References

BUERGER, M. J. (1957). *Z. Kristallogr.* **109**, 42–60.
BUERGER, M. J. (1960). *Z. Kristallogr.* **113**, 52–56.
GRUBER, B. (1973). *Acta Cryst.* A**29**, 433–440.
GRUBER, B. (1989). *Acta Cryst.* A**45**, 123–131.
KŘIVÝ, I. & GRUBER, B. (1976). *Acta Cryst.* A**32**, 297–298.
SANTORO, A. & MIGHELL, A. D. (1970). *Acta Cryst.* A**26**, 124–127.

Acta Cryst. (1995). A**51**, 945–947

Use of globic scattering factors for protein structures at low resolution. By D. Y. GUO, G. DAVID SMITH, JANE F. GRIFFIN and DAVID A. LANGS, *Hauptman–Woodward Medical Research Institute (formerly the Medical Foundation of Buffalo), 73 High Street, Buffalo, NY* 14203, *USA*

(*Received* 9 *September* 1994; *accepted* 24 *July* 1995)

Abstract

At 3 to 4 Å resolution, the electron density of a protein may be modeled by a continuous chain of 'globs' representing the amide region of the peptide backbone and the side-chain residues. Group scattering factors are derived from a *trans* planar CαC=ONCα backbone segment and most favored side-chain conformer for 18 different amino acids. Trial calculations indicate that the phase error and crystallographic residual comparing the atomic and 'globic' models rapidly decrease from high to low resolution. At 3 Å resolution, the phase error is approximately 80°. These results indicate that the electron density of a protein composed of N amino acid residues may be adequately modeled by $2N$ globs at low resolution.

Introduction

David Harker pointed out in 1953 that 'globs', *i.e.* the total electron density of clusters of atoms within a local region of the molecular envelope, may provide a more accurate representation of the average intensity for a protein at low resolution than the sum of the square of its atomic structure-factor magnitudes, and suggested using this concept in data reduction to obtain a better scale and temperature factor (Harker, 1953). This analysis is only appropriate for macromolecules that are well separated by solvent boundaries in large unit cells. Similar efforts to obtain better scaling of the normalized E values for the *MULTAN* program make use of much smaller chemically rigid molecular fragments of unknown position and orientation (Main, 1976); the contribution of these fragments to the average intensity utilizes an expression derived by Debye (1915). Podjarny and co-workers used three group scatterers (phosphate, ribose, nucleic acid bases) to assign peaks in low-resolution MIR maps of $tRNA^{Met}$ prior to using other methods to extend phases to data at higher resolution and fill in low-order terms that were not reliably determined by the MIR process (Podjarny & Yonath, 1977; Podjarny, Schevitz & Sigler, 1981; Podjarny & Faerman, 1982).

The concept of atomicity for small structures is useful since, at atomic resolution, the atoms of the structure are recognizable in an electron-density map. For proteins, however, a more useful concept is 'globicity', which is based on the fact that 'globs', consisting of groups of atoms in the unit cell, are the only recognizable features in low-resolution electron-density maps. In the situation that one cannot confidently fit the known protein sequence to a low-resolution density map, globs may offer a viable alternative for model fitting and real-space phase improvement.

Table 1. *Glob scatterers for the common amino acid residues*

Column 4 lists the one-letter code used for simplified identification purposes, Z is the number of electrons in the chemical group and R is the residual of fit between the Debye group scattering factor and its analytical exponential form as defined by equation (4).

No.	Chemical name	Three-letter symbol	One-letter symbol	No. of Atoms	Z	R
0	Peptide	O=C<(Cα—, N—)	X	4	27	0.0013
1	Glycine	Gly	G	H atom		
2	Alanine	Ala	A	C atom		
3	Cysteine	Cys	C	2	22	0.0002
4	Serine	Ser	S	2	14	0.0009
5	Valine	Val	V	3	18	0.0006
6	Threonine	Thr	T	3	20	0.0004
7	Proline	Pro	P	3	18	0.0005
8	Isoleucine	Ile	I	4	24	0.0005
9	Leucine	Leu	L	4	24	0.0022
10	Methionine	Met	M	4	34	0.0001
11	Asparagine	Asn	N	4	27	0.0016
12	Aspartate	Asp	D	4	28	0.0013
13	Glutamine	Gln	Q	5	33	0.0003
14	Glutamate	Glu	E	5	34	0.0001
15	Lysine	Lys	K	5	31	0.0012
16	Histidine	His	H	6	38	0.0004
17	Phenylalanine	Phe	F	7	42	0.0001
18	Arginine	Arg	R	7	45	0.0014
19	Tyrosine	Tyr	Y	8	50	0.0007
20	Tryptophan	Trp	W	10	61	0.0013

At low resolution, each glob may be treated as a spherically averaged cluster as its shape will be insufficiently resolved to determine the orientation of the underlying chemical group. The group scattering factors of these spherically averaged clusters may be analytically defined as a nine-coefficient exponential expression in $\sin\theta/\lambda$ as shown previously (Cromer & Waber, 1965). Globs chosen from such a tabulation would logically correspond to the *trans* planar peptide segments in the backbone of polypeptide chain and the most favored conformations of the 20 amino acid side chains. Metal ions and ordered solvent water would still be regarded as single atoms.

A trace of the protein backbone of a polypeptide chain undoubtedly plays a major role in initial phasing, and it may be more straightforward to represent it as a strand of globs that are individual peptide segments, as compared to fitting a polyalanine model to the density. The geometry and dimensions of the peptide bond were given by Pauling, Corey & Branson (1951) and well described later (Schulz & Schirmer, 1979). For

© 1995 International Union of Crystallography
Printed in Great Britain – all rights reserved

ISSN 0108-7673
© 1995

Table 2. *Coefficients for the nine-parameter analytical fit of the Debye group scattering factors for the trans peptide and common amino acid side chains in their most favored conformations*

Values were fit over the range $0 \le \sin\theta/\lambda \le 0.2\,\text{Å}^{-1}$.

No.	Symbol	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4	c
0	X	35.19627	41.41863	0.19677	32.43376	3.21733	−7.30809	5.44058	−13.06057	−17.09161
3	C	31.16827	16.67323	4.00823	79.61819	−2.13560	33.04341	4.25983	−16.17433	−15.29927
4	S	20.72474	14.29200	1.51669	72.28514	−0.12100	67.05477	1.83400	−16.00539	−9.95547
5	V	27.48518	30.36808	0.99853	157.80148	2.81672	−5.74692	5.14800	−14.70492	−18.44238
6	T	30.34349	29.13166	0.95397	146.79330	2.69414	−5.70274	4.94526	−14.72278	−18.93233
7	P	28.78002	26.67951	1.43885	132.16763	2.02693	−4.85408	4.90488	−16.30359	−19.14580
8	I	30.92209	57.61759	−3.11301	47.54528	1.64416	−1.21085	3.60459	8.50597	−9.07199
9	L	32.46183	46.65597	−0.89377	29.85442	2.39080	−10.54924	5.44747	−13.41108	−15.45786
10	M	29.95495	78.16905	−0.71425	49.55961	1.35400	6.04091	1.79172	−1.14173	1.61178
11	N	36.73970	40.02905	0.19528	30.53796	3.94872	−8.19174	6.01751	−13.41582	−19.94956
12	D	37.24300	39.21637	0.08376	30.53387	3.92990	−6.58359	5.87636	−12.71961	−19.17064
13	Q	34.60658	65.94930	−0.36076	29.55709	1.21835	59.62092	1.65039	−14.07278	−4.11623
14	E	35.71180	64.98331	−0.28205	29.83458	1.27687	45.54034	1.71388	−12.98251	−4.42298
15	K	32.03548	105.41551	−3.01596	13.40090	0.95251	20.01938	1.73431	−25.76262	−0.68289
16	H	39.13257	67.47704	−0.11685	29.82956	1.36013	4.74581	1.69898	−1.66763	−4.10394
17	F	45.88143	80.61031	−0.70143	10.95995	1.17212	47.67025	1.56655	−17.46447	−5.92206
18	R	43.88660	141.88795	−2.02919	22.98510	1.17511	3.65697	1.81350	−11.31751	0.18658
19	Y	50.22390	111.22536	0.90052	15.24231	1.17803	9.79488	1.48135	7.94098	−3.76967
20	W	70.08553	109.98412	−8.06115	13.39345	1.04523	13.15240	2.65379	−33.13556	−4.67442

a five-atom model of a *trans* $C\alpha_iC{=}ONC\alpha_{i+1}$ peptide glob, the occupation of each $C\alpha$ is taken as 0.5. An alternative model of the same glob can be defined by four atoms, with the occupation of $C\alpha_i$ equal to 1 and $C\alpha_{i+1}$ to zero. The two models are essentially the same in the low-resolution region but the first one will give a slightly broader electron-density distribution.

Analysis

The Debye equation for the X-ray scattering from a spherically averaged rigid chemical group consisting of N atoms is

$$F(s) = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} f_i f_j \sin(4\pi r_{ij} s)/(4\pi r_{ij} s)\right]^{1/2}, \quad (1)$$

where s is $\sin\theta/\lambda$, r_{ij} is the magnitude of the distance between the ith and jth atoms whose atomic scattering factors are f_i and f_j. Following Harker, let $F_g(s)$ be the Debye scattering factor of the gth glob in the unit cell that is positioned at $\mathbf{r}_g$, such that structure factors may be expressed as

$$F_{\mathbf{h}} = \sum_{g=1}^{G} F_g(s)\exp(2\pi i\mathbf{h}\cdot\mathbf{r}_g). \quad (2)$$

The Debye scattering factors (1) can be expressed analytically as a nine-coefficient exponential function (Cromer & Waber, 1965):

$$F(s) = \sum_{i=1}^{4} a_i \exp(-b_i s^2) + c \quad (3)$$

by means of a simplex refinement procedure to fit the coefficients a_i, b_i and c. The residual minimized at each step of the refinement over the range of s is

$$R = \left(\left\{\sum_{i=1}^{n}[F_g(s_i) - F_{gc}(s_i)]^2\right\}\Big/\left[\sum F_g(s_i)^2\right]\right)^{1/2}, \quad (4)$$

where $F_g(s)$ is obtained from (1) and $F_{gc}(s)$ is the value computed from (3). The refinement is initiated by first refining a_1, b_1 and c and setting all other parameters to zero. Since at s equal to zero the sum of the four a_i and c must equal Z, the total number of electrons in the glob, a_1 may first be approximated by Z, b_1 as one and c as zero. It is best first to refine the values of a_1, b_1 and c to convergence and then extend the refinement to include a_2 and b_2 then a_3 and b_3 and finally a_4 and b_4. Table 1 lists the amino acid globs and R values from the exponential fit. The nine coefficients of the isotropic scattering factors are given in Table 2. The scattering factors for the side chains used the most favored conformation in the 'ideals' file of the *PROFFT* program (Hendrickson, 1985). The scattering-factor curves for the amino acids are shown in Fig. 1, where the one-letter amino acid code identifies each residue.

Finally, data are presented here for the 1.5 Å room-temperature crambin protein structure (Hendrickson & Teeter, 1981). This structure crystallizes in space group $P2_1$; 327

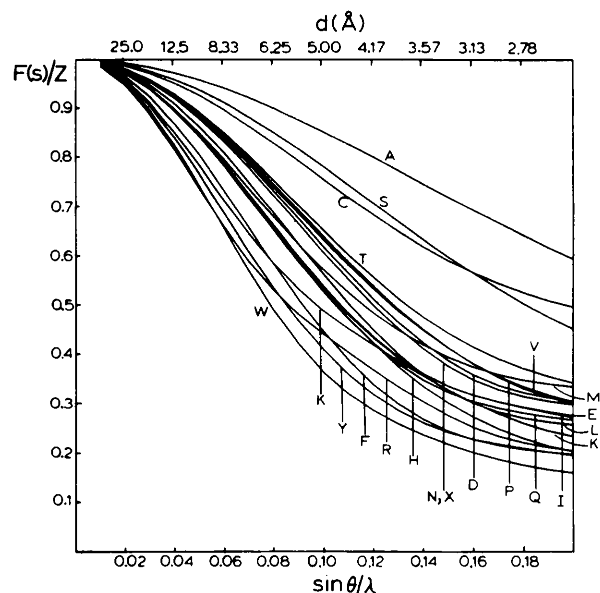

Fig. 1. Scattering-factor curves for the 20 amino acids. One-letter codes identify the residue, *cf.* Table 1. The scattering factor for a C atom is identical to that of alanine (A) shown as the uppermost curve.

non-H independent protein atoms were determined, including 202 C, 55 N, 64 O and 6 S atoms. The globic scattering factors are used to calculate phases $\varphi_{\mathbf{h}}(\text{glob})$ by means of (2). The phase error between the atomic and globic computed structure factors is

$$\Delta\varphi = \left(\sum_{\mathbf{h}} w_{\mathbf{h}} |\varphi_{\mathbf{h}}(\text{atom}) - \varphi_{\mathbf{h}}(\text{glob})|\right) \Big/ \left(\sum W_{\mathbf{h}}\right), \quad (5)$$

where $w_{\mathbf{h}} = |F_{\mathbf{h}0}|$ and $\varphi_{\mathbf{h}}(\text{atom})$ is the phase computed from the refined atomic coordinates. The average phase errors were calculated for resolution shells of 50 reflections, shown in Fig. 2. The error distribution clearly shows that the phase errors decrease rapidly when d's are longer than about 3.0 Å. A marked reduction of the globic phase error is achieved when the side-chain positions, in addition to the polypeptide backbone, have been modeled. Further reduction in phase error by modeling the ordered solvent structure is modest in comparison, *i.e.* less than 10°. Calculations performed for triclinic lysozyme, haemoglobin and erabutoxin reveal a similar pattern over the resolution range indicated in Fig. 2 (data not shown).

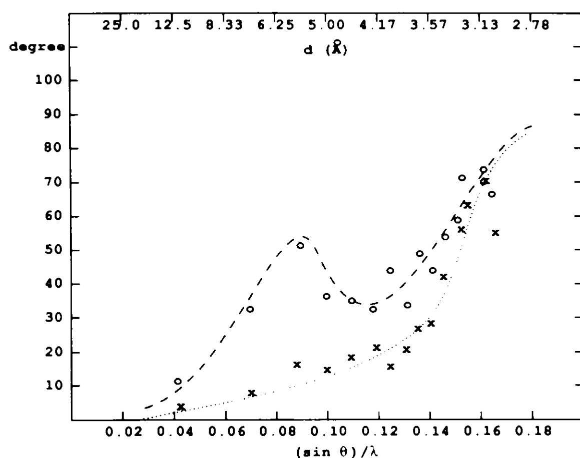

Fig. 2. Globic phase-error distribution of crambin. ····· Phasing by both main polypeptide-chain globs and side-chain globs. ----- Phasing by only polypeptide-chain globs. The peak in the latter curve at ~ 5.6 Å correlates well with the average closest separation distance between the side-chain globs of adjacent amino acid residues which were not included in the structure-factor calculation of the polypeptide-chain model.

In summary, globs may be advantageously used in low-resolution electron-density maps in which an atomic model cannot be confidently fitted.

Research support provided by NIH Grant GM46733 is gratefully acknowledged.

References

CROMER, D. T. & WABER, J. T. (1965). *Acta Cryst.* **18**, 104–109.

DEBYE, P. (1915). *Ann. Phys. (Leipzig)*, **46**, 809–832.

HARKER, D. (1953). *Acta Cryst.* **6**, 731–736.

HENDRICKSON, W. A. (1985). *Methods in Enzymology*, Vol. 115, *Diffraction Methods for Biological Macromolecules*, Part B, edited by H. W. WYCKOFF, C. H. W. HIRS & S. N. TIMASHEFF, pp. 252–270. New York: Academic Press.

HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.

MAIN, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 97–105. Copenhagen: Munksgaard.

PAULING, L., COREY, R. B. & BRANSON, H. R. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 205–211.

PODJARNY, A. D. & FAERMAN, C. (1982). *Acta Cryst.* **A38**, 401–407.

PODJARNY, A. D., SCHEVITZ, R. W. & SIGLER, P. B. (1981). *Acta Cryst.* **A37**, 662–668.

PODJARNY, A. D. & YONATH, A. (1977). *Acta Cryst.* **A33**, 655–661.

SCHULZ, G. E. & SCHIRMER, R. H. (1979). *Principles of Protein Structure*, pp. 18–19. New York: Springer-Verlag.

Acta Cryst. (1995). **A51**, 947–949

Non-magnetic twin laws. By J. SCHLESSMAN and D. B. LITVIN, *Department of Physics, The Pennsylvania State University - Berks Campus, PO Box* 7009, *Reading, PA* 19610-6009, *USA*

(*Received* 3 *February* 1995; *accepted* 4 *July* 1995)

Abstract

Twin laws are groups that express the symmetry relationships between two simultaneously observed domain states (*domain pair*) and are used to determine physical properties that can distinguish between the observed domains. A tabulation is presented of all possible non-magnetic twin laws, that is, all possible *symmetry groups* and *twinning groups* of the domain pair. Additional information is provided related to determining twin laws. This includes the coset and double-coset decomposition of point groups, the indexing and point-group symmetry of domain states, permutations of domain states, and a classification of domain states.

1. Introduction

Crystalline domains can arise in phase transitions from a high-symmetry phase of symmetry **G** to a low-symmetry phase of symmetry **F**. The bulk structures of these domains in polydomain samples are referred to as *domain states*. Two domain states have the same crystal structure and differ only in their spatial orientation. Because of this difference in spatial orientation, when simultaneously observing two domain states, the two domain states can exhibit different physical properties (see *e.g.* Litvin & Litvin, 1990). In this paper, *domain states* will refer to *single-domain states* (Janovec, Richterova & Litvin, 1993), as we do not take into account any rotations of

©1995 International Union of Crystallography
Printed in Great Britain – all rights reserved