Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Sheldrick's 1.2 Å rule and beyond

Richard J. Morris and Gérard Bricogne*

Global Phasing Ltd, Sheraton House, Castle Park, Cambridge CB3 0AX, England

Correspondence e-mail: gb10@globalphasing.com

Received 8 November 2002
Accepted 20 January 2003

An average profile of squared normalized structure factors as a function of resolution, $\langle|E|^2\rangle(d^*)$, calculated from a large ensemble of high-resolution protein models, is presented. An interpretation is given that provides a structural explanation for Sheldrick's 1.2 Å rule for the applicability of direct methods. The implications for the potential effectiveness of extended direct methods, incorporating stereochemical knowledge, are discussed.

1. Introduction

Terms such as low, medium and high resolution have been used for a long time in crystallography and have, despite their somewhat imprecise definition, been useful in describing the level of detail in electron-density maps. One resolution limit that stands out among the rest in terms of a precise operational definition is that of atomic resolution. In 1990, Sheldrick wrote

> Experience with a large number of structures has led us to formulate the empirical rule that if fewer than half the number of theoretically measurable reflections in the range 1.1 to 1.2 Å are 'observed' [*i.e.* have $F > 4\sigma(F)$], it is very unlikely that the structure can be solved by direct methods . . . This rule simply reflects the assumption of resolved atoms, which is often invoked in direct methods.

(Sheldrick, 1990). The application of direct methods to macromolecules (Miller *et al.*, 1994; Miller & Weeks, 1998; Sheldrick, 1998; Usón & Sheldrick, 1999) has confirmed the stringency of this rule and it is now well established in the crystallographic literature. In this communication, we present a structural interpretation for Sheldrick's rule and show that it corresponds to intrinsic properties of organic molecules, especially those of proteins.

2. Wilson statistics, standard normalized structure factors and direct methods

If atoms are distributed independently and uniformly within the unit cell, the distribution of $F(\mathbf{h})$ can readily be shown to be Gaussian (Wilson, 1949, 1950). Standard normalized structure-factor amplitudes (Hauptman & Karle, 1953), defined by

$$|E(\mathbf{h})|^2 = \frac{|F(\mathbf{h})|^2}{\varepsilon_{\mathbf{h}}\sigma_2(\mathbf{h})}, \qquad (1)$$

where $\sigma_2(\mathbf{h}) = \sum f_j^2(\mathbf{h})$ and $\varepsilon_{\mathbf{h}}$ is the statistical weight of $\mathbf{h}$, have the property that $\langle|E(\mathbf{h})|^2\rangle = 1$. Deviations correspond to a departure from uniformity and/or independence, *i.e.* to structural features. $|E|$ values play a prominent role in the theory of direct methods in that the reliability of triple-phase relationships depends on the corresponding triple products of $|E(\mathbf{h})|$ values. Harker (1953), Hauptman (1965) and Main (1976) have shown how structural knowledge could be used to improve data normalization, but the subsequent derivations of phase relationships remained based on the standard random-atom model. A more radical approach was proposed by Bricogne (1994, 1995, 1997*a*,*b*), replacing the random-atom model by a random-fragment model, in which selected fragments are randomly placed according to a given probability distribution of positions and orientations. This type of approach has been nicknamed the micromolecular-replacement method (μMR). Its starting point is the multipole expansion for the transform of an arbitrary molecular fragment as

$$F(\mathbf{h}) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} b_{lm}(d^*_{\mathbf{h}})Y_{lm}(\theta_h, \varphi_h), \qquad (2)$$

which was used by Lattman (1989) to compute small-angle scattering curves for proteins. $Y_{lm}(\theta_h, \varphi_h)$ denotes the spherical harmonics as a function of the angular spherical coordinates of the reciprocal-space vector $\mathbf{h}$ and the b_{lm} values are the expansion coefficients. In the process of implementing the μMR approach (Bricogne, 1993) it was noted that these curves, normalized into $|E|^2$ profiles as a function of d^*,

$$|E|^2(d^*) = \frac{1}{\sigma_2(d^*)}\sum_{l=0}^{\infty}\sum_{m=-l}^{l}|b_{lm}|^2(d^*), \qquad (3)$$

exhibited a pronounced peak at a resolution of around 1.1–1.2 Å, with $|E|^2$ typically reaching a value of 1.3 or higher for all organic molecules examined. The universal occurrence of this peak led to the conjecture that it was associated with 'Sheldrick's rule' (Sheldrick, 1990),

although no obvious structural interpretation was at hand.

3. $|E|^2$ profiles

Theoretical $|E|^2$ profiles have since been computed according to (3) for 700 good-quality ($R_{\rm factor}$ < 0.2) high-resolution (<2.0 Å; Hooft *et al.*, 1996) protein chains from the Protein Data Bank (Bernstein *et al.*, 1977; Berman *et al.*, 2000) as well as for nucleic acid structures and a number of small molecules. A detailed analysis of their similarities and differences and of secondary-structure influence will be discussed elsewhere (manuscript in preparation). The data presented here were obtained for structures without H atoms. The inclusion of H atoms has a negligible effect of the $|E|^2$ values in the range of interest presented here; however, the departure from the equal-atom model increases and the radial pair distribution as defined above may no longer be regarded as a purely geometric term (the equal-atom structure approximation holds well for organic molecules without considering H atoms).

The correctness of the multipole expansion equation has been confirmed both in analytical terms (Stuhrmann, 1970) and also by direct comparison with computed $|E|^2$ profiles that we have obtained using Debye's formula (Debye, 1915),

$$I(d^*) = \sum_i f_i^2(d^*) + \sum_i \sum_{i\neq j} f_i(d^*) f_j(d^*) \mathrm{sinc}(2\pi d^* r^{ij}), \quad (4)$$

and again normalizing by σ_2. The sinc function is defined as $\mathrm{sinc}(x) = \sin(x)/x$.

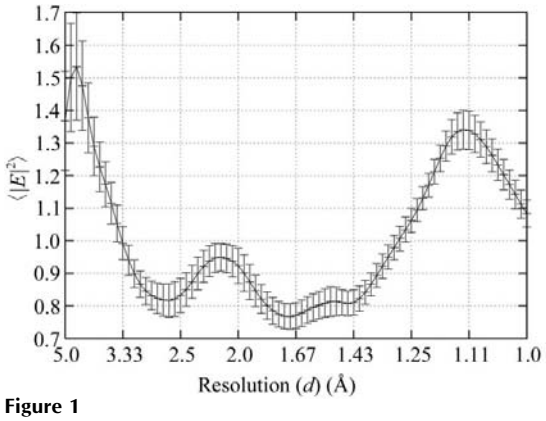

Figure 1
Averaged squared normalized structure-factor amplitudes over 700 protein structures with standard deviations calculated from the population of individual $|E|^2$ profiles.

Table 1
Approximate distances commonly observed in proteins.

Atom	Atom	Distance (Å)	Difference (Å)
C^i	O^i	1.23	
			1.15
C^i_α	N^{i+1}	2.38	
C^i	N^i	1.32	
			1.13
C^i_α	C^{i+1}	2.45	
C^i_α	C^i	1.52	
			1.08–1.18
$C^i_{\alpha,\beta}$	$C^i_{\gamma,\delta}$	2.6–2.7	
			1.11–1.21
C^i_α	C^{i+1}_α	3.81	

3.1. Sheldrick's 1.2 Å rule for direct methods

One of the most striking features of the calculated $|E|^2$ profiles is the pronounced maximum at approximately 1.1 Å (Fig. 1). The emergence of this peak can be understood by recalling that radial $|E|^2$ profiles and the radial pair distribution function are related by a sinc transformation – a spherically averaged form of the Fourier relationship between the intensity distribution and the Patterson function,

$$|E|^2(d^*) = \frac{1}{\sigma_2}\int_0^R P(r)\mathrm{sinc}(2\pi d^* r) r^2\,\mathrm{d}r \quad (5)$$

$$= 1 + \int_0^R p(r)\mathrm{sinc}(2\pi d^* r)\,\mathrm{d}r, \quad (6)$$

where R is the largest interatomic distance and $P(r)$ is the radial Patterson function. $p(r)$ is known as the radial pair distribution function (the spherically averaged origin-removed Patterson function) and for truly equal atom structures gives the number of atoms with a separation within $(r, r + dr)$. A substantial contribution to the peak around $d = 1.1$ Å simply arises from the fact that the sinc function shows a maximum in this region for typical bonding distances of about $r = 1.5$ Å, as can readily be seen by plotting $\mathrm{sinc}(2\pi d^* r)$ as a function of d. However, this peak in $|E|^2$ is greatly enhanced owing to the interference of interatomic distances that give rise to an approximately 1.1 Å repetitive structure in the radial pair distribution function. Table 1 lists some typical distances found within proteins. It can be seen that every protein contains distance beats of approximately 1.1 Å. The relevance of this observation to direct methods is as follows. After a medium-resolution peak at about 4.5 Å, $|E|^2$ values are systematically depressed below 1.0 from about 3.5 Å onwards and only at about 1.25 Å does the expectation value for $|E|^2$ start to exceed unity again, thus giving an increase in good strong $|E|$ values for direct methods. A resolution of about 1.2 Å is sufficient to reproduce a radial distance distribution with separated peaks of the above-mentioned type and to therefore provide both the atomicity and more importantly the stereochemical regularities for the successful application of direct methods.

4. Discussion

The 1.2 Å limit required for the successful application of direct methods and the definition of atomic resolution have long seemed plausible from a simple argument of the observation-to-parameter ratio (Dauter *et al.*, 1997). In this article, we have presented a structural basis for Sheldrick's 1.2 Å rule in terms of (i) typical bonding distances that directly give rise to a maximum of ∼1.2 Å through the sinc function in Debye's formula and (ii) the interference between various sinc-function contributions as a consequence of distance beats that occur in the radial pair distribution function in proteins and typical organic molecules. The incorporation of data above 1.2 Å is in effect simply introducing the structural information that is missing in the stereochemistry-free foundations of classical direct methods. Below 1.0 Å the $|E|^2$ values fluctuate closely around unity, indicating that no further major structural information is encoded in this region (see also the $|E|$ filtering method of Gilmore & Brown, 1988). The μMR approach was proposed to overcome the above limitations by producing stereochemically aware structure-factor statistics and likelihood functions and thereby significantly relaxing the data-resolution requirements. The underlying theoretical foundations of this approach have been published (Bricogne, 1995, 1997*a*,*b*) and implementation is under way.

GB acknowledges support from the Swedish NFR in the form of a Tage Erlander Guest Professorship at Uppsala University (1992–93) during which the first phase of this work was carried out. We wish to thank Eric Blanc, Pietro Roversi, Gwyndaf Evans and Marc Schiltz for detailed discussions and helpful suggestions.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
Bricogne, G. (1993). Unpublished results.
Bricogne, G. (1994). *Trans. Am. Crystallogr. Assoc.* **30**, 41.
Bricogne, G. (1995). *AIP Conference Proceedings*, Vol. 330, *ECCC1 Computational Chemistry*, edited by F. Bernardi & J. L. Rivail, pp. 449–470. Woodbury, New York: American Institute of Physics.
Bricogne, G. (1997*a*). *Methods Enzymol.* **276**, 361–423.
Bricogne, G. (1997*b*). *Methods Enzymol.* **277**, 14–18.
Dauter, A., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
Debye, P. (1915). *Ann. Phys. (Leipzig)*, **46**, 809.
Gilmore, C. J. & Brown, S. R. (1988). *Acta Cryst.* A**44**, 1018–1021.
Harker, D. (1953). *Acta Cryst.* **6**, 731–735.
Hauptman, H. (1965). *Z. Kristallogr.* **121**, 1–8.
Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem, I. The Centrosymmetric Crystal. Am. Crystallogr. Assoc. Monograph No. 3*. Dayton, Ohio: Polycrystal.
Hooft, R. W. W. & Sander, C. & Vriend, G. (1996). *J. Appl. Cryst.* **29**, 714–716.
Lattman, V. (1989). *Proteins Struct. Funct. Genet.* **5**, 149–155.
Main, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 97–105. Copenhagen: Munksgaard.
Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Miller, R. & Weeks, C. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 463–468. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.
Stuhrmann, H. B. (1970). *Acta Cryst.* A**26**, 297–306.
Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
Wilson, A. J. C. (1950). *Acta Cryst.* **3**, 258–261.