Structure

X-rays, electrons, and neutrons as probes of atomic matter
--Manuscript Draft--

Manuscript Number:	STRUCTURE-D-22-00198R
Full Title:	X-rays, electrons, and neutrons as probes of atomic matter
Article Type:	Theory
Keywords:	X-ray crystallography; cryo-EM; micro-ED; neutron diffraction; electron density; ED; electrostatic potential; ESP; nuclear coherent scattering length; NCSL; Thomas-Fermi; Bethe-Mott
Corresponding Author:	Matthias Bochtler, PhD IIMCB Warszawa, POLAND
First Author:	Matthias Bochtler, PhD
Order of Authors:	Matthias Bochtler, PhD
Abstract:	X-rays, electrons and neutrons interact differently with matter and probe different properties. X-rays detect electron density (ED). Electrons measure the electrostatic potential (ESP) of electrons and nuclei. Neutrons measure the nuclear coherent scattering length (NCSL). The differences between NCSL maps and the other maps are well known. In contrast, ED and ESP maps are tacitly expected to be similar or even identical, as evidenced by the description of micro-ED and cryo-EM maps as ``densities''. Here, I demonstrate that the implicit assumption of ED and ESP equivalence is wrong, but contains a grain of truth. Based on Density Functional Theory (DFT), the Bethe-Mott (BM) relation and the Thomas-Fermi (TF) and Cromer-Mann (CM) atomic models, I show that ED and ESP maps are indeed more similar to each other than to NCSL maps. Nonetheless, peak and integrated map values depend differently on atom order number and on contributions from electrons in inner and outer CM shells. ED and ESP maps also differ in the sign and relative magnitude of excess charge effects.
Additional Information:	
Question	Response
Standardized datasets<p>A list of datatypes considered standardized under Cell Press policy is available here. Does this manuscript report new standardized datasets?	No
Original Code<p>Does this manuscript report original code?	No

Prof. Matthias Bochtler
International Institute of Molecular and Cell Biology
Trojdena 4, 02-109 Warsaw, Poland
Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
Tel: 0048 22 5970732
Fax: 0048 22 5970715
e-mail: mbochtler@iimcb.gov.pl

Dr Karin Kühnel
Editor-in-Chief, Structure

2022-11-22

Dear Dr. Kühnel,

I apologize that it has taken me so long to get back to you with what seemed to be a relatively simple revision of my manuscript on the maps obtained using X-rays, electrons and neutrons as the probe. The delay was caused by issues unrelated to the manuscript: I had to undergo several surgeries in the summer/autumn, and I am only slowly catching up now.

The major issue in this revision was the referee's request to better justify the qualitative statements on the effects of excess charge on ED and ESP maps at low and high resolution. I believe that I have been able to do this in a very satisfactory way, using both RHF based model calculations, and a comparison of experimental ED and ESP maps. The results of these calculations, which fully support the analytical results, are presented in the new Fig. 9. I was also able to support my conclusions by more literature search. All charge effects that I describe have previously been noticed by other authors. The new version of the manuscript does a better job referencing the prior work on this specific aspect.

With one exception, the other referee comments were largely editorial, or based on misunderstandings that hopefully could be clarified by rewriting some text passages. The referee has a point that my generalization from CM shells to quantum chemical shells was too general. I still believe that the essence of the argument is correct (when orbitals with different

INTERNATIONAL INSTITUTE OF MOLECULAR AND CELL BIOLOGY IN WARSAW
4 Ks. Trojdena Street, 02-109 Warsaw, Poland; phone: +48 22 597 07 00, fax: +48 22 597 07 15
e-mail: secretariat@iimcb.gov.pl, **www.iimcb.gov.pl**

angular momentum are averaged over). However, I was not able to provide a rigorous mathematical demonstration. As the generalization is peripheral to the main argument of the manuscript, I have simply removed it from the manuscript.

On an editorial level, I have tried to comply with Structure format requirements. In particular, I have separated text, figures and Tables, and I have removed the numbering of paragraphs. I hope that you will find the new version of the manuscript suitable for publication.

With best regards

Prof. Matthias Bochtler
International Institute of Molecular and Cell Biology
Trojdena 4, 02-109 Warsaw, Poland
Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
Tel: 0048 22 5970732
Fax: 0048 22 5970715
e-mail: mbochtler@iimcb.gov.pl

Reviewer #1: Matthias Bochtler provided the third version of his manuscript, titled "X-rays, electrons, and neutrons as probes of atomic matter". In his response, he argues that the manuscript was difficult to read because it requires advanced mathematics. However, it is not the equations that are difficult to read, it is the text between the equations. The manuscript is lengthy to read and seems to jump between topics - it is still hard to discover the line of thought, with some sections seemingly repeating what was stated earlier.

Both reviewers did not understand the term EO in version 2 of the manuscript. I still do not. A mathematical definition might help, where words fail to explain the concept.

There is agreement with all referees that the ESP is the electrostatic potential due to nuclei and electrons. From the EO point of view, the potential is attributed to the electrons, the contribution of the nucleus is treated implicitly. To make the point clearer, the EO picture is now introduced with an additional panel in Fig. 2A. In the context of the Cromer-Mann model, the ESP is expressed in terms of a_i and b_i, which can be seen as an example of the EO point of view.

p. 20: "eq. 5 predicts..." where does eq. 5 come from? p. 21 line 16: typo with 'y' rather than 'x' in integral?

There was a glitch in equation referencing. The text should be: "Eq. 4 for predicts…" The argument substitution is now written out explicitly so that the reader can verify mathematical correctness. For the second comment, I assume that page 12 (not 21), line 16 is meant. There is no typo. The 'y' in the integral is correct.

In Eq. 6, tilde(rho(q)) = f(s), while in Eq. 16: tilde(rho(q)) = f_X(q), while q != s

The difference is between the notations of physicists and mathematicians. Throughout, I use f for the form factor, which can be understood as a function of q or s, which is common practice in the physics community. Of course, in the mathematical sense, this makes f a different function when the argument is q and when the argument is s. To avoid the issue, I would have to introduce different symbols for form factors dependent on whether the function argument is q or s. I doubt that this would be helpful.

The text leading from Eq. 18 to 19 is not very clear. Maybe it could be replaced with a logically consistent sequence of equations for better readability (same applies in many sections, e.g. 6.1 could this way be shortened to a few lines).

I struggle to reword this to make it clearer. The problem is analogous to a famous problem in astronomy, that is described by the exact same mathematics, because the Coulomb and gravitational potentials have the same mathematical form. For a long time, it was unclear whether the gravitational potential from a spherical earth was identical to the gravitational potential that would be caused by an equal point mass within the center. Differential geometry can then be used to show that:
(a) outside the spherical earth, the statement is true.

(b) inside the spherical earth, the statement is only true with modification. The gravitational potential is the potential from a point mass, but of the mass equal to the mass inside the sphere that contains the observation point.

Section 4.4.5 seems a lengthy discussion about Eq. (1). It states Eq. (20), points out that this Eq. (20) cannot be evaluated, while being more accurate than (1), and returns to Eq. (1) for a qualitative description of the 2/3 power relation between ESP and ED. It seems this is just a repetition of what has been said.

The derivation of the power law relating ESP and ED from the CM model was necessary because of referees' doubts about the derivation from DFT and TF theory. Note that only "routine numerical integration of equation (20) is impossible", but the integral can actually be evaluated by a combination of analytical and numerical methods. As the

Eq. 22 assumes a spherical symmetric potential. This is appropriate for a single atom (or ion), but for molecules, this approximation should not hold. Maybe it refers to the integrated intensity around one atom. In this case, the upper integration limit would not be infinity. In case only a single atom is considered, Coulomb's law is valid, and the intermediate step (the double integral) seems an unnecessary complication.

The referee's argument looks convincing, but it is nonetheless incorrect. In fact, I made the same mistake originally. Following the referee's line of reasoning, one obtains a result that disagrees with the reciprocal space result by exactly a factor of 2. This discrepancy has puzzled me for a long time. The referee's argument and my original own reasoning overlook that only the potential gradient, but not the potential itself is related to the charge within. Hence, one integration is necessary to go from the potential gradient to potential itself. The other integration is needed to quantify the charge that goes into the calculation of the potential gradient, i.e. the charge "within".

p. 28 argues "In case of the ESP, electrons contribute proportionally to the bj values of their CM shells. Therefore, all space ESP integrals depend strongly on the quantum chemical shell structure of the atoms". In response to the review, this line of thought is reasoned with the statement "The chemical orbitals can be expanded as a sum of Gaussians". This statement and the above conclusion are incorrect. The concept of 'Gaussian orbitals' include Gauss functions, but in combination with a polynomial term r^l. In contrast to the CM expansion, Gaussian orbitals therefore do not peak at the origin, but at a certain point away from the origin (nucleus), as one would expect for chemical shell structure, but not like CM shells. Whether or not ESP integrals depend on the quantum chemical shell (which seems plausible), this connection to CM expansion does not justify this claim.

I am grateful to the referee for spotting this problem. When I made the argument, I had the radial charge distribution in mind, which contains an additional $4\pi r^2$ factor and is 0 at the nucleus for the Cromer-Mann shells and the quantum chemical shells, irrespective of n and l. As the expansion is indeed problematic (also because of intrinsic widths of CM and quantum chemical shells), I have removed the claim that the inner/outer shell statements generalize from CM shells to quantum chemical shells, as well as any explanations of other effects that use this argument.

The discussion on integrated potential and shell structure would benefit from previous work, notably, Acta Cryst.(1994). A50,33-45 (O'Keefe, Spence, "On the Average Coulomb Potential (Phi_o) and Constraints on the Electron Density Crystals"

Indeed, the reference contains not only the Bethe-Mott formula, but also an equation for the integrated ESP (with infinite integration radius), albeit given in Fourier space, that I was unaware of (their equation (8)). The work of O'Keefe and Spence is now appropriately referenced.

p.28, "This increase can be understood as a consequence of the addition of electrons, ". Electrons are added also within the same row of the periodic table. When moving to the next period, the first electron is added at a much larger distance (higher energy level) compared with the addition within a period, resulting in a weaker shielding.

With the removal of the generalization of the influence of inner and outer electrons from CM shells to quantum chemical shells, the statement has been removed altogether. Instead, I simply argue with the much increased mean square radius, which was earlier given as an alternative explanation.

p. 30: 'equation X': is this a typo with missing reference?

Indeed, this is an error that has been introduced with the rewrite of the paragraph on the power law. The proper equation is now referenced.

Section 4.5.2.1 defines rho_B and rho_R (without further comment), but they are not used anywhere else in the manuscript (other than the appendix). What is their purpose within the manuscript?

In this section, I discuss peak map values for finite B-factor (rho_B) or finite resolution (rho_R), or finite integration radius (rho_Rhat). The subscripts were introduced at the request of the other referee. Elsewhere, they are not needed, because integration is to infinity, and a resolution or B-factor limit is not used.

p. 28: "The data show that integrated and peak map values depend in opposite ways on inner and outer shell electrons of neutral atoms. Inner electrons are more important for the ED, outer electrons for the ESP" How do the data show these opposite dependency? The graphs are different, but how are they related to inner and outer shell electrons (c.f. comment above on Gaussian orbitals).

The abscissa is in units of (scaled) sqrt(b), a measure for the size of the CM shell. For the validity of the extrapolation from CM shells to quantum chemical shells, see the response to the comment on Gaussian orbitals.

p. 36: "Therefore, negative excess charge increases the ED and positive excess charge decreases it" is this a sloppy sentence? positive charge (i.e. a proton) does not alter the electron density.

Note that I am referring to "excess" charge, defined as charge not present in the neutral atom, so nuclear charge is not an option.

p. 38: "(Fig. 5, column ED)" typo, Fig 6 referred to instead?

Indeed, this is a typo. Fig. 6 was meant.

p.39: "Of course, the greater relative contribution of excess charge to map values at low than high resolution does not mean that partial charge is more accurately measured at low resolution". This may or may not be right, but to test this hypothesis, maps calculated from shells (!, high and low resolution) would need to be juxtaposed. The manuscript argues by maps from the entire resolution range by providing two references, one with high resolution data, one with low resolution data. (Yonekura et al. 2017, Yip et al. 2020).

In my eyes, the best way to demonstrate the validity of the statement would be a juxtaposition of

(a) a high-resolution cryo-EM map

(b) an originally high-resolution cryo-EM map, with resolution degraded by a resolution cutoff

(c) an originally high-resolution cryo-EM map, with resolution degraded by addition of a B-factor

(d) an originally low resolution cryo-EM map

The manuscript makes three claims about excess charge:

(a) it has opposite effect on the ED and ESP

(b) it affects the ESP more than the ED

(c) it is more noticeable in low resolution than in high resolution maps.

In the meantime, I have found reports in the literature on all three effects. The earlier observations are now appropriately referenced in the text. To further support the point, I have carried out RHF calculations in GAMESS and then calculated ED and ESP maps for high and low B-factors (Fig. 9A). Additionally, I have compared charged aspartate and glutamate residues with uncharged asparagine and glutamine residues in low and high resolution experimental ED and ESP maps (Fig. 9B). The results of this computational analysis are fully consistent with the theoretical expectations.

Overall this text my contain some useful, or new information, but this is not obvious to find, and much of the reasoning does not seem to hold. As mentioned above, replacing paragraphs with logically consistent equations but be of benefit. In order to test this hypothesis, one would have to compare data in a low resolution shell (!) with data at high resolution shell (!). While the claim may or may not be valid, the reasoning of the author compares two entire data sets. However, both reviewer seem to judge the quality of this manuscript quite differently. I suggest it is now time for the editor to make a decision, without further rounds of revision.

I see a dilemma here. Lecturing on the content of this paper, I have noticed that I have a much better audience response if I focus on the results and emphasize conclusions. On the other hand, the reaction of the referees shows that many statements in the manuscript are definitely novel, and are even met with considerable skepticism, showing that omission of many derivations (as in the original version of the manuscript) is not possible at this stage. Perhaps it would be useful to write up a more streamlined review of the conclusions after publication of the original paper. Alternatively, I would welcome it very much if Structure accompanied the paper by a Commentary that could help guide the reader.

X-rays, electrons, and neutrons as probes of atomic matter

Matthias Bochtler

IIMCB, Trojdena 4, 02-109 Warsaw, Poland and IBB, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland
mbochtler@iimcb.gov.pl

Abstract

X-rays, electrons and neutrons interact differently with matter and probe different properties. X-rays detect electron density (ED). Electrons measure the electrostatic potential (ESP) of electrons and nuclei. Neutrons measure the nuclear coherent scattering length (NCSL). The differences between NCSL maps and the other maps are well known. In contrast, ED and ESP maps are tacitly expected to be similar or even identical, as evidenced by the description of micro-ED and cryo-EM maps as "densities". Here, I demonstrate that the implicit assumption of ED and ESP equivalence is wrong, but contains a grain of truth. Based on Density Functional Theory (DFT), the Bethe-Mott (BM) relation and the Thomas-Fermi (TF) and Cromer-Mann (CM) atomic models, I show that ED and ESP maps are indeed more similar to each other than to NCSL maps. Nonetheless, peak and integrated map values depend differently on atom order number and on contributions from electrons in inner and outer CM shells. ED and ESP maps also differ in the sign and relative magnitude of excess charge effects.

- X-rays, electrons and neutrons probe different aspects of matter.

- The resulting maps are ED, ESP and NCSL maps.

- ESP maps are intrinsically broadest, NCSL maps are intrinsically sharpest.

- Excess charge has opposite effects on ED and ESP maps, and no effect on NCSL maps.

Keywords: X-rays, electrons, neutrons, electron density (ED), electrostatic potential (ESP), nuclear coherent scattering length (NCSL), X-ray crystallography, electron diffraction, cryo-EM, neutron crystallography.

Introduction

X-rays, electrons and neutrons as the probe

Macromolecular structures can be studied using X-rays, electrons or neutrons (Shull and Wollan 1948). Diffraction can be used for all three types of probes. Reciprocal space "imaging" is practiced in the fields of X-ray crystallography (Drenth 1999), of 3D-ED techniques (Gemmi et al. 2019) including micro-ED (Clabbers, Shiriaeva, and Gonen 2022), and of neutron crystallography (Blakeley, Langan, and Niimura 2008). "Conventional" imaging is only practical using electrons as probes, because good lenses are only available for electrons, but not X-ray photons or neutrons. Real space imaging is practiced in the fields of cryo-electron microscopy (cryo-EM) (Dubochet 2012; Bai, McMullan, and Scheres 2015) and of cryo-electron tomography (cryo-ET) (Baumeister 2020). For diffraction techniques with incident wavelength λ, the physical resolution limit is $\lambda/2$ ($\lambda/\sqrt{2}$ for a flat detector perpendicular to the beam). For cryo-EM imaging and cryo-ET, the Abbe limit applies (Abbe 1873), given by Scherzer as 0.6λ or 0.8λ depending on the setup (Scherzer 1949). For diffraction and imaging, the best possible resolution is therefore of the order of the wavelength. To achieve atomic or near-atomic resolution, the wavelength must therefore be less than a few Å.

X-rays as the probe

X-rays as probes of matter are typically used in the $1.6 - 0.5$ Å wavelength range, corresponding to photon energies in the $8 - 25$ keV range (Fig. 1A, left). The long wavelength limit is close to the limit set by the resolution criterion and is additionally limited by the photoelectric effect (Hall 1936). This effect becomes much more pronounced at long wavelengths (Saloman, Hubbell, and Scofield 1988) and causes radiation-induced damage without contributing to the useful signal (except in anomalous scattering experiments). The short wavelength limit is set by the ratio of desirable coherent to undesirable incoherent scattering, which becomes increasingly unfavourable at shorter wavelengths (Wentzel 1927; Cromer 1969). Since the incident wavelengths are of the order of atomic distances, the scattering is typically wide-angle (Fig. 1B, left). In the long wavelength limit, the cross section for coherent scattering is of the order of $(Zr_e)^2$, where $r_e \approx 2.8$ fm is the classical electron radius (Fig. 1C, left). As this value is very small, thick samples such as macromolecular crystals can be probed. Despite the small cross section, very thin samples can also be studied thanks to the high photon flux and beam brilliance of modern synchrotrons and X-FEL sources (Shin 2021). In the classical picture, coherent X-ray scattering can be understood as the radiation from induced dipoles (Jackson 1998). Compared to single electrons, the nucleus feels only a slightly stronger force (Z times, where Z is the atomic order number). However, it reacts with much greater inertia, due to the larger mass (Nm_n/m_e times, where $m_n/me \approx 2000$ is the nucleon/electron mass ratio, and $N \approx 2Z$ the number of nucleons). Therefore,

2

an X-ray beam sees only the electron density (ED) but is blind to the nuclear density to a very good approximation (Fig. 1D, left).

Electrons as the probe

Electrons as a probe of matter are typically used in the $4.0 - 2.0$ pm wavelength range, corresponding to kinetic energies in the $100 - 300$ keV range (Fig. 1A, middle). Short wavelengths are used, because the larger cross sections for longer wavelengths would lead to undesirable multi-bounce scattering in all but the thinnest samples (Lenz 1954). On the short wavelength side, the cryo-EM contrast at 2.0 pm is already low and would decrease further at even shorter wavelengths. As the probe wavelength is much shorter than atomic distances, scattering is focused into a narrow forward cone (Fig.1B, middle). The cross section for scattering is orders of magnitude larger than for X-rays as the probe (Henderson 1995). At 300 keV, the electron scattering cross section is of the order of $(0.1a_0)^2$ for the lighter elements, where $a_0 \approx 0.52$ Å is the Bohr radius (Fig. 1C, middle). Due to the large cross section, electrons as a probe are only suitable for very thin specimens, such as micro-crystals in micro-ED and other 3D-ED techniques, or molecules embedded in thin layers of amorphous ice (cryo-EM) or very thin specimens (cryo-ET). Electrons as charged probes respond to the nuclei and the electrons in a sample. Several authors have emphasized that electrons measure the electostatic potential (ESP), as one would expect, justifying the conclusion either by a physical picture of the probe-sample interaction and empirical observations (Marques, Purdy, and Yeager 2019), or by detailed calculations of image formation and reconstruction (Rullgård et al. 2011) (Fig 1D, middle).

Neutrons as the probe

Neutrons as probes of matter are typically used in the $6.0 - 0.7$ Å pm wavelength range, corresponding to thermal kinetic energies in the $2 - 150$ meV range. The long wavelength limit is set by the maximum resolution achievable, and to some extent by radiation damage due to neutron capture, primarily by the ^{1}H and ^{14}N nuclei in the sample (Henderson 1995). The short wavelength limit appears to be primarily related to the contraction of the diffraction pattern with decreasing wavelength and the resulting limits on beam divergence. Tighter limits reduce the usable neutron flux, which is typically time limiting in neutron diffraction experiments. As the probe wavelengths and atomic distances are similar, the scattering is wide angle (Fig. 1B, right). In the useful wavelength range, the cross section for coherent neutron scattering is independent of wavelength and happens to be of similar magnitude as the cross section for coherent X-ray scattering, despite the very different physical origin (Sears 1992) (Fig. 1C, right). As the cross section is small and the neutron flux is typically limiting, the technique is best suited to very thick samples (large crystals). Neutrons interact with nuclei by short-range nuclear forces (Squires 1996). The cross section for scattering of neutrons by electrons is orders of magnitude smaller

3

(Condon 1936; Foldy 1958) unless the electrons are unpaired and spin ordered (Lynn 1936). For non-magnetic biological samples it is therefore reasonable to conclude that neutrons see only the nuclei but are blind to the electrons (Fig. 1D, right). In the following I will use the term "nuclear coherent scattering length" (NCSL) maps for maps of non-magnetic samples obtained using thermal neutrons as a probe (Fig. 1D, right).

X-rays, electrons and neutrons as probes of ED, ESP, and NCSL

The concept that X-rays, electrons and neutrons measure ED, ESP and NCSL, respectively, is useful and will form the basis of this work. It is an idealisation applicable when confounding factors such as incoherent scattering, absorption in the sample and multiple scattering (dynamical diffraction) can be neglected or corrected for (Wentzel 1927; Cromer 1969; Saloman, Hubbell, and Scofield 1988; Clabbers and Abrahams 2018; Schoenborn and Nunes 1972). Multiple scattering is a particularly severe limitation for electrons as the probe, because of the large scattering cross section. In diffraction experiments, it causes a paucity of weak reflections (Clabbers et al. 2019) and degrades hydrogen visibility in ESP maps (Palatinus et al. 2017; McCusker 2017; Clabbers et al. 2019). In the following I will discuss expectations for "ideal" ED and ESP maps, with occasional reference to NCSL maps for comparison. The focus on "ideal" maps is useful, because it drastically simplifies the analysis, and still explains actual features of experimental maps measured using X-rays, electrons and neutrons as probes.

Methods

Conventions

Key abbreviations are listed in Table 1. Throughout this work, Hartree atomic units are used. The unsigned electron charge e, the electron mass m_e, and the reduced Planck constant $\hbar = h/(2\pi)$ are set to 1, and the vacuum permittivity ϵ_0 is set to $1/(4\pi)$. Note that in these units, the speed of light is $1/\alpha \approx 137$, where α is Sommerfeld's fine structure constant, and not 1 as in many other systems of "natural" units. The unit of length in Hartree atomic units is the Bohr, $a_0 \approx 0.52$ Å, used for example for resolution R and wavelength λ. The unit of energy is the Hartree, $E_0 = \hbar^2/(m_e a_0^2) \approx 27.2$eV. EDs and ESPs are measured in units of $e/a_0^3 \approx 1.08 \cdot 10^{12}$C/m^3 and $E_0/e \approx 27.2$V, respectively. The ED is denoted as a positive number density. Fourier transform pairs are defined asymmetrically with the $1/(2\pi)^3$ factor assigned to the transformation from reciprocal to real space. A tilde denotes a reciprocal space quantity. Integrals without explicit bounds are taken over all real or reciprocal space. For scattering events, λ is the wavelength of the incident beam, θ is the Bragg angle, and 2θ is the angle of deflection of the probe beam. The momentum change in a scattering event in atomic units is $q = 4\pi\sin\theta/\lambda$. In connection with the Cromer-Mann model, I also use $s = \sin\theta/\lambda = q/(4\pi)$. Both q and s are given in reciprocal

4

Bohrs. The resolution R is defined as the smallest distance between crystal lattice planes that scatter discrete spots, i.e. $R = \lambda/(2\sin\theta) = 1/(2s)$. The gauge (arbitrary offset) for all potentials is chosen such that the potentials are 0 at a large distance from the sample. $\Phi(r)$ is used interchangeably with the ESP for the electrostatic potential from the nucleus and all electrons. Φ_{el} is the electrostatic potential from the electrons only. Φ_{TF} is the electrostatic potential seen by one electron due to the nucleus and all *other* electrons, in the TF approximation. In the large Z limit, for which the TF model is intended, Φ_{TF} and $\Phi = ESP$ need not be distinguished.

Data sources

Revised Cromer-Mann coefficients were taken from the Int. Tables for X-ray crystallography, vol 4, pages 99-101, as digitalized by the CCP4 project (Agirre et al. 2023).

Calculations

Numerical work for Figures 3, 5 and 7 was done in python, using the numpy (Harris et al. 2020) and scipy packages (Virtanen et al. 2020). The relative contributions of electrons to ED and ESP maps (from the EO point of view) are expressed in terms of weights w_{ED} and w_{ESP}. These weights were calculated using symbolic integration as implemented in the maxima package (Maxima 2022). For numerical tests of the power law relating ESP and ED, equation (20) had to be evaluated. As an analytical expression for the integral could not be found, integration was done semi-numerically. For this purpose, the integration range was partitioned into a small qr region, where numeric integration was unproblematic, and a large qr region, where numeric integration was hindered by poor convergence. In this region, the terms in the sum of equation (20) can be neglected, and the integral can be evaluated analytically noting that $\int_0^\infty dx \cdot \sin(x)/x = \pi/2$. The latter formula can be derived in the $y \to 0$ limit from the Laplace transform identity $e^{-xy}\sin(x)/x = \arctan(1/y)$ (Bronstein and Semendjajew 1991). For the calculation of ESP maps, short peptides were first geometry-optimized using Avogadro (Hanwell et al. 2012). HF calculations were done for a 6-31G(d) basis set using GAMESS (Barca et al. 2020). The ED and ESP were calculated based on the GAMESS output using multiwfn (Lu and Chen 2012; Zhang and Liu 2021). Multiwfn generated maps were read into numpy arrays using cubetools (P. R. Vaidyanathan, unpublished). Convolutions to apply B-factors to maps were calculated in reciprocal space for computational efficiency. Maps were output in .mrc format using EMDA (Warshamanage, Yamashita, and Murshudov 2022) and displayed using PYMOL (Schroedinger, unpublished). Map values at atom positions in experimental maps were obtained by interpolation using MAPMAN (Kleywegt and Jones 1996).

Results

Manuscript organization

This work summarizes theoretical expectations for the shared features and differences of ED, ESP and NCSL maps. The differences between NCSL maps on the one hand and ED and ESP maps on the other hand are widely appreciated. In contrast, the differences between ED and ESP maps do not appear to have been systematically explored. Attention has been paid primarily to the accuracy of map derived models (Wlodawer, Li, and Dauter 2017), rather than to differences in the maps themselves. With this work, I aim to fill this gap. After a brief statement of conventions, I start out with the Thomas-Fermi (TF) and Cromer-Mann (CM) atom models. While the material on the TF model is of course not novel, the interpretation of the CM coefficients in real space appears new (albeit it is certainly implicit in the work of Cromer and Mann). I then proceed to clarify the relationship between the ED and the ESP. The equations relating the potentials due to electrons Φ_{el} with the electron density ρ are of course well known. It is less well appreciated, however, that Φ_{el} cannot be equated with the ESP, because the nuclear contribution is missing. In fact, the nuclear contribution dominates and the ESP is best regarded as a nuclear potential that is imperfectly screened by the electrons. I also test the applicability of the TF theory power law relating ESP and ED for neutral atoms using the CM coefficients derived from Hartree-Fock theory. The results confirm that ESP and ED are approximately related by a power law, with an exponent not very different from 1, explaining why ED and ESP maps are more similar to each other than to NCSL maps. I quantify map values for spherical neutral atoms, either at the nuclear positions ("peak map values"), or in a spherical region around the nuclei ("integrated map values"). I deduce the dependence on the atom order number and disentangle the contributions of electrons in inner and outer CM shells. The dependence of the integrated and peak map values on the atom order number is certainly implicit in prior work, but I am unaware of any juxtaposition of results for ED and ESP maps. To my knowledge, the discussion of the relative contributions of electrons in inner and outer CM shells to the ED and ESP is novel. In the end, I present qualitative corrections for the neutral atom theory when charge neutrality does not hold.

Atom models

Qualitative properties of ED, ESP and NCSL maps are best understood in the context of simple atom models. In this work, two types of models for spherical symmetric atoms are considered, the Thomas-Fermi (TF) model and the Cromer-Mann (CM) model.

The Thomas-Fermi (TF) model

The TF model smooths over the quantum chemical shell structure of electrons and treats them as a a continuous gas or quantum liquid (Hohenberg and Kohn

6

1964) and can therefore be seen as a precursor of the density functional theory (Becke 2014). The central ingredient of the TF model is a power law (derived in the first paragraph of the Suppl.) that relates the potential Φ_{TF} due to the nuclear charge and the charge of all other electrons on the one hand and the electronic charge density on the other hand.

$$\Phi_{TF}(\vec{r}) = \frac{1}{2}(3\pi^2)^{2/3}\rho^{2/3}(\vec{r}) \tag{1}$$

For spherical, neutral atoms, this formula can be combined with the Poisson equation (Jackson 1998) to derive the TF equation (Thomas 1927):

$$\Delta\Phi_{TF}(r) = \frac{8\sqrt{2}}{3\pi}\Phi_{TF}^{3/2}(r) \tag{2}$$

This equation can be solved for all atoms (all Z) in terms of a universal function χ that fulfills $\sqrt{x}\chi'' = \chi^{3/2}$ and the conditions $\chi(0) = 1$ and $\chi(x) \to \infty$ as $x \to \infty$ (Landau and Lifshitz 1977). With the abbreviation $C = (3\pi/4)^{2/3}/2$, the expressions for the potential and the electron density are (Landau and Lifshitz 1977):

$$\Phi_{TF}(r) = \frac{Z}{r}\chi\left(\frac{rZ^{1/3}}{C}\right) \tag{3}$$

$$\rho(r) = \left(\frac{Z}{r}\right)^{3/2}\frac{\sqrt{8}}{3\pi^2}\chi^{3/2}\left(\frac{rZ^{1/3}}{C}\right) \tag{4}$$

In these equations, $\rho(r)$ can be directly identified with the ED. $\Phi_{TF}(r)$ is the potential felt by an electron of the atom, due to the interactions with the nucleus and all *other* electrons. By contrast, the ESP or $\Phi(r)$ is the potential felt by an electron of the probe beam, which is due to the interactions the nucleus and *all* electrons of the atom. For atoms with many electrons, for which the TF model is intended, the difference is unimportant and $\Phi_{FT}(r)$ can be identified with the ESP. For atoms with just a few electrons, the TF approximation becomes poor and the identification of $\Phi_{TF}(r)$ with the ESP becomes questionable. Numerical calculations presented later in this work show that even in this low Z regime, the relationship between ESP and ED is still quite well described by a power law, albeit with an exponent closer to 1 than TF theory would predict. Qualitatively, equations (3) and (4) express that all atoms "look alike" from the perspective of an X-ray photon or electron beam, except for charge and size scale factors. This is a consequence of the smoothing over the quantum chemical shell structure by the inhomogeneous electron gas approximation of the TF-model. Within the limits of this approximation, equation (4) predicts a power-law decrease of the mean squared radius $\langle r^2 \rangle$ with Z.

$$\langle r^2 \rangle \propto Z^{-2/3} \tag{5}$$

The Cromer-Mann (CM) model

The CM model is based Hartree Fock (HF) calculations that respect the chemical shell structure of atoms. Cromer and Mann calculated form factors, which

7

they then approximated as sums of Gaussian functions, according to the formula:

$$\tilde{\rho}(q) = f(s) = \sum_{i=j}^{4} a_i e^{-b_i s^2} + c = \sum_{j=1}^{5} a_j e^{-b_i s^2}$$

$$a_5 = c, \quad b_5 = 0, \quad \sum_{j=1}^{5} a_j = Z \tag{6}$$

where θ is the Bragg angle, i.e. half the angle of beam deflection, λ the wavelength, and $s = \sin(\theta)/\lambda$, which relates to the exponent q in the Fourier transform according to $q = 4\pi s$. This reciprocal space expression can be converted into an expression for the real space density $\rho(q)$ using the spherical Fourier transform for Gaussians.

$$\rho(r) = \sum_j \rho_j(r) \text{ where } \rho_j(r) = a_j \left(\frac{4\pi}{b_j}\right)^{3/2} e^{-4\pi^2 r^2/b_j} \tag{7}$$

For $b_j = 0$, the expression for $\rho_j(r)$ is technically undefined. However, it can easily be shown that for $b_j \to 0$, $\rho_j(r) \to a_j \delta^3(r)$, where $\delta^3(r)$ is the 3-dimensional delta function. The result shows that the CM model approximates both the reciprocal space form factors and the real space densities as sums of Gaussians. This justifies an interpretation of the $\rho_j(r)$ as CM electron shells, which together account for the ED. The CM shells are of course different from the usual quantum-chemically defined electron shells. The CM shells are not associated with quantum numbers and can be filled with a non-integral number of electrons. They are spherically symmetric and have their maximum electron density at the nuclear position, which is not true in general for quantum chemical shells (Pauling and Wilson Jr 1985). The a_j values play the role of weights in real and reciprocal space. As the a_j values sum to Z according to equation (6), the a_j values can be understood as the number of electrons in the CM shells in real space. To get a feel for the interpretation of the b_j values, note that the total (non-integer) number of electrons $4\pi r^2 dr a_j e^{-4\pi^2 r^2/b_j}$ in an interval dr is maximal when $r = \sqrt{b_j}/(2\pi)$. This observation suggests a possible relationship between the mean squared radius $\langle r^2 \rangle$ and the b_j value of a CM shell. Such a formula can indeed be derived using the formula $\int_0^\infty x^n e^{-x^2} dx = \frac{1}{2} \int_0^\infty y^{\frac{n+1}{2}-1} e^{-y} dy = \frac{1}{2} \Gamma(\frac{n+1}{2})$, where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ is the well-known Γ function. Note also that $\Gamma(5/2) = 3/2 \cdot \Gamma(3/2)$, because of the recursion $\Gamma(z) = (z-1)\Gamma(z-1)$. Hence:

$$\langle r^2 \rangle_j = \frac{\int_0^\infty 4\pi r^4 \rho_j(r) dr}{\int_0^\infty 4\pi r^2 \rho_j(r) dr} = \frac{a_j \frac{4\pi}{b_j^{3/2}} (\frac{b_j}{4\pi^2})^{5/2} \Gamma(5/2)}{a_j \frac{4\pi}{b_j^{3/2}} (\frac{b_j}{4\pi^2})^{3/2} \Gamma(3/2)} = \frac{3}{8\pi^2} b_j \tag{8}$$

The equation shows that the mean squared radius grows with the b_j. Mathematically, this result is identical to the expression that relates the isotropic

mean square displacement (in all three spatial directions, hence the factor 3) to the temperature factor (Bahar, Atilgan, and Erman 1997). However, the b_j values do not describe thermal motion. They rather reflect the smearing of the electrons in the CM shells. The mean squared radius of the whole atom is a weighted average of the mean squared radii of the CM shells, where the shell weights reflect the number of electrons in the shells.

$$\langle r^2 \rangle = \frac{\int_0^\infty 4\pi r^4 \rho(r) dr}{\int_0^\infty 4\pi r^2 \rho(r) dr} = \frac{3}{8\pi^2} \frac{\sum_{j=0}^{5} a_j b_j}{\sum_{j=0}^{5} a_j} \tag{9}$$

ED, ESP, and NCSL maps

At atomic resolution, NCSL maps should have point-like atomic signals in the absence of thermal motion or blurring due to limited resolution. In contrast, ED and ESP atomic signals are intrinsically broadened, even in the absence of blurring from thermal motion or limited resolution. The ED and ESP maps are related in that they both depend entirely (ED maps) or partially (ESP maps) on the electron cloud. In the following, these relationships will be explored in more detail. The key conclusion is a confirmation of the power law of equation (1) relating ESP and ED, with exponent 2/3 for large Z and between 2/3 and 1 for smaller Z. ED and ESP maps are similar because they are related by a power law with an exponent close to 1. As the exponent in the power law is smaller than 1, atomic signals are intrinsically broader for ESP than for ED maps, which can be understood in terms of the long-range nature of the Coulomb interactions. Of course, atom intrinsic signals are further broadened by thermal motion (B factor effects) and limited resolution for all three map types.

Electron electrostatic potential and ED

The electron density and the electrostatic potential due to the electrons, a quantity different from ESP (see below), are interdependent. In real space, the relations are (Jackson 1998):

$$\Phi_{el}(\vec{r}) = - \int \frac{\rho(\vec{r'})}{|\vec{r} - \vec{r'}|} d^3 r' = -\rho(\vec{r}) \star \frac{1}{|\vec{r}|} \tag{10}$$

$$\Delta \Phi_{el}(\vec{r}) = \vec{\nabla}\vec{\nabla} \Phi_{el}(\vec{r}) = 4\pi\rho(\vec{r}) \tag{11}$$

In the equation (10) the $*$ denotes a convolution. The equation follows directly from the Coulomb law and the superposition principle. In the equation (11) the operators Δ and $\vec{\nabla}$ denote the Laplace and Nabla (gradient) operators respectively. Geometrically, this equation states that the electron density is related to the three-dimensional curvature of the electron electrostatic potential. The Fourier transform gives the corresponding expressions in reciprocal space:

$$\tilde{\rho}(\vec{q}) = \int \rho(\vec{r}) e^{i\vec{q}\vec{r}} d^3 r = -\frac{1}{4\pi} q^2 \tilde{\Phi}_{el}(\vec{q}) \tag{12}$$

9

$$\tilde{\Phi}_{el}(\vec{q}) = \int \Phi_{el}(\vec{r})e^{i\vec{q}\vec{r}}d^3r = -4\pi\frac{\tilde{\rho}(\vec{q})}{q^2} \tag{13}$$

These reciprocal space equations can be derived from equation (10), remembering that the Fourier transform of a convolution of two functions is the product of their Fourier transforms, and that the Fourier transform of $1/|\vec{r}|$ is $(4\pi)/q^2$. Alternatively, they can be derived from equation (11), since each application of the $\vec{\nabla}$ operator in real space corresponds to a multiplication by $i\vec{q}$ in reciprocal space, and since $i^2 = -1$.

ESP and ED, the electrons only (EO) point of view

The equations in the previous paragraph do *not* describe the relationship between ED and ESP, despite a widespread perception to the contrary. While ED is indeed ρ, the ESP is the sum of the potentials of the electrons *and* nuclei.

$$ESP = \Phi = \Phi_{el} + \Phi_{nuc} \tag{14}$$

Despite this physical reality, it is sometimes useful to assign the potential to the electrons alone. The electrons must then be divided into a "balanced charge" group of electrons with a total charge equal to the combined charge of the nuclei, and an "excess charge" group of electrons (or holes for positive partial charges) that can contain a non-integer number of electrons and has no counterbalancing nuclear charge. For the electrons in the "balanced charge" group, the potential is the potential normally associated with them, plus the potential of a unit positive charge at the associated nucleus. Ambiguities in the partitioning of electrons between balanced and excess charge groups, and to nuclei within the balanced charge group, are inconsequential. They only affect the partitioning of the ESP into contributing terms, which are non-observable anyway. Since the potentials are additive, the total ESP is not affected by any of the ambiguities. In the following, I will refer to the description of the ESP in terms of the electrons alone, i.e. with implicit nuclear charge, as the "electrons only" (EO) point of view (Fig. 2A).

Exact reciprocal space relation between ESP and ED

Exact relations between ESP and ED are simpler in reciprocal than in real space. For simplicity, assume that a spherical neutral atom is placed at the origin. The ESP is then the sum of potential from the nucleus, Z/r, and from the electrons Φ_{el}. The Fourier transform of $1/|\vec{r}|$ is $(4\pi)/q^2$. The Fourier transform of Φ_{el} in the spherical case is $-4\pi\tilde{\rho}/q^2$ according to equation (13). As the Fourier transform is linear, it follows that:

$$\tilde{\Phi}(q) = 4\pi\frac{Z - \tilde{\rho}(q)}{q^2} \tag{15}$$

The first and second terms in the numerator of the equation represent the nuclear and electronic contributions to the ESP, respectively. For spherically symmetric atoms, $\tilde{\rho}(q)$ and $\tilde{\Phi}(q)$ are linearly related to the form factor for X-ray

10

scattering f_X, and the form factor for electron scattering f_e that is termed the Born electron scattering amplitude in the volume C of the International Tables of Crystallography.

$$f_X(q) = \tilde{\rho}(q) \tag{16}$$

$$f_e(q) = \tilde{\Phi}(q)/(2\pi) \tag{17}$$

Therefore, equation (15) is equivalent to the celebrated Bethe-Mott equation (Bethe 1930; Mott 1965) relating X-ray and electron scattering form factors.

Exact real space relation between ESP and ED

The exact relation between ED and ESP is more complicated in real space than in reciprocal space. First consider the contribution of the electrons to the potential gradient, in the spherically symmetric case. The Gauss flux theorem (Gauss 1877) relates the surface integral of the potential gradient to the charge enclosed by this surface:

$$\int_S \vec{\nabla}\Phi_{el}d\vec{S} = 4\pi \int_V \rho(\vec{r})d^3r \tag{18}$$

The equation shows that $\vec{\nabla}\Phi_{el}(\vec{r})$ remains unchanged when the electronic charge inside the sphere of radius r is collapsed into a point charge at the center and the charge outside this sphere is ignored (Fig. 2B). Now include the nuclear charge in the consideration. At any observation point $\vec{r}$, the contributions to the potential gradient $\vec{\nabla}\Phi$ by electrons inside the sphere of radius r and by the associated nuclear charges cancel according to Fig. 2B. The rest of the nuclear charge remains uncompensated. Hence, the potential gradient $\vec{\nabla}\Phi$ is equal to the potential that would result if the electrons outside the sphere of radius r were collapsed into the nucleus with inverse charge, and all other charge was removed (Fig. 2C). With the usual $\Phi(\infty) = 0$ gauge, one gets:

$$\Phi(r) = \int_r^\infty |\vec{\nabla}\Phi|dr = \int_r^\infty \frac{\int_\mu^\infty 4\pi\nu^2\rho(\nu)d\nu}{\mu^2}d\mu \tag{19}$$

The equation indicates that for neutral atoms the potential at a point $\vec{r}$ is only dependent on the electrons *outside* a sphere of radius r from the EO point of view.

Approximate real space relation between ESP and ED

The exact relation between ED and ESP in equation (19) is non-local, and therefore not very intuitive. By contrast, equation (1), which is based on the TF model, has a simple interpretation. It expresses a power law relationship between the ED and ESP, provided the quantum chemical shell structure can be ignored and $\Phi_{TF}(r)$ can be identified with the ESP, which is permissible in the large Z limit for which TF theory is intended. The applicability and accuracy of the power law relating ED and ESP can be tested using the CM

11

model. The expression for the ED can be taken directly from equation (7). A semi-analytically tractable (see Methods) expression for the ESP is obtained by plugging equation (6) into equation (15) and applying a radial Fourier transform.

$$\Phi(r) = \frac{1}{r} \cdot \frac{2}{\pi} \int_0^\infty \left(Z - \sum_{j=0}^5 a_j e^{-(qr)^2 \cdot b_j/(4\pi r)^2} \right) \frac{\sin(qr)}{qr} d(qr) \qquad (20)$$

By eliminating r from the expressions (7) and (20) for the ED and ESP, one can relate map values over several orders of magnitude, independent of the TF approximation and the identification of Φ_{TF} with Φ. For large Z atoms, the dependence $\ln(\text{ESP})$ on $\ln(\text{ED})$ is well described by a straight line of slope $2/3$, as equation (1) of TF theory predicts. For second row elements, and to a lesser extent third row elements, with smaller Z, $\ln(\text{ESP})$ still depends linearly on $\ln(\text{ED})$ to a reasonably good approximation, but the slope is somewhere between $2/3$ and 1 (Fig. 3).

The power law relating ESP and ED with exponent close to 1 indicates that the maps for neutral spherical atoms are similar, except for scale factors. Both the ED and the ESP decrease gradually with increasing distance from the nucleus. As the exponent of the power law is less than 1, the ESP is slightly broader than the ED profile, particularly for atoms with larger order number Z, provided that B-factor blurring is absent or equal. This difference may be understood as a consequence of the long-range of the Coulomb potential. The intrinsic width of the ED and ESP map peaks has to be contrasted with the delta-function shape of the NCSL in the absence of thermal broadening. Formally, the delta function NCSL depends on the ED by a power law with an exponent approaching infinity, which is very different from the exponent close to 1 relating ESP and ED. At least for neutral atoms, it can therefore be concluded that ED and ESP maps are much more similar to each other than to NCSL maps (Fig. 4A).

Due to the reciprocity of lengths in real and reciprocal space, slightly broader ESP than ED profiles in real space translate to form factors that are slightly broader for X-ray than for electron scattering. In both cases, they fall off with inverse resolution, or with increasing momentum transfer $4\pi\sin\theta/\lambda$. In contrast, the form factor for neutron scattering (the scattering length) is a constant, independent of resolution or momentum transfer (Fig. 4B).

All space ED and ESP integrals

The ED integral over all space is simply the electron count Z.

$$\int \rho(\vec{r}) d^3 r = Z \qquad (21)$$

The ESP integral over all space can be calculated in reciprocal space with equation (15), as detailed in the Supplement, or in real space by integration by parts

12

with reference to equation (19):

$$\int_0^\infty 4\pi r^2 dr \Phi(r) = \int_0^\infty 4\pi \frac{r^3}{3} \cdot \frac{\int_r^\infty 4\pi \nu^2 \rho(\nu) d\nu}{r^2} dr =$$
$$\frac{4\pi}{3} \int_0^\infty \frac{r^2}{2} 4\pi r^2 \rho(r) dr = 2\pi Z \frac{\langle r^2 \rangle}{3} \tag{22}$$

Equations (21) and (22) are general for neutral atoms and neutral, spherical atoms, respectively. Taking into account that $\int_0^\infty 4\pi r^2 dr \Phi(r) = \tilde{\Phi}(0) = 2\pi f_e(0)$ according to equation (17), equation (22) is equivalent to the results of Ibers and O'Keeffe and Spence for the electron scattering form factor in forward direction (Ibers 1958; O'Keeffe and Spence 1994). For the CM model, equations (21) and (22) specialize to:

$$\int_0^\infty 4\pi r^2 dr \rho(r) = \sum_{j=0}^5 a_j \tag{23}$$

$$\int_0^\infty 4\pi r^2 dr \Phi(r) = \frac{1}{4\pi} \sum_{j=0}^5 a_j b_j \tag{24}$$

Equations (23) and (24) show that electrons contribute equally to the all space integral in case of the ED, but not the ESP. In case of the ESP, electrons contribute proportionally to the b_j values of their CM shells. Therefore, all space ESP integrals depend strongly on the quantum chemical shell structure of the atoms. With each new row in the periodic table, the all space ESP integral rises rather abruptly. This increase can be attributed to the strongly increasing mean square radius $\langle r^2 \rangle$ when a shell with a new main quantum number (K, L, M...) starts to fill. The integration limit ESP decreases with Z as deeper d- and f-quantum chemical shells are filled. In these situations, the shrinkage of the mean square atom radius with Z dominates over the additional ESP contribution from newly added electrons. For p-shells the effects depend on the main quantum number. Decreases of the integration limit ESP are seen for low main quantum numbers, while increases are seen for high main quantum numbers (Fig. 5).

Integrated and peak map values

Maps are usually displaced as isosurfaces at a given contour level, chosen somewhat arbitrarily with respect to the root mean square deviation (rmsd) of the map. For theoretical description, integrated and peak map values are more tractable. However, there is a problem at the atomic centers. ESP peak values diverge at the atomic center. So do ED values, but only in the TF and CM approximations (when $a_i > 0$ and $b_i = 0$). Fortunately, the ED and ESP divergences are weak. According to the equations (4) and (3), the ED and ESP diverge as $r^{-3/2}$ and r^{-1}, respectively. This divergence is weaker than the

growth of the differential sphere volume $4\pi r^2 dr$ with r. Therefore, the map values become finite as soon as any kind of averaging is involved. Map values that are integrated to a maximum radius $\hat{R}$, and peak map values in the presence of B-factor broadening or finite resolution R are effectively averaged, with or without some weighting. These differences should matter in the general case, but not in the $\hat{R}, B, R \to 0$ and $\hat{R}, B, R \to \infty$ limiting cases. In the following, I refer to these limits as the "peak limit" and the "integration limit" respectively. In the latter limit, the integrated map values become equal to the all space integrals discussed earlier. In the following, I calculate values for the peak and integrated ED and ESP, derive the dependence on the atom order number Z, and clarify the contribution of inner and outer electrons.

ED and ESP Z-dependence in the limits (TF model)

Some scaling rules for integrated and peak map values for neutral spherical atoms can be deduced independently of detailed atom models. In atomic units, the ED is equal to Z in the integration limit. Moreover, the ESP is proportional to Z in the peak limit, because the potential close to the nucleus is Z/r, since almost all the electron density is outside the sphere of radius r and therefore does not contribute to the ESP at r. The remaining rules require the power law of equation (1), or the TF model. In the peak limit, the proportionality of the ESP to Z implies that the ED is proportional to $Z^{3/2}$. In the integration limit, the power law cannot be applied, because a power law for the summands does not imply a power law for the sum. In this limit, the ESP should be proportional to $Z\langle r^2 \rangle$ according to equation (22). Since $\langle r^2 \rangle \propto Z^{-2/3}$ according to equation (5), the ESP should grow as $Z^{1/3}$ in the integration limit, in reasonable agreement with CM calculations (Fig. 4B). Taken together, the scaling rules suggest that the ED is more dependent on Z than the ESP, and that the dependence of both the ED and the ESP is stronger in the peak than in the integration limit (Table 2).

ED and ESP Z-dependence in general (CM model)

Calculations of the Z-dependence of integrated and peak map values in the general case are easier to do in the CM model than the TF model. Integrated ED map values can be calculated analytically. For integrated ESP map values, only a numerical solution was found. For peak map values, analytical solutions were found in all cases. The necessary calculations are tedious. In the main text and Fig. 6, I state only the results. The full derivations are presented in the Supplement.

ED: For the ED one gets with reference to Fig. 6:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \rho(r) = \sum_{j=1}^{5} a_j w_{ED}(b_j, \hat{R}) \qquad (25)$$

14

$$\rho_B(0) := \frac{1}{(2\pi)^3} \int_0^\infty dq \cdot 4\pi q^2 \tilde{\rho}(q) e^{-Bs^2} = \left(\frac{4\pi}{B}\right)^{3/2} \sum_{j=1}^{5} a_j w_{ED}(b_j, B) \quad (26)$$

$$\rho_R(0) := \frac{1}{(2\pi)^3} \int_0^{2\pi/R} dq \cdot 4\pi q^2 \tilde{\rho}(q) = \frac{4\pi}{3} \frac{1}{R^3} \sum_{j=1}^{5} a_j w_{ED}(b_j, R) \quad (27)$$

ESP: For the ESP, calculations require equation (15). One gets, also with reference to Fig. 6:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \Phi(r) = 2\pi \hat{R}^2 \sum_{j=1}^{5} a_j w_{ESP}(b_j, \hat{R}) \quad (28)$$

$$\Phi_B(0) := \frac{1}{(2\pi)^3} \int_0^\infty dq \cdot 4\pi q^2 \tilde{\Phi}(q) e^{-Bs^2} = \frac{4\sqrt{\pi}}{\sqrt{B}} \sum_{j=1}^{5} a_j w_{ESP}(b_j, B) \quad (29)$$

$$\Phi_R(0) := \frac{1}{(2\pi)^3} \int_0^{2\pi/R} dq \cdot 4\pi q^2 \tilde{\Phi}(q) = \frac{4}{R} \sum_{i=j}^{5} a_i w_{ESP}(b_j, R) \quad (30)$$

Interpretation of the w_{ED} and w_{ESP}

The w_{ED} or w_{EPS} in equations (25-30) measure the fractional contribution of an electron to the ED or ESP (from the EO point of view) compared to what would be expected for a unit point charge at the origin. For any fixed $\hat{R}, B, R$, the $\sum a_j w_j$ sum with w_j values from equations (25-30) is proportional to the integrated or peak map values, with a proportionality constant that does not depend on Z or the b_j values. Results for typical $\hat{R}, B, R$ are shown in Fig. 7. The data show that the ED grows fairly steadily with Z, regardless of the $\hat{R}, B, R$. In comparison, the ESP grows less smoothly with Z. Ripples in the ESP dependence on Z in comparison to the overall growth are pronounced in the large $\hat{R}, B, R$ regime. This is expected, since the integrated and peak map values in this regime become proportional to $\sum a_j b_j$, to which outer CM shell electrons make strong contributions. The ED/ESP ratio grows with Z as anticipated. The power laws of Table 2 imply growth with a slope between $1/2$ and $2/3$ on a double logarithmic plot. The CM results are closer to the exponent $1/2$. Further calculations show that this is the case because the $ESP \propto Z^{1/3}$ regime is hard to reach. Moreover, the ESP integral depends on a contribution from the region of large r, where the TF approximation is poor (Landau and Lifshitz 1977).

15

b_j dependent electron contributions to ESP and ED

Electrons in small b_j and large b_j CM shells generally do not contribute equally to peak and integrated map values. The expressions of equations (25-30) depend on a b_j independent prefactor, and the b_j dependent w_{ED} or w_{ESP} terms. The prefactors in equations (25-30) are chosen to represent the contribution of a point-like unit charge at the nucleus. With this choice, the w_{ED} and w_{ESP} can take values between 0 and 1 for the whole (nominal) b_j range from 0 to ∞. The w_{ED} and w_{ESP} values can thus be interpreted as weights that measure the contribution of electrons in different CM shells to integrated or peak map values. The plots in Fig. 6 show that electrons in inner CM shells contribute more to peak or integrated ED map values than electrons in outer CM shells. In the case of the ESP, the absolute value of the weights decreases with b_j only for excess charges. From the EO point of view, the opposite is true for electrons in neutral atoms. Those in inner CM shells contribute less than those in outer CM shells to peak and integrated ESP values. Qualitatively, this result can be understood as a consequence of incomplete cancellation of nuclear and electron potentials. The smaller the b_j, the better the cancellation, and the smaller the remaining net potential. However, contributions become (nearly) b_j independent in the large and small $\hat{R}, B, R$ regimes for ED and ESP maps, respectively. While the large $\hat{R}, B, R$ limit is easily reached, the small $\hat{R}, B, R$ limit is not, because some of the CM b_j values are always small compared to typical $\hat{R}, B, R$ values. One b_j value may even be 0, making the limit completely inaccessible. Therefore, the conclusion that large b_j electrons contribute more to peak and integrated ESP values than small b_j electrons is always true in practice. In contrast, the statement that small b_j electrons contribute more to peak and integrated ED than large b_j electrons needs to be qualified. In the large $\hat{R}, B, R$ regime, electrons contribute equally to the ED (Fig. 6 and Table 3).

NCSL

On the atomic scale, the nuclei are point-like. Moreover, nuclear interactions with neutrons have a very short range (Segre 1977). Therefore, the scattering of neutrons from nuclei is of s-wave type (Landau and Lifshitz 1977) and therefore isotropic, and the form factor does not depend on $\sin(\theta)/\lambda$. It can be described by a single number, the nuclear coherent scattering length (NCSL), which determines integrated and peak NCSL map values. For isotope mixtures, the scattering length can be averaged. Table 4 lists the values for biologically important atoms, assuming natural isotope abundances (Sears 1992). The NCSL is positive for most elements, but shows no systematic dependence on Z. The most prominent example of a biologically relevant element with a negative NCSL is ^{1}H. Since the scattering length of ^{2}H is positive, the hydrogen signal in NCSL maps can be tuned continuously from negative to positive values by increasing deuteration. The NCSL of ^{2}H is similar in magnitude to that of other elements. Therefore, the hydrogen atoms in deuterated biological samples are very well

16

resolved. Deuteration of biological samples also minimizes unwanted neutron capture because the neutron capture cross section of ^{2}H is exceptionally small (Sears 1992). Therefore, biological samples are usually deuterated for neutron diffraction experiments.

Excess charge

Neutral atom theory requires corrections when excess charges are present. Such charges can be either "true" net charges, or excess charges from bond polarization. "True" net charges are typically compensated by ions from the buffer. From an electrostatic point of view, the charge pairs are similar to partial charge pairs, except that there is no covalent bond. The quantitative discussion in this manuscript is limited to the simplest case of spherically symmetric excess charge. Spherical symmetry may be a good approximation for "true" net charge in cations (e.g. Na^+, K^+, Mg^{2+}, Ca^{2+}) or anions (e.g Cl^-), but is clearly a poor description for charges from bond polarization. The qualitative arguments, however, are more general, and therefore conclusions are expected to be qualitatively applicable also to non-spherically symmetric excess charge. For a quantitative treatment, calculations that go beyond the spherical atom approximation are required (Kulik, Chodkiewicz, and Dominiak 2022).

Excess charge has opposite effects on ED and ESP

In the case of the ED, excess charge simply adds to the charge of the electron cloud of the neutral atom. Therefore, negative excess charge increases the ED and positive excess charge decreases it. In the case of the ESP, excess negative charge adds to the shielding of the nuclear charge, while excess positive charge reduces the shielding. Therefore, excess negative charge decreases the ESP and excess positive charge increases it. The effect explains the poor visibility of negatively charged amino acid carboxylate groups (Yonekura et al. 2015; Wang 2017b), the weaker than expected signal for phosphate groups (Wang and Moore 2017), and the excellent visibility of metal carbocations (Wang et al. 2021a) in ESP maps. Hydrogen atoms carrying partial positive charge should be more visible in ESP than in ED maps. This difference was indeed noted when high resolution ESP maps obtained by cryo-EM became available for comparison with ED maps of the same resolution (Yip et al. 2020; Nakane et al. 2020). In the absence of magnetic scattering, neutrons as a probe are completely insensitive to excess charge, except for indirect effects of charge on geometry (Fig. 8A).

Excess charge affects ESP more drastically than ED

Excess charge perturbs the quantum chemical shell structure of atoms only on the periphery, inner quantum chemical shells are expected to be filled and largely unperturbed. Excess charge is therefore expected to be associated with a b_{ex} value (or several b_{ex} values) that are comparable to the largest b_j values for electrons of the neutral atom. For the ED, large b_j values are associated with

17

small contributions to integrated or peak map values (Fig. 6, column ED). For ED maps, excess charge therefore competes with all neutral atom electrons, on less or equal terms. For the ESP, potentials due to electrons and associated nuclear charges cancel partly. This cancellation is particularly good for inner CM shell electrons. Therefore, excess charge mostly competes with outer CM shell electrons only. Additionally, competition is on better or equal terms, except for peak values when the resolution is very good and the temperature factor very low $(B > b/(1.7)^2)$ (Fig. 6, column ESP). As excess charge effectively competes more favorably and with fewer neutral atom electrons in case of the ESP than in case of the ED, the ESP is more sensitive to excess charge than the ED. This difference between ED and ESP maps has been noted earlier, based on comparisons of both simulated and experimental maps (Hirai et al. 2007; Wang 2017a). In a materials science context, the result has been stated as a finding in reciprocal space. "Below a critical scattering vector, s ($|s| = sin(\theta)/\lambda$), ranging typically from 0.2 to 0.6 $\overset{\circ}{\text{A}}^{-1}$ depending on the atomic number, electron diffraction is more sensitive to valence charge densities than X-ray diffraction" (Zheng et al. 2005). As $R = 1/(2s)$, critical scattering vectors of 0.2 $\overset{\circ}{\text{A}}^{-1}$ and 0.6 $\overset{\circ}{\text{A}}^{-1}$ correspond to resolutions of 2.5 Å and 0.83 Å, respectively. As most ESP maps have only Fourier components below this threshold, the statement implies that ESP maps are more sensitive to charge than ED maps at the same resolution (Fig. 8B).

Excess charge has a relatively larger effect on the ESP at low than at high resolution

At low resolution, when the temperature factors are high, or when a large integration radius is chosen, the difference between the point-like nuclear charge and the more diffuse electron cloud of neutral atoms matters little, and the potentials compensate quite well. Such compensation does not occur for excess charges. Hence, excess charges make an important contribution to low resolution ESP maps. For high resolution, low temperature factors, and when the integration radius is small, the compensation of nuclear and electronic contributions to the potential occurs only to a lesser extent. The potential for the neutral atom (at the nucleus) is therefore larger. On the scale of this larger neutral atom contribution, the contribution of excess charges to the ESP appears smaller. Thus, excess charges have a more striking effect on ESP map appearance at low than at high resolution, even though the high resolution map is objectively more informative (higher resolution data can always be truncated). The resolution dependence of excess charge effects on ESP map appearance has been noted previously by other authors (Hirai et al. 2007; Wang 2017b) (Fig. 8B).

A comparison of actual ED and ESP maps at low and high resolution

To illustrate charge effects independent of potentially confounding radiation damage effects, theoretical ED and ESP maps were calculated based on RHF atom models (Fig. 9A). The dipeptides Glu-Arg and Glu-Gln were chosen to

18

compare anionic (Glu), cationic (Arg) and neutral (Gln) side chains. To avoid spurious charge effects at the termini, the dipeptides were assumed to be amidated at the C-terminus and formylated at the N-terminus. To simulate low and high resolution maps, the ideal ED and ESP maps from the quantum chemical calculations were convoluted with point-spread functions corresponding to temperature factors of 45 Å^2 and 10 Å^2, respectively. The simulated maps are consistent with the theoretical expectation that charge has opposite effects on ED and ESP maps, affects ESP maps more than ED maps, and is more noticeable in low than in high resolution ESP maps (note the Glu carboxylate oxygen atoms outside the map isocontour surface, black arrows) (Fig. 9A).

To illustrate charge effects for experimental maps, available apoferritin ED and ESP maps were analyzed (Fig. 9B). Apoferritin was selected as a model protein because of the availability ED and ESP maps of similar resolution in the low (Wang et al. 2021b; Hamdi et al. 2020) and high (Val et al. 2012; Yip et al. 2020) resolution regimes. Asp/Asn and Glu/Gln pairs were chosen for analysis, because of their different charge and near isostericity, which should result in similar temperature factors. To further minimize temperature factor effects, average peak map values were determined for all Asp and Asn OD1 and all Glu and Gln OE1 atoms. For the low resolution ESP map, the Asp(OD1)/Asn(OD1) and Glu(OE1)/Gln(OE1) ratios of averages were small, reflecting poor carboxylate visibility. For the other maps, ratios were closer to 1, reflecting similar visibility of carboxylates and carboxamides. Note that experimental Asp(OD1)/Asn(OD1) and Glu(OE1)/Gln(OE1) ratios may be even lower than theoretically expected due to radiation damage (Petrova et al. 2009; Shelley and Garman 2022; Hattne et al. 2018) and the common practice to offset maps to a mean at or near 0. At least qualitatively, the theoretically expected effects of charge are also seen in the experimental maps (Fig. 9B).

Discussion

This work describes theoretical expectations for ideal ED, ESP, and NCSL maps measured using X-rays, electrons, and neutrons as probes (Fig. 1). Atoms in ED and ESP maps have intrinsic widths, whereas atoms in NCSL maps are effectively point-like (at the atomic scale) in the absence of broadening due to thermal motion or limited resolution. DFT theory (or TF theory in the special case of neutral spherical atoms) suggests that ESP and ED are related by a power law with an exponent different from, but similar to, 1 (Figs. 3 and 4). Thus, the (scaled) maps are similar but not identical. As a consequence of the long-range nature of electrostatic interactions, atoms appear intrinsically wider in ESP maps than in the ED maps. Due to Fourier reciprocity, the reverse is true for the form factor profiles (Fig. 4). On average, integrated and peak map values increase with atom order number Z for both ED and ESP maps, but the increase is stronger for ED than ESP maps. There is no systematic Z-dependence for NCSL maps (Table 2). For ESP, but not ED maps, integrated or peak map values can decrease in some Z regions, for large integration radius,

19

high B-factor, or poor resolution (Figs. 5 and 7). This effect, together with charge effects, predicts that ESP maps should have slightly stronger integrated map values for carbon than for oxygen, while the reverse should be true for ED maps. Electrons in inner CM shells contribute equally or more to the ED than electrons in outer CM shells. The opposite is true for ESP maps (from the EO point of view) (Table 3 and Fig. 6). Excess or partial charges have opposite effects on the ED and ESP maps, and no effect on the NCSL maps. In ESP maps, they are measured more accurately at higher resolution, but their effect on map isosurfaces appears stronger at lower resolution (Fig. 8 and 9). Overall, this paper highlights the differences between ED, ESP, and NCSL maps, but also shows that the former two are more similar to each other than the latter, which is in line with the common perception.

Acknowledgement

Author contributions

The author is responsible for all work.

Declaration of Interests

The author declares no competing interests.

References

Abbe, E. (1873). Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. Archiv für Mikroskopische Anatomie. Archiv für Mikroskopische Anatomie 9, 413–418.

Agirre, Jon et al. (2023). The *CCP*4 suite: integrative software for macromolecular crystallography. Acta Crystallographica Section D 79, 449–461.

Bahar, I., A. R. Atilgan, and B. Erman (1997). Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Folding and Design 2, 173–181.

Bai, X.-c Xiao-chen, G. McMullan, and S. H. W. Scheres (2015). How cryo-EM is revolutionizing structural biology. TIBS 40, 49–57.

Barca, Giuseppe M. J. et al. (2020). Recent developments in the general atomic and molecular electronic structure system. The Journal of Chemical Physics 152, 154102.

Baumeister, W. (2020). The promise and the challenges of cryo-electron tomography. FEBS Letters 594, 3243–3261.

Becke, A. D. (2014). Perspective: Fifty years of density-functional theory in chemical physics. J. Chem. Phys. 140, 18A301.

Bethe, H. (1930). Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. Annalen der Physik 397, 325–400.

Blakeley, M. P., P. Langan, and A. Niimura N. and Podjarny (2008). Neutron crystallography: opportunities, challenges, and limitations. Current Opinion in Structural Biology 18, 593–600.

Bronstein, I. N. and K. A. Semendjajew (1991). Taschenbuch der Mathematik. Stuttgart: Teubner Verlagsgesellschaft.

Clabbers, M. T. B and J. P. Abrahams (2018). Electron diffraction and three-dimensional crystallography for structural biology. Crystallography Reviews 24, 176–204.

Clabbers, M. T. B., A. Shiriaeva, and T. Gonen (2022). MicroED: conception, practice and future opportunities. IUCrJ 9, 169–179.

Clabbers, M. T. B. et al. (2019). Reducing dynamical electron scattering reveals hydrogen atoms. Acta Crystallogr A Found Adv. 75, 82–93.

Condon, E. U. (1936). Note on electron-neutron interaction. Phys. Rev. 49, 727–734.

Cromer, D. T. (1969). Compton Scattering Factors for Aspherical Free Atoms. J. Chem. Phys. 50, 4857–4859.

Drenth, J. (1999). Principles of Protein X-ray Crystallography. New York: Springer.

Dubochet, J. (2012). Cryo-EM—the first thirty years. J. Microscopy 245, 221–224.

Foldy, L. L. (1958). Neutron-electron interaction. Reviews of Modern Physics 30, 471–481.

Gauss, Carl Friedrich (1877). Theoria attractionis corporum sphaeroidicorum ellipticorum homogeneorum, methodo nova tractata. *Werke: Fünfter Band*. Berlin, Heidelberg: Springer Berlin Heidelberg, 279–286.

Gemmi, M. et al. (2019). 3D Electron Diffraction: The Nanocrystallography Revolution. ACS Cent. Sci. 8, 1315—1329.

Hall, H. (1936). The Theory of Photoelectric Absorption for X-Rays and gamma-Rays. Rev. Mod. Phys. 358–397, 8.

Hamdi, F. et al. (2020). 2.7 Å cryo-EM structure of vitrified M. musculus H-chain apoferritin from a compact 200 keV cryo-microscope. PLOS One 15, e0232540.

Hanwell, M. D. et al. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics 4, 17.

Harris, Charles R. et al. (Sept. 2020). Array programming with NumPy. Nature 585, 357–362.

Hattne, Johan et al. (2018). Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. Structure 26.5, 759–766.e4.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Q Rev Biophys 28, 171–193.

Hirai, T. et al. (2007). Simulation of charge effects on density maps obtained by high-resolution electron crystallography. Journal of Electron Microscopy 56, 131–140.

Hohenberg, P. and W. Kohn (1964). Inhomogeneous Electron Gas. Phys. Rev. B 136, 864–871.

Ibers, J. A. (1958). Atomic scattering amplitudes for electrons. Acta Cryst. 11, 178–183.

Jackson, J. D. (1998). Classical electrodynamics. New York: John Wiley and Sons.

Kleywegt, G. J. and T. A. Jones (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. Acta Cryst. D 52, 826–828.

Kulik, M., M. L. Chodkiewicz, and M. Dominiak (2022). Theoretical 3D electron diffraction electrostatic potential maps of proteins modeled with a multipolar pseudoatom data bank. Acta Cryst. D 78, 1010–1020.

Landau, L. D. and E. M. Lifshitz (1977). Quantum Mechanics: Non-Relativistic Theory. Vol. 3 (3rd ed.) Pergamon Press.

Lenz, F. (1954). Zur Streuung mittelschneller Elektronen in kleinste Winkel. Zeitschrift für Naturforschung A 9, 185–204.

Lu, T. and F. Chen (2012). Multiwfn: A Multifunctional Wavefunction Analyzer. J. Comput. Chem. 33, 580–592.

Lynn, J. W. (1936). Magnetic neutron scattering. J. Appl. Physics 75, 6806–6810.

Marques, M. A., M. D. Purdy, and M. Yeager (2019). CryoEM maps are full of potential. Curr. Opin. Struct. Biol. 58, 214–223.

Maxima (2022). Maxima, a Computer Algebra System. Version 5.44.0.

McCusker, L. B. (2017). Electron diffraction and the hydrogen atom. Science 355, 136–136.

Mott, N. (1965). The theory of atomic collisions. Oxford: Clarendon Press.

Nakane, Takanori et al. (2020). Single-particle cryo-EM at atomic resolution. Nature 587, 152–156.

O'Keeffe, M. and J. C. H. Spence (1994). On the average Coulomb potential (Φ_0) and constraints on the electron density in crystals. Acta Cryst. A 50, 33–45.

Palatinus, L. et al. (2017). Hydrogen positions in single nanocrystals revealed by electron diffraction. Science 355, 166–169.

Pauling, L. and E. B. Wilson Jr (1985). Introduction to Quantum Mechanics with Applications to Chemistry. New York: Dover Publications, Inc.

Petrova, Tatiana et al. (2009). X-Ray-Radiation-Induced Cooperative Atomic Movements in Protein. Journal of Molecular Biology 387, 1092–1105.

Rullgård, H. et al. (2011). Simulation of transmission electron microscope images of biological specimens. J. Microsc. 243, 234–256.

Saloman, E. B., J. H. Hubbell, and J. H. Scofield (1988). X-ray attenuation cross sections for energies 100 eV to 100 keV and elements Z = 1 to Z = 92. Atomic Data and Nuclear Data Tables 38, 1–196.

Scherzer, O. (1949). The Theoretical Resolution Limit of the Electron Microscope. Journal of Applied Physics 20, 20.

Schoenborn, B. P. and A. C. Nunes (1972). Neutron scattering. Annual Review of Biophysics and Bioengineering 1, 529–552.

Sears, V. F. (1992). Neutron scattering lengths and cross sections. Neutron News 3, 26–37.

Segre, E. (1977). Nuclei and Particles. Reading, Mass: W. A. Benjamin.

Shelley, K. L. and E. F. Garman (2022). Quantifying and comparing radiation damage in the Protein Data Bank. Nat. Comm. 13, 1314.

Shin, S. (2021). New era of synchrotron radiation: fourth-generation storage ring. AAPPS Bulletin 31, 21.

Shull, C. G. and E. O. Wollan (1948). X-Ray, Electron, and Neutron Diffraction. Science 108, 69–75.

Squires, G. L. (1996). Introduction to the theory of thermal neutron scattering. Dover publications.

Thomas, L. H. (1927). The calculation of atomic fields. Mathematical Proceedings of the Cambridge Philosophical Society 23, 542–548.

Val, N. de et al. (2012). Structural analysis of haemin demetallation by L-chain apoferritins. J. Inorg. Biochem. 112, 77–84.

Virtanen, Pauli et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272. DOI: 10.1038/s41592-019-0686-2.

Wang, J. (2017a). Experimental charge density from electron microscopic maps. Protein Science 26, 1619–1626.

— (2017b). On the appearance of carboxylates in electrostatic potential maps. Protein Science 26, 396–402.

Wang, J. and P. B. Moore (2017). On the interpretation of electron microscopic maps of biological macromolecules. Protein Science 26, 122–129.

23

Wang, Jimin et al. (2021a). Identification of Mg$^2+$ ions next to nucleotides in cryo-EM maps using electrostatic potential maps. Acta Crystallographica Section D 77, 534–539.

Wang, Y. et al. (2021b). Structural insights for the stronger ability of shrimp ferritin to coordinate with metal ions as compared to human H-chain ferritin. Int. J. Mol. Sci. 22, 7859.

Warshamanage, R, K. Yamashita, and G. N. Murshudov (2022). EMDA: A Python package for Electron Microscopy Data Analysis. Journal of Structural Biology 214, 107826.

Wentzel, G. (1927). Zur Theorie des Comptoneffekts. Zeitschrift für Physik 43, 1—8.

Wlodawer, A., M. Li, and Z. Dauter (2017). High-Resolution Cryo-EM Maps and Models: A Crystallographer's Perspective. Structure 25, 1589–1597.

Yip, K.M. et al. (2020). Atomic-resolution protein structure determination by cryo-EM. Nature 587, 157–161.

Yonekura, K. et al. (2015). Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. Proc. Nat. Acad. Sci. 112, 3368–3373.

Zhang, J. and T. Liu (2021). Efficient evaluation of electrostatic potential with computerized optimized code. Phys.Chem. Chem. Phys. 23, 20323–20328.

Zheng, J.C. et al. (2005). On the sensitivity of electron and X-ray scattering factors to valence charge distributions. J. Appl. Cryst. 38, 648–656.

Figure legends

Figure 1: X-rays, electrons, and neutrons as atomic resolution probes of matter. (A) Typical energies and wavelengths (B) Deflection of incident waves/particles (C) Cross sections (D) Sensitivity to atomic components. Contributions from electrons in blue, from nuclei in red, and combined contributions in purple.

Figure 2: Nuclear and electronic contributions to the ESP. (A) "Electrons Only" (EO) point of view with implicit nuclear charge. From the conventional point of view, the ESP is dominated by the nuclear contribution. Nonetheless, it is often convenient to describe the ESP exclusively in terms of features of the electron cloud. For this description, each electron is associated with one unit charge (one proton) in the nucleus, and its contribution to the potential gradient includes the contribution of the (implicit) proton. (B) Potential gradient due to the electrons only, for spherically symmetric electron clouds. According to Gauss's flux theorem (Gauss 1877; Jackson 1998), the electric field, and its inverse, the potential gradient $\vec{\nabla}\Phi_{el}$ depend only on the charge inside the sphere of radius r. The potential gradient $\vec{\nabla}\Phi_{el}$ at $\vec{r}$ remains therefore unchanged (curved dark blue arrow) when the charge inside the sphere of radius r (dark blue) is collapsed into a point charge, and the charge outside the sphere of radius r (light blue) is removed. (C) Potential gradient $\vec{\nabla}\Phi$ due to the electrons and the nucleus, for spherically symmetric electron clouds. According to (B), the contribution of the electrons inside r (dark blue) to the potential gradient cancels perfectly with the contributions of the associated nuclear charges. Therefore, their contribution to the potential gradient is 0 from the EO point of view. The electrons outside the sphere of radius r (light blue) do not contribute to the potential gradient from the conventional point of view. However, their associated nuclear charges do. From the EO point of view, these electrons are therefore responsible for the potential gradient $\vec{\nabla}\Phi$. This gradient remains unchanged (curved light blue arrow) if the outer electrons (light blue) are collapsed with inverted charge into the nucleus, and all other charges (dark blue and red) are removed. From the EO point of view, the potential gradient at $\vec{r}$ can therefore be attributed to the electrons outside the sphere of radius r (albeit with inverted sign). Qualitatively, this argument explains why electrons in outer CM shells contribute more to the ESP than those in inner CM shells. For a quantitative treatment, the potential and its gradient have to be distinguished, as is done in the text.

Figure 3: Test of the power law relating ED and ESP. Black lines show the results of CM calculations for atoms with order number 6 (carbon), 10, 20, .., 90. Note that arbitrary offsets have been added to avoid overlap (3, 6, .., 27 for Z=10, 20, ..90) to avoid overlap between the curves. The thick continuous grey lines shows the expectation based on TF or lowest order DFT theory, i.e. they have a slope of 2/3. The dotted line has a slope of 1, which would make ESP and ED maps identical except for scale factors. For large Z, the calculated slopes are very close to the expected value of 2/3. For small Z, the CM model suggests some deviations from the power low. The dependence in the log-log plot can be roughly modeled with slopes between 2/3 and 1. The curves are calculated for the radial range from 0.05 to 3 Bohrs (0.26 Å to 1.56 Å). For smaller values, the CM Gaussians model the cusp of the ED too inaccurately. For larger values, the numerical integration becomes increasingly unreliable.

Figure 4: Maps and form factors for X-rays, electrons and neutrons as the probe. (A) Real space. In the absence of B-factor blurring, atoms should appear point-like in NCSL maps, but not in ED and ESP maps. Equation (1) predicts that atoms appear slightly broader in ESP than ED maps when B-factor blurring is absent or equal. (B) Reciprocal space. Form factors are related to (spherical) atom map values by a radial Fourier transform. Because of the inverse relationship between lengths in real and reciprocal space, form factors are broadest (resolution independent) for neutrons as the probe, and slightly sharper for electrons than for X-rays as the probe.

Figure 5: Dependence of map values on Z for neutral spherical atoms in the integration limit. The ordinate shows the integrated map values. Calculations were performed using the revised CM coefficients from the International Tables for X-ray crystallography. (A) Linear plot of the dependence of the integrated ESP on Z. The integrated ESP increases strongly when a new main quantum number shell (K, L, M...) is started and tends to decrease in between, except when p electrons are added. (B) Double logarithmic plot of the dependence of the integrated ESP on Z. The TF model predicts the straight dashed line with a slope of 1/3 (the offset was fitted, not calculated). This slope is in reasonable agreement with the CM model. The sensitivity of the ESP map values contrasts with the insensitivity of the ED map values to the quantum chemical shell structure of the atoms in the integration limit.

Figure 6: ED (left) and ESP (right, EO point of view) contributions of a unit charge, relative to the contribution of a unit point charge at the nucleus. For the ED, the black line applies to electrons in neutral atoms ($b = b_j$) and to excess electrons ($b = b_{ex}$). For the ESP, the black line applies to electrons in neutral atoms ($b = b_j$), and the grey dotted line to excess positive charge ($b = b_{ex}$). Grey dotted and black lines add to 1 (as the potentials cancel, except for the potential from a unit charge at the nucleus). Dashed lines indicate the midpoint between the small and large b_j regimes, where $w=1/2$. The data show that electrons in inner CM shells are more important for the integrated and peak ED map values than those in outer CM shells. For the ESP, the reverse is true. From the EO point of view with implicit nuclear charge, outer CM shell electrons contribute more than inner CM shell electrons.

Figure 7: Z-dependence of integrated and peak map values for selected integration radii $\hat{R}$, B-factors B, and resolution cutoffs R, for neutral spherical atoms. To display the plots for different $\hat{R}$, B or R values in the same graph, only the $\sum_j a_j w_j$ is plotted. To get to the map values, the ordinate values in the plots must be multiplied by the prefactors in the equations (25-30). The data show that the dependence on Z is relatively smooth for the ED in any regime and for the ESP in the peak limit. The effects of the chemical shell structure of the atoms are most evident for the ESP in the regime close to the integration limit (second and third column, red line). In this regime, the map values decrease with Z in some regions of the plot. The ED/ESP ratio plots (right panels) confirm the TF model prediction that the ED is more strongly Z-dependent than the ESP.

Figure 8: Qualitative excess charge effects (A) Correction to the neutral atom model due to excess positive ("+", orange) or negative ("-", green) charge for X-rays, electrons and neutrons as the probe. Excess charge has opposite effects on ED and ESP and affects ESP more strongly than ESP. (B) Comparison of the effect of excess charge on ESP at high and low resolution. Poor resolution degrades the neutral atom signal more strongly than the excess charge signal. The schematic diagrams apply for peak and integrated map values.

Figure 9: Effects of excess charge on maps (A) Calculated ED and ESP maps for Glu-Arg and Glu-Gln dipeptides. To avoid spurious charges at the termini, aminotermini were capped by formyl groups and carboxytermini were amidated. Note the poor visibility of Glu carboxylate oxygen atoms in high temperature factor (i.e. low resolution) ESP, but not ED maps. With usual contouring, the effect is much less visible in high resolution maps. (B) Experimental low and high resolution ED and ESP maps for selected Asp, Asn, Glu, and Gln residues of apoferritins. As the structures were determined for apoferritins from different animal species, conserved residues were chosen as examples. For the calculation of Asp(OD1)/Asn(OD1) and Glu(OE1)/Gln(OE1) ratios, average OD1 and OE1 map values for all Asp, Asn, Glu, Gln residues of an apoferritin subunit were used. No attempt was made to second-guess map offsets chosen by the original authors. Note the very small ratios for ESP maps at low resolution.

Table legends
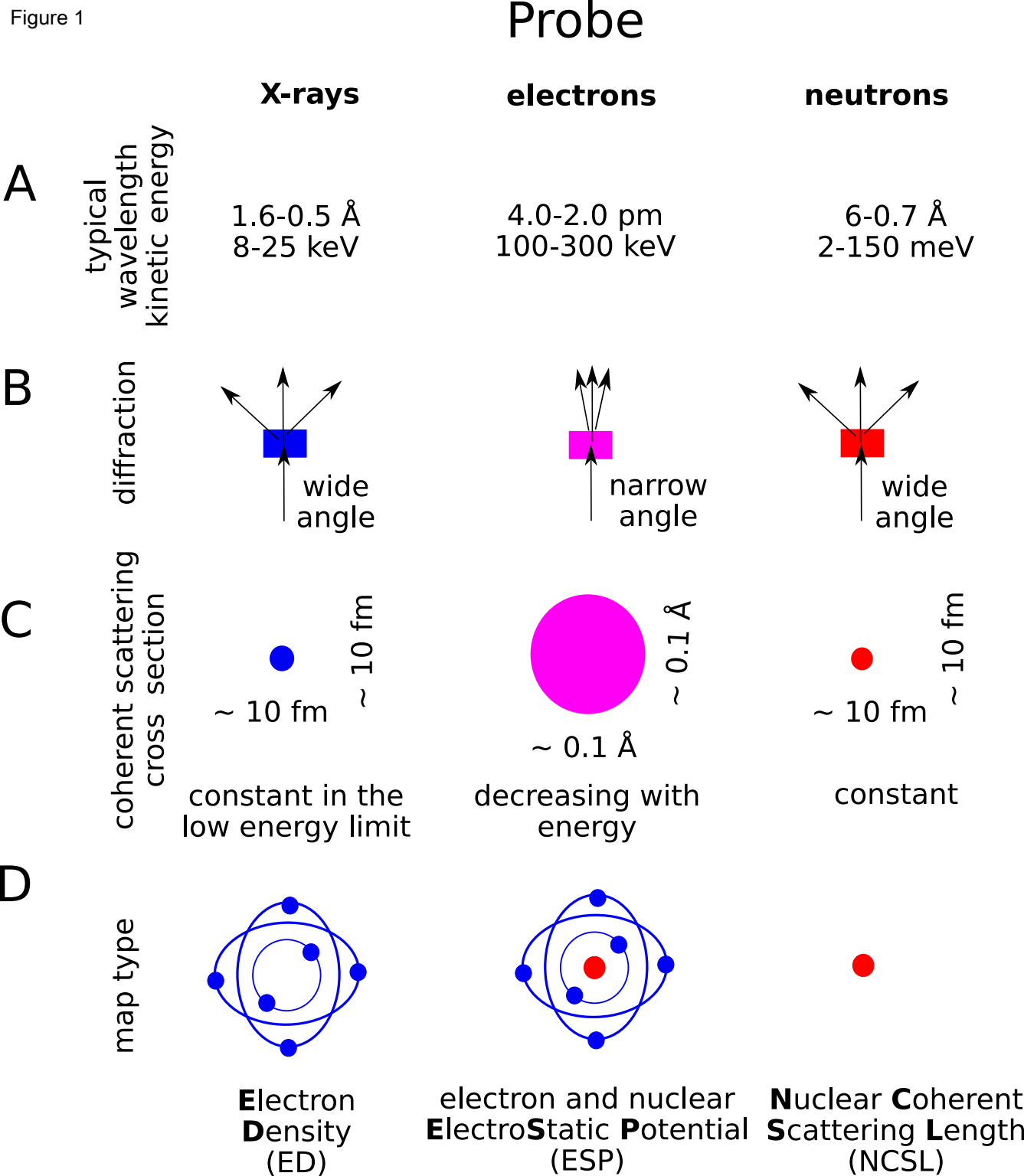
Table 1: Frequently used abbreviations in this work. ED, ESP and NCSL maps are measured using X-rays, electrons, and neutrons as probes, respectively.

Table 2: Power laws relating integrated or peak map values to Z. The exponents apply to the limiting cases of very large and small effective averaging areas. Exponents of 1 are independent of the atom model and should apply to any neutral atom. The other a values are calculated for the TF model that averages over the quantum chemical shell structure of atoms. For the NCSL, there is no systematic dependence of integrated or peak map values on Z. The Z-dependence of integrated and peak map values is strongest for the ED, intermediate for the ESP, and absent for the NCSL.
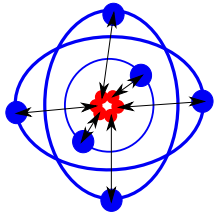
Table 3: b_j dependent contribution of a CM shell electron to the ED and ESP (from the EO point of view). Electrons with small b_j values reside in inner CM shells, those with large b_j values in outer CM shells. In practice, the regime of equal contributions from inner and outer CM shell electrons is not achievable for the ESP. Therefore, large b_j electrons always contribute more to the integrated and peak ESPs than small b_j electrons.

Table 4: Coherent neutron scattering lengths l in fm (not atomic units!) of biologically important elements, listed in order of increasing Z. For hydrogen, the ^{1}H and ^{2}H coherent scattering lengths are given separately. For the other elements, isotope weighted averages are given. The values are taken from a systematic compilation of scattering data for all elements (Sears 1992).

Figure 1

Probe

	X-rays	**electrons**	**neutrons**

A typical wavelength kinetic energy

1.6-0.5 Å
8-25 keV

4.0-2.0 pm
100-300 keV

6-0.7 Å
2-150 meV

B diffraction

wide angle

narrow angle

wide angle

C coherent scattering cross section

~ 10 fm
~ 10 fm

~ 0.1 Å
~ 0.1 Å

~ 10 fm
~ 10 fm

constant in the low energy limit

decreasing with energy

constant

D map type

Electron **D**ensity
(ED)

electron and nuclear **E**lectro**S**tatic **P**otential
(ESP)

Nuclear **C**oherent **S**cattering **L**ength
(NCSL)

Figure 20

A: BO picture

B: $\vec{\nabla}\Phi_{el}(\vec{r})$ (electrons only)
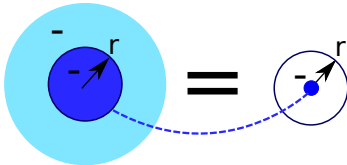
C: $\vec{\nabla}\Phi(\vec{r}) = \vec{\nabla}ESP$ (electrons and nuclei)

Figure 3
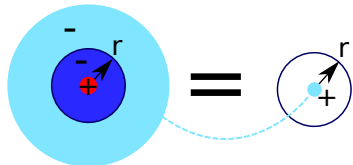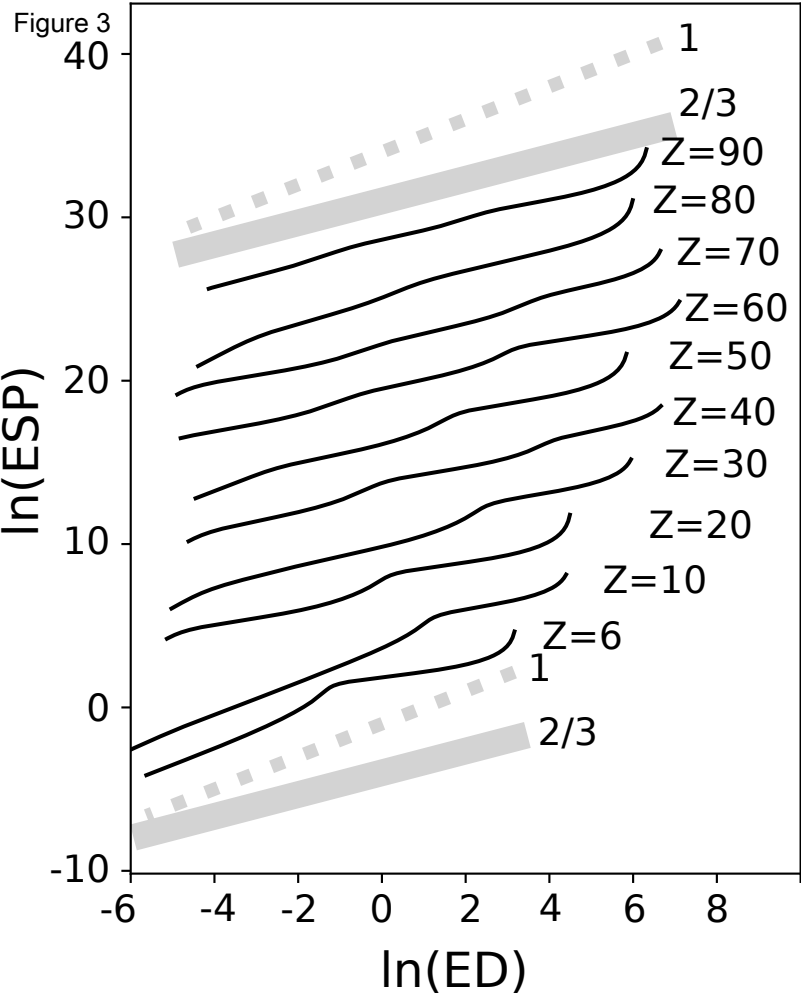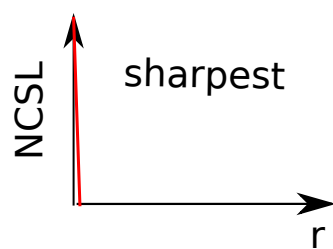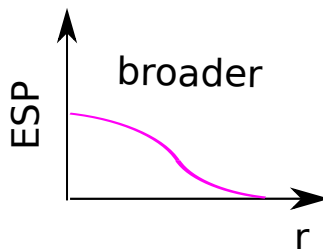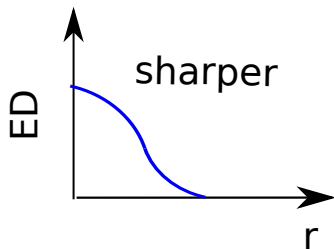
Figure 4

Probe

X-rays · electrons · neutrons

Figure 5

Z-dependence of the all space ESP integral
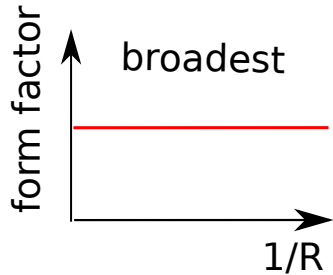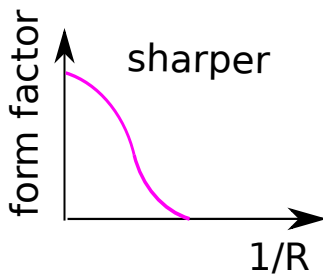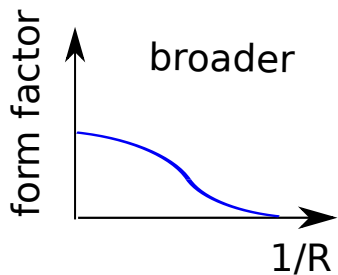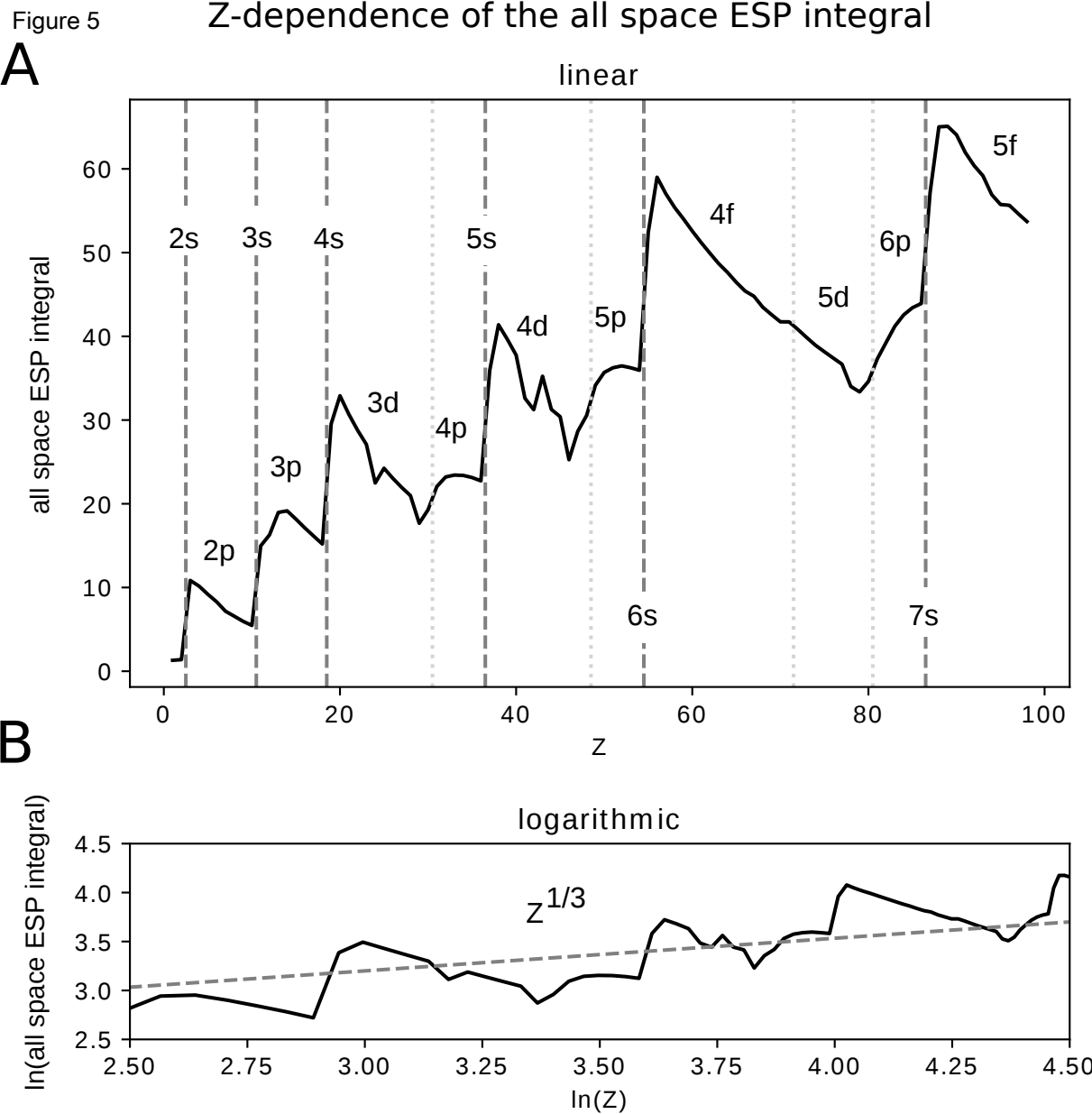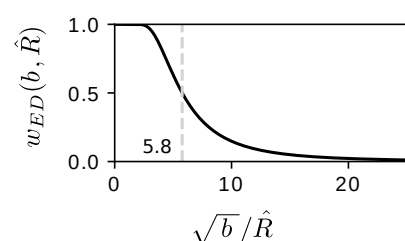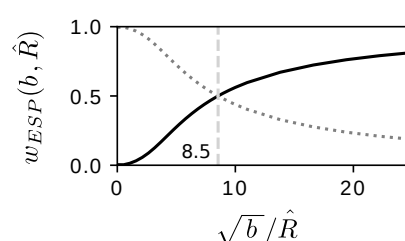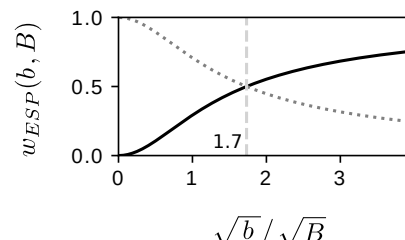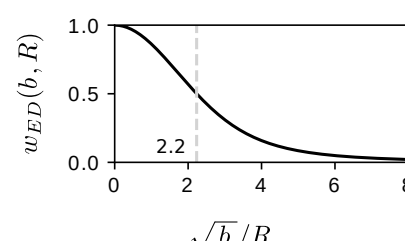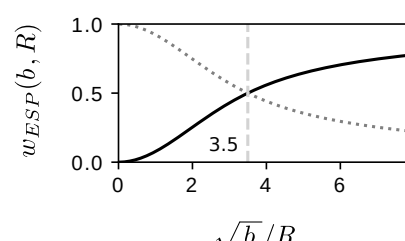
Figure 6	b dependent electron contributions to integrated and peak map values	
	ED	ESP

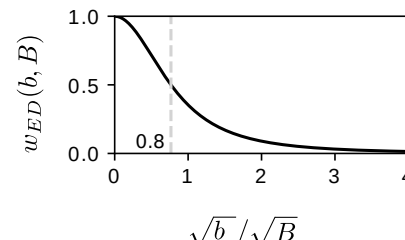|	ED	ESP
integrated value $\gamma = 2\pi\hat{R}/\sqrt{b}$	$w_{ED}(b,\hat{R}) = \mathrm{erf}\,\gamma - 2\gamma\mathrm{e}^{-\gamma^2}/\sqrt{\pi}$ $\approx \begin{cases} 1 & \text{if } \sqrt{b} \ll 2\pi\hat{R} \\ 32\pi^{5/2}/3 \cdot (b/\hat{R}^2)^{-3/2} & \text{if } \sqrt{b} \gg 2\pi\hat{R} \end{cases}$ *(plot: $w_{ED}(b,\hat{R})$ vs $\sqrt{b}/\hat{R}$, marker at 5.8)*	$w_{ESP}(b,\hat{R})$ $\approx \begin{cases} 1/(8\pi^2) \cdot b/\hat{R}^2 & \text{if } \sqrt{b} \ll 2\pi\hat{R} \\ 1 & \text{if } \sqrt{b} \gg 2\pi\hat{R} \end{cases}$ *(plot: $w_{ESP}(b,\hat{R})$ vs $\sqrt{b}/\hat{R}$, marker at 8.5)*
smooth resolution cutoff (B-factor)	$w_{ED}(b,B) = (1 + b/B)^{-3/2}$ $\approx \begin{cases} 1 & \text{if } b \ll B \\ (b/B)^{-3/2} & \text{if } b \gg B \end{cases}$ *(plot: $w_{ED}(b,B)$ vs $\sqrt{b}/\sqrt{B}$, marker at 0.8)*	$w_{ESP}(b,B) = 1 - (1 + b/B)^{-1/2}$ $\approx \begin{cases} 1/2 \cdot b/B & \text{if } \sqrt{b} \ll \sqrt{B} \\ 1 & \text{if } \sqrt{b} \gg \sqrt{B} \end{cases}$ *(plot: $w_{ESP}(b,B)$ vs $\sqrt{b}/\sqrt{B}$, marker at 1.7)*
peak value — **sharp resolution cutoff** $\beta = \sqrt{b}/(2R)$	$w_{ED}(b,R) = 3\sqrt{\pi}\left(\mathrm{erf}\,\beta - 2\beta\mathrm{e}^{-\beta^2}/\sqrt{\pi}\right)/(4\beta^3)$ $\approx \begin{cases} 1 & \text{if } \sqrt{b} \ll 2R \\ 6\sqrt{\pi} \cdot (b/R^2)^{-3/2} & \text{if } \sqrt{b} \gg 2R \end{cases}$ *(plot: $w_{ED}(b,R)$ vs $\sqrt{b}/R$, marker at 2.2)*	$w_{ESP}(b,R) = 1 - (\sqrt{\pi}\,\mathrm{erf}\,\beta)/(2\beta)$ $\approx \begin{cases} 1/12 \cdot b/R^2 & \text{if } \sqrt{b} \ll 2R \\ 1 & \text{if } \sqrt{b} \gg 2R \end{cases}$ *(plot: $w_{ESP}(b,R)$ vs $\sqrt{b}/R$, marker at 3.5)*

Figure 7

Integrated and peak map values (without Z and b_j independent prefactors)

$$\sum_{j=0}^{5} a_j w_j$$

			ED	ESP	ED/ESP
integrated value	integration radius $\hat{R}$ (Å)	$\hat{R}$ (Å) 0.2 0.5 1.0 2.0 4.0			
peak value	smooth resolution cutoff B-factor B (Å^2)	B (Å^2) 3 5 10 30 100			
	sharp resolution cutoff resolution R (Å)	R (Å) 1.0 1.5 2.0 3.0 4.0			

Figure 8

Figure 9

Charge effects on ED and ESP map appearance at low and high resolution

A — Theoretical maps (GAMESS, Multiwfn)

temperature factor 45 Å^2 / temperature factor 10 Å^2

B — Experimental maps

2.7 Å resolution:
$$\frac{Asp(OD1)}{Asn(OD1)} = 0.73$$
$$\frac{Glu(OE1)}{Gln(OE1)} = 0.68$$
PDB-id 7ek7

1.15 Å resolution:
$$\frac{Asp(OD1)}{Asn(OD1)} = 0.78$$
$$\frac{Glu(OE1)}{Gln(OE1)} = 1.46$$
PDB-id 2v2p

ESP 2.7 Å:
$$\frac{Asp(OD1)}{Asn(OD1)} = \mathbf{0.06}$$
$$\frac{Glu(OE1)}{Gln(OE1)} = \mathbf{0.02}$$
PDB-id 6sht

ESP 1.15 Å:
$$\frac{Asp(OD1)}{Asn(OD1)} = 0.78$$
$$\frac{Glu(OE1)}{Gln(OE1)} = 0.39$$
PDB-id 7a6a

Table 1	Abbreviations
ED (ρ)	Electron Density
ESP (Φ)	ElectroStatic Potential
NCLS	Nuclear Coherent Scattering Length
EM	Electron Microscopy
TF model	Thomas-Fermi model (electron fluid approximation)
HF model	Hartree-Fock model (discrete quantum orbitals)
CM model	Cromer-Mann model (Gaussian "orbitals")
EO point of view	"Electrons Only" point of view (implicit nuclear charge)
All space integral	Integral extended over all space
Z	Atomic order number (number of electrons or protons)

Table 2	Z-dependence	
	integration limit (large $\hat{R}, B, R$)	peak limit (small $\hat{R}, B, R$)
ED	Z	$Z^{3/2}$
ESP	$Z^{1/3}$	Z
NCLS	Z^0	Z^0

Table 3

	b_j dependent electron contributions	
	integration limit (large $\hat{R}, B, R$)	peak limit (small $\hat{R}, B, R$)
ED	small b_j = large b_j	small b_j > large b_j
ESP	small b_j < large b_j	(small b_j = large b_j)

Table 4	^{1}H	^{2}H	C	N	O	Na	Mg	P	S	Cl	K	Ca	Zn
l(fm)	-3.7	6.7	6.6	9.4	5.8	3.6	5.4	5.1	2.8	9.6	3.7	4.7	5.7

Supplement to "X-rays, electrons, and neutrons as probes of atomic matter"

Matthias Bochtler

Derivation of the power law relating potential and density

The Thomas-Fermi (TF) (Thomas 1927) model treats the electron cloud as an inhomogeneous electron gas (Hohenberg and Kohn 1964). In a volume dV there are $2(4\pi/3)p_F^3 dV/(2\pi)^3$ quantum states with a momentum smaller than the Fermi momentum p_F. In this formula, the pre-factor 2 represents the two spin states of an electron, and $(4\pi/3)p_F^3$ is the volume of a sphere with momentum p_F. The factor $(2\pi)^3$ in the denominator is the phase space volume occupied by a quantum state in atomic units. Equating the number of states with momentum up to p_F to the number of electrons in the volume element ρdV gives an expression for the largest (Fermi) momentum for a given density ρ, $p_F = (3\pi^2\rho)^{1/3}$. The equation can also be understood in terms of Heisenberg's uncertainty relation (Heisenberg 1927). The higher the electron density, the more confined a single electron is in space, and therefore the higher its momentum (the exponent $1/3$ arises because the cube root of the density is an inverse "confinement" length). In atomic units, the kinetic energy corresponding to momentum p_F is $E_{kin} = p_F^2/2 = (3\pi^2\rho)^{2/3}/2$. More sophisticated analysis shows that this expression for the dependence of the kinetic energy E_{kin} on ρ is the leading term of more complicated expressions that also include the gradient of ρ (Weizsäcker 1935), or even non-local functionals (Becke 2014). For the sake of simplicity, I will continue here with only the leading term. For a stationary state, no energy should be gained by shifting the position of an electron. Therefore, the sum of the kinetic and potential energies should be constant. From the condition that the potential and kinetic energies should vanish far away from the molecule, it can be further deduced that the constant should be 0. Taking into account that the electron has a charge of -1 in atomic units, it follows that $E_{kin}(\vec{r}) = \Phi_{TF}(\vec{r})$. Substitution then leads to equation (1) of the main text. This approximate equation is very general. It does not depend on the assumptions of spherical atoms, but it does assume charge neutrality. In practice, charges in biological macromolecules are neutralised by atoms from the solution. The inclusion of these ions makes the equation applicable to charged macromolecules. Alternatively, variants of equation (1) of the main text (with an offset in the potential) are available that are applicable in the presence of

charge (Landau and Lifshitz 1977).

All space ESP integral from the BM rule

The calculation of the all (real) space ESP integral can be done in reciprocal space. The calculation starts from the observation that the integral of the ESP over all real space is related to the Fourier transform of the the ESP at the origin, which is turn is given by the BM equation. For the ESP, the integrated map value is given by:

$$\int \Phi(\vec{r})d^3r = \tilde{\Phi}(0) = 4\pi \cdot \lim_{q\to 0} \frac{Z - \tilde{\rho}(q)}{q^2} \tag{1}$$

In order to calculate this limit, it is necessary to know $\tilde{\rho}(q)$ not only at the origin, but also for small $\vec{q}$. In this limit, the $e^{i\vec{q}\vec{r}}$ in the Fourier transform formula for $\tilde{\rho}(q)$ can be Taylor expanded to $1 + i\vec{q}\vec{r} - (\vec{q}\vec{r})^2/2 + ... = 1 + iqr\cos\psi - q^2r^2\cos^2\psi/2 + ...$, where ψ is the angle between $\vec{q}$ and $\vec{r}$. Since the Fourier transform of a spherically symmetric real function centered on the origin is real, the first order term in the Taylor expansion (with the pre-factor i) cannot contribute to the Fourier transform. Considering that the differential volume element is $dV = 2\pi r^2 dr \sin\psi d\psi$, one has for neutral atoms and small q:

$$\tilde{\rho}(q) = \int_0^\infty dr\, 2\pi r^2 \rho(r) \int_0^\pi d\psi\, \sin\psi \left(1 - \frac{q^2r^2}{2}\cos^2\psi\right) = Z\left(1 + q^2\frac{\langle r^2\rangle}{6}\right) \tag{2}$$

One therefore has:

$$\int \Phi(\vec{r})d^3r = \lim_{q\to 0}\tilde{\Phi}(q) = 2\pi Z\frac{\langle r^2\rangle}{3} \tag{3}$$

For neutral, spherical atoms, the result is exact because q does not appear in the equation, and the small q approximation was only necessary to take the $q \to 0$ bound. The result could also have been derived by applying L'Hopital's rule twice to equation (1). It confirms equation (22) of the main text, which was derived based on real space considerations.

Integrated and peak intensities

In the following, derivations for equations (25-30) of the main text and the expressions for w_{ED} and w_{ESP} in main text Fig. 6 will be derived. Note that scaling is always chosen so that the w_{ED} and w_{ESP} take on values between 0 and 1. If $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$ denotes the well-known Γ function, one has:

$$\int_0^\infty x^n e^{-x^2}dx = \frac{1}{2}\int_0^\infty y^{\frac{n+1}{2}-1}e^{-y}dx = \frac{1}{2}\Gamma(\frac{n+1}{2}) \tag{4}$$

2

With the definition of the error function by $\mathrm{erf}(X) = \frac{2}{\sqrt{\pi}} \int_0^X \mathrm{e}^{-x^2} dx$, one can show:

$$\int_0^X x^2 \mathrm{e}^{-x^2} dx = \frac{\sqrt{\pi}}{4} \left(\mathrm{erf}(X) - \frac{2}{\sqrt{\pi}} X \mathrm{e}^{-X^2} \right) \tag{5}$$

This formula can be readily proven by comparing the values at $X = 0$ and the demonstration that the derivatives with respect to X of the left and right sides of the equation are equal.

ED

Integrated map value

Using main text equation (7) for $\rho(r)$, one has:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \rho(r) =$$

$$4\pi \sum_{j=1}^5 a_j \left(\frac{4\pi}{b_j} \right)^{3/2} \int_0^{\hat{R}} dr \cdot r^2 \mathrm{e}^{-4\pi^2 r^2 / b_j} = \tag{6}$$

$$\frac{4}{\sqrt{\pi}} \sum_{j=1}^5 a_j \int_0^{2\pi \hat{R}/\sqrt{b_j}} d\nu \cdot \nu^2 \mathrm{e}^{-\nu^2}$$

With reference to equation (2), and the abbreviation $\gamma_j = 2\pi \hat{R}/\sqrt{b_j}$, one therefore has:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \rho(r) = \sum_{j=1}^5 a_j w_{ED}(b_j, \hat{R})$$

$$w_{ED}(b, \hat{R}) = \mathrm{erf}(\gamma) - \frac{2}{\sqrt{\pi}} \gamma \mathrm{e}^{-\gamma^2} \tag{7}$$

$$\gamma = 2\pi \hat{R}/\sqrt{b}$$

Peak map value, B-factor broadening

Using main text equation (6) for $\tilde{\rho}(r)$, one has:

$$\rho_B(0) = \frac{1}{(2\pi)^3} \int_0^\infty dq \cdot 4\pi q^2 \tilde{\rho}(q) \mathrm{e}^{-Bs^2} =$$

$$32\pi \sum_{j=1}^5 a_j \int_0^\infty ds \cdot s^2 \mathrm{e}^{-(B+b_j)s^2} \tag{8}$$

Using $\int_0^\infty t^2 \mathrm{e}^{-t^2} dt = \frac{1}{2}\Gamma(\frac{3}{2}) = \sqrt{\pi}/4$, it follows that:

$$\rho_B(0) = (4\pi)^{3/2} \sum_{j=1}^5 \frac{a_j}{(B + b_j)^{3/2}} \tag{9}$$

3

Hence:

$$\rho_B(0) = \left(\frac{4\pi}{B}\right)^{3/2} \sum_{j=1}^{5} a_j w_{ED}(b_j, B)$$

$$w_{ED}(b, B) = (1 + b/B)^{-3/2}$$

(10)

Peak map value, blurring by limited resolution

Using main text equation (6) for $\tilde{\rho}(r)$, one has with $q = 4\pi s$:

$$\rho_R(0) = \frac{1}{(2\pi)^3} \int_0^{2\pi/R} dq \cdot 4\pi q^2 \tilde{\rho}(q) =$$

$$32\pi \sum_{j=1}^{5} a_j \int_0^{1/(2R)} ds \cdot s^2 e^{-b_j s^2} =$$

$$32\pi \sum_{j=1}^{5} \frac{a_j}{b_j^{3/2}} \int_0^{\sqrt{b_j}/(2R)} dx \cdot x^2 e^{-x^2} =$$

$$\frac{4\pi}{R^3} \sum_{j=1}^{5} \frac{a_j}{\left(\frac{\sqrt{b_j}}{2R}\right)^3} \left(\int_0^{\sqrt{b_j}/(2R)} dx \cdot x^2 e^{-x^2}\right)$$

(11)

Using equation (2) and the abbreviation $\beta_j = \frac{\sqrt{b_j}}{2R}$, one therefore has:

$$\rho_R(0) = \frac{4\pi}{R^3} \sum_{j=1}^{5} a_j \cdot \frac{\sqrt{\pi}}{4} \cdot \frac{\operatorname{erf}(\beta_j) - \frac{2}{\sqrt{\pi}}\beta_j e^{-\beta_j^2}}{\beta_j^3}$$

(12)

For large β_j, the last fraction in equation (7) approaches 0. To derive the small β_j limit, it is necessary to Taylor expand the error function around 0. As can be readily verified from the definition, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\left(x - \frac{x^3}{3} + ...\right)$. The Taylor expansion of the exponential is $e^{-x^2} = 1 - x^2 + ...$ For small β_j, the last fraction in equation (7) therefore converges towards $\frac{4}{3\sqrt{\pi}}$. It is therefore convenient to write the expression for $\rho_R(0)$ as:

$$\rho_R(0) = \frac{4\pi}{3} \frac{1}{R^3} \sum_{j=1}^{5} a_j \cdot \frac{3\sqrt{\pi}}{4} \cdot \frac{\operatorname{erf}(\beta_j) - \frac{2}{\sqrt{\pi}}\beta_j e^{-\beta_j^2}}{\beta_j^3}$$

(13)

4

Hence:

$$\rho_R(0) = \frac{4\pi}{3}\frac{1}{R^3}\sum_{j=1}^{5} a_j \cdot w_{ED}(b_j, R)$$

$$w_{ED}(b, R) = \frac{3\sqrt{\pi}}{4} \cdot \frac{\mathrm{erf}(\beta) - \frac{2}{\sqrt{\pi}}\beta e^{-\beta^2}}{\beta^3} \tag{14}$$

$$\beta = \frac{\sqrt{b}}{2R}$$

ESP

Integrated map value

Taking into account that $\sum_{j=1}^{5} a_j = Z$, the integrated nuclear potential is:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \Phi_{nuc}(r) = \int_0^{\hat{R}} dr \cdot 4\pi r^2 \sum_{j=1}^{5} a_j/r^2 = 2\pi\hat{R}^2 \sum_{j=1}^{5} a_j \tag{15}$$

The gradient of the electronic potential at a point r depends upon the charge inside a sphere of radius a (see Fig. 2 of the main text). This charge is given by equation (4). With the gauge that the potential vanishes at infinity, one has:

$$\Phi_{el}(r) = -\sum_{j=0}^{5} a_j \int_r^{\infty} \frac{\mathrm{erf}(\nu_j) - 2\nu_j e^{-\nu_j^2}/\sqrt{\pi}}{\tilde{r}^2} d\tilde{r} \quad \text{where } \nu_j = 2\pi\tilde{r}/\sqrt{b_j} \tag{16}$$

Variable substitution yields:

$$\Phi_{el}(r) = -\sum_{j=0}^{5} a_j \frac{2\pi}{\sqrt{b_j}} \int_{\mu}^{\infty} \frac{\mathrm{erf}(\nu_j) - 2\nu_j e^{-\nu_j^2}/\sqrt{\pi}}{\nu_j^2} d\nu_j \quad \text{where } \mu = 2\pi r/\sqrt{b_j} \tag{17}$$

The integrated electronic potential is therefore:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \Phi_{el}(r) = -\int_0^{\hat{R}} dr \cdot 4\pi r^2 \sum_{j=1}^{5} a_j \frac{2\pi}{\sqrt{b_j}} \int_{\mu_j}^{\infty} \frac{\mathrm{erf}(\nu) - 2\nu e^{-\nu^2}/\sqrt{\pi}}{\nu^2} d\nu$$

$$\text{where } \mu_j = 2\pi r/\sqrt{b_j} \tag{18}$$

Another round of variable substitution yields:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \Phi_{el}(r) = -\sum_{j=1}^{5} a_j \frac{b_j}{\pi} \int_0^{\gamma_j} \mu^2 \left(\int_{\mu}^{\infty} \frac{\mathrm{erf}(\nu) - 2\nu e^{-\nu^2}/\sqrt{\pi}}{\nu^2} d\nu \right) d\mu =$$

$$-2\pi\hat{R}^2 \cdot 2 \sum_{j=1}^{5} a_j \frac{b_j}{(2\pi\hat{R})^2} \int_0^{\gamma_j} \mu^2 \left(\int_{\mu}^{\infty} \frac{\mathrm{erf}(\nu) - 2\nu e^{-\nu^2}/\sqrt{\pi}}{\nu^2} d\nu \right) d\mu$$

$$\text{where } \gamma_j = 2\pi\hat{R}/\sqrt{b_j} \tag{19}$$

5

Taking into account that $b_j/(2\pi\hat{R})^2 = 1/\gamma_i^2$, and that $\Phi = \Phi_{nuc} + \Phi_{el}$, it follows from equations (12) and (16) that:

$$\int_0^{\hat{R}} dr \cdot 4\pi r^2 \Phi(r) = 2\pi\hat{R}^2 \sum_{j=1}^5 a_j w_{ESP}(b_j, \hat{R})$$

$$w_{ESP}(b, \hat{R}) = 1 - 2\frac{\int_0^\gamma \mu^2 \left(\int_\mu^\infty \frac{\mathrm{erf}(\nu) - 2\nu e^{-\nu^2}/\sqrt{\pi}}{\nu^2} d\nu \right) d\mu}{\gamma^2} \qquad (20)$$

$$\gamma = \frac{2\pi\hat{R}}{\sqrt{b}}$$

To my knowledge, this double integral can only be evaluated numerically. In the limit $\hat{R} \to 0$, w_{ESP} approaches 1, as should be, since only the nuclear charge matters in this limit. To derive the opposite limit of large $\hat{R}$, i.e $2\pi\hat{R} \gg \sqrt{b_i}$, note that:

$$\int_0^\infty dr \cdot 4\pi r^2 \Phi(r) = \sum_{j=0}^5 \frac{1}{4\pi} a_j b_j \qquad (21)$$

according to equation (24) of the main text. This can be rewritten as:

$$\int_0^\infty dr \cdot 4\pi r^2 \Phi(r) = 2\pi\hat{R}^2 \sum_{j=0}^5 a_j \frac{b_j}{8\pi^2 \hat{R}^2} \qquad (22)$$

Therefore, one can conclude that:

$$w_{ESP}(b, \hat{R}) \approx \frac{b}{8\pi^2 \hat{R}^2} \text{ for } 2\pi\hat{R} \gg \sqrt{b} \qquad (23)$$

Peak map value, B-factor broadening

$$\Phi_B(0) = \frac{1}{(2\pi)^3} \int_0^\infty dq \cdot 4\pi q^2 \tilde{\Phi}(q) e^{-Bs^2} \qquad (24)$$

Using the Bethe-Mott relationship, $\sum_{j=1}^5 a_j = Z$, and $q = 4\pi s$, it follows that:

$$\Phi_B(0) = \frac{2}{\pi} \int_0^\infty dq (Z - \tilde{\rho}(q)) e^{-Bs^2} =$$

$$8 \sum_{j=1}^5 a_j \int_0^\infty (1 - e^{-b_j s^2}) e^{-Bs^2} ds = \qquad (25)$$

$$8 \sum_{j=1}^5 a_j \int_0^\infty \left(e^{-Bs^2} - e^{-(b_j+B)s^2} \right) ds$$

6

Using $\int_0^\infty e^{-t^2} dt = \frac{1}{2}\Gamma(\frac{1}{2}) = \sqrt{\pi}/2$, it follows that:

$$\Phi_B(0) = 4\sqrt{\pi} \sum_{j=1}^{5} a_j \left(\frac{1}{\sqrt{B}} - \frac{1}{\sqrt{b_j + B}} \right) \tag{26}$$

Hence:

$$\Phi_B(0) = \frac{4\sqrt{\pi}}{\sqrt{B}} \sum_{j=1}^{5} a_j w_{ESP}(b_j, B)$$

$$w_{ESP}(b, B) = 1 - (1 + b/B)^{-1/2} \tag{27}$$

Peak map value, blurring by limited resolution

$$\Phi_R(0) = \frac{1}{(2\pi)^3} \int_0^{2\pi/R} dq \cdot 4\pi q^2 \tilde{\Phi}(q) \tag{28}$$

Using the Bethe-Mott relationship, $\sum_{j=1}^{5} a_j = Z$, and $q = 4\pi s$, it follows that:

$$\Phi_R(0) = \frac{2}{\pi} \int_0^{2\pi/R} dq \cdot (Z - \tilde{\rho}(q)) = \frac{2}{\pi} \sum_{j=1}^{5} a_j \int_0^{2\pi/R} dq \cdot (1 - e^{-b_j s^2}) =$$

$$8 \sum_{j=1}^{5} a_j \int_0^{1/(2R)} ds \cdot (1 - e^{-b_j s^2}) = \sum_{j=1}^{5} a_j \cdot \left(\frac{4}{R} - \frac{8}{\sqrt{b_j}} \int_0^{\sqrt{b_j}/(2R)} dx \cdot e^{-x^2} \right) =$$

$$\sum_{j=1}^{5} a_j \cdot \left(\frac{4}{R} - \frac{4\sqrt{\pi}}{\sqrt{b_j}} \text{erf}\left(\frac{\sqrt{b_j}}{2R} \right) \right) = \frac{4}{R} \sum_{j=1}^{5} a_j \cdot \left(1 - \frac{\sqrt{\pi}}{2} \frac{\text{erf}(\frac{\sqrt{b_j}}{2R})}{\frac{\sqrt{b_j}}{2R}} \right) \tag{29}$$

Hence:

$$\Phi_R(0) = \frac{4}{R} \sum_{i=j}^{5} a_i w_{ESP}(b_j, R)$$

$$w_{ESP}(b, R) = 1 - \sqrt{\pi}\text{erf}(\beta)/(2\beta) \tag{30}$$

$$\beta = \frac{\sqrt{b}}{2R}$$

The author declares no competing interests.