Acta Cryst. (1976). A32, 778

A Packing Function for Delimiting the Allowable Locations of Crystallized Macromolecules

BY WAYNE A. HENDRICKSON AND KEITH B. WARD

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C. 20375, *U.S.A.*

(*Received* 8 *October* 1975; *accepted* 24 *January* 1976)

A function is proposed for evaluating the likelihood of packing arrangements of macromolecules in crystals. It is based on the simple principle that the constituent molecules of a crystal should not interpenetrate. This packing function has been successfully applied in the structure solution of a hemerythrin by using the molecular shape previously determined for myohemerythrin.

An increasingly important method in protein crystallography is the use of molecular search techniques to solve crystal structures composed of molecules that are nearly isostructural with a known molecular structure. A typical procedure uses the rotation function of Rossmann & Blow (1962) to determine the orientation of the known molecule in the unknown crystal and then finds the position of the properly oriented molecule by the translation function of Tollin (1966) or Crowther & Blow (1967). One criterion frequently invoked to establish the validity of a resulting proposed structure is the reasonableness of its packing arrangement. Indeed the modes of packing available to a given macromolecule of globular shape in the lattice of a particular crystal are usually quite limited. Thus an analysis of packing could also be used *a priori* to delimit the allowable locations of crystallized macromolecules. It is the purpose of this note to describe a simple, general packing function which proved to be crucial to the solution of the structure of *Phascolopsis gouldii* hemerythrin B (Ward, Hendrickson & Klippenstein, 1975) when attempts with a conventional translation function had failed to yield the structure.

In order to deal quantitatively with the crystal packing of macromolecules, it is first necessary to describe the shape of the molecules from which the crystal is supposed to be composed. This can be done by defining a molecular shape function by

$$M(\mathbf{x})=\begin{cases} 1 \text{ if } \mathbf{x} \text{ is intramolecular} \\ 0 \text{ if } \mathbf{x} \text{ is elsewhere} . \end{cases} \quad (1)$$

Such a shape function can be determined by any of several different means. Among the possibilities are definition through an analytic functional form such as an ellipsoid, evaluation of the van der Waals envelope of an atomic model, or row-by-row delimitation of a single contiguous molecule from an electron-density map.

A second prerequisite to an analysis of packing is a definition of the transformations required to place the known molecule into the unknown crystal structure. Each independent molecule of the crystal structure must be transformed from points given by $\mathbf{x}$ in the coordinate frame of the known molecular structure to points given by $\mathbf{x}'$ in the coordinate frame of the unknown crystal. This transformation can be accomplished by the relation

$$\mathbf{x}'=\mathsf{R}\mathbf{x}+\mathbf{t} . \quad (2)$$

Here $\mathbf{t}$ is a translation vector along the three axes of the unknown crystal and R is the rotational transformation matrix which reorients the known molecule. The matrix R is in general a function of three orientation angles and the parameters of the two coordinate frames. Rossmann & Blow (1962) give the elements of R in terms of typical variables and parameters. Once a molecule has been transformed by (2), it remains to generate the other molecules related to it by the crystallographic symmetry operations of the unknown crystal. This can be done by applying the transformations,

$$\mathbf{x}_i'=\mathsf{A}_i\mathbf{x}'+\mathbf{d}_i \quad (3)$$

where A_i is the rotation matrix and $\mathbf{d}_i$ is the translation vector which relate points in the ith molecule of the crystal to equivalent ones in the unique molecule.

An evaluation of the molecular packing in a proposed crystal structure can be based on the simple principle that the constituent molecules of a crystal should not interpenetrate. Packing arrangements that minimize the intersection of the molecular spaces from the several molecules of the crystal which impinge on a given unit cell present the most probable crystal structures. Maximization of the union of molecular spaces is equivalent to minimization of the intersection. Thus the volume represented by the union of all intramolecular space in a proposed crystal structure can be taken as a measure of likelihood for that packing arrangement. With definitions of the molecular shape function and coordinate transformations already in hand, this idea can readily be expressed in functional form. For the case of a crystal with a single molecule of known structure in the asymmetric unit, a suitably normalized measure of packing likelihood is the packing function given by

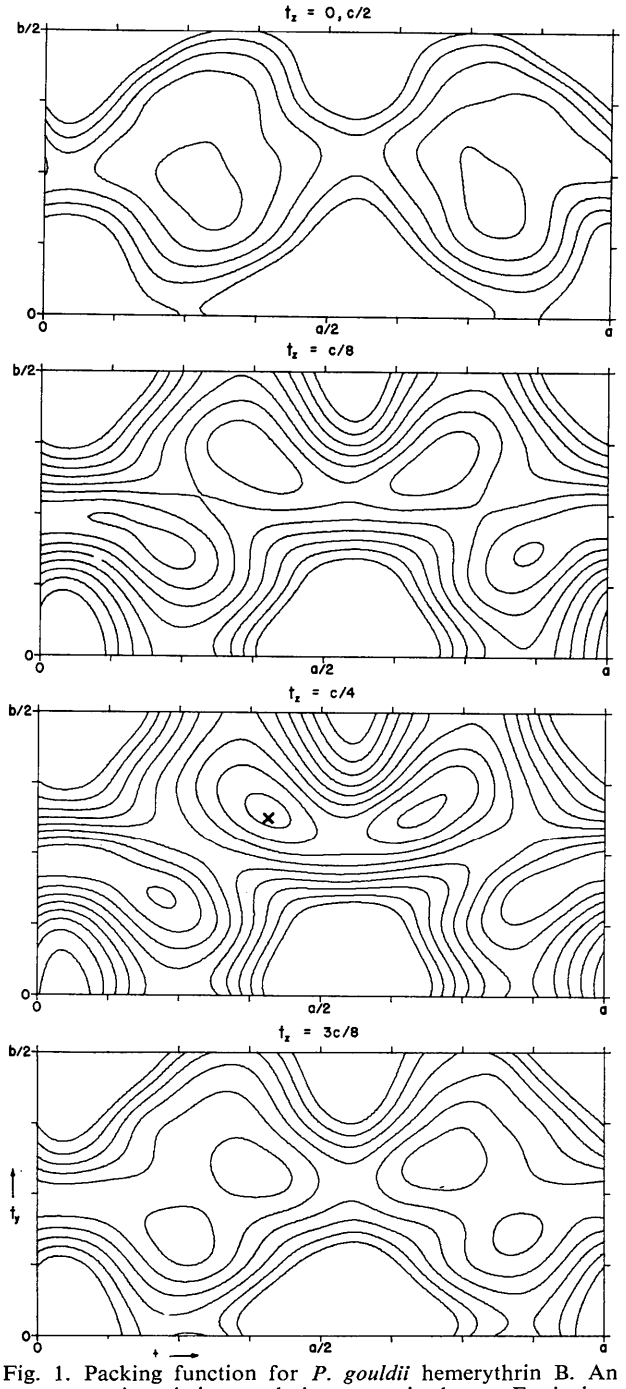

Fig. 1. Packing function for *P. gouldii* hemerythrin B. An asymmetric unit in translation space is shown. Equivalent points are related by translation between permissible crystallographic origins. In this space group, *P*422, these occur at the points of 422 symmetry which are at (0,0,0 or $\frac{1}{2}$) and ($\frac{1}{2},\frac{1}{2}$,0 or $\frac{1}{2}$). The lowest contour is drawn at 0·50 and higher contours are drawn on intervals of 0·05. The minimum value of the function on this grid ($a/16, b/16, c/8$) is 0·15 and the maximum value is 0·91. The × in the section at $c/4$ marks the projection of the true translation position (0·407, 0·316, 0·270) from the final refinement of orientation and position. This map was computed at Eulerian angles, $(\theta_1,\theta_2,\theta_3)$, of (90, −35, −172·5°) as compared to the final refinement angles of (100·3, −40·8, −177·6°).

$$P(\mathrm{R},\mathbf{t}) = \frac{1}{V_M}\iiint_{V_{au}}\left[\bigcup_{i=1}^{N_S} M_i(\mathrm{A}_i\mathbf{x}'+\mathbf{d}_i)\right]\mathrm{d}\mathbf{x}' \quad (4)$$

where

$$V_M = \iiint_{-\infty}^{\infty} M(\mathbf{x})\mathrm{d}\mathbf{x} = \iiint_{-\infty}^{\infty} M(\mathbf{x}')\mathrm{d}\mathbf{x}'. \quad (5)$$

V_M, the molecular volume of the known structure, normalizes P to a maximum possible value of 1·0. V_{au} signifies integration limits corresponding to an asymmetric unit of the crystal structure and M_i is the shape function (1) for the ith of the N_S crystallographically related molecules in the unit cell. The translation elements, $\mathbf{d}_i$, are taken so as to place all points within a single unique unit cell. The union operation gives a value of 1 to any point in the crystal which lies in the intramolecular space of one or more molecules and a value of 0 to all other points.

A possible generalization of this packing function would be to have $M(\mathbf{x})$ be other than a two-valued function. For example, it could take the values of the electron density function. An appropriate redefinition of the union operation would then be required.

Although the packing function is in principle dependent both on parameters of orientation and of translation, in usual practice it can be made a function of solely one kind of parameter. Often R will be known, *e.g.* from rotation function calculations, and P need only be a function of translations. In other cases $\mathbf{t}$ may be known, *e.g.* from the anomalous scattering location of a heavy-atom center, and P need only be a function of orientation angles. However, unless the molecular shape is highly asymmetric, the packing function is not likely to be very sensitive to orientation. In the case of crystals with more than one molecule per asymmetric unit it may be necessary to generalize P to dependence on multiple sets of orientations and translations. This can readily be done by including the added molecules within the union operation in (4) and by modifying (5) with an appropriate multiplicity factor. The computational expense in evaluating such a multiparameter function may, however, render it useless.

In practical application the packing function is evaluated numerically. First, $M(\mathbf{x})$ is generated on an appropriate grid in the coordinate frame of the known molecule. Next, the shape function for one molecule of the crystal, $M_1(\mathbf{x}')$, is found on a similar grid in the crystal coordinate frame by transformation from $M(\mathbf{x})$ according to (2). Then a third array encompassing an asymmetric unit of the crystal is filled as the union of $M_1(\mathbf{x}')$ and all symmetry-related molecules. Finally, the integrals of (4) and (5) are approximated by summations over the grids of the asymmetric unit and an isolated molecule respectively. Depending upon which kinds of parameters are to be varied in the packing function, certain parts of this procedure are then repeated at successive sample points in these variables.

The range of angles and translations that need be explored to evaluate a complete packing function may

Table 1. *Evaluation of packing function peaks for hemerythrin*

The location of packing function maxima and the peak values were determined by interpolation from the grid of the function shown in Fig. 1. Structure factors were calculated by Fourier inversion of model structures composed from the electron density of isolated myohemerythrin molecules appropriately oriented and positioned and adjusted to the proper solvent density level. The orientation at positions *A*–*E* was at Eulerian angles of (90, −35, −172·5°) whereas the orientation at the 'Final' position was (100·3, −40·8, −177·6°). The *R* value given is simply $R=\sum|F_o-kF_c|/\sum F_o$ where k is the least-squares scale factor relating the calculated structure factor amplitudes, F_c, to the observed amplitudes, F_o. There are 36 reflections in the d spacing range of $\infty > d > 20$ Å and 184 reflections in the $20 > d > 10$ Å range.

Position	t_x	t_y	t_z	Packing function value	*R* value $\infty > d > 20$ Å	*R* value $20 > d > 10$ Å
Final	0·407	0·316	0·270		0·28	0·34
A	0·405	0·328	0·236	0·92	0·31	0·56
B	0·675	0·327	0·234	0·91	0·45	0·65
C	0·220	0·170	0·213	0·91	0·51	0·70
D	0·885	0·181	0·213	0·90	0·33	0·61
E	0·055	0·236	0·208	0·90	0·50	0·64

depend both on the crystal symmetry and on the molecular symmetry. The asymmetric unit in translation space alone depends on the location of the permissible origins in a given space group and may be greater than the crystallographic asymmetric unit. The asymmetric unit in rotation space alone will depend on the symmetry of the search molecule. The range in parameters for a combined rotational and translational search could be more restricted than for the asymmetric units of the separate searches.

In the packing function application to hemerythrin, only translation parameters were treated as variables. The fast rotation function of Crowther (1972) had been used to determine approximate Eulerian angles for orienting an isolated myohemerythrin molecule (Hendrickson, Klippenstein & Ward, 1975) in the tetragonal *P*422 cell of hemerythrin. The asymmetric unit of these crystals includes one protomer from each of two hemerythrin octamers. Thus a version of (4) which had been modified as above to accommodate two molecules per asymmetric unit was applied. However, in this case, the packing function remained a function of only three variables since an exact relationship between the two crystallographically independent protomers is determined by a special element of local symmetry (Ward, Hendrickson & Klippenstein, 1975). The shape function for myohemerythrin was determined by manual isolation of one molecule from the 5·5 Å resolution electron-density map. This was done on a grid of $1\cdot75\times1\cdot67\times1\cdot57$ Å with the origin at a grid point near the dimeric iron center. The transformed shape function was evaluated on a similar grid ($a/64$, $c/32$) in the hemerythrin cell of $a=104\cdot82$, $c=54\cdot08$ Å. A relatively coarse grid ($a/16$, $c/8$) sufficed for the packing function itself which is of fairly low resolution. The entire packing function is shown in Fig. 1. Each of the five local maxima in this translation map was tested by structure factor calculations for low-order reflections as reported in Table 1. The packing function peak which was thus identified as the most likely position, *A*, eventually led to the solution of the structure. Peaks *B*, *C* and *D* correspond to false solutions wherein the molecules are centered near correct positions but are wrongly oriented. Peak *E* corresponds to the incorrect quaternary arrangement suggested by a speculation based on the myohemerythrin structure (Hendrickson, Klippenstein & Ward, 1975).

We thank Jerome Karle for his critical reading of the manuscript. K.B.W. is a recipient of an N.R.C. Resident Research Associateship.

References

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, p. 174–177. New York: Gordon and Breach.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

HENDRICKSON, W. A., KLIPPENSTEIN, G. L. & WARD, K. B. (1975). *Proc. Natl. Acad. Sci. U.S.* **72**, 2160–2164.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

TOLLIN, P. (1966). *Acta Cryst.* **21**, 613–614.

WARD, K. B., HENDRICKSON, W. A. & KLIPPENSTEIN, G. L. (1975). *Nature, Lond.* **257**, 818–821.