Studies of Monoclinic Hen Egg-White Lysozyme.† IV. X-ray Refinement at 1.8 Å Resolution and a Comparison of the Variable Regions in the Polymorphic Forms

S. T. Rao and M. Sundaralingam*

Laboratory of Biological Macromolecular Structure, 012 Rightmire Hall, Departments of Chemistry and Biochemistry, The Ohio State University, 1060 Carmack Road, Columbus, OH 43210, USA

Abstract

Monoclinic crystals of hen egg-white lysozyme (E.C. 3.2.1.17, HEL) grown at low pH in the presence of $NaNO_3$ belong to space group $P2_1$ with unit-cell dimensions, $a = 28.0$, $b = 62.5$, $c = 60.9$ Å and $\beta = 90.8°$ with two molecules in the asymmetric unit. 1.8 Å resolution intensity data, collected on a CAD-4 diffractometer, contained 17 524 reflections with $F > 3\sigma$ (93% complete). Our earlier preliminary 1.8 Å model was refitted and refined using *X-PLOR* to an R value of 0.189. The deviations in the model from ideal geometry are 0.013 Å in bond lengths and 2.8° in bond angles. The r.m.s. deviation in the backbone atoms between the two molecules is 0.42 Å. A comparison of HEL in different polymorphic crystal forms reveals that the prominent structural variability among them resides in two exposed regions 45–50 and 65–73 which are also regions of lattice contacts.

1. Introduction

Hen egg-white lysozyme (HEL) has been crystallized in several crystal forms (Steinrauf, 1959). Since the first determination of the structure of the tetragonal form (Blake *et al.*, 1965), the structures in the triclinic, monoclinic and orthorhombic forms have also been determined. Except for the orthorhombic form which has been studied only at low resolution (Artymiuk, Blake, Rice & Wilson, 1982), high-resolution structures of the tetragonal (Diamond, 1974; Kundrot & Richards, 1987), triclinic (Ramanadham, Sieker & Jensen, 1990) and monoclinic (Yu, Rao & Sundaralingam, 1989) crystal forms have been determined. We grew the monoclinic crystals in the presence of $NaNO_3$ and identified two closely related crystal forms which grew in the same vial from unbuffered solutions. Subsequently we found that one form grew at pH < 6.8 (low pH form or M1) and the other at pH > 7.0 (high pH or M2 form), each containing two molecules in the asymmetric unit (Hogle *et al.*,

1981). We have studied the M1 form, which diffracted to higher solution than the M2 form. The structure of M1 at 4 Å resolution was determined by MIR method (Hogle *et al.*, 1981) and initially refined at 2.5 Å resolution (Rao, Hogle & Sundaralingam, 1983). The results of the preliminary refinement at 1.8 Å resolution have been reported (Yu *et al.*, 1989). Recently, monoclinic crystals at high pH and high temperature have been obtained (Harata, 1994). The crystals were grown from solutions containing NaCl and 5%(v/v) 1-propanol at 313 K. This form (which we call M2') is reminiscent of our high pH M2 form, which was grown in the presence of $NaNO_3$ at room temperature and has very similar unit-cell dimensions and two molecules in the asymmetric unit. In both M1 and M2' forms, the two molecules are related by a pseudo *B*-centering, the pseudo symmetry being stronger in M2'. The structure of the M2' form has been determined and refined at 1.72 Å resolution (Harata, 1994). A third monoclinic form of HEL (a low-humidity form, which we call M3) containing one molecule in the asymmetric unit has been studied at 1.75 Å resolution (Madhusudan, Kodandapani & Vijayan, 1993). The crystals of M3 were formed by growing crystals of the M1 form and the enclosing the crystals in an environment of reduced humidity. The transformation to the M3 form was complete in about 15–20 h. The two molecules related by a pseudo *B*-centering in the M1/M2' forms are related by an exact lattice translation in the M3 form. In addition to describing the details of completion of the refinement at 1.8 Å resolution using *X-PLOR* (Brünger, 1992a), a comparison of the seven HEL molecules in the different polymorphic crystal forms, two each in M1 and M2' forms, one each in M3, tetragonal and triclinic forms, has shown the correspondence between variable regions of the molecule and the regions of lattice contacts.

2. Experimental

Crystallization conditions and the intensity-data collection strategies (Hogle *et al.*, 1981; Yu *et al.*, 1989) have been described previously. Two crystals were used to collect the intensity data between 2.5 and 1.8 Å resolution using a Enraf–Nonius CAD-4 diffractometer

---

Table 1. *Crystal data and refinement statistics*

Crystal system, space group	Monoclinic, $P2_1$
Unit-cell constants (Å, °)	$a = 28.0, b = 62.5,$
	$c = 60.9, \beta = 90.8$
No. of molecules in asymmetric unit	2
Intensity data collecton	CAD-4 diffractometer
Temperature (K)	293
Resolution (Å)	1.8
No. of reflections used in the refinement $F > 3\sigma$ (% completeness)	17524 (93)
Program used for refinement	X-PLOR, Version 3.0
Parameter file	param19.pro
Final R value	0.189
Model	
Protein residues	1–129 (molecule A),
	201–329 (molecule B)
Water molecules	111
Nitrate ions	6
Positional error (Å)	0.18
R.m.s. deviations from ideal values	
Bond lengths	0.013
Bond angles (°)	2.8
Dihedral angles (°)	25.1
Improper angles (°)	2.2
$\langle B \rangle$ for protein atoms (Å^2)	
Molecule A	20.0
Molecule B	18.3
$\langle B \rangle$ for water molecules (Å^2)	32.5
R.m.s. B values and target (Å^2)	
Backbone (bonded)	1.4/5.0
Backbone (angle)	2.1/5.0
Side chain (bonded)	1.6/5.0
Side chain (angle)	2.4/5.0

at room temperature (293 K). 150 scaling reflections collected on each crystal were used to scale the data sets and the $R_{merge}(F)$ values were in the range 0.018–0.025. A total of 17524 reflections in the resolution range 8–1.8 Å had $F > 3\sigma(F)$ (93% of total) and were used in the present refinement studies (Table 1).

3. Refinement

The preliminary refinement studies at 1.8 Å resolution were carried out using PROLSQ (Hendrickson, 1985). This model has now been further improved by refitting into minimum bias (Read, 1986) cumulative omit maps (Bhat & Cohen, 1984) and by refinement using the X-PLOR program (Brünger, 1992a). Three rounds of refitting the starting model followed by conjugate-gradient refinement using X-PLOR resulted in an R value of 0.25. Solvent sites were selected for the next round of refinement if the difference electron density was $> 3\sigma$, and in the omit map the density was $> 1\sigma$ and the site was within 3.4 Å of either a polar atom of the protein or an already characterized solvent. Those that refined to B values > 70 Å^2 were eliminated from the model. The free R factor (Brünger, 1992b) was monitored using 10%

of the reflections during the solvent selection and it fell smoothly from 0.295 to 0.264 as the solvents were added. A total of 111 solvents and six nitrate ions were included in the final model and the R value was 0.189. The coordinates have been deposited with the Protein Data Bank (Bernstein *et al.*, 1977).*

4. Results and discussion

The stereochemical quality of the model and details of the refinement are contained in Table 1. The r.m.s. deviations in bond lengths and bond angles from ideal values are 0.013 Å and 2.3°, respectively. σ_A (Read, 1986) as well as Luzzati (1952) plots indicate an estimated coordinate error of 0.18 Å. In general, the omit electron density for residues of molecule B was stronger than for those of molecule A. The electron density was consistently clear for the backbone in both molecules, except for the relatively weak electron densities over residues S72, A82, S85, D87, R125, C127, R128 and L129 in molecule A, and T247, G249, T269, N303, N3134, D319, L329 in molecule B. The electron density over the side chains were also uniformly strong, except for a few side chains where the electron density beyond the C_β atom was weak: R21, N41, R45, D48, R68, S72, N113, A122, C127, R128, L129 (in molecule A), N244, R245, N246, R261, R268, T269, S272, R273, D319, N321, C327, R328, L329 (in molecule B). These residues are located in the exposed loop regions and at the C-termini of the molecules.

A plot of the average B values for the backbone atoms in the two molecules, against the residue number is shown in Fig. 1. The average thermal parameters for
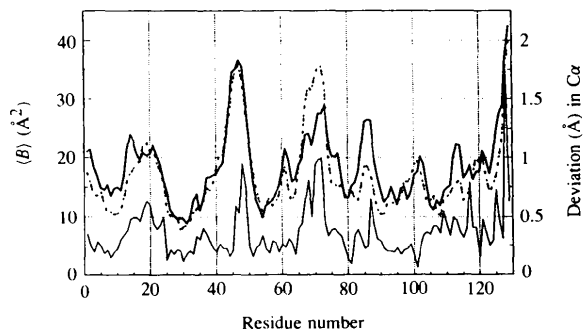
Fig. 1. Plot of the average thermal parameters for backbone atoms (molecule A, solid line and molecule B, dashed line) against residue number. Deviations of corresponding C_α atoms between the two molecules are also shown at the bottom of the figure (thin solid line). Note that regions of large deviations are also regions of higher $\langle B \rangle$ or mobility.

backbone atoms and all atoms in the two molecules are 18.6, 20.0 Å^2 for molecule A and 17.1 and 18.3 Å^2 for molecule B, respectively. The trend in the variation of $\langle B \rangle$ is similar in the two molecules and, as expected, the regions with high $\langle B \rangle$ values correspond to the residues with weak/poor electron densities. Structures of HEL in different polymorphic crystal forms exhibit a similar behaviour in their $\langle B \rangle$ values.

The r.m.s. deviations between the two molecules in the present study, the two molecules in our earlier 2.5 Å study, the two molecules of the M2′ form, the low-humidity M3 form, the tetragonal and the triclinic structures are tabulated in Table 2. The r.m.s. deviations in the backbone atoms range from 0.42 to 0.83 Å, with the two molecules in the present study being most similar. The largest changes in the present model from our previous 2.5 Å model are in the N-terminal residues 1–3, the loop residues 47–54, 70–71, 101–102 and the C-terminal residues 127–129, all in regions of high $\langle B \rangle$. The present high-resolution refinement has made the two

molecules more similar than in our earlier 2.5 Å analysis. The deviations in the corresponding C_α atoms of the two molecules of M1 are also shown in Fig. 1. It is seen that the regions exhibiting large deviations (structural variability) also have large $\langle B \rangle$ values (large mobility). A superposition of the C_α trace of the two molecules of M1 is shown in Fig. 2(a) and a superposition of all seven HEL structures in Fig. 2(b).

The deviations between the two molecules of M1 form are prominent in two loop regions 45–50 (Fig. 3a) and 62–73 (Fig. 3b). The region 46–49 connects the first two strands of the antiparallel β-sheet structure, forming part of the active-site cleft along with the region of residues 100–120. These two regions come together by a hinge motion when a substrate analog is bound. W62 residue plays an important role in substrate binding, with the indole moiety being parallel to pyranose sugar residue (Blake, Mair, North, Phillips & Sarma, 1967; Strynadka & James, 1991). The orientation of the indole group of W62 is opposite in the two molecules of M1, with
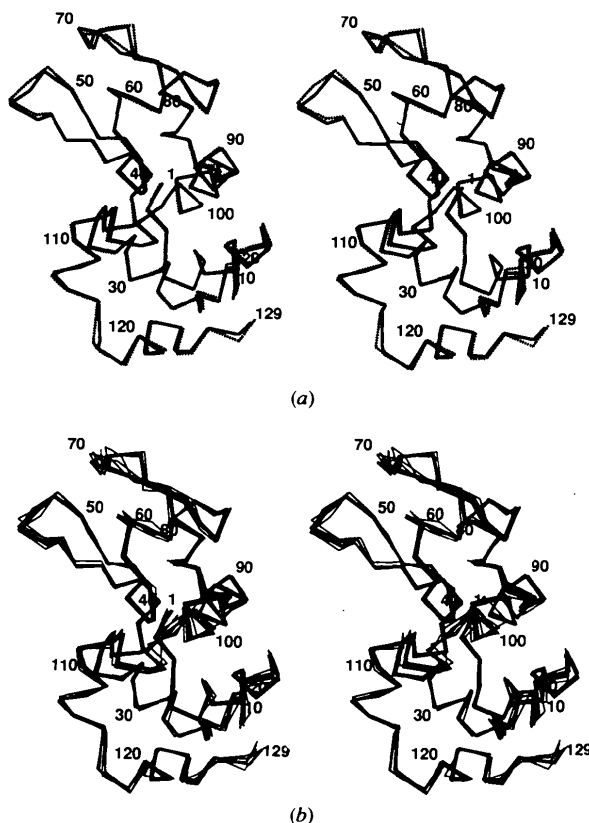

(a)

(b)

Fig. 2. (a) Stereo diagram of the superposition of the C_α traces of molecule A (solid line) and molecule B (dashed line). The r.m.s. deviation between the C_α atoms is 0.42 Å. Notice the large deviations near the N- and C-termini and in the two loop regions 45–50 and 62–73. (b) superposition of the C_α trace of seven HEL molecules. The r.m.s. atomic deviations between any two models lie in the range 0.42–0.73 Å. In addition to the two loop regions in (a) the loop region 100–104, after the third helix is also variable.

(a)

(b)
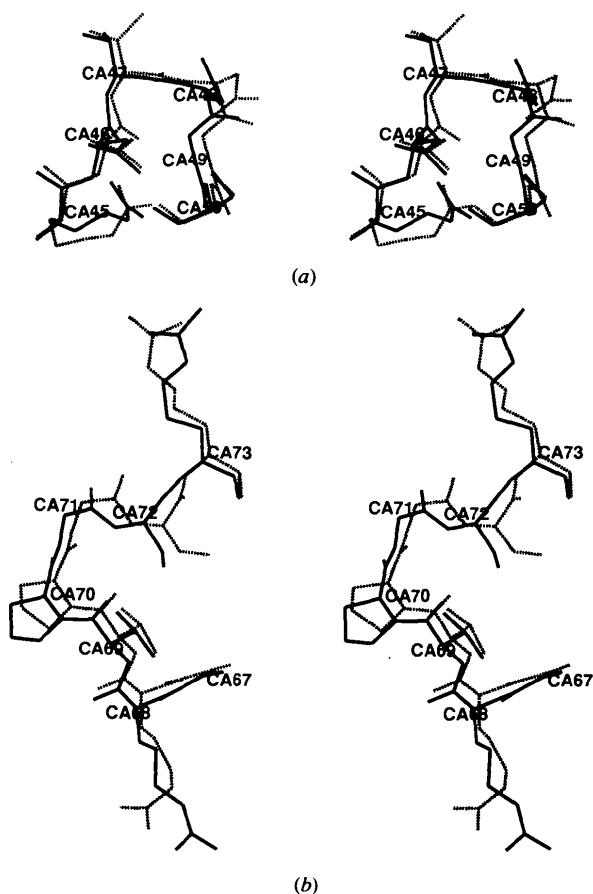
Fig. 3. Stereo figures showing the two regions of largest deviations between the two molecules. Molecule A is drawn with solid lines and molecule B is drawn with dashed lines. (a) Region 45–50 around the type I turn between the first two strands of the β-sheet region. (b) Loop region 67–73 around P70 an G71.

Table 2. *R.m.s. deviations (in Å) between different HEW lyzosyme structures*

Upper triangle contains r.m.s. deviations for backbone atoms and the lower triangle for the Cα atoms.

Molecule	PDB code	MolA M1	MolB M1	MolA M1 (2.5 Å)	MolB M1 (2.5 Å)	MolA M2'	MolB M2'	Mono M3	Tetra-gonal	Tri-clinic
M1-Mol A	*	—	0.42	0.66	0.70	0.53	0.48	0.49	0.45	0.67
M1-Mol B	*	0.42	—	0.73	0.63	0.65	0.56	0.55	0.54	0.68
M1-Mol A (2.5 Å)	1LYM	0.51	0.62	—	0.83	0.68	0.70	0.74	0.62	0.83
M1-Mol B (2.5 Å)	1LYM	0.63	0.55	0.76	—	0.84	0.71	0.87	0.61	0.92
M2'-Mol A	1LYS	0.49	0.60	0.62	0.80	—	0.63	0.58	0.59	0.63
M2'-Mol B	1LYS	0.42	0.48	0.62	0.65	0.63	—	0.63	0.49	0.78
M3-Mono	1LMA	0.42	0.51	0.62	0.73	0.71	0.58	—	0.59	0.55
Tetragonal	2LYZ	0.39	0.51	0.54	0.60	0.59	0.49	0.50	—	0.73
Triclinic	2LZT	0.63	0.64	0.73	0.85	0.63	0.78	0.49	0.65	—

* Present study. The PDB codes and references for others are: M1(2.5 Å) (1LYM, Rao *et al.*, 1983), M2' (1LYS, Harata, 1994) M3 (Madhusudan *et al.*, 1993), tetragonal (2LYZ, Diamond, 1974) and triclinic (2LZT, Ramanadham *et al.*, 1990).

Table 3. *Comparison of backbone torsion angles (°)*

For PDB codes and references of different structures, see footnote to Table 2.

Residue	M1-MolA φ	M1-MolA ψ	M1-MolB φ	M1-MolB ψ	M2'-MolA φ	M2'-MolA ψ	M2'-MolB φ	M2'-MolB ψ	M3 φ	M3 ψ	Tetragonal φ	Tetragonal ψ	Triclinic φ	Triclinic ψ
Arg45	−74	128	−69	131	−84	123	−85	135	−96	120	−88	135	−93	116
Asn46	−96	163	−112	144	−103	179	−95	154	−83	170	−101	172	−94	157
Thr47	−61	−36	−23	−47	−63	−59	−51	−32	−75	−19	−69	−20	−70	−13
Asp48	−45	−23	−86	20	−85	44	−81	7	−97	19	−94	7	−85	2
Gly49	93	−4	77	13	72	6	94	−8	87	−11	102	−23	96	−15
Ser50	−90	164	−107	153	−88	167	−93	164	−91	158	−65	161	−88	158
Gly67	67	12	71	−1	79	−3	73	−7	76	2	62	6	74	−1
Arg68	−131	14	−139	9	−133	10	−116	26	−134	15	−136	23	−123	1
Thr69	−106	116	−99	129	−94	126	−142	151	−110	130	−120	102	−101	126
Pro70	−59	132	−56	133	−47	−52	−17	−130	−66	152	−51	−45	−64	143
Gly71	60	40	64	25	−127	58	−108	68	58	31	−55	−33	74	23
Ser72	−84	144	−60	106	−80	148	−72	176	−58	131	−28	120	−56	127
Arg73	−106	177	−78	169	−117	162	−106	175	−113	−16	−117	143	−105	−21

$\chi_2 = -88$ and 104°, respectively. This is similar to that found for the two molecules of the M2' form. The side chain of W62 appears to be highly flexible in the absence of substrate and the flexibility is probably needed to bind the substrate and to release the products. On the other hand, the indole ring of the adjacent W63, not contacting the substrate sugar ring, has the same orientation in all the structures. The loop region N65 to R73 is most variable and the residues T69 and P70 show large deviations in the backbone torsion angles (Table 3). The region S100–G104, subsite *A* for substrate binding, is mobile in both the molecules of M1 but their structures are quite similar. This is in contrast to the two molecules of the M2' form which have different structures. The structures of the polymorphic forms show that this is also a variable region (Fig. 2*b*). NMR studies on HEL (Smith, Sutcliffe, Redfield & Dobson, 1993) have shown that the well defined regions of the structure in solution, with small deviations from the mean structure, correspond to regions of low variability in the polymorphic crystal forms. Relatively large r.m.s. deviations from the NMR mean structure are found in the regions 46–49 and 68–70, similar to the large variability seen in the polymorphic HEL crystal structures. The NMR study also found the side chain of W62 to be totally disordered.

The intermolecular distances (< 3.2 Å) in M1 between the polar atoms of the protein are listed in Table 4. Though there are some differences in the hydrogen-bonding schemes involving the two molecules, their environments are quite similar, due to the pseudo-symmetry in the crystal. It is interesting that in all the crystal forms – three monoclinic, triclinic and tetragonal – the regions of variability and higher ⟨*B*⟩ are also regions of intermolecular contacts. The intermolecular contacts generally reduce the mobility of the residues involved. In the polymorphic crystal forms of HEL the molecular ions involved in intermolecular crystal contacts are the same; perhaps the polymorphic crystal forms arise from the subtle changes in the intermolecular interactions of the loop regions. H15 and E35 have been

Table 4. *Intermolecular hydrogen bonds in monoclinic HEW lyzosyme with distances < 3.2 Å*

Residue numbers for molecule A are 1–129, and for molecule B 201–329.

Atom 1 Residue 1		Atom 2 Residue 2		Distance (Å)	Symmetry	Atom 1 Residue 1		Atom 2 Residue 2		Distance (Å)	Symmetry
NH2	R5	O	D101	3.1	2, 0 0 1	NH2	R205	O	D301	2.9	2, 1 0 2
NH2	R5	OD2	D101	3.0	2, 0 0 1						
N	C6	OD1	N103	2.7	2, 0 0 1						
O	G16	NH1	R114	2.7	1,-1 0 0						
OG1	T47	NZ	K97	3.0	1, 1 0 0						
						NH2	R261	OD2	D319	3.0	2, 1-1 2
OD1	D101	NH2	R125	3.1	2, 0-1 1	OD2	D301	NH2	R325	3.0	2, 1-1 2
OD2	D101	NE	R125	2.5	2, 0-1 1	OD1	D301	NE	R325	3.0	2, 1-1 2
						ND2	N303	O	G326	2.9	2, 1-1 2
O	N106	NH1	R128	2.5	2, 0-1 1						
						NH1	R312	O	R328	2.9	2, 1-1 2
						NH2	R312	O	R328	2.7	2, 1-1 2
OD1	N19	NE2	Q241	3.2	1,-1 0 0						
OG	S24	NE2	Q241	3.0	1,-1 0 0						
NH2	R68	O	G221	3.1	1, 0 0-1						
OG	S81	O	N313	2.9	1,-1 0-1						
OD2	D87	NH2	R273	2.8	2, 0 0 1						
N	D119	OD2	D287	3.0	1, 0 0 0						

Symmetry code: 1, x, y, z; 2, $-x, 0.5 + y, -z$, followed by lattice translations in the three directions.

shown to be involved in the pH-dependent self association of lysozyme in solution (Shindo, Cohen & Rupley, 1977). These residues in both the molecules in M1 and the M2′ crystal forms are very similar with no direct intermolecular contacts within 3.2 Å, suggesting no direct influence of these residues in the formation of the two pH-dependent M1 and M2′ crystal forms.

A total of 111 solvent sites (water molecules) were identified with ⟨B⟩ of 32.5 Å². 20 protein backbone amide N atoms and 30 backbone carbonyl O atoms take part in water interactions in molecule A and 16 and 33 sites in molecule B, respectively. The corresponding numbers for the N and O atoms in the side chains are ten and 24 for molecule A and 11 and 18 for molecule B,
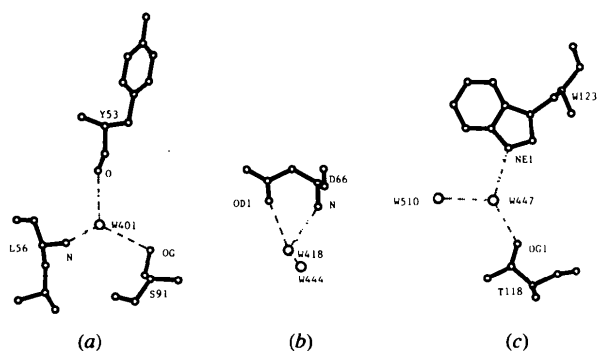

Fig. 4. Three hydration sites common to the two molecules. (a) W401 bridges the carbonyl O atom of Y53 and amide N atom of L56 at a turn and also hydrogen bonds to the side chain OG of S91. (b) W418 bridging the carbonyl O atom and side chain OD1 of D66 and also interacting with W444. (c) W447 forming hydrogen bonds to side-chain atoms OG1 of T118 in a turn and NE1 of W123 of the C-terminal helix and also to W510.

respectively. As found in other proteins (Baker & Hubbard, 1984) the water interactions with protein O atoms are more numerous than with protein N atoms. In other words, the hydrogen bonds involving the protein atoms are probably stronger when the water molecule participates as a donor rather than as an acceptor. Of the 111 water molecules, 94 are in the first hydration shell, nine in the second shell and the remaining in the third shell. Of the 94 waters in the first shell, 52 interact with more than one residue on the same molecule while seven are involved in forming water bridges between neighboring molecules. 24 water molecules associated with molecule A are superposable on the corresponding sites in molecule B within a distance of 1.25 Å and an average deviation of 0.38 Å. 22 of these water molecules form intramolecular water bridges and stabilize the structure (Fig. 4). The corresponding 24 water sites on molecules A and B of M2′ and on the HEL molecule in M3 have average deviations of 0.42, 0.43 and 0.52 Å respectively. Many of these sites therefore may be expected to be also occupied in solution.

The transformation relating the coordinates of the two molecules in M1 form is,

$$A = \begin{bmatrix} 0.9738 & -0.1931 & 0.1201 \\ 0.2055 & 0.9734 & -0.1017 \\ -0.0972 & 0.1237 & 0.9875 \end{bmatrix} B + \begin{bmatrix} -14.38 \\ 4.35 \\ -31.36 \end{bmatrix}$$

This corresponds to an orientational difference of 14° between the two molecules (Table 2) and translations of 14.38 and 31.36 Å parallel to the a and c axes. The translations are very close to $a/2$ (14.0 Å) and $c/2$ (30.45 Å), respectively, but the orientations of the two molecules are significantly different. In the high pH M2′

form, the orientational difference of the two molecules is only 4° and the pseudo *B* centering is stronger. In the M3 form, the pseudo *B* centering becomes an exact lattice translation. It is interesting that while the pseudo centering becomes increasingly prominent as one proceeds from M1⟶M2'⟶M3 monoclinic crystal forms and the solvent volume content in the crystals decreases (34, 32 and 24%, respectively), more of the solvent is ordered.

References

Artymiuk, P. J., Blake, C. C. F., Rice, D. W. & Wilson, K. S. (1982). *Acta Cryst.* B**38**, 778–783.

Baker, E. N. & Hubbard, R. E. (1984). *Prog. Biophys. Mol. Biol.* **44**, 97–179.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rogers, J. B., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bhat, T. N. & Cohen, G. H. (1984). *J. Appl. Cryst.* **17**, 244–248.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London)*, **206**, 757–763.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. R. Soc. London Ser. B*, **167**, 365–377.

Brünger, A. L. (1992a). *X-PLOR 3.0 Manual.* Yale University, New Haven, CT, USA.

Brünger, A. L. (1992b). *Nature (London)*, **355**, 472–474.

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

Harata, K. (1994). *Acta Cryst.* D**50**, 250–257.

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

Hogle, J., Rao, S. T., Mallikarjunan, M., Beddell, C., McMullan, R. K. & Sundaralingam, M. (1981). *Acta Cryst.* B**37**, 591–597.

Kundrot, C. E. & Richards, F. M. (1987). *J. Mol. Biol.* **193**, 157–170.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Madhusudan, Kodandapani, R. & Vijayan, M. (1993). *Acta Cryst.* D**49**, 234–245.

Ramanadham, M., Sieker, L. & Jensen, L. H. (1990). *Acta Cryst.* B**46**, 63–69.

Rao, S. T., Hogle, J. & Sundaralingam, M. (1983). *Acta Cryst.* C**39**, 237–240.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Shindo, H., Cohen, J. S. & Rupley, J. A. (1977). *Biochemistry*, **16**, 3879–3882.

Smith, L. J., Sutcliffe, M. J., Redfield, C. & Dobson, C. M. (1993). *J. Mol. Biol.* **229**, 930–944.

Steinrauf, L. K. (1959). *Acta Cryst.* **12**, 77–79.

Strynadka, N. C. & James, M. N. G. (1991). *J. Mol. Biol.* **220**, 401–424.

Yu, C., Rao, S. T. & Sundaralingam, M. (1989). *The Immune Response to Structurally Defined Proteins: The Lysozyme Model*, edited by S. Smith-Gill & E. Sercarz, pp. O25–O38. New York: Adenine Press.