Article

Toward Real Real-Space Refinement of Atomic Models

Alexandre G. Urzhumtsev [1,2] **and Vladimir Y. Lunin** [3]

1   Centre for Integrative Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS–INSERM-UdS, 1 rue Laurent Fries, BP 10142, 67404 Illkirch, France
2   Faculté des Sciences et Technologies, Université de Lorraine, BP 239, 54506 Vandoeuvre-lès-Nancy, France
3   Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, 142290 Pushchino, Russia
*   Correspondence: sacha@igbmc.fr

Abstract: High-quality atomic models providing structural information are the results of their refinement versus diffraction data (reciprocal-space refinement), or versus experimental or experimentally based maps (real-space refinement). A proper real-space refinement can be achieved by comparing such a map with a map calculated from the atomic model. Similar to density distributions, the maps of a limited and even inhomogeneous resolution can also be calculated as sums of terms, known as atomic images, which are three-dimensional peaky functions surrounded by Fourier ripples. These atomic images and, consequently, the maps for the respective models, can be expressed analytically as functions of coordinates, atomic displacement parameters, and the local resolution. This work discusses the practical feasibility of such calculation for the real-space refinement of macromolecular atomic models.

Keywords: real-space refinement; refinement programs; atomic images; map calculation; shell decomposition; inhomogeneous resolution; CPU time

1. Introduction

Even though structural biology deals with biological objects of different complexities, sizes, and levels, very impressive results and information have been obtained from macromolecular studies at the atomic level. Two principal methods for such studies, X-ray or neutron crystallography (MX) and cryo-electron microscopy (cryo-EM), describe macromolecular models in terms of positions $\mathbf{r}_n$, $n = 1, 2, \ldots, N$, of the atomic centers and of the uncertainties of these positions. Below, we discuss only an isotropic uncertainty characterized for each atom by its own atomic displacement parameter (ADP) B_n.

The experimental information of these methods is available in different terms. In cryo-EM, the experiment gives the maps of the electrostatic scattering potential $\rho_{obs}(\mathbf{r})$ as a result of 3D reconstruction from 2D experimentally observed projections. These maps have a limited resolution, which usually varies from one region to another [1]. In MX, the experiment results in a set of Fourier coefficients $\mathbf{F}_{obs}(\mathbf{s})$ of the electron density distribution, or rather the magnitudes $F_{obs}(\mathbf{s}) = |\mathbf{F}_{obs}(\mathbf{s})|$ of these complex values, which are also known as structure factors. After all, these data are converted into maps $\rho_{obs}(\mathbf{r})$ of a limited resolution. This procedure consists of several steps, but it is applied only once for a certain period of work and this experimentally based map is used then for validation and the improvement of atomic models. In what follows, we refer to both experimental distributions $\rho_{obs}(\mathbf{r})$ as a 'density distribution'. In both cases, this function is considered in a crystal: a real one in MX and a virtual one, containing an isolated macromolecular object per unit cell, in single-particle cryo-EM.

Atomic models are refined by their best fit to the experimental data; for a recent review, see [2]. For such model-to-data comparisons, model information is expressed in the same terms as the data, $\mathbf{F}_{calc}(\mathbf{s}; \{\mathbf{r}_n, B_n\})$ or $\rho_{calc}(\mathbf{r}; \{\mathbf{r}_n, B_n\})$, and some score function

is calculated. Reducing this function value is expected to be an indicator of a model improvement. Depending on the type of information and respective score functions, structural biologists talk about reciprocal-space refinement

$$\min_{\{r_n, B_n\}} f_{reciprocal}(F_{calc}(s; \{r_n, B_n\}); F_{obs}(s)) \tag{1}$$

and real-space refinement

$$\min_{\{r_n, B_n\}} f_{real}(\rho_{calc}(r; \{r_n, B_n\}); \rho_{obs}(r)). \tag{2}$$

Each of these two types of refinement has its features (e.g., [2,3]) based on the properties of structure factors and density distributions. Each atom, and more generally each piece of the density distribution, contributes to all structure factors. By this reasoning, an appropriate reciprocal-space refinement requires the coordinates and diffraction parameters of all atoms in the unit cell [4], as well as a contribution from all disordered regions, in particular that from the bulk solvent [5]. Differing from this, each atom significantly contributes to the density only in a relatively small region of the space around the atomic center. As a consequence, to calculate an accurate density distribution in the vicinity of an atom, one requires the parameters of only a few atoms: those close to the atom under consideration. Except for the atoms at the molecule surface, especially in the regions partially occupied [6], this density is not influenced by the disordered solvent either [7]. This suggests real-space refinement as the method of choice [8], especially in earlier studies of work when only a partial atomic model is available. Then, the refinement of the inner part of such a model can be performed, ignoring the missed parts of the model and a contribution from the disordered regions, and this step can be completed later with reciprocal-space refinement [3].

The minimization of the chosen score function is usually performed iteratively and is ruled by its gradient with respect to the atomic parameters $\{r_n, B_n\}$. While $\rho_{obs}(r)$ is obtained only once for a whole refinement procedure in cryo-EM, or is updated from time to time in MX, the maps $\rho_{calc}(r; \{r_n, B_n\})$ are calculated for each model tried during real-space fit.

To obtain model structure factors $F_{calc}(s; \{r_n, B_n\})$ or model density maps $\rho_{calc}(r; \{r_n, B_n\})$ from the atomic parameters, refinement programs use transition modules from one parameter level to the next one. Refinement can require several consecutive transition steps, as Figure 1 shows [9,10]. These calculations should provide sufficiently accurate structure factors or a map and be fast enough to make such calculations useful in practice. An efficient algorithm to calculate the gradient is automatically defined by inverting the scheme of the function calculation [9].

The scheme relating different model levels (Figure 1) is hierarchic and direct. This means that one can routinely pass from a previous level of model data to the next one and not necessarily the opposite way. In particular, one can generate a density map on any fine grid from an atomic model and then calculate a set of Fourier coefficients (structure factors) of this grid function. On the contrary, given a set of structure factors, one cannot recover the exact density distribution at any fine grid but only an approximation to it, a map of a limited resolution, calculated as a Fourier series with this set. While the architecture of the reciprocal-space atomic refinement programs is quite established, this is not yet the case for real-space refinement programs. In this work, we discuss the overall scheme and practical steps for such procedures.
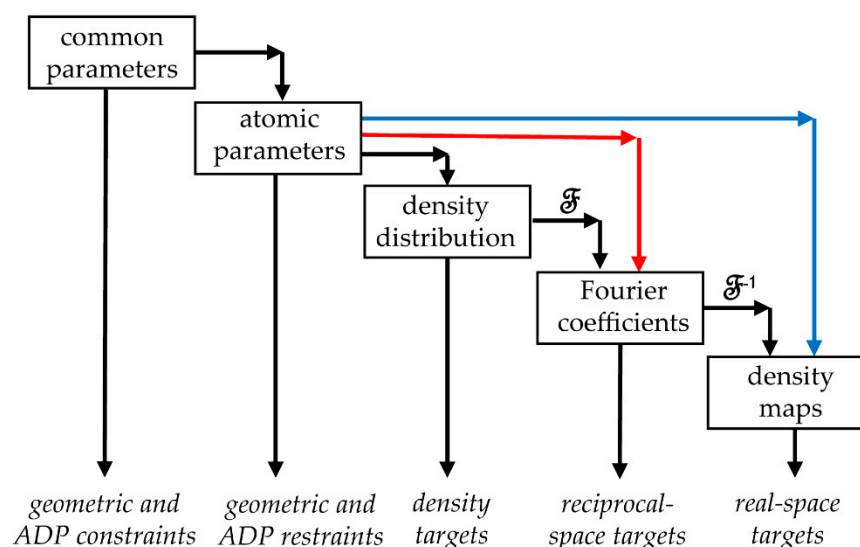
Figure 1. Levels of macromolecular parameterization in MX and cryo-EM. By 'density distribution', we consider various kinds of scalar functions in space, such as an electron or nuclear scattering density distribution in crystallography or scattering electrostatic potential in cryo-EM, etc. The term 'density map' stands for maps of any of these distributions. Atomic parameters are usually the coordinates of the centers of atoms and their displacement parameters, ADP. Common parameters may be dihedral angles [11–13], rigid-body parameters [14], common ADP values for all atoms of the residue [15] or TLS parameters [16,17], or something else, describing common features of an atomic group. Black arrows show the step-by-step hierarchic recalculation of the model parameters; the red and blue arrows illustrate alternative direct calculations of structure factors and maps from model parameters.

2. Results

2.1. Gaussian Atomic Model

In MX and cryo-EM, the atomic scattering factor is a Fourier transform of a density distribution of an immobile isolated atom placed in the origin, and it is usually approximated by a weighted sum of a few Gaussian functions, $K_{Gauss} \sim 1$–5 [18–23]. Coefficients of this sum depend on the diffraction method, given the chemical type of atoms, and eventually on the atomic environment [24]. In what follows, to simplify the illustrations and unless the opposite is written, we consider a 'movable' virtual single-Gaussian atom n of a unit 'charge' for which its scattering function (structure factor with index **s** when the atom is placed in the crystal origin, $\mathbf{r_n} = \mathbf{0}$) is equal to

$$\mathbf{F_n}(\mathbf{s}; B_n) = \exp\left[-\frac{(b_A + B_n)|\mathbf{s}|^2}{4}\right]. \tag{3}$$

Here, b_A is the parameter of this immobile Gaussian atom, representing the rate of decrease in the atomic scattering factor with resolution (or, respectively, the width of the peak of the atomic density, as shown below), and B_n is its isotropic atomic displacement parameter describing the variation in the position of this atom in time during the experiment or in space over equivalent copies. The density $\rho_n^0(\mathbf{r}; B_n)$ corresponding to this atom is also Gaussian

$$\rho_n^0(\mathbf{r}; B_n) = g(\mathbf{r}; b_A; B_n) = \left(\frac{4\pi}{b_A + B_n}\right)^{3/2} \exp\left(-\frac{4\pi^2|\mathbf{r}|^2}{b_A + B_n}\right). \tag{4}$$

Functions $\mathbf{F_n}(\mathbf{s}; B_n)$ and $\rho_n^0(\mathbf{r}; B_n)$ are spherically symmetric functions decreasing with the distance $s = |\mathbf{s}|$ and $r = |\mathbf{r}|$ to the origin. The values of b_A and B_n, typical in structural

biology, are of order of $10^1 - 10^2$ Å^2 [25,26]. For $b_A + B_n \approx 40$ Å^2, the value of $\rho_n^0(\boldsymbol{r}; B_n)$ at $|\boldsymbol{r}| = 2.5$ Å decreases to about 0.002 times the function value in the origin. By this reasoning, when generating a density distribution as a sum of atomic densities, the atomic contributions are cut out beyond $r = |\boldsymbol{r}| > r_{dens}$ with $r_{dens} \sim 2.5 - 3.0$ Å.

2.2. Schemes of Reciprocal-Space Refinement

For an atomic model, its structure factors can be directly calculated from atomic coordinates and displacement parameters, isotropic or anisotropic. In this procedure (red arrows in Figure 1), each of N atoms of the model directly contributes to each of the structure factors making, for large macromolecules, the total number of computing operations too high. Modern macromolecular refinement programs obtain these values as Fourier coefficients of the respective density distribution calculated on a regular grid as a function of the model parameters [27,28]. In this two-step procedure (black arrows in Figure 1), the number of operations is independent of the number of structure factors in the first step and is independent of the number of atoms in the second step. The grid size is a common factor which influences the number of operations for both steps.

The two-step scheme is faster but introduces errors in the calculated values of model structure factors. First, the density is generated within a sphere centered in the atomic position and with the radius r_{FT}. For a virtual Gaussian atom (4), the error in the Fourier transform of the atomic density due to this distance cut-off, being expressed with rescaled parameters $x = r/\sqrt{B}$, $X = r_{FT}/\sqrt{B}$, $\mathbf{t} = \mathbf{s}\sqrt{B}$, and $t = s\sqrt{B}$, is

$$\Delta(X) = |\boldsymbol{F}_{exact}(\boldsymbol{t})| - \left|\boldsymbol{F}_{integral}(\boldsymbol{t}; X)\right| =$$
$$exp\left(-\frac{t^2}{4}\right) - \frac{8\sqrt{\pi}}{t} \int_0^X (2\pi x) exp\left(-4\pi^2 x^2\right) sin(2\pi x t) dx. \tag{5}$$

For a given B value, discrepancy (5) is a non-monotonous function of the distance cut-off X (Figure 2) which suggests that the optimally chosen r_{FT} value may be eventually different from r_{dens}. The error becomes small for $X \sim 0.4$, which, for typical B values, means $r_{FT} \sim 2.5 - 3.5$ Å. Increasing r_{FT} increases, as a cube, the number K_{grid} of the grid points to which each atom contributes and the CPU time.
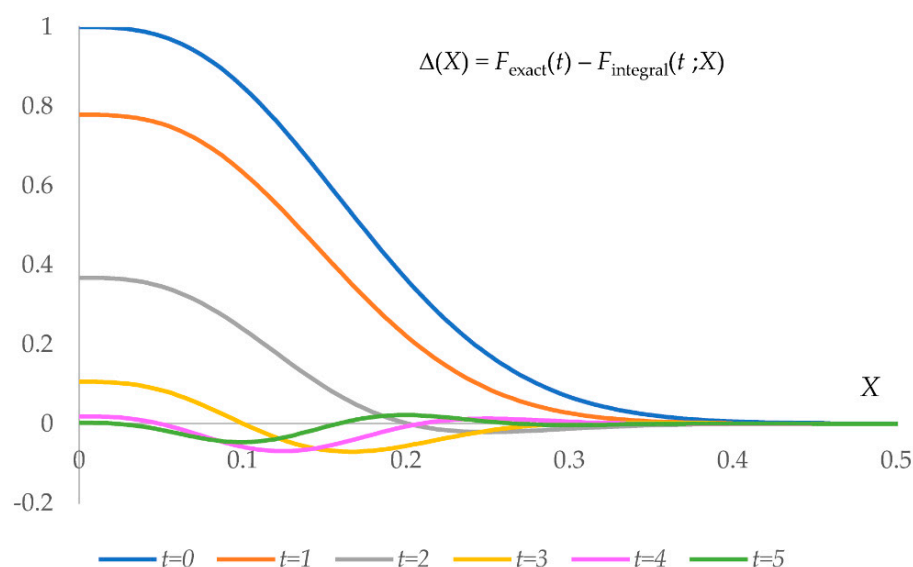

Figure 2. Error in the Fourier transform of a density of a Gaussian virtual atom. Error $\Delta(X)$ is given as function (5) of the dimensionless truncation radius, $X = r_{FT}/\sqrt{B}$, and for different values of the parameter $t = s\sqrt{B}$. $\Delta(0)$ is equal to the exact value $|\mathbf{F}(\boldsymbol{s}; B)|$ for the respective $s\sqrt{B}$.

Extra errors in structure factors occur due to the substitution of the integral Fourier transform by the discrete Fourier transform (DFT) using a finite regular grid. Increasing the grid step improves the accuracy but increases K_{grid}, again as a cubic function. When using FFT [29], a compromise between accuracy and the computation time has been discussed [28,30]. At conventional resolutions $D \sim 2 - 3$ Å, with the standard choice of $r_{FT} \sim 2.5 - 3.5$ Å and the grid step equal to $D/4$ or $D/3$, the two steps require CPU time of the same order of magnitude, $T_{density} \sim T_{FT}$. Exact values, e.g., those shown in Section 2.5, also depend on other parameters, for example, the relative unit cell volume per atom [31].

2.3. Schemes of Real-Space Refinement

Real-space refinement compares the model map of a density distribution with an experimental one [11]. For an appropriate comparison, the former map should reproduce the imperfections of the latter. The main sources of imperfections of the maps are their limited resolution and an uncertainty in atomic positions. In MX, maps may also be influenced by missed or downweighed reflections. Usually, at the stage of real-space refinement, eventual experimental errors in the map values are neglected. Similar to reciprocal-space refinement, different procedures can be envisaged to obtain a density map from an atomic model.

First, following the principal scheme (Figure 1), given an atomic model, one generates a respective model density and then applies two consecutive Fourier transforms. The grid for the density should be sufficiently fine to assure accurate structure factors. A similar number of calculations is required to obtain a gradient of a real-space score function with respect to the atomic parameters [9]. In total, using this procedure makes real-space refinement more time-consuming than the reciprocal-space one.

Instead, the model map can be calculated directly from an atomic model as a sum of atomic contributions (blue arrows in Figure 1)

$$\rho^d(\boldsymbol{r}) = \sum_{n=1}^{N^{atoms}} \rho_n^d(\boldsymbol{r} - \boldsymbol{r}_n; B_n, D) . \tag{6}$$

Here, $\rho_n^d(\boldsymbol{r} - \boldsymbol{r}_n; B_n, D)$ is no longer an atomic density but its image at a given resolution. To realize such a procedure, one needs to express these images as a function, ideally an analytic one, of the atomic coordinates, isotropic displacement parameter B_n, and the resolution D. While both increasing the B_n value and decreasing the resolution somewhat similarly blur the central peak of the atomic contribution, their effects are different at a distance to the atomic center. Atomic images $\rho_n^d(\boldsymbol{r}; B, D)$ are oscillating functions. Their central peak is surrounded by spherically symmetric waves of a decreasing amplitude, known as Fourier ripples.

To avoid the difficulty of modeling atomic images, some programs [32–35] deal only with the map values in the atomic centers, making the refinement of B_n values impossible. Some authors model only the central peak [36,37] or take the exact atomic density instead of its limited-resolution image [11]. To keep the ripples, the atomic images are either precalculated for some grid of B_n values [38] or parametrized using a step approximation to scattering functions [39,40].

2.4. Map as an Analytic Function

Fourier ripples are the result of the resolution truncation independent of how this truncation has occurred, explicitly or implicitly. The effect of ripples coming from neighboring atoms is prominent at low and medium resolution; moreover, at subatomic resolution, this effect can strongly bias density deformation maps [41]. The amplitude of these ripples decreases, as a function of the distance to the center, much slower than the atomic density itself. The number of atom contributions to a given point increases with the same rate, giving an important cumulative effect of the ripple truncation [42]. For this reason, to

calculate the maps accurately, atomic images should include at least a few Fourier ripples before being cut out at some truncation distance r_{map}.

To model oscillating images, Urzhumtsev and Lunin [43] suggested decomposing them into a weighted sum of spherically symmetric terms

$$\Omega(\boldsymbol{x}; \mu, \nu) = \frac{1}{|\boldsymbol{x}|\mu} \sqrt{\frac{1}{4\pi\nu}} \left[\exp\left(-\frac{4\pi^2(|\boldsymbol{x}| - \mu)^2}{\nu} \right) - \exp\left(-\frac{4\pi^2(|\boldsymbol{x}| + \mu)^2}{\nu} \right) \right]. \quad (7)$$

Each such term represents a uniform distribution on the spherical surface of the radius μ blurred with a Gaussian function with a parameter ν. Thanks to the features of function (7), an image of a normalized virtual Gaussian atom (4), placed in the origin, with any value of its atomic displacement parameter B_n and at any resolution D is

$$g^d(\boldsymbol{r}; b_A; B_n, D) = \frac{4\pi}{3} \sum_{m=1}^{M} \kappa^{(m)} \Omega\left(\boldsymbol{r}; \mu^{(m)} D, b_A + B_n + \nu^{(m)} D^2 \right). \quad (8)$$

Here, $\mu^{(m)}, \nu^{(m)},$ and $\kappa^{(m)}$ are coefficients of the decomposition of the three-dimensional interference function

$$3\frac{sin(2\pi|\boldsymbol{x}|) - (2\pi|\boldsymbol{x}|)cos(2\pi|\boldsymbol{x}|)}{(2\pi|\boldsymbol{x}|)^3} \approx \sum_{m=1}^{M} \kappa^{(m)} \Omega\left(\boldsymbol{x}; \mu^{(m)}, \nu^{(m)} \right). \quad (9)$$

into the sum over $\Omega(\boldsymbol{x}; \mu, \nu)$ terms (shell decomposition) [43].

The number M of terms in (9) is defined by r_{map}. With (8), the resolution in (6) may be individual for each atomic image, $D = D_n$. This value becomes a parameter of an atomic model, characterizing how confidently B_n and $\boldsymbol{r}_n$ values are found from the given map. When an atomic density $\rho_n^d(\boldsymbol{r}, B, D)$ is represented by a few Gaussians, its image is a respective weighted sum of (8), one per Gaussian.

We illustrated this latter option with Figure 3, which shows a simulated inhomogeneous-resolution map. This map is directly calculated as (6)–(8) for a protein model of IF2 [44] placed in a virtual unit cell in space group P1, similar to cryo-EM models. Here, the resolution was artificially assigned as 2 Å in the center of the molecule, increasing, as a function of distance, up to 5 Å at its periphery.

Actually, the shell decomposition into a sum of term (7) can be applied to any spherically symmetric oscillating function in space. In particular, an atomic image at a given resolution D for any B_n can be directly represented as

$$\rho_n^d(\boldsymbol{r}; B_n, D) = \frac{4\pi}{3} \sum_{m=1}^{M} C^{(m)} \Omega\left(\boldsymbol{r}; R^{(m)}, B_n + B^{(m)} \right). \quad (10)$$

where the coefficients $R^{(m)}, B^{(m)},$ and $C^{(m)}$ are calculated for an immobile atom and are universal for all atoms of the given chemical type. Representation (10) reduces the number of terms in comparison with (8) and (9), and thus accelerates calculations while the resolution becomes no more variable.
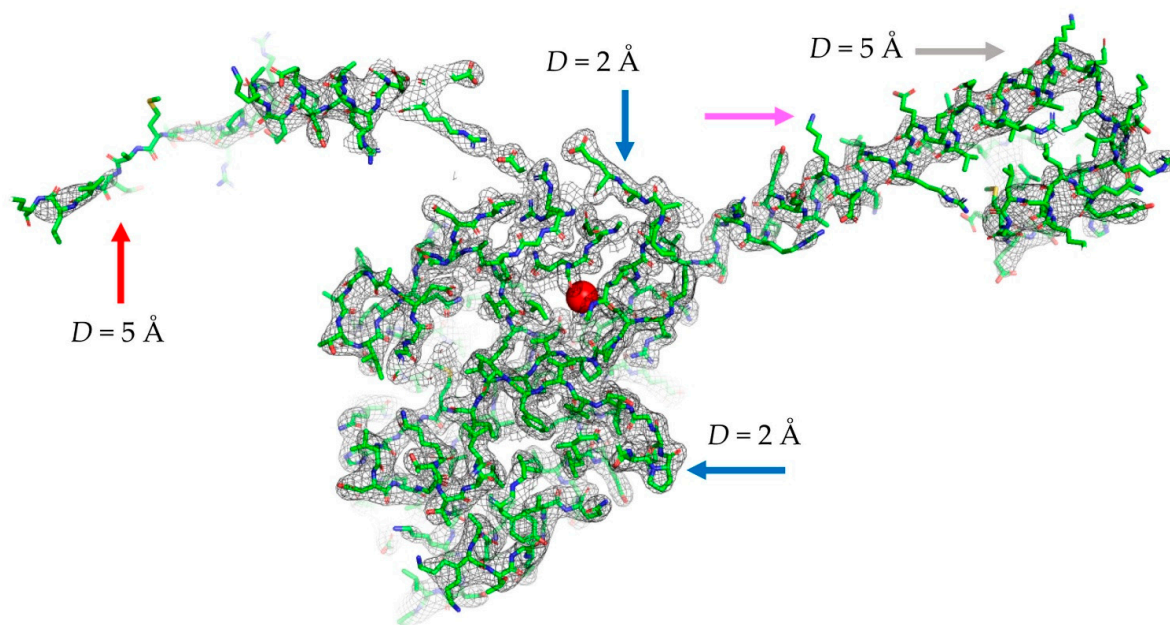
Figure 3. Map of an inhomogeneous resolution calculated in a single run. Map resolution varies from 2 Å around the molecular center (red sphere) to 5 Å at the periphery. Color arrows indicate the regions of a high resolution and small ADP (blue), high resolution and large ADP (magenta), low resolution and small ADP (grey), and low resolution and large ADP (red). Figure has been prepared using Pymol [45].

2.5. Comparison of Schemes of Real-Space Refinement

Similar to the structure factor calculation, now we have two ways to obtain an accurate model map: a step-by-step numeric and a direct analytic. The former consists of three steps. First, $T_{density}$ time is required to calculate the exact density distribution on a regular and sufficiently fine grid, with each atom contributing within a sphere of a given radius r_{FT}. Second, FFT is applied to this function requiring T_{FFT_SF} time. Finally, one more FFT is applied to the obtained Fourier coefficients to produce a map of the required resolution on a regular grid, which is usually coarser than the initial one, requiring $T_{FFT\_map} \leq T_{FFT\_SF}$. The map errors become unacceptably large when a too large step $h_{density}$ of the initial grid or a too short r_{FT} are taken. For conventional resolutions $D \sim 2 - 3$ Å, the standard values are $h_{density} \sim D/3 - D/4$ Å and $r_{FT} \sim 2.5 - 3.5$ Å with no need for artificial manipulations with displacement parameters and increased r_{FT}, which may be required for lower resolutions [28,30].

The final map is calculated with the same step as the experimental one, usually $h_{map} \sim D/2 - D/3$ Å. The alternative, direct map calculation consists of a single step requiring CPU time T_{direct}. This value depends on the grid step h_{map} of the map, the same as above, and on the truncation distance r_{map} for the atomic images. To obtain accurate maps, this distance has been recommended to be $kD/2$, with k equal to 4 or 5, or a higher integer [42]. The sum over $\Omega(\boldsymbol{x}; \mu, \nu)$ should include the terms significantly contributing up to this distance.

To compare the computational efficiency of the two ways to calculate the model maps, we made a numeric experiment with the IF2 model [44] placed in a virtual unit cell with the sides $80 \times 120 \times 100$ Å in space group P1 remining a cryo-EM case. A conventional five-Gaussian approximation to the atomic density was used [23]. We made calculations at the resolution of 2 Å, with varying grid steps and truncation radii. We used the original crystallographic FFT program [46] and our own fast-written rather non-optimized programs to obtain the model density distributions and to calculate directly the limited-resolution maps. The CPU time varies with the computer, compiler, and degree of the algorithm optimization. Additionally, for the same grid, $T_{density}$ and T_{direct} can vary if the model contains more or

fewer atoms. This means that when comparing $T_{density} + T_{FFT\_SF} + T_{FFT\_map}$ with T_{direct}, as obtained below, some margins should be considered.

Figure 4 shows CPU time, as a function of the grid step and truncation radius, for the components of the three-step map calculation. $T_{density}$ is near proportional to the number of Gaussians in the approximation. For the given example, using a single-Gaussian approximation, not used in practice, reduces the respective values by four (not shown).
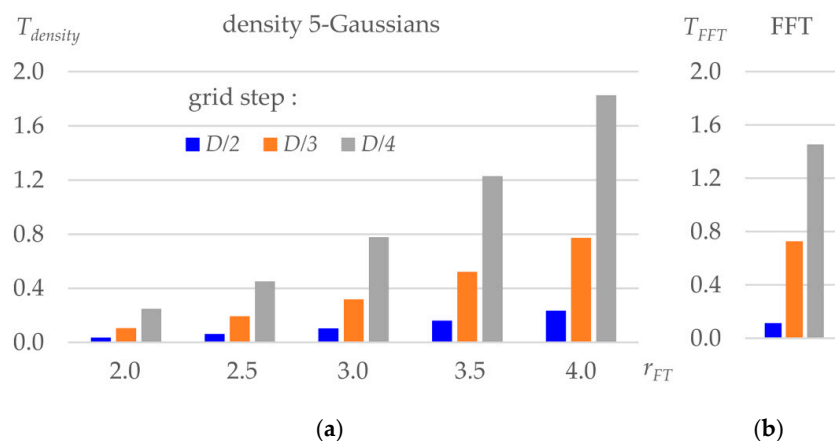

(a)   (b)

Figure 4. CPU time for the three-step map calculation for different grid steps expressed as a part of the resolution D. (**a**) CPU time, in seconds, to calculate a density distribution for the test protein model using different truncation distance r_{FT}. (**b**) CPU time to calculate FFT on a grid as defined in (**a**).

Figure 5 shows CPU time, as a function of the grid step and truncation radius r_{map}, for the direct map calculation. We calculated the map for both options, when the resolution is fixed and the simplified decomposition (10) is used, and for the variable-resolution option. The latter multiplies CPU time roughly by four, as it was for the density calculation with multi-Gaussians. One should note that when increasing r_{map} from 4 Å to 5 Å, we not only increase the number of grid points to which each atom contributes but also increase the number of terms in (8) and (10). Inversely, we reduce one term when shortening r_{map} to 3 Å, the distance of which is not recommended except at early refinement iterations [42].
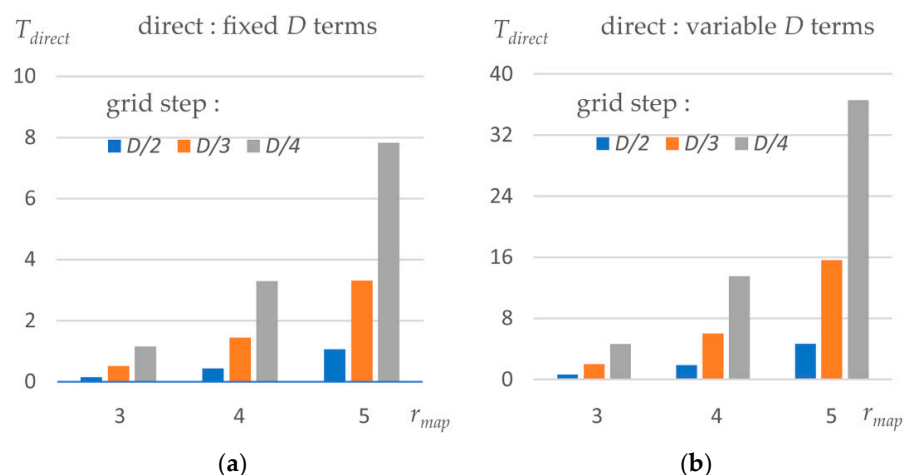

(a)   (b)

Figure 5. CPU time for the single-step map calculation, as a function of the truncation distance and the grid step, expressed as a part of the resolution D. (**a**) CPU time, in seconds, to calculate a density map for the test protein model using the simplified decomposition (10) of atomic images at a fixed resolution. (**b**) The same using the variable-resolution terms (8).

3. Discussion

Figure 6 compares CPU time for the different sets of parameter values eventually applicable in practice, i.e., giving sufficiently accurate maps while not requiring excessive time. The results are shown when the resulted map is calculated on the grid with the step $D/2$, $D/3$, or $D/4$; the last group is not expected to be used at the refinement step and is given for reference.
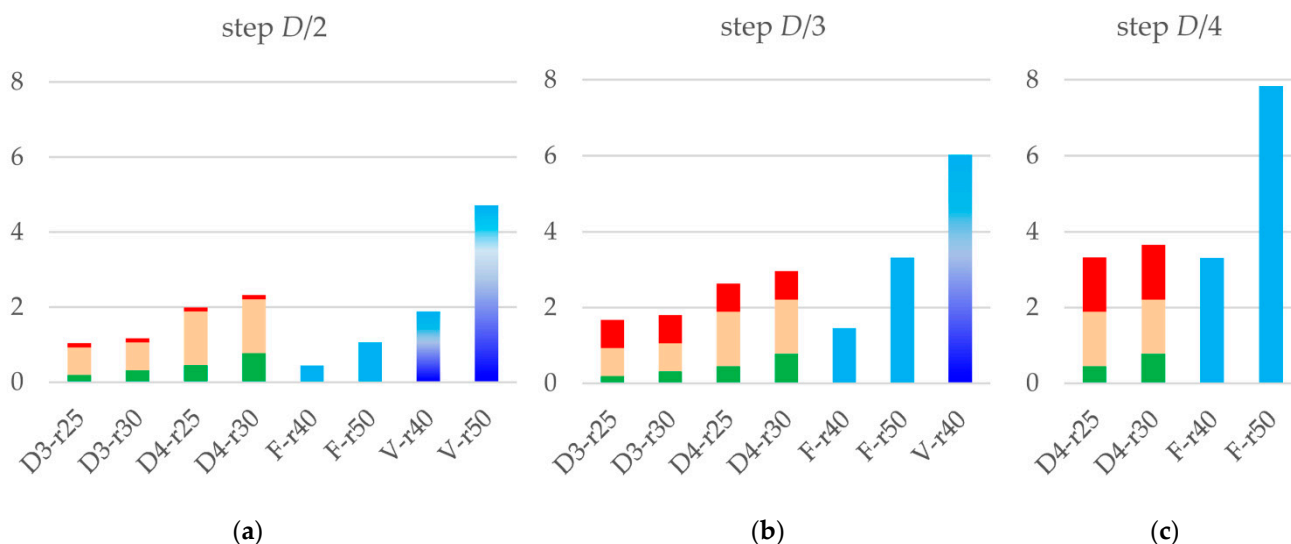

(**a**)   (**b**)   (**c**)

Figure 6. CPU time, in seconds, for different values of parameters used to calculate the model map. Multicolor columns represent the multi-step calculation with the green part for $T_{density}$, beige for T_{FFT_SF}, and red for T_{FFT_map}. Index 'D' indicates the grid step for the density, as a part of the resolution, $D/3$ or $D/4$; index 'r' value is equal to the truncation distance times ten. Blue and variable-blue columns stand for T_{direct} for the fixed-resolution and variable-resolution decompositions, respectively, as indicated by 'F' and 'V' letters. The grid step of the resulted model map is equal to (**a**) $D/2$; (**b**) $D/3$; and (**c**) $D/4$.

The direct map calculation gives the results roughly for the same time or faster than the multi-step procedure, even when this gain is not of an order of magnitude. Using the fixed-resolution image decomposition, especially with $r_{map} = 4$ Å for the chosen resolution of 2 Å, is advantageous and can be used as a default option for real-space refinement. Using $r_{map} = 5$ Å at the final refinement iteration is also acceptable and recommended.

A particularly important property of the suggested procedure is the possibility to routinely calculate the maps of an inhomogeneous resolution from atomic models (Figure 3). Figure 6 shows that such calculation still be computationally efficient when using $r_{map} = 4$ Å and the output grid step $D/2$. Theoretically speaking, maps calculated on such grids contain all information which is contained in the maps with a finer grid and, therefore, may be sufficient for real-space refinement. Calculations of an inhomogeneous-resolution map on a finer grid or with a larger truncation radius may make $T_{density}$ larger than the total time of the three-step calculation, but this is the price for the possibility to introduce and refine individual atomic resolution D_n, and Figure 6 shows that this price is not excessive.

From a qualitative point of view, the mathematical features of (7) lead to a new concept when the local resolution is associated with atoms. As a consequence, it can be included into the list of the parameters to be refined and reported as the result of real-space refinement. A feature of the particular map used for refinement is that it characterizes the confidence of the atomic parameters. Another important point is that the discrepancy between the experimental and the model maps becomes an analytic function of all these parameters, and all necessary partial derivatives required for real-space refinement become analytic functions as well.

Concluding, the features discussed above make the real-space refinement of atomic coordinates and atomic displacement parameters feasible without appealing to reciprocal-space data and tools.

References

1. Cardone, G.; Heymann, J.B.; Steven, A.C. One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *J. Struct. Biol.* **2013**, *184*, 226–236. [CrossRef]
2. Urzhumtsev, A.G.; Lunin, V.Y. Introduction to crystallographic refinement of macromolecular atomic models. *Crystallogr. Rev.* **2019**, *25*, 164–262. [CrossRef]
3. Brown, A.; Long, F.; Nicholls, R.A.; Toots, I.; Emsley, P.; Murshudov, G. Tools for molecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr.* **2015**, *D71*, 136–153.
4. Lunin, V.Y.; Afonine, P.V.; Urzhumtsev, A.G. Likelihood-based refinement. I. Irremovable model errors. *Acta Crystallogr.* **2002**, *A58*, 270–282. [CrossRef] [PubMed]
5. Kostrewa, D. Bulk Solvent Correction: Practical Application and Effects in Reciprocal and Real Space. *Jt. CCP4 ESF-EACBM Newsl. Protein Crystallogr.* **1995**, *34*, 9–22.
6. Afonine, P.V.; Adams, P.D.; Sobolev, O.V.; Urzhumtsev, A. A mosaic bulk-solvent model improves density maps and the fit between model and data. *bioRxiv*. 2021. Available online: https://www.biorxiv.org/content/10.1101/2021.12.09.471976v1.full (accessed on 6 October 2022).
7. Afonine, P.V.; Urzhumtsev, A.; Adams, P.D. On the analysis of residual density distribution on an absolute scale. *Comput. Cryst. Newsl.* **2012**, *3*, 43–46.
8. Palmer, C.M.; Aylett, C.H.S. Real space in cryo-EM: The future is local. *Acta Crystallogr.* **2022**, *D78*, 136–143. [CrossRef] [PubMed]
9. Lunin, V.Y.; Urzhumtsev, A. Program construction for macromolecule atomic model refinement based on the fast Fourier transform and fast differentiation algorithms. *Acta Crystallogr.* **1985**, *A41*, 327–333. [CrossRef]
10. Urzhumtsev, A.G.; Lunin, V.Y. Fast differentiation algorithm and efficient calculation of the exact matrix of the second derivatives. *Acta Crystallogr.* **2001**, *A57*, 451–460. [CrossRef]
11. Diamond, R. A real-space procedure for proteins. *Acta Crystallogr.* **1971**, *A27*, 436–451. [CrossRef]
12. Abagyan, R.A.; Totrov, M.M.; Kuznetsov, D.A. Icm—A new method for protein modeling and design—Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488–506. [CrossRef]
13. Rice, L.M.; Brünger, A.T. Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins Struct. Funct. Genet.* **1994**, *19*, 277–290. [CrossRef] [PubMed]
14. Afonine, P.V.; Grosse-Kunstleve, R.W.; Urzhumtsev, A.G.; Adams, P.D. Automatic multiple-zone rigid-body refinement with a large convergence radius. *J. Appl. Crystallogr.* **2009**, *42*, 607–615. [CrossRef]
15. Merritt, E.A. To B or not to B: A question of resolution? *Acta Crystallogr.* **2012**, *D68*, 468–477. [CrossRef] [PubMed]
16. Cruickshank, D.W.J. The analysis of the anisotropic thermal motion of molecules in crystals. *Acta Crystallogr.* **1956**, *9*, 754–756. [CrossRef]
17. Schomaker, V.; Trueblood, K.N. On the rigid-body motion of molecules in crystals. *Acta Crystallogr.* **1968**, *B24*, 63–76. [CrossRef]
18. Doyle, P.A.; Turner, P.S. Relativistic Hartree-Fock X-ray and electron scattering factors. *Acta Crystallogr.* **1968**, *A24*, 390–397. [CrossRef]
19. Agarwal, R.C. A new least-squares refinement technique based on the fast Fourier transform algorithm. *Acta Crystallogr.* **1978**, *A41*, 327–333. [CrossRef]
20. Waasmaier, D.; Kirfel, A. New analytical scattering-factor functions for free atoms and ions. *Acta Crystallogr.* **1985**, *A34*, 791–809. [CrossRef]
21. Peng, L.-M. Electron atomic scattering factors and scattering potentials of crystals. *Micron* **1999**, *30*, 625–648. [CrossRef]

22. Grosse-Kunstleve, R.W.; Sauter, N.K.; Adams, P.D. CCTBX news. *Newsl. IUCr Comm. Crystallogr. Comput.* **2004**, *3*, 22–31.

23. Brown, P.J.; Fox, A.G.; Maslen, E.N.; O'Keefe, M.A.; Willis, B.T.M. Intensity of diffracted intensities. In *International Tables for X-ray Crystallography*; Prince, E., Ed.; Springer: Dordrecht, The Netherlands, 2006; Volume C, pp. 554–595.

24. Marques, M.A.; Purdy, M.D.; Yeager, M. CryoEM maps are full of potential. *Curr. Opin. Struct. Biol.* **2019**, *58*, 214–223. [CrossRef]

25. Carugo, O. B-factor accuracy in protein crystal structures. *Acta Crystallogr.* **2022**, *D78*, 69–74. [CrossRef] [PubMed]

26. Masmaliyeva, R.C.; Murshudov, G.N. Analysis and validation of macromolecular B values. *Acta Crystallogr.* **2019**, *D75*, 505–518. [CrossRef] [PubMed]

27. Sayre, D. The Calculation of Structure Factors by Fourier Summation. *Acta Crystallogr.* **1951**, *4*, 327–333. [CrossRef]

28. Ten Eyck, L.F. Efficient structure-factor calculation for large molecules by the fast Fourier transform. *Acta Crystallogr.* **1977**, *A33*, 486–492. [CrossRef]

29. Cooley, J.W.; Tukey, J.W. An algorithm for machine calculation of complex Fourier series. *Math. Comput.* **1965**, *19*, 297–301. [CrossRef]

30. Navaza, J. On the computation of structure factors by FFT techniques. *Acta Crystallogr.* **2002**, *A58*, 568–573. [CrossRef] [PubMed]

31. Afonine, P.V.; Urzhumtsev, A. On a fast and accurate calculation of structure factors at a subatomic resolution. *Acta Crystallogr.* **2004**, *A60*, 19–32. [CrossRef] [PubMed]

32. Rossmann, M.G. Fitting atomic models into electron-microscopy maps. *Acta Crystallogr.* **2000**, *D56*, 1341–1349. [CrossRef] [PubMed]

33. Rossmann, M.G.; Bernal, R.; Pletnev, S.V. Combining electron microscopic with x-ray crystallographic structures. *J. Struct. Biol.* **2001**, *136*, 190–200. [CrossRef]

34. Emsley, P.; Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr.* **2004**, *D60*, 2126–2132. [CrossRef] [PubMed]

35. Afonine, P.V.; Poon, B.K.; Read, R.J.; Sobolev, O.V.; Terwilliger, T.C.; Urzhumtsev, A.G.; Adams, P.D. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr.* **2018**, *D74*, 531–544.

36. Lunin, V.Y.; Urzhumtsev, A. Improvement of protein phases by coarse model modification. *Acta Crystallogr.* **1984**, *A40*, 269–277. [CrossRef]

37. Mooij, W.T.M.; Hartshorn, M.J.; Tickle, I.J.; Sharff, A.J.; Verdonk, M.L.; Jhoti, H. Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem* **2006**, *1*, 827–838. [CrossRef] [PubMed]

38. DiMaio, F.; Song, Y.; Li, X.; Brunner, M.J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T.; Cheng, Y.; Baker, D. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **2015**, *12*, 361–365. [CrossRef] [PubMed]

39. Chapman, M.S. Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. *Acta Crystallogr.* **1995**, *A51*, 69–80. [CrossRef]

40. Chapman, M.S.; Trzynka, A.; Chapman, B.K. Atomic modeling of cryo-electron microscopy reconstructions—Joint refinement of model and imaging parameters. *J. Struct. Biol.* **2013**, *182*, 10–21. [CrossRef] [PubMed]

41. Afonine, P.V.; Lunin, V.Y.; Muzet, N.; Urzhumtsev, A. On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps. *Acta Crystalogr.* **2004**, *D60*, 260–274. [CrossRef]

42. Urzhumtsev, A.; Urzhumtseva, L.; Lunin, V.Y. Direct calculation of cryo EM and crystallographic model maps for real-space refinement. *bioRxiv* **2022**. Available online: https://doi.org/10.1101/2022.07.17.500345 (accessed on 6 October 2022). [CrossRef]

43. Urzhumtsev, A.G.; Lunin, V.Y. Analytic representation of inhomogeneous-resolution maps of three-dimensional scalar fields. *bioRxiv* **2022**. Available online: https://doi.org/10.1101/2022.03.28.486044 (accessed on 6 October 2022). [CrossRef]

44. Simonetti, A.; Marzi, S.; Fabbretti, A.; Myasnikov, A.G.; Hazemann, I.; Jenner, L.; Urzhumtsev, A.; Gualerzi, C.O.; Klaholz, B.P. Crystal structure of the protein core of translation initiation factor IF2 in apo, GTP and GDP forms. *Acta Crystallogr.* **2013**, *D69*, 925–933.

45. Schrödinger, L.; DeLano, W.L. Pymol. Version 2.5.2. 2020. Available online: http://www.pymol.org (accessed on 6 October 2022).

46. Ten Eyck, L.F. Crystallographic fast Fourier transforms. *Acta Crystallogr.* **1973**, *A33*, 183–191. [CrossRef]