REAL-SPACE REFINEMENT USING RSREF

ERIC BLANC, ZHI CHEN & MICHAEL S. CHAPMAN†
Department of Chemistry & Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306-3015, USA.

Abstract: Real-space methods for the refinement of macromolecular structures are briefly presented with regard to their most common application, when "experimental" phases are accurate. Recent improvements extend their use to initial stages of protein refinements, when phases are poor. RSRef can also be used interactively to hasten model (re)building and improve the starting point for conventional refinement.

1. Introduction

Although real-space methods of refinement were successfully applied to some of the first protein structures ([11], [18] for example), macromolecular structures are now usually refined by reciprocal space methods ([4] and [13]), as they are independent of the experimental phases, that are poorly determined. By contrast, real-space methods rely implicitly on phase information, as they minimize a residual based on the squared difference between observed and calculated electron densities: the observed electron density is calculated using experimental phases (from isomorphous replacement, symmetry averaging, *etc.*) and/or phases from a preliminary atomic model, in which case refinement may be biased towards the preliminary model.

Nevertheless, there are several niches for which real-space refinement is well-suited:

1. Manual optimization: A lot of time is spent manually optimizing a structure before or between rounds of reciprocal-space refinement. Because real-space refinement is a local method, *i.e.* a small part of the structure can be refined independently from the rest, it is a very fast and therefore very efficient way to keep a good fit between the model and the observed map while interactively

† To whom correspondence should be addressed.

S. Fortier (ed.), Direct Methods for Solving Macromolecular Structures, 513–519.

remodeling. For this reason, real-space refinement has been implemented within interactive molecular modeling programs (see [2] and [15] for example).

2. Accurate phases: When accurate phases are available (from MAD or averaging for example) Rees and Lewis [16], and Arnold and Rossmann [1] have shown that reciprocal-space refinement was improved through the addition of explicit phase restraints. These phase-restrained reciprocal-space refinements are near-equivalents of the real-space methods described.

3. Viruses: The very high non-crystallographic symmetry (ncs) present in icosahedral viruses has made real-space refinement the method of choice for optimizing such structures. The benefits of ncs are two fold. First, phases are of unusual quality because of the high order averaging (typically 15 to 60 fold) applied to the observed map (see accompanying articles [8] and [9]). As already pointed out, refinement is improved by the addition of phase information, if accurate. Secondly, real-space refinement is much faster than reciprocal-space methods, partly because it is carried out on a non-crystallographic asymmetric unit only, which contains typically 15 to 60 fold fewer atoms than the crystallographic asymmetric unit used by reciprocal space methods. It should also be noted that real-space refinement fits the model to the complete diffraction data set at once, whereas it is common to improve the efficiency of reciprocal-space virus structure refinement by using only alternating subsets of the *ca.* 1 million reflections on each cycle.

For conventional protein refinement, for example a protein structure with poor "observed" phases from MIR or molecular replacement, real-space methods are not widely used, even though it has been recently shown [7] that refinements could be improved by alternating real- and reciprocal-space methods. New techniques are currently under development to increase the power of real-space methods during the initial stages of refinement. These new methods, briefly described in section 3, take advantage of the large convergence radius of real-space refinement [12], of the much shorter range of interdependence of atoms, and of the relatively low computing requirements of real-space refinement.

2. Theory

The real-space refinement minimizes a least-squares residual S with respect to a set of atomic parameters $\{p_n\}$, where n runs over the refined atoms. The residual depends on atomic positions ξ_n, displacement parameters B_n and occupancies O_n; all these individual atomic parameters will be generically referred to as p_n. S is thus given by:

$$S = \int_V \left[\rho_{\text{obs}}(\mathbf{x}) - \rho_{\text{calc}}(\mathbf{x};\{p_n\})\right]^2 dx\,, \qquad (1)$$

where ρ_{obs} and ρ_{calc} are respectively the observed and calculated electron densities (put on an absolute scale) evaluated at any point **x** in the crystal. The integration domain V depends on the positions of the refined atoms: V is the volume occupied by spheres of given radius R_n centered on atom position ξ_n. It must be noted that this dependency is not taken into account during the minimization process.

The calculated electronic density is assumed to be a linear superposition of individual atomic contributions:

$$\rho_{\text{calc}}(\mathbf{x};\{p_n\}) = \sum_m O_m \rho_m(\|\mathbf{x} - \xi_m\|; B_m) \ , \qquad (2)$$

where $\rho_m(r)$ is the electron density of atom m at a distance r from its center. The individual atomic densities ρ_m are weighted by an atomic occupancy factor O_m and depend upon a displacement parameter B_m, which is usually assumed isotropic. In principle, all atoms in the crystal contribute to $\rho_{calc}(\mathbf{x})$, but as their distance to the point **x** increases, their contribution to the sum becomes vanishingly small. In practice, atoms for which $\|\mathbf{x} - \xi_m\|$ is larger than a cut-off radius R_m are not used in (2). It implies that unrefined atoms contribute to (2) if they are close to the refined region, or more precisely if their distance to a grid point in V is smaller than R_m.

Unlike previous implementations of real-space refinement, RsRef [6] explicitly accounts for resolution limits of the experimental data: ρ_m is calculated by the Fourier transform of the apparent atomic form factor smeared by the "temperature" factor $\exp(-B_m \|\mathbf{h}\|^2/4)$ (f_m), and restricted to the resolution limits d^*_{low} and d^*_{high}. With $h = \|\mathbf{h}\|$, ρ_m is given by

$$\rho_m(r) = \frac{2}{r}\int_{d^*_{low}}^{d^*_{high}} h\, f_m(h; B_m)\, \sin(2\pi h r)\, dh \qquad (3)$$

for r smaller than R_m, usually independent of the atom. For larger r, ρ_m is set to 0.

The resolution limits are responsible for the truncation ripple seen in electronic density maps, even for very accurate experimental or model phases. Figure 1a shows the comparison of the apparent density of a carbon atom "seen" at infinite resolution or only between 10^{-1} and 3^{-1} Å^{-1}. The difference between these two formulae is enhanced when the density is weighted by r^2 (fig. 1b), proportional to the surface of a spherical shell of radius r, and according to the total contribution to refinement of all grid points at a given r.

For macromolecular refinement at medium resolution, real-space refinement must be used with stereochemical information, that can be provided in several ways. RsRef is written as a module of the TNT package [17], using TNT's stereochemical restraints and minimizer.

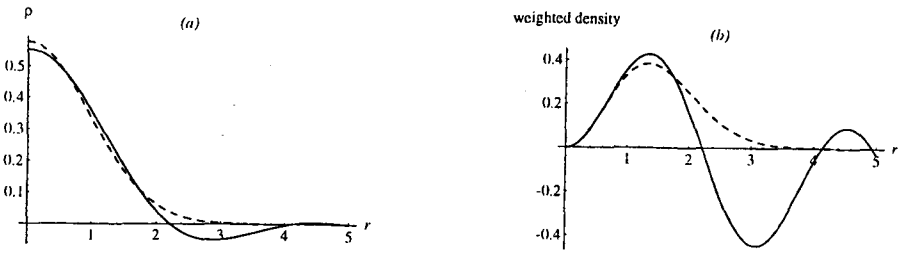

Figure 1: Effects of resolution limits on the apparent atomic electronic density. (a) Apparent electronic density of a carbon atom smeared by a displacement factor of B = 15.0, as a function of the distance from the atomic center r in Å. The continuous line shows the density for data collected between 10^{-1} and 3^{-1} Å^{-1}, the dashed line shows the apparent density at infinite resolution (sum of gaussians). (b) shows the densities of (a) weighted by the square of the radius, which is proportional to the contribution of each spherical shell to the total density.

3. Recent Developments

3.1. INITIAL REFINEMENT

When phases are poor, real-space methods suffer from their dependence to phases. While unlikely to replace reciprocal-space methods, it has recently been shown [7] that real-space methods complement the more usual methods effectively. Starting with an unrefined initial model and using a poor MIR-phased map, alternating real- and reciprocal-space refinements lead to an improvement of the model (monitored by R_{free}) nearly as great as can be realized through many hours of labor-intensive rounds of interactive model building and re-refinement. One efficient protocol for initial refinement iterates real- and reciprocal-space refinements with $2F_o$-F_c map calculation. It demonstrates the complementarity of real- and reciprocal space refinements, as the improvement realized by either method alone is smaller.

During the initial refinement, real-space methods likely improve the conditioning of the optimization by increasing the data to parameter ratio with phase information. Conditioning is also improved by restricting the interdependencies of atoms to local interactions, leading to improved convergence. Reduced interdependence also helps avoid overfitting (frequently observed in reciprocal-space refinements) by forcing a good *local* fit of the model and the experimental data. In a complementary way, by uncoupling atoms from phases, and reducing phase bias, the reciprocal-space refinement benefits real-space refinement.

Eventually however, progress of real-space methods is limited by the bias of the calculated phases. Several methods are under development to overcome this limitation.

During initial stages of model building, the calculation of phases from a partial or preliminary model has long been a method of improving experimental phases. It was

soon recognized that these calculated phases have the potential to bias electron density maps towards this model, parts of which might be incorrect. Omit maps have been used to reduce such bias, but remaining bias has been the cause of model errors in several structures [3]. The source of bias in omit maps is refinement, which can cause atoms outside the omit region to adjust away from their correct positions to compensate for errors elsewhere and minimize the overall residual [14]. Real-space local refinements are thus well suited for reducing remaining bias, because mutually compensating adjustments of atoms from different regions of the structure are no longer possible. Implementations of such protocols are undergoing preliminary tests.

Complementing the efforts above, real-space refinement has been implemented with simulated annealing molecular dynamics replacing least-squares optimization. Preliminary tests indicate wide convergence radius for this refinement. Moreover, the implicit phase restraints of real-space refinement should decrease the risk of finding an incorrect structure with a low crystallographic residual when using powerful molecular dynamics searches. It also offers the possibility of fast local refinements.

3.2. DETERMINING THE QUALITY OF REAL-SPACE REFINED MODELS

Although standard free *R*-factor [5] have been used to date, recently it has become apparent that they do not show real-space refinement in fair light. Real-space methods are very sensitive to missing data. The common practice of removing part of the data set for cross-validation purposes is much more detrimental to refinement in real-space than in reciprocal-space, because missing reflections are effectively set to zero in the map used for real-space refinement, whereas they are simply ignored by reciprocal space methods. The real-space refined model is then optimized to reproduce amplitudes of zero for reflections in the cross-validation test set. This has deleterious effects on the model, but even for models of equal quality, the agreement between test and calculated amplitudes is artifactually reduced [10]. The harmful effects are mostly mitigated by using for map calculations, approximations to the test-set amplitudes that are independent of the actual observed value. Effective approximations are the average amplitudes of working set reflections at similar resolution. This procedure almost eliminates most of the systematic discrepancy between R_{free} calculated after real- and reciprocal-space refinements.

4. Conclusion

Although most appropriate when experimental phases are accurate, real-space refinement is a useful tool for many purposes, even during the initial stages of protein refinement, when phases are usually poor. The main reason for this usefulness comes from the uncoupling of the atoms belonging to different regions of the structure: because real-space refinement is local, overfitting is reduced, as is model bias in omit maps. Methodological developments are ongoing to efficiently use real-space refinement in the early stages of macromolecular structure determination.

This local nature of real-space methods results in relatively expeditious local refinements, which makes them well-suited for interactive modeling. The tutorial on real-space methods will focus on this aspect, namely by showing how RsRef [6] is used in conjunction with O [15] for model rebuilding in various map qualities.

5. Acknowledgements & References

This work was supported by the National Science Foundation.

1. Arnold, E. and Rossmann, M.G. (1988). The Use of Molecular-Replacement Phases for the Refinement of the Human Rhinovirus 14 Structure, *Acta Crystallographica* **A44**, 270-282.

2. Blanc, E. and Chapman, M.S. (1997). RsRef: Interactive Real-Space Stereochemically Restrained Refinement during Model Building, **accepted for publication in** *Journal of Applied Crystallography*.

3. Brändén, C.I. and Jones, T.A. (1990). Between Objectivity and Subjectivity, *Nature* **343**, 687-690.

4. Brünger, A.T. (1992). *X-PLOR A System for X-ray Crystallography and NMR*, Yale University, New Haven.

5. Brünger, A.T. (1992). Free R Value: a Novel Statistical Quantity for Assessing the Quality of Crystal Structures, *Nature* **355**, 472-475.

6. Chapman, M.S. (1995). Restrained Real-Space Macromolecular Atomic Refinement using a New Resolution-Dependent Electron-Density Function, *Acta Crystallographica* **A51**, 69-80.

7. Chapman, M.S. and Blanc, E. (1997). Potential Use of Real-Space Refinement in Protein Structure Determination, *Acta Crystallographica* **D53**, 203-206.

8. Chapman, M.S. (1997). Introduction to the Use of Non-Crystallographic Symmetry in Phasing, in Fortier, S. (ed.), *NATO Advanced Study Institute on Direct Methods for Solving Macromolecular Structures*, Kluwer Academic Publishers, Dordrecht, The Netherlands

9. Chapman, M.S., Blanc, E., McKenna, R., Munshi, R., Rossmann, M.G. and Tsao, J. (1997). Use of Non-Crystallographic Symmetry for *ab initio* Phasing of Virus Structures, in Fortier, S. (ed.), *NATO Advanced Study Institute on Direct Methods for Solving Macromolecular Structures*, Kluwer, Academic Publishers, Dordrecht, The Netherlands

10. Chen, Z., Blanc, E. and Chapman, M.S. (1997). Improved free R Factors for Cross-Validation of Macromolecular Structure, **submitted to** *Acta Crystallographica*.

11. Diamond, R. (1974). Real-space Refinement of the Structure of Hen Egg-white Lysozyme, *Journal of Molecular Biology* **82**, 371-391.

12. Diamond, R. (1985). Real-Space Refinement, *Methods in Enzymology* **115**, 237-252.

13. Hendrickson, W.W. (1985). Stereochemically Restrained Refinement of Macromolecular Structures, *Methods in Enzymology* **115**, 252-270.

14. Hodel, A., Kim, S.H. and Brünger, A.T. (1992). Model Bias in Macromolecular Crystal Structures, *Acta Crystallographica* **A48**, 851-858.

15. Jones, T.A., Zou, J.Y., Cowan, S.W., Kjeldgaard, M. (1991), Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models, *Acta Crystallographica* **A47**, 110-119.

16. Rees, D.C. and Lewis, M. (1983). Incorporation of Experimental Phases in a Restrained Least-Squares Refinement, *Acta Crystallographica* **A39**, 94-97.

17. Tronrud, D.E., Ten Eyck, L.F. and Matthews, B.W. (1987). An Efficient General-Purpose Least-Squares Refinement Program for Macromolecular Structures, *Acta Crystallographica* **A43**, 489-501.

18. Watenpaugh, K. D., Sieker, L.C., Herriot, J.R. and Jensen, L.H. (1973). Refinement of the Model of a Protein: Rubredoxin at 1.5 Å Resolution, *Acta Crystallographica* **B29**, 943-956.