SHORT COMMUNICATIONS

Acta Cryst. (1995). A51, 943–945

A refined algorithm for the reduced-cell determination.

By LIANG ZUO,* JACQUES MULLER, MARIE-JEANNE PHILIPPE and CLAUDE ESLING, *Laboratoire de Métallurgie des Matériaux Polycristallins, ISGMP, Université de Metz, F-57045 Metz CEDEX 01, France*

(*Received* 9 *February* 1995; *accepted* 18 *May* 1995)

Abstract

A new algorithm is proposed to convert an arbitrary primitive cell of various Bravais lattices into the reduced (Niggli) cell. The geometrical basis of this algorithm is briefly discussed.

Buerger cell

Consider a primitive unit cell U of a given lattice defined by the three non-coplanar vectors **a**, **b**, **c**, with the cell matrix

$$\begin{pmatrix} \mathbf{a}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c} \\ \mathbf{b}\cdot\mathbf{c} & \mathbf{c}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} \end{pmatrix} = \begin{pmatrix} A & B & C \\ D & E & F \end{pmatrix}. \quad (1)$$

U is called a Buerger cell if, and only if, the sum S of the lengths of **a**, **b**, **c** yields an absolute minimum value (Gruber, 1989), *i.e.*

$$S = |\mathbf{a}| + |\mathbf{b}| + |\mathbf{c}| = \text{abs min} \quad (2)$$

(the symbol abs before min indicates that all primitive cells of the lattice have to be taken into account).

Expression (2), although it is well defined, does not suggest a direct evaluation of the Buerger cell. Nevertheless, we may replace it by the following inequalities:

$$\begin{aligned} &\text{(i)} && A \le B \le C \\ &\text{(ii)} && |F| \le A/2 \\ &\text{(iii)} && |D| \le B/2 \\ &\text{(iv)} && |E| \le A/2 \\ &\text{(v)} && |D \pm E| \le (A + B \pm 2F)/2. \end{aligned} \quad (3)$$

A brief geometrical account for the conditions (i)–(v) is given as follows: the basis vectors **a**, **b**, **c** are ordered in such a way that **c** is not shorter than **b** and **b** is not shorter than **a** [(i)]; the perpendicular projection of **b** onto **a** is necessarily within the halves of ±**a** [(ii)]; the perpendicular projection of the end point of **c** on the plane containing **a** and **b** is closer to the origin O than all other lattice points on that plane, *i.e.* within the two-dimensional Voronoi domain around O [(iii)–(v)].

By means of an algorithm proposed by Buerger (1957, 1960) and refined by Gruber (1973), it is possible to transform an arbitrary primitive cell into a Buerger cell. The main feature of the algorithm is the use of an iterative process subject to the individual conditions (i)–(v). Since the sum S diminishes at every step, a Buerger cell would be reached eventually.

An alternative procedure is described here which is considered to be an immediate consequence of the above geometrical interpretation of the Buerger cell. The crucial point of this procedure is to satisfy simultaneously conditions (iii)–(v). Now let us go into detail.

We can start from any given primitive cell with the original basis vectors **a**, **b**, **c**. The basis vectors are first modified to fulfil the conditions (i) and (ii) (Gruber, 1973). In the next step, the perpendicular projection of **c** on the **ab** net plane is constructed (Fig. 1). It is essential to find the net point N closest to the tip P of the resultant vector **OP** in the net. The lattice vector **ON** needs to be subtracted from **c**. In this way, a new basis vector **c**, which is shorter than the old one, is found without altering the other two basis vectors **a** and **b**; meanwhile, it generally meets the conditions (iii)–(v).

There exist limiting cases where two, three or even four net points have respective closest distances of the same value, *i.e.* P lying on the borders of a Voronoi domain. Since such geometrical situations are essentially related to the ambiguities of the Buerger cell, any of these net points can be chosen as long as the reduced cell has not been involved. Otherwise, a unique choice is to be made by imposing additional conditions, *e.g.* Gruber's geometrical criteria used in the present reduction procedure. Furthermore, the resulting **c** may be shorter than **b**. If this occurs, the whole procedure has to be repeated until the basis vectors **a**, **b**, **c** form a Buerger cell. It is easily seen that the present procedure, besides being logically reasonable, reduces the sum S very effectively, especially when starting from a highly deformed primitive cell. Moreover, it opens a direct way to the determination of the reduced cells, as will be shown in the next section.

By convention, a general Buerger cell is normalized by means of the relations (Santoro & Mighell, 1970; Gruber,

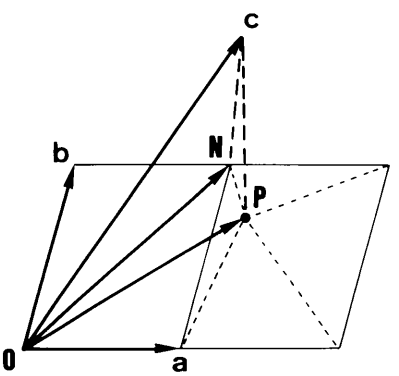

Fig. 1. Perpendicular projection of the basis vector **c** onto the net plane containing the basis vectors **a** and **b**. N is the nearest net point to the tip P of the resultant vector **OP**.

* Permanent address: Department of Materials Science and Engineering, Northeastern University, Shenyang 110006, People's Republic of China.

1973):

$$\begin{aligned} &\text{if} \quad A = B \quad \text{then} \quad |D| \le |E| \\ &\text{if} \quad B = C \quad \text{then} \quad |E| \le |F| \\ &\text{either} \quad D, E, F > 0 \\ &\text{or} \quad D, E, F \le 0. \end{aligned} \tag{4}$$

Geometrically speaking, the first two conditions make it possible to label the cell edges uniquely when two of them are equal, and the latter two to keep the interaxial angles $\alpha = \angle(\mathbf{b}, \mathbf{c}), \beta = \angle(\mathbf{c}, \mathbf{a}), \gamma = \angle(\mathbf{a}, \mathbf{b})$ all acute or all obtuse. In what follows, we shall always use this normalized description without saying it explicitly.

Niggli cell

Gruber (1989) has demonstrated that four types of reduced cell may be chosen out of the same set of Buerger cells. Basically, the four variants of reduced cells do not differ in principle from one another. Here we consider only one of them, that is, the widely used Niggli cell. Apparently, if the Buerger cell of a given lattice is unique, it must be the Niggli cell. In general, a Buerger cell is seen to be the Niggli cell if, and only if, the additional extremal condition is true for a Buerger cell with maximum deviation δ (Gruber, 1989):

$$\delta = |\pi/2 - \alpha| + |\pi/2 - \beta| + |\pi/2 - \gamma| = \text{rel max}, \tag{5}$$

where the symbol rel before max means that all possible Buerger cells have to be taken into account. With (4), δ can be written

$$\begin{aligned} \delta = {} & 3\pi/2 - \arccos\left[|D|/(BC)^{1/2}\right] - \arccos\left[|E|/(CA)^{1/2}\right] \\ & - \arccos\left[|F|/(AB)^{1/2}\right] \\ = {} & \text{rel max}. \end{aligned} \tag{6}$$

As a formal definition, the relative extremal condition (5) allows an easy geometrical interpretation on the Niggli cell. The question arises whether it is convenient to be incorporated into the main extremal condition (2). This seems rather troublesome for most existing procedures. Of course, the analytical form of the Niggli criterion may be applied as a substitute (Santoro & Mighell, 1970). In our procedure, however, the condition (5) or (6) is taken as a useful means for resolving the ambiguities of the Buerger cell, as already mentioned. To be precise, we may temporarily refer to the preceding discussion on the reduction of the basis vector **c**. Since at that step both **a** and **b** are kept fixed, the three-dimensional reduction problem reduces to a two-dimensional one. If, now, more than one net point has the smallest distance from P, only their second and third terms in (6) need to be calculated and compared so as to yield a unique solution. The same holds for the reduction of **b** while **a** and **c** remain unchanged. Note that expression (5) of the present procedure may be replaced by some other relative extremal conditions such as minimum deviation, maximum surface or minimum surface (Gruber, 1989). Again, the determination of the reduced cells is simple and straightforward.

Algorithm

A complete algorithm for determining the Niggli cell is formulated here in Gruber's (1973) notation. It has been tested for the different cases according to the 44 Niggli lattice characters. The algorithm starts from any given primitive cell with the characteristic parameters A, B, C, D, E, F. After being normalized with points G1–G4, the reduction is first made for the basis vector **b** while the other two basis vectors **a** and **c** remain unchanged. This is done with points G5–G7. Then **c** is reduced with respect to the tentatively fixed **a** and **b** through G8–G13. The most interesting fact is that both the main and the relative extremal conditions are implemented in a rather logically compact formulation (with no violation of each other). The reduction procedure will end definitely after a few iterations as is guaranteed by the property $i, j = 0$. At this point, it outputs the characteristic parameters of the Niggli cell.

Algorithm

G0 Input: A, B, C, D, E, F.

G1 If A > B or A = B, |D| > |E|, exchange A ↔ B, D ↔ E.

G2 If B > C or B = C, |E| > |F|, exchange B ↔ C, E ↔ F and go to the point G1.

G3 If DEF > 0, let D = |D|, E = |E|, F = |F|.

G4 If DEF ≤ 0, let D = −|D|, E = −|E|, F = −|F|.

G5 Let u = F/A, i = entier(u)†, u = u − i.

G6 If u > 0.5 or u = 0.5, |D − E| > |D|, let i = i + 1.

G7 If i ≠ 0, let B = B + i²A − 2iF, D = D − iE, F = F − iA and go to the point G1.

G8 Let X = AB − F², u = (BE − DF)/X, v = (AD − EF)/X, i1 = entier(u), j1 = entier(v), u = u − i1, v = v − j1, X1 = A − 2uA − 2vF, X2 = B − 2vB − 2uF, X3 = X1 + X2 + 2F, w = −1, X = amin(0, X1, X2, X3)‡.

G9 If X = 0, let i = i1, j = j1, w = Q(i, j)§.

G10 If X = X1, Q(i1 + 1, j1) ≥ w, let i = i1 + 1, j = j1, w = Q(i, j).

G11 If X = X2, Q(i1, j1 + 1) ≥ w, let i = i1, j = j1 + 1, w = Q(i, j).

G12 If X = X3, Q(i1 + 1, j1 + 1) ≥ w, let i = i1 + 1, j = j1 + 1, w = Q(i, j).

G13 If i ≠ 0 or j ≠ 0, let C = C + i²A + j²B + 2(ijF − iE − jD), D = D − iF − jB, E = E − iA − jF and go to the point G1.

G14 Output: A, B, C, D, E, F.

A comparison of the present algorithm with that by Křivý & Gruber (1976) was made through a numerical example. The results show that our reduction procedure yields a much shorter list of the intermediate results and hence converges more quickly to the solution.

LZ is grateful for the support from the National Education Committee of China and from the Ministry of Metallurgical Industry of China.

† entier(X) is an intrinsic function that returns the greatest integer whose value does not exceed the value of the argument X, *i.e.* 0 ≤ X−entier(X) < 1.

‡ amin(X1, X2, ..., Xn) is an intrinsic function that returns the minimum value among the argument list; there must be at least two arguments.

§ Q(i, j) is an external function that is defined by Q(i, j) = pq − [(1 − p²)(1 − q²)]^1/2, p = |D − iF − jB|/(BC)^1/2, q = |E − iA − jF|/(CA)^1/2.

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$<^{\text{Cα—}}_{\text{N—}}$	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