Hydrogens and hydrogen-bond networks in macromolecular MicroED data

Max T.B. Clabbers\textsuperscript{a,b}, Michael W. Martynowycz\textsuperscript{a,b}, Johan Hattne\textsuperscript{a,b}, Tamir Gonen\textsuperscript{a,b,c,*}

\textsuperscript{a} Department of Biological Chemistry, University of California, Los Angeles, CA 90095, United States
\textsuperscript{b} Howard Hughes Medical Institute, University of California, Los Angeles, CA 90095, United States
\textsuperscript{c} Department of Physiology, University of California, Los Angeles, CA 90095, United States

\textbf{A B S T R A C T}

Microcrystal electron diffraction (MicroED) is a powerful technique utilizing electron cryo-microscopy (cryo-EM) for protein structure determination of crystalline samples too small for X-ray crystallography. Electrons interact with the electrostatic potential of the sample, which means that the scattered electrons carry information about the charged state of atoms and provide relatively stronger contrast for visualizing hydrogen atoms. Accurately identifying the positions of hydrogen atoms, and by extension the hydrogen bonding networks, is of importance for understanding protein structure and function, in particular for drug discovery. However, identification of individual hydrogen atom positions typically requires atomic resolution data, and has thus far remained elusive for macromolecular MicroED. Recently, we presented the ab initio structure of triclinic hen egg-white lysozyme at 0.87 Å resolution. The corresponding data were recorded under low exposure conditions using an electron-counting detector from thin crystalline lamellae. Here, using these subatomic resolution MicroED data, we identified over a third of all hydrogen atom positions based on strong difference peaks, and directly visualize hydrogen bonding interactions and the charged states of residues. Furthermore, we find that the hydrogen bond lengths are more accurately described by the inter-nuclei distances than the centers of mass of the corresponding electron clouds. We anticipate that MicroED, coupled with ongoing advances in data collection and refinement, can open further avenues for structural biology by uncovering the hydrogen atoms and hydrogen bonding interactions underlying protein structure and function.

\textbf{A R T I C L E  I N F O}

Keywords:
Microcrystal electron diffraction
MicroED
Cryo-EM
Hydrogens
Atomic structure

1. Introduction

Microcrystal electron diffraction (MicroED) has been successful in structure determination of crystalline biological specimens using electron cryo-microscopy (cryo-EM) \cite{Nannenga, Shi, Yonekura, Clabbers, Rodriguez, Sawaya, Xu}, including novel structures \cite{Clabbers, Rodriguez, Sawaya, Xu}, as well as difficult to crystallize membrane proteins in detergent and lipid mixtures \cite{Liu, Martynowycz, Martynowycz, Martynowycz}. As electrons interact more strongly with matter than X-rays \cite{Henderson}, the crystal volume required for useful diffraction is typically about a million times smaller. Electrons are scattered by the electrostatic potential and the strength of scattering depends on the charged state of atoms \cite{Cowley}. The effects of charge distribution are already apparent at moderate to low resolution \cite{Yonekura, Yonekura}, and the charged state of residues in macromolecules has previously been investigated using electron crystallography \cite{Kimura, Mitsuoka, Yonekura, Yonekura}. Accurately identifying the positions of hydrogen atoms and visualizing their hydrogen bonding networks are crucial for understanding protein structure and function such as resolving precise drug or ligand binding interactions \cite{Clabbers, Martynowycz, Purdy} or elucidating mechanisms for substrate transfer in membrane protein transporters and channels \cite{Gonen, Liu, Gonen}. In single-particle cryo-EM imaging, individual hydrogen atom positions were localized from reconstructions of apoferritin at 1.2 Å resolution \cite{Maki-Yonekura, Nakane, Nakane} and the GABA\textsubscript{A} receptor at 1.7 Å resolution \cite{Nakane}. Here, hydrogen atoms were identified by omitting them from the image for model building or through density modification.

Electrostatic potential maps obtained from electron scattering can provide strong contrast for identifying hydrogen atoms, which has enabled localizing hydrogens in electron diffraction structures of small molecule organics and peptide fragments \cite{Clabbers, Dorset, Gallagher-Jones, Gruene, Jones, Palatinus, Rodriguez, Sawaya}. Identifying the positions of hydrogen atoms and visualizing their resulting hydrogen bonding networks are crucial for understanding protein structure and function such as resolving precise drug or ligand binding interactions \cite{Clabbers, Martynowycz, Purdy} or elucidating mechanisms for substrate transfer in membrane protein transporters and channels \cite{Gonen, Liu, Gonen}. In single-particle cryo-EM imaging, individual hydrogen atom positions were localized from reconstructions of apoferritin at 1.2 Å resolution \cite{Maki-Yonekura, Nakane, Nakane} and the GABA\textsubscript{A} receptor at 1.7 Å resolution \cite{Nakane}. Here, hydrogen atoms were identified by omitting them from the image for model building or through density modification.

\textsuperscript{*} Corresponding author at: Department of Biological Chemistry, University of California, Los Angeles, CA 90095, United States.
\textit{E-mail address:} tgonen@g.ucla.edu (T. Gonen).

https://doi.org/10.1016/j.jsbxx.2022.100078

Received 18 August 2022; Received in revised form 1 November 2022; Accepted 7 November 2022
Available online 10 November 2022

2590-1524/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
model and inspecting the peaks in a calculated $mF_o - DF_c$ difference map following refinement in Servalcat based on crystallographic refinement routines implemented in REFMAC5 (Murshudov et al., 2011; Yamashita et al., 2021).

Visualizing hydrogen atoms in macromolecular X-ray crystallography generally requires (sub-) atomic resolution data. The accuracy of localizing hydrogens varies with local structural flexibility that is reflected by the temperature factors. Typically, crystals of macromolecules are more disordered than peptides or small molecules and have a much higher solvent content. Therefore, in absence of high-quality and atomic resolution data, identification of hydrogen atoms in macromolecular MicroED structures has thus far remained elusive.

Recently, we reported the structure of triclinic hen egg-white lysozyme at 0.87 Å resolution from electron-counted MicroED data (Martynowycz et al., 2022). MicroED data were collected from 16 crystal lamellae and the structure was phased ab initio as described previously (Martynowycz et al., 2022). Following density modification, individual atoms could be resolved at sub-Angström resolution, enabling automated model building of the entire structure without reference to a previously determined homologous model (Martynowycz et al., 2022).

The improvement in data accuracy and resolution in this study compared to previous efforts was realized by combining focused ion-beam milling to produce approximately 300 nm thin crystalline lamellae ideal for cryo-EM at 300 kV (Martynowycz et al., 2021a), and collecting data in electron-counting mode at a significantly reduced total exposure of only 0.64 eÅ$^{-2}$ per crystal (Martynowycz et al., 2022). A low exposure rate is required for electron counting as it ensures that the rate of scattered electrons remains within the linear range of the camera. Lowering the total exposure also reduces the effects of radiation damage that can affect the structural integrity of the protein and the ability to localize hydrogen atoms (Hattne et al., 2018; Leapman and Sun, 1995).

Here, we set out to further refine the $ab\ initio$ model resulting from automated building against the subatomic resolution MicroED data to closely examine the individual hydrogen atom positions. We demonstrate that over a third of all hydrogen atoms can be identified from strong difference peaks, the most complete view of a macromolecular hydrogen network visualized by electron diffraction to date. We describe the hydrogen bonding interactions that are observed, as well as the charged states of residues and hydrogen bond networks. Furthermore, analysis of the hydrogen bond lengths from the MicroED data reveals that these are more accurately described by the inter-nuclei distances. The results illustrate that MicroED can provide accurate structural information on hydrogen atoms and hydrogen bonding interactions.

2. Materials and methods

2.1. Crystallization and sample preparation

Crystalline lamellae of triclinic lysozyme were prepared as described previously (Martynowycz et al., 2022). Briefly, crystals of hen egg-white lysozyme (Gallus gallus) were grown by dissolving 10 mg/ml protein in a solution of 0.2 M sodium nitrate and 50 mM sodium acetate at pH 4.5. After incubation overnight at 4°C an opaque suspension was observed. After further incubation for one week at room temperature a crystalline slurry containing microcrystals appeared. Samples were prepared by depositing 3 μl of the crystalline slurry onto a glow-discharged EM grid (Quantifoil, Cu 200 mesh, R2/2 holey carbon). Excess liquid was blotted away and the sample was vitrified using a Leica GP2 vitrification robot. Grids were transferred to an Aquilos dual-beam FIB/SEM (Thermo Fisher) and crystals were milled to lamellae with an optimal thickness of approximately 300 nm as described previously (Martynowycz et al., 2021a; Martynowycz et al., 2022).

2.2. Data collection and processing

Electron-counted MicroED data were collected on a Titan Krios 3G TEM (Thermo Fisher) operated at 300 kV as described previously (Martynowycz et al., 2022). Briefly, the TEM was set up for low exposure data collection using a 50 μm C2 aperture, spot size 11, and a beam diameter of 25 μm. A 100 μm SA aperture was used, corresponding to an area of 2 μm diameter on the specimen. Crystalline lamellae were continuously rotated over a range of 84° at a rotation speed of 0.2°/s over 420 s with a total exposure of approximately 0.64 eÅ$^{-2}$ per dataset. Data were recorded on a Falcon 4 direct electron detector (Thermo Fisher) in electron-counting mode operating at an internal frame rate of 250 Hz. Data from 16 crystal lamellae were integrated using XDS (Kabsch, 2010) and scaled and merged in AIMLESS (Evans and Murshudov, 2013). The structure was phased $ab\ initio$ by placing a three-residue idealized α-helical fragment using PHASER (McCoy et al., 2007) followed by density modification in ACORN (Fošak et al., 2000).

The entire structure was built automatically using BUCCANEER (Cowan, 2006) and refined in REFMAC5 (Murshudov et al., 2011) using electron scattering factors.

2.3. Identification of hydrogen atoms

The structure was manually inspected and remodeled using Coot (Emsley et al., 2010), and re-refined with REFMAC5 (Murshudov et al., 2011) using electron scattering factors. Hydrogen atoms were added in idealized riding positions. A hydrogen-only omit map was calculated from the final structural model by REFMAC5 (Murshudov et al., 2011). Peaks in the $mF_o - DF_c$ difference map at a threshold $>2.0σ$ above the mean were identified and listed using PEAKMAX in the CCP4 software package (Winn et al., 2011, p. 4). Difference peaks that fell within 0.5 Å of the idealized distance for the known positions were assigned as hydrogen atoms.

2.4. Figure and table preparation

Figures were prepared using ChimeraX, and the matplotlib library in Python 3.6. Figures were arranged in PowerPoint. Tables were arranged in Excel.

3. Results

3.1. Identifying hydrogen atoms in macromolecular MicroED data

First, the structural model of triclinic lysozyme resulting from automated building was refined using electron scattering factors, isotropic atomic displacement parameters, and the default riding hydrogen model in REFMAC5 (Murshudov et al., 2011). Twelve alternate side-chain conformations were modeled upon visual inspection using Coot (Emsley et al., 2010), and their occupancies were refined. The model was then refined using anisotropic B-factors until convergence (Supplementary Table 1). A crystallographic $mF_o - DF_c$ difference map was calculated using a model without hydrogen atoms (Yamashita et al., 2021). Peaks in the difference hydrogen omit map at greater than or equal to 2.0σ above the mean were then identified using PEAKMAX (Winn et al., 2011), and those within 0.5 Å distance from any idealized riding position were identified as potential hydrogen atoms. In this manner, we located 376 out of 1067 possible hydrogen atoms corresponding to about 35% of the entire structure. Within 10 Å of the model including solvent regions, 1369 peaks are identified at a threshold $\geq2.0σ$. Lowering the threshold to 1.0σ revealed a total of 562 hydrogen atom positions, approximately 53%. At contour levels below 2.0σ, the difference map is noisier, increasing the chance of false positives and making it more challenging to unambiguously identify peaks as hydrogen atoms. Nevertheless, these results constitute the most complete hydrogen bonding network visualized to date by macromolecular
Overall, the protein main chain is expected to be more rigid than the side chains; we consequently expect more hydrogen atoms to be found in the backbone than in the protein side chains. At the 2.0σ threshold, we identified 61 out of 141 possible Ca-H hydrogens and 76 out of 127 peptide N–H hydrogen bonds corresponding to approximately 43% and 60% of the entire backbone structure, respectively (Table 1, Supplementary Tables 2, 3 and 8). The backbone hydrogen atoms are structurally important and can be involved in forming and stabilizing secondary structural elements via hydrogen-bonding interactions. For example, the structure of lysozyme has two short antiparallel β-strands and we could identify three strong difference peaks at > 3.0σ indicating the positions of those hydrogen atoms involved in hydrogen bonding interactions (Fig. 1a, Supplementary Video 1). The average N–H distance in the β-strands is 1.14(26) Å, and the distance between the amide group hydrogen donor and carbonyl acceptor is 2.76(9) Å (Table 2).

Interestingly, whereas the Asp52 and Gly54 N–H distances are close to the idealized positions, the difference peak for the Asn44 N–H is located at an almost equal distance between the donor and the Asp52 carbonyl acceptor (Fig. 1a, Table 2). The structure of lysozyme is further composed of several short helices, and we could identify a total of 15 hydrogen bonding interactions in the three major α-helices (Table 2). For example, in the longest 12-residue α-helix we identified 6 out of 10 possible hydrogen bonds based on strong difference peaks at > 2.7σ (Fig. 1b, Supplementary Video 2).

The average hydrogen atom peptide N–H distance for the α-helices is 0.97(14) Å with an average distance between donor and acceptor of 2.84(13) Å (Table 2).

Higher flexibility and alternate conformations can affect localizing hydrogen atoms in the side chains. Nevertheless, we could successfully localize side-chain hydrogen atoms in the data and identify several hydrogen-bonding interactions between side-chain atoms (Fig. 2, Supplementary Table 2). For example, a difference peak at 2.4σ can be resolved between His15-NE2 and Thr89-OG1 indicating a possible hydrogen bond based on strong difference peaks at > 2.7σ (Fig. 2a). As expected at pH 4.5, the data show the solvent-exposed histidine to be protonated at ND1, although the hydrogen distance and angle are different from idealized geometry (Fig. 2a). Another example of hydrogen bonding interactions is illustrated for Tyr53-OH acting as a hydrogen donor to Asp66-OD1 with a strong difference peak at 3.4σ (Fig. 2b, Supplementary Table 2).

In single-particle cryo-EM, it was previously observed that acidic side chains were poorly resolved at moderate to low resolution owing to radiation damage and due to the rapid falloff of the electron scattering factors for negatively charged atoms at lower scattering angles (Maki-Yonekura et al., 2021; Yonekura et al., 2018, 2015). In the MicroED data, the acidic aspartate and glutamate residues and their negatively charged side-chain carboxyl groups are generally well resolved (Fig. 2c). Additionally, clear difference peaks at > 2.3σ were identified in the data for the amide side-chain nitrogen for asparagine and glutamine residues, making it possible to clearly distinguish between the nitrogen and oxygen atoms of the side-chain amide group (Fig. 2c).

Difference peaks were also identified for several water molecules that are involved in hydrogen bonding interactions with the protein backbone and side chains (Fig. 2d, Supplementary Video 3). Such hydrogen bonding networks can act as long-range proton transfer wires. For example, a water molecule is coordinated with the adjacent Ser91, Leu56, and Tyr53 residues and shows two strong difference peaks at > 2.7σ (Fig. 2d). Two additional water molecules show hydrogen atom peaks at > 2.2σ and are involved in hydrogen bonding interactions with each other and residues of the neighboring protein backbone (Fig. 2d).

The O–H hydrogen bond lengths and angles of the water molecules are reasonably close to ideal values, except for one of the differences peaks for w1001 which is significantly shorter at 0.64 Å. The distance between the w1001-O proton donor and the Tyr53-O proton acceptor is however close to ideal values at 2.75 Å.

### 3.2. Hydrogen bond distances

The sheer numbers of hydrogen atoms visualized in this study allow us to measure and report hydrogen bond distances in a way previously not possible in cryo-EM (Supplementary Tables 2-10; Fig. 3). Electrons are scattered by the potential field generated from electron clouds and the nuclei. The peaks in an electrostatic potential map are therefore expected to reflect the inter-nuclei distances more than distances between centers of mass of electron clouds as observed in X-ray diffraction. Whereas for non-hydrogen atoms the centroids and nuclei coincide, for hydrogen atoms the centroid of the electron cloud does not match the nucleus. We refined the structure using the default riding hydrogen model, although the number of observations for each type is rather limited and the standard deviations from the mean value are quite large (Table 1, Fig. 3). Furthermore, elongation of the hydrogen bond model based on hydrogen distances between the electron cloud centroids using restraints derived from X-ray scattering. We analyzed the identified hydrogen atom difference peaks in the data at ≥ 2.0σ and calculated the average distance for each of the hydrogen bond types (Table 1, Fig. 3, Supplementary Tables 2-10). The interpretation for some bond types is insufficient for a rigorous statistical analysis. We do however find an average Ca-H distance for the main chain of 1.11(13) Å for 61 hydrogen bonds, compared to idealized values of 0.98 and 1.10 Å for X-ray and neutron diffraction, respectively (Table 1, Fig. 3, Supplementary Table 3). The average distance for all N–H bonds is 1.03(16) Å for 83 observations, compared to idealized values of 0.86 and 1.01 Å for X-ray and neutron diffraction, respectively (Supplementary Tables 2 and 8). Interestingly, the distances for the amide N–H bonds that are involved in hydrogen bonding interactions with neighboring residues are slightly longer compared to those that are not involved in such electrostatic interactions (Table 1, Fig. 3).

These results suggest an elongation of the hydrogen bond lengths compared to the electron cloud centroid distances assumed in the riding hydrogen model, although the number of observations for each type is rather limited and the standard deviations from the mean value are quite large (Table 1, Fig. 3). Furthermore, elongation of the hydrogen bond distances may be reflected by higher B-factors (Nakane et al., 2020; Yamashita et al., 2021), or may represent some excited state (Gallagher-Jones et al., 2018). Nevertheless, we find an overall trend that the Co-H and N–H bond lengths are closer inter-nuclei distances (Gruene et al., 2014; Williams et al., 2018). This observation agrees with previous

### Table 1

| Hydrogen bonds | No. observations | X-H(Å) | X-H<sub>ray</sub>(Å) | X-H<sub>neutron</sub>(Å) |
|----------------|------------------|--------|----------------------|-------------------------|
| Cα-H          | 61               | 1.11   | 0.98                 | 1.10                    |
| Cα<sub>amide-H</sub> | 14          | 1.25   | 0.98                 | 1.10                    |
| Cα<sub>amine-H</sub> | 17          | 1.13   | 0.93                 | 1.08                    |
| CH<sub>2</sub>   | 99               | 1.17   | 0.97                 | 1.09                    |
| CH<sub>3</sub>   | 77               | 1.09   | 0.96                 | 1.06                    |
| N–H            | 44               | 1.02   | 0.86                 | 1.01                    |
| N–H<sub>O</sub> | 38               | 1.05   | 0.86                 | 1.01                    |
| NH<sub>2</sub>   | 13               | 1.08   | 0.89                 | 1.03                    |
| NH<sub>2</sub>   | 3                | 1.13   | 0.86                 | 1.01                    |
| O–H            | 10               | 1.13   | 0.82                 | 0.98                    |

<sup>a</sup> Mean observed hydrogen bond lengths measured for hydrogen atoms difference peaks at ≥ 2.0σ, standard deviations are listed in parenthesis. Values for individual hydrogen bond distances are listed in Supplementary Tables 2-10.

<sup>b</sup> Idealized hydrogen bond lengths between electron cloud centroids used in X-ray diffraction (Gruene et al., 2014).

<sup>c</sup> Idealized inter-nuclei hydrogen bond lengths used in neutron diffraction (Gruene et al., 2014).
Fig. 1. Hydrogen atoms and bonding interactions in secondary structure elements. Difference peaks for individual hydrogen atoms are displayed as green spheres with their $\sigma$ values shown for (a) two short anti-parallel $\beta$-strands (residues 42–45 and 51–54, respectively), and (b) an $\alpha$-helix (residues 88–101). Hydrogen atoms were assigned from a hydrogen-only omit map for peaks at $\geq 2.0\sigma$ that are within 0.5 Å distance from their idealized position. Hydrogen bonding interactions are indicated by dashed black lines and their respective bond distances and angles are listed in Table 2. Electrostatic potential 2mFo-DFc maps are contoured at 4.0 $\sigma$(blue) and mFo-DFc difference maps are shown at 2.5 $\sigma$(red). 

Table 2: Hydrogen bond distances and angles for secondary structure.

| Donor-H–Acceptor | Diff. peak $\sigma$ (Å) | D-H (Å) | H–A (Å) | D–A (Å) | D-H–A (°) |
|------------------|------------------------|---------|---------|---------|-----------|
| $\beta$-strands  |                        |         |         |         |           |
| Asn$^{34}$–N–H–Asp$^{72}$, O       | 3.13       | 1.41   | 1.46   | 2.86   | 170.74    |
| Asp$^{52}$–N–H–Asn$^{54}$, O       | 4.33       | 0.90   | 1.93   | 2.75   | 149.88    |
| Gly$^{94}$–N–H–Thr$^{95}$, O       | 4.48       | 1.10   | 1.59   | 2.67   | 168.44    |
| $\alpha$-helices |                        |         |         |         |           |
| Ala$^{10}$–N–H–Glu$^{27}$, O       | 2.58       | 0.78   | 2.40   | 2.89   | 122.19    |
| Met$^{14}$–N–H–Leu$^{18}$, O       | 2.44       | 0.90   | 1.87   | 2.70   | 153.21    |
| Lys$^{28}$–N–H–Ala$^{30}$, O       | 3.44       | 1.21   | 1.59   | 2.76   | 162.32    |
| Arg$^{26}$–N–H–Ala$^{28}$, O       | 2.39       | 1.08   | 2.02   | 2.82   | 128.88    |
| Val$^{35}$–N–H–Leu$^{36}$, O       | 2.09       | 0.77   | 2.39   | 2.94   | 129.35    |
| Cys$^{42}$–N–H–Gly$^{50}$, O       | 2.45       | 1.00   | 1.85   | 2.74   | 146.82    |
| Ala$^{38}$–N–H–Asn$^{72}$, O       | 3.23       | 1.10   | 1.87   | 2.79   | 138.11    |
| Lys$^{74}$–N–H–Val$^{85}$, O       | 3.03       | 0.86   | 2.01   | 2.81   | 153.99    |
| Phe$^{74}$–N–H–Cys$^{70}$, O       | 3.21       | 0.79   | 2.18   | 2.95   | 169.63    |
| Val$^{82}$–N–H–Ile$^{86}$, O       | 3.00       | 0.98   | 1.79   | 2.77   | 174.55    |
| Asn$^{93}$–N–H–Thr$^{95}$, O       | 3.67       | 0.96   | 1.79   | 2.74   | 168.88    |
| $\alpha$-helices |                        |         |         |         |           |
| Ala$^{109}$–N–H–Ser$^{118}$, O      | 2.74       | 0.87   | 1.86   | 2.73   | 172.72    |
| Lys$^{146}$–N–H–Val$^{152}$, O      | 4.21       | 1.05   | 1.76   | 2.81   | 174.10    |
| Ile$^{153}$–N–H–Cys$^{156}$, O      | 3.04       | 1.13   | 1.72   | 3.15   | 152.39    |
| Val$^{155}$–N–H–Ala$^{157}$, O      | 3.06       | 1.09   | 1.95   | 3.03   | 171.83    |

4. Conclusions

The results demonstrate that hydrogen atom positions can be accurately identified in macromolecular MicroED data. As with X-ray crystallography, this will typically require atomic resolution data or better (Dauter et al., 1997; Howard et al., 2004; Kosinska Eriksson et al., 2013; Ogata et al., 2015; Walsh et al., 1998; Wang et al., 2007). In comparison, the structure of triclinic lysozyme was determined previously using X-ray diffraction at 120 K and room temperature to 0.93 and 0.95 Å resolution, respectively (Walsh et al., 1998). The single-crystal low-temperature X-ray structure is of high quality and generally has more clearly visible hydrogen atoms than the room temperature model merged from three crystal datasets. Difference maps contoured at 1.9$\sigma$ show hydrogen atoms in residues within the better-defined regions of the structure, and at 1.8$\sigma$ contour level, 77 out of 127 peptide N–H atoms (61 %) are identified (Walsh et al., 1998). The number of hydrogen atoms localized in the low-temperature structure is similar to the MicroED structure at comparable resolution, even though the intensity and model statistics are worse (Supplementary Table 1, Supplementary Fig. 1) (Martynowycz et al., 2022). Out of a total of 112 waters in the MicroED structure, 65 are within 1 Å distance from waters located in the X-ray map. Within 2 Å, a total of 97 waters match with those located using the X-ray data. The lower accuracy of the MicroED data can in part be attributed to non-isomorphism from merging of 16 crystal datasets and lower completeness in the highest resolution shells (Supplementary Fig. 1). Additional factors that contribute to the errors are multiple scattering interactions and absorption that can affect the accuracy of the intensities and increase the background noise. Furthermore, inaccurate modeling of the electron form factors and the electrostatic potential in structure refinement can contribute to higher model R-factors. Compared to X-ray diffraction, electrons are expected to provide better contrast for identifying hydrogen atoms at a similar resolution as the scattering factors fall off less steeply with decreasing atomic number. The lighter hydrogen atoms are therefore expected to be better resolved next to the heavier atoms, which might explain why we can identify many hydrogen atoms even though the MicroED data appear noisier. This is further supported by a comparison between apoferritin models from X-ray crystallography and single-particle cryo-EM showing that hydrogen atoms are more clearly visible in the latter (Yamashita et al., 2021). More recently, a significantly higher resolution structure of triclinic lysozyme was solved.
ab initio at 0.65 Å by X-ray diffraction (Wang et al., 2007). At this resolution, approximately 31 % of all hydrogen atoms in main and side chains could be identified at 3.0e or higher. We would anticipate major improvements in hydrogen atom localization in MicroED data upon further improving data quality and increasing the resolution.

Previously, hydrogen atoms were successfully identified in protein complexes by single-particle cryo-EM. In comparison to the results presented here, these studies reported that about 70 % of the expected number of hydrogen atoms could be identified above a threshold level of 2.0e using hydrogen-only omit maps from atomic resolution reconstructions of apoferritin (Maki-Yonekura et al., 2021; Yamashita et al., 2021). Remarkably, about 17 % of possible hydrogen atoms could be identified from data as low as 1.84 Å resolution (Yamashita et al., 2021). In imaging, the phase information is retained during reconstruction, and images are filtered to remove noise and to select a specific conformational state. The resolution is therefore a local feature of the map whereas the B-factor is a global parameter applied in map sharpening or blurring. This is unlike a crystallographic map, where the structural flexibility or disorder is modeled locally using alternate conformations and per-atom refined B-factors. In crystallography, resolving detailed features such as hydrogen atoms is affected by local conformational state. The resolution is therefore a local feature of the map whereas the B-value is a global parameter applied in map sharpening or blurring. This is unlike a crystallographic map, where the structural flexibility or disorder is modeled locally using alternate conformations and per-atom refined B-factors. In crystallography, resolving detailed features such as hydrogen atoms is affected by local conformation and per-atom refined B-factors.

Energy filtering does not exclude multiple elastic scattering which may affect the measured kinematic intensities (Cowley, 1995; Fujiwara, 1959). For any typical hydrated protein crystal, these effects are suggested to be far less detrimental to data quality compared to inelastic scattering (Lyatechko and Abrahams, 2019; Martynowycz et al., 2021a). Dynamical structure refinement can enhance the localization of hydrogen atoms in small molecule structures (Palatinus et al., 2017), but its implementation is computationally expensive and has yet to be extended to macromolecules that include bulk solvent. In recent experiments, recording MicroED data using a direct electron detector in electron-counting mode significantly improved data quality, and we...
Data availability

Coordinates and structure factors have been deposited to the PDB under accession code 7ULY. Maps have been deposited to the EMDB under accession code EMD 26596.

Acknowledgements

This study was supported by the National Institutes of Health P41GM136508, and the Department of Defense HDTRA1-21-1-0004. The Gonen laboratory is supported by funds from the Howard Hughes Medical Institute. M.T.B.X collected data, processed data, analyzed data, wrote the paper, prepared figures. M.W.M collected data, processed data, analyzed data, wrote the paper, prepared figures. J.H. processed data, analyzed data, wrote the paper. T.G conceived the project, wrote the manuscript, prepared figures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsb.2022.100078.

References

Clabbers, M.T.B., Gruene, T., van Genderen, E., Abrahams, J.P., 2019. Reducing dynamical electron scattering reveals hydrogen atoms. Acta Crystallogr. Sect. Found. Adv. 75, 82–93. https://doi.org/10.1107/S2653273519001316.

Clabbers, M.T.B., Fisher, S.Z., Coisson, M., Zou, X., Xu, H., 2020. Visualizing drug binding interactions using microcrystal electron diffraction. Commun. Biol. 3, 417. https://doi.org/10.1038/s42003-020-01155-1.

Clabbers, M.T.B., Holmes, S., Muusse, T.W., Vajhaila, P.R., Thygesen, S.J., Malde, A.K., Hunter, D.J.B., Croll, T.I., Fucekiger, L., Nanton, J.D., Rahaman, M.H., Aquila, A., Hunter, M.S., Liang, M., Yoon, C.H., Zhao, J., Zatepina, N.A., Abbey, B., Siersecki, E., Gambin, Y., Stacey, K.J., Darmanin, C., Kobe, B., Xu, H., Ve, T., 2021. MyD88 TIR domain higher-order assembly interactions revealed by microcrystal electron diffraction and serial femtosecond cryocrystallography. Nat. Commun. 12, 2578. https://doi.org/10.1038/s41467-021-22590-6.

Cowley, J.M., 1995. Diffraction physics. North-Holland, Amsterdam.

Cowan, K., 2006. The Baxcunner software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. 62, 1002–1011. https://doi.org/10.1107/S0974440906022116.

Dauter, Z., Lanzin, V.S., Wilson, K.S., 1997. The benefits of atomic resolution. Curr. Opin. Struct. Biol. 7, 681–688. https://doi.org/10.1016/S0959-440X(97)80078-4.

Dorset, D.L. (Ed.), 1995. Structural Electron Crystallography. Springer US, Boston, MA.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0974474510007495.

Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. 69, 1204–1214. https://doi.org/10.1107/S0974440713006061.

Foadi, J., Woolfson, M.M., Dodson, E.J., Wilson, K.S., Jia-xing, Y., Chao-de, Z., 2000. A flexible and efficient procedure for the solution and phase refinement of protein structures. Acta Crystallogr. D Biol. Crystallogr. 56, 1137–1147. https://doi.org/10.1107/S097444070000932K.

Fujikawa, K., 1999. Application of higher order Born approximation to multiple elastic scattering of electrons by crystals. J. Phys. Soc. Jpn. 14, 1513–1524. https://doi.org/10.1143/JPSJ.14.1513.

Gallagher-Jones, M., Glyn, C., Boyer, D.R., Martynowycz, M.W., Hernandez, E., Miao, J., Zee, C.-T., Novikova, I.V., Goldschmidt, J., McFarlane, H.T., Helgesen, G.F., Evans, J.E., Sawaya, M.R., Cascarino, D., Eisenberg, D.S., Kon, T., Rodriguez, J.A., 2018. Sub-Ångström cryo-EM structure of a prion protofibril reveals a polar clasp. Nat. Struct. Mol. Biol. 25, 131–134. https://doi.org/10.1038/s41594-017-0018-0.

Gonen, T., Cheng, V., Sitz, F., Hiroski, Y., Fujivoshi, Y., Harrison, S.C., Walz, T., 2005. Lipid-protein interactions in double-layered two-dimensional AQP0 crystals. Nature 438, 633–638. https://doi.org/10.1038/nature04321.

Gruene, T., Hahn, H.W., Laebben, A.V., Meilleur, F., Sheldrick, G.M., 2014. Refinement of macromolecular structures against neutron data with SHEXL2013. J. Appl. Crystallogr. 47, 466–470. https://doi.org/10.1107/S1600576713027659.

Gruene, T., Wennmacher, J.T.C., Zaubitzer, C., Holstein, J.J., Heidler, J., Fecteau-Lebreton, A., De Carlo, S., Müller, E., Golde, K.N., Regeni, I., Li, T., Santioso-Quinones, G., Steinfeld, G., Handschin, S., van Genderen, E., van Bokhoven, J.A., Clever, G.H., Pantelic, R., 2018. Rapid Structure determination of microcrystalline molecular compounds using electron diffraction. Angew. Chem. Int. Ed. 57, 16313–16317. https://doi.org/10.1002/anie.201811318.

Guo, H., Franken, E., Deng, Y., Benelkhir, S., Single Lexzcano, G., Janssen, B., Yu, L., Ripstein, Z.A., Tan, Y.Z., Rubinstein, J.L., 2020. Electron-event representation data enable efficient cryoEM file storage with full preservation of spatial and temporal resolution. IUCrJ 7, 860–869. https://doi.org/10.1107/S205522252000929X.

expect further benefits from faster readout and better electron-counting algorithms using electron-event representation (Guo et al., 2020; Nakane et al., 2020).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Kimura, Y., Vassylyev, D.G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., Henderson, R., 1995. The potential and limitations of neutrons, electrons and X-rays for high-resolution electron crystallography. Nature 378, 152–156. https://doi.org/10.1038/31586-020-2829-0.

Nannenga, B.L., Shi, D., Hatte, J., Reyes, E.F., Gonen, T., 2014a. Structure of catalase determined by MicroED. eLife 3, e03600.

Nannenga, B.L., Shi, D., Leslie, A.G.W., Gonen, T., 2014b. High-resolution structure determination by continuous-rotation data collection in MicroED. Nat. Methods 11, 927–930. https://doi.org/10.1038/nmeth.3043.

Ogata, H., Nishikawa, K., Lubitz, W., 2015. Hydrogen detected by subatomic resolution protein crystallography in a [NiFe] hydrogenase. Nature 520, 571–574. https://doi.org/10.1038/nature14110.

Palatinus, L., Brázdíl, P., Bouillay, P., Perez, O., Klementová, M., Petit, S., Eigner, V., Zaarour, M., Míntova, S., 2017. Hydrogen positions in single nanocrystals revealed by electron diffraction. Science 355, 166–169. https://doi.org/10.1126/science.aak9652.

Purdy, M.D., Shi, D., Chrustowicz, J., Hatte, J., Gonen, T., Yeager, M., 2018. MicroED structures of HIV-1 Gag CTD-SP1 reveal binding interactions with the maturation inhibitor bevirimat. Proc. Natl. Acad. Sci. 115, 13258–13263. https://doi.org/10.1073/pnas.1806806115.

Rodríguez, J.A., Ivanova, M.I., Sawaya, M.R., Cascio, D., Reyes, E.F., Shi, D., Sangwan, S., Guenther, E.L., Johnson, L.M., Zhang, M., Jiang, L., Arbing, M.A., Nannenga, B.L., Hatte, J., Whitelegge, J., Brewster, A.S., Messerschmidt, M., Boutet, S., Sauter, N.K., Gonen, T., Eisenberg, D.S., 2015. Structure of the core toxic of α-synuclein from invisible crystals. Nature 525, 486–490. https://doi.org/10.1038/nature15368.

Sawaya, M.R., Rodríguez, J., Cascio, D., Collazo, M.J., Shi, D., Reyes, E.F., Hatte, J., Gonen, T., Eisenberg, D.S., 2016. Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. Proc. Natl. Acad. Sci. 113, 11222–11226. https://doi.org/10.1073/pnas.1602871113.

Shi, D., Nannenga, B.L., Iadanza, M.G., Gonen, T., 2013. Three-dimensional electron crystallography of protein microcrystals. eLife 2, e01345.

Walsh, M.A., Schneider, T.R., Sieker, L.C., Dauter, Z., Lázár, V., Wang, J., Dauter, M., Alkhare, R., Joachimiak, A., Dauter, Z., 2007. Triclinic lysozyme at 0.65 Å resolution. Acta Crystallogr. D Biol. Crystallogr. 63, 1254–1268. https://doi.org/10.1107/S0907444907504224.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prianti, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, J.B., Chen, V.B., Jain, S., Lewis, S.M., Arendall, W.B., Snavely, J., Adams, P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C., 2018. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 27 (1), 293–315.

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., Nicholson, S.J., Murshudov, G.N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP 4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. 67, 235–242. https://doi.org/10.1107/S0907444910507794.

Xu, H., Lebrette, H., Clabbers, M.T.B., Zhao, J., Giuse, J.J., Zou, X., Högberg, M., 2019. Solving a new R2lox protein structure by microcrystal electron diffraction. Sci. Adv. 5, eaaq4621. https://doi.org/10.1126/sciadv.aaq4621.

Yamashita, K., Palmer, C.M., Burnley, T., Murshudov, G.N., 2021. Cryo-EM single-particle structure refinement and map calculation using Servalcat. Acta Crystallogr. Sect. Struct. Biol. 77, 1282–1291. https://doi.org/10.1107/S0907444921009475.

Yonekura, K., Matsuoka, R., Yamashita, Y., Tomita, M., Toyoshima, C., 2015. Electron crystallography of ultrathin 3D protein crystals: Atomic model with charges. Proc. Natl. Acad. Sci. 112, 3368–3373. https://doi.org/10.1073/pnas.1500724112.

Yonekura, K., Matsuoka, R., Yamashita, Y., Tomita, I., Ikekuchi, M., Kidera, A., Maki-Yonekura, S., 2018. Ionic scattering factors of atoms that compose biological molecules. J. Phys. Chem. Lett. 9, 348–353. https://doi.org/10.1021/acs.jpclett.8b00527.

Yonekura, K., Ishikawa, T., Maki-Yonekura, S., 2019. A new cryo-EM system for electron 3D crystallography by eEDF. J. Struct. Biol. 206, 243–253. https://doi.org/10.1016/j.jsb.2019.03.009.