Theoretical 3D ED electrostatic potential density maps of proteins modeled with multipolar pseudoatom data bank

Marta Kulik, Michał Leszek Chodkiewicz and Paulina Maria Dominiak *

Biological and Chemical Research Centre, Department of Chemistry, University of Warsaw, Żwirki i Wigury 101, Warszawa, 02-089, Poland.

E-mail: pdomin@chem.uw.edu.pl

electron diffraction; multipolar scattering factors; electrostatic potential density maps; transferable aspherical atom model (TAAM); micro-electron diffraction; cryo-electron microscopy; electron microscopy; macromolecular structure; proteins; quantum crystallography

Abstract

The availability of atomic resolution experimental maps of electrostatic potential from 3D electron diffraction (3D ED) extends the possibility of investigating the electrostatic potential beyond the determination of the non-hydrogen atom positions. However, accurate tools to calculate this potential for macromolecules, without reaching to expensive quantum calculations, are lacking. The University at Buffalo Data Bank (UBDB) gathers atom types that can be used to calculate the accurate electrostatic potential maps via structure factor calculations. Here, we apply the Transferable Aspherical Atom Model with UBDB to investigate the theoretically obtained potential maps for lysozyme and proteinase K and compare them with experimental maps from 3D ED. UBDB reproduces the molecular electrostatic potential of molecules within

their entire volume better than the neutral spherical models used in the popular Independent Atom Model. Additionally, we calculate the electron density maps for the studied proteins. The atomic displacement parameters affect the shape of the electrostatic potential density maps to a much higher extent than the electron density maps. The computational method presented in this study could potentially facilitate the interpretation of the less-resolved regions of the cryo-electron density maps and pave the way for distinguishing between different ions/water molecules in the active sites of macromolecules in high resolution structures, which is of interest for drug design purposes.

1. Introduction

The enormous advancement in the field of cryo-electron microscopy (cryo-EM) expanded the possibilities of obtaining high resolution structures (Kühlbrandt, 2014). A similar progress is noticeable in 3D electron crystallography (3D ED) methods, in particular in micro-electron diffraction (microED) that uses nanocrystals as a sample (Shi *et al.*, 2013; Nannenga *et al.*, 2014; Nannenga & Gonen, 2018). At the same time, deep understanding and theoretical modeling of the density maps derived in all those experiments lags behind.

From the physical point of view, the observed density is the electrostatic potential of the studied sample. The map of the same electrostatic potential should be extracted from 3D ED experiments. The electrostatic potential map is shaped by the electrons scattered both by the positively charged atom nuclei and the negatively charged electron cloud. In contrast, the electron density maps obtained in X-ray crystallography are shaped by the X-rays scattered only by the negatively charged electron cloud (Marques *et al.*, 2019). It is worth to note that the electrons are scattered by the

matter more efficiently than X-rays (Dorset, 1991), thus smaller amount of sample and shorter time of radiation is needed in electron diffraction than in X-ray diffraction experiments. Scattering of electrons depends on atomic charges and scattering angles: the amplitudes are always positive for non-negatively charged atoms but for negatively-charged atoms at low scattering angles the amplitude values become negative (Marques *et al.*, 2019). Thus, the electron scattering factors of charged atoms need careful treatment and parametrization (Yonekura & Maki-Yonekura, 2016). As a result, the obtained electrostatic potential maps may have negative or zero values at the negatively charged functional groups, for example it was observed that the amplitudes of the peaks corresponding to the phosphate groups in RNA are significantly smaller than the peaks representing their bases (Wang & Moore, 2017).

Frequently used and the simplest model applied in X-ray diffraction, called Independent Atom Model (IAM), is based on spherical scatterers located at atom positions. More sophisticated and advanced methods involve aspherical modeling of the scatterer, such as using the multipole expansion in spherical harmonics. One of those methods, based on the Hansen–Coppens equation for modeling the electron density, served as the cornerstone of the data bank of aspherical atom types called University at Buffalo Data Bank (UBDB) (Dominiak *et al.*, 2007; Jarzembska & Dominiak, 2012; Kumar *et al.*, 2019). Currently, the successor of UBDB is developed under the name Multipolar Atom Types from Theory and Statistical clustering (MATTS) data bank (soon to be published). Apart from UBDB, two other data banks of the electron density parameters for atom types used in X-ray crystallography have gained significant popularity: ELMAM (Pichon-Pesme *et al.*, 2004; Domagała *et al.*, 2012) and Invariom (Dittrich *et al.*, 2013). Usage of such atom types to recreate the electron density of the sample in an accurate way is justified as its parameters, derived from theoretical or experi-

mental atom positions in one chemical environment, can be used in a similar chemical environment - these are transferable aspherical atom model (TAAM) parameters. The superiority of TAAM over the simple IAM has been well proven over the years (Bąk *et al.*, 2011; Dittrich *et al.*, 2006; Dittrich *et al.*, 2013; Jha *et al.*, 2020). TAAM was also used to provide a deeper understanding of the electrostatic interactions within many protein and nucleic acid systems (Malińska *et al.*, 2014; Kulik *et al.*, 2015; Kumar & Dominiak, 2021).

Here, we study the electrostatic potential density maps of two model proteins, solved with 3D ED at relatively high resolution, close to 1.8 Å. The experimental density maps are compared with the maps calculated using IAM and TAAM, based on the UBDB atom types. We can compare their features and relate them to the features visible in electron density maps calculated at the same resolution, including thermal smearing effects.

2. Methods

2.1. Theoretical background

In general, in the spherical models, the electron scattering factors are customarily approximated as a sum of Gaussians. On the other hand, the multipole model based on the Hansen-Coppens formalism uses the sum of Slaters and spherical harmonics to parametrize the electron density, that corresponds to the spherical Bessel functions and spherical harmonics to parametrize the X-ray scattering factors and then the analytical expressions for the electron scattering factors can be derived. This is the reason why the electron scattering factors obtained within the Hansen-Coppens formalism are exact, in contrast to the approximations with the Gaussian fitting.

2.1.1. Independent Atom Model (IAM). The scattering potential and the electrostatic potential produced by the electrons scattered on a sample are considered equivalent (Peng, 1999). High energy elastic electron scattering on a group of well-separated atoms generates the Coulomb electrostatic potential $V(\mathbf{r})$ that depends not only on the distribution of the electron density $\rho_n(\mathbf{r}')$ but also on the positions of the atomic nuclei $\mathbf{R}_n$ and the atomic number Z (Ghermani *et al.*, 1993):

$$V(\mathbf{r}) = \sum_n \left(\frac{Z}{|\mathbf{r} - \mathbf{R}_n|} - \int_0^\infty \frac{\rho_n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}' - \mathbf{R}_n|} d^3\mathbf{r}'\right) \tag{1}$$

According to the kinematical approximation, we assume that the electron scattering amplitudes are proportional to the Fourier transform of the potential distribution (Cowley *et al.*, 2006). To calculate the spherical electron scattering factors f^e_{IAM}, it is possible to use the spherical X-ray scattering factors f^x_{IAM} calculated with a quantum mechanical method such as the atomic multiconfiguration Dirac-Fock code (Rez *et al.*, 1994) and then use the Mott-Bethe formula based on the Born approximation (Mott & Massey, 1965):

$$f^e_{IAM}(\mathbf{h}) = \frac{m_0 e^2}{8\pi^3 \hbar^2 \epsilon_0} \frac{Z - f^x_{IAM}(\mathbf{h})}{\mathbf{h}^2} \tag{2}$$

Here, $|\mathbf{h}| = \frac{2sin(\theta)}{\lambda}$, where θ and λ stand for one half of the scattering angle and the electron wavelength, respectively. m_0 and e are the mass and electron charge, whereas ϵ_0 is the permittivity of vacuum. The electron scattering factors have been parametrized for all neutral atoms (Peng *et al.*, 1996) and are gathered in the International Tables for Crystallography (2006), Vol. C, as the a_i and b_i values of the approximations as sums of five Gaussians:

$$f^e_{IAM}(s) = \sum_{i=0}^{n} a_i exp(-b_i s^2), \tag{3}$$

where $s = \frac{sin(\theta)}{\lambda}$.

2.1.2. Transferable Aspherical Atom Model (TAAM). First, let us look at the multipole model of electron density, based on the Hansen Coppens equation, in which the total atom charge density is divided into three terms: spherical core and valence electrons terms ρ_{core} and ρ_{val}, represented by Slater-type functions, and aspherical multipole expansion term, represented by both Slater-type radial functions (R_l) and a finite spherical harmonic expansion in a nucleus-centered local frame (Y_{lm}) (Hansen & Coppens, 1978):

$$\begin{aligned}\rho_{TAAM}(\mathbf{r}) = P_{core}\rho_{core}(r) + P_{val}\kappa^3\rho_{val}(\kappa r) \\ + \sum_{l=0}^{l_{max}} \kappa'^3 R_l(\kappa' r) \sum_{m=-l}^{l} P_{lm} Y_{lm}(\theta,\phi)\end{aligned} \tag{4}$$

The ρ_{core} and $\rho_{valence}$ terms are spherically averaged, normalized to one electron, and have to be multiplied by P_{core} and P_{val} parameters, representing the electron populations of core and valence electrons. P_{lm} represents the population of multipole densities. κ and $\kappa^{'}$ parameters reflect the expansion and contraction of the spherical valence shell and the aspherical part, respectively. The atomic multipolar scattering factors for X-ray scattering $f^x_{TAAM}(\mathbf{h})$ can be derived directly basing on the parameters from the Hansen-Coppens equation (Hansen & Coppens, 1978):

$$\begin{aligned}f^x_{TAAM}(\mathbf{h}) = P_{core}f_{core}(h) + P_{val}f_{val}(\frac{h}{\kappa}) \\ +4\pi \sum_{l=0}^{l_{max}} i^l J_l(\frac{h}{\kappa'}) \sum_{m=-l}^{l} P_{lm} Y_{lm}(\theta,\phi)\end{aligned} \tag{5}$$

The $f_{core}(h)$ and $f_{val}(\frac{h}{\kappa})$ stand for the atomic form factors from the core and spherically-averaged valence electron densities, whereas the $J_l(\frac{h}{\kappa'})$ denote the l-th order Fourier-Bessel transforms of Slater radial functions. Then, the Mott-Bethe formula (Equation 2) can be used to transform the aspherical X-ray scattering factors f^x_{TAAM} to the aspherical electron scattering factors f^e_{TAAM} , in a similar way as in the

IAM but without using approximations with Gaussian fitting.

2.1.3. Structure factors. For both IAM and TAAM, the structure factors $F_e(\mathbf{h})$ for a crystal in electron diffraction experiment can be expressed in the following way (Chodkiewicz *et al.*, 2018):

$$F^e(\mathbf{h}) = \sum_{a \in asu} occ_a m_a \sum_{\{\mathbf{R}_k|\mathbf{t}_k\}} f^e(\mathbf{R}_k^\mathsf{T}\mathbf{h}) T(\mathbf{R}_k^\mathsf{T}\mathbf{h}) exp[2\pi i \mathbf{h}^\mathsf{T}(\mathbf{R}_k\mathbf{r}_a + \mathbf{t}_k)] \qquad (6)$$

The sum of contributions from each atom a in the asymmetric unit with occupancy occ_a, multiplicity m_a and the sum of contributions from each symmetry-equivalent atom using symmetry operations $\{\mathbf{R}_k|\mathbf{t}_k\}$ are calculated for the atoms with the aspherical form factor $f_a(\mathbf{R}_k^\mathsf{T}\mathbf{h})$. Note, that this equation can be also applied to the spherical form factor case. $T(\mathbf{R}_k^\mathsf{T}\mathbf{h})$ denotes the temperature factor and for isotropic thermal vibrations is equal to (Peng 2005):

$$T(\mathbf{R}_k^\mathsf{T}\mathbf{h}) = exp(-B\mathbf{h}^2), \qquad (7)$$

where B is the Debye–Waller B-factor, equivalent to the isotropic atomic displacement parameters in the small molecule crystallography. The X-ray diffraction structure factors $F^x(\mathbf{h})$ can be calculated in the same way using the $f^x(\mathbf{R}_k^\mathsf{T}\mathbf{h})$ form factors.

The refinements of small organic molecules with TAAM result in the B-factors that are closer to the reference ones than in the refinements with IAM (Gruza *et al.*, 2020). In the case of electron diffraction data, the B-factors generated with IAM were overall too small comparing to the reference data, and for the X-ray diffraction data, they were too large. Using the calculated structure factors for different scattering models, the apparent difference between the overall B-factors can be estimated from the equation:

$$ln\frac{|F_{TAAM}(s)|^2}{|F_{IAM}(s)|^2} = ln(k^2) - 2\Delta B s^2, \qquad (8)$$

where k is a scale factor between $|F_{TAAM}(s)|$ and $|F_{IAM}(s)|$, whereas $\Delta B = B_{IAM} - B_{TAAM}$.

2.2. Calculations

The experimental 3D ED data sets containing electrostatic potential maps with fitted atomic coordinates for lysozyme (*Gallus gallus*) structure at 1.8 Å resolution PDB ID 5k7o, EMD-8217 (de la Cruz *et al.*, 2017), and for proteinase K (*Parengyodontium album*) at 1.75 Å resolution PDB ID 5i9s, EMD-8077 (Hattne *et al.*, 2016), were downloaded from RCSB PDB (Berman *et al.*, 2000) and Unified Data Resource for 3DEM (Lawson *et al.*, 2015) databases. Hydrogen atoms were added to the protein structures based on geometry with Molprobity (Williams *et al.*, 2018) and adjusted to ensure the catalytically competent protonation state at pH 4.7 for lysozyme and pH 8 for proteinase K. Hydrogen atoms in water molecules were added with Chimera (Pettersen *et al.*, 2004), considering the clashes and hydrogen bond formation. The lengths of all bonds between the hydrogen and non-hydrogen atoms were extended to match the typical distances observed in neutron diffraction data. In proteinase K, the side chain of the partially missing residue ARG 64 was rebuilt in Maestro 11.9 (Schrödinger Release 2019-1: Maestro, Schrödinger, LLC, New York, NY, 2019). The atomic B-factors, equal to 120% of the B-factors of the closest non-hydrogen atoms, were assigned to all hydrogens in proteins except for the methyl group hydrogens. In the case of the hydrogen atoms in methyl groups and water molecules, 150% of the B-factors of the closest non-hydrogen atoms were used. Next, the file format was changed to SHELX style (Sheldrick, 2015) with Mercury (Macrae *et al.*, 2020) and used with the LSDB program (Volkov *et al.*, 2004) to transfer the UBDB2018 atom type parameters (Kumar *et al.*, 2019). In the proteinase K structure, the UBDB parameters were manually adjusted for the S atoms from SO_4^{2-} molecules and for water molecules

401, 408 and 480. The latter water molecule was located at the symmetry element, so the multipole parameters were multiplied by $\frac{1}{2}$. The structure factors in xd.fou files of XD format (Volkov *et al.*, 2016) for X-ray diffraction for the TAAM model were calculated using UBDB2018 parameters with the software being an extension of the DiSCaMB library (Chodkiewicz *et al.*, 2018). Then, they were converted using the Mott-Bethe formula (Equation 2) to arrive at the TAAM model for electron diffraction. The coefficients for analytical Gaussian approximation to scattering factors for the IAM model for both X-ray (Doyle & Turner, 1968; Fox *et al.*, 1989) and electron scattering (Peng *et al.*, 1996) were directly taken from the International Tables for Crystallography (2006), Vol. C, Tables 6.1.1.4 and 4.3.2.3, respectively. The reflection indices necessary for the structure factors calculations were generated with Python 3.7 with 100% completeness up to given resolution, taking into account the experimental unit cell dimensions. An exemplary Python script to generate the reflection indices for lysozyme is provided in the Supplementary Information, Section S1. The experimental reflection indices and structure factors were used only to generate the Wilson plots. The experimental density maps deposited in the Unified Data Resource for 3DEM were used for the comparison of the electrostatic potential map features with the calculated maps. The experimental structure factors were not used for the experimental map recalculation as the purpose of this project was to compare the theoretical maps with the deposited ones and in the future a similar study will be performed for the Cryo-EM-derived density maps, where recalculating the experimental density map from the experimental structure factors is not feasible. Fourier maps were calculated in the WinXD2016 package (Volkov *et al.*, 2016) and their format was changed to situs format using in-house scripts for visualization in Chimera (Pettersen *et al.*, 2004). To arrive at the e/Å units, the calculated potential density maps values were recalculated (for more details, see Section S2 in the Supporting Information).

2.3. Analysis

Two different approaches to the data analysis were applied. The first approach was based on the data in the real space, taking into account the experimental deposited 3D ED density maps EMD-8217 (de la Cruz *et al.*, 2017), EMD-8077 (Hattne *et al.*, 2016) and the voxel values of the calculated electrostatic potential and electron density maps. For the visualization, all the density maps were scaled to match the standard deviation of the voxel values of the experimental density maps and the mean voxel value was shifted to zero. All the density maps were cut around the protein with a 3 Å margin before scaling. Cutting, scaling, the visualization of the cross-sections and sigma contours of the maps were done in Chimera (Pettersen *et al.*, 2004). The calculation of the map correlation coefficients around mean (CC) and rank CC for the quantile rank-scaled maps (CCr) between two grid functions were based on the equations 4 and 17 from (Urzhumtsev *et al.*, 2014). The calculations were done in PHENIX 1.14 (Liebschner *et al.*, 2019) with the Map sigma level comparison tool. To compare the experimental and calculated density maps in a quantitative manner close to atom positions, the covalent radius averaging method was used. In this method, the averaging over the grids sampled within the covalent radius distance from atom positions is performed. Sampling was done in Chimera every 0.1 Å using the original deposited experimental maps with ca. 0.6 Å voxels and the calculated and scaled IAM and TAAM maps with 0.3 Å voxels. More details about the sampling, together with the covalent atom radii for different elements and the resulting number of grid points are available in the Supporting Information, Section S3. The rebuilt residue ARG 64 in proteinase K was not taken into account in this analysis. All the atom names follow the standard nomenclature present in the PDB structures of proteins, except for the

oxygen atoms in the water molecules, named Owat.

The second approach focused on the reciprocal space information, with detailed analysis of the relations between calculated structure factors. Wilson plots were plotted with reciprocal squared resolution ($1/d^2$) shells averaged for each 0.01 Å^{-2} bin. In crystallography, the reliability factor (R-factor) usually measures the agreement between the amplitudes of the structure factors from a model and from the X-ray diffraction data. Here, it was used to compare two models (TAAM and IAM). The Fourier shell correlation (FSC) was calculated over all structure factors in 0.1 Å^{-1} reciprocal resolution (1/d) bins according to the formula shown in equation 1 in (Nicholls *et al.*, 2018).

3. Results and Discussion

The electron scattering factors were used to calculate the 3D electrostatic potential density maps eTAAM and eIAM. The cross-section through the maps is shown in Figure 1a. The two calculated maps both correspond well with the experimental map from 3D ED, for example in the region of the disulfide bonds, visible in dark red. The same visualizations done for the electron density maps generated with the X-ray scattering factors xTAAM and xIAM do not reveal significant differences when compared with the eTAAM and eIAM maps. In all the calculated maps after rescaling the background is dominated by the negative values and noise from the Fourier truncation errors is seen. At the first glance, we would not be able to tell the difference between the maps of the electrostatic potential and electron density. However, the differences are well visible after taking a closer look at single amino acid residues, in particular the charged ones in Figure 1b. The maps were calculated in two ways: with thermal smearing effects expressed by experimental B-factor values (with B) and without those

effects (w/o B). Regardless of taking into account thermal smearing, the electrostatic potential map contour encompasses less volume around oxygen atoms in the negatively charged acidic side chains when the eTAAM is used in comparison with the eIAM. For the positively charged arginine residue, the eTAAM map contour is larger around the nitrogen atoms than the eIAM contour. This dependence is consistent for all other negatively and positively charged amino acids in lysozyme and proteinase K. It is with agreement with our expectation as the electron scattering factors of the negatively charged oxygens become negative at low resolution range, which decreases the positive contributions to the Coulomb potential coming from the atomic cores. In contrast, the neutral phenylalanine contours do not reveal visible differences between the eTAAM and eIAM models. In the electron density contour maps, the dependence on the side chain charges is negligible and both xTAAM and xIAM maps are strikingly similar. It is also in line with the theoretical expectations as the X-ray scattering factors are always positive and are not influenced so much by the charge differences.

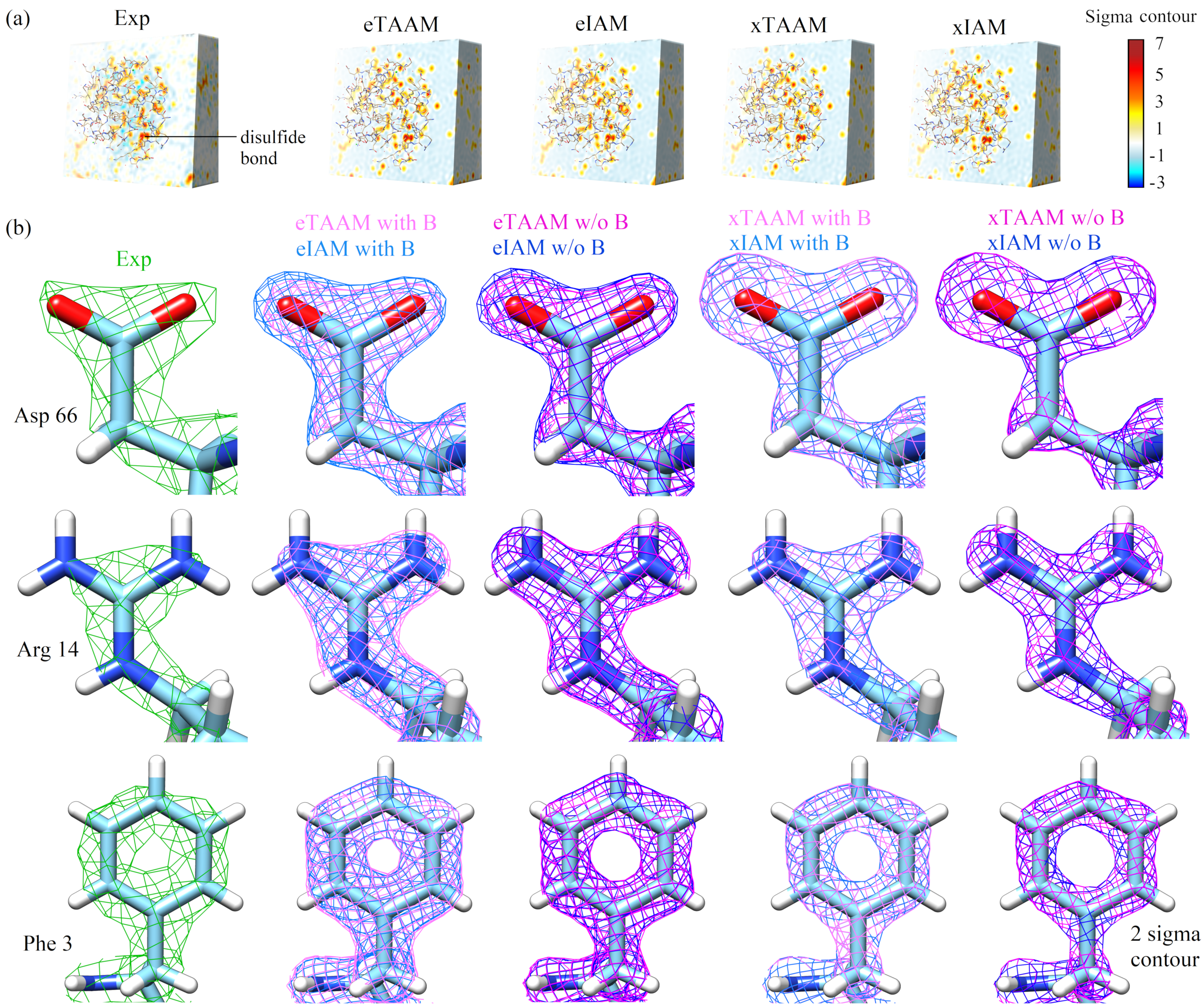

Fig. 1. Experimental (Exp) and calculated electrostatic potential density maps (TAAM - based on the Transferable Aspherical Atom Model, IAM - based on the Independent Atom Model) of the lysozyme structure at 1.8 Å. eTAAM and eIAM maps were calculated using the electron diffraction scattering factors, whereas xTAAM and xIAM maps were calculated using the X-ray diffraction scattering factors. The voxel values of all calculated maps are scaled to the standard deviation of the experimental density map and the average value of zero, then their sigma contours are shown. (a) Cross-sections through the electrostatic potential density maps colored by sigma contour values. The protein structure heavy atoms and water molecules oxygen atoms are shown as licorice and small spheres, respectively. (b) The contour electrostatic potential density map for chosen amino acid side chains from the lysozyme structure. The maps in light colors include the experimental B-factor values and take into account the thermal smearing effects (with B), while the maps in dark colors neglect the thermal smearing effects (w/o B).

All the maps shown in Figure 1 are scaled to match the distribution of the values of the experimental 3D ED map and the average value of the voxel is equal to zero. Note that the experimental voxel size is close to 0.6 Å, whereas the voxel size of the calculated maps is 0.3 Å. If we resample the calculated grids on the 0.6 Å grids, it is possible to investigate the CC values around mean between each two maps. The results for the lysozyme and proteinase K experimental and calculated maps are shown in the Supporting Information, Section S4, in Tables S2 and S3. CC values for both proteins between the experimental and calculated density maps range from 0.75 to 0.79. The CC values measured between the experimental and calculated density maps extracted within the protein region with minimum solvent content are all higher than 0.92. Nevertheless, by looking at the CC, it is not possible to differentiate the maps calculated with the electron diffraction factors from the maps calculated with the X-ray diffraction factors. Also, there is no significant difference in CC between the IAM and TAAM resampled maps or the maps with or without the B-factors. To check if the histogram equalization of the maps would bring new insights to the analysis, we have also added CCr to the Tables S2 and S3. The CCr values comparing the experimental and TAAM/IAM maps appeared to be systematically higher for the density maps with the protein-only regions than for the maps with high solvent contribution. Moreover, it shows slightly higher values for the TAAM/IAM full maps calculated without taking into account the B-factor values than for the maps calculated with B-factors. Both the CC and CCr correlation coefficients computed for the comparison between different e/xTAAM and e/xIAM models with/without B-factors indicate that the xTAAM and xIAM maps are practically indistinguishable with the coefficients close to 1.00, whereas the introduction of different treatment of the B-factors may lower the CC and CCr values to 0.95 and 0.80, respectively.

The correlation coefficient analysis does not provide satisfactory explanation of the changes in the density maps generated with TAAM/IAM models, different B-factor treatment and electron/X-ray diffraction. To avoid the visual inspection of hundreds of amino acids in the various experimental and calculated density maps, we have performed a high-throughput radial analysis of the map values close to atom positions, namely sampled within the covalent radius from the atoms, for example within 0.8 Å from the positions of CA atoms. The details about the sampling method and the full list of the covalent radii with the number of atoms in each structure are available in the Supporting Information, Section S4. Figure 2a gathers the boxplots for the average density values in lysozyme and proteinase K for chosen atoms with the standard atom nomenclature as in PDB data files. For compatibility with the experimental density maps, the calculated eTAAM and eIAM maps were scaled. It is visible that the eTAAM density maps tend to have closer values to the experimental maps than eIAM, except for the hydrogen atoms, for which none of the models corresponded well with the experimental values. It may stem from the fact, that the hydrogen atoms in the proteins were not present in the deposited structures. Thus, they may have been absent during the refinement of the experimental data, resulting in deformed experimental maps. Alternatively, the fluctuations of the hydrogen atoms and the corresponding B-factors assigned in the model may have been underestimated.

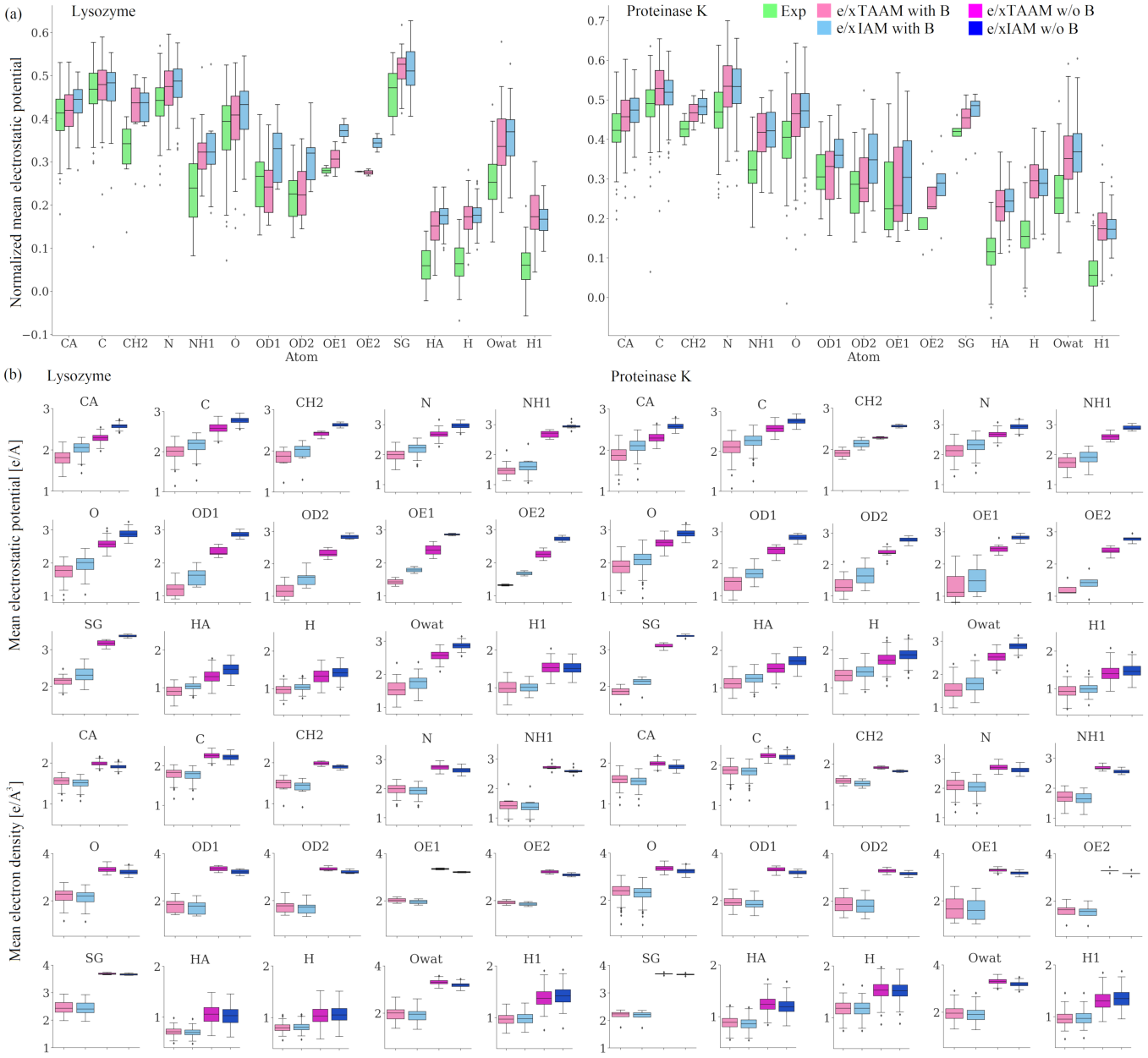

Fig. 2. (a) Boxplots for the normalized average values of the electrostatic potential around chosen atom positions, calculated for experimental (Exp) and scaled electrostatic density maps (eTAAM and eIAM) for lysozyme and proteinase K. For details about the sampling and choice of the atoms, see Section S3 in Supporting Information. (b) Boxplots for the unscaled average values of the electrostatic potential (eTAAM and eIAM) and electron density (xTAAM and xIAM) around chosen atom positions. The light and dark colors indicate taking into account and neglecting the thermal smearing effects, respectively.

Such a high-throughput analysis of many amino acids allows us to follow certain trends in the unscaled density maps. Figure 2b presents the boxplots for the electrostatic potential (eTAAM and eIAM) and electron density (xTAAM and xIAM) maps, calculated for the same atoms in two systems, both with and without the thermal smearing effects. Not surprisingly, the value ranges of the maps taking into account the B-factor values in the calculations are lower than for the static models. This is visible well in the maps from electron and X-ray diffraction. Additionally, the mean eTAAM values are higher than the mean eIAM values but for the maps derived with the X-ray scattering factors, this is not the case. The overall differences between the two models are small but consistent throughout the full dataset. The IAM model does not take into account the deformation of the density coming from the influence of the local chemical environment. Note that averaging around the atom positions pictures a radial overview of the map features, whereas the TAAM model is aspherical. The analysis of the density along the bonds would be more appropriate to get insight into the aspherical character of the density but due to large voxel size, the sampling along the bonds would contain very few data points.

Careful observation of the graphs in Figure 2b allows to see an irregular behavior of the proteinase K density maps around OE1 atoms. The static maps are characterized with very small diversification of the eTAAM and eIAM map values, whereas the maps with thermal smearing show large range of acquired values. Similar discrepancy, but to a lower extent, is seen in the xTAAM and xIAM electron density maps. Inspecting the B-factor values of the OE1 atoms in proteinase K indicates that there is one atom in GLU 48 with strikingly low B-factor (Figure 3). Visualizing the structural vicinity of that atom helps in understanding the differences visible in the previous graphs. This atom creates a stronger hydrogen bond with the surrounding protein

residues, is better stabilized and its movements are restricted. This influences the shape of the experimental and calculated scaled density maps, both in electron and X-ray diffraction. The contours of the static density maps are not affected by the presence of this hydrogen bond as they do not take into account the B-factor values in the calculations. That observation underlines the importance of having the correctly determined B-factors in the structures deposited to the PDB Data bank. It is worth to mention that currently there are no tools for the B-factor validation in the structures determined with Cryo-EM and those tools are urgently needed.

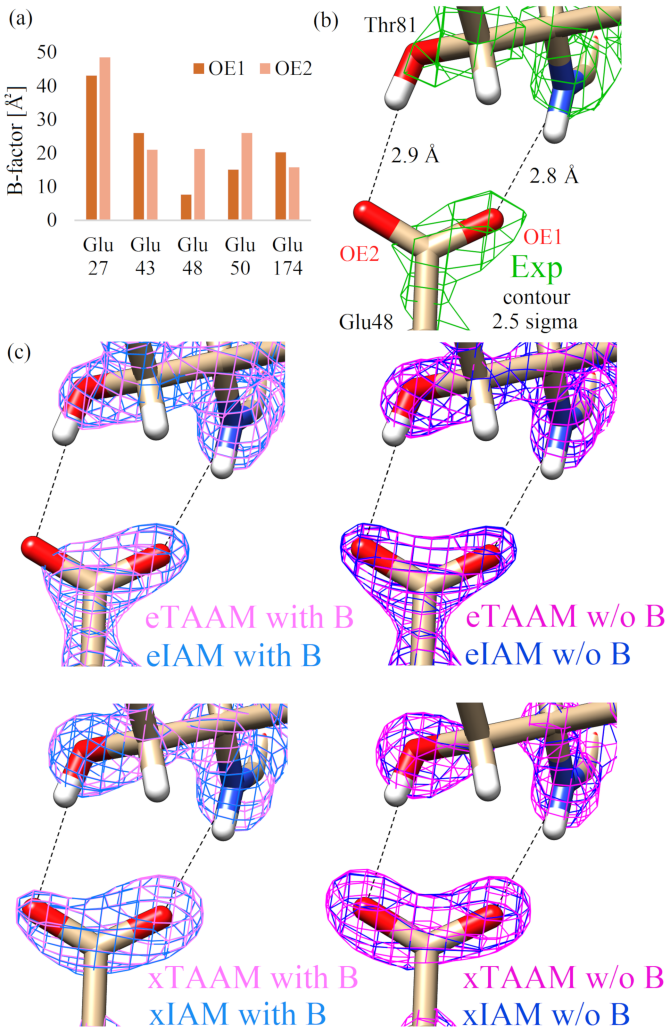

Fig. 3. (a) B-factor values for the OE1 and OE2 atoms in Glu residues in Proteinase K. (b) Experimental electrostatic potential density map, (c) calculated electrostatic potential density maps and (d) electron density maps for the Glu 48 side chain with two hydrogen bonds marked with dashed lines. All density maps shown at 2.5 sigma contour.

Table 1. *R-factors (R_1 and R_2) calculated between different variants of the lysozyme structure factors.*

	F_1	F_2	$R_1 = \frac{\sum \lvert\lvert F_1\rvert - \lvert F_2\rvert\rvert}{\sum \lvert F_1\rvert}$	$R_2 = \frac{\sum \lvert\lvert F_2\rvert - \lvert F_1\rvert\rvert}{\sum \lvert F_2\rvert}$
Impact of the scattering model	eTAAM with B	eIAM with B	0.13	0.11
	eTAAM w/o B	eIAM w/o B	0.12	0.11
	xTAAM with B	xIAM with B	0.04	0.04
	xTAAM w/o B	xIAM w/o B	0.04	0.05
Impact of thermal smearing	eTAAM with B	eTAAM w/o B	0.61	0.38
	xTAAM with B	xTAAM w/o B	0.65	0.39
	eIAM with B	eIAM w/o B	0.61	0.38
	xIAM with B	xIAM w/o B	0.64	0.39
Impact of electron/ X-ray diffraction	eTAAM with B	xTAAM with B	2.39	0.71
	eTAAM w/o B	xTAAM w/o B	2.46	0.71
	eIAM with B	xIAM with B	1.95	0.66
	eIAM w/o B	xIAM w/o B	2.00	0.67

In order to quantify the impact of the TAAM/IAM model, thermal smearing and the electron/X-ray scattering factors on the structure factors, we have calculated the R-factors, presented in Table 1. The R-factors calculated between the eTAAM and eIAM structure factors are higher than those for xTAAM and xIAM, which underlines the fact that for electron diffraction the choice of the model plays a more significant role than for X-ray diffraction. Nevertheless, all the values are lower than 0.13 so the choice of the model does not apply large changes to the structure factors. Upon applying thermal smearing, the structure factors show larger deviation, while the largest impact on the structure factors comes from switching between electron and X-ray scattering factors.

Analysis of FSC shown in Figure 4 between different models allows us to follow the trends in the structure factor values indicated by the R-factor in separate resolution shells. Thus, by looking at the top panel in Figure 4, it is straightforward that the highest deviations in the structure factors between eTAAM and eIAM with B-factors are present in the low resolution reflections. On the other hand, the introduction of the thermal smearing affects the high resolution structure factor values. A very interesting trend in the structure factors is observed when electron or X-ray scattering factors are used. TAAM is more sensitive to the change from electron to X-ray scattering factors

in the low resolution region, while IAM in the medium and high resolution region.

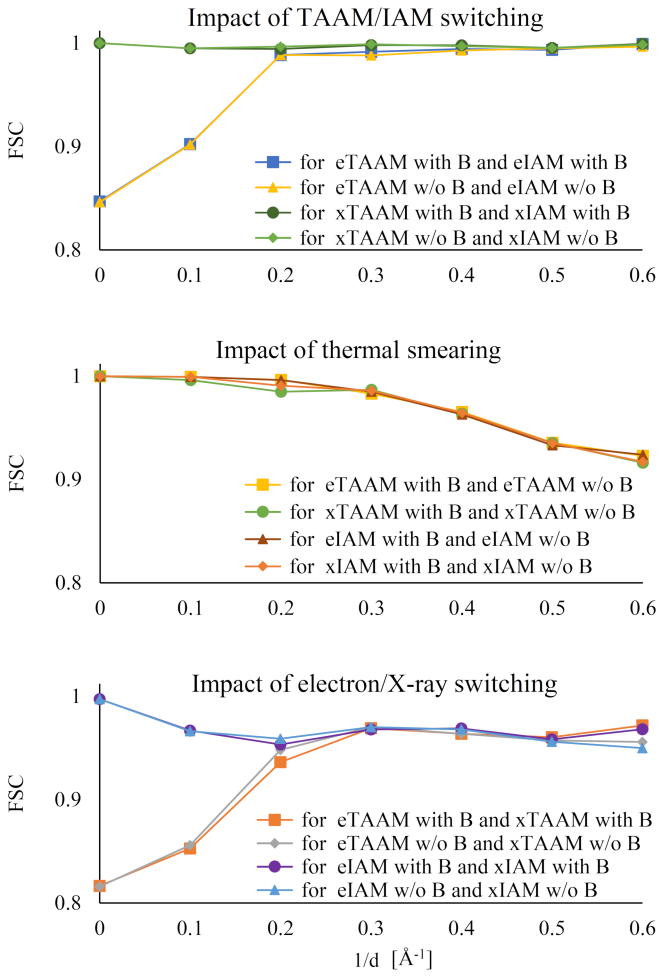

Fig. 4. Fourier shell correlation calculated for structure factors in 0.1 $Å^{-1}$ reciprocal resolution bins for various TAAM and IAM models.

The Wilson plots show the squares of the structure factors generated for each model change with respect to the inverse square of the resolution (Figure 5a,b). As expected, the higher is the resolution, the lower are the structure factor amplitudes. A local minimum is observed in the region around 7 Å resolution, followed by a local maximum at around 4 Å. The differences between the TAAM and IAM models are visible best in the low resolution region. The eTAAM squared structure factors are lower than the corresponding values for eIAM. However, the xTAAM and xIAM squared structure

factors show an opposite trend but the differences are small. When including the effect of the B-factors on the structure factors in the model calculation with thermal smearing, the high resolution structure factors diminish, which is visible in both calculated and experimental structure factors.

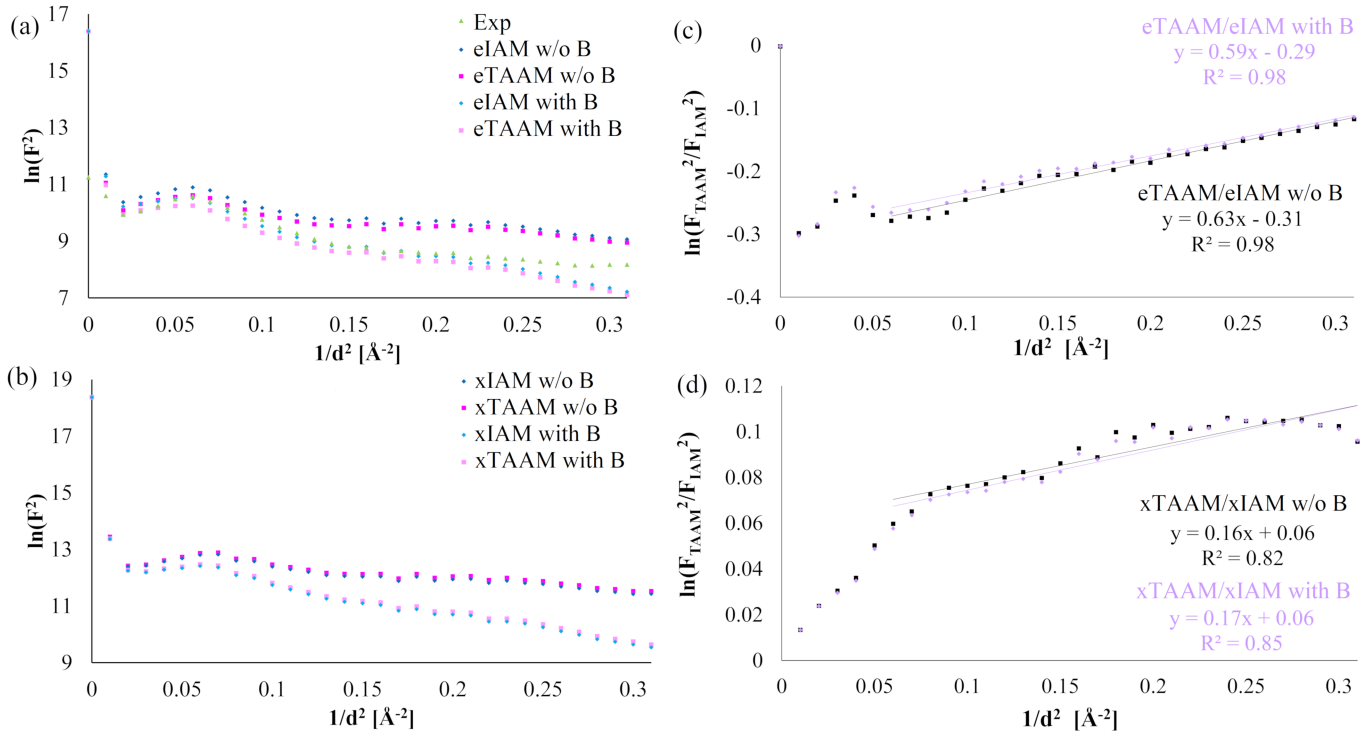

Fig. 5. Wilson plots for (a) electron and (b) X-ray diffraction structure factors for the lysozyme. The experimental squared structure factors were scaled to match the local minimum value of the eTAAM with B data points and were truncated according to the resolution range shown in the plot. (c) Relation between squared eTAAM and eIAM structure factors with and without thermal smearing. (d) Relation between squared xTAAM and xIAM structure factors with and without thermal smearing. In all plots, the resolution shells were averaged for each 0.01 $Å^{-2}$ bin.

It is possible to picture the relation between the squared eTAAM and eIAM structure factors (Figure 5c). The largest deviation appears in the region of the resolution around 3.5 - 4.5 Å (ca. 0.08 - 0.05 $Å^{-2}$) and is visible again in the low resolution. Including the thermal smearing effects in the calculations impacts slightly the slope of the fitted line. Then, by analysis of this slope and using Equation 8, we can calculate

the apparent change in B-factors. For the electron diffraction structure factors with thermal smearing ΔB is equal to -1.18 Å^2, while without thermal smearing to -1.24 Å^2. The corresponding values for the X-ray diffraction are equal to -0.41 Å^2 and -0.38 Å^2, respectively (Figure 5d). These results show that the experimentally obtained B-factors are underestimated.

The calculations of theoretical electrostatic density maps for macromolecules may potentially help in understanding the structural features of the solved macromolecular complexes, such as the presence of charged ions and water molecules. Future work on this project includes the calculations of the electrostatic potential density maps at different resolutions with the comparison between chosen experimental datasets from 3D ED and Cryo-EM.

4. Conclusions

We have developed a method to calculate theoretical electrostatic potential density maps with high accuracy via structure factor calculation. The method is based on a Transferable Aspherical Atom Model (TAAM) and is derived from the Hansen-Coppens multipole model with the atom type parameters transferred from the University at Buffalo Data Bank. The calculated TAAM maps for electron diffraction (eTAAM) at 1.8 Å correspond well with the experimental density maps of lysozyme and proteinase K. The density maps based on the Independent Atom Model (IAM), using the approximated electron scattering factors (eIAM), are not as sensitive to the charged amino acids as the eTAAM maps. For comparison, we have also calculated the corresponding maps calculated using the X-ray scattering factors (xTAAM and xIAM, respectively). The density measured around atoms positions reveal that in general the eTAAM maps show lower values than the eIAM, whereas for the xTAAM and xIAM maps the trend is opposite. Moreover, the differences between eTAAM and eIAM maps

are larger than between the xTAAM and xIAM maps. The high-throughput analysis of densities measured around atoms in amino acids can reveal interesting structural features, for example hydrogen bonds that stabilize the structure and visible differences in the shape of the density maps, which underline the importance of B-factors, especially for the electrostatic potential density maps.

Acknowledgements

The authors thank Małgorzata Cabaj for the calculation of the UBDB parameters for the S atoms in SO_4^{2-} molecules.

Funding information

The following funding is acknowledged: the National Centre of Science (Poland) (grant OPUS No. UMO-2017/27/B/ST4/02721 to Paulina Maria Dominiak) and PL-Grid Infrastructure through grant UBDB2019.

References

Bąk, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B. & Dominiak, P. M. (2011). *Acta Cryst.* A**67**, 141–153.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Research*, **28**(1), 235–242. **URL:** *https://doi.org/10.1093/nar/28.1.235*

Chodkiewicz, M. L., Migacz, S., Rudnicki, W., Makal, A., Kalinowski, J. A., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Adams, P. D. & Dominiak, P. M. (2018). *J Appl Crystallogr*, **51**(Pt 1), 193–199.

Cowley, J. M., Goodman, P., Vainshtein, B. K., Zvyagin, B. B. & Dorset, D. L. (2006). In *International Tables for Crystallography*, pp. 276–345. International Union of Crystallography. **URL:** *https://doi.org/10.1107/97809553602060000558*

de la Cruz, M. J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., Sawaya, M. R., Cascio, D., Weiss, S. C., Kim, S. K., Hinck, C. S., Hinck, A. P., Calero, G., Eisenberg, D. & Gonen, T. (2017). *Nat Methods*, **14**(4), 399–402.

Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* D**62**, 1325–1335.

Dittrich, B., Hübschle, C. B., Pröpper, K., Dietrich, F., Stolper, T. & Holstein, J. J. (2013). *Acta Cryst.* B**69**, 91–104.

Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch., C. (2012). *Acta Cryst.* A**68**, 337–351.

Dominiak, P. M., Li, X., Coppens, P., Messerschmidt, M. & Volkov, A. (2007). *J. Chem. Theory Comput.* **3**, 232–247.

Dorset, D. L. (1991). *Electron Diffraction Structure Analysis of Organic Crystals*, pp. 1–10. Dordrecht: Springer Netherlands.

Doyle, P. A. & Turner, P. S. (1968). *Acta Crystallographica Section A*, **24**(3), 390–397.

Fox, A. G., O'Keefe, M. A. & Tabbernor, M. A. (1989). *Acta Crystallographica Section A*, **45**(11), 786–793.

Ghermani, N.-E., Bouhmaida, N. & Lecomte, C. (1993). *Acta Crystallographica Section A*, **49**(5), 781–789.
URL: *https://doi.org/10.1107/S0108767393003538*

Gruza, B., Chodkiewicz, M. L., Krzeszczakowska, J. & Dominiak, P. M. (2020). *Acta Crystallogr A Found Adv*, **76**(Pt 1), 92–109.

Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.

Hattne, J., Shi, D., de la Cruz, M. J., Reyes, F. E. & Gonen, T. (2016). *J Appl Crystallogr*, **49**(Pt 3), 1029–1034.

Jarzembska, K. N. & Dominiak, P. M. (2012). *Acta Cryst.* A**68**, 139–147.

Jha, K. K., Gruza, B., Kumar, P., Chodkiewicz, M. L. & Dominiak, P. M. (2020). *Acta Crystallographica Section B*, **76**(3), 296–306.
URL: *https://doi.org/10.1107/S2052520620002917*

Kühlbrandt, W. (2014). *Science*, **343**(6178), 1443–1444.

Kulik, M., Goral, A. M., Jasiński, M., Dominiak, P. M. & Trylska, J. (2015). *Biophys. J.* **108**, 655–665.

Kumar, P. & Dominiak, P. M. (2021). *Molecules*, **26**(13).
URL: *https://www.mdpi.com/1420-3049/26/13/3872*

Kumar, P., Gruza, B., Bojarowski, S. A. & Dominiak, P. M. (2019). *Acta Crystallogr A Found Adv*, **75**(Pt 2), 398–408.

Lawson, C. L., Patwardhan, A., Baker, M. L., Hryc, C., Garcia, E. S., Hudson, B. P., Lagerstedt, I., Ludtke, S. J., Pintilie, G., Sala, R., Westbrook, J. D., Berman, H. M., Kleywegt, G. J. & Chiu, W. (2015). *Nucleic Acids Research*, **44**(D1), D396–D403.
URL: *https://doi.org/10.1093/nar/gkv1126*

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 Crystallographica Section D*, **75**(10), 861–877.
URL: *https://doi.org/10.1107/S2059798319011471*

Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M. & Wood, P. A. (2020). *Journal of Applied Crystallography*, **53**(1), 226–235.
URL: *https://doi.org/10.1107/S1600576719014092*

Malińska, M., Jarzembska, K. N., Goral, A. M., Kutner, A., Woźniak, K. & Dominiak, P. M. (2014). *Acta Cryst.* D**70**, 1257–1270.

Marques, M. A., Purdy, M. D. & Yeager, M. (2019). *Curr Opin Struct Biol*, **58**, 214–223.

Mott, N. F. & Massey, B. S. (1965). *The theory of atomic collisions*. Oxford University Press.

Nannenga, B. L. & Gonen, T. (2018). *Emerg Top Life Sci*, **2**(1), 1–8.

Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. (2014). *Nat Methods*, **11**(9), 927–930.

Nicholls, R. A., Tykac, M., Kovalevskiy, O. & Murshudov, G. N. (2018). *Acta Crystallographica Section D*, **74**(6), 492–505.
URL: *https://doi.org/10.1107/S2059798318007313*

Peng, L.-M. (1999). *Micron*, **30**(6), 625–648.

Peng, L.-M., Ren, G., Dudarev, S. L. & Whelan, M. J. (1996). *Acta Crystallographica Section A*, **52**(2), 257–276.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J Comput Chem*, **25**(13), 1605–1612.

Pichon-Pesme, V., Jelsch, C., Benoît, G. & Lecomte, C. (2004). *Acta Cryst.* A**60**, 204–208.

Rez, D., Rez, P. & Grant, I. (1994). *Acta Crystallographica Section A*, **50**(4), 481–497. **URL:** *https://doi.org/10.1107/S0108767393013200*

Sheldrick, G. M. (2015). *Acta Crystallographica Section A*, **71**(1), 3–8. **URL:** *https://doi.org/10.1107/S2053273314026370*

Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *Elife*, **2**, e01345.

Urzhumtsev, A., Afonine, P. V., Lunin, V. Y., Terwilliger, T. C. & Adams, P. D. (2014). *Acta Crystallographica Section D*, **70**(10), 2593–2606. **URL:** *https://doi.org/10.1107/S1399004714016289*

Volkov, A., Li, X., Koritsánszky, T. & Coppens, P. (2004). *J. Phys. Chem. A*, **108**, 4283–4300.

Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. & Koritsánszky, T., (2016). `XD2016` - a computer program package for multipole refinement, topological analysis of charge densities and evaluation of intermolecular energies from experimental and theoretical structure factors.

Wang, J. & Moore, P. B. (2017). *Protein Sci*, **26**(1), 122–129.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018). *Protein Sci*, **27**(1), 293–315.

Yonekura, K. & Maki-Yonekura, S. (2016). *Journal of Applied Crystallography*, **49**(5), 1517–1523.

Synopsis

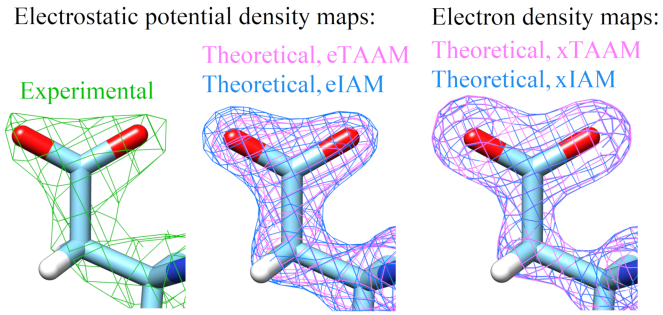

Fig. 6. Accurate electrostatic potential and electron density maps of proteins are calculated based on the transferable aspherical atom model with the pseudoatom databank and compared with the experimental data.