Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

(*Received 5 July* 1979; *accepted 15 October* 1979)

Abstract

A method is described for deriving three-dimensional coordinates from stereodiagrams of molecular architecture. The accuracy of the method is tested for cytochrome b_5 (86 C_α atoms) and tomato bushy stunt virus (311 C_α atoms). The coordinates were reconstructed to 1·9 Å and 2·6 Å r.m.s. deviation of their original values, respectively. The ethics of the procedure are discussed.

Introduction

The publication and illustration of molecular detail often takes the form of ball-and-stick representations in stereodiagrams. This practice has been greatly aided by Johnson's (1970) *ORTEP* program. These stereodiagrams are frequently published by protein and nucleic acid crystallographers long before the actual coordinates are publicly distributed. This practice has created an unusual situation where scientists publicly present their results, but do not attempt to provide sufficient information for others to use quantitatively their 'published data'.

In this paper a technique is described with which the three-dimensional coordinates can be determined from stereodiagrams. This is followed by a discussion of the ethics of non-publication of relevant data and the use of the present technique to extract the missing information.

The technique

A stereodrawing consists of two projections of a three-dimensional object on to a plane. The object is viewed from a given distance but is rotated by a small angle $+\varphi$ and $-\varphi$ to create the left and right projections. The viewing distance, v, is normally about 20 in (~0·5 m) or at infinity. The total angular separation 2φ is usually about 5°, which is the average angle subtended by the eyes at the normal focal plane.

Let an atom of the object be at position x,y,z relative to a Cartesian axial frame.

Let the object be rotated by $\pm\varphi$ about the y axis and viewed along z (Fig. 1).

Let the coordinates of the projected atom be at x_L,y_L and x_R,y_R in the left and right stereodiagrams, respectively.

If the viewing distance is at infinity, then

$$x_L = x\cos\varphi + z\sin\varphi,$$

$$y_L = y,$$

and

$$x_R = x\cos\varphi - z\sin\varphi,$$

$$y_R = y.$$

It follows that

$$\left.\begin{aligned} x &= \frac{x_L + x_R}{2\cos\varphi}, \\ y &= \frac{y_L + y_R}{2}, \\ z &= \frac{x_L - x_R}{2\sin\varphi}. \end{aligned}\right\} \quad (1)$$

If, however, the viewing distance is not at infinity, an object which has a length d and is in front of the projection plane will appear to have a larger length D (Fig. 2) within the plane, where

$$D = \frac{d}{1 - \frac{z}{v}}. \quad (2)$$

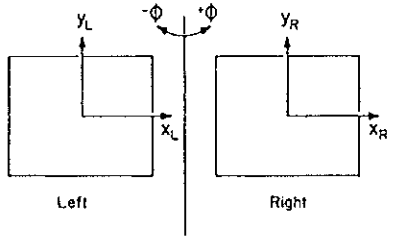

Fig. 1. Definition of coordinates in viewing a stereodiagram. The z axis is perpendicular to the page.

0567-7408/80/040819-05$01.00

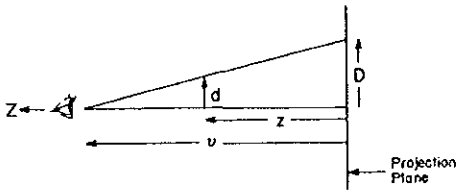

Fig. 2. An object of length d at distance z in front of the projection plane when viewed at a distance v from this plane will appear to have a projected length D where $D = d/[1 - (z/v)]$.

Since x_L,y_L,x_R and y_R are measured projected lengths on the stereodiagram, their values must be corrected with the expression (2). Hence, combining (1) and (2), it is clear that

$$x = \frac{x_L + x_R}{2\cos\varphi}\left(1 - \frac{z}{v}\right),$$

$$y = \frac{y_L + y_R}{2}\left(1 - \frac{z}{v}\right),$$

$$z = \frac{x_L - x_R}{2\sin\varphi}\left(1 - \frac{z}{v}\right).$$

By solving these equations it will be found that

$$\left.\begin{aligned} x &= \frac{x_L + x_R}{2\cos\varphi} \times q, \\ y &= \frac{y_L + y_R}{2} \times q, \\ z &= \frac{x_L - x_R}{2\sin\varphi} \times q, \end{aligned}\right\} \qquad (3)$$

where

$$q = \frac{1}{1 + \left(\dfrac{x_L - x_R}{2v\sin\varphi}\right)}.$$

As v tends to infinity, q will approach unity and equations (3) and (1) become equivalent. Furthermore, as v approaches $-[(x_L - x_R)/2\sin\varphi]$ an instability is reached since the atom will then project at infinity on the viewing plane.

The expressions (3) can, however, be used only with a knowledge of φ and v, parameters rarely supplied with stereodiagrams. Accordingly some criterion is necessary to determine these parameters; that is, the depth measurements (along z) must be correlated to their horizontal (along x and y) counterparts. A reasonable criterion used for analyzing stereodiagrams of polypeptide backbones is that all C_α–C_α distances are of equal length. Additional criteria could use well known constants of such objects as the α-helix or β-pleated sheet; for example, the distance between every fourth, eighth, *etc.* C_α atom within an α-helix. The longer the depth measurements, the greater will be the accuracy of the determination of φ and v, although the C_α–C_α criterion has been found to be sufficient.

The criterion can be stated as requiring the minimum value of

$$E = \sum_{i=1}^{N} (kR_i - S_i)^2, \qquad (4)$$

where S_i is the anticipated distance between two atoms (say C_α–C_α = 3·84 Å), R_i is the distance derived from the stereodiagram (equations 3) in terms of the measurement units, and k is a scale factor which relates the units used in R to those used for S. The length of each R_i will depend on the selected values of φ and v. Thus, a two-dimensional search must be made for the minimum of E in terms of these variables. Likely search limits are $1{\cdot}5° \leq \varphi \leq 6°$ and $10'' \leq v \leq 50''$, with suitable steps of $\Delta\varphi = 0{\cdot}25°$ and $\Delta v = 5''$.

The best value of k for a given φ and v is given when $\partial E/\partial k = 0$, that is,

$$k = \frac{\sum_N RS}{\sum_N R^2},$$

which can be used to evaluate E.

Once φ and v have been found, the atomic coordinates can be calculated from expressions (3). However, the values of z will be subject to a good deal of error, as they depend on the differences $(x_L - x_R)$. In contrast, x and y are essentially averages of the left and right measurements. The accuracy in z can be regained to some extent by invoking the same criteria as used in the determination of φ and v. An iterative least-squares procedure can be set up to adjust the coordinates so as to equalize all C_α–C_α distances.

There may be a residual systematic error if the stereo angle φ has been incorrectly estimated. If φ_C and φ_E are the correct and estimated angles, respectively, then it can be shown that the z parameters will be in systematic error according to the ratio of $(\sin\varphi_C/\sin\varphi_E)$ or approximately φ_C/φ_E. For instance, if $\varphi_C = 2{\cdot}5°$ and $\varphi_E = 2{\cdot}75°$, then the molecule will be compressed in the ratio of 0·91 along z. Hence, even a small error in the determination of φ will produce a significant systematic error in z. This is a consequence of the lack of physical information inherently residing in the small rotation angle between the left and right images. Other systematic errors might arise in the photographic reduction due to lens aberrations.

The steps in the procedure can be summarized as follows.

(i) Determine x_L, y_L and x_R, y_R from the stereodiagram. [One of these sets must now be rotated and

translated in order to minimize $\sum(y_L - y_R)$ to compensate for any fortuitous rotation and translation of the diagrams with respect to each other.]

(ii) Find the center of gravity of the left and the right coordinate sets and refer x_L, y_L and x_R, y_R to these origins. (One of the referees points out that some improvement might be obtained by referring both diagrams to the same origin and by considering the translation factor as a further variable along with v and φ.)

(iii) Search for the minimum in E (expression 4) to obtain the angular separation φ and viewing distance v.

(iv) Compute x,y,z from expressions (3).

(v) Refine x, y and z given φ and v from (4).

In practice it has been found that some atomic coordinates are particularly poor due to overlapping in one or other of the projections. Such inaccurate coordinates can interfere in the search procedure for φ and v. However, these can readily be eliminated from the search procedure by using the C_α–C_α criterion in conjunction with reasonable test values of φ and v (*e.g.* $\varphi = 3°$, $v = \infty$). The test must be applied immediately preceding step (iii).

Results

The procedure was tested by comparing 'calculated' coordinates determined from published stereodiagrams against the corresponding 'observed' sets obtained from the original investigators. Two examples were chosen: one easy, where each half-diagram could be readily followed, and one difficult, where each monoprojection contained many overlapping atoms and bonds. A stereodiagram of cytochrome b_5 (Mathews, Levine & Argos, 1972) represented the easy example while tomato bushy stunt virus (TBSV) protein subunit (Harrison, Olson, Schutt, Winkler & Bricogne, 1978) provided the difficult case. Original coordinates were kindly supplied by Drs Scott Mathews and Steve Harrison, respectively.

The stereodiagrams were photographed without change of size. The resultant transparencies were digitized on an Optronics film scanner with a 100 μm raster. The optical densities were then listed on a line printer where each density was represented by a single character, but those below a given threshold were shown simply as an asterisk. Thus, the output was essentially binary where the bond lines of the original stereodiagrams were easily recognizable on a much enlarged scale. The molecular line drawing could then be followed in most places. Consultation of the original stereopair was able to resolve the remaining ambiguities. The x and y coordinates of all atoms could then be read in terms of raster steps.

In Table 1 are listed the results for the analysis of the cytochrome b_5 and TBSV stereodiagrams. Shown are

Table 1. *Determination of atomic coordinates from stereodiagrams*

		Before refinement			Search procedure			After refinement			
	n	$\{\Sigma(y_L - y_R)^2/n\}^{1/2}$ (Å)	$\{\Sigma(x_L - x_R)^2/n\}^{1/2}$ (Å)	$\{\Sigma(R - 3{\cdot}84\ Å)^2/(n-1)\}^{1/2}$ (Å)	Number of atomic positions rejected as too inaccurate	φ (°)	v (m)*	$\{\Sigma(R - 3{\cdot}84\ Å)^2/(n-1)\}^{1/2}$ (Å)	$\{\Sigma(x_o - x_c)^2/n\}^{1/2}$ (Å)	$\{\Sigma(y_o - y_c)^2/n\}^{1/2}$ (Å)	$\{\Sigma(z_o - z_c)^2/n\}^{1/2}$ (Å)
Cytochrome b_5	86	0·12	0·61	0·78	1	3·0	30	0·29	0·92	0·84	1·49
TBSV	308	0·32	1·26	1·16	43	4·25	10	0·53	1·17	1·05	2·92

Notes:
(1) n is the number of atoms in the polypeptide chain according to the stereodiagram analysis.
(2) x_L, y_L and x_R, y_R are coordinates of atoms on left and right stereodrawings. Although originally measured in microns, they have been scaled to correspond to Å here for comparison purposes.
(3) R is a calculated C_α–C_α distance.
(4) x_o, y_o, z_o are the 'observed' atomic coordinates obtained from the original investigators.
(5) x_c, y_c, z_c are the 'calculated' atomic coordinates determined from the stereodiagrams.
(6) Rejected atoms are omitted only in the search procedure for finding φ and v.

* 1 in = 25·4 mm.

Table 2. *Two-dimensional exploration of angular separation (φ) and viewing distance (v) for coordinates taken from a cytochrome b_5 stereo pair*

v (in)*	φ (°)									
	2·00	2·25	2·50	2·75	3·00	3·25	3·50	3·75	4·00	4·25
10	0·914	0·842	0·798	0·774	0·766	0·767	0·775	0·786	0·799	0·813
20	0·862	0·804	0·770	0·754	0·751	0·756	0·766	0·779	0·794	0·809
30	0·856	0·800	0·768	0·753	0·750	0·756	0·766	0·779	0·794	0·809
40	0·856	0·800	0·768	0·753	0·751	0·756	0·766	0·780	0·794	0·810
∞	0·862	0·806	0·774	0·758	0·755	0·760	0·770	0·782	0·797	0·811

Note: Numbers represent r.m.s. deviations in Å for calculated C_α—C_α distances from 3·84 Å.

* 1 in ≡ 25·4 mm.

the r.m.s. deviation between the measured x_L and x_R and the measured y_L and y_R coordinates. The latter pair should be the same and thus give an estimate of the error in the experimental determination of coordinates. The former must be significantly different since variation in the x coordinates gives the determination of the unknown z parameters. Thus $p = [\sum(x_L - x_R)^2/\sum(y_L - y_R)^2]^{1/2}$ is a measure of the power of the technique when applied to a given diagram. It will be observed that the determinations of φ and v (Table 2) give reasonable results when based only on the C_α—C_α distances. Inclusion of α-helical parameters gave essentially the same results. The r.m.s. deviation of all C_α—C_α distances from 3·84 Å was improved from 0·78 to 0·29 Å for cytochrome b_5 and from 1·16 to 0·53 Å for TBSV with the refinement of the x, y and z parameters.

Comparison of the observed and calculated coordinates was performed by a least-squares procedure (Rao & Rossmann, 1973; Rossmann & Argos, 1975) which obtains the best fit of two molecules in space. In this case, the two molecules were the 'observed' (coordinates from original investigator) and 'calculated' (coordinates from stereodiagram) structures. While the z parameters (depth) do have a systematic error due to a small inaccuracy of estimating φ, no substantial error was found (Table 1). The larger error in z for TBSV reflects the larger molecular thickness along z so that the systematic error will be greater at the extremities of the molecule. A comparison of the original diagram of cytochrome b_5 (Mathews *et al.*, 1972) and one drawn from the calculated coordinates is shown in Fig. 3.

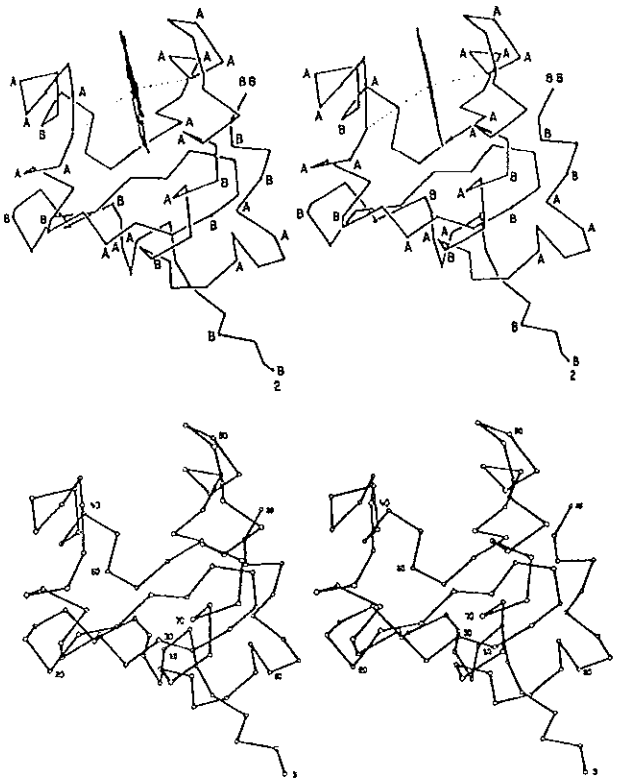

Fig. 3. Comparison of the original stereodiagram (top) as published by Mathews *et al.* (1972) with a stereodiagram drawn from the calculated coordinates.

The TBSV calculated coordinates were tested for their usefulness in showing structural equivalence which involves the topological superposition of two or more protein domains. Argos, Tsukihara & Rossmann (1980) have recently suggested structural analogy among the β-barrels comprising the two TBSV domains and concanavalin A. The necessary methodology has been reviewed by Rossmann & Argos (1978). The intent here is to demonstrate the utility of stereodiagram coordinates, even in a case as complex as TBSV, but not to discuss the functional and evolutionary implications of these comparisons. Table 3 shows the number of topologically equivalenced C_α atoms determined by using the observed and calculated coordinates. While the calculated coordinates gave somewhat fewer equivalences with a slightly higher r.m.s. deviation, the same homologous secondary structural spans were easily recognized.

Table 3. *Analysis of the utility of the coordinates derived from a stereodiagram of a TBSV subunit*

	'Observed' coordinates		'Calculated' coordinates	
	Number of equivalences	r.m.s. deviation (Å)	Number of equivalences	r.m.s. deviation (Å)
TBSV(*S*)–TBSV(*P*)	69	3·8	63	3·6
TBSV(*P*)–concanavalin A	68	3·4	58	3·6
TBSV(*S*)–concanavalin A	82	3·2	77	3·6

Notes:

(1) 'Observed' refers to coordinates from original investigator. 'Calculated' refers to coordinates derived from the published stereodiagrams.
(2) Coordinates of concanavalin A were obtained from the *GAPSOM Atlas* (Feldmann, 1975) and refer to the work of Hardman & Ainsworth (1972, 1973).
(3) (*S*) and (*P*) refer to the surface and protruding domains of the TBSV subunit (Harrison *et al.*, 1978).

The ethics question

The tradition of science is to gather and publish facts. Others may wish either to verify the facts by repeating the observations or to use these results to obtain a fundamental understanding of Nature in terms of a unifying concept or correlation. The accepted practice is to extract information from the literature, acknowledge its source, and to build upon it. The trend to withhold coordinates appears to be at odds with this long-standing tradition of scientific endeavor and exchange. Furthermore, coordinates are sometimes given only to close associates thus stifling a healthy public debate. Nevertheless, the present authors foresee that the technique published here may be considered a 'sharp' practice by some, although it is only extracting information from publications. This is evidenced by resistance to suggestions that coordinates be deposited with the Brookhaven Data Bank upon publication of high-resolution structures (*cf. Instructions to Authors* of the *Journal of Biological Chemistry*, 1979; *Crystallography of Molecular Biology*, 1976).

This situation appears to have arisen as many years of intensive effort are required by groups of scientists to determine tertiary structures. The researchers wish to announce their basic findings and yet withhold their quantitative results for some time in order to either (i) digest and utilize the coordinates for further scientific interpretations and thus reap the benefits of many years of effort or (ii) perhaps refine their atomic positions before release for the public domain to avoid erroneous deductions. Whichever is the case, it is clear that early publications announcing crystallographically determined structures often omit the detailed results. Hence, if the deduction of coordinates from published stereodiagrams represents a feat not intended by authors, controversy will obviously result.

The question of accuracy may not generally be sufficient justification for withholding the quantitative results. Even quite inaccurate coordinates can give information on such topics as polypeptide topology and possible gene duplication. Authors need simply state what they have done and how they have arrived at their results. As far as possible, they should give limits of error. It is those who use the coordinates outside the limits of accuracy who are to be held culpable. However, this does not mitigate the first author's responsibility in only publishing stereodiagrams whose features can be regarded with reasonable confidence. It can hardly be Tycho Brahe's fault if others arrive at unacceptable concepts of the solar system.

A fair and just solution to the problems raised here is imperative. Clearly the original author should always be approached before resorting to the extraction technique given here. Furthermore, the source of the coordinates, whether obtained directly or from a stereodiagram, must always be stated as recognition of the degree of error. In any event the continued absence of coordinates and now perhaps even stereodiagrams can only result in retardation of scientific advancement.

We would like to thank Sharon Wilder for help in the preparation of the manuscript. The work was supported by grants from the National Science Foundation (No. PCM78-16584) and the National Institutes of Health (Nos. AI 11219 and GM 10704) to MGR and by a grant from the National Science Foundation (No. PCM77-20287) and a Faculty Research Award from the American Cancer Society (No. FRA173) to PA.

References

ARGOS, P., TSUKIHARA, T. & ROSSMANN, M. G. (1980). Submitted for publication.

Crystallography of Molecular Biology (1976). 'Ettore Majorana' Centre for Scientific Culture, International School of Crystallography, Course III, Erice, Trapani, p. 9.

FELDMANN, R. J. (1975). *GAPSOM – Global Atlas of Protein Structure on Microfiche*. Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20014.

HARDMAN, K. D. & AINSWORTH, C. F. (1972). *Biochemistry*, 11, 4910–4919.

HARDMAN, K. D. & AINSWORTH, C. F. (1973). *Biochemistry*, 12, 4442–4448.

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature* (*London*), **276**, 368–373.

JOHNSON, C. K. (1970). *ORTEP*. Report ORNL-3794, second revision. Oak Ridge National Laboratory, Tennessee 37830.

Journal of Biological Chemistry (1979). **254**, 1–11 (*Instructions to Authors*).

MATHEWS, F. S., LEVINE, M. & ARGOS, P. (1972). *J. Mol. Biol.* **64**, 449–464.

RAO, S. T. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.

ROSSMANN, M. G. & ARGOS, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.

ROSSMANN, M. G. & ARGOS, P. (1978). *Mol. Cell Biochem.* **21**, 161–182.