Structure Refinement at Atomic Resolution

Mariusz Jaskolski

Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University

and Center for Biocrystallographic Research, Institute of Bioorganic Chemistry,

Polish Academy of Sciences, Poznan, Poland

Abstract

X-Ray diffraction data at atomic resolution, i.e. beyond 1.2 Å, provide the most detailed and reliable information we have about the structure of macromolecules, which is especially important for validating new discoveries and resolving subtle issues of molecular mechanisms. Refinement at atomic resolution allows reliable interpretation of static disorder and solvent structure, as well as modeling of anisotropic atomic vibrations and even of H atoms. Stereochemical restraints can be relaxed or removed, providing unbiased information about macromolecular stereochemistry, which in turn can be used to define improved conformation-dependent libraries, and the surplus of data allows estimation of least-squares uncertainties in the derived parameters. At ultrahigh resolution it is possible to study charge density distribution by multipolar refinement of electrons in non-spherical orbitals.

Key words Atomic resolution, Stereochemical restraints, Conformation-dependent stereochemical libraries, H atoms, Multipolar refinement, Charge density, Standard uncertainties

1 Introduction

Strictly speaking, crystallographic resolution refers to the diffraction data and electron density maps (as their Fourier transform) and not to models, which are only interpretations of electron

density. It is defined as the minimum *d*-spacing in Bragg's Law ($\lambda = 2d_{min}\sin\theta_{max}$), corresponding to the maximum glancing angle θ_{max} at which statistically significant reflection intensities are still observed. It can be shown that the d_{min} limit corresponds almost exactly to the minimal separation of two points that can be distinguished in electron density maps generated by Fourier transformation.

On the somewhat arbitrary scale of resolution intervals (Fig. 1), the point at d_{min} = 1.2 Å is defined by Sheldrick [1] as atomic resolution. This choice, also supported by rigorous argument [2], is quite intuitive as it allows resolution of all non-H atoms, including the shortest (1.2 Å) C=O bond. For ultrahigh resolution we will use the 0.8 Å mark, corresponding approximately to the limit of Cu *K*α data.

2 Collecting atomic-resolution data

If a single advice is to be offered, it would be: *always get the highest resolution during your diffraction experiment*. This will ease all subsequent steps, will help reduce model bias, and will authenticate any unusual features discovered in the structure. However, collecting meaningful high-resolution data is not equivalent to "visiting" high-order *hkl* indices without statistically meaningful intensity signal. As a rule of thumb, we used to expect the average signal-to-noise ratio ($\langle I/\sigma(I)\rangle$) in the highest resolution shell to be at least 2, which is roughly equivalent to having ~50% of the data in that shell with $I > 2\sigma(I)$. The above criteria are rather conservative but will guarantee a high-quality data set. A resolution-oriented approach might, however, push d_{min} to the extreme limit, where adding more observations ceases to add information [3]. Statistically, this would correspond to a correlation coefficient CC_{true} between the experimental and ideal noise-free data of ~0.4. Karplus & Diederichs showed [4] that CC_{true} can be estimated by $CC_{½}$, which measures the correlation between two half-sets and should be acceptable even down to ~0.1. It is also good to have the last resolution shell as

complete as possible but incomplete resolution shells should not be rejected! On the contrary, every single reflection is precious and should always be included, particularly at high resolution. If completeness in the last resolution shell is poor, it is possible to estimate effective resolution (as opposed to nominal resolution) by finding that d_{eff} at which a reciprocal-lattice sphere of radius $1/d_{eff}$ would be filled completely. One can also estimate the optical resolution d_{opt} by a Gaussian analysis of the Patterson map based on optical principles, as proposed by Vaguine [5] and implemented in SFCHECK. One should remember that completeness of the lowest-resolution shells is important as well.

The outlier rejection criterion ($I < -3\sigma(I)$) used by data reduction programs should not be manipulated to "improve" the data set. Likewise, no σ-cutoff should be applied to select reflections for the refinement. However, in algorithms that use $|F_o|$ for refinement, the data will be effectively truncated at $0\sigma(I)$, by eliminating negative intensities during the $|F_o| = \sqrt{I}$ conversion. From this point of view, refinement algorithms based on reflection intensities, e.g. in SHELXL [6], are preferred.

The second most important parameter of a good data set is high redundancy, which will always improve data quality unless compromised by radiation damage. In addition to reducing random errors, multiple observations can be used to estimate standard deviations of intensity measurements. R_{merge} as a resolution-limiting criterion is not recommended as it deteriorates at high symmetry and with high redundancy. Better, redundancy-independent parameters (e.g. R_{rim}) were proposed by Diederichs & Karplus [7] and Weiss & Hilgenfeld [8,9].

If the dynamic range of the detector is insufficient to reliably record very strong and very weak data at the same time, it may be necessary to measure the strongest, low-resolution data in a (first) quick pass. At ultrahigh resolution, three overlapping runs may be necessary, e.g. ∞ - 2.0 Å, 2.4 - 1.0 Å and 1.5 - 0.7 Å, with relative 1:10:100 exposure.

3 Model refinement at high resolution step-by-step

3.1 Full or stepping resolution?

Historically, high order refinements were carried out with careful gradual extension of resolution. This strategy, dictated by limited computer power, is no longer necessary if the starting model is very good. When only an approximate model is available, starting the refinement at ~2 Å and even inclusion of a rigid-body step may be advisable to increase the radius of convergence.

3.2 Atomic Displacement Parameters (ADPs)

The first round of refinement at full resolution is done with isotropic atomic displacement parameters (ADPs, historically called *B*-factors), and is followed by model adjustment in electron density maps and inclusion of the most evident solvent molecules. Switching from isotropic (1 parameter per atom) to anisotropic (6 parameters per atom) model at this stage more than doubles the number of model parameters and brings about a dramatic decrease of the *R* factors (up to 0.05). Individual anisotropic ADPs are used for both the macromolecule and solvent atoms. They should not be mixed with TLS (Translation, Libration, Screw-motion) parameters, which describe concerted anisotropic motions of rigid structural fragments at medium resolution. The subsequent steps of the refinement protocol are listed in Table 1.

3.3 Prudent expansion of model parameters

If the resolution is 1.2 Å or higher, there is no question of the validity of using individual anisotropic ADPs. However, in the gray zone of 1.3-1.5 Å, the optimal strategy could be less obvious. The Protein Anisotropic Refinement Validation and Analysis (PARVATI) tool [10] and server (http://www.bmsc.washington.edu/parvati/) may be used to guide the optimal

choice of strategy in such cases. The question of expanding the ADP model from iso- to anisotropic is in fact part of a more general optimization problem, namely at what point the expansion of model parameters is no longer statistically justified by the experimental data and thus should be treated as overinterpretation. There is a rigorous “Occam’s razor”-type statistical *R*-factor ratio test for such hypotheses introduced by Hamilton [11] but its application to restrained macromolecular refinements is not obvious, although practical solutions have been proposed [12]. Merritt used the Hamilton *R*-factor ratio test to guide the iso/aniso decision [13] and concluded that at 1.5 Å the anisotropic model ceases to be valid, but also warned that proper statistical analysis should not be replaced by this rule of thumb.

3.4 Multiple conformations

As the increasing resolution permits distinction between closely spaced alternate occupancies, the proportion of fragments that are modeled in dual (or exceptionally triple) conformation increases as well. This only applies to static disorder. Dynamic disorder can be reduced by collecting the diffraction data at low temperature. Fractional occupancies of light atoms (C/N/O) are considered from ca. 0.2, or exceptionally from 0.1 at ultrahigh resolution, i.e. from electron density contribution equivalent to one H atom.

At 0.9 Å resolution or better, stereochemical restraints of well-ordered fragments may be gradually relaxed, or even removed altogether at ultrahigh resolution. However, disordered or multiple-conformation fragments should remain restrained as they are poorly defined by diffraction.

4 Application and validation of stereochemical restraint libraries

The geometrical (and other) restraints are extra equations that supplement the set of experimental equations, acting as “springs” that tie the model parameters (such as bond

lengths) to some pre-defined targets. The toughness of the spring is dictated by the variance (error estimate) of the target value. The restraints represent, therefore, some prior knowledge, which may be correct or not. It is thus important to be critical of such information and validate it whenever possible. Once wrong information has been fed to the system, it is very difficult to weed out. At lower resolution, the use of stereochemical restraints is absolutely necessary, simply to improve the data/parameter (d/p) ratio. At 1.2 Å, the d/p is ~3 even for anisotropic models and approaches 5 at 1.0 Å, making restraints dispensable from the mathematical point of view. However, while they may be relaxed in well ordered segments, flexible areas (e.g. side chains) still need to be restrained. At ultrahigh resolution, for well ordered structures, the refinement is highly overdetermined and stereochemical restraints may be eliminated altogether, as illustrated, for example, by the structure of Z-DNA at 0.55 Å resolution [14]. Under strict control of stereochemical restraints at lower resolution, model deviations from the target values should not exceed the uncertainties of the target estimates. In the case of protein bond lengths [15,16], this is on the order of 0.015-0.020 Å [17]. At very high resolution, the results are dominated by the diffraction terms and the root-mean-square deviations (r.m.s.d.'s) from the target values are likely to reflect errors in the targets themselves. Deviations as high as 0.02-0.03 Å could be still acceptable.

The target values were compiled by analyzing small-molecule databases about 20 years ago, for proteins by Engh & Huber [15,16] and for nucleic acids by several authors [18-20]. Although they are largely correct, some adjustments might be necessary. For example, the dictionary entry for the protein C-N peptide bond may need re-evaluation [17] and the peptide group planarity is most certainly enforced too strictly, distorting the adjacent φ/ψ backbone torsion angles and deteriorating the overall Ramachandran geometry [21]. The nucleic-acid parameters for the phosphate group and the valence angles at the guanine glycosidic bond also should be re-examined [14]. The situation is now very interesting because not only is the

small-molecule CSD database [22] over 10 times larger than when originally used for target evaluation, but we now have a subset of ultrahigh resolution structures in the PDB [23] with minimal target bias, from which the targets can be derived independently. Attempts to revise the Engh & Huber libraries have been already published. For example, Malinska et al. showed [24] that the imidazole ring of histidine can be restrained according to its protonation status, deduced from a trial refinement without restraints (at high resolution) or even from its H-bonding pattern (at lower resolution). In addition to covalent geometry, other model parameters, such as the ADPs or non-bonded contacts, are also restrained. Main-chain torsion angles should be left unrestrained to ensure bias-free model validation via Ramachandran plots.

5 Conformation-dependent stereochemical restraints

Macromolecular models refined at ultrahigh resolution are largely independent of the stereochemical targets (even if restraints have been included) and can be used for their validation and improvement. It has been noted in a number of studies that some (especially angular) parameters of such models have surprisingly wide distributions that could be correlated with the conformation and other characteristic features (e.g. H-bonding) of the macromolecules [17]. For instance, the N-Cα-C angle of the polypeptide backbone has a wide spread [21] and is correlated not only with residue type but also with the local φ/ψ backbone conformation [25]. By modeling main-chain bond distances and angles in proteins characterized at 1 Å resolution or better, as functions of the φ/ψ torsion angles, Tronrud & Karplus [26] were able to create a conformation-dependent stereochemical library (CDL) that leads to better models at lower resolution and, when applied at higher resolution, does not distort the models from the diffraction-driven target (r.m.s.d. for bonds ~0.007-0.010 Å) but, indeed, improves the results [27,28].

6 Unrestrained refinement vs. disorder

It is normal to relax the restraints at atomic resolution and restraint-free refinement is mathematically possible at ultrahigh resolution. However, from the point of view of the d/p ratio it is somewhat contradictory that the degree of discrete (static) disorder that can be modeled by fractional-occupancy conformations increases with resolution, thus demanding more parameters. As demonstrated for the case of BPTI, the percent of disordered residues that are seen even in the same crystal structure increases with resolution [29,21] and reaches 21% at 0.86 Å. In the 0.66 Å crystal structure of human aldose reductase, one-third of all residues were modeled in multiple conformations [30]. This makes the improvement of d/p less spectacular, as many parameters have to be invested in poorly defined fragments, and requires the retention of stereochemical restraints in multiple-conformation areas. The disorder is usually visible in the macromolecule and in the solvent region, and it is often found to form correlated networks, which should be identified and refined with common occupancy.

7 Treatment of H atoms

The X-ray scattering power of the H atom is very low and, therefore, H atoms are normally omitted in modeling macromolecular crystal structures. Although at high θ angles the scattering cross section diminishes further, paradoxically H atoms can be better visualized using high-resolution data because the disproportion to C/N/O scattering is less drastic. Besides, H atoms in X-H bonds, which are 0.9-1.1 Å long, can be delineated only when the resolution reaches this level. Even if the H atoms are not fully resolved by diffraction, it is still advisable to include their contribution to F_c to improve agreement with F_o and to remove bias in the location of the parent atoms (on which the H atoms are “riding”) that otherwise

suffer form the “expanded” skeleton syndrome. The positions of most H atoms in proteins and nucleic acids are easily generated from the skeleton of the remaining atoms, and their contribution at atomic resolution will typically decrease the *R* factor by ca 0.01. The H atoms in $-NH_3^+$ and $-CH_3$ groups, in the ambiguously protonated His residues, -OH groups and (possibly) carboxylic groups cannot be generated blindly and have to be analyzed individually, usually based on logical H-bond circuits. Generation of H atoms with fractional occupancy is not sensible. Generation of H atoms in water molecules cannot be done fully automatically, although there are algorithms that claim to challenge even neutron scattering data. Considering the high proportion of water molecules with fractional occupancy, it is doubtful if en bloc generation of water H atoms would be meaningful. Those special cases where water H atoms are important and are clearly defined in electron density should be dealt with manually.

With the overwhelming overdeterminacy at ultrahigh resolution, full refinement of H-atom parameters (x,y,z,B_{iso}) is possible as in small-molecule crystallography. Such tests have been carried out but the minimal gain (e.g. negligible drop of the *R* factor) does not justify the massive effort needed to verify the results. It is thus concluded that even at very high resolution, conventional refinement should include riding H atoms and, if necessary, only the key H atoms should be refined individually. Some protocols place or shift H atoms along the X-H bonds to neutron distances [31]. While this procedure yields a geometrically correct model, it is not necessarily compatible with X-ray refinement of spherical atoms. Moreover, normalization of H atoms in very short (and thus of key importance) hydrogen bonds may be simply unjustified [32].

8 Electron density maps at atomic resolution

Work with electron density maps at better than atomic resolution is very gratifying

because they show most of the atoms as well resolved spheres (Fig. 2). The electron density maps can use $2F_o$-F_c coefficients, or $3F_o$-$2F_c$ coefficients as recommended by Lamzin & Wilson [33], but the difference is not very obvious. At very high resolution even F_o maps can be used as series termination effects are negligible. For difference maps, σ_A-derived coefficients are usually used [34]. For methodological correctness, electron density maps should be contoured in absolute e/Å^3 units, but the values from Fourier summation are not absolute because of the missing strong low-order terms and the unknown $F(000)$ term. Owing to the low noise level of accurate maps, the σ unit frequently used for map contouring is usually low and meaningful features correspond to high-level contours.

9 R_{free} validation

Calculation of R_{free} [35] is the standard way for validating crystallographic models and the process of their generation. Although refinement at atomic resolution is usually not frustrated with profound strategic ambiguities, some decisions are clearly validated by reference to R_{free}. It is enough to set aside 1000-2000 test reflections, rather than applying the 5-10% rule, which could be very wasteful concerning the large data set size at atomic resolution. One should ensure that the test reflections are randomly selected from the entire data set, i.e. including the highest resolution shell as well. When the model has been completed, the test reflections should be included in the working set for a final round of refinement and for the generation of final electron density maps. This will further improve the final d/p ratio and reduce series truncation errors in the Fourier transform, i.e. will lead to better results, which is the ultimate goal of any high-resolution study.

10 Estimation of standard uncertainties

The program particularly well-suited to refining ultrahigh resolution structures is the least-squares oriented SHELXL. Most of the refinement cycles in SHELXL are done using the conjugate-gradient algorithm, which, in the interest of speed, circumvents the inversion of the least-squares matrix. The last refinement cycle (for diagnostic purposes, without application of parameter shifts) should be calculated in the full (or blocked) -matrix least-squares mode to estimate the standard uncertainties (s.u.) in the atomic parameters. This is done for all reflections but without restraints and usually for positional parameters only (to obtain s.u.'s of geometrical parameters). If the problem is prohibitively large (over 100 residues), the matrix can be blocked into 50-residue segments (with 5-residue overlap) that will be refined in alternating cycles. Accurately estimated s.u.'s are a treasure trove because they allow meaningful interpretation of model geometry. For instance, it is possible to gauge significant vs. insignificant geometry differences, or even evaluate potential errors in the stereochemical standards. At ultrahigh resolution, the s.u.'s in bond lengths are as low as in small-molecule crystallography. In the 0.55 Å structure of Z-DNA, these values are 0.002-0.004 Å [14], while in the 0.86 Å structure of BPTI they are on the order of 0.005-0.02 Å [29].

11 Multipolar refinement and deformation density studies vs. “interatomic scatterers”

At ultimate resolution, higher than 0.7 Å, one may contemplate charge (or deformation) density studies and multipolar refinement. Deformation density studies aim at mapping deviations of atomic electron clouds from the classic (but incorrect in covalent molecules) spherical independent-atom model (IAM) (Fig. 3). Such studies require data of very high resolution and are rare even in small-molecule crystallography. In multipolar expansion, the atomic electrons are partitioned into core and valence shells, and the latter ones are described by multipolar functions [36]. Depending on the level of multipolar expansion, multipolar

atoms require from 3 (for H) to 27 (heavy elements) extra parameters, in addition to the usual 3 coordinates and 6 anisotropic ADPs. Usually, the first round of refinement uses reflections from the high-resolution shell only, to distill thermal motion parameters of non-H atoms from the electron distribution functions. Subsequent cycles refine the multipolar parameters of a subset of atoms with excellent order and low thermal motion. Even if ultrahigh resolution is not achieved, deformation density studies are still possible using libraries of transferable multipolar atom models derived from experiment (ELMAM) [37] or by theoretical calculations [38].

Experimental charge-density studies of macromolecules are extremely rare and are limited to aldose reductase, a protein of 316 residues, analyzed at 0.66 Å resolution [39], and to crambin (46 residues), analyzed at 0.54 Å [40] and at 0.48 Å [41]. No charge-density studies for nucleic acids have been published so far, but one study is underway (Fig. 3).

As an alternative to the rigorous multipolar refinement, a simple-minded approximation has been proposed to use “pseudo-atom” scatterers at midpoints of covalent bonds that would take care of the bonding electrons [42]. This simplistic approach is not quite on a par with accurate high-resolution studies.

12 Solvent structure

As a rule of thumb, one should be allowed to model up to (3-|d_{min}|) water molecules per residue [43]. For ultrahigh resolution structures, for which the Matthews fractional volume of solvent [44] is usually low, it is often possible to locate nearly all water molecules, although the count is complicated because many solvent molecules in atomic-resolution structures show a high degree of disorder, populating many sites with partial occupancy. In fact, modeling the outer hydration shell in high-resolution macromolecular structures is usually the most frustrating step, well justifying the opinion that “macromolecular refinement against

high-resolution data is never finished, only abandoned" [6]. Despite the near-complete atomic interpretation of the solvent region at high resolution, it is common to include in the refinement a bulk-solvent correction, for instance based on Babinet's principle (Fourier transforms of a mask and its complement have the same amplitudes, but opposite phases), which affects only very low-resolution (d > 15 Å) data.

The site occupation factors of (even all) water molecules could be refined together with their ADPs but a more prudent approach is to fix them after manual or automatic adjustment. After a round of occupancy (occ) refinement, one would (i) eliminate phantom molecules (occ < 0.2), (ii) fix those refined to occ > 0.9 at 1.0, (iii) couple the occupancies of alternate sites (O···O distance < 2 Å), and (iv) let the remaining occupancies refine freely. A water molecule that is retained in the model should have clear $2F_o$-F_c electron density at the 1σ level, should form at least one reasonable hydrogen bond (2.3 - 3.2 Å), and should not have prohibitively short contacts, e.g. with C atoms; however, the possibility of forming C-H···O hydrogen bonds (which are usually long, C···O ~3 Å) should not be overlooked.

Water molecules are not to be confused with metal cations. Although such species can be isoelectronic (e.g. $H_2O/Na^+/Mg^{2+}$), metal cations are likely to form shorter bonds (e.g. Mg···O ~2 Å), do not have typical proton donors (such as amide N-H) in their coordination sphere, and will often have more than four ligands, e.g. six in the case of octahedral Mg coordination.

13 Benefits of atomic resolution

The benefits of atomic-resolution macromolecular structures have been discussed in several reviews, e.g. [45-47]. They are certainly worth the considerable effort that must be invested in the experiment, computations, and interpretation of the results. By improving the d/p ratio, high-resolution data help to remove model bias, which blights crystallographic structures solved by molecular replacement. More reflections and better resolving power allow accurate

interpretation of multiple conformations, yielding more realistic models and better agreement with the experiment. Unusual stereochemical features are best confirmed at atomic resolution. Refinement with relaxed or eliminated stereochemical restraints is the surest way to the discovery of new phenomena that could be masked by data paucity at low resolution and/or prejudiced ideas about the result. Restraint-free refinement can ultimately produce accurate dictionaries of macromolecular stereochemistry, to be used as restraints at lower resolution. Restraint-free refinement with sufficiently high d/p ratio allows the application of full-matrix least-squares, from which the standard uncertainties of the geometrical parameters can be estimated. The determination of both, the parameters and their error estimates, places the discussion of macromolecular geometry at an entirely new, statistically sound level, and is only possible in crystallography. Although H atoms have only minimal contribution to X-ray scattering and are normally omitted from models of macromolecular structure, they are often of key importance for understanding the functioning of macromolecules, e.g. in enzyme catalysis or fine-tuned intermolecular recognition. Any sensible experimental interpretation of H atoms requires X-ray diffraction data of very high resolution. Indeed, there is evidence suggesting that careful ultrahigh resolution X-ray analysis could be superior in this respect to macromolecular neutron diffraction, which requires prohibitively large crystals (~1 mm^3), deuterated solvent or even perdeuterated protein, and is normally limited to only medium resolution. Even if H atoms are not visualized in electron density maps, their location is easily deduced from the framework of the C/N/O atoms and even in the toughest cases (such as the O-H groups) their placement can be often predicted in atomic-resolution structures not only from H-bond networks but also from the patterns of bond lengths involving the heavier atoms [48]. Also, solvent molecules, which are often disordered and not amenable to accurate modeling at lower resolution, can get sensible interpretation at atomic resolution. Finally, when ultimately high resolution data are available, it is possible to interpret the

macromolecular structure at a level of detail that goes far beyond the localization of atoms. Such charge density studies, which involve refinement of multipolar parameters as mock orbitals, are still rare but they are beginning to unveil a fascinating inner world of the macromolecules at the level of electrons in atoms, in interatomic bonds and in intermolecular interactions. Charge density studies also provide a better estimate of the electrostatic properties of biological macromolecules, which are important for understanding their function.

14 Popular refinement programs

Refinement at very high resolution is usually carried out in SHELXL, which uses conventional, accurate structure-factor summations [6]. Test calculations with least-squares targets seem to indicate that the results of SHELXL and REFMAC [49] are similar. However, newer versions of programs such as REFMAC or phenix.refine [50] allow refinement against maximum-likelihood targets, not available in the least-squares oriented SHELXL algorithm, and it is yet to be seen if this offers any benefit at high resolution. SHELXL does have, however, its advantages, which include (i) very versatile definition of stereochemical (and other) restraints, (ii) flexible definition and refinement of free variables (assigned to selected groups of parameters), (iii) strictly enforced refinement on intensities rather than structure factor amplitudes, and (iv) the possibility to estimate standard uncertainties in the refined parameters (and in the derived geometrical parameters of the model) through the explicit inversion of the Hessian matrix. Protein multipolar refinement is only possible in a dedicated program, MoPro, developed by Jelsch et al. [51].

References

1. Sheldrick GM (1990) Phase Annealing in SHELX-90: Direct Methods for larger structures. Acta Crystallogr A 46:467–473

2. Bricogne G, Morris RJ (2003) Sheldrick's 1.2 Å rule and beyond. Acta Crystallogr D 59:615–617
3. Evans P (2012) Resolving some old problems in protein crystallography. Science 336:986–987
4. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. Science 336:1030–1033
5. Vaguine AA, Richelle J, Wodak SJ (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallogr D 55:191–205
6. Sheldrick GM (2008) A short history of SHELX. Acta Crystallogr A 64:112–122
7. Diederichs K, Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nature Struct Biol 4:269–274
8. Weiss M, Hilgenfeld R (1997) On the use of the merging R factor as a quality indicator for X-ray data. J Appl Crystallogr 30:203–205
9. Weiss M (2001) Global indicators of X-ray data quality. J Appl Crystallogr 34:130–135
10. Merritt EA (1999) Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Crystallogr D 55:1109–1117
11. Hamilton WC (1965) Significance tests on the crystallographic R factor. Acta Crystallogr 18:502–510
12. Bacchi A, Lamzin VS, Wilson KS (1996) A self-validation technique for protein structure refinement: the extended Hamilton test. Acta Crystallogr D 52:641–646
13. Merritt EA (2012) To B or not to B: a question of resolution? Acta Crystallogr D 68:468–477
14. Brzezinski K, Brzuszkiewicz A, Dauter M, Kubicki M, Jaskolski M, Dauter Z (2011) High regularity of Z-DNA revealed by ultra high-resolution crystal structure at 0.55 Å. Nucleic Acids Res 39:6238–6248
15. Engh RA, Huber R (1991) Accurate bond and angle parameters for X-ray protein structure refinement. Acta Crystallogr A 47:392–400
16. Engh RA, Huber R (2001) International Tables for Crystallography, vol. F, edited by Rossmann MG, Arnold E, pp. 382–392. Dordrecht: Kluwer Academic Publishers
17. Jaskolski M, Gilski M, Dauter Z, Wlodawer A (2007) Stereochemical restraints revisited: how accurate are refinement targets and how much should protein structures be allowed to deviate from them? Acta Crystallogr D 63:611–620
18. Parkinson G, Vojtechovsky J, Clowney L, Brünger AT, Berman HM (1996) New parameters for the refinement of nucleic acid-containing structures. Acta Crystallogr D 52:57–64
19. Clowney L, Jain SC, Srinivasan AR, Westbrook J, Olson WK, Berman HM (1996) Gemoetric parameters in nucleic acids: nitrogenous bases. J. Am. Chem. Soc. 118:509–518
20. Gelbin A, Schneider B, Clowney L, Hsieh S-H, Olson WK, Berman HM (1996) Gemoetric parameters in nucleic acids: sugar and phosphate constituents. J. Am. Chem. Soc. 118:519–529
21. Addlagatta A, Czapinska H, Krzywda S, Otlewski J, Jaskolski M (2001) Ultrahigh-resolution structure of a BPTI mutant. Acta Crystallogr D 57:649–663
22. Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58:380–388
23. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
24. Malinska, M, Dauter, M, Kowiel, M, Jaskolski, M, Dauter, Z (2015) Protonation and geometry of histidine rings. Acta Crystallogr D:1444–1454
25. Berkholz DS, Shapovalov MV, Dunbrack RLJ, Karplus PA (2009) Conformation dependence of backbone geometry in proteins. Structure 17:1316–1325
26. Tronrud DE, Karplus PA (2011) A conformation-dependent stereochemical library improves crystallographic refinement even at atomic resolution. Acta Crystallogr D 67:699–706
27. Moriarty NW, Tronrud DE, Adams PD, Karplus PA (2014) Conformation-dependent backbone geometry restraints set a new standard for protein crystallographic refinement. FEBS J 281:4061–4071

28. Moriarty NW, Tronrud DE, Adams PD, Karplus PA (2015) A new default restraint library for the protein backbone in Phenix: a conformation-dependent geometry goes mainstream. Acta Crystallogr D 72:176–179

29. Czapinska H, Otlewski J, Krzywda S, Sheldrick GM, Jaskolski M (2000) High resolution structure of bovine pancreatic trypsin inhibitor with altered binding loop sequence. J Mol Biol 295:1237–1249

30. Howard ER, Sanishvili R, Cachau RE, Mitschler A, Chevrier B, Barth P, Lamour V, Van Zandt M, Sibley E, Bon C, Moras D, Schneider TR, Joachimiak A, Podjarny A (2004) Ultrahigh resolution drug design I: Details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. Proteins 55:792–804

31. Allen FH (1986) A systematic pairwise comparison of geometric parameters obtained by X-ray and neutron diffraction. Acta Crystallogr B 42:515–522

32. Olovsson I, Jaskolski M (1986) Reaction Coordinate for the Proton Transfer in Some Hydrogen Bonds. Pol J Chem 60:759–766

33. Lamzin VS, Wilson KS (1997) Automated refinement for protein crystallography. Methods Enzymol 277:269–305

34. Read RJ (1997) Model phases: probabilities and bias. Methods Enzymol 277:110–128

35. Brünger AT (1992) Free R-value: a novel statistical quantity for assessing the accuracy of the crystal structures. Nature 335:472–475

36. Hansen NK, Coppens P (1978) Testing aspherical atom refinements on small-molecule data sets. Acta Crystallogr A 34:909–921

37. Zarychta B, Pichon-Pesme V, Guillot B, Lecomte C, Jelsch C (2007) On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures. Acta Crystallogr A 63:108–125

38. Koritsanszky T, Volkov A, Coppens P (2002) Aspherical-atom scattering factors from molecular wave functions. 1. Transferability and conformation dependence of atomic electron densities of peptides within the nultipole formalism. Acta Crystallogr A 58:464–472

39. Guillot B, Jelsch C, Podjarny A, Lecomte C (2008) Charge-density analysis of a protein structure at subatomic resolution: the human aldose reductase case. Acta Crystallogr D 64:567–588

40. Jelsch C, Teeter MM, Lamzin V, Pichon-Pesme V, Blessing RH, Lecomte C (2000) Accurate protein crystallography at ultra-high resolution: Valence electron distribution in crambin. Proc Natl Acad Sci USA 97:3171–3176

41. Schmidt A, Teeter M, Weckert E, Lamzin VS (2011) Crystal structure of small protein crambin at 0.48 Å resolution. Acta Crystallogr F 67:424–428

42. Afonine PV, Grosse-Kunstleve RW, Adams PD, Lunin VY, Urzhumtsev AG (2007) On macromolecular refinement at subatomic resolution with interatomic scatterers. Acta Crystallogr D 63:1194–1197

43. Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J 275:1–21

44. Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33:491–497

45. Dauter Z, Lamzin V, Wilson K (1997) The benefits of atomic resolution. Curr Opin Struct Biol 7: 681–688

46. Thaimattam R, Jaskolski M (2004) Synchrotron radiation in atomic-resolution studies of protein structure. J Alloys Comp 362:12–20

47. Jaskolski M (2013) High resolution macromolecular crystallography. In: Advancing Methods for Biomolecular Crystallography, edited by Read R, Urzhumtsev AG, Lunin VY, pp. 259–275. Springer

48. Wlodawer A, Li M, Gustchina A, Dauter Z, Uchida K, Oyama H, Goldfarb NE, Dunn BM, Oda K (2001) Inhibitor complexes of the *Pseudomonas* serine-carboxyl proteinase. Biochem 40:15602–15611

49. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) Acta Crystallogr D 67:355–367

50. Adams PD, Afonine PV, Bunkóczi B, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D 66:213–221

51. Jelsch C, Guillot B, Lagoutte A, Lecomte C (2005) Advances in protein and small-molecule charge-density refinement methods using MoPro. J Appl Crystallogr 38:38–54

52. Rosenbaum G, Ginell SL, Chen JC (2015) Energy optimization of a regular macromolecular crystallography beamline for ultra-high-resolution crystallography. J Synchrotron Radiat. 22:172–174

53. Wang J, Dauter M, Alkire R, Joachimiak A, Dauter Z (2007) Triclinic lysozyme at 0.65 Å resolution. Acta Crystallogr D 63:1254–1268

Table 1. Stages of macromolecular refinement at atomic resolution. The steps listed in parentheses are only possible at ultrahigh resolution.

Step	**Action**
1	Include reflections at full resolution
2	Isotropic ADPs
3	Correction of model errors, evident solvent molecules
4	Bulk-solvent correction
5	Anisotropic ADPs
6	Modeling of disorder
7	Riding H atoms
8	Partial water molecules
9	Refine/adjust occupancies
10	Relax/(remove) restraints
(11)	(H-atoms refined)
12	Include all reflections (work+test)
(13)	(Multipolar refinement)
14	Full-matrix least-squares

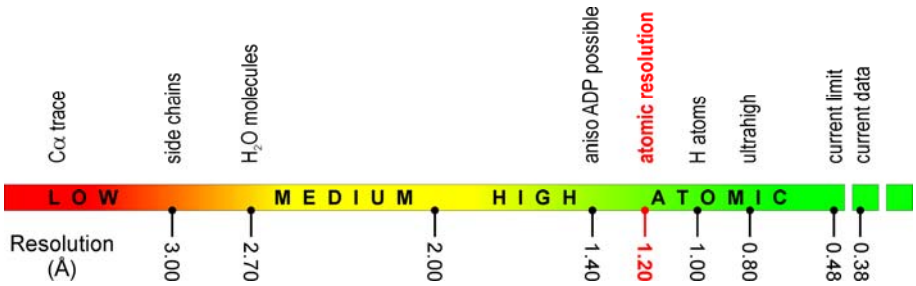

Fig. 1. In an arbitrary division of crystallographic resolution into descriptive ranges, only the criterion of atomic resolution (1.2 Å) has precise definition [1]. The annotations indicate the justified level of interpretation. The highest-resolution structure in the PDB (3nir) is at 0.48Å for crambin [41]. A 0.38-Å data collection for the same crystal form was announced in the literature [52] but no structure has been reported yet.

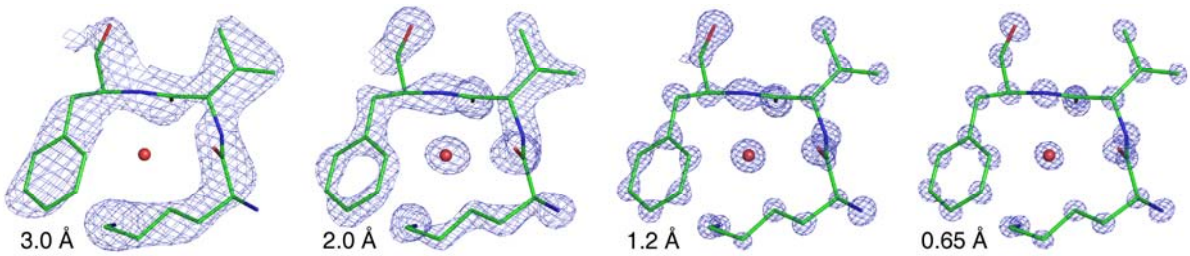

Fig. 2. $2mF_o$-DF_c electron density map (blue) calculated at four levels of resolution (3.0, 2.0, 1.2, 0.65 Å) for the Lys1-Val2-Phe3 fragment (sticks) of triclinic lysozyme (PDB code 2vb1) [53], contoured, respectively, at the following level: 1.5σ (0.7 *e*/Å^3), 2.0σ (1.3 *e*/Å^3), 3.0σ (2.2 *e*/Å^3) and 3.6σ (3.0 *e*/Å^3). The absolute contours (in parentheses) were estimated using as *F*(000) the total count of electrons in the model in the unit cell. The following numbers of reflections were used for each map generation: 1909 (3.0 Å), 6519 (2.0 Å), 30325 (1.2 Å) and 185045 (0.65 Å). Note that the electron density of a water molecule (red sphere) becomes apparent only at ~2.7 Å resolution. Also note that the presented maps, generated from artificially truncated (1.2, 2.0 and 3.0 Å) but otherwise ultrahigh (0.65 Å) resolution data, are better than typical maps calculated for data extending only to (and thus losing statistical significance at) such a resolution. The σ level for map contouring was estimated from electron density distribution in the entire unit cell, and not around the illustrated fragment. Figure provided by Z. Dauter.

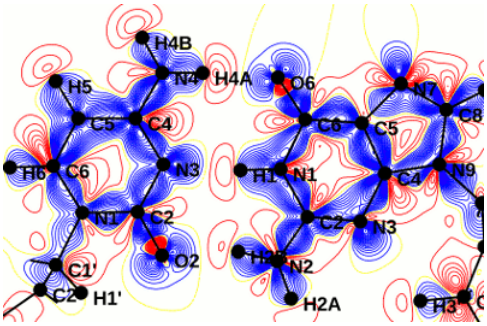

Fig. 3. Static deformation density of a C•G base pair in a Z-DNA structure after multipolar refinement at 0.55 Å resolution. Static deformation density represents charge distribution calculated for atoms at rest as the sum of all multipolar contributions after subtraction of the spherical IAM approximations. The figure is therefore an illustration of the asphericity of real atoms in molecules. Note, for example, the electrons in covalent bonds or in the lone pairs (“rabbit ears”) of the oxygen atoms. The contours (solid blue - positive, dash red - negative) are drawn with 0.05 e/Å^3 increment, starting from 0 e/Å^3 (dash green contour). Figure provided by M. Kubicki (unpublished results).