ations from periodicity everywhere in the material. These deviations appear in the diffraction patterns as internal modulations. The internal modulations are then connected uniquely (in the present case) to the external modulations through Umklapp processes.

The rules for the occupancy of the atomic sites are thus found through the sequences of atomic layering over a large spatial range. One of these modulation periodicities agrees well with the second set of observed modulations. The appearance of this periodicity is suggestive that the limit of the occupancy sequences should be taken for successively larger periodic modulations. This generalization determines the entire sequence throughout the material (consistent with what can be seen from the experiment).

It is significant that the atomic scale models constructed *via* experimental analysis lead to periodic (modulated) structures rather than aperiodic ones. However, the fact that successively larger periodic arrays are required to complete the description of the occupancy function suggests that the true structure is at the limit of the periodic approximations. This view yields some physical insight into the meaning of almost periodic functions as applied to the Al-Mn icosahedral phase. The unusual form of the structure factor (occupancy function) calculation relevant to the model developed in this series of papers closely resembles the Fourier transform associated with almost periodic functions. This analysis, which was built up of periodic (modulated) sequence models can be used to help define the relationship between classical three-dimensional periodicity and almost (or 'quasi') periodicity in a physical system.

As the physical model is built up along the ⟨111⟩ axes according to the derived occupancy rules, it is possible that long-range flaws may occur. Such flaws would disrupt the perfect Fibonacci sequence while being consistent with the short-range requirements such as the rule that A may have as a neighbor either A or B, but B may only have A as a neighbor. Of course, such flaws would be very difficult to detect experimentally for reasons related to the difficulty of separating periodic and almost periodic lattices.

Viewed thus, the icosahedral phase belongs firmly under the classification of crystal rather than glass. Only in so far as icosahedral packing is relevant to many glasses (and this is considerable), does this structure tell us something new about glasses.

The clue to the discovery of the relationship between the reference lattice we derived and the Fibonacci sequence lay in the fact that two colinear sets of six modulations were derived from the electron diffraction patterns, and the ratio of their magnitudes was close to τ^3. In this regard, it is interesting that it was suggested by Mackay (1976) that a three-dimensional icosahedral construction be attempted based on $\tau^3 - \tau^{-3} = 4$. Similarly, Mackay (1982) noted that inflation of the three-dimensional model results in a coincidence of every other point, while in this paper we suggest that the construction requires that only half of the sites are occupied according to the derived occupancy sequence rule.

References

BANCEL, P. A., HEINEY, P. A., STEPHENS, P. W., GOLDMAN, A. I. & HORN, P. M. (1985). *Phys. Rev. Lett.* **54**, 2422-2425.
KURIYAMA, M., LONG, G. G. & BENDERSKY, L. (1985). *Phys. Rev. Lett.* **55**, 849-851.
LONG, G. G. & KURIYAMA, M. (1986). *Acta Cryst.* A**42**, 156-164.
MACKAY, A. L. (1976). *Phys. Bull.* pp. 495-497.
MACKAY, A. L. (1982). *Physica (Utrecht)*, **114**A, 609-613.
NELSON, D. R. & SACHDEV, S. (1985). *Phys. Rev. B*, **32**, 689-695.
REITER, G. & MOSS, S. C. (1982). *Phys. Rev. B*, **26**, 3520-3522.

Acta Cryst. (1986). A**42**, 172-178

Automated Calculation of Coordinate Transformations for the Superposition of Protein Structures

BY RICHARD B. HONZATKO

Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50011, *USA*

(*Received* 22 *April* 1985; *accepted* 22 *November* 1985)

Abstract

A computer program is described that performs a superposition of two protein structures. The program calculates a coordinate transformation that minimizes the root-mean-square deviation between atoms representing homologous structure in the two proteins. All atoms of the main chain and those atoms of side chains that bear common labels contribute to the calculation of the transformation. Required input by the user is either a small set of integers representing the sequence numbers of spatially equivalent residues in the two proteins and/or the initial and terminal residues of homologous elements of secondary struc-

0108-7673/86/030172-07$01.50

ture. After using the starting set of homologies to calculate an initial transformation, the program discards the user input and then determines the full set of homologous residues by application of simple criteria. The superposition that results is the point of departure of a search for alternative transformations that represent superpositions of merit. The computation time for the superposition of two structures of 150 residues is approximately 20 s on a VAX 11/780 and rises linearly with the size of the problem. Thus, the program is inexpensive in computer time and applicable to even the largest of macromolecules whose three-dimensional structures are known.

Introduction

Structural homology among proteins has been the focus of numerous investigations. The studies of Rossmann & Argos (1977) on the protein folds of dehydrogenases and kinases, of Richardson, Richardson & Thomas (1976) on the immunoglobins and superoxide dismutase, of Hendrickson & Love (1971) on the hemoglobins and of Weaver, Grutter, Remington, Gray, Isaacs & Matthews (1985) on the lysozymes exemplify a body of literature primarily concerned with the comparison of structure. If one of the goals is the calculation of a coordinate transformation that superimposes the structures under investigation, then the comparison requires ultimately the pairing of homologous residues in the two proteins. The assignment of corresponding residues is essential for the implementation of least-squares programs that minimize the root-mean-square distance between corresponding atoms by making adjustments to a coordinate transformation.

Several techniques are given in the literature for the assignment of corresponding residues in different protein structures. Rossmann & Argos (1976) use the distance of separation of α-carbons and a 'progression rule' to designate pairs of corresponding residues. Remington & Matthews (1978) run a probe structure of a specified number of residues along another structure possessing homology. The corresponding residues that produce the lowest root-mean-square difference gives the best superposition in the Remington & Matthews approach. More recently, Murthy (1984) and Liebman (1982) have introduced rapid techniques for the comparison of structure. The approach of Murthy has a number of features in common with that of the method of Rossmann & Argos. The method of Liebman, however, differs significantly from other techniques in that the comparison of structure is by way of two-dimensional distance plots. Liebman's approach provides a wealth of information, but unfortunately cannot provide the mathematical transformation relating two sets of coordinates. Of the methods above only the approach of Liebman incorporates, albeit indirectly, information that pertains to the relative orientation of residues of the polypeptide. The method described here incorporates information concerning the relative orientation of residues in the polypeptide and provides, in addition, a coordinate transformation for the superposition.

The Fortran program *SUPERIMP* searches for an optimal set of residue pairs that represents the structural homology between two proteins. *SUPERIMP* does not search for general homology between a given protein and a data bank of structures, but rather makes a detailed comparison of proteins already known to exhibit some homology. The program develops initially an approximate superposition based on homologous secondary structure. *SUPERIMP* in turn refines the initial superposition by identifying spatially equivalent residues, first in regions of secondary structure and then throughout the full extent of the polypeptide under consideration. The program examines whether the result is unique by searching for alternative superpositions of merit. Provided here are the conceptual basis of the program and examples of its performance.

Description of the program

The methods of Rossmann & Argos (1976) and Remington & Matthews (1978) spend the largest fraction of the total computer time in seeking homology between structures where its existence is unlikely. As most superpositions necessarily align elements of secondary structure, a transformation developed from matching the secondary structure of two proteins should provide an approximate superposition. The calculation of the initial transformation by *SUPERIMP*, although developed independently, is similar in concept to the approach of Murthy (1984). The initial transformation calculated by *SUPERIMP* derives entirely from a small set of spatially equivalent residues and/or segments of structure specified by the user. The user may equivalence all atoms of individual residues or just their α-carbons. Spatially equivalent segments are generally elements of secondary structure (β-strands or α-helices), but can be, in fact, any sequence of residues whose α-carbons fall close to a line. The number of α-carbons in spatially equivalent segments need not be equal. *SUPERIMP* represents elements of secondary structure as pairs of points designating the amino and carboxyl termini of the element. The program employs the least-squares technique of Kabsch (1978) in the calculation of a coordinate transformation based upon the initial set of points. If a point and its transformed mate deviate from each other beyond a specified limit, that pair is rejected and the coordinate transformation is calculated again. Of the initial set of spatially equivalent points, the program requires

only four pairs of points to fall within the specified limit.

The transformation developed from the initial set of points places the majority of α-carbons of the probe to within 5·0 Å of corresponding residues of the specimen. (In the following the term 'probe' refers to the structure that is transformed onto a related structure known as the 'specimen'.) In order to refine the initial transformation and compare single residues of the probe and specimen, the program erects a local reference frame at α-carbons within elements of secondary structure specified by the user. Two axes of a local frame connect the α-carbon of origin to neighboring α-carbons. The third axis is the cross product of the other two axes. In order to compare the local frame of a probe residue to the frames of specimen residues, *SUPERIMP* calculates the quantity

$$d_L = \sum_{i=1}^{3} |P^i - S^i|^2,$$

where P^i and S^i are the normalized vectors that represent the local frames of the probe and the specimen residues in the coordinate space of the specimen structure. d_L can vary between 0 and 12 Å^2. For adjacent residues in an ideal β-strand d_L is 4·72 Å^2. Values for d_L between the local frames of residue i and the $i\pm1$, $i\pm2$, $i\pm3$ and $i\pm4$ residues of an ideal α-helix are respectively 4·72, 7·78, 2·00 and 0·94 Å^2. Provided the initial transformation brings the α-carbon of the probe to within 5 Å of the correct α-carbon of the specimen, the comparison of local frames will yield at worse a twofold ambiguity in the match of probe and specimen residues in β-strands and α-helices.

SUPERIMP breaks ambiguities by choosing the correspondence between probe and specimen that gives the least value for

$$w_L d_L + w_\alpha d_\alpha,$$

where d_L is as defined above and d_α is the sum of squared distances between the $i-1$, i and $i+1$ α-carbons of the probe and corresponding α-carbons of the specimen. For a typical superposition, the root-mean-square displacement between corresponding α-carbons is 2·5 Å giving a value of 18·75 Å^2 for d_α. The typical values for d_L are close to 1·0 Å^2. Thus, equal emphasis of the two terms in the equation above requires w_L and w_α to have values of approximately 1·0 and 0·05. In practice equal emphasis of d_L and d_α ($w_L = 1{\cdot}0$ and $w_\alpha = 0{\cdot}05$) or complete reliance on d_L ($w_L = 1{\cdot}0$ and $w_\alpha = 0{\cdot}0$) generate the best results. Assignments of spatially equivalent residues that derive from a complete reliance on d_α, however, are always inferior to those assignments based on the equal weighting of d_α and d_L. In addition to the criteria above, the user may specify cut-off distances for the maximum separation of α-carbons of residues of the probe and specimen. The program behaves well using a limit of 5·0 Å, but limits of 8·0 Å or more usually cause a deterioration of the performance of *SUPERIMP*.

The criteria employed for breaking ambiguities cause two types of error in the assignment of equivalent residues. The first is the multiple assignment of a specimen residue to more than one residue of the probe and the second is the violation of monotonicity with regard to sequence number in the assignment of residues of the probe to residues of the specimen. Routines exist in *SUPERIMP* that automatically check for errors in assignment. If a residue of the specimen is assigned to more than one residue of the probe only the assignment that gives the least value for the quantity $w_L d_L + w_\alpha d_\alpha$ is retained. A second routine removes correspondences that violate a monotonic increase in the sequence of paired residues of the probe and specimen.

The process of matching reference frames, eliminating ambiguity and rejecting errant assignments yields a definitive set of correspondences between residues of secondary structure. After calculating a new transformation, the program repeats the process above until the number of equivalenced residues is constant. Typically, convergence occurs after five cycles. The program automatically includes not only equivalenced α-carbons, but also all atoms of the main chain and those atoms of the side chain bearing common labels.

After pairing residues of secondary structure, *SUPERIMP* determines the first and last residues of gap regions, residues of the probe and specimen that have yet to be equivalenced. The program includes as a gap any probe residue of secondary structure that is without an assigned residue from the specimen. The non-equivalenced residues that define gap regions are processed in the same manner as the regions of secondary structure. The program matches local frames, resolves ambiguities and rejects errant assignments. After the calculation and the application of a new transformation, *SUPERIMP* cycles through the sequence of steps given above until the number of equivalenced atoms is constant. As for the matching of residues in secondary elements, all atoms of homologous residues bearing the same atom label contribute to the calculation of the coordinate transformation. Typically five iterations will produce a stable set of correspondences. After achieving a constant number of pairs, the program updates the gap regions to reflect the equivalencing of new residue pairs. The program searches then for additional pairs of residues in the redefined gap regions. The process of redefinition of gap regions, followed by searches for additional residue pairs continues until the number of equivalent residues is a constant.

The initial transformation developed by *SUPERIMP* will bring the center of mass of the

endpoints of secondary structure for the probe and specimen into coincidence. If the best superposition requires a significant separation of these centers of mass, then convergence to the best answer is unlikely; refinement leads to a false minimum.

SUPERIMP will move out of false solutions by implementation of a search routine. The search routine systematically displaces the probe by modifying the translation vector of the coordinate transformation of the 'false' solution. If the secondary structure shows a preferred orientation, shifts in translation vectors are consistent with that preferred direction. *SUPERIMP* determines an anisotropic distribution in translation vectors by first calculating the quantities

$$[\sum (\Delta S_x)^2]^{1/2}, \quad [\sum (\Delta S_y)^2]^{1/2}, \quad [\sum (\Delta S_z)^2]^{1/2}. \quad (1)$$

In the above, ΔS_x, ΔS_y and ΔS_z are the projections along the x, y and z axes of the line segment that connects the initial and final points of an element of secondary structure. The summation is over $2N$ elements of structure, N from the probe and N from the specimen. The maximum value of the three quantities above are scaled, then, to an outer and inner radius for the search. The program accepts shifts in the translation vector that fall between two elliptical surfaces given by the equations

$$X^2/[\sum (\Delta S_x)^2]_{\text{out}} + Y^2/[\sum (\Delta S_y)^2]_{\text{out}} + Z^2/[\sum (\Delta S_z)^2]_{\text{out}} < 1$$

$$X^2/[\sum (\Delta S_x)^2]_{\text{in}} + Y^2/[\sum (\Delta S_y)^2]_{\text{in}} + Z^2/[\sum (\Delta S_z)^2]_{\text{in}} > 1,$$

where the subscripts 'out' and 'in' refer to the quantities of (1) scaled to the outer and inner radius of the search. After each displacement, the program repeats the matching of local reference frames between residues in secondary elements and gap regions. The result is either convergence to the former superposition or convergence to a new solution. An example of multiple solutions as determined by the search routine appears in Fig. 1. In the search, *SUPERIMP* modifies only translation parameters. No search through the Eulerian angles is necessary as the angle parameters already reflect the similar orientation of secondary structure that must occur for any reasonable superposition. However, if one has made improper assignments of homology between elements of secondary structure, then the search routine may not be effective as the Eulerian angles may be far from their ideal values.

A serious problem in the superposition of structure lies in the mispairing of residues or spans of residues, due to insertions or deletions of amino acids from the primary structure of the probe or specimen. *SUPERIMP* pairs residues on the basis of local homology in structure between probe and specimen. If deletions and insertions of amino acids cause the structures of the probe and specimen to diverge, then *SUPERIMP* will not pair residues. Poor agreement of local reference frames automatically disqualifies tentative assignments between residues within either gap regions or elements of secondary structure. Random deletions of residues from the structure of the probe or specimen do not perturb the pairing of residues that border the deleted amino acid.

Results

We present results of eight superpositions that test the performance of the program. The tests are the superposition of lamprey hemoglobin (Honzatko, Hendrickson & Love, 1985) onto sperm whale myoglobin (Phillips, 1980), myohemerythrin (Hendrickson & Ward, 1975) onto cytochrome c′ (Weber, Howard, Xuong & Salemme, 1981), flavodoxin (Smith, Burnett, Darling & Ludwig, 1977) onto lactate dehydrogenase (White *et al.*, 1976), glyceraldehyde phosphate dehydrogenase (Murthy, Garavito, Johnson & Rossmann, 1980) onto lactate dehydrogenase, liver alcohol dehydrogenase (Eklund *et al.*, 1976) onto lactate dehydrogenase, malate dehydrogenase (Hill, Tsernoglou, Webb & Banaszak, 1972) onto lactate dehydrogenase, the variable domain of the light chain of the immunoglobulin REI (Epp, Lattman, Schiffer, Huber & Palm, 1975) onto superoxide dismutase (Tanier, Getzoff, Beem, Richardson & Richardson, 1982), and phage T4

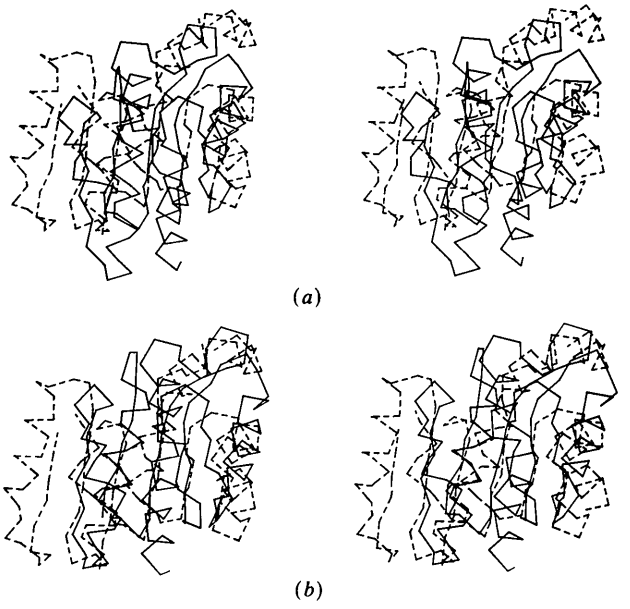

Fig. 1. Alternative solutions for the superposition of flavodoxin (solid line) onto the coenzyme domain of lactate dehydrogenase (dashed line). The two superpositions differ essentially by a translation of flavodoxin relative to lactate dehydrogenase in the direction of individual strands of the central β-sheet. The root-mean-square displacement of atoms of flavodoxin in the two solutions is 5·52 Å. The statistics of (*b*) are in Table 1.

lysozyme (Remington, Ten Eyck & Matthews, 1977) onto hen egg-white lysozyme (Diamond, 1974). Coordinates for all of the structures are in the Protein Data Bank at Brookhaven. The application of *SUPERIMP* to the pairs of molecules above yields from 40 to 90% homology. In Table 1 we give a complete list of correspondences chosen by the program, in order to facilitate the comparison of the results from *SUPERIMP* with other methods of superposition.

The central processor times listed in Table 1 are for single superposition trials on a VAX 11/780. The bulk of the computer time goes to the input of coordinates. For all of the structures listed in Table 1, the superposition time is approximately 10 s or less. A single superposition in the study of lysozyme, for instance, requires approximately 14 s of computer time for the input of coordinates and 6 s for the actual superposition. The two searches performed in the context of the lysozyme study required 8 and 4 min of central processor time on the VAX 11/780. The search of 8 min performed 69 superpositions; the 4 min search performed 41 trials. The computer time expended for individual superpositions during the two searches was approximately 6 s. *SUPERIMP* checks for the spatial equivalence of all atoms of each structure. Thus, typically the program brings into register a set of atoms which is five- to sixfold larger than the set of α-carbons considered by alternative

Table 1. *Statistics for the superposition of proteins*

Subject probe/specimen	Alignment of secondary structure (User specified) Probe	Specimen
Lamprey hemoglobin/ sperm whale myoglobin	Helices *A, B, C, D, E*	Helices *A, B, C, D, E*
Myohemerythrin/ cytochrome c′	Helices *A, B, C, D*	Helices *A, B, C, D*
Variable domain REI/ superoxide dismutase	β-strands *A, B, C, D, E, F, G*	β-strands *A, B, C, D, E, F, G*
Glyceraldehyde phosphate dehydrogenase/lactate dehydrogenase	β-strands *A, B, C, D, E, F*	β-strands *A, B, C, D, E, F*
Liver alcohol dehydrogenase/lactate dehydrogenase	β-strands *A, B, C, D, E, F*	β-strands *A, B, C, D, E, F*
Flavodoxin/lactate dehydrogenase	β-strands *A, B, D, E, F* and helix *A*	β-strands *A, B, D, E, F* and helix *A*
Malate dehydrogenase/ lactate dehydrogenase	β-strands *A, B, C, D, E, F*	β-strands *A, B, C, D, E, F*
Phage T4 lysozyme/ hen egg-white lysozyme	β-strands *A, B, C*; helices *A, C*	β-strands *A, B, C*; helices *B, C*

Alignment of residues by sequence number (program generated) Probe	Specimen
11-88, 91-92, 94-128, 129, 132-149	2-79, 80-81, 82-116, 128, 129-146
19-36, 37, 39-41, 43-61, 62, 70-90, 93-110	10-27, 35, 35-38, 39-57, 59, 83-104, 108-126
3-5, 7-8, 15, 18-24, 27-28, 30-31, 33-39, 40, 43-45, 46-47, 49-50, 64-66, 68, 69-74, 76, 79, 80, 81, 82-83, 84-90, 97-98, 99, 101-107	13-15, 16-17, 22, 28-34, 37-38, 39-40, 41-47, 49, 60-62, 81-82, 83-84, 85-87, 88, 91-96, 98, 103, 105, 108, 110-111, 113-119, 140-141, 143, 144-150
1-22, 26, 28, 29-36, 39-42, 44, 66, 67, 69, 70-73, 76, 82-84, 88-96, 99, 105, 107, 108-111, 113, 114-120, 141-147	22-43, 45, 48, 50-56, 58-61, 62, 66, 69, 75, 77-80, 84, 85-88, 90-98, 120, 124, 125, 127-130, 131, 133-139, 158-164
193-214, 215, 216, 218-228, 229-235, 239-240, 251, 253, 260, 262-268, 269, 270, 272-274, 276, 277-282, 284, 287-293, 295, 301, 302-303, 305, 307, 312, 314-318, 321-322, 323	22-43, 45, 47, 48-58, 48-58, 60-66, 79-80, 84, 86, 90, 91-97, 99, 120, 121-123, 124, 126-131, 132, 133-139, 140, 149, 152-153, 154, 155, 159, 160-164, 166-167, 169
1-4, 6, 13-28, 29-34, 29-34, 36, 39-54, 59, 60, 61, 63, 65-70, 72-73, 76, 80-87, 96-105, 106-108, 109, 110-111, 114-115	24-27, 28, 29-45, 48-53, 48-53, 54, 84-99, 112, 114, 117, 118, 119-124, 125-126, 129, 133-140, 142-151, 153-155, 158, 160-161, 162-163
3, 6-16, 17-24, 31, 33-39, 41, 43, 45-58, 60, 61-63, 64-67, 70, 75-83, 85, 88-89, 94, 96-102, 104-107, 108, 111-112, 115-120, 122, 123-133, 134, 136-139, 141, 142-146, 147-169, 171, 173-175, 176, 178-192, 197-203, 207-209, 210-211, 213-228, 230-231, 233-237, 239-250, 251-253, 258-266, 269-270, 271, 273, 274-284, 286-288, 289, 290-292, 294-313	22, 23-33, 35-42, 45, 47-53, 54, 55, 56-69, 70, 73-75, 77-80, 84, 89-97, 98, 101-102, 114, 115-121, 122-125, 128, 131-132, 133-138, 139, 141-151, 153, 154-156, 157, 159-163, 165-187, 188, 189-191, 193, 194-208, 209-215, 220-222, 226-227, 228-243, 244-245, 246-250, 251-262, 265-267, 269-277, 278-279, 281, 282, 284-294, 296-298, 300, 302-304, 305-324
3-13, 16-19, 22-33, 40-41, 45, 49-51, 56-73, 101-103, 105, 146, 148-149, 153, 157-158, 162	27-41, 43-48, 51-54, 56-60, 67, 70, 73, 75-77, 80-82, 85, 88-101, 105-108, 113-114, 116-117, 119-120, 122

Table 1 (*cont.*)

Central processor time Vax 11/780 (s)	Number of paired α-carbons (r.m.s. separation in Å)*	Number of paired atoms (r.m.s. separation in Å)
21	134 (1·41)	821 (1·50)
12	81 (2·32)‡	-
16	65 (2·88)	330 (2·82)
25	78 (2·29)	414 (2·34)
27	87 (2·27)	460 (2·39)
21	83 (2·49)	443 (2·55)
34	252 (2·37)	-
20	62 (2·93)	311 (2·78)

Local frame (r.m.s. Å)	Percent homology†	Coordinate transformation Euler angles $\theta_1, \theta_2, \theta_3$ (°)/ translation parameters t_x, t_y, t_z (Å)		
0·51	89	113·48,	81·46,	−8·83/
		−13·03,	29·99,	−21·34
0·79	66	4·53,	62·31,	−9·89/
		−4·33,	−30·99,	42·34
1·22	50	−173·26,	166·08,	133·91/
		−52·54,	34·38,	−6·82
1·17	52	62·31,	8·53,	−40·98/
		5·63,	36·61,	−11·72
0·97	61	4·51,	32·34,	−177·91/
		18·41,	48·72,	−10·17
1·04	58	95·90,	72·07,	125·75/
		8·47,	32·42,	−25·77
1·00	77	−79·22,	122·52,	78·95/
		−21·65,	−15·33,	−14·76
1·09	42	−138·13,	64·06,	−158·51/
		−10·86,	56·87,	44·50

* Root-mean-square separation of endpoints of local reference frames taken over all homologous pairs.

† Percent homology defined as 2·0 × (total number of homologous residues)/(total number of residues in the probe and specimen).

‡ Residue 82 is absent from the sequence of lactate dehydrogenase. As a consequence, *SUPERIMP* makes no assignment to residues 81 and 83 because local reference frames are undefined for these residues. For a continuous numbering of the sequence, *SUPERIMP* equivalences two more residues than reported in Table 1.

methods of superposition. By considering only α-carbons *SUPERIMP* expends under 2 s of computer time for each superposition for proteins of approximately 150 residues.

As the literature concerning the comparison of hen egg-white lysozyme and phage T4 lysozyme report correspondences between residues (Weaver *et al.*, 1985; Rossmann & Argos, 1976), we provide here a detailed account of the superposition of lysozymes. The initial transformation for the superposition of hen egg-white lysozyme (HEWL) onto phage T4 lysozyme (PT4L) derives from the specification of five elements of secondary structure. Helices of residue range 3–11 and 60–80 of PT4L were equivalenced to helices of residue range 25–35 and 89–95 of HEWL. β-strands of residue range 14–20, 23–28 and 37–34 of PT4L were equivalenced to β-strands of residue range 42–46, 50–54 and 57–60 of HEWL. Of the ten pairs of points in the initial set, the program rejected the two pairs corresponding to the carboxyl and amino termini of the helices of residue range 60–80 (PT4L) and 89–95 (HEWL). Rejection of the above pairs stems from the large discrepancy in length between the corresponding helices of HEWL and PT4L. The initial transformation accepted by *SUPERIMP* matched corresponding points to a root-mean-square deviation of 2·37 Å. Within the five elements of secondary structure the program selected 30 pairs of spatially equivalent residues, whose α-carbons were within 5·0 Å of each other and whose reference frames matched to within 2·0 Å^2. Subsequently, *SUPERIMP* located another 24 spatially equivalent pairs in gap regions whose α-carbons fell within 5·0 Å of each other and whose discrepancy in local frames fell below 5·0 Å^2. Thus, the initial run of *SUPERIMP* matched 54 α-carbons and 290 atoms yielding root-mean-square discrepancies of 2·13 and 2·30 Å, respectively.

A search for other superpositions of merit, using the solution above (solution *C*) as a point of departure, revealed two superpositions (*A* and *D*) that differed significantly from *C*. Solution *A* equivalenced 62 α-carbons; the results of solution *A* appear in Table 1. Solution *D* equivalenced 50 α-carbons and 248 atoms with root-mean-square discrepancies of 2·99 and 2·92 Å. As solution *A* revealed the largest number of spatially equivalent residues, we used solution *A* as a starting point for another search. Solution *B* resulted, equivalencing 56 α-carbons and 281 atoms with root-mean-square deviations of 2·77 and 2·62 Å. The solutions *A*, *B*, *C* and *D* account for 108 of the 110 trials performed during the two search operations. The remaining two trials produced solutions clearly inferior to any of the four above.

Solutions *B*, *C* and *D* differ from *A* essentially by a relative displacement of either one of two helices. Solution *B* has the first helix of PT4L (residues 3–11) shifted relative to solution *A* by four residues toward the amino terminus of the corresponding helix of HEWL (residues 25–35). On the other hand, solution *D* has the first helix of PT4L shifted by four residues toward the carboxyl terminus of the corresponding helix of HEWL. Solution *C*, however, shows no difference in the alignment of the first helix of PT4L relative to solution *A*. Instead, solution *C* has the second helix of PT4L (residues 60–80) offset by four residues toward the amino terminus of the corresponding helix of HEWL (residues 85–99).

Solution *A* is similar to the results of Weaver *et al.* (1985) and Rossmann & Argos (1976). The discrepancy in the total number of spatially equivalent residues as specified by *SUPERIMP* (62) from the total number of Weaver *et al.* (74) and Rossmann & Argos (78) resides in the 5·0 Å limit placed on the distance of separation of spatially equivalent residues. 18 correspondences in the work of Weaver *et al.* are above 5·0 Å and 22 correspondences are beyond the 5·0 Å limit in the alignment of Rossmann & Argos. The number of spatial equivalences common to any two of the three independent studies is 42. Only four correspondences common to the studies of Weaver *et al.* and Rossmann & Argos occur between α-carbons separated by more than 5·0 Å. Thus, the 5·0 Å limit used in *SUPERIMP* has little impact on the set of spatial equivalences that are reproduced by independent methods of superposition. Furthermore, only 35 correspondences are common to all three methods of superposition. If one uses consistency among independent methods of superposition as a criterion of homology, then all three methods of superposition overestimate homology.

One of the greatest disadvantages of *SUPERIMP* lies in its requirement for an initial set of points. In practice, however, if the user of *SUPERIMP* has a suspicion of homology between proteins, then the assignment of equivalent elements of secondary structure does not entail considerable labor. For all of the test cases presented here, the initial and final residues of each element of secondary structure come directly from the file stored in the Protein Data Bank. Only for the superpositions of lysozyme and flavodoxin onto lactate dehydrogenase did a search reveal a superposition of greater merit than that obtained from the initial transformation based on secondary structure. Aside from the speed of superposition, a strength of *SUPERIMP* lies in its rapid convergence to one member of a small set of plausible solutions. On the basis of the trials here, as the homology between two proteins rises to 70% and above, the set of plausible solutions has one member. For proteins of low homology the number of solutions increases. The worst case encountered thus far is that of the lysozymes.

Support for this work came from the NIH (GM33828) the ONR (N00014-84-G-0094) and the

Petroleum Research Fund (16269-G4). We thank W. A. Hendrickson for his critical reading of the manuscript and for his application of the program *SUPERIMP* in an independent study.

References

DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371-391.
EKLUND, H., NORDSTROM, B., ZEPPEZAUER, E., SODERLUND, G., OHLSSON, I., BOWIE, T., SODERBERG, B.-O., TAPIA, O., BRANDEN, C.-I. & AKESON, A. (1976). *J. Mol. Biol.* **102**, 27-101.
EPP, O., LATTMAN, E. E., SCHIFFER, M., HUBER, R. & PALM, W. (1975). *Biochem.* **14**, 4943-4952.
HENDRICKSON, W. A. & LOVE, W. E. (1971). *Nature (London) New Biol.* **232**, 197-203.
HENDRICKSON, W. A. & WARD, K. B. (1975). *Biochem. Biophys. Res. Commun.* **66**, 1349-1356.
HILL, E., TSERNOGLOU, D., WEBB, L. & BANASZAK, L. J. (1972). *J. Mol. Biol.* **72**, 577-591.
HONZATKO, R. B., HENDRICKSON, W. A. & LOVE, W. E. (1985). *J. Mol. Biol.* **184**, 147-164.
KABSCH, W. (1978). *Acta Cryst.* A**34**, 827-828.
LIEBMAN, M. N. (1982). In *Molecular Structure and Biological Activity*, edited by J. GRIFFIN & W. L. DUAX, pp. 193-212. New York: Elsevier.
MURTHY, N. R. N. (1984). *FEBS Lett.* **168**, 97-102.
MURTHY, N. R. N., GARAVITO, R. M., JOHNSON, J. E. & ROSSMANN, M. G. (1980). *J. Mol. Biol.* **138**, 859-872.
PHILLIPS, S. E. V. (1980). *J. Mol. Biol.* **142**, 531-534.
REMINGTON, S. J. & MATTHEWS, B. W. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2180-2184.
REMINGTON, S. J., TEN EYCK, L. F. & MATTHEWS, B. W. (1977). *Biochem. Biophys. Res. Commun.* **75**, 265-270.
RICHARDSON, J. S., RICHARDSON, D. C. & THOMAS, K. A. (1976). *J. Mol. Biol.* **102**, 221-235.
ROSSMANN, M. G. & ARGOS, P. (1976). *J. Mol. Biol.* **105**, 75-95.
ROSSMANN, M. G. & ARGOS, P. (1977). *J. Mol. Biol.* **109**, 99-129.
SMITH, W. W., BURNETT, R. M., DARLING, G. D. & LUDWIG, M. L. (1977). *J. Mol. Biol.* **117**, 195-225.
TANIER, J. A., GETZOFF, E. D., BEEM, K. M., RICHARDSON, J. S. & RICHARDSON, D. C. (1982). *J. Mol. Biol.* **160**, 181-217.
WEAVER, L. H., GRUTTER, M. G., REMINGTON, S. J., GRAY, T. M., ISAACS, N. W. & MATTHEWS, B. W. (1985). *J. Mol. Evol.* **21**, 97-111.
WEBER, P. C., HOWARD, A., XUONG, N. H. & SALEMME, F. R. (1981). *J. Mol. Biol.* **153**, 399-424.
WHITE, J. L., HACKERT, M. L., BUEHNER, M., ADAMS, M. J., FORD, G. C., LENTZ, P. J. JR, SMILEY, I. E., STEINDEL, S. J. & ROSSMAN, M. G. (1976). *J. Mol. Biol.* **102**, 759-779.

Acta Cryst. (1986). A**42**, 178-184

The Role of the Crystal Rotation Axis in Experimental Three- and Four-Beam Phase Determination

BY BEN POST* AND P. P. GONG

Polytechnic Institute of New York, Brooklyn, New York, USA

AND LISA KERN AND JOSHUA LADELL

Philips Laboratories, North American Philips Corporation, Briarcliff Manor, New York, USA

(*Received* 19 *August* 1985; *accepted* 4 *November* 1985)

Abstract

The geometry of four-beam diffraction and procedures for generating it systematically are described. These utilize relatively simple Renninger-type experimental arrangements. The four reciprocal-lattice points involved in each four-beam interaction are located at the corners of rectangles or symmetrical trapezoids in reciprocal space. One of the sides, or a diagonal, of each such quadrilateral serves as the axis of the azimuthal rotation of the crystal. Experiments designed to compare the relative merits of different types of rotation axes have been carried out. It is found that axes of twofold (or higher) symmetry provide advantages over alternate arrangements for experimental phase determination. Four-beam interactions are then generated systematically and in greater abundance than in all other n-beam interactions combined ($n > 2$). Such interactions usually provide stronger phase indications than comparable three-beam interactions. The experiments also showed that, although the phase of an 'invariant' quartet is clearly invariant to the choice of unit-cell origin, it is not necessarily invariant to a change of rotation axis from one two-fold axis to another.

I. Introduction

A. Four-beam diffraction

The use of four-beam diffraction data for the experimental determination of X-ray reflection

* This author's work was supported by the Division of Materials Research of the National Science Foundation and in part by the Joint Services Electronics Program of the US Defense Department.

0108-7673/86/030178-07$01.50