	X-ray RT	X-ray 100 K	Neutron RT		X-ray RT	X-ray 100 K	Neutron RT
2	14.23	8.43	10.75	**55**	13.23	7.5	10.29
3	13.39	7.9	10.31	**57**	10.69	6.57	7.39
4	11.77	7.57	8.24	**58**	11.35	6.97	7.86
6	13.32	8.76	10.08	**59**	11.52	7.55	8.25
8	12.34	7.44	8.97	**60**	10.54	7.33	7.55
9	12.3	7.65	9.38	**64**	11.4	8.13	8.15
10	13.32	8.16	10.57	**66**	12.08	8.96	8.94
11	13.23	7.97	9.98	**69**	12.96	9.65	9.63
12	13.24	7.79	10.82	**73**	17.56	9.49	11.96
20	15.17	9.7	12.77	**74**	11.35	8.49	8.48
22	15.84	8.42	12.56	**75**	11.12	7.49	8.15
23	14.02	7.89	10	**76**	9.98	6.94	6.99
25	14.74	7.89	11.91	**77**	12.19	8.09	8.97
26	12.11	7	8.8	**78**	13.17	7.91	9.75
27	12.54	7.36	8.73	**79**	12.07	7.76	8.38
29	12.19	7.1	8.27	**80**	10.47	7.16	7.35
30	11.88	7.03	7.79	**82**	11.47	7.47	8.59
31	11.46	6.75	7.95	**83**	13.1	7.78	10.55
32	11.01	6.77	7.38	**84**	14.64	8.05	11.61
33	14.9	7.86	11.5	**88**	16.02	9.27	13.49
34	14.94	8.43	11.14	**90**	12.49	7.86	9.26
35	12.64	7.07	9	**91**	11.09	6.87	8.33
36	11.89	7.26	8.86	**92**	14.15	8.87	10.77
38	13.43	7.57	9.78	**93**	15.18	8.84	11.88
40	12.67	7.51	9.13	**94**	10.27	6.85	7.15
42	11.87	7.58	8.72	**95**	10.96	7.19	7.78
50	11.86	9.52	8.44	**96**	15.05	9.24	12.29
51	10.38	7.14	6.79	**105**	14.27	8.28	10.37
52	12.05	8.56	8.99	**108**	13.28	7.59	9.42
53	10.36	6.52	7.11	**111**	15.08	8.3	10.45
54	9.95	6.36	6.8	**115**	14.85	8.88	10.09

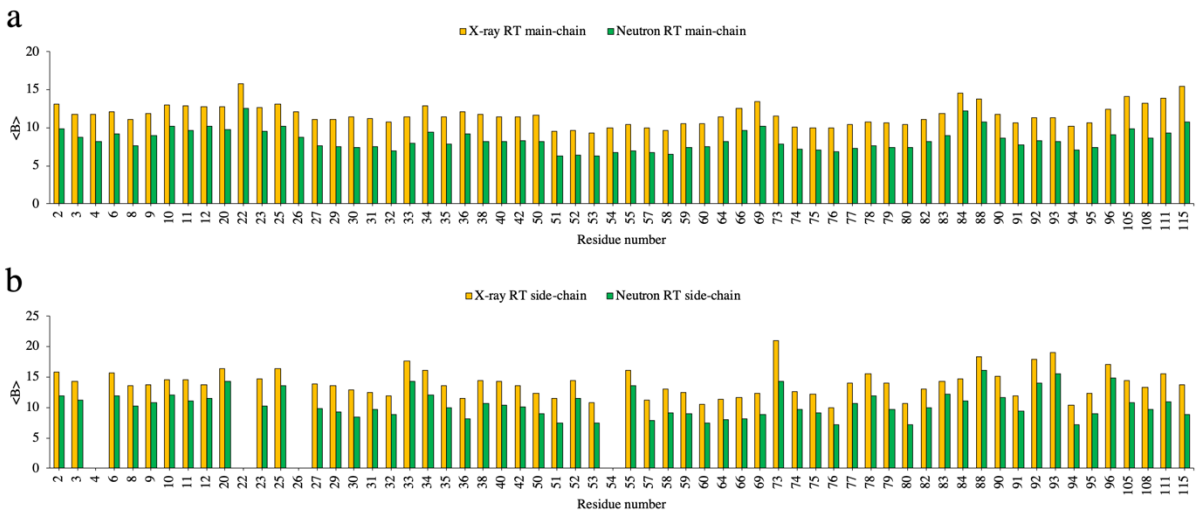


Figure S3

Mean B_{eq} for the main-chain (a) and side-chain (b) atoms of the well-ordered residues in D-HEWL RT X-ray and neutron models.

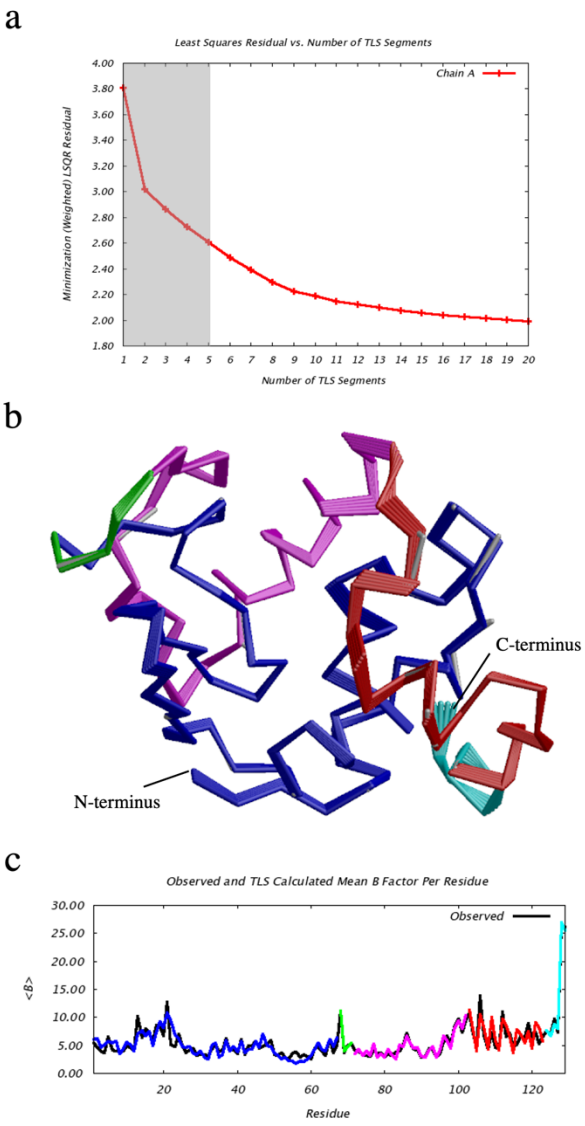


Figure S4

TLS model obtained for the 0.65 Å resolution X-ray structure of HEWL (PDB entry 2vb1, Wang *et al.* (2007). Acta Crystallogr. D63, 1254–1268), using the TLSMD web server. The least squares residuals are presented for the optimal partitioning up to 20 segments (a). The three-dimensional model is shown with the optimal solution using 5 segments as highlighted by color (b). The mean B_{eq} for each residue is presented for both the experimental model (black) and the fitted TLS model with 5 segments (c).

Chapter 7

Conclusions and future perspectives

In this study, it was shown that perdeuterated lysozyme could be produced in significant amounts using the *E. coli* expression system and subsequent *in vitro* protein refolding. As described in Manuscript 1, the increment in protein yield was of more than 3-fold, compared to the previously used system of *P. pastoris*. The perdeuterated refolded HEWL variant was demonstrated to be thermally stable and enzymatically active, and to have an identical overall fold to that of unlabelled HEWL from *G. gallus*. Minor differences found in protein structure could be related to the observed changes in biophysical properties, although the deuteration and refolding effects could not be completely separated since the variants were produced in different expression systems. Ultimately, the work reported in Manuscript 1 represented the overcoming of the two primary challenges of this project: i) the production of significant quantities of perdeuterated lysozyme; ii) and the optimization of crystallization conditions to yield high quality triclinic crystals.

Manuscript 2 contributed to the clarification of questions that arose from Manuscript 1. It was valuable to quantify the isolated effects of deuteration and *in vitro* refolding to protein structure, thermal stability and activity. It was shown that the macromolecular and solvent isotope effects have distinct impacts in protein thermal stability, being consistent for each HEWL variant. Interestingly, it was found that the *in vitro* refolding disturbs protein thermal stability to a greater extent than either of the deuteration effects. This observation was corroborated by significant changes in enzymatic activity which were explained by variations in protein conformation. Surprisingly, it was observed that at the structural level, the perdeuterated refolded variant is closer to the unlabelled HEWL from *G. gallus*, than to the hydrogenated refolded protein. Additionally, to our knowledge, this work constitutes the first report of a full Asn103 peptide-plane flip observed in a crystal structure of HEWL. This structural change seems to be responsible for a different conformation of the enzyme's binding cleft which disturbs protein function. The reported work suggests that protein perdeuteration plays a central role in changing protein folding dynamics, which in this case was favorable to the production of a structure closer to the native HEWL. These observations emphasize the importance of obtaining structural data to validate *in vitro* refolding approaches and to justify eventual biophysical discrepancies observed with respect to the native variant. Finally, the effects of deuteration (macromolecular and solvent) in protein dynamics have not been extensively studied to date, however, the underlying slower dynamics could help mimic the crowded cellular environment in which proteins are naturally folded, as suggested by our data.

While HEWL remains one of the most studied enzymes in structural biology, the description of its active site configuration, in the active state, has been elusive to X-ray crystallography, due to the inability of identifying H positions from electron density maps. This lack of knowledge has hindered the interpretation of the enzyme's function, especially in terms of characterizing the interactions made by Asp52 and its subsequent role in stabilizing the product from the breakage of the substrate's glycosidic bond. The work presented in Manuscript 3 shows the orientations of important active site residues and water molecules from atomic resolution neutron data. It appears that Asp52 Oδ1 is

stabilized by a short H-bond to Asn59, while Oδ2 displays some flexibility due to dynamical and weaker H-bonds to Asn46, Asn44 and W306. Interestingly, it is also shown that the water molecules near the active site, including the catalytic water H-bonded to the protonated Glu35, are mostly transient (*i.e.*, not fully occupying their positions), emphasizing the high degree of dynamics essential to the accommodation of the enzymatic substrate. These findings demonstrate that valuable structural and dynamical information can be extracted from complete atomic resolution neutron diffraction data. Additionally, it is shown that perdeuteration can be essential for NMX studies, by allowing the measurement of complete data to higher resolutions, and also by providing relevant structural information otherwise precluded from hydrogenated or partially deuterated samples. In the context of the thesis' project, this work has contributed with the first of two complete atomic resolution neutron datasets, which could allow the description of atomic thermal motion in protein crystals. It also proved that the perdeuterated lysozyme crystals being produced possess high diffracting quality and sufficient volume for further atomic resolution neutron experiments.

Ultimately, the chief objective of this PhD thesis is the description of atomic thermal motion in protein crystals, and the assessment of the potential bias in the ADPs refined from X-ray diffraction data. Manuscript 4 attempts to answer this important question, however, the lack of a reliable model for the 100 K neutron data hinders any definitive conclusions. Nevertheless, the analysis presented for the room temperature data suggests that there are systematic differences between the refined ADPs from X-ray and neutron data, on the order of 3 $Å^2$. The TLS models show an improved and biologically more realistic fitting of the ADPs from neutron data which indicates that these might be less affected by errors stemming *e.g.*, from inaccurate modelling of disorder. Neutron data seems to allow a better understanding of structural disorder, due to the additional information from the positions of H/D atoms, especially in residues such as Arg, Trp, Asn, or Gln. Ultimately, this preliminary study highlights potential biologically relevant information which can be extracted from ADPs obtained from X-ray data, when using those from neutron data as reference. Further efforts are necessary to resolve the issues surrounding the 100 K neutron dataset. The subsequent analysis may corroborate the initial observations made for the room temperature data and provide additional knowledge regarding the potential bias of ADPs from X-ray models. It will be also important to analyze the anisotropic ADPs for D atoms, since this type of information is only possible from neutron data. To our knowledge, these data from protein crystals were not obtained before and this study represents a chance of estimating the motion of D atoms from experimental data. Importantly, because commonly X-ray crystal structures, at high resolutions, rely on the inclusion of H atoms in idealized positions and treated as riding atoms, meaning that their ADPs are fixed and strictly dependent on the ADPs for the heavier atoms to which they are bonded. This *a priori* knowledge was obtained from small molecules, however, it has not been explored by protein crystallography.

Additionally, neutron and X-ray data has been measured on hydrogenated D_2O soaked HEWL, prior to the start of this PhD project. Data reduction using XDS showed that these datasets are complete

to atomic resolution, and preliminary neutron refinements enabled the estimation of anisotropic ADPs for the majority of the protein atoms, including D atoms in exchangeable positions. Once the model refinements from X-ray data are finished, an identical analysis to that presented in Manuscript 4, will be valuable for two purposes: i) corroborate the systematic differences in X-ray and neutron derived ADPs found for perdeuterated HEWL; ii) and contribute to the current understanding of the effects of perdeuteration in protein dynamics.

References

Adrian, M., Dubochet, J., Lepault, J. & McDowall, A. W. (1984). *Nature*. **308**, 32–36.

Afonine, P. V., Mustyakimov, M., Grosse-Kunstleve, R. W., Moriarty, N. W., Langan, P. & Adams, P. D. (2010). *Acta Crystallogr.* **D66**, 1153–1163.

Anfinsen. C. B. (1973). *Science (80-.).* **181**, 223–230.

Anfinsen, C. B., Sela, M. & Cooke, J. P. (1962). *J. Biol. Chem.* **237**, 1825–1831.

Armstrong, N., Lencastre, A. De & Gouaux, E. (1999). *Protein Sci.* **8**, 1475–1483.

Artero, J., Härtlein, M., McSweeney, S. & Timmins, P. (2005). *Acta Crystallogr.* **D61**, 1541–1549.

Assmann, G. M., Wang, M. & Diederichs, K. (2020). *Acta Crystallogr.* **D76**, 636–652.

Bader, R. F. W. (1994). Atoms in Molecules - A Quantum Theory Oxford University Press.

Bai, N., Roder, H., Dickson, A. & Karanicolas, J. (2019). *Sci. Rep.* **9**, 1–15.

Batas, B. & Chaudhuri, J. B. (1996). *Biotechnol. Bioeng.* **50**, 16–23.

Batas, B., Schiraldi, C. & Chaudhuri, J. B. (1999). *J. Biotechnol.* **68**, 149–158.

Van Den Bedem, H., Bhabha, G., Yang, K., Wright, P. E. & Fraser, J. S. (2013). *Nat. Methods*. **10**, 896–902.

Berger, L. R. & Weiser, R. S. (1957). *Biochim. Biophys. Acta*. **26**, 517–521.

Bergfors, T. (2003). *J. Struct. Biol.* **142**, 66–76.

Berman, H., Westbrook, J., Feng, Z., Gilliand, G., Bhat, T., Weissig, H., Shindyalov, I. & Bourne, P. (2000). *Nucleic Acids Res.* **28**, 235–242.

Berns, D. S. (1963). *Biochemistry*. **2**, 1377–1380.

Biter, A. B., De La Peña, A. H., Thapar, R., Lin, J. Z. & Phillips, K. J. (2016). *Sci. Rep.* **6**, 1–9.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature*. **206**, 757–761.

Blake, C. C., Johnson, L. N., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1967). *Proc. R. Soc. London. Ser. B. Biol. Sci.* **167**, 378–388.

Blakeley, M. P. (2009). *Crystallography Reviews*, Vol. pp. 157–218.

Blessing, R. H. (1995). *Acta Crystallogr.* **B51**, 816–823.

Bon, C., Lehmann, M. S. & Wilkinson, C. (1999). *Acta Crystallogr.* **D55**, 978–987.

Bragg, W. L. (1913). *R. Soc.* **89**, 248–277.

Breyton, C., Gabel, F., Lethier, M., Flayhan, A., Durand, G., Jault, J. M., Juillan-Binard, C., Imbert, L., Moulin, M., Ravaud, S., Härtlein, M. & Ebel, C. (2013). *Eur. Phys. J. E*. **36**,.

Brockwell, D., Yu, L., Cooper, S., McCleland, S., Cooper, A., Attwood, D., Gaskell, S. J. & Barber, J. (2001). *Protein Sci.* **10**, 572–580.

Burgess, R. R. (2009). *Methods in Enzymology*, Vol. *463*, pp. 259–282. Elsevier Inc.

Campbell, R. A., Tummino, A., Varga, I., Milyaeva, O. Y., Krycki, M. M., Lin, S. Y., Laux, V., Haertlein, M., Forsyth, V. T. & Noskov, B. A. (2018). *Langmuir*. **34**, 5020–5029.

Chen, H., Deng, J., Cui, Q., Chanda, B. & Henzler-wildman, K. (2021). *PNAS.* **118**, e20172280118.

Chen, J. C. H., Hanson, B. L., Fisher, S. Z., Langan, P. & Kovalevsky, A. Y. (2012). *PNAS.* **109**, 15301–15306.

Chinte, U., Shah, B., DeWitt, K., Kirschbaum, K., Pinkerton, A. A. & Schall, C. (2005). *J. Appl. Crystallogr.* **38**, 412–419.

Clore, G. M. & Gronenborn, A. M. (1994). *Methods Enzymol.* **239**, 349–363.

Cowieson, N. P., Wensley, B., Listwan, P., Hume, D. A., Kobe, B. & Martin, J. L. (2006). *Proteomics*. **6**, 1750–1757.

Cuypers, M. G., Mason, S. A., Blakeley, M. P., Mitchell, E. P., Haertlein, M. & Forsyth, V. T. (2013). *Angew. Chemie - Int. Ed.* **52**, 1022–1025.

Cuypers, M. G., Trubitsyna, M., Callow, P., Forsyth, V. T. & Richardson, J. M. (2013). *Nucleic Acids Res.* **41**, 2020–2033.

D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. (2014). *Acta Crystallogr.* **F70**, 1117–1126.

D'Arcy, A., Mac Sweeney, A. & Haber, A. (2003). *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59**, 1343–1346.

Darmanin, C., Strachan, J., Adda, C. G., Ve, T., Kobe, B. & Abbey, B. (2016). *Sci. Rep.* **6**, 25345.

Dauter, Z., Jaskolski, M. & Wlodawer, A. (2010). *J. Synchrotron Radiat.* **17**, 433–444.

Debye, P. (1914). *Ann. Phys.* **43**, 49.

Deller, M. C., Kong, L. & Rupp, B. (2016). *Acta Crystallogr.* **F72**, 72–95.

Dong, A., Kendrick, B., Kreilgård, L., Matsuura, J., Manning, M. C. & Carpenter, J. F. (1997). *Arch. Biochem. Biophys.* **347**, 213–220.

Dubochet, J. & McDowall, A. W. (1981). *J. Microsc.* **124**, RP3–RP4.

Dunne, O., Weidenhaupt, M., Callow, P., Martel, A., Moulin, M., Perkins, S. J., Haertlein, M. & Forsyth, V. T. (2017). *Eur. Biophys. J.* **46**, 425–432.

Durbin, S. D. & Feher, G. (1986). *J. Cryst. Growth*. **76**, 583–592.

Efimova, Y. M., Haemers, S., Wierczinski, B., Norde, W. & van Well, A. A. (2006). *Biopolymers*. **85**, 264–273.

Epstein, H. F., Schechter, A. N., Chen, R. F. & Anfinsen, C. B. (1971). *J. Mol. Biol.* **60**, 499–508.

Epstein, L. A. & Chain, E. (1940). *Br. J. Exp. Pathol.* **21**, 339–355.

Faruqi, A. R., Cattermole, D. M., Henderson, R., Mikulec, B. & Raeburn, C. (2003). *Ultramicroscopy*. **94**, 263–276.

Fischer, M. (2021). *Q. Rev. Biophys.* **54**, 1–15.

Fisher, S. J., Blakeley, M. P., Howard, E. I., Petit-Haertlein, I., Haertlein, M., Mitschler, A., Cousido-Siah, A., Salvay, A. G., Popov, A., Muller-Dieckmann, C., Petrova, T. & Podjarny, A. (2014). *Acta Crystallogr.* **D70**, 3266–3272.

Fleming, A. (1922). *Proc. R. Soc. B*. **93**, 306–317.

Foglia, F., Hazael, R., Simeoni, G. G., Appavou, M. S., Moulin, M., Haertlein, M., Trevor Forsyth, V., Seydel, T., Daniel, I., Meersman, F. & Mcmillan, P. F. (2016). *Sci. Rep.* **6**, 1–9.

Forneris, F., Orru, R., Bonivento, D., Chiarelli, L. R. & Mattevi, A. (2009). *FEBS J.* **276**, 2833–2840.

Fraser, J. S., Van Den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N. & Alber, T. (2011). *PNAS.* **108**, 16247–16252.

Gajdos, L., Forsyth, V. T., Blakeley, M. P., Haertlein, M., Imberty, A., Samain, E. & Devos, J. M. (2020). *Glycobiology*. **00**, 1–8.

Gamble, T. R., Clauser, K. R. & Kossiakoff, A. A. (1994). *Biophys. Chem.* **53**, 15–26.

Gao, K., Oerlemans, R. & Groves, M. R. (2020). *Biophys. Rev.* **12**, 85–104.

Garman, E. (2003). *Curr. Opin. Struct. Biol.* **13**, 545–551.

Geders, T. W., Gustafson, K. & Finzel, B. C. (2012). *Acta Crystallogr.* **F68**, 596–600.

Gisriel, C., Coe, J., Letrun, R., Yefanov, O. M., Luna-Chavez, C., Stander, N. E., Lisova, S., Mariani, V., Kuhn, M., Aplin, S., Grant, T. D., Dörner, K., Sato, T., Echelmeier, A., Cruz Villarreal, J., Hunter, M. S., Wiedorn, M. O., Knoska, J., Mazalova, V., Roy-Chowdhury, S., Yang, J. H., Jones, A., Bean, R., Bielecki, J., Kim, Y., Mills, G., Weinhausen, B., Meza, J. D., Al-Qudami, N., Bajt, S., Brehm, G., Botha, S., Boukhelef, D., Brockhauser, S., Bruce, B. D., Coleman, M. A., Danilevski, C., Discianno, E., Dobson, Z., Fangohr, H., Martin-Garcia, J. M., Gevorkov, Y., Hauf, S., Hosseinizadeh, A., Januschek, F., Ketawala, G. K., Kupitz, C., Maia, L., Manetti, M., Messerschmidt, M., Michelat, T., Mondal, J., Ourmazd, A., Previtali, G., Sarrou, I., Schön, S., Schwander, P., Shelby, M. L., Silenzi, A., Sztuk-Dambietz, J., Szuba, J., Turcato, M., White, T. A., Wrona, K., Xu, C., Abdellatif, M. H., Zook, J. D., Spence, J. C. H., Chapman, H. N., Barty, A., Kirian, R. A., Frank, M., Ros, A., Schmidt, M., Fromme, R., Mancuso, A. P., Fromme, P. & Zatsepin, N. A. (2019). *Nat. Commun.* **10**, 1–11.

Givol, D., Goldberger, R. F. & Anfinsen, C. B. (1964). *J. Biol. Chem.* **239**, 3114–3117.

Grage, S. L., Keleshian, A. M., Turdzeladze, T., Battle, A. R., Tay, W. C., May, R. P., Holt, S. A., Contera, S. A., Haertlein, M., Moulin, M., Pal, P., Rohde, P. R., Forsyth, T. V., Watts, A., Huang, K. C., Ulrich, A. S. & Martinac, B. (2011). *Biophys. J.* **100**, 1252–1260.

Gruene, T., Hahn, H. W., Luebben, A. V., Meilleur, F. & Sheldrick, G. M. (2014). *J. Appl. Crystallogr.* **47**, 462–466.

Haber, E. & Anfinsen, C. B. (1962). *J. Biol. Chem.* **237**, 1839–1844.

Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O. & Forsyth, V. T. (2016). *Methods in Enzymology*, Vol. *566*, pp. 113–157.

Harp, J. M., Timm, D. E. & Bunick, G. J. (1998). *Acta Crystallogr.* **D54**, 622–628.

Harrington, W. F. & von Hippel, P. H. (1961). *Arch. Biochem. Biophys.* **92**, 100–113.

Hassell, A. M., An, G., Bledsoe, R. K., Bynum, J. M., Carter, H. L., Deng, S. J. J., Gampe, R. T., Grisard, T. E., Madauss, K. P., Nolte, R. T., Rocque, W. J., Wang, L., Weaver, K. L., Williams, S. P., Wisely, G. B., Xu, R. & Shewchuk, L. M. (2007). *Acta Crystallogr.* **D63**, 72–79.

Hattori, A., Crespi, H. L. & Katz, J. J. (1965). *Biochemistry*. **4**, 1213–1225.

Hazemann, I., Dauvergne, M. T., Blakeley, M. P., Meilleur, F., Haertlein, M., Van Dorsselaer, A., Mitschler, A., Myles, D. A. A. & Podjarny, A. (2005). *Acta Crystallogr.* **D61**, 1413–1417.

Held, J. & Van Smaalen, S. (2014). *Acta Crystallogr.* **D70**, 1136–1146.

Holdgate, G. A., Anderson, M., Edfeldt, F. & Geschwindner, S. (2010). *J. Struct. Biol.* **172**, 142–157.

Jain, R., Rice, W. J., Malik, R., Johnson, R. E., Prakash, L., Prakash, S., Ubarretxena-Belandia, I. & Aggarwal, A. K. (2019). *Nat. Struct. Mol. Biol.* **26**, 955–962.

Jasnin, M., Tehei, M., Moulin, M., Haertlein, M. & Zaccai, G. (2008). *Eur. Biophys. J.* **37**, 613–617.

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C., Duan, Y., Yu, J., Wang, L., Yang, K., Liu, F., Jiang, R., Yang, X., You, T., Liu, X., Yang, X., Bai, F., Liu, H., Liu, X., Guddat, L. W., Xu, W., Xiao, G., Qin, C., Shi, Z., Jiang, H., Rao, Z. & Yang, H. (2020). *Nature*. **582**, 289–293.

Johnson, L. N. & Phillips, D. C. (1965). *Nature*. **208**, 761–763.

Josts, I., Nitsche, J., Maric, S., Mertens, H. D., Moulin, M., Haertlein, M., Prevost, S., Svergun, D. I., Busch, S., Forsyth, V. T. & Tidow, H. (2018). *Structure*. **26**, 1072-1079.e4.

Kabsch, W. (2010). *Acta Crystallogr.* **D66**, 133–144.

Kalinin, Y., Kmetko, J., Bartnik, A., Stewart, A., Gillilan, R., Lobkovsky, E. & Thorne, R. (2005). *J. Appl. Crystallogr.* **38**, 333–339.

Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.

Katz, J. J. & Crespi, H. L. (1966). *Science (80-.).* **151**, 1187–1194.

Kehlenbeck, D.-M., Josts, I., Nitsche, J., Busch, S., Forsyth, V. T. & Tidow, H. (2019). *Biol. Chem.* **400**, 1509–1518.

Kita, A. & Morimoto, Y. (2016). *Mol. Biotechnol.* **58**, 130–136.

Kita, A. & Morimoto, Y. (2020). *J. Appl. Crystallogr.* **53**, 837–840.

Kneller, D. W., Phillips, G., O'Neill, H. M., Jedrzejczak, R., Stols, L., Langan, P., Joachimiak, A., Coates, L. & Kovalevsky, A. (2020). *Nat. Commun.* **11**, 7–12.

Koetzle, T. F. & McIntyre, G. J. (2012). *Characterization of Materials*, Vol. pp. 2192–2205.

Koritsanszky, T. S. & Coppens, P. (2001). *Chem Rev*. **101**, 1583–1627.

Koruza, K., Lafumat, B., Nyblom, M., Mahon, B. P., Knecht, W., McKenna, R. & Fisher, S. Z. (2019). *Acta Crystallogr.* **D75**, 895–903.

Koruza, K., Lafumat, B., Végvári, Knecht, W. & Fisher, S. Z. (2018). *Arch. Biochem. Biophys.* **645**, 26–33.

Koshland Jr., D. E. (1953). *Biol. Rev.* **28**, 416–436.

Kovermann, M., Rogne, P. & Wolf-Watz, M. (2016). *Q. Rev. Biophys.* **49**, 1–43.

Kriminski, S., Kazmierczak, M. & Thorne, R. E. (2003). *Acta Crystallogr.* **D59**, 697–708.

Kuhlman, B. & Raleigh, D. P. (1998). *Protein Sci.* **7**, 2405–2412.

Kumari, P., Ghosh, D., Vanas, A., Fleischmann, Y. & Wiegand, T. (2021). *PNAS*. **118**, e2012171118.

Laux, V., Callow, P., Svergun, D. I., Timmins, P. A., Forsyth, V. T. & Haertlein, M. (2008). *Eur. Biophys. J.* **37**, 815–822.

Legrand, L., Riès-Kautt, M. & Robert, C. (2002). *Acta Crystallogr.* **D58**, 1564–1567.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C.,

Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Crystallogr.* **D75**, 861–877.

Liebschner, D., Afonine, P. V., Moriarty, N. W., Langan, P. & Adams, P. D. (2018). *Acta Crystallogr.* **D74**, 800–813.

Liu, D., Mao, Y., Gu, X., Zhou, Y. & Long, D. (2021). *PNAS*. **118**, e2024725118.

Liu, X., Hanson, B. L., Langan, P. & Viola, R. E. (2007). *Acta Crystallogr.* **D63**, 1000–1008.

Madsen, A. Ø. (2006). *J. Appl. Crystallogr.* **39**, 757–758.

Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. (1995). *Nat. Struct. Biol.* **2**, 852–855.

Manzoni, F., Saraboji, K., Sprenger, J., Kumar, R., Noresson, A. L., Nilsson, U. J., Leffler, H., Fisher, S. Z., Schrader, T. E., Ostermann, A., Coates, L., Blakeley, M. P., Oksanen, E. & Logan, D. T. (2016). *Acta Crystallogr.* **D72**, 1194–1202.

Maric, S., Skar-Gislinge, N., Midtgaard, S., Thygesen, M. B., Schiller, J., Frielinghaus, H., Moulin, M., Haertlein, M., Forsyth, V. T., Pomorski, T. G. & Arleth, L. (2014). *Acta Crystallogr.* **D70**, 317–328.

Maric, S., Thygesen, M. B., Schiller, J., Marek, M., Moulin, M., Haertlein, M., Forsyth, V. T., Bogdanov, M., Dowhan, W., Arleth, L. & Pomorski, T. G. (2015). *Appl. Microbiol. Biotechnol.* **99**, 241–254.

Mason, S. A., Bentley, G. A. & McIntyre, G. J. (1984). *Neutrons in Biology*, Vol. pp. 323–334.

Matsumoto, S., Ishida, S., Araki, M., Kato, T., Terayama, K. & Okuno, Y. (2021). *Nat. Mach. Intell.* **3**, 153–160.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

McPherson, A. & Delucas, L. J. (2015). *Npj Microgravity*. **1**, 15010.

Mehrabi, P., Schulz, E. C., Agthe, M., Horrell, S., Bourenkov, G., von Stetten, D., Leimkohl, J. P., Schikora, H., Schneider, T. R., Pearson, A. R., Tellkamp, F. & Miller, R. J. D. (2019). *Nat. Methods*. **16**, 979–982.

Meilleur, F., Contzen, J., Myles, D. A. A. & Jung, C. (2004). *Biochemistry*. **43**, 8744–8753.

Meyer, K., Hahnel, E. & Steinberg, A. (1946). *J. Biol. Chem.* **163**, 733–740.

Meyer, K., Palmer, J. W., Thompson, R. & Khorazo, D. (1936). *J. Biol. Chem.* **113**, 479–486.

Millero, F. J., Dexter, R. & Hoff, E. (1971). *J. Chem. Eng. Data*. **16**, 85–87.

Miranker, A., Radford, S. E., Karplus, M. & Dobson, C. M. (1991). *Nature*. **349**, 633–636.

Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T. & Dobson, C. M. (1993). *Science (80-.).* **262**, 896–900.

Moulin, M., Strohmeier, G. A., Hirz, M., Thompson, K. C., Rennie, A. R., Campbell, R. A., Pichler, H., Maric, S., Forsyth, V. T. & Haertlein, M. (2018). *Chem. Phys. Lipids*. **212**, 80–87.

Munshi, P., Madsen, A. Ø., Spackman, M. A., Larsen, S. & Destro, R. (2008). *Acta Crystallogr.* **A64**, 465–475.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R. & Scheres, S. H. W. (2020). *Nature*. **587**, 152–156.

Nichols, P. J., Falconer, I., Griffin, A., Mant, C., Hodges, R., McKnight, C. J., Vögeli, B. & Vugmeyster, L. (2020). *Protein Sci.* **29**, 1641–1654.

Niimura, N. & Bau, R. (2008). *Acta Crystallogr.* **A64**, 12–22.

Niimura, N., Minezaki, Y., Nonaka, T., Castagna, J., Cipriani, F., Hghj, P., Lehmann, M. S. & Wilkinson, C. (1997). *Nat. Struct. Biol.* **4**, 909–914.

Nitsche, J., Josts, I., Heidemann, J., Mertens, H. D., Maric, S., Moulin, M., Haertlein, M., Busch, S., Forsyth, V. T., Svergun, D. I., Uetrecht, C. & Tidow, H. (2018). *Commun. Biol.* **1**, 1–10.

Oksanen, E., Chen, J. C.-H. & Fisher, S. Z. (2017). *Molecules*. **22**,.

Paliy, O., Bloor, D., Brockwell, D., Gilbert, P. & Barber, J. (2003). *J. Appl. Microbiol.* **94**, 580–586.

Palmer, I. & Wingfield, P. T. (2004). *Curr Protoc Protein Sci*. **6**, 1–18.

Pandey, S., Bean, R., Sato, T., Poudyal, I., Bielecki, J., Cruz Villarreal, J., Yefanov, O., Mariani, V., White, T. A., Kupitz, C., Hunter, M., Abdellatif, M. H., Bajt, S., Bondar, V., Echelmeier, A., Doppler, D., Emons, M., Frank, M., Fromme, R., Gevorkov, Y., Giovanetti, G., Jiang, M., Kim, D., Kim, Y., Kirkwood, H., Klimovskaia, A., Knoska, J., Koua, F. H.

M., Letrun, R., Lisova, S., Maia, L., Mazalova, V., Meza, D., Michelat, T., Ourmazd, A., Palmer, G., Ramilli, M., Schubert, R., Schwander, P., Silenzi, A., Sztuk-Dambietz, J., Tolstikova, A., Chapman, H. N., Ros, A., Barty, A., Fromme, P., Mancuso, A. P. & Schmidt, M. (2020). *Nat. Methods*. **17**, 73–78.

Pedersen, B. P., Buch-Pedersen, M. J., Preben Morth, J., Palmgren, M. G. & Nissen, P. (2007). *Nature*. **450**, 1111–1114.

Petit-Haertlein, I., Blakeley, M. P., Howard, E., Hazemann, I., Mitschler, A., Haertlein, M. & Podjarny, A. (2009). *Acta Crystallogr.* **F65**, 406–409.

Pflugrath, J. W. (2015). *Acta Crystallogr.* **F71**, 622–642.

Phillips, D. C. (1967). *PNAS*. **57**, 484–495.

Pillon, M. C., Frazier, M. N., Dillard, L. B., Williams, J. G., Kocaman, S., Krahn, J. M., Perera, L., Hayne, C. K., Gordon, J., Stewart, Z. D., Sobhany, M., Deterding, L. J., Hsu, A. L., Dandey, V. P., Borgnia, M. J. & Stanley, R. E. (2021). *Nat. Commun.* **12**, 1–12.

Radford, S. E., Dobson, C. M. & Evans, P. A. (1992). *Nature*. **358**, 302–307.

Ramanathan, A., Savol, A., Burger, V., Chennubhotla, C. S. & Agarwal, P. K. (2014). *Acc. Chem. Res.* **47**, 149–156.

Ramos, J., Laux, V., Haertlein, M., Boeri Erba, E., McAuley, K. E., Forsyth, V. T., Mossou, E., Larsen, S. & Langkilde, A. E. (2021). *IUCrJ*. **8**,.

Ramos, J., Muthukumaran, J., Freire, F., Paquete-Ferreira, J., Otrelo-Cardoso, A. R., Svergun, D., Panjkovich, A. & Santos-Silva, T. (2019). *Int. J. Mol. Sci.* **20**,.

Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. (2013). *Acta Crystallogr.* **F69**, 209–214.

Romoli, F., Mossou, E., Cuypers, M., Van Der Linden, P., Carpentier, P., Mason, S. A., Forsyth, V. T. & McSweeney, S. (2014). *Acta Crystallogr.* **F70**, 681–684.

Rosenblum, G., Steen, P. E. Van Den, Cohen, S. R., Gu, J., Frenkel, J., Sertchook, R., Slack, N., Strange, R. W., Opdenakker, G. & Sagi, I. (2007). *Structure*. **15**, 1227–1236.

Russi, S., González, A., Kenner, L. R., Keedy, D. A., Fraser, J. S. & Van Den Bedem, H. (2017). *J. Synchrotron Radiat.* **24**, 73–82.

Salton, M. R. J. & Ghuysen, J. M. (1959). *Biochim. Biophys. Acta*. **36**, 552–554.

Sanishvili, R. G., Margoliash, E., Westbrook, M. L., Westbrook, E. M. & Volz, K. W. (1994). *Acta Crystallogr.* **D50**, 687–694.

Sasisanker, P., Oleinikova, A., Weingärtner, H., Ravindra, R. & Winter, R. (2004). *Phys. Chem. Chem. Phys.* **6**, 1899–1905.

Sattler, M. & Fesik, S. W. (1996). *Structure*. **4**, 1245–1249.

Schmidt, A., Teeter, M., Weckert, E. & Lamzin, V. S. (2011). *Acta Crystallogr.* **F67**, 424–428.

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

Senda, M., Hayashi, T., Hatakeyama, M., Takeuchi, K., Sasaki, A. T. & Senda, T. (2016). *Cryst. Growth Des.* **16**, 1565–1571.

Sheldrick, G. M. (2015). *Acta Crystallogr.* **C71**, 3–8.

Shugar, D. (1952). *Biochim. Biophys. Acta*. **8**, 302–309.

Sieker, L. C. (1988). *J. Cryst. Growth*. **90**, 31–38.

Smits, A. H. & Vermeulen, M. (2016). *Trends Biotechnol.* **34**, 825–834.

Steinrauf, L. K. (1959). *Acta Crystallogr.* **12**, 77–79.

Strynadka, N. C. J. & James, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). *Methods Enzymol.* **185**, 60–89.

Stura, E. A. & Wilson, I. A. (1990). *Methods A Companion to Methods Enzymol.* **1**, 38–49.

Svergun, D. I., Richard, S., Koch, M. H. J., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). *PNAS*. **95**, 2267–2272.

Mac Sweeney, A. & D'Arcy, A. (2003). *J. Appl. Crystallogr.* **36**, 165–166.

Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R. & Jansonius, J. N. (1981). *J. Mol. Biol.* **147**, 465–469.

Van Tilbeurgh, H., Egloff, M. P., Martinez, C., Rugani, N., Verger, R. & Cambillau, C. (1993). *Nature*. **362**, 814–820.

Varga, K., Aslimovska, L., Parrot, I., Dauvergne, M. T., Haertlein, M., Forsyth, V. T. & Watts, A. (2007). *Biochim. Biophys.*

Acta - Biomembr. **1768**, 3029–3035.

Vidal, O., Robert, M. C., Arnoux, B. & Capelle, B. (1999). *J. Cryst. Growth*. **196**, 559–571.

Vijayakrishnan, S., Kelly, S. M., Gilbert, R. J. C., Callow, P., Bhella, D., Forsyth, T., Lindsay, J. G. & Byron, O. (2010). *J. Mol. Biol.* **399**, 71–93.

Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. & Bignon, C. (2004). *Protein Sci.* **13**, 2782–2792.

Waldie, S., Lind, T. K., Browning, K., Moulin, M., Haertlein, M., Forsyth, V. T., Luchini, A., Strohmeier, G. A., Pichler, H., Maric, S. & Cárdenas, M. (2018). *Langmuir*. **34**, 472–479.

Waldie, S., Moulin, M., Porcar, L., Pichler, H., Strohmeier, G. A., Skoda, M., Forsyth, V. T., Haertlein, M., Maric, S. & Cárdenas, M. (2019). *Sci. Rep.* **9**, 1–11.

Waldie, S., Sebastiani, F., Browning, K., Maric, S., Lind, T. K., Yepuri, N., Darwish, T. A., Moulin, M., Strohmeier, G., Pichler, H., Skoda, M. W. A., Maestro, A., Haertlein, M., Forsyth, V. T., Bengtsson, E., Malmsten, M. & Cárdenas, M. (2020). *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids*. **1865**, 158769.

Waller, I. (1923). *Zeitschrift Für Phys.* **17**, 398–408.

Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Crystallogr.* **D54**, 522–546.

Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Crystallogr.* **D63**, 1254–1268.

Weichenberger, C. X., Afonine, P. V., Kantardjieff, K. & Rupp, B. (2015). *Acta Crystallogr.* **D71**, 1023–1038.

Weik, M., Kryger, G., Schreurs, A. M. M., Bouma, B., Silman, I., Sussman, J. L., Gros, P. & Kroon, J. (2001). *Acta Crystallogr.* **D57**, 566–573.

Wenzel, M., Lenk, H.-P. & Schütte, E. (1962). *J. Biol. Chem.* **327**, 13.

Wildegger, G. & Kiefhaber, T. (1997). *J. Mol. Biol.* **270**, 294–304.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 295–315.

Willis, M. S., Hogan, J. K., Prabhakar, P., Liu, X., Tsai, K., Wei, Y. & Fox, T. (2005). *Protein Sci.* **14**, 1818–1826.

Wood, K., Gallat, F. X., Otten, R., Van Heel, A. J., Lethier, M., Van Eijck, L., Moulin, M., Haertlein, M., Weik, M. & Mulder, F. A. A. (2013). *Angew. Chemie - Int. Ed.* **52**, 665–668.

Yee, A. W., Moulin, M., Breteau, N., Haertlein, M., Mitchell, E. P., Cooper, J. B., Erba, E. B. & Forsyth, V. T. (2016). *Angew. Chemie - Int. Ed.* **55**, 9292–9296.

Zhang, M., Gui, M., Wang, Z. F., Gorgulla, C., Yu, J. J., Wu, H., Sun, Z. yu J., Klenk, C., Merklinger, L., Morstein, L., Hagn, F., Plückthun, A., Brown, A., Nasr, M. L. & Wagner, G. (2021). *Nat. Struct. Mol. Biol.* **28**, 258–267.

Zhou, X. E., Gao, X., Barty, A., Kang, Y., He, Y., Liu, W., Ishchenko, A., White, T. A., Yefanov, O., Han, G. W., Xu, Q., De Waal, P. W., Suino-Powell, K. M., Boutet, S., Williams, G. J., Wang, M., Li, D., Caffrey, M., Chapman, H. N., Spence, J. C. H., Fromme, P., Weierstall, U., Stevens, R. C., Cherezov, V., Melcher, K. & Xu, H. E. (2016). *Sci. Data*. **3**,.

Zhuravleva, A. & Korzhnev, D. M. (2017). *Prog. Nucl. Magn. Reson. Spectrosc.* **100**, 52–77.