ARNONE, A. & PERUTZ, M. F. (1974). *Nature (London)*, **249**, 34–36.
DEREWENDA, Z. S. (1989). *Acta Cryst.* A**45**, 227–234.
DRABKIN, J. (1949). *Arch. Biochem. Biophys.* **21**, 224–236.
FERMI, G., PERUTZ, M. F. & SHAANAN, B. (1984). *J. Mol. Biol.* **175**, 159–174.
FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.
JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
LIDDINGTON, R. C. (1986). DPhil Thesis, Univ. of York, England.
LIDDINGTON, R. C., DEREWENDA, Z., DODSON, G. & HARRIS, D. (1988). *Nature (London)*, **331**, 725–728.
LUISI, B. F. (1986). PhD Thesis, Univ. of Cambridge, England.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MONOD, J., WYMAN, J. & CHANGEUX, J. P. (1965). *J. Mol. Biol.* **12**, 88–118.
RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. (1963). *J. Mol. Biol.* **7**, 95–99.
SHAANAN, B. (1983). *J. Mol. Biol.* **171**, 31–59.
WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A**35**, 901–911.
WONACOTT, A. J. (1980). *MOSCO. A Suite of Programs for On-Line Evaluation of Integrated Intensities on Small-Angle Rotation/Oscillation Photographs.* Unpublished.

Acta Cryst. (1990). B**46**, 418–425

Molecular-Replacement Structure Determination of Two Different Antibody:Antigen Complexes

BY STEVEN SHERIFF,* EDUARDO A. PADLAN, GERSON H. COHEN AND DAVID R. DAVIES

National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

(*Received* 12 *July* 1989; *accepted* 9 *January* 1990)

Abstract

We have used molecular replacement to determine the structures of two antibody:antigen complexes, HyHEL-5 Fab:lysozyme and HyHEL-10 Fab:lysozyme by orienting and locating the C_L:C_H1 domains, Fv and lysozyme. We used the model of McPC603 as our probe for the C_L:C_H1 domains and Fv. In the case of HyHEL-5 Fab:lysozyme, there were two closely related crystal forms and the top peak in the rotation function was correct in five out of six cases (two crystals and three probes). The top peak in the Crowther–Blow translation function was also correct in five out of six cases. We used the program *BRUTE* [Fujinaga & Read (1987). *J. Appl. Cryst.* **20**, 517–521] to put together the three pieces and thereby solve the relative origin problem in space group $P2_1$. In the case of HyHEL-10 Fab:lysozyme the top peak in the rotation function was correct for the Fv and lysozyme, but even with an appropriate model (the C_L:C_H1 domains of HyHEL-5) it was no better than the seventh peak (72% of the top peak) for C_L:C_H1. In the Crowther–Blow translation function, the top peaks were correct on two of three Harker sections for the Fv domains and lysozyme. The Crowther–Blow translation function was unable to locate C_L:C_H1, when using McPC603 C_L:C_H1 as the probe; however, when HyHEL-5 C_L:C_H1 was used, the top peak in all three sections was correct.

* Current address: Squibb Institute for Medical Research, PO Box 4000, Princeton, NJ 08543-4000, USA.

Introduction

Molecular replacement can be used to determine the location of a known structure in an unknown crystal. This use of molecular replacement was first demonstrated at low resolution by the determination of seal myoglobin using the structure of sperm-whale myoglobin (Tollin, 1969) and at intermediate resolution by the determination of carboxypeptidase B using the structure of carboxypeptidase A (Schmid, Lattman & Herriott, 1974; Schmid & Herriott, 1976). Since that time molecular replacement has become a routine technique for determining the structure of closely related macromolecules or the same macromolecule in a different unit cell. Several workers have used models which account for only part of the macromolecule in the crystal, *e.g.* the use of *Streptomyces griseus* protease B in the determination of the SGPB-ovomucoid inhibitor complex (Fujinaga, Read, Sielecki, Ardelt, Laskowski & James, 1982) and the use of trypsin in the determination of a trypsinogen–Kazal-type inhibitor complex (Bolognesi *et al.*, 1982).

More recently, Cygler, Boodhoo, Lee & Anderson (1987) and Cygler & Anderson (1988*a*,*b*) have used the C_L:C_H1 domains and the Fv (V_L:V_H) domains of the McPC603 Fab (Satow, Cohen, Padlan & Davies, 1986) to independently determine the location of these two pieces of structure for an unknown Fab, HED10. Building on this result we have determined the structures of two different antibody:antigen complexes, HyHEL-5 Fab:lysozyme and HyHEL-10 Fab:lysozyme. We determined the orientation and

location of three substructures, the C_L:C_H1 domains, the Fv domains and lysozyme. We demonstrate that we are able to locate by molecular replacement a substructure (lysozyme) that accounts for only 23% of the entire structure. We report here our experiences with the molecular replacement structure determination of HyHEL-5 Fab:lysozyme and HyHEL-10 Fab:lysozyme.

The monoclonal antibodies HyHEL-5 and HyHEL-10 (Smith-Gill, Wilson, Potter, Prager, Feldmann & Mainhart, 1982) bind to different parts of the surface of lysozyme (Davies, Sheriff & Padlan, 1988). The structure of the HyHEL-5 Fab:lysozyme complex in one crystal form has been published with $R = 0{\cdot}245$ and r.m.s. deviation from ideal bond lengths of 0·012 Å (Sheriff *et al.*, 1987). These coordinates have been deposited in the Protein Data Bank (Bernstein *et al.*, 1977) (Reference: 2HFL). Since that time the model of HyHEL-5 Fab:lysozyme in this crystal form has been refined to $R = 0{\cdot}199$ and the model of the HyHEL-5 Fab:lysozyme in a second closely related crystal form has been refined to $R = 0{\cdot}198$ (Sheriff *et al.*, unpublished results). Coordinates for these latter two models and data from crystals of the complexes will be deposited in the Protein Data Bank upon submission of a manuscript summarizing the results of these refinements. The structure of the HyHEL-10 Fab:lysozyme complex has been published with $R = 0{\cdot}24$ and r.m.s. deviation from ideal bond lengths of 0·011 Å (Padlan, Silverton, Sheriff, Cohen, Smith-Gill & Davies, 1989). The coordinates and structure factors for the HyHEL-10 Fab:lysozyme complex have been deposited in the Protein Data Bank (Reference: 3HFM and R3HFMSF, respectively).

Materials and methods

Crystallization and data collection

HyHEL-5 Fab:lysozyme complex was crystallized as described by Sheriff *et al.* (1987). Crystals of the HyHEL-5 Fab:lysozyme complex grow with differing *b*-cell axis lengths, ranging unpredictably between 65 and 75 Å, with symmetry consistent with space group $P2_1$. Two sets of data were collected: (1) from a crystal with cell dimensions of $a = 54{\cdot}9$, $b = 65{\cdot}2$, $c = 78{\cdot}6$ Å and $\beta = 102{\cdot}4°$; (2) from a crystal with cell dimensions of $a = 54{\cdot}8$, $b = 74{\cdot}8$, $c = 79{\cdot}0$ Å and $\beta = 101{\cdot}8°$ (Sheriff *et al.*, 1987).

HyHEL-10 Fab:lysozyme complex was crystallized and crystals were grown larger by repeated macroseeding as described by Silverton, Padlan, Davies, Smith-Gill & Potter (1984). The crystals exhibit the symmetry of space group $P2_12_12_1$ with $a = 57{\cdot}47$, $b = 118{\cdot}73$, $c = 137{\cdot}68$ Å and one Fab:lysozyme complex per symmetric unit. Data were collected for HyHEL-10 Fab:lysozyme complex as described by Padlan *et al.* (1989).

Molecular-replacement programs

We used a predecessor to the *MERLOT* package of programs assembled by Fitzgerald (1988), which included the fast-rotation function of Crowther (1972), the rotation function of Lattman & Love (1970), which uses a finely sampled transform of the model and avoids the need for integration radii, and the translation function of Crowther & Blow (1967). In addition we used the program *BRUTE* (Fujinaga & Read, 1987). We made minor modifications to *BRUTE* to run it on our Numerix NMX-432 array processor rather than the Floating Point Systems array processor for which it was originally written. We also modified the output of *BRUTE* so that it wrote correlation-coefficient and *R*-value map files compatible with the *MAPOUT* program in the *MERLOT* program package (Fitzgerald, 1988).

Probes

Following the procedure of Cygler & Anderson (1988*a*,*b*), independent rotation and translation searches were performed using probes for the lysozyme, Fv and C_L:C_H1 portions of the structures. We used the model of McPC603 (Satow *et al.*, 1986) which we oriented so that the axis of the hinge between Fv and C_L:C_H1 was parallel to the *z* axis. This allowed us to observe most of the difference in the elbow directly in the γ rotation-function angle (Cygler & Anderson, 1988*a*), since the γ angle is the first applied to the rotating body.

HyHEL-5 Fab:lysozyme

We used two probes from McPC603 (Protein Data Bank File Reference 1MCP) Fab, C_L:C_H1 and Fv. We removed the following residues from the model of V_L: 27C–31 and 91–95; and V_H: 30–31, 52B–54, 61–64, 96–100I and 101 [the numbering is that of Kabat, Wu, Reid-Miller, Perry & Gottesman (1987)]. We also used tetragonal lysozyme deposited by Diamond (1974) in the Protein Data Bank (Reference: 2LYZ).

For calculation of structure factors used to generate harmonic coefficients for the fast-rotation function, Fv and C_L:C_H1 were placed in boxes 74 × 111 × 82 Å, which was chosen so as to encompass the maximum dimensions of the entire Fab, plus the maximum radius of integration tested (27 Å) plus 1 Å. Later tests with smaller boxes for Fv and C_L:C_H1, yielded nearly indentical results, although a shoulder on the first peak, which was 'picked' as a peak at 84% of the height of the first peak, appeared in rotation-function maps with C_L:C_H1 and

HyHEL-5 Fab:lysozyme complex long *b*-cell axis data. Lysozyme was placed in a box 66 × 61 × 72 Å, which was chosen by the above criteria for the fast-rotation function harmonic coefficient calculation. For calculation of a finely sampled transform for the rotation function of Lattman & Love (1970) and for the Lattman implementation of the Crowther–Blow translation function, C_L:C_H1 and Fv were placed in boxes of 150 Å in each dimension, and lysozyme was placed in a box 130 Å in each dimension. We used an overall *B* factor of 17 Å^2 for all structure-factor calculations.

HyHEL-10 Fab:lysozyme

The probe used for the search for lysozyme was that deposited by Diamond (1974) in the Protein Data Bank (Reference: 6LYZ). Various molecules were used as probes for the Fab portion of the complex; these include the Fv and C_L:C_H1 modules of McPC603 Fab (Protein Data Bank File Reference 1MCP), J539 (1FBJ), NEW (3FAB), KOL (1FB4) and, later, the C_L:C_H1 of HyHEL-5 (2HFL). Of these, only the Fv of McPC603 and the C_L:C_H1 of HyHEL-5 gave good results. In addition, a modified Fv was obtained from McPC603 Fv by excising the appropriate number of residues from the hypervariable loops based on the known sequence differences between McPC603 and HyHEL-10 (McPC603 is longer by three residues in H2, six in H3 and six in L1). This 'pruned' McPC603 Fv proved useful by confirming the results obtained with the unmodified McPC603 Fv. To generate harmonic coefficients for the Crowther fast-rotation function, the probe molecule was placed in a box with dimensions equal to that of the probe plus the radius of integration plus 1 Å in each direction. Various combinations of reflection data, resolution and radius of integation were tried; a radius of integration of 27 Å and reflections between 10·0 and 4·5 Å spacings having $F \geq 5\sigma_F$ and $F \geq \langle F \rangle$ gave good results. For the Lattman–Love rotation search and for the Lattman implementation of the Crowther–Blow translation function, the probe molecules were placed in boxes of 150 Å in each dimension. An overall *B* factor of 15 Å^2 was used in all structure-factor calculations.

Structure determination of HyHEL-5:lysozyme complex

We used the fast-rotation function (Crowther, 1972) to determine the angles needed to rotate the probes so that they generated the structure of HyHEL-5. The results are summarized in Table 1 for 24 Å integration radius and all of the data from 10 to 4 Å spacings. We obtained similar results to those listed in Table 1 with a 27 Å integration radius and data from 10 to 4·5 Å spacings. The results show that in a case with good data the correct peak was the first peak in five out of six cases (three probes and two data sets). In the sixth case, that of lysozyme and the long *b*-cell axis data set, the correct peak was the second peak. This latter result was unchanged by changing to a 27 Å radius of integration or selecting data that were greater than $2\sigma_I$. A peak corresponding to the first peak with the long *b*-cell axis data did appear in a list of the peaks for the short *b*-cell axis data, but it was 12th highest at 83% of the maximum and an r.m.s. of 3·6.

Table 1. *Results of the fast-rotation function using C_L:C_H1, Fv and lysozyme probes with HyHEL-5 Fab:lysozyme complex data*

Data used to calculate harmonic coefficients were from 10–4 Å spacings. No origin removal was performed. Radius of integration was 24 Å.

Probe	Peak rank	α (°)	β (°)	γ (°)	Peak height R.m.s.*	Peak height % max.
(*a*) Long *b*-cell axis data						
C_L:C_H1	1	105	73	142	5·6	100†
	2	101	83	103	4·1	73
Fv	1	104	74	173	4·6	100†
	2	105	105	125	3·7	81
	3	113	103	164	3·7	81
Lysozyme	1	105	66	204	4·9	100
	2	108	90	300	4·3	86†
	3	23	125	55	4·1	82
(*b*) Short *b*-cell axis data						
C_L:C_H1	1	108	74	148	5·5	100†
	2	86	60	149	4·1	74
Fv	1	105	75	171	4·0	100†
	2	45	30	255	3·9	97
Lysozyme	1	108	88	303	4·3	100†
	2	153	80	270	4·0	92

* R.m.s. is the root-mean-square deviation from the mean.
† Correct peak.

The success of determining the correct translation is highly dependent on how accurate the rotation angles are. We used the rotation function of Lattman & Love (1970) to 'refine' the rotation angles. We used seven steps in each angle around the angle determined by the Crowther fast-rotation function and, successively, used 4, 2 and 1° increments. The results of this procedure are noted in Table 2.

To improve our confidence in the results we used the Crowther–Blow translation function and the program *BRUTE* to determine the translation of the individual pieces. The results are summarized in Table 2. With the exception of Fv and lysozyme (rotation peak 2) with the long *b*-cell axis data, the results were discouraging because there was little correlation between the techniques. However, analysis of the results following the structure determination showed that the Crowther–Blow translation function yielded the correct answer in five out of six cases and the second peak was correct in the sixth case. The correlation coefficient in *BRUTE* was correct in two out of six cases and the third peak was correct in a third case. The *R* value in *BRUTE* was

Table 2. *Results of Crowther–Blow translation function and program BRUTE using C_L:C_H1, Fv and lysozyme probes with HyHEL-5 Fab:lysozyme complex data*

For the Crowther–Blow translation-function data from 10–4 Å spacings were used with no origin removal. The translation function was sampled every 0·02 of a unit-cell edge, *i.e.* approximately every 1·1 Å along *u* and every 1·6 Å along *w*. For *BRUTE*, data from 5–4 Å spacings were used. The correlation coefficient and *R* value were sampled every 1 Å in both directions.

							Program *BRUTE*									
			Translation function				Correlation coefficient					*R* value				
Probe	Rotation angles (°)	Peak rank	*x**	*z**	R.m.s.†	% max.	*x**	*z**	Value	R.m.s.†	% max.	*x**	*z**	Value	R.m.s.†	% max.
(*a*) Long *b*-cell axis data																
C_L:C_H1	104, 73, 143	1	0·19	0·38	3·7	100‡	0·39	0·32	0·1596	4·2	100	0·02	0·32	0·5304	3·1	100
		2	0·45	0·45	3·6	99	0·13	0·09	0·1435	3·2	90	0·36	0·08	0·5310	3·0	100
Fv	103, 73, 172	1	0·29	0·31	5·3	100‡	0·28	0·31	0·1517	4·4	100‡	0·29	0·30	0·5342	3·4	100‡
		2	0·04	0·34	3·4	64	0·25	0·23	0·1398	3·7	92	0·17	0·34	0·5353	3·1	100
Lysozyme	108, 90, 302§	1	0·42	0·29	5·5	100‡	0·42	0·29	0·1381	4·6	100‡	0·43	0·30	0·5242	4·3	100‡
		2	0·30	0·09	3·3	61	0·04	0·32	0·1137	3·2	82	0·41	0·48	0·5279	3·5	99
(*b*) Short *b*-cell axis data																
C_L:C_H1	106, 75, 151	1	0·07	0·41	3·5	100‡	0·16	0·09	0·1426	3.5	100	0·10	0·11	0·5245	3·8	100
		2	0·25	0·02	3·0	86	0·49	0·01	0·1387	3·3	97	0·06	0·40	0·5286	3·1	99‡
		3	0·48	0·02	2·8	81	0·06	0·40	0·1345	3·1	94‡	0·36	0·48	0·5304	2·8	99
Fv	103, 74, 171	1	0·16	0·34	3·6	100‡	0·46	0·07	0·1539	3·1	100	0·47	0·08	0·5285	3·5	100
		2	0·50	0·41	3·2	90	0·29	0·17	0·1511	3·0	98	0·09	0·50	0·5290	3·3	100
Lysozyme	108, 90, 302	1	0·13	0·34	3·3	100	0·46	0·38	0·1363	3·2	100	0·05	0·26	0·5353	3·2	100‡
		2	0·05	0·26	3·0	90‡	0·29	0·31	0·1339	3·1	98	0·39	0·49	0·5354	3·2	100
		3	0·05	0·04	2·9	89	0·21	0·15	0·1304	2·9	96	0·19	0·46	0·5373	2·9	100

* Fractions of a unit-cell edge.
† R.m.s. is the root-mean-square deviation from the mean.
‡ Correct peak.
§ Second peak in rotation function list; see Table 1(*a*).

correct in three out of six cases and was the second peak in a fourth case. However, the *R* value provided little discrimination between the first and second peaks. On the other hand the translation function provided considerable discrimination in four of the five cases in which it was correct, but it also provided good discrimination in the case in which it was incorrect. The correlation coefficient provided good discrimination in the two cases when it was correct, but it also provided good discrimination in a case when it was incorrect.

Following this stage we used *BRUTE* to rotate the probes in successively 2 and 1° increments around the orientation determined by the rotation function. From this we found a match between the translation predicted by the correlation coefficient and the translation function for Fv with short *b*-axis data. In no other case did this procedure bring the first peak of either the correlation coefficient or the *R* value into agreement with the Crowther–Blow translation function. However, as noted in Table 3 we did use the improved orientations obtained for the probes by this procedure.

In space group $P2_1$, the origin is not defined along the *y* axis, so with three probes we had a relative origin problem similar to that which exists for multiple isomorphous replacement in space groups where the origin is not defined in at least one direction. To piece together our structure we used *BRUTE*, which has the facility to hold a probe stationary, while moving another probe to determine the best correlation coefficient and *R* value.

We started with the long *b*-cell axis data and held the Fv probe stationary while we separately investigated the behavior of the lysozyme and C_L:C_H1 probes. The results are summarized in Table 3(*a*)(i). In both cases the correlation coefficient and *R* value were far above the second peak and were correct (relative to Table 2 the result for C_L:C_H1 is an origin shift to *x*, *z* = 0, 1/2). At this point we had determined the structure for the long *b*-cell axis, but to see the effect of two stationary probes, we held Fv and lysozyme stationary and moved C_L:C_H1. This result is summarized in Table 3(*a*)(ii).

With the short *b*-cell axis data we did not recognize the correlation for lysozyme between the second Crowther–Blow translation function peak and the *BRUTE R*-value first peak until after the structure was determined. Instead we used the result for Fv that upon 'refining' the rotation the *BRUTE* correlation coefficient corresponded to the Crowther–Blow translation function peak. We held the Fv probe stationary while we moved the C_L:C_H1 and lysozyme probes, Table 3(*b*)(i). The correlation coefficient clearly distinguished between the first and second peaks in both cases and the first peak was correct. Although the *R*-value peak was correct for the C_L:C_H1 probe, the fourth peak was the correct one for lysozyme. Relative to the results in Table 2, C_L:C_H1 shows an origin shift of *x*, *z* = 0, 1/2 and lysozyme shows an origin shift of *x*, *z* = 1/2, 0. Again at this stage we had determined the structure, but we examined the effect of holding two probes stationary, Fv and C_L:C_H1 in this case, and moving the third

Table 3. *Results of program BRUTE holding one or two probes stationary while moving another probe with HyHEL-5 Fab:lysozyme complex data*

Data from 5–4 Å spacings were used. The correlation coefficient and R value were sampled every 1 Å in x, y and z.

Moving probes	Rotation angles (°)	Peak rank	Correlation coefficient x^*	y^*	z^*	Value	R.m.s.†	% max.	R value x^*	y^*	z^*	Value	R.m.s.	% max.
(*a*) Long *b*-cell axis data														
(i) Stationary probe: Fv; 101·3, 73·2, 171·0; x^* = 0·28, y^* = 0·00, z^* = 0·31														
C_L:C_H1	104, 73, 143	1	0·20	0·88	0·88	0·2599	7·2	100‡	0·19	0·88	0·88	0·5030	6·1	100‡
		2	0·74	0·44	0·73	0·2057	4·4	79	0·05	0·94	0·80	0·5130	4·5	98
Lysozyme	108·0, 88·1, 301·4	1	0·42	0·59	0·29	0·2588	7·8	100‡	0·42	0·59	0·29	0·4858	9·1	100‡
		2	0·54	0·50	0·46	0·2041	4·6	79	0·39	0·92	0·14	0·5098	4·4	95
(ii) Stationary probes: Fv; 101·3, 73·2, 171·0°; x^* = 0·28, y^* = 0·00, z^* = 0·31 Lysozyme; 108·0, 88·1, 301·4°; x^* = 0·42, y^* = 0·59, z^* = 0·29														
C_L:C_H1	104, 73, 143	1	0·19	0·88	0·88	0·3468	8·2	100‡	0·19	0·88	0·88	0·4692	8·7	100‡
		2	0·46	0·11	0·62	0·2906	5·1	84	0·68	0·11	0·08	0·4900	4·9	96
(*b*) Short *b*-cell axis data														
(i) Stationary probe: Fv; 105·4, 76·4, 169·5°; x^* = 0·16, y^* = 0·00, z^* = 0·34														
C_L:C_H1	106, 75, 151	1	0·06	0·89	0·90	0·2402	5·3	100‡	0·06	0·89	0·90	0·5114	4·7	100‡
		2	0·88	0·32	0·48	0·2296	4·6	96	0·33	0·98	0·51	0·5135	4·4	100
Lysozyme	108·0, 88·1, 301·4	1	0·55	0·63	0·26	0·2312	4·8	100‡	0·29	0·30	0·31	0·5174	4·3	100
		2	0·08	0·28	0·55	0·2179	4·2	94	0·22	0·41	0·43	0·5176	4·2	100
		3	0·39	0·44	0·48	0·2166	4·2	94	0·89	0·02	0·63	0·5177	4·2	100
		4	0·17	0·29	0·58	0·2132	4·0	92	0·55	0·63	0·26	0·5184	4·1	100‡
(ii) Stationary probes: $V_L + V_H$; 105·4, 76·4, 169·5°; x^* = 0·16, y^* = 0·00, z^* = 0·34 $C_L + C_H1$; 106·0, 75·0, 151·0°; x^* = 0·06, y^* = 0·89, z^* = 0·90														
Lysozyme	108·0, 88·1, 301·4	1	0·55	0·63	0·26	0·3235	6·6	100‡	0·54	0·63	0·26	0·4958	5·1	100‡
		2	0·30	0·06	0·31	0·2779	4·1	86	0·86	0·26	0·55	0·4967	3·7	100

* Fractions of a unit-cell edge.
† R.m.s. is the root-mean-square deviation from the mean.
‡ Correct peak.

probe, lysozyme. The results are shown in Table 3(*b*)(ii), where the first peak for both the correlation coefficient and R value are correct, but there is much better discrimination by the correlation coefficient.

Structure determination of HyHEL-10:lysozyme complex

The results of the Crowther search are summarized in Table 4. The orientation angles obtained by the Crowther search were further refined using the Lattman–Love procedure; the refined values were then used as input into the translation function. The correctness of each solution was assessed on the basis of packing considerations, closeness of the angles in the case of the Fv and C_L:C_H1 modules, and the existence of an internally consistent set of vectors from three Harker sections in the translation search.

The search for lysozyme was straightforward. The highest peak in the rotation function (Table 4) was refined to angles of 172·58, 61·84 and 33·87°. Using these values, the translation search yielded an internally consistent set of vectors as shown in Table 5.

The Crowther search for the Fv and C_L:C_H1 portions of the structure yielded conflicting results with the different probes used. The peaks that were common to the various search results did not give good translation-function solutions, nor did they pack well in the unit cell. The highest peak from the search with McPC603 Fv, although confirmed by the

Table 4. *Results of the fast-rotation function using C_L:C_H1, Fv and lysosyme probes with HyHEL-10 Fab:lysozyme complex data*

Data used to calculate harmonic coefficients were from 10–4·5 Å spacings. No origin removal was performed. Radius of integration was 27 Å.

Probe	Peak rank	α (°)	β (°)	γ (°)	Peak height R.m.s.	% max.
Lysozyme	1	170·00	65·00	35·00	4·1	100·0†
	2	82·50	35·00	145·00	3·5	86·0
McPC603 Fv	1	52·03	63·44	321·11	4·0	100·0†
	2	86·19	61·46	19·90	4·0	99·3
'Pruned'	1	51·94	63·44	321·13	4·2	100·0†
McPC603 Fv	2	111·31	69·38	11·17	4·0	94·7
HyHEL-5	1	115·00	60·00	45·00	4·3	100·0
C_L:C_H1	⋮	⋮	⋮	⋮	⋮	⋮
	7	52·50	60·00	305·00	3·1	72·3†

* R.m.s. is the root-mean-square deviation from the mean.
† Correct peak.

results with the 'pruned' McPC603 Fv, was only marginally higher than the next peak (Table 4). Nevertheless, this peak, which refined to 127·55, 244·27, 319·65° (having chosen, from packing considerations, the equivalent orientation: 180 − α, β + 180, γ), when used in the translation search, gave the sets of peaks shown in Table 6.

No satisfactory solution was found in the search for C_L:C_H1 using the corresponding module from the known Fab stuctures McPC603, J539, NEW, or KOL. An exhaustive search using the McPC603 C_L:C_H1 around the Fv rotation-function solution

Table 5. *Results of Crowther–Blow translation function with lysozyme with HyHEL-10 Fab:lysozyme complex data*

Data from 8–4 Å spacings were used with no origin removal. The translation function was sampled every 1/72 of a unit-cell edge, *i.e.* approximately 0·8 Å along *u*, 1·6 Å along *v* and 1·9 Å along *w*.

Orientation: 172·58, 61·84, 33·87°

Harker section	Peak rank	x*	y*	z*	Peak height R.m.s.†	% max.
z = 1/2	1	0·2916	0·1111	0·5000	4·5	100·0‡
	2	0·6042	0·3194	0·5000	3·1	68·0
y = 1/2	1	0·2917	0·5000	0·4862	5·2	100·0‡
	2	0·0486	0·5000	0·4375	3·5	67·7
x = 1/2	1	0·5000	0·2848	0·4931	6·6	100·0
	2	0·5000	0·2570	0·4931	6·3	95·0
	3	0·5000	0·1112	0·4862	4·7	71·2‡
	4	0·5000	0·3056	0·0070	4·0	59·9

* Fractions of a unit-cell.
† R.m.s. is the root-mean-square deviation from the mean.
‡ Correct peak.

Table 6. *Results of Crowther–Blow translation function with McPC603 Fv with HyHEL10 Fab:lysozyme complex data*

Data from 8–4 Å spacings were used with no origin removal. The translation function was sampled every 1/72 of a unit-cell edge, *i.e.* approximately 0·8 Å along *u*, 1·6 Å along *v* and 1·9 Å along *w*.

Orientation: 127·55, 244·27, 319·65°

Harker section	Peak rank	x*	y*	z*	Peak height R.m.s.†	% max.
z = 1/2	1	0·3680	0·2362	0·5000	4·8	100·0‡
	2	0·4167	0·2362	0·5000	4·5	93·8
y = 1/2	1	0·3681	0·5000	0·3750	5·5	100·0‡
	2	0·1042	0·5000	0·4792	3·7	67·6
x = 1/2	1	0·5000	0·2639	0·0208	4·7	100·0
	2	0·5000	0·2362	0·4445	4·1	87·2
	3	0·5000	0·2570	0·4584	4·0	85·5
	4	0·5000	0·2362	0·3750	3·9	82·2‡
	5	0·5000	0·2014	0·4653	3·8	81·4

* Fractions of a unit-cell edge.
† R.m.s. is the root-mean-square deviation from the mean.
‡ Correct peak.

yielded a probable orientation of 127·82, 239·59, 306·91° for this part of the structure. These angles, however, did not give a satisfactory translation-function solution. After the HyHEL-5 Fab:lysozyme complex structure was determined by molecular replacement [see above and Sheriff *et al.* (1987)] its C_L:C_H1 module was used as a probe in the searches. The peak in the Crowther search that turned out to be correct (Table 4) was only the seventh highest; the orientation angles refined to 128·54, 240·10, 308·38°, which are essentially those found above for McPC603 C_L:C_H1. The translation search using the HyHEL-5 C_L:C_H1, on the other hand, yielded a solution as shown in Table 7(*a*).

The inability of McPC603 C_L:C_H1 to give a satisfactory translation function solution is puzzling in view of the closeness of the orientation angles found for this module and for that of HyHEL-5 which gave an internally consistent set of vectors. The C_L's of

Table 7. *Results of Crowther–Blow translation function with HyHEL-5 C_L:C_H1 with HyHEL-10 Fab:lysozyme complex data*

Data from 8–4 Å spacings were used with no origin removal. The translation function was sampled every 1/72 of a unit-cell edge, *i.e.* approximately 0·8 Å along *u*, 1·6 Å along *v* and 1·9 Å along *w*.

Harker section	Peak rank	x*	y*	z*	Peak height R.m.s.†	% max.
(*a*) Orientation: 128·54, 240·10, 308·38°						
z = 1/2	1	0·1250	0·4514	0·5000	4·8	100·0‡
	2	0·3194	0·2500	0·5000	3·4	71·7
y = 1/2	1	0·1250	0·5000	0·2292	6·0	100·0‡
	2	0·4445	0·5000	0·0278	3·3	55·4
x = 1/2	1	0·5000	0·4514	0·2292	4·3	100·0‡
	2	0·5000	0·2848	0·4514	4·3	99·8
(*b*) Orientation: 127·82, 239·59, 306·91°						
z = 1/2	1	0·1250	0·4514	0·5000	4·8	100·0‡
	2	0·3680	0·1736	0·5000	3·4	70·9
y = 1/2	1	0·1250	0·5000	0·2292	6·2	100·0‡
	2	0·2292	0·5000	0·0972	3·2	52·4
x = 1/2	1	0·5000	0·2848	0·4514	4·7	100·0
	2	0·5000	0·4514	0·2292	4·4	92·7‡
	3	0·5000	0·2708	0·0000	3·5	74·2

* Fractions of a unit-cell edge.
† R.m.s. is the root-mean-square deviation from the mean.
‡ Correct peak.

McPC603, HyHEL-5 and HyHEL-10 have the same sequence and probably the same structure. However, the C_H1's of HyHEL-5 and HyHEL-10 are the same, but differ from that of McPC603. Using HyHEL-5 C_L:C_H1, but with the orientation angles for McPC603 C_L:C_H1 (127·82, 239·59, 306·91°), the translation search gave the vectors shown in Table 7(*b*). The peaks corresponding to the correct solution are essentially the same as those shown in Table 7(*a*). However, when McPC603 C_L:C_H1 was used with the orientation angles for HyHEL-5 (128·54, 240·10, 308·38°), no translation-function solution was obvious.

An Fab:lysozyme complex was assembled from lysozyme, McPC603 Fv and HyHEL-5 C_L:C_H1, using *FRODO* (Jones, 1978) to choose visually from among the various symmetry-related possibilities the one that was most plausible. Many possibilities exist since for each component module of the complex (1) there are four rotational possibilities, (2) for every translation-function solution, x, $x + 1/2$ is also a solution (similarly for y and z), and (3) there are four equivalent positions of the space group. The criteria used in choosing a plausible composite structure were (1) that the Fv and the C_L:C_H1 modules be so disposed as to result in a reasonable Fab structure, (2) that there be close contact without undue interpenetration, between the Fv and the lysozyme, and (3) that the resulting crystal packing be reasonable. One composite structure was chosen and this had reasonable intermolecular and crystal contacts. The orientation and position of the various components of the assembled complex were then refined with CORELS (Sussman, 1985) allowing the V_L, V_H, C_L

and C_H1 domains, and lysozyme to move independently. The *CORELS* refinement was performed in three steps: first with 10·0–8·0 Å data (518 reflections), then with 10·0–7·0 Å data (1017 reflections), and finally with 10·0–6·0 Å data (1893 reflections), using only the reflections with $F \geq 5\sigma_F$. At the end of this refinement, the R value was 0·47 and the correlation coefficient was 0·40.

At this point the Fv was modified to conform to the HyHEL-10 sequence. Residues common to HyHEL-10 and McPC603 were retained; those that were different were replaced by Ala. Further, nine residues were excised from the L1 of McPC603, six from H2, eight from H3, two from the V_L–C_L switch region, and three from the V_H–C_H1 switch. The structure was then subjected to restrained least-squares refinement with *PROLSQ* (Hendrickson & Konnert, 1980; Hendrickson, 1985; Cohen, 1986) to an R value of 0·40 for the 12 501 reflections between 10·0 and 3·0 Å spacings with $F \geq 3\sigma_F$. The fit of the structure to a $2F_o - F_c$ map was then examined using *FRODO* (Jones, 1978). It turned out that the orientation chosen for lysozyme was wrong, while those for the Fab modules were correct. The Fab portion of the complex fitted the electron density very well; density was present for most of the side chains omitted during refinement and for the segments excised from the model. The fit of lysozyme to density, on the other hand, was sporadic. *CORELS* was therefore run on all four orientational possibilities, combined with two translational choices that gave reasonable intermolecular contacts. Seven of these possibilities gave R values ranging from 0·46 to 0·50; one gave an R value of 0·41. This last possibility was subjected to *PROLSQ* refinement, which very quickly resulted in an R value of 0·33; the fit of this model to a $2F_o - F_c$ map was now very good for all portions of the structure. Alternating rounds of *PROLSQ* refinement and model rebuilding resulted in a final R value of 0·24 with good geometry (Padlan *et al.*, 1989).

Discussion

We have shown that we can determine by molecular-replacement methods the structures of antibody:antigen complexes by orientating and locating three independent parts – Fv, C_L:C_H1 and lysozyme. Although Fv and C_L:C_H1 each contain about 40% of the structure, lysozyme contains only 23% of the structure. Nevertheless we were able to orient and locate lysozyme as easily as the larger fragments. This is undoubtedly due to the α-helices present in lysozyme, because experiments to test the limits of the method with just Fab's showed that a single domain (basically an antiparallel β-sheet structure) was an insufficient model (see also Cygler & Anderson, 1988*a*). We have presented a fairly easy case, HyHEL-5 Fab:lysozyme, and one that was quite difficult, HyHEL-10 Fab:lysozyme. The ease of determination of the HyHEL-5 Fab:lysozyme structure was in part due to the two data sets from closely related crystal forms. Comparison of the rotation-function angles for lysozyme suggested that the second peak was the correct one for the long *b*-axis crystal. This was a great aid when determining the translation function for lysozyme in this crystal form. HyHEL-10 Fab:lysozyme proved difficult because the model for C_L:C_H1 for the translation function had to be nearly identical and until the coordinates for HyHEL-5 Fab were available there was no appropriate model. The greater difficulty in analyzing the crystal structure of the HyHEL-10 complex may simply stem from the fact that the intensity data for this complex are of lesser quality [the R value relating symmetry-related reflections, R_{symm}, was 0·066 (Padlan *et al.*, 1989)] compared to those for the HyHEL-5 complex ($R_{symm} = 0{\cdot}044$) (Sheriff *et al.*, 1987).

To help orient the Fv and C_L:C_H1 domains, we found that prior orientation of the elbow bend so that the axis was parallel to the z axis was useful in the case of HyHEL-5 Fab:lysozyme. When this is done nearly all of the difference in elbow bend shows up in the γ rotation-function angle and the α and β angles are nearly identical (Table 1). This made it easy to choose the correct peak in the case of the short *b*-axis form data and the Fv probe, where there was not much discrimination between the first and second peaks. Moreover, this information provided us with early estimates of the elbow bend in the two crystal forms: 164° for the long *b*-axis form and 156° for the short *b*-axis form [these numbers were obtained by taking the elbow bend for McPC603 (Satow *et al.*, 1986) and adding the difference between the γ rotation-function angles for C_L:C_H1 and Fv – Table 1]. These early estimates have withstood the test of refinement and are essentially unchanged at 162 and 154°, respectively, and the difference in the elbow-bend angles also confirmed our belief that the principal difference between the protein structures in the two crystal forms was the elbow-bend angle.

We found that the Crowther–Blow translation function worked very well. When we began this work we were under the impression that the Crowther–Blow translation function often failed for lack of sufficiently accurate rotation-function angles. While we would agree with that general assessment, we found that we were able to obtain accurate enough rotation-function angles, probably as a result of our use of the Lattman rotation function and in one case program *BRUTE* to 'refine' the angles generated by the Crowther fast-rotation function. With HyHEL-5

Fab:lysozyme, the Crowther–Blow translation function correctly determined the location in the *xz* plane in five out of six cases and in the sixth case, the second highest peak was correct. With HyHEL-10 in two cases (lysozyme and Fv) the top peak was correct on two of the Harker sections with excellent to moderate discrimination and either the third (lysozyme) or fifth (Fv) peak was correct on the third Harker section. The same section ($x = 1/2$) was poor in both cases and this might have been a result of inadequate sampling, which was 1/72 of a cell edge or the equivalent of 1·6 and 1·9 Å, along the two longest cell edges. In the case of C_L:C_H1, the Crowther–Blow translation function correctly determined the location on all three Harker sections, although the discrimination was quite poor on the $x = 1/2$ section. In this third case the proper model for C_L:C_H1 was crucially important for the translation function. On the other hand, rotation-function angles which were slightly different had very little effect on the translation function (Table 7).

The correlation coefficient and *R* value as calculated by *BRUTE* with the resolution range we used were considerably less accurate in determining the location of the fragments of the HyHEL-5 Fab:lysozyme complex. The correlation coefficient often provided good discrimination between peaks, but was correct only in the best two cases for the Crowther–Blow translation function. The *R* value provided very little discrimination compared to the correlation coefficient, but did a slightly better job of finding the correct result. On the other hand, *BRUTE* provided a way to piece together the various fragments. In these cases the correlation coefficient provided good to excellent discrimination and was correct in all six cases tried (see Table 3). The *R* value was correct in the three cases tried with the long *b*-axis form data providing excellent discrimination on the basis of the r.m.s. deviation from the mean, but only moderate discrimination on the basis of the percent of the maximum value. However, with the short *b*-axis form data, the *R* value provided poor discimination and in one case the fourth peak was the correct peak. Cygler & Desrochers (1989) have shown that *BRUTE* works considerably better when a wider resolution range than 5–4 Å is used. However, this increases the computation time considerably.

We were able to orient and locate a rather small substructure, lysozyme, which was only 23% of the complete structure. We believe that α-helices that were oriented in a nonparallel manner provided the key to this result. Cygler & Anderson (1988*a*) and S. Sheriff (unpublished results) found it very difficult to properly orient a single antibody domain, which contains no α-helices, when using the rotation function on data from an Fab where a single domain represents about 25% of the structure. Given our results it should be possible to locate an even smaller substructure as long as it has a higher helical content than lysozyme.

We thank Paula Fitzgerald, Mirek Cygler and Masao Fujinaga for helpful discussions.

References

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BOLOGNESI, M., GATTI, G., MENEGATTI, E., GUARNERI, M., MARQUART, M., PAPAMOKOS, E. & HUBER, R. (1982). *J. Mol. Biol.* **162**, 839–868.

COHEN, G. H. (1986). *J. Appl. Cryst.* **19**, 486–488.

CROWTHER, R. A. (1972). In *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon & Breach.

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

CYGLER, M. & ANDERSON, W. F. (1988*a*). *Acta Cryst.* A**44**, 38–45.

CYGLER, M. & ANDERSON, W. F. (1988*b*). *Acta Cryst.* A**44**, 300–308.

CYGLER, M., BOODHOO, A., LEE, J. S. & ANDERSON, W. F. (1987). *J. Biol. Chem.* **262**, 643–648.

CYGLER, M. & DESROCHERS, M. (1989). *Acta Cryst.* A**45**, 563–572.

DAVIES, D. R., SHERIFF, S. & PADLAN, E. A. (1988). *J. Biol. Chem.* **263**, 10541–10544.

DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.

FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.

FUJINAGA, M., READ, R. J., SIELECKI, A., ARDELT, W., LASKOWSKI, M. JR & JAMES, M. N. G. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 4868–4872.

HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). In *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 614–617.

KABAT, E. A., WU, T. T., REID-MILLER, M., PERRY, H. M. & GOTTESMAN, K. S. (1987). *Sequences of Proteins of Immunological Interest*, 4th ed. Bethesda: National Institutes of Health.

LATTMAN, E. E. & LOVE, W. E. (1970). *Acta Cryst.* B**26**, 1854–1857.

PADLAN, E. A., SILVERTON, E. W., SHERIFF, S., COHEN, G. H., SMITH-GILL, S. J. & DAVIES, D. R. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 5938–5942.

SATOW, Y., COHEN, G. H., PADLAN, E. A. & DAVIES, D. R. (1986). *J. Mol. Biol.* **190**, 593–604.

SCHMID, M. F. & HERRIOTT, J. R. (1976). *J. Mol. Biol.* **103**, 175–190.

SCHMID, M. F., LATTMAN, E. E. & HERRIOTT, J. R. (1974). *J. Mol. Biol.* **84**, 97–101.

SHERIFF, S., SILVERTON, E. W., PADLAN, E. A., COHEN, G. H., SMITH-GILL, S. J., FINZEL, B. C. & DAVIES, D. R. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 8075–8079.

SILVERTON, E. W., PADLAN, E. A., DAVIES, D. R., SMITH-GILL, S. J. & POTTER, M. (1984). *J. Mol. Biol.* **180**, 761–765.

SMITH-GILL, S. J., WILSON, A. C., POTTER, M., PRAGER, E. M., FELDMANN, R. J. & MAINHART, C. R. (1982). *J. Immunol.* **128**, 314–322.

SUSSMAN, J. L. (1985). *Methods Enzymol.* **115**, 271–303.

TOLLIN, P. (1969). *J. Mol. Biol.* **45**, 481–490.