Acta Cryst. (1988). A44, 226-230

Improving the Interpretation of Translation Functions by a Simple Map-Correlation Procedure

BY C. C. WILSON

Neutron Division, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, England

AND P. TOLLIN

Carnegie Laboratory of Physics, University of Dundee, Dundee DD1 4HN, Scotland

(*Received* 3 *April* 1987; *accepted* 28 *October* 1987)

Abstract

A procedure for improving the reliability of the Q (translation) functions is presented. The procedure involves correlating between maps calculated using different sections of the reflection data, these being spherical shells divided according to d spacing. The peaks in the Q function representing the true shifts of the fragment are found to be the most stable under such a correlation procedure. The modification has been incorporated into a computer program.

Introduction

Traditionally the weakest part of Patterson-function interpretation procedures for crystal structure solution has been the location of oriented fragments in the cell. Much effort has been put into improving the reliability of translation functions and several techniques have been employed to do so (Karle, 1972; Langs, 1975; Doesburg & Beurskens, 1983; Bruins Slot & Beurskens, 1984; Egert, 1983).

In our own Patterson-method routines (Wilson & Tollin, 1985, 1986) we use the Q functions of Tollin (Tollin & Cochran, 1964; Tollin, 1966) and perform the calculations in reciprocal space. These functions find the location of an oriented fragment with respect to symmetry elements individually. For certain space groups, therefore, there is a built-in degeneracy in the Q-function determination, *e.g.* in $P2_12_12_1$ three Q functions determine six coordinates and hence each shift is located with twofold degeneracy.

This degeneracy can be exploited in the interpretation of the maps by using cross comparison between them. Often such an interpretive procedure can eliminate ambiguities in the Q-function calculations. However, the intrinsic problems of translation functions mean that even such a procedure can give incorrect answers, owing to the coincidence of spurious peaks from each of the maps. Further, if no such degeneracy exists, there is no means to check the results indicated by a Q function other than by trial and error of the resulting model in Fourier or tangent-recycling procedures.

We have found that in unfavourable circumstances the correct translational shifts in a Q-function calculation may not correspond to the highest peak in the map. Indeed, the solution representing the correct shifts can sometimes be well down the peak list. Even in degenerate cases this situation can arise, as will be shown below.

In this paper we discuss ways of improving the interpretation of Q maps, thus reducing the possibility of an incorrect translation being indicated and used in an attempted structural solution.

'Improving' Q maps

As was noticed by Karle (1972) and ourselves (Wilson & Tollin, 1985), some improvement is often gained in the calculation of translation functions (in our case the Q functions) when only the outer half [higher $(\sin\theta)/\lambda$, smaller d spacing] of the data, or some outer portion, is used in the calculations. Such data will be referred to as 'cut-off' data. The effect on the appearance of a Q map resulting from the use of cut-off data will now be discussed.

A typical feature of the Q functions is the appearance of bands of density in the map, regions of considerable linear extent where positive density is found (Fig. 1*a*). These are not universal, but are very common, especially when the model being used is small in relation to the asymmetric unit. The peaks in the Q map representing the required shifts arise from this general plateau region.

However, when cut-off data are used, these bands tend to break up and to become narrower (Fig. 1*b*). In the new map the peaks will be more likely to appear as islands in a sea of negative density, and are thus qualitatively more obvious. In this sense, the maps calculated with cut-off data are improved. A rationale for this improvement is now given.

The Q functions (Tollin & Cochran, 1964; Tollin, 1966) are defined as

$$Q(\mathbf{R}_0) = \sum_{\mathbf{h}} |F_s(\mathbf{h})|^2 \sum_{j,j'=1}^{n} \cos 2\pi \mathbf{h} . [\mathbf{r}_j + \mathbf{R}_0 - T(\mathbf{r}_{j'} + \mathbf{R}_0)],$$

0108-7673/88/020226-05$03.00

where n is the number of atoms in the model, $F_s(\mathbf{h})$ is the sharpened structure factor, and $\mathbf{r}_j$ are the coordinates of the model. $T(\mathbf{R})$ is the symmetry operator for the particular symmetry element under investigation.

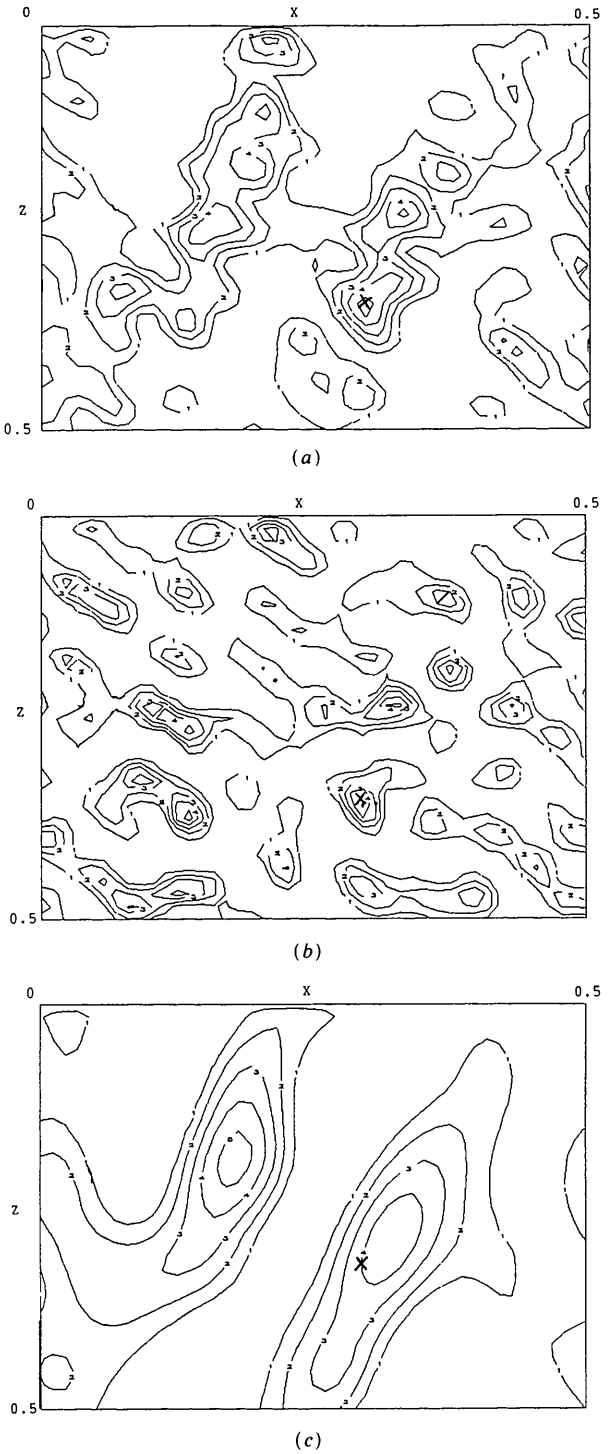

Fig. 1. The $Q(X_0, Z_0)$ function for 2',3',5'-tri-*O*-acetyladenosine (TAA) calculated using different sections of the data. (*a*) All data used. (*b*) Data with $d < 1{\cdot}5$ Å used. (*c*) Data with $d > 2{\cdot}0$ Å used.

The Q function can thus be viewed as a Patterson sum function whose 'frequency' components depend on the interatomic vectors. This function should have peaks when the $\mathbf{r}_j$, $\mathbf{r}_{j'}$ are related by

$$\mathbf{r}_j = T(\mathbf{r}_{j'} + \mathbf{R}_0) - \mathbf{R}_0.$$

As was shown by Tollin (1966) the equation is evaluated for each symmetry element $T(x, y, z)$ separately. The frequently found bands of density mentioned, which increase the noise and unreliability of the Q functions, can be explained in terms of the above Patterson summation. If different spherical shells of data are used in the summation the contribution of each successive shell represents higher-frequency contributions to the maps, *i.e.* the very longest d spacings given the lowest-frequency components, and thus the early terms in the summation give a general envelope of density in the calculated Q map (Fig. 1*c*), whereas the later terms, corresponding to shorter d-spacing data, give higher-frequency variations.

Elimination of areas of broad positivity in the Q map by use of just some outer section of the data should leave the detailed information more clearly indicated, and should reduce the number of spurious peaks produced. However, because the low-frequency components do contain useful information in addition to the above possibly spurious detail, it is possible that the map calculated from cut-off data, while much clearer and easier to interpret, may have the relative heights of the peaks distorted. That is, it is possible for the relative intensity of the 'true' peak to reduce when a data cut-off is applied, as was observed by Karle (1972) and in the TAA example below. The true peak should remain prominent in such cases but this height variation is likely to be the reason for the 'outer half' improvement not being universal. The following general statements can be made about the calculation of a Q function from cut-off data: (i) the *quality* of the map is likely to improve (increased signal/noise, less spurious peaks); but (ii) the *accuracy* may diminish (relative peak heights). It should be noted that in many cases the accuracy is improved as well.

A correlation procedure

As a consequence of (ii) above, the only safe way to exploit the improvement gained by using cut-off data is to correlate between maps calculated with different sections of the data. This relies on the fact that spurious features will come and go but the true peak should be relatively stable. 'Stable' in this context means that the peak will remain close to the same position, *and will be a peak*, in all maps. No constraint should be imposed [again because of (ii)] on the

height of the peak in each individual map other than that this should remain significant.

Such a correlation procedure should eliminate most of the possible solutions by (i), and despite (ii) the true peak would be expected to have a high summed value over all the maps, since it should be the most consistently indicated. It would be hoped that any remaining spurious peaks would be of lower intensity, and certainly that the hierarchy of possible solutions should be more reliable than that from just one map.

As the correlation has been applied in the current procedure, three maps have been calculated and correlated. These are respectively the full Q function (no restriction on d_{max}), a set restricted by a d_{max} chosen according to the resolution of the data, and a third map with d_{max} at an intermediate value. In all maps $d_{min} = 0$ and thus the high-resolution data are always used.

The procedure of selecting peaks in each map and then correlating between the peaks found has advantages over performing a straight summation. The latter procedure does not discriminate against regions in a map where no peak is present, and gives too high a weight to general regions of positive density which may be considerably removed from a peak position. Thus spurious details are eliminated more convincingly using the peak-correlation method chosen.

On a practical note, the peak position is allowed to shift by $\simeq 0{\cdot}3$ Å before the map correlation fails. This allows for sampling errors (the grid normally used is at about 0·1 Å spacing) and small fluctuations between maps.

In addition, following the example of Beurskens, Gould, Bruins Slot & Bosman (1987) in three dimensions, we calculate and eliminate excluded regions in the 2D Q maps. These correspond to shifts which would lead to unacceptable clashes (< about 2 Å) of the search fragment with a symmetry equivalent. As can be seen in Fig. 2, which shows the $Q(X_0, Z_0)$ function for seal myoglobin (Scouloudi, 1969; Scouloudi & Baker, 1978; Tollin, 1976), often these excluded regions can be very substantial indeed. In the map shown in Fig. 2, only the regions *within* the heavy lines represent allowed shifts for the fragment. Thus it can be seen that the number of possible solutions is severely restricted in this case. The true peak is indicated by the cross (×).

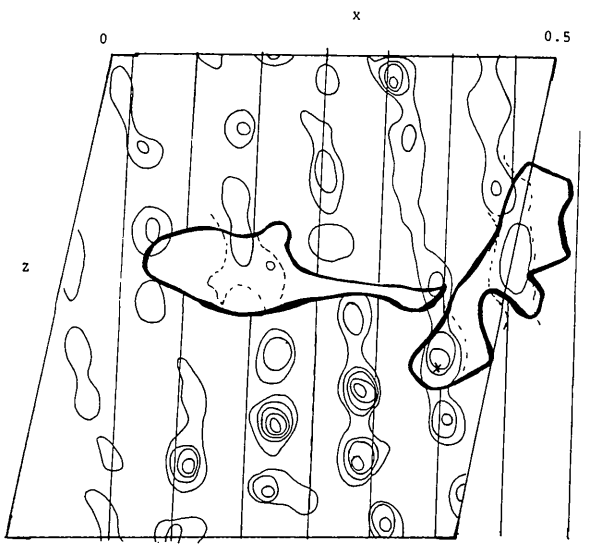

Fig. 2. Seal myoglobin $Q(X_0, Z_0)$ function showing excluded regions. Only the areas inside the bold lines are possible solutions for the Q functions. × marks the correct solution.

In the examples studied the excluded regions are taken into account in all maps calculated.

Examples

The results of several tests of the procedure outlined are shown in Table 1. The improvements rendered by the suggested map-correlation technique can be seen quite clearly. The following points should be noted about the nature of the improvements.

(i) There is a reduction (often substantial) in the total number of possible solutions.

(ii) There are improvements (at times dramatic) in the results for individual maps, where the map-correlation procedure often moves the correct solution substantially higher in the list of possible solutions.

(iii) There are improvements in the 'discrimination' of the solutions in many cases, *i.e.* the ratio of the correct solution value to that of the highest incorrect solution.

(iv) For TAU (example 3), which has the degeneracy discussed above, even in the combined map $[Q(X_0, Y_0, Z_0)]$ the uncorrelated procedure can rank the true solution some way down the list of possibles (*e.g.* at $N = 75$ this was only ranked fifth). The correlated procedure gives the true solution as first in all cases.

(v) For TAA (example 1) there is a relative worsening of correlated results as N increases to >200 – owing to the use of very weak reflections as part of cut-off data sets. The nature of the procedure means that the data in such cases may be 'oversharpened' for the very high $(\sin\theta)/\lambda$ reflections which need to be used in this case, since sharpened $|F_s(\mathbf{h})|^2$ values are normally employed in our methods. This overemphasis of possibly weak high-angle reflections can cause problems even if $|E_\mathbf{h}|$ values are used. An ideal procedure would reduce the number of reflections used as more severe cut-off criteria are introduced, and this can easily be chosen as an option.

While the improvements in the interpretation of the individual maps are important, it is the improvement in the final combined solution which is more significant, implying that results obtained from translation functions employing the new procedure should be more reliable.

Table 1. *Improvements gained by using the correlation procedure outlined on four test structures (results from PATMET)*

Excluded regions: atoms approaching to within 2 Å.

Number of reflections (N)	Position of correct peak (without/with correlation) $Q(Y_0, Z_0)$	$Q(X_0, Z_0)$	$Q(X_0, Y_0)$	Number of possible solutions	Position of correct full solution $[Q(X_0, Y_0, Z_0)]$	Relative heights (correct: highest incorrect)
Example 1. TAA ($C_{16}H_{19}N_5O_7$); search fragment C_7N_4; $p^2 = 0{\cdot}37$, $p_1^2 = 0{\cdot}092$. Data sections used in calculations: *A* full data, $B\,d < 2$ Å, $C\,d < 1{\cdot}75$ Å						
75	1/1	2/1	11/1	17/1	1/1	240:216/–
125	1/1	1/1	11/2	19/2	1/1	247:243/740:557
175	1/1	1/1	7/2	21/1	1/1	247:226/–
200	1/1	1/1	6/3	27/1	1/1	248:221/–
300	1/2	1/1	7/3	17/4	1/1	242:221/683:574
Example 2. DAZA ($C_5H_5NO_2$); search fragment C_5NO_2; $p^2 = 0{\cdot}96$, $p_1^2 = 0{\cdot}24$. Data sections used in calculations: *A* full data, $B\,d <$ 1·75 Å, $C\,d < 1{\cdot}4$ Å						
50	1/1	1/1	1/1	30/19	1/1	277:211/772:609
100	1/1	1/1	1/1	22/9	1/1	256:183/733:532
Example 3. TAU ($C_{15}H_{18}N_2O_9$); search fragment $C_4N_2O_2$; $p^2 = 0{\cdot}29$, $p_1^2 = 0{\cdot}073$. Data sections used in calculations: *A* full data, $B\,d < 1{\cdot}75$ Å, $C\,d < 1{\cdot}5$ Å						
75	7/3	7/1	7/1	16/2	5(*)/1	198:231/682:608
150	4/1	3/1	5/1	18/4	1/1	222:215/696:572
200	5/1	4/1	7/1	16/4	1/1	206:198/678:554
300	4/1	2/1	3/1	16/4	1/1	210:196/682:526
500	4/1	2/1	1/1	11/3	1/1	229:184/690:533
Example 4. TAG ($C_{16}H_{19}N_5O_8$); search fragment C_7N_4; $p^2 = 0{\cdot}34$, $p_1^2 = 0{\cdot}085$. Data sections used in calculations: *A* full data, $B\,d < 1{\cdot}75$ Å, $C\,d < 1{\cdot}5$ Å						
75	2(*)/1					85:89/283:221
150	1/1					91:90/289:222
200	1/1					89:88/287:223

References for test structures:
1. 2',3',5'-Tri-*O*-acetyladenosine. Wilson, Tollin & Howie (1986).
2. 3-Deazauracil. Low & Wilson (1983).
3. 2',3',5'-Tri-*O*-acetyluridine. Low & Wilson (1984).
4. 2',3',5',Tri-*O*-acetylguanosine. Wilson, Low & Tollin (1985).

* Cases where the ranking of the correct solution in the *overall* map has been improved.

Of course, if there is no combination of maps to be made [situations where there is no degeneracy such as in $P2_1$ (see example 4, TAG)] the individual map improvement then becomes equally important and the above results give confidence that these will also be more reliable. In addition, in cases where only a limited data set is available (small N), this procedure can be expected to be of great assistance in the interpretation of Q maps, on the evidence of those structures in Table 1.

Finally, it should be noted that the search fragments used in these translation-function examples are often small, typically consisting of <10% of the total scattering in the cell (p_1^2), which is less than the fraction required in many other translation-function procedures (Beurskens, Gould, Bruins Slot & Bosman, 1987). The combination of being able to use few reflections and a small search fragment renders the procedure outlined very useful in cases where only poor data are available, or when only a small part of the stereochemistry of the compound is known with sufficient reliability to form a model on which to base a translation search procedure.

Concluding remarks

It is hoped that the above map-correlation procedure will render more reliable the interpretation of Q-function results and lead to less time being spent (i) trying incorrect translational shifts in a trial-and-error Fourier recycling procedure or (ii) pursuing an incorrectly defined or oriented model on the assumption that the translation-function results are incorrect. This procedure has been incorporated as an option into the Patterson-methods program *PATMET* (Wilson & Tollin, 1985, 1986).

References

BEURSKENS, P. T., GOULD, R. O., BRUINS SLOT, H. J. & BOSMAN, W. P. (1987). *Z. Kristallogr.* In the press.

BRUINS SLOT, H. J. & BEURSKENS, P. T. (1984). *Acta Cryst.* A**40**, 701-704.

DOESBURG, H. M. & BEURSKENS, P. T. (1983). *Acta Cryst.* A**39**, 368-376.

EGERT, E. (1983). *Acta Cryst.* A**39**, 936-940.

KARLE, J. (1972). *Acta Cryst.* B**28**, 820-824.

LANGS, D. A. (1975). *Acta Cryst.* A**31**, 543-550.

LOW, J. N. & WILSON, C. C. (1983). *Acta Cryst.* C**39**, 1688-1690.

LOW, J. N. & WILSON, C. C. (1984). *Acta Cryst.* C**40**, 1030-1032.
SCOULOUDI, H. (1969). Personal communication.
SCOULOUDI, H. & BAKER, E. N. (1978). *J. Mol. Biol.* **126**, 637-660.
TOLLIN, P. (1966). *Acta Cryst.* **21**, 613-614.
TOLLIN, P. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML & B. SEDLACEK, pp. 212-221. Copenhagen: Munksgaard.
TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322-1324.
WILSON, C. C., LOW, J. N. & TOLLIN, P. (1985). *Acta Cryst.* C**41**, 1123-1125.
WILSON, C. C. & TOLLIN, P. (1985). *PATMET*. Program for determination of orientation and position of a known fragment in the unit cell. Univ. of Dundee, Scotland.
WILSON, C. C. & TOLLIN, P. (1986). *J. Appl. Cryst.* **19**, 411-412.
WILSON, C. C., TOLLIN, P. & HOWIE, R. A. (1986). *Acta Cryst.* C**42**, 697-700.

International Union of Crystallography

Acta Cryst. (1988). A**44**, 230

Commission on Powder Diffraction

At its 14 August 1987 meeting in Perth, the General Assembly of the IUCr established a Commission on Powder Diffraction. This is an action widely welcomed and considered by many to be long overdue.

In early 1986, the IUCr Executive Committee established an *ad hoc* committee to assess world-wide interest and, if it be sufficient, to prepare specific proposals for the General Assembly to consider in determining whether to establish a Commission on Powder Diffraction. The Terms of Reference proposed by the committee and the Executive Committee and accepted by the General Assembly are:

i. To advice the IUCr in organizing or sponsoring meetings, schools and Congress sessions on powder diffraction and related subjects.

ii. To promote and coordinate scientific exchange between countries in the field of powder diffraction.

iii. To cooperate with other IUCr Commissions on matters concerning powder diffraction.

iv. To cooperate with other international bodies interested in powder diffraction and allied subjects.

v. To promote useful interactions of the IUCr with the large world-wide body of X-ray and neutron powder diffractionists.

iv. To promote the scientific growth and development of the field of powder diffraction.

Since the members of the *ad hoc* committee became the initial members of the Commission, they were able to meet three times during the Congress to make plans for Commission projects. Among those being given first consideration are

Program exchange 'bank'.

Satellite meeting for the 1-990 IUCr Congress at Bordeaux.

Workshop on the Rietveld method (to be held between August 1988 and August 1989).

Newsletter.

Round-robin with the Rietveld method involving both X-ray and neutron data and several samples.

New book(s), possibly resulting from workshop(s).

Powder diffraction sessions at 1990 IUCr Congress at Bordeaux.

The members of the new Commission on Powder Diffraction are:

R. A. Young	USA (Chairman)
Z. Bojarski	Poland
R. J. Hill	Australia
A. W. Hewat	France
J. I. Langford	UK (Secretary)
P.-E. Werner	Sweden
T. Yamanaka	Japan.

In addition, Dr L. Frevel serves as the JCPDS-appointed representative to the Commission.

Acta Cryst. (1988). A**44**, 230

Crystallographic Databases

Crystallographic Databases - Information Content, Software Systems, Scientific Applications (221 pp., A5 format), edited by F. H. Allen, G. Bergerhoff and R. Sievers, has recently been published by the Data Commission of the International Union of Crystallography. It contains descriptions of the Cambridge Structural Database (organic and metal-organic compounds), the Inorganic Crystal Structure Database (ICSD), the Metals Crystallographic Data File, the Protein Data Bank, the JCPDS Power Diffraction File, the Database of Order-Disorder Structures, and the NBS Biological Macromolecule Crystallization Database.

The various software systems currently available are also considered: CSD (Cambridge), CRYSTIN (ICSD), NBS*SEARCH, JCPDS Search/Match, Chemical Databank System (SERC, UK), and the Canadian Scientific Numeric Database Service (CAN/SND; NRC Ottawa).

Additionally there is an overview of crystal structure analysis for non-specialists, a survey of printed information sources in crystallography, and a bibliography of scientific applications.

This is the first publication to be devoted to this important new area of crystallography.

Personal orders (US$20.00, £12.50, Swiss Fr30, Netherlands Guilders 40; payment with order, air-mail postage included) should be addressed to Dr J. N. King, Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. [Institutional order forms will be fulfilled and invoiced at an additional cost of US$5.00 or the equivalent.]