Acta Cryst. (1998). D**54**, 767–779

Structures of Monoclinic Lysozyme Iodide at 1.6 Å and of Triclinic Lysozyme Nitrate at 1.1 Å

L. K. STEINRAUF†

Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis IN 46202-5122, USA, and Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan 80424, Republic of China on Taiwan. E-mail: lsteinra@mail.nsysu.edu.tw

(*Received 12 March 1997; accepted 14 November 1997*)

Abstract

Hen egg-white lysozyme is one of the most thoroughly studied of enzymes and has been the subject of study by many methods, including X-ray crystallography. The present work extends the X-ray crystallography to higher resolution, includes the positions of the anions, and examines the contacts of the neighbors in greater detail. Data were collected at room temperature on a Rigaku R-axis area detector with rotating-anode X-ray generator to 1.6 Å resolution for monoclinic lysozyme iodide at pH 4.0, to 1.8 Å for monoclinic lysozyme iodide at pH 8.0, and to 1.1 Å resolution for triclinic lysozyme nitrate at pH 4.5. The structures have been refined by *SHELX*93 with the expected number of anion sites being accounted for. Two regions of the protein have been found to be variable: residues 65–75 and 99–104. Except for 65–75 and 99–104, lysozyme is a very stable molecule with the crystal forms being held together by the electrostatic contacts of the anions and by layers of water molecules. The anion positions can be described as paired half sites, each half being contributed by a different lysozyme molecule. The many different crystal forms of lysozyme may be due to different combinations of the many such half sites on the surface. A hypothesis is presented for lysozyme in the different crystal forms and which may be extended to behavior in solution. Suggestions for future crystallographic research are proposed, involving anions of different shape and charge.

1. Introduction

Hen egg-white lysozyme was the third protein and the first enzyme to be determined by X-ray crystallography (Blake *et al.*, 1962; Stanford *et al.*, 1962; Dickerson *et al.*, 1962). The first determination was completed by the group of Phillips & Blake (Blake *et al.*, 1965; Johnson & Phillips, 1965) using the method of isomorphous replacement on tetragonal crystals grown at pH 4.7 in 5% sodium chloride. Crystal structure determinations (Blake *et al.*, 1967; Ford *et al.*, 1974; Perkins *et al.*, 1978) of the tetragonal form of lysozyme with substrates in place have identified the binding sites of three (A, B and C) of the six sugar residues of the substrate. Some flexibility of tetragonal lysozyme was found with the binding of substrates (Kelly *et al.*, 1979). The influence of urea at 1.9–1.5 Å resolution in the presence of up to 9 *M* urea was studied by Pike & Acharya (1994). Among the findings were peptide chain differences in the 69–74 region and the 100–103 region, including a peptide flip between Arg73 and Asn74. The triclinic form of lysozyme at pH 4.5 and 2% sodium nitrate was first crystallized by Steinrauf (1959). The structure was solved with molecular replacement by Joynson *et al.* (1970), and refined to 2.5 Å resolution by Moult *et al.* (1976).

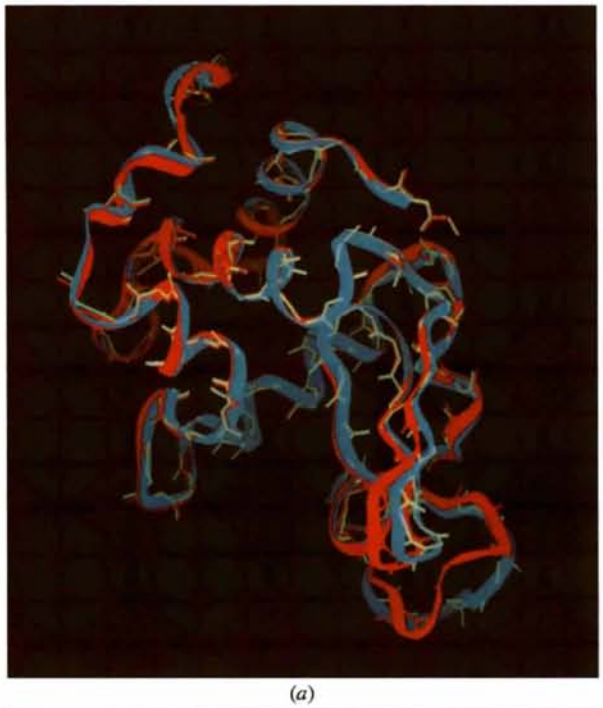
The triclinic structure, including the binding of a trisaccharide molecule (sites D, E and F), has since been refined to 1.5 Å resolution by the group of Jensen (Kurachi *et al.*, 1976; Ramanadham *et al.*, 1981, 1990; Hodsdon *et al.*, 1990).

The monoclinic form, which has two independent lysozyme molecules in the asymmetric unit, was first described by Crick (1953) and again by Steinrauf (1959). Crystals were initially grown from pH 4.7 and 3% sodium nitrate. The structure, solved by isomorphous replacement, was reported by the group of Sundaralingam, first at 4 Å resolution, then at 2.5 Å, and finally at 1.8 Å (Hogle *et al.*, 1981; Rao *et al.*, 1983; Yu *et al.*, 1989; Rao & Sundaralingam, 1996). This was the first observation of two independent conformations of lysozyme in the same unit cell.

A low-humidity form of monoclinic lysozyme nitrate, which had also been noticed before, was given the well deserved attention by Madhusudan *et al.* (1993), who solved the structure at 1.75 Å resolution. In this modification the two independent molecules have assumed crystallographic identity with one molecule in the asymmetric unit. The structure was found to be well defined with average *B* values for the main-chain atoms, side-chain atoms, and 172 water sites to be 8.1, 13.8, and 29.2 $Å^2$, respectively. Two nitrate sites were found. One was hydrogen bonded to the side chain of Arg14, the N atom of Ser72, and *via* a water to the O atom of Ser72, and which may be similar to iodide 10 of monoclinic

† Current address: the Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung, Taiwan 80424, Republic of China on Taiwan.

lysozyme iodide as reported here. The second was hydrogen bonded to the N atoms of residues 5, 6 and 7; *via* a water molecule to the N atom and the guanidium of Arg5; and to the ND1 atom of Asn106.



(*a*)

(*b*)

Fig. 1. Ribbon drawings of lysozyme A and lysozyme B superimposed. The two molecules coincide except for the 65–75 loop.

A variation of the monoclinic structure has been noticed by several investigators. Harata (1994) has now investigated this at 1.7 Å using crystals from 10% sodium chloride and 5% 1-propanol at pH 7.6. He observed differences in the two independent lysozyme molecules 45–50 region, as well as the 67–73 and the 99–104 regions.

Room-temperature and low-temperature studies were conducted on both monoclinic lysozyme nitrate and tetragonal lysozyme chloride by Young & Dewan (1993) at 298 K and also at 100 K. Data were taken to 1.9 Å resolution for all cases. For both the monoclinic and tetragonal forms, refinement of the structures gave approximately twice as many water molecules from the low-temperature data, 406 at low temperature and 191 at room temperature for the monoclinic, and 237 at low temperature and 100 at room temperature for the tetragonal. In all crystal forms the molecule of lysozyme retains its form quite well, with the helices, the turns, and the beta strands intact. A ribbon drawing is shown in Fig. 1 with the two independent molecules of monoclinic lysozyme iodide being superimposed.

2. Materials and methods

2.1. *Crystallization*

Lyophilized salt-free hen egg-white lysozyme from Worthington Biochemicals was used without further purification. Batch crystallizations as described by Steinrauf (1959), gave crystals suitable for X-ray diffraction. The crystals of monoclinic lysozyme iodide were obtained from solutions buffered with 100 m*M* acetate buffer, pH 4.7. When data from these crystals proved the feasibility of finding the iodide ions, it was decided to explore the effects of pH. Accordingly, data were taken from two new crystallizations. The initial pH of the solutions was adjusted to (1) 4.0 with 1.0 *M* hydriodic acid, and (2) to pH 8.0 with 1.0 *M* sodium hydroxide. The initial protein concentration was 10 mg ml^{-1} and the concentration of sodium iodide was 5%(*w*/*w*). A spot of toluene (about 0.01 ml) was placed on the cover lid of the crystallization vessel to discourage the growth of mold. The characteristic blade-shaped crystals of monoclinic lysozyme appeared spontaneously after 2–4 d. The data from crystals at pH 4.7 were not used further.

Triclinic crystals were grown in the presence of bromophenol blue. Monoclinic crystals had first grown in batch at pH 4.5 and 5% sodium nitrate. To these crystals a small amount of bromophenol blue was added to see if the crystals would be stained. After several months at room temperature it was observed that the monoclinic crystals were not stained at all and were dissolving, but very dark blue crystals were growing. These crystals were triclinic.

Table 1. *Data-collection information and statistics for monoclinic lysozyme iodide and triclinic lysozyme nitrate*

One single-crystal was used for each of the lysozyme iodide data. Three crystals were used for the triclinic data. The crystal symmetry of the iodide is monoclinic of space group $P21$ with two molecules of lysozyme per asymmetric unit. The triclinic crystals were indexed by the *MSC* program as pseudo $A1$ with dimensions 31.97 (1), 65.08 (3), 27.25 (3) Å, 94.57 (5), 111.56 (5), 82.98 (4)°, but refined as exact $A1$. After refinement the results were returned to the standard $P1$ triclinic lysozyme unit cell (Steinrauf, 1959). The standard deviations are based on measurements on three crystals. Data reduction was carried out using the R-axis software supplied by MSC.

Crystal form	Monoclinic lysozyme iodide		Triclinic lysozyme
	pH 4		pH 8
Number of crystals used	1	1	3
Number of frames	60	60	3 × 100 each
Exposure per frame (min)	20	20	20
1σ	46116	68962	216397
Unique reflections	14257	22327†	36599
R_{merge} (%) Total	8.01	7.61	9.49
Full	7.65	7.22	8.99
Partial	9.10	10.48	11.14
Percentage of possible reflections	68 to 1.8 Å	63 to 1.6 Å	78.8 to 1.1 Å (88.6 to 1.3 Å)
$R+$	4.46	4.88	6.06
$R-$	3.63	4.59	6.27
$R\pm$	7.11	6.56	6.63
a (Å)	27.9424 (7)	27.8936 (4)	27.189 (10)
b (Å)	63.2130 (22)	63.1535 (5)	31.988 (14)
c (Å)	60.2168 (19)	60.2275 (8)	34.250 (18)
α (°)	90	90	88.35 (5)
β (°)	90.0295 (18)	90.1172 (2)	108.66 (4)
γ (°)	90	90	111.57 (6)
Mosaic spread	0.33	0.30	0.30

† 44 071 unique reflections when the Friedel pairs were separated.

2.2. *Data collection and reduction*

2.2.1. *Monoclinic lysozyme iodide.* Monoclinic crystals, approximately 0.2 × 0.3 × 0.5 mm, were mounted in capillary tubes and examined on the Rigaku R-axis II image-plate detector. The crystal-to-plate distance was 82 mm, the rotating-anode generator was operated at 50 kV and 100 mA, and the exposure time was 20 min for 60 frames of 3° each. The number of reflections, maximum resolution, per cent coverage, and R_{merge} for crystals are given in Table 1 for the monoclinic crystals at pH 4 and the pH 8 and for the triclinic crystals.

2.2.2. *Triclinic lysozyme.* Crystals of the triclinic form, about 0.6 mm on edge, were mounted in capillary tubes and examined on the R-axis. Three crystals were used to collect date to 1.1 Å resolution, using a crystal-to-plate distance of 78 mm and a 2θ angle of 48°. After the 180° of oscillations the capillary was repositioned using the 45° jig and data were collected, and yet again with a 90° rotation of the jig. Crystals were very stable for up to 300 h of exposure and could be mounted with very little mother liquid. The water ring on the plates was very light, and the diffraction spots were quite sharp. Small areas of diffuse spots and heavy background were frequently found; the *MSC* (Molecular Structure Corporation, 1993) program was allowed to process the data without any intervention. Although reflections were frequently observed to the edge of the plate at the 48° position, higher settings of 2θ moved the shadow of the beam stop off the plate. The *MSC* processing program required the location of the beam stop.

The *MSC* program initially indexed the data as twice the expected unit-cell volume. The resulting space group was pseudo $A1$ with the $k+l$ odd reflections very weak, only a few being above 3σ. Much later when a precession camera became available, photographs showed no signs of the $k+l$ odd reflections.

After the data collection and reduction, it was discovered that the R-axis had a slight misalignment of the optical and the mechanical 2θ rotations. Because of this and because of the need to observe the shadow of the beam stop when using the data-processing program, no higher settings of 2θ were attempted.

2.3. *Structure determination and refinement*

2.3.1. *Monoclinic lysozyme iodide.* The data from monoclinic lysozyme iodide, pH 8, were refined first using the coordinates of the protein atoms from the structure of monoclinic lysozyme nitrate (Yu *et al.*, 1989) as the initial model. The two independent lysozyme molecules were refined as rigid bodies by *X-PLOR* (Brünger, 1992*a*) using data from 6 to 3 Å resolution. The results were passed to the *Protin-PROLSQ* (Hendrickson, 1985) constrained least-squares refinement program and the *FRODO* system of model rebuilding (Jones, 1985). The iodide ions were immediately obtained from the electron-density maps and added to the model. A considerable amount of rebuilding was necessary, especially in the 64–76 stretch of the second molecule. Reflections were added to the 3 Å data in groups of 2000–3000 at a time as the

Table 2. *Steps in the refinement of monoclinic lysozyme iodide and triclinic lysozyme nitrate*

Monoclinic lysozyme iodide

The estimated r.m.s. error by Luzzati plot is 0.10. Water molecules either isotropic or anisotropic and the occupancy is either fixed or variable. R values are given in per cent. The free R is that of Brünger, in which 10% of the reflections, chosen at random, are always excluded from the refinements. The $R_{4\sigma}$ is the R value for those reflections which are greater than four times the standard deviation. R overall is summed over all reflections, including those used for R_{free}. Exactly the same number of reflections, 44 071, was used for all runs, and the same 4369 reflections were excluded to form R_{free}. The Friedel pairs were averaged in the first run and the average used for both pair. Also, exactly the same reflections were used to obtain the free R for all runs.

Protein	Iso	Iso	Iso	Iso	Aniso	Aniso	Aniso†
Iodide	Iso	Aniso	Aniso	Aniso	Aniso	Aniso	Aniso
Water	Iso/fx	Iso/fx	Iso/fx	Iso/vr	Iso/fx	Iso/vr	Iso/fx
Anomol?	No	No	Yes	Yes	Yes	Yes	Yes
$R_{4\sigma}$	14.62	13.86	13.46	13.14	10.55	9.46	7.98
$R_{overall}$	16.87	16.02	15.67	15.42	12.81	11.78	10.58
R_{free}	19.61	18.56	18.22	18.05	15.77	14.79	14.04

Triclinic lysozyme nitrate

The estimated r.m.s. error by Luzzati plot is 0.07. Refinement by *SHELX*93 with all data to 1.1 Å resolution, finally using isotropic or anisotropic temperature parameter refinement on all atoms. Several residues were assigned positions for disordered atoms, and several other residues were given variable occupancy parameters.

Indexed as	$P1$‡	$P1$	$P1$	$P1$
	Isotropic	Isotropic	Anisotropic	Anisotropic
	Double cell	No disorder	No disorder	Disorder
$R_{4\sigma}$	14.45	14.88	10.70	9.89
$R_{Overall}$ (%)	18.17	15.77	11.81	11.01
R_{free} (%)	19.73	17.81	15.18	14.16

† H atoms added. ‡ The initial refinement assumed the doubled unit cell and two molecules of lysozyme. Thereafter, 3619 reflections were excluded to form R_{free}.

Table 3. *Occupancy and temperature factors for monoclinic lysozyme iodide crystals grown at pH 4 and pH 8*

Iodide	pH 4	pH 8
1	0.74/32	0.63/34
2	0.35/28	0.31/30
3	0.64/22	0.51/26
4	0.50/31	0.33/44†
5	0.35/28	0.27/46
6	0.76/27	0.56/31
7	0.39.32	0.31/34
8	0.80/26	0.74/29
9	0.35/34	0.24/54†
10	0.61/50	0.54/49
11	0.32/31	0.23/28
12	0.39/26	0.23/29
13	0.36/34	0.25/35
14	0.30/51	0.24/46
15	0.67/39	0.69/47
16	0.96/23	0.90/26
17	0.30/45	0.23/46

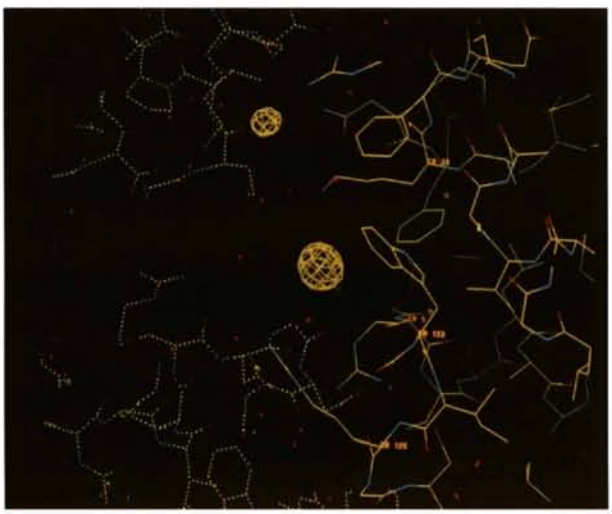
† Associated with His15 or Lys1.

refinement proceeded. Water molecules were added when observed in reasonable positions. Using data to 1.6 Å resolution, the refinement terminated at R = 23% with the iodide ion sites showing large amounts of anisotropy.

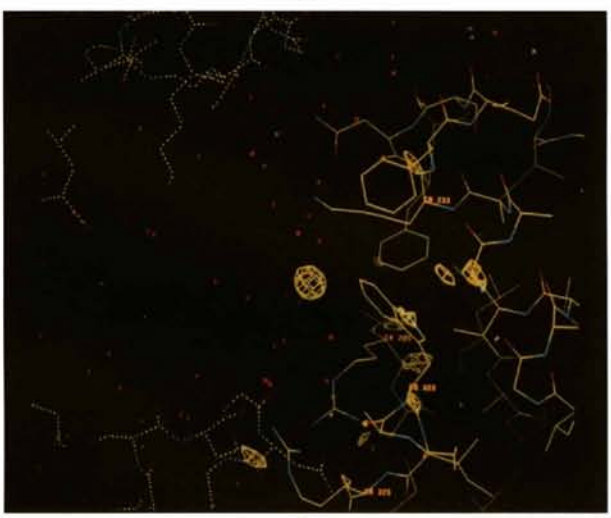
Refinement of the iodide structures was then passed to *SHELX*93 (Sheldrick *et al.*, 1997). With all atoms having individual isotropic temperature factors, the refinement proceeded to R = 16.87% on all data. When the iodide ions were allowed to become anisotropic, the R value was lowered to 16.02%. During the collection of the data the Friedel pairs were measured separately, but up to now the pairs had been averaged. When the Friedel pairs were separated, the R value dropped to 15.67%. Allowing the water molecules to have variable occupancy decreased the R to 15.42%. However, allowing the protein atoms to become anisotropic gave R = 12.81%. Allowing the water molecules to have anisotropic temperature factors dropped the R value to 11.78%, and adding hydrogen gave an R_{all} = 10.58 and $R_{4\sigma}$ = 7.98%. In all cases the free R value paralleled the overall R value; complete details are given in Table 2. One side chain, Lys13, showed more than one atom in doubled position, but that only by 0.6 Å. The estimated root-mean-square error in the coordinates obtained by Luzzati plot (Luzzati, 1952) was 0.10 Å. The two molecules in the independent unit will be referred to as lysozyme A and lysozyme B; the single molecule in the triclinic structure will be called lysozyme T.

When refinement of the data at pH 8 was complete, the results were passed to the data from pH 4. The only important difference was that the iodide ions were more fully occupied. This would be expected since the histidine and N-terminal residues should be more fully ionized at the lower pH. Results are given in Table 3. The positions and the temperature factors of the protein atoms and the water molecules were the same.

A comparison of refinement of the determination by Harata (1994) of the otherwise very similar monoclinic lysozyme chloride/propanol at 1.7 Å shows the effect of the heavy iodide atoms. The ability of the refinement to model the nonspherical shape of the iodide atoms allowed the final R value to become 0.0798 (at 4σ) as compared to 0.187 (at 3σ) for the chloride/propanol structure. That the iodide ions contribute very strongly to the structure will be seen in Fig. 5, the radial distribution of the R value. The iodide ions are modeled very accurately, but have high temperature factors. Thus, the contribution is lost around 2 Å, after which the protein atoms dominate the R value which increases strongly. The radial distribution of the triclinic structure, which has no heavy atoms, is more regular.



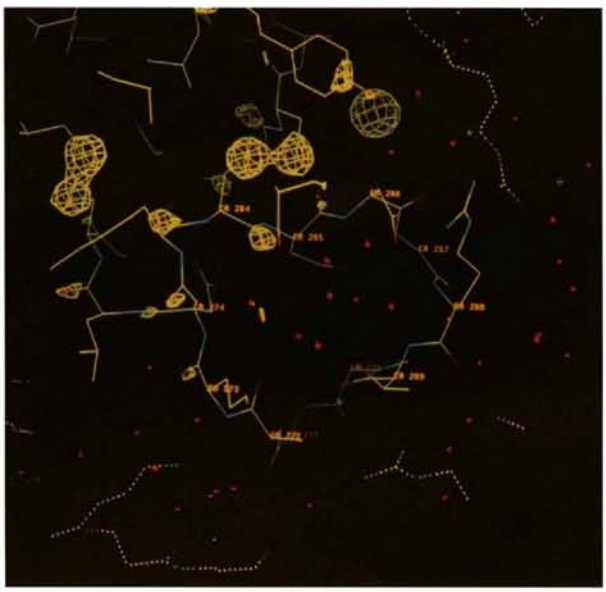
(*a*)



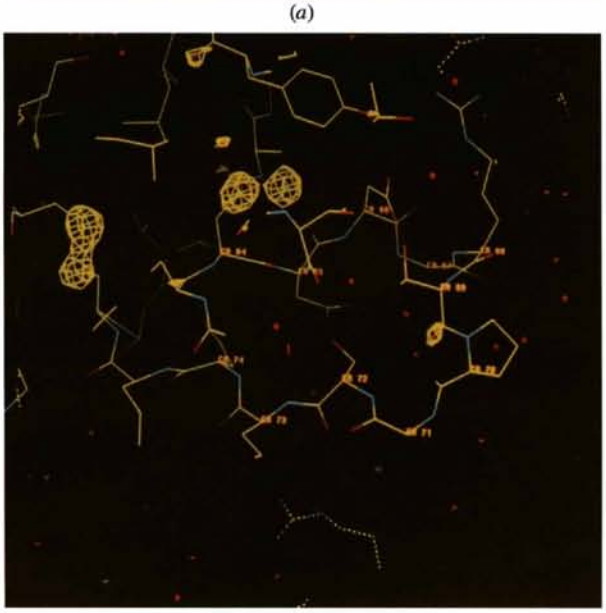
(*b*)

Fig. 2. (*a*) Lysozyme A (solid lines) in the vicinity of iodide 3. (*b*) Lysozyme B (solid lines) in the vicinity of iodide 7 which is the equivalent of iodide 3. Neighbors are in dotted lines. The smaller sphere in Fig. 1(*a*) is iodide 12 which has no counterpart.

2.3.2. *Triclinic lysozyme*. The data-processing program usually gave a unit cell of the triclinic form with $a = 31.97$, $b = 65.08$, $c = 27.24$ Å, $\alpha = 94.97$, $\beta = 111.54$, $\gamma = 82.57°$, in space group $P1$, with exactly twice the volume of the known triclinic cell. Although reflections with $k+l$ odd were very weak, the larger unit cell was accepted and the structure was solved with *X-PLOR* (Brünger, 1992*a*,*b*) by using molecule A of the monoclinic lysozyme iodide, but without the iodide ions. A single orientation was found from the rotation function, and the translation search gave the position of the second molecule at $0,\frac{1}{2},\frac{1}{2}$ from the first. Refinement was initiated



(*a*)



(*b*)

Fig. 3. (*a*) Close-up representation of the 65–75 loop for lysozyme A. (*b*) Close-up representation of the loop for lysozyme B.

Table 4. *Contacts around the anions for monoclinic lysozyme iodide*

Occupancy/B (Å^2)	Distance (Å)	Contact atom of residue (or water)
I-1 short football, between two lysozyme molecules, both having negative charges		
0.6/34.0	3.50	water (weak) which is hydrogen bonded to Arg112 ($x-1, y, z$)
	3.95	peptide of ArgA21
	4.19	OG of SerB81 ($x-1, y, z$)
	4.80	CB of AspA79 ($x-1, y, z$)
	5.50	NE of ArgA21
	5.81	NH2 of ArgA12 ($x-1, y, z$)
I-2 sphere, on lysozyme B but no positive charges nearby, positive charges on neighbor 8–10 Å away.		
0.9/39.4 equivalent to I-13	3.32	OG of SerB24
	3.41	N of GlyB26
	3.64	CG of GlnB121 (weak)
	4.15	CD1 of IluB124
	5.03	water (weak)
	9.65	NZ of LysB13
I-3 sphere, mostly with lysozyme A,		
0.5/26.4 equivalent to I-7	3.75	NZ of LysA33
	3.89	CD of ArgA5
	4.06	NE of ArgA5
	4.18	O of AspA101 ($-x, y+\frac{1}{2}, -z+1$)
	4.23	CD2 of PheA38
	4.45	CZ3 of TrpA123
	7.73	NH2 of ArgA21
I-4 football, between two lysozyme molecules.		
0.3/44.3 equivalent to I-9	2.94	O of GlyB71 ($-x, y+\frac{1}{2}, -z+1$)
	3.31	OD2 of AsnA87
	3.51	N of IluA88
	3.75	CG1 of IluA88
	3.87	CA of SerB72 ($-x, y+\frac{1}{2}, -z+1$)
	4.08	NH1 of ArgA14
	4.00	⊥ CB, CG of AsnA103
	4.22	CD1 of HisA15
	4.59	CZ of PheA3
	6.01	NZ of LysA1
I-5 sphere, between two lysozyme molecules		
0.3/46.2	3.00	Water
	3.19	Water
	3.64	O of AsnA106
	3.79	CB of CysA6 ($-x, y-\frac{1}{2}, -z+1$)
	4.03	OD & ND2 of AsnA103
	5.11	NH1 of ArgA128 ($-x, y-\frac{1}{2}, -z+1$)
I-6 short football, in a pocket made by two lysozyme molecules.		
0.6/30.6 equivalent to I-10	3.51	water
	3.57	NZ of LysA96 ($x+1, y, z$)
	3.72	ND2 of AsnA113
	4.82	O of ValA109
	4.37	CE2 of TyrA20 ($x+1, y, z$)
	4.47	CD of ArgA114
I-7 sphere, in a pocket on lysozyme B, open to water		
0.3/34.0 equivalent to I-3	3.20	water
	3.92	NZ of LysB33
	4.00	CD1 of PheB38
	4.15	CZ3 of TrpB123
	4.38	CD of ArgB5
I-8 doorknob, in a pocket of lysozyme B, but with close contacts with neighboring lysozyme B. I-17 is possibly an alternate site		
0.7/28.5	3.47	N of GlyB126 ($-x+1, y-\frac{1}{2}, -z+2$)
	3.67	NE1 of TrpB63
	4.44	NH2 of ArgB61
	4.52	CG of ArgB125 ($-x+1, y-\frac{1}{2}, -z+2$)
	4.57	CD1 of TrpB62
	4.69	CB of AsnB59
	4.75	CD of ArgB128 ($-x+1, y-\frac{1}{2}, -z+2$)

Table 4. (*cont.*).

Occupancy/B (Å^2)	Distance (Å)	Contact atom of residue (or water)
I-9 short football, pocket made of two lysozyme molecules		
0.2/53.6 equivalent to I-4	3.64	N of IluB88
	3.84	CG1 of IluB88
	4.16	OD1 of AspB87
	4.20	CE1 of HisB15
	4.81	CZ of PheB3
	6.10	NH2 of ArgB14
	6.39	NZ of LysB1
I-10 football, between two lysozyme molecules		
0.5/48.9 equivalent to I-6	3.59	NH1 of ArgB21 ($x+1, y, z$)
	3.95	ND2 of AsnB113
	4.16	NH2 of ArgB114
	4.27	CD1 of TyrB20 ($x+1, y, z$)
	4.32	CD of ArgB114
	4.93	O of AsnB19 ($x+1, y, z$)
I-11 sphere, pocket from lysozyme A with single contact from lysozyme B and neighboring lysozyme B		
0.2/27.9	3.33	OH of TyrA23
	3.62	Water
	3.98	ND2 of AsnA106
	4.05	OD1 of AsnB77
	4.23	CH2 of TrpA111
	4.37	NZ of LysA116
	4.55	NE of ArgB45 ($x-1, y, z$)
I-12 sphere, between two lysozyme molecules with positive charges from both.		
0.2/28.8	3.82	NH1 of ArgA114
	3.95	⊥ CA of ArgA14
	4.50	NZ of LysA13
	4.61	CE2 of PheA34 ($x-1, y, z$)
	5.86	NZ of LysA33 ($x-1, y, z$)
13 sphere, pocket on lysozyme A plus contact with Gln from B′		
0.3/35.1 equivalent to I-2	2.99	NE2 of GlnA121
	3.03	Water (weak)
	3.35	OG of SerA24
	3.39	N of GlyA26
	4.65	OE1 of GlnB41 ($x-1, y, z$)
I-14 doorknob, between two lysozyme molecules		
0.2/45.9	3.44	OH of TyrB23
	3.87	ND2 of AsnB106
	3.87	CH2 of TrpB111
	4.29	NE of ArgA45
	5.11	NZ of LysB116
	5.55	NH1 of ArgA68
I-15 long football, pocket made of two lysozyme molecules		
0.7/46.6 equivalent to I-16	3.60	ND2 of AsnA74
	4.11	⊥ CD of ProA79
	4.22	NZ of LysB116
	4.22	ND2 of AsnA74
	6.12	NH2 of ArgA45 ($x-1, y, z$)
I-16 fat doorknob, pocket of two lysozyme molecules		
0.9/25.6 equivalent to I-15	3.42	NH1 of ArgA112
	3.52	NZ of LysA116
	3.66	ND2 of AsnB74
	3.81	ND2 of AsnB65
	4.10	⊥ CD of ProB79
	4.64	ND2 of AsnB77
I-17 sphere, deep pocket of two lysozyme molecules.		
0.2/46.2	3.56	NE1 of TrpB62
	4.06	NE of ArgB61
	4.12	⊥ CA of ArgB125 ($-x, y-\frac{1}{2}, -z+2$)
	4.35	O of GlnB119′

with 2.0 Å resolution data using *PROLSQ*. Several side chains needed rebuilding and the 100–106 loop was very weak and confused. Data were added in shells of reflections in small increments. Refinement ceased at 1.35 Å and $R = 0.18$ at which point, the addition of additional reflections caused the program to anti-refine. Refinement was now shifted to *SHELX*93, and proceeded smoothly. Using the same weighting factors

Table 5. *Contacts around the anions for triclinic lysozyme nitrate*

Anion/Occ/B (Å^2)		Distance	Receptor(s)
1	O1	2.85	Water
0.8/15.9–20.6		3.78	CB of Pro79
		3.34	N of Cys80
	O3	2.96	NE of Arg21 ($x+1, y, z$)
2	O1	2.98	NZ of Lys116 ($x-1, y-1, z$)
0.8/17.0–27.6		3.32	⊥ N of Ilu78
		3.48	⊥ N of Pro79
	O2	3.02	NZ of Lys116 ($x-1, y, z-1$)
		3.21	⊥ NE of Arg112 ($x-1, y, z-1$)
	O3	3.25	ND2 of Asn65
3	O1	3.39	NZ of Lys33
0.8/26.5–34.2	O2	2.99	water
		3.16	NH1 of Arg73 ($x-1, y, z-1$);
		3.62	CE2 of Phe38
		3.68	CH2 of Trp62 ($x-1, y, z-1$)
	O3	3.08	NH2 of Arg73 ($x-1, y, z-1$)
		3.03	NZ of Lys33
4	O1	2.43	ND2 of Asn106 (weak)
0.5/27.5–30.0		3.25	Water
	O2	2.69	ND2 of Asn106 (weak)
		3.21	N of Met105
		3.29	N of Asn106
	O3	2.49	OH of Tyr23
		3.30	⊥ NE of Arg45 ($x+1, y, z$)
5	O1	2.60	water
0.8/25.9–31.7		2.94	N of Gly26
		3.32	⊥ N of Leu25
	O2	3.32	NE2 of Gln41 ($x+1, y, z$)
	O3	3.43	CA of Gln121
6	O1	3.28	ND2 of Asn114
0.8/45.2–50.8	O2	2.83	water
		3.5	CA-CB of Ala110
	O3	3.4	ND2-CB of Asn113

as had been used in the iodide, the refinement converged for 1.1 Å resolution with $R_{4\sigma} = 0.101$ with all atoms anisotropic, and several side chains, Lys1, Arg21, Thr43, Arg45, Arg61, Ser85, Ser86, Gly102, Asp119, Arg125 and Arg128 showing one or more atoms in doubled positions. Also, several side chains and some of the residues in the 100–106 loop, and residues 128 and 129 were given partial occupancy. Details are given in Table 2. Why is there so much more disorder found in the triclinic structure; is it the increased resolution makes such details more discernible? Perhaps the main reason is that the mass of the iodide ions restricts thermal vibrations of the surrounding side chains to small displacements. Certainly most of the disordered residues, most often arginine, of the triclinic are found associated with iodide ions in the monoclinic structure. The Luzzati plot gave 0.07 as the estimated error. The space group, the unit-cell parameters, and the data-collection statistics are given in Table 1, and the course of the refinement is followed in Table 2. The course of the atoms of the main chain are identical to that given by Jensen (Ramanadham *et al.*, 1990) in spite of the fact that the crystals, the data collection, the resolution, the starting model, and the method of refinement were all quite different. The only differences seen are in the more ephemeral regions such as the side chains of Asn104, Arg125, and Arg128.

Six nitrate ions were identified (Table 6), but the only empty region large enough to contain the bromophenol blue did not display any features that would allow the identification of that molecule. We had previously used bromophenol blue and other dyes as aids in the crystallization of proteins (Steinrauf *et al.*, 1991), and often these molecules were found poorly or not at all in the resulting structure. The crystal structure of tetragonal lysozyme chloride with bromophenol blue as a guest has been reported by Madhusudan & Vijayan (1992). For that structure the guest had been located by difference Fourier maps.

3. Results and discussion

It was expected that the locations of the anions would be found in the structure of monoclinic lysozyme iodide. This was based on the observation that there was significant anomalous dispersion in the diffraction data. This could come only from atoms in fixed positions with significant contributions of anomalous scattering, and would not be observed if the iodine atoms were highly disordered. It would be reasonable to expect that the

Table 6. *φ and ψ angles for the residues of the variable regions*

Variable regions: 68–73 and 99–104 for triclinic lysozyme nitrate (this work), molecules A and B of monoclinic lysozyme nitrate (Rao & Sundaralingam, 1996), molecules A and B of monoclinic lysozyme chloride (Harata, 1994), and molecules A and B of monoclinic lysozyme iodide (this work).

				Angle φ/ψ			
Residue	Triclinic	Nitrate A	Nitrate B	Chloride A	Chloride B	Iodide A	Iodide B
Arg68	−125/−3	−131/14	−139/19	−133/10	−116/26	−128/15	−96/−8
Thr69	−102/128	−106/116	−99/129	−94/126	−142/151	−123/117	−163/67
Pro70	−64/143	−59/132	−56/133	−47/−52	−17/−130	−59/143	−65/144
Gly71	74/23	60/40	64/25	−127/58	−108/168	62/−22	82/14
Ser72	−63/130	−84/144	−60/106	−80/148	−72/170	−53/149	−139/149
Arg73	−104/−30	−106/177	−78/169	−117/162	−106/175	−120/−22	28/−122
Val99	−74/−32	−69/−22	−70/−17	−69/−24	−69/24	−68/−20	−72/−11
Ser100	−68/−25	−94/6	−88/0	−82/−9	−84/9	−99/15	−97/5
Asp101	−90/40	−86/2	−89/9	−116/−59	−90/−3	−96/16	−86/−1
Gly102	75/19	117/−6	114/−7	−100/110	128/−24	−96/7	114/−8
Asn103	−100/154	−117/14	−122/9	17/57	110/8	−130/15	−116/8
Gly104	−95/−148	52/−139	49/−2	59/−142	57/−141	55/−148	46/−140

iodide ion sites would be close to the positively charged residues, arginine, lysine and histidine. This has usually been found to be true, but as may be seen in Table 4, the iodide ions are in contact with other kinds of residues, also, and both hydrogen bonding and hydrophobic contacts are important.

Two unexpected features have been found. Some of the iodide sites are different around the two independent lysozyme molecules, and the 65–75 loop of the two molecules is very different. These two observations are of course related, but which is the cause and which the effect? Since the differences in the 65–75 loop are more pronounced in the monoclinic iodide than in the monoclinic chloride/propanol as reported by Harata (1994), perhaps we can assume that the size and shape of the anions are responsible for producing changes in the flexible regions of lysozyme. Changes in the 65–75 loop have been found from other studies as well. The X-ray crystal structure by Kundrot & Richards (1987) of tetragonal lysozyme chloride at hydrostatic pressure of 1000 atmospheres ($\sim 10^5$ Pa) showed changes in that loop. Studies by Ford *et al.* (1974), Perkins *et al.* (1978) and Kelly *et al.* (1979) on the binding of substrates to tetragonal lysozyme chloride showed changes in both the 70–74 region and the 101–103 region. Figs. 3(*a*) and 3(*b*) show the 65–75 loop for lysozyme A and B as viewed from the same direction. For example, it is easy to see the five-member ring of Pro70 in Fig. 3(*a*), but this has moved to edge on in Fig. 3(*b*).

The individual temperature factors were considerably lower for the triclinic lysozyme nitrate than for the iodide. The average B values for the peptide/side-chain atoms were 13.1/17.3 Å^2 for the triclinic, 23.7/27.6 Å^2 for the iodide at pH 8, and 23.1/26.7 Å^2 for the iodide at pH 4. As would be expected the electron density is much better resolved for triclinic lysozyme. Fig. 4(*a*) shows the electron density of the first three residues of the N-terminal end for monoclinic lysozyme A; Fig. 4(*b*) shows the same residues for triclinic lysozyme.

The pattern of thermal motion along the peptide chain was very similar for all three crystals, low along the helices, and higher in the loops. Several residues having a branched side chain were found to have one atom of the branch with a much higher temperature factor. This atom was always pointing away into the solvent, and the other atom was hydrogen bonded to the chain nearby. A typical example would be an Asp or Asn residue with the OD1 hydrogen bonded to peptide N atom two residues farther on, and the OD2 or ON2 moving freely. This was found for Asp18 hydrogen bonding to N of Leu25, Asn38 hydrogen bonding to N of Trp41, Asp48 hydrogen bonding to N of Ser50, and Asn65 hydrogen bonding to N of Gly67. These were found in the triclinic and in both molecules of the iodide at pH 8 and pH 4. In a similar manner the side chain of Arg5 makes hydrogen bonds with the O atoms of both Arg125 and Trp123. The other side of the side chain makes no hydrogen bonds in the triclinic structure, but in the iodide structures it makes both a salt link and a hydrogen bond with Asp101 of the neighboring molecule and is prevented from moving. This link is one of the key stabilizations of the monoclinic form.

3.1. *Electrostatic interactions*

There are several important interactions between acidic and basic side groups. These are not all the same in the different molecules. Glu7 makes an intramolecular salt bridge to Lys1 in both molecules A and B, but makes contact with a neighbor in molecule T. Asp18 makes a hydrogen bond with the N atom of Leu24 and is about 5 Å from Lys13 in all molecules. Glu35 makes a hydrogen bond to the N atom of Ala110 in all molecules. Asp48 makes a salt link to Arg61 in all molecules with the details being slightly different. Asp66 is more variable, but makes bonds with Arg68 in all molecules, and has a hydrogen bond to the OH of Tyr53 in molecules A and T. When the 65–75 loop swings, the Arg68 maintains

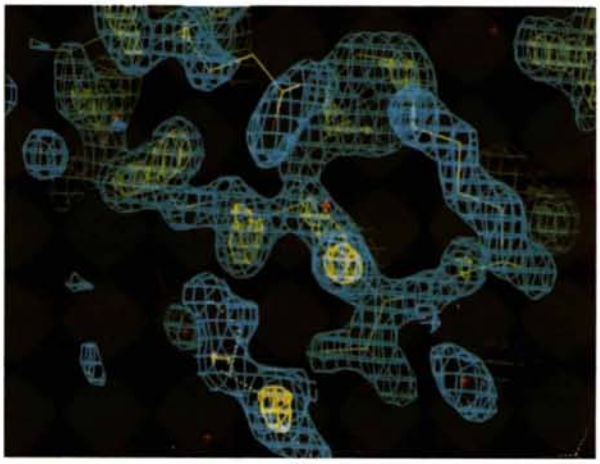
contact with Asp66, but loses contact with Tyr53. Asp87 makes a hydrogen bond to OG1 of Thr98 in all forms, and also makes a hydrogen bond with the N of Arg101 in molecule B. Asp101 is quite different, making a hydrogen bond with Glu7 in molecule T, but in molecules A and B it makes strong double hydrogen bonds with the Arg125 of a neighbor. For molecules A and B, Asp119 makes a good salt bond to Arg73 of a neighbor, but has an intramolecular salt link to Arg125 in molecule T.

The iodide and nitrate sites are usually sandwiched between two lysozyme molecules, but for some sites it is clear that the anion is bound to one lysozyme more than the other. Thus, lysozyme can be considered to contain half sites for anion binding. Such half sites usually contain one or two positively charged side chains. However, there are a few sites that have no nearby positive charges.

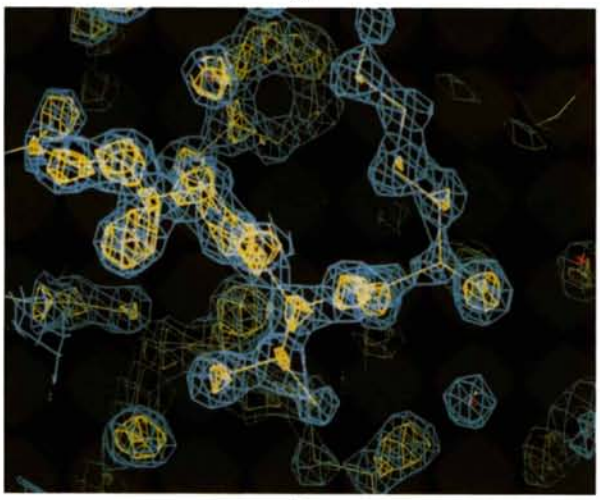
It was observed very early from precession photographs that the monoclinic lysozyme iodide was less B centered than was the nitrate. This lack of perfect centering of the electron density is shown in Figs. 2(*a*) and 2(*b*), in which the neighbors (dotted lines) are considerable farther away from lysozyme B than from lysozyme A. The figures show a common site for the iodide ion, and show that lysozyme A has an additional iodide site. Comparison of the iodide sites show the following similarities: iodides 2 = 13, 3 = 7, 4 = 9, 6 = 10 and 15 = 16, leaving seven iodides not paired. Of course even the pairing by binding to the same residues on the other lysozyme molecule does not provide crystallographic equivalence. The pair may have different occupancy or different shape, running from football to sphere to doorknob. In Figs. 2(*a*) and 2(*b*) and the following figures, the residues of lysozyme A are numbered from 1 to 129 and the residues of lysozyme B run from 201 to 329.

The triclinic structure yielded six definite sites of nitrate ions. It might be expected that some of the half sites from the iodide could be utilized for nitrate. Indeed, nitrate 1 nestles above the ring of Pro79 (the beginning of a γ-helix) just as do iodides 15 and 16, respectively. There are some similarities in the placement of nitrate 4 with iodides 11 and 14, and of nitrate 5 with iodide 13.

Lysozyme can be considered to be a huge cation with many potential half sites on its surface. In order for



(*a*)



(*b*)

Fig. 4. The N-terminal part of lysozyme, right to left, Arg, Val, Phe. (*a*) Monoclinic lysozyme iodide at 1.6 Å resolution. (*ab*) Triclinic lysozyme nitrate at 1.1 Å resolution.

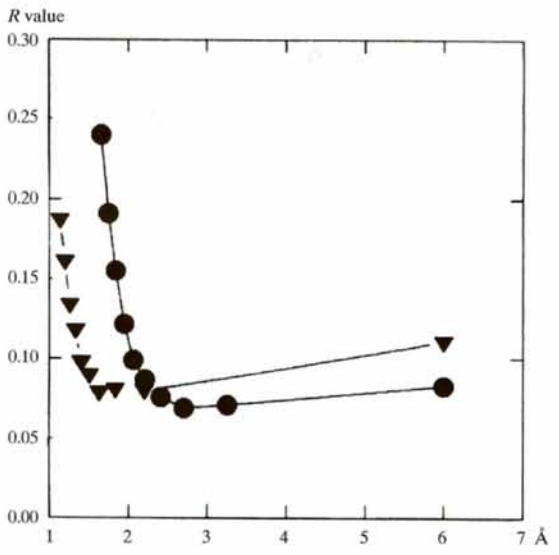


Fig. 5. The radial distribution of R value for triclinic lysozyme nitrate, 1.1 Å resolution, $R = 0.0988$ (triangles), and monoclinic lysozyme iodide at 1.6 Å resolution, $R = 0.0798$ (squares). If iodide atoms where not in the monoclinic structure, the R values in the 2.0–7.0 range would be higher (probably around 0.12) and the R values in the 1.6–1.8 range would be lower than found here.

crystallization to be possible the required number of anions must find appropriate half sites to make a crystal lattice. These paired half sites with the contained anion, and the salt bridges mentioned previously, are the principal contacts between neighbors. It is not surprising to find that some of the positively charged residues can take part in more than one half site. It is thus easy to see how many different crystal forms may be obtained using different anions. In fact an even simpler system is well known. The natural dodecapeptide valinomycin binds a single monovalent cation and has two sites for an accompanying anion. Several different crystal forms of valinomycin have been discovered (Hamilton *et al.*, 1981; Steinrauf *et al.*, 1982) depending on the combinations of anion and cation. Such a concept in the crystal may extend to solution. This leads to the idea that an anion in solution may move across the surface of lysozyme, being shuttled from one half site to another.

3.2. *Conformation*

The structure of lysozyme is dominated by the six segments of helix, 4–15, 24–37, 79–84 (γ), 88–99, 108–115 and 119–125 (γ). Connecting these are β-strands 1–3, 38–40, 42–46, 50–54 and 57–60; turns at 41, 47–49, 55–56, 61–62 and 85–87; and rigid loops or strands 16–23, 63–66, 74–78, 116–118 and 126–127. Variations in conformation are seen at 47–49, and especially in the long flexible loops 67–73 and 100–107, and at 128–129. Disulfide bonds connect 6–127, 30–115, 64–80, and 76–94 provide rigidity to those regions. The 67–73 loop is seen in the ribbon drawings of lysozyme A and B superimposed (Figs. 1*a* and 1*b*). Differences for the anion sites of lysozyme A and B, however, are not restricted to this loop region.

Some of the more important differences for the φ/ψ angles are given in Table 5 for lysozyme A, B, and T. It can be seen that there is a twist between residues 68 and 69, and at 73 for lysozyme B. Comparison with a similar table given by Harata (1994) suggests that monoclinic lysozyme chloride/propanol is an intermediate and that the iodide is the extreme case. It might be noted in Table 5 that lysozyme A is identical to lysozyme T in this region, and the structure by Harata is also identical except for Gly71, although it has been pointed out by Jensen (Ramandham *et al.*, 1990) that details such as the orientation of a glycine are often not apparent until the rest of a structure has been well refined. The same region in lysozyme B is more variable. The deviations are much smaller in monoclinic lysozyme nitrate. The deviations start with Thr69 and end with Gly71 for lysozyme chloride/propanol, while for lysozyme iodide the changes start at Arg68 and extend to Arg73 with Pro70 and Gly71 moving as a unit. The application of molecular modeling to the 65–75 loop may well be a fruitful subject for future research.

It is to be expected that the lysozyme molecules of the triclinic and monoclinic crystal forms would have different contacts with neighbors. One very significant difference is found in the 99–104 loop. This region is stabilized in the monoclinic structure by hydrogen bonds to the neighbors and a salt bridge between Asp101 and Arg5 of the neighbor. The electron density is quite good in the monoclinic structure and the residues are well located. In the triclinic structure the residue Asp101 has a good hydrogen-bonding distance with Glu7 of the neighbor (x, y, $z+1$). Since this must mean that one of these residues has lost the charge, it is not surprising that the triclinic crystal form cannot be made above pH 4.7. Which of the two residues is in the carboxylic acid form cannot be determined absolutely on the basis of C—O bond distances, but tends to favor Glu7, which would be expected on the basis of the slightly higher pK for glutamate. The other intramolecular hydrogen bonds are absent in the lysozyme T and the electron density of the 102–104 range is weaker than in lysozyme A or B. The φ/ψ angles in Table 5 show a rotation at 100–101 and at 103–104. The 101–104 region has been shown to be sensitive to chemical attack in solution (Yanada, 1990). The reagent tetranitromethane will attack Gly104 with oxidative bond cleavage. The last two residues of the monoclinic structure are of stronger electron density than those of lysozyme T. This is at least partly because the salt bridge between the carboxy terminus and NZ of Lys13 is not complete in lysozyme T.

3.3. *Anions*

16 iodide ion positions were located in the first difference Fourier map and one more thereafter. Each of the two lysozyme molecules contains two glutamyl, seven aspartyl, six lysyl, and 11 arginyl residues which would give a net positive charge of 8 if the single histidyl residue is not counted. Thus, the correct number of major iodide sites are found, but the occupancy of these sites need not be full. Indeed the electron density at some of these sites is significantly weaker than for others, as may be appreciated from the list of site occupancies and temperature factors in Table 3. It is well known that the absorption of X-rays by heavy elements such as iodine will cause the observed electron density of atoms to be more uncertain. That, and the strong anisotropic motion of some of the sites, makes a more precise accounting of the iodide ions not worth pursuing. The 17 iodide sites are quite prominent, some showing a nearly spherical shape, others a cigar shape, and others a disk shape.

As would be expected, the iodide sites are often located near arginyl or lysyl side chains. However, the iodide sites also show a distinct predilection for hydrophobic areas of the protein. Hydrogen bonding from the protein to the iodide is also common, with the side chain

of asparagine and the peptide N atom being the most common. In the crystal structures of arginine dihydriodide and tyrosine hydriodide (Seely, 1965) hydrogen-bond distances of 3.63–3.81 Å were found from the guanidium N atoms, and 3.46–3.58 Å from the charged amino groups. The bond from the phenol O atom to the iodide ion was shorter, 3.39 Å.

Table 4 gives the contacts from each of the 17 iodide sites to the nearest side chains. Two iodide sites are between molecules with positive charges on one side only. Six iodide sites are between molecules, finding a small pocket in one of the molecules, and having positive charges from both molecules. Each lysozyme molecule has three self-contained anion sites. These are the same for both independent lysozyme molecules and are also found in triclinic lysozyme nitrate. Two iodide sites are contained in a pocket within one lysozyme molecule but containing no positive charges. The positive charges are donated by the neighboring molecule. These two sites are the same on both lysozyme molecules. In summary each lysozyme molecule had four sites for iodide ions for a total of eight sites. The other nine sites are between molecules and are different on the two independent lysozyme molecules. The equivalent sites on the two lysozyme molecules are given in Table 4 and are numbered 1–17. Of these iodide 2 = 13, iodide 3 = 7: iodide 4 = 9, iodide 15 = 16, and perhaps iodide 6 = 10. This leaves iodides 1, 5, 8, 11, 12, 14 and 17 without any equivalent on the other molecule. Instability of the 65–75 loop has been detected in previous structure determinations. In all such determinations that loop has had high temperature factors, indicating a loose structure. The binding of substrates to crystals of lysozyme produced shifts in the peptide atoms of residues 70–75 (among others).

Two sites, the histidine side chain and the N-terminal amino group, might be expected to change in charge during the change from pH 8 to pH 4. The occupancies of sites 4 and 9 have increased at pH 4, and the temperature factors have decreased. However, most of the sites show increased occupancies and decreased temperature factors, indicating that the influence is over the entire molecule and not localized.

3.4. *Water molecules*

The number of sites for water molecules found in the structures per lysozyme molecule is 148 (296 for the two molecules) for the monoclinic structure and 189 for the triclinic, which is a moderate favoring for the superior resolution of the triclinic. However, the number of water sites of full occupancy is only 30 per lysozyme in the monoclinic *versus* 81 in the triclinic. This is at least in part due to the overall lower temperature factor for the triclinic structure. These results are in agreement with the observations on lysozyme at low temperature, where refinement of data taken at 178 K produced nearly twice as many water sites as data from room temperature. This suggests that the use of low-temperature data is more important than high resolution for the purpose of locating water molecules.

3.5. *Future research*

Although the crystallography of lysozyme is 50 years old, there are still some interesting questions to be asked. The effects of changing the anion in the monoclinic form has now been surveyed for the nitrate at low pH, the iodide at both high and low pH, and the chloride/propanol at high pH which seems to be similar to the nitrate at high pH. For the chloride/propanol structure the anion positions were not reported. The solubility of lysozyme in various salts seems to follow the sequence, sulfate > chloride > nitrate > bromide > iodide > thiocyanate, which suggests that some anions fit the binding sites on lysozyme better than others.

What would be the consequences of changing the charge of the anions? In fact this has partly been performed. Steinrauf (1959) reported that lysozyme could be crystallized from sodium sulfate, and he found the crystals from sulfate to be similar in appearance and with cell dimensions close to that of the nitrate and iodide. In retrospect this is surprising. How can the divalent sulfate be accommodated in the sites for the monovalent nitrate and iodide? Of course the extent of isomorphism of the sulfate exists only in the memory of the author. Another interesting question is whether the monoclinic lysozyme iodide will undergo the transition at lower humidity that is observed with monoclinic lysozyme nitrate. Precession films of the iodide show much less pseudo centering than do those of the nitrate. A picture of the iodide has been used by Moore in his *Physical Chemistry* textbook (1972).

It has also been observed by the author that crystals of triclinic lysozyme iodide can be grown with careful seeding. Precession films of the triclinic showed much stronger anomalous differences than did those of monoclinic lysozyme iodide. The reason for this is not immediately apparent.

Crystals of lysozyme grown from sodium thiocyanate solutions are isomorphous to the monoclinic lysozyme nitrate and iodide. Most of the iodide sites as described in Table 4 are highly asymmetric. The thiocyanate ion has a dipole moment, and therefore would the anions have a preferred directional orientation in these sites? At the resolution of the present data, the S atom in the thiocyanate ion should be easily identified thereby giving the orientation. This study is now under way.†

† Atomic coordinates and structure factors have been deposited with the Protein Data Bank at Brookhaven National Laboratory (References: 1LKR, 1KLS).

I wish to thank Dr Tom Hurley for help with various programs, and I received much needed help from George Sheldrick with the use of *SHELX*93. Thanks are due to Drs Michael Chiang and David Hsiao for their help in preparing the manuscript.

References

Blake, C. C. F., Fenn, R. H., North, A. C. T., Phillips, D. C. & Polijak, R. J. (1962). *Nature (London)*, **196**, 1173–1176.

Blake, C. C. F., Johnson, L. N., Muir, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. R. Soc. Ser. B*, **167**, 378–388.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London)*, **206**, 757–761.

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

Brünger, A. T. (1992*b*). *Nature (London)*, **355**, 472–475.

Crick, F. H. C. (1953). *Acta Cryst.* **6**, 221–222.

Dickerson, R. E., Reddy, J. M., Pinkerton, M. & Steinrauf, L. K. (1962). *Nature (London)*, **196**.

Ford, L. O., Johnson, Machin, P. A., Phillips, D. C. & Tjian, R. (1974). *J. Mol. Biol.* **88**, 349–371.

Hamilton, J. A., Steinrauf, L. K., & Sabesan, M. N. (1981). *J. Am. Chem. Soc.* **103**, 5880–5885.

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

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

Hodsdon, J. M., Brown, G. M., Sieker, L. C. & Jensen, L. H. (1990). *Acta Cryst.* B**46**, 54–62.

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

Johnson, L. N. & Phillips, D. C. (1965). *Nature (London)*, **206**, 761–763.

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

Joynson, M. A., North, A. C. T., Sarma, V. R., Dickerson, R. E. & Steinrauf, L. K. (1970). *J. Mol. Biol.* **50**, 137–142.

Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N. G. & Phillips, D. C. (1979). *Nature (London)*. **282**, 875–878.

Kundrot, C. E. & Richards, F. M. (1987). *J. Mol. Biol.* **93**, 157–170.

Kurachi, K., Sieker, L. C. & Jensen, L. H. (1976). *J. Mol. Biol.* **101**, 11,24.

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

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

Madhusudan & Vijayan, M. (1992). *Protein Eng.* **5**, 399–404.

Molecular Structure Corporation (1993). *MSC*, Molecular Structure Corporation, The Woodlands, TX, USA.

Moore, W. J. (1972). *Physical Chemistry*, 4th ed., p. 853. Englewood Cliffs, New Jersey: Prentice-Hall.

Moult, J., Yonath, A., Traub, W., Smilansky, A., Podjarny, A., Rabinovich, D. & Saya, A. (1976). *J. Mol. Biol.* **100**, 179–195.

Perkins, S. J., Johnson, L. N., Machin, P. A. & Phillips, D. C. (1978). *Biochem. J.* **173**, 607–616.

Pike, A. C. W. & Acharya, R. A. (1994). *Protein Sci.* **3**, 706–710.

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

Ramanadham, M., Sieker, L. C., Jensen, L. H. & Birkens, B. J. (1981). *Acta Cryst.* A**37**, C-33.

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

Rao, S. T. & Sundaralingam, M. (1996). *Acta Cryst.* D**52**, 170–175.

Seely, O. J. (1965). PhD thesis, University of Illinois, Urbana, USA.

Sheldrick, G. (1997). *Methods Enzymol.* **277**, 319–343.

Stanford, R. H., Marsh, R. E. & Corey, R. B. (1962). *Nature (London)*, **196**, 8–13.

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

Steinrauf, L. K., Cao. Y. J., Hamilton, J., Murrell, J., Liepnieks, J. J. & Benson, M. D. (1991). *Biochem. Biophys. Res. Commun.* **179**, 804–809.

Steinrauf, L. K., Hamilton, J. A. & Sabesan, M. N. (1982). *J. Am. Chem. Soc.* **104**, 4085–4091.

Yanada, H, Yamashita, T., Domoto, H. & Imoto, T. (1990). *J. Biochem.* **108**, 432–440.

Young, A. C. M. & Dewan, J. C. (1993). *J. Appl. Cryst.* **26**, 309–319.

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. 25–38. New York: Adenine Press.

Acta Cryst. (1998). D**54**, 780–798

Structure of a Non-psychrophilic Trypsin from a Cold-Adapted Fish Species

HANNA-KIRSTI SCHRØDER,[a] NILS P. WILLASSEN[b] AND ARNE O. SMALÅS[a]*

[a]*Protein Crystallography Group, Institute of Chemistry, Faculty of Science, University of Tromsø, N-9037, Tromsø, Norway, and* [b]*Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037, Tromsø, Norway. E-mail: arne.smalas@chem.uit.no*

(*Received 4 October 1997; accepted 2 December 1997*)

Abstract

The crystal structure of cationic trypsin (CST) from the Atlantic salmon (*Salmo salar*) has been refined at 1.70 Å resolution. The crystals are orthorhombic, belong to space group $P2_12_12_1$, with lattice parameters a = 65.91, b = 83.11 and c = 154.79 Å, and comprise four molecules per asymmetric unit. The structure was solved by molecular replacement with *AMoRe* and refined with *X-PLOR* to an R value of 17.4% and R_{free} of 21.5% for reflections $|F| > 3\sigma_F$ between 8.0 and 1.7 Å resolution. The four non-crystallographic symmetry (NCS) related molecules in the asymmetric unit display r.m.s. deviations in the range 0.31–0.74 Å for main-chain atoms, with the largest differences confined to two loops. One of these is the calcium-binding loop where the electron-density indicates a calcium ion for only one of the four molecules. In order to find structural rationalizations for the observed difference in thermostability and catalytic efficiency of CST, anionic salmon trypsin (AST) and bovine trypsin (BT), the three structures have been extensively compared. The largest deviations for the superimposed structures occur in the surface loops and particularly in the so-called 'autolysis loop'. Both the salmon enzymes possess a high methionine content, lower overall hydrophobicity and enhanced surface hydrophilicity, compared with BT. These properties have so far been correlated to cold-adaptation features, while in this work it is shown that the non-psychrophilic cationic salmon trypsin shares these features with the psychrophilic anionic salmon trypsin.

1. Abbreviations

CST, cationic salmon trypsin; AST, anionic salmon trypsin crystal form 1 (PDB entry 2TBS); AST-II, anionic salmon trypsin crystal form 2 (PDB entry 1BIT); BT, bovine trypsin (PDB entry 3PTB); NCS, non-crystallographic symmetry; Mol *A*–*D*, molecules *A*–*D* of cationic salmon trypsin; r.m.s., root-mean-squares; res., residue(s); $R_{merge} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |\langle I_h \rangle| \times 100\%$.

2. Introduction

The digestive proteinase trypsin is found in multiple forms in many species, both mammalian *e.g.* cattle (le Huerou *et al.*, 1990; Titani *et al.*, 1975), rat (Craik *et al.*, 1984; Fletcher *et al.*, 1987) and other vertebrates *e.g.* chicken (Wang *et al.*, 1995). In particular many marine species are known to have a rather complex isoenzyme pattern. Capelin (Hjelmeland & Raa, 1982), Antarctic krill (Osnes & Mohr, 1985), cod (Gudmundsdóttir *et al.*, 1993), anchovy (Martínes *et al.*, 1988) and crayfish (Kim *et al.*, 1994) have all been identified with multiple forms of trypsin. While the predominant isoforms in mammals are cationic, only anionic variants have been reported for the above-mentioned marine species. To our knowledge, Atlantic salmon and Chum salmon are the only fish species for which a cationic isoform has been isolated and characterized (Male *et al.*, 1995; Outzen *et al.*, 1996; Uchida *et al.*, 1986), although anionic isoforms seem to dominate in both species. Little is known about the possible functions of the multiple isoenzymes in an organism, but studies by Torrissen (1987) showed that there were variations in trypsin isoenzyme patterns between families of Atlantic salmon and one particular genotype could, for example, be associated with fish size (Torrissen, 1987, 1991). In Atlantic salmon five different clones of trypsin have been identified and sequenced (Male *et al.*, 1995), and judging from the amino-acid sequences one of these is cationic and four are anionic isoforms. One cationic and one main anionic fraction of salmon trypsin have been isolated and characterized (Outzen *et al.*, 1996), and the results from this study show that the two trypsin fractions possess notable differences in both activity and stability. The cationic form (CST) resembles the cationic bovine trypsin (BT), while the anionic form (AST) is considerably less stable at high temperature and low pH, and about 20–30 times more efficient (in terms of k_{cat}/K_m) in hydrolysing an amide substrate, compared to the cationic salmon and bovine enzymes. The GdnHCl-induced unfolding experiments showed that the free energy of unfolding of CST, AST and BT were 10.5, 4.7 and 14.8 kcal mol^{-1} (1 kcal = 4.184 kJ), respectively. Unfolding by GdnHCl is believed to monitor the hydrophobic stability of a