Acta Cryst. (1996). D**52**, 631–640

Structure of Diferric Duck Ovotransferrin at 2.35 Å Resolution

AHMAD RAWAS, HILARY MUIRHEAD* AND JOHN WILLIAMS

Department of Biochemistry and Molecular Recognition Centre, University of Bristol, Bristol BS8 1TD, England. E-mail: muirhead@bsa.bristol.ac.uk

(Received 21 July 1995; accepted 3 January 1996)

Abstract

The structure of diferric duck ovotransferrin (DOT) has been determined and refined at a resolution of 2.35 Å. The DOT structure, which contains two iron binding sites, is similar to the known transferrin and lactoferrin structures. The two iron-binding sites, one in the N-terminal lobe and one in the C-terminal lobe of the molecule, are similar but not identical. The main differences between the three known structures lie in the relative orientations of the N- and C-lobes with respect to each other. In the DOT structure the large aromatic side chain of Phe322 in the N-lobe packs against the conserved residue Gly387 in the C-lobe. This interaction is at the centre of the interface between the two lobes and could play a crucial role in determining their relative orientation. Other differences between the structures occur in the surface loops and in the peptide connecting the two lobes. The final crystallographic model consists of 5309 protein atoms (686 residues), two Fe^{3+} ions, two (bi)carbonate ions and three carbohydrate moities. 318 water molecules have been added to the model. The final *R* factor is 0.22 for 25 400 observed reflections between 10 and 2.35 Å resolution.

1. Abbreviations

DOT, duck ovotransferrin; HOT, hen ovotransferrin; HST, human serum transferrin; HMT, human melanotransferrin; RST, rabbit serum transferrin; HLT, human lactoferrin; MLT, mouse lactoferrin; MST, *Manduca sexta* transferrin; r.m.s., root-mean-square; NAG, *N*-acetylglucosamine; FUC, fucose.

2. Introduction

Transferrins are metal-binding glycoproteins of molecular weight 80 kDa. The single polypeptide chain contains nearly 700 amino-acid residues (for reviews see Brock, 1985, Crichton, 1991). The N-terminal and C-terminal halves of the chain fold up to form two homologous compact lobes (Anderson *et al.*, 1987; Bailey *et al.*, 1988, Fig. 1, Table 1). Most transferrins are diferric and each lobe carries one high-affinity iron-binding site. These two sites are similar but not identical. Diferric transferrins include serum transferrin (the iron-transport protein found in serum), ovotransferrin (from egg white) and lactoferrin (found in milk and other secretions, as well as leucocytes). The membrane protein human melanotransferrin (HMT) and *Manduca sexta* transferrin from tobacco hornworm (MST) have only one iron-binding site. Sequence homologies indicate that in each case the iron-binding site is located in the N-lobe (Baker *et al.*, 1992, Bartfeld & Law, 1990). Diferric transferrin molecules bind reversibly two Fe^{3+} ions per molecule, concomitantly with two CO_3^{2-} anions and have many biological functions. The primary function of serum transferrin is the transport of iron in the plasma of vertebrates while ovotransferrin delivers iron to chick embryo red cells *in vitro*. Thus, serum transferrin and ovotransferrin have similar biological functions. Lactoferrin protects cells from damage by free radicals by binding potentially catalytic free iron. In addition the strong affinity for iron leads to the removal of iron from the extra-cellular environment hence depriving bacteria of iron which is essential for their growth.

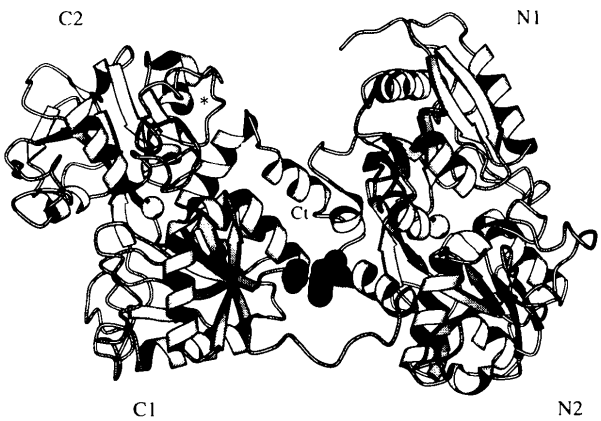

Fig. 1. A ribbon representation of duck ovotransferrin drawn with the program *MOLSCRIPT* (Kraulis, 1991). The positions of the two Fe^{3+} ions are indicated by surface filled circles in the clefts between the two domains N1 and N2 in the N-lobe and C1 and C2 in the C-lobe. The * labels the position at which the carbohydrate chain is attached (Asn473 on helix H5_C). Residues Phe322 (helix H11_N) and Gly387 (helix H2_C) in the inter-lobe contact are shown in c.p.k. representation. Ct indicates the carboxyl terminus of the DOT molecule.

© 1996 International Union of Crystallography
Printed in Great Britain – all rights reserved

ISSN 0907-4449 © 1996

Table 1. *Secondary structure in duck ovotransferrin*

The nomenclature used in defining the structure is that described by Anderson *et al.* (1987, 1989) and Bailey *et al.* (1988). The β-strands are labelled from *a* to *k*, helices are labelled from 1 to 12. β-strands *e*1 and *j*1 form a two-stranded antiparallel sheet which forms an interdomain link. Two extra strands (*x*) have been located in domains N2 and C2 (residues 232–234 and 574–576). Lobe N, domain N1; residues 1–91, 248–332; domain N2; residues 94–244. Lobe C, domain C1; residues 345–431, 589–672; domain C2; residues 433–586. Interdomain links; 91–94, 244–248; 431–433, 586–589. Interlobe links; 333–344, 672–686.

N-lobe		C-lobe
6–11	*a*	345–351
14–26	1	352–365
33–39	*b*	368–374
42–50	2	376–386
55–59	*c*	390–395
61–68	3	396–404
75–81	*d*	408–412
91–94	*e*1	431–433
95–98	*e*2	433–437
106–109	4	446–447
112–116	*f*	451–455
122–135	5	462–475
149–155	6	481–483
158–161	*g*	487–488
190–200	7	523–534
205–208	*h*	537–540
211–216	8	543–549
224–228	*i*	567–570
232–234	*x*	574–576
245–248	*j*1	586–589
251–255	*j*2	592–596
262–274	9	600–614
305–308	*k*	643–647
315–320	10	653–657
322–332	11	659–671
	12	675–683

Complete amino-acid sequences have been determined for a number of transferrins; hen ovotransferrin (HOT, Williams, Elleman, Kingston, Wilkins & Kuhn, 1982), duck ovotransferrin (DOT, R. W. Evans, personal communication), human serumtransferrin (HST, Yang *et al.*, 1984), rabbit serum transferrin (RST, Banfield *et al.*, 1991), HMT (Rose *et al.*, 1986), human lactoferrin (HLT, Metz-Boutigue *et al.*, 1984), mouse lactoferrin (MLT, Pentecost & Teng, 1987) and MST from tobacco hornworm (Bartfeld & Law, 1990). Sequence comparisons show that the N-terminal lobe of one transferrin is more similar to the N-terminal lobe of another transferrin than to its own or other C-terminal lobes suggesting that all modern transferrins have resulted from the same early gene duplication in a primitive ancestor (Williams, 1982). The sequences of the N-lobes (iron-binding lobes) of HMT and MST are more homologous to other transferrins than are their non-iron-binding C-lobes. Sequence homologies (identical residues in corresponding positions) are between 49 and 59% for a number of different transferrins (HST, HOT and HLT) while internal homologies between the N-lobe and the C-lobe lie between 37 and 46% (Metz-Boutigue *et al.*, 1984).

The diferric transferrins contain 30 half-cystine residues in homologous positions, giving 15 disulfide bridges in common (Metz-Boutigue *et al.*, 1984). Each lobe contains six disulfide bridges in homologous positions and these are numbered N1–6 and C1–6. There are a further three disulfide bridges in the C-lobe numbered 7–9. Serum transferrin contains an additional four bridges, two in each lobe (N10–11, C12–13) while lactoferrin contains one extra bridge in the C-lobe (C13). The primary structure, content and position of the carbohydrate moieties are specific for different transferrin molecules. In HOT one glycan chain is present located in the C-lobe linked to Asn473 (Dorland *et al.*, 1979, Williams *et al.*, 1982). HST and human lactoferrin possess two glycan chains per molecule located at Asn437 and Asn611 for HST (MacGillivray *et al.*, 1983) and at Asn137 and Asn490 for HLT (Metz-Boutigue *et al.*, 1984) in the N and C-lobes respectively. The carbohydrate content of RST is similar to that of the isolated C-lobe of ovotransferrin and consists of a single glycan chain (Heaphy & Williams, 1982; Evans, Aitken & Patel, 1988), and located at Asn491 (Bailey *et al.*, 1988).

This paper describes the determination of the crystal structure of diferric duck ovotransferrin at a resolution of 2.35 Å.

3. Methods and results

3.1. *Protein crystallization*

The procedure used in protein crystallization has been described (Rawas, Moreton, Muirhead & Williams, 1989). Crystallization was performed by a modification of the traditional batch method using quartz capillary tubes of diameter 1.5 mm. Reddish-brown crystals of duck diferric ovotransferrin were obtained, within a week, from 12–18% of polyethylene glycol 6000 in 0.01 *M* sodium acetate buffer (pH 5.8) at 291 K. The space group is $P2_12_12_1$ and the unit-cell dimensions are $a = 49.6$, $b = 85.6$, $c = 178.7$ Å giving one molecule in the crystallographic asymmetric unit. The volume per unit mass V_m is 2.4 Å^3 Da^{-1} which falls within the range (1.60 to 3.53 Å^3 Da^{-1}) observed for protein crystals (Matthews, 1968). This corresponds to a solvent content of 48%.

3.2. *Data collection*

Three-dimensional X-ray data were collected on the SERC X-ray synchrotron source at Daresbury using an Arndt–Wonacott rotation camera. Data were collected at 2.3 Å resolution for the native protein and four heavy-atom derivatives – sodium aurichloride, platinum potassium chloride, mercuric acetate and iridium chloride. The native data set, covering an oscillation

range of 90° was collected from two crystals using a wavelength of 1.488 Å. In order to measure the anomalous scattering of the Fe atoms a data set was collected near the iron absorption edge (1.725 Å) over an oscillation range of 60°. The films were scanned with an Optronics P1000 microdensitometer and the data processed using the *CCP4* suite of programs (Collaborative Computational Project, Number 4, 1994). The R_{sym}'s on intensities for individual packs were in the range 5.5–7.0% and 5.2–12.0% for the anomalous and native data, respectively. The length of the *c* axis means that, since the crystals were mounted about the *b* axis, the *R* factors for the oscillation range from 60 to 90° are worse than those from 0 to 60°. The overall R_{sym}'s were 6.4% for a total of 24 028 independent reflections (having 15 831 anomalous measurements from the same batch) for the anomalous data, and 5.9% for a total of 29 846 unique reflections of 90% completeness to 2.35 Å resolution for the native data, 27 564 reflections with 86% completeness had $I > \sigma$.

3.3. *Multiple isomorphous replacement, MIR*

Difference Patterson and difference Fourier syntheses were used to find the heavy-atom sites. The platinum derivative had one major and three minor sites. The major site had an occupancy about four times that of the minor sites. The gold and iridium derivatives had eight and ten sites, respectively. The phases were calculated for 15 268 reflections to a resolution of 3.0 Å using multiple isomorphous replacement with anomalous scattering. The average figure of merit was 0.5. An electron-density map was calculated at a resolution of 3.0 Å. The boundary between the molecule and solvent regions was determined by an automated procedure (Wang, 1985). The solvent mask was updated every three cycles of solvent flattening. At the end of the 12th cycle the mean difference from the MIR phases was 39.7° and the average figure of merit was 0.84. An attempt was made to interpret the map although it was of poor quality. There are two major reasons for the poor quality of this map. Firstly, it was essentially a single isomorphous derivative map with only one major site. Secondly, molecules related by the twofold screw axes parallel to the unit-cell axes *a* and *b* have a number of very close contacts leading to errors in the determination of the molecular boundary. Later inspection of the molecular envelope of the DOT structure revealed that a major error in the calculation of the molecular envelope by the automated procedure was a shift of approximately 9 Å along the *c* axis.

When the coordinates of HLT (Anderson *et al.*, 1987) and RST (Bailey *et al.*, 1988) became available it was decided to solve the structure using molecular replacement. In the final refined structure of DOT the major site for the platinum derivative is within 3.5 Å of the side chain of Met626.

3.4. *The molecular replacement method*

The *MERLOT* package developed by Fitzgerald (1988) was used to determine the structure of duck ovotransferrin by molecular replacement. Six major steps were followed.

(*a*) Throughout the molecular-replacement studies, the N-terminal residues 1–4 which precede the first β-strand, residues 333–344 linking the two lobes, residues 672–686 which include the final α-helix, the two Fe^{3+} ions, and the two CO_3^{2-} ions were omitted from the search model. All other protein atoms were included and a temperature factor of 10 Å^2 was applied to all atoms. The N- and C-lobes were defined as residues 1–332 and 345–672, respectively (Table 1).

(*b*) The Crowther fast-rotation function (Crowther, 1977) was used to obtain the initial rotational orientations of the N and C lobes separately. The rotation function was calculated using data from 8–4 Å and search intervals of 2.5°. The test structure data of HLT was prepared by placing the structures of the N and C lobes and of the whole molecule in *P*1 unit cells of 120 × 120 × 120 and 150 × 150 × 150 Å, respectively. The rotation function did not contain any obvious solution when using the whole HLT molecule, whereas different single peaks about 25% above the level of the next highest peak were obtained when using either only the N-lobe or only the C-lobe (Fig. 2). These two peaks correspond to the two highest peaks in the rotation function using the whole molecule as the test structure and it became apparent that the relative orientations of the N and C lobes are different for the HLT and DOT structures.

(*c*) The Lattman rotation function (Lattman, Nockolds, Krestinger & Love, 1971) was used to obtain more precise orientations of the N- and C-lobes. The

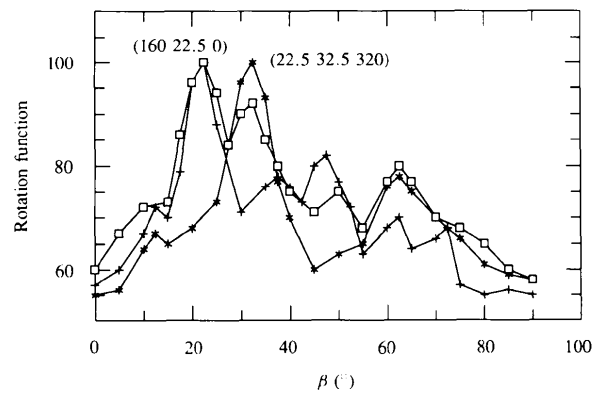

Fig. 2. The Crowther rotation function, calculated using the coordinates of HLT as the search model. (*a*) The complete molecule (□). (*b*) The N-lobe (+). (*c*) The C-lobe (*). The highest peaks of the rotation functions are at ($\alpha = 160$, $\beta = 22.5$, $\gamma = 0.0°$) and ($\alpha = 22.5$, $\beta = 32.5$, $\gamma = 320°$) corresponding to the N- and C-lobes respectively.

rotation search was conducted around the peak positions at 0.5° intervals.

(*d*) The molecular positions were determined using the translation function of Crowther & Blow (1967). In this search the asymmetric unit of the DOT structure was assumed to have two independent molecules (N- and C-lobes) related by a defined rotation. The corresponding peaks of the cross vectors between different molecules were about 42% higher than the next highest peak.

(*e*) The *R*-factor minimization procedure was used to further refine the rotation/translation function parameters (Ward, Wishner, Lattmann & Love, 1975).

(*f*) In order to check the solution the rotated and translated N and C-lobes were placed as one molecule in a *P*1 unit cell of 150 × 150 × 150 Å and the rotation/translation functions were calculated. The highest peak in each search was about 42% above the level of the next highest peak.

3.5. *Constrained refinement*

The refined coordinates of the N- and C-lobes of the HLT model were treated as two rigid groups and refined using the *CORELS* refinement program (Sussman, Holbrook, Church & Kim, 1977). All 1627 reflections in the resolution range 10–6 Å were used in the refinement. Six cycles of *CORELS* refinement only lowered the *R* factor from 0.47 to 0.46, the shifts along the unit-cell axes, *a*, *b* and *c*, were about 0.08, 0.006 and 0.23 Å for both lobes, and the rotational/translational shifts in (φ, θ, ρ) were (0.73°, 0.33°, −0.088 Å) and (0.82°, −0.09°, 1.48 Å) for the N- and C-lobes, respectively.

3.6. *Amino-acid replacement, model building and refinement*

At this stage the HLT sequence was replaced by the hen ovotransferrin sequence (Williams *et al.*, 1982) using the *REPLACE* and *REFINE* options of the program *FRODO* (Jones, 1978). Coordinates were retained for the atoms in common between the two sequences and calculated for new atoms. These initial coordinates with the HOT sequence in the DOT structure were refined using the molecular dynamics method of the program *X-PLOR* 2.1 (Brünger, 1988). Four groups of rigid bodies (N1, N2, C1 and C2) were defined (Table 1) and 15, 25 or 25 steps of rigid-body refinement were performed using the whole molecule, N- and C-lobes, or domains N1, N2, C1, C2, respectively, as rigid bodies. This was followed by several cycles of energy minimization and molecular-dynamics refinements. A slow-cooling option was employed and the temperature of the system was reduced from 2000 to 300 K over a 2 ps time period. The refinement was performed according to the example given in the *X-PLOR* V2.1 manual (Brünger, 1988). After three cycles of refinement the *R* factor for data between 8.0 and 3.0 Å resolution was 0.25. The HOT sequence was replaced by the duck ovotransferrin sequence as it became available. After four cycles of molecular dynamics refinement using all data between 10.0 and 2.35 Å the *R* factor dropped from 0.35 to 0.25.

Table 2. *Refinement statistics and the standard deviations for the final coordinates with the target values in parentheses using the program PROLSQ (Konnert & Hendrickson, 1980)*

Resolution limits (Å)	10.0–2.35
Final *R* factor	0.22
Free *R* factor	0.32
No. of reflections ($I > 3\sigma$)	25400
No. of protein atoms	5300
No. of solvent molecules	318
No. of sugar residues	2 NAG, 1 FUC
No. of ions	$2\,Fe^{3+}$, $2\,CO_3^{2-}$
Average *B* factor (Å^2)	
All atoms	28.7
Protein atoms, N-lobe	27.7 for the main-chain atoms
	28.7 for the side-chain atoms
Protein atoms, C-lobe	26.9 for the main-chain atoms
	28.6 for the side-chain atoms
Bond lengths (Å)	0.023 (0.023)
Angle distances (Å)	0.047 (0.030)
Planarity (Å)	0.026 (0.030)
Chiral volume (Å^3)	0.197 (0.120)
Non-bonded contacts (Å)	
Single torsion contacts	0.180 (0.200)
Multiple torsion contacts	0.237 (0.200)
Torsion angles (°)	
Peptide plane	7.4 (5.0)
Staggered	22.7 (10.0)
Orthonormal	27.9 (35.0)

In order to optimize the geometry a least-squares refinement using the stereochemically restrained techniques of the program *PROLSQ* (Konnert & Hendrickson, 1980) was performed including refinement of individual temperature factors. Alternating cycles of *PROLSQ* refinement and manual rebuilding were used in the refinement process. After each cycle of *PROLSQ* refinement, the $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ electron-density maps were inspected and improvement of the model was achieved by manually correcting the misplaced atoms using *FRODO* (Jones, 1978).

Early in the process the peptide link between the N- and C-lobes, corresponding to residues 333–344, and some segments of the surface loops had high temperature factors greater than 60 Å^2. Attempts to model these parts of the structure included deletion of part of the structure and reintroduction of residues on the basis of subsequent $2|F_o| - F_c|$ and $|F_o| - |F_c|$ difference maps. The final model was refined to give an *R* factor of 0.24. The introduction of 318 water molecules reduced the *R*

factor to 0.23 for all 27564 observed reflections between 10.0 and 2.35. The final R factor is 0.22 for 25400 reflections between 10.0 and 2.35 Å and $I > 3\sigma$. The r.m.s. differences between bond lengths and angle distances for the model and ideal values are 0.023 and 0.047 Å, respectively. Table 2 lists the standard deviations and the geometrical parameters after the final least-squares refinement. All non-glycine residues with the exception of two Leu residues have conformational angles (φ, ψ) within the allowed regions (Ramakrishnan & Ramachandran, 1965, Fig. 3).

The expected coordinate error estimated from Luzzati plots (Luzzati, 1952) is between 0.3 and 0.35 Å. The average main-chain B values are plotted, as a function of residue number in Fig. 4.

3.7. *Structural comparisons*

The DOT structure was compared with the RST and HLT structures using both complete molecules and individual lobes and domains. The relative orientation of the N-lobe to the C-lobe, the comparisons of the various structures, and transformations between the coordinates were performed using the *LSQKAB* programe of the *CCP*4 suite of programs (Collaborative Computational Project, Number 4, 1994).

When 530 Cα's, 77% of all residues, including all those involved in regular secondary structure in both lobes (61%) are superimposed there is an average displacement of 1.35 Å between DOT and RST compared with an average displacement of 2.45 Å between DOT and HLT. The r.m.s. deviations between the atomic positions in each of the four domains (N1, N2, C1, C2), the two lobes (N, C) and the structures (N and C) of HLT and RST with DOT are shown in Table 3. In order to superimpose the C-lobe on the N-lobe in the DOT structure a rotation of 169° followed by a translation of 27 Å is required. The corresponding figures for HLT and RST are rotations of 180 and 167° followed by translations of 25 and 23 Å, respectively (Anderson *et al.*, 1989; Bailey *et al.*, 1988), Table 3. In each case the value of the translation will depend on the precise location of the rotation axis. If the N-lobes of each of the three structures, DOT, HLT and RST are superimposed using 247 Cα atoms, the C-lobes of the HLT and RST structures require rotations of 19.2 and 10.4°, respectively in order to superimpose with the C-lobe of the DOT structure (Table 3).

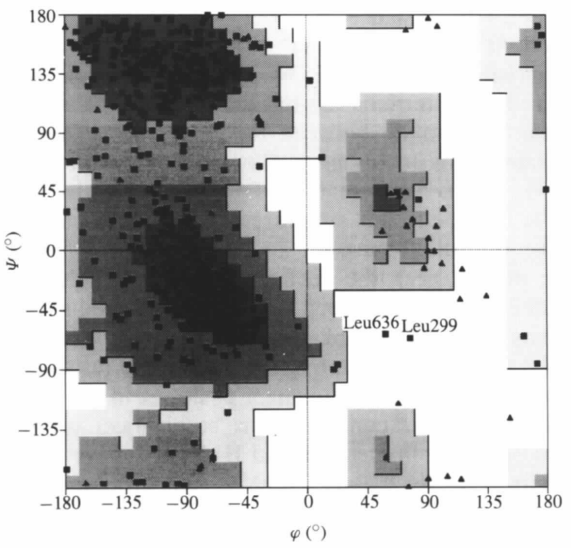

Fig. 3. The (φ, ψ) Ramachandran plot for the final refined model of DOT. Glycine residues are indicated by black triangles and other residues by black squares. The regions of the plot shown in dark grey indicate the fully allowed (φ, ψ) combinations, those in lighter grey indicate the additionally allowed (φ, ψ) combination, and the regions shaded in light grey indicate the generously allowed regions of the conformational space. Out of 605 non-glycine and non-proline residues 470 (77.7%) had torsion angles in the fully allowed regions of the plot, with a further 113 (18.7%) in the additionally allowed regions. The plot was calculated with the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993).

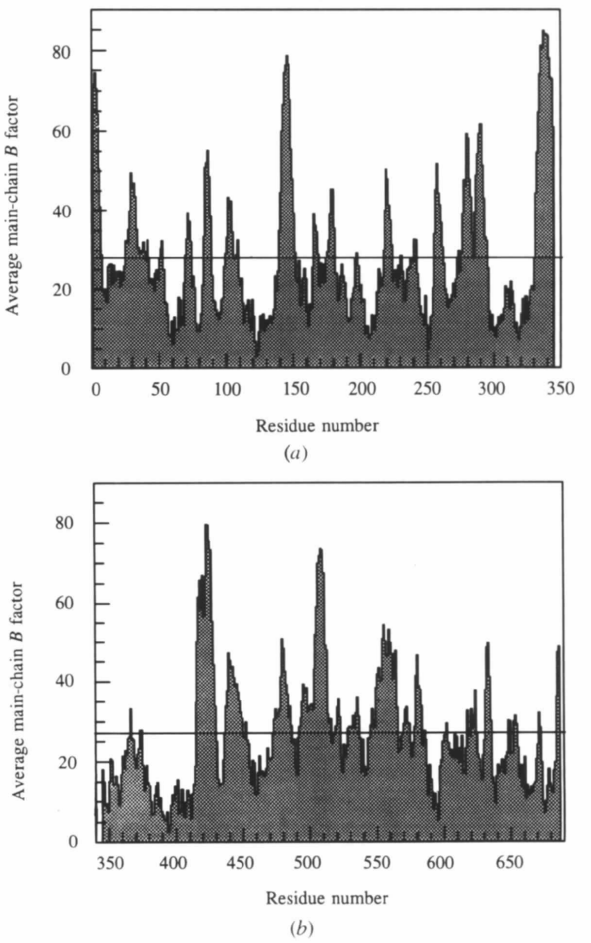

Fig. 4. Plots of the average main-chain B factor (Å^2) against residue number. For residue range 1–344 and 345–686 in (*a*) and (*b*), respectively. The horizontal line is the average B factor.

Table 3. *Superposition of domains, lobes and the complete structures of HLT and RST on to DOT structure*

		N1	N2	C1	C2	N	C	N&C
No. of Cα atoms used		127	120	130	131	247	261	530
R.m.s. deviation (Å)	HLT	0.59	0.60	0.63	0.62	0.82	0.73	2.45
	RST	0.63	0.66	0.59	0.98	0.78	0.86	1.35

	Rotation (°) of the C-lobe to superimpose on the N-lobe	R.m.s. (Å) deviation between the atomic positions in the N- and C-lobes	No. of Cα atoms used in each lobe
DOT	169	1.33	239
HLT	180	1.25	285
RST	167	1.60	147

	Rotation (°) of the C-lobes of HLT and RST to superimpose on the C-lobe of DOT*	R.m.s. (Å) deviation between the atomic position	No. of Cα atoms
DOT	—	—	266
HLT	19.2	0.73	266
RST	10.4	0.65	266

* The HLT and RST structures were superimposed on the DOT structure using the N-lobes only prior to the calculations.

4. Discussion

4.1. *Fold of the polypeptide chain*

The final model includes all 686 residues in the polypeptide chain of DOT. A number of surface loops are ill defined as is the peptide link between the N- and C-lobes. There is interpretable density for three carbohydrate moieties of a single glycan chain on the DOT structure at Asn473; this is the equivalent residue to that glycosylated in HOT (Heaphy & Williams, 1982). An *N*-acetylglucosmaine (NAG) moiety is covalently linked through C3 to the ND2 of Asn473 at the C-terminal end of helix 5 (Table 1) in the C-lobe. The position is indicated in Fig. 1.

The fold of the polypeptide chain in DOT is similar to that found in other transferrins as is the internal homology between the N- and C-lobes (Fig. 1 and Table 1). The nomenclature used in describing the structure is that defined by Anderson *et al.* (1987, 1989) and Bailey *et al.* (1988). The only difference in nomenclature is that strands *e* and *j* are defined as *e*1, *e*2, *j*1 and *j*2, where *e*1 and *j*1 form the interdomain link consisting of a two-stranded antiparallel sheet, *e*2 is in domain 2 and *j*2 is in domain 1. In addition residues 232–234 and 574–576 form an extra β-strand in domain 2 of each lobe (labelled *x* in Table 1). This means that both domains contain a mixed six-stranded β-sheet. Thus the domains have similar super-secondary structures consisting of the mixed β-sheet with connecting helices packed on both sides of the sheet. The two lobes comprise residues 1–332 (N-lobe) and 345–672 (C-lobe). Each lobe consists of two domains N1 (1–91, 248–332) and N2 (94–244); C1 (345–431, 589–672) and C2 (433–586), respectively, with the iron site in the cleft between them approximately 15 Å from the protein surface. The protein ligands for the Fe^{3+} and CO_3^{2-} ions, which are the same in both lobes, are the phenolate O atoms of Tyr191 (Tyr524) from helix 7 of domain 2 and Tyr92 (Tyr431) from an interdomain connecting strand (*e*1), a carboxylate O atom from Asp60 (Asp395) from strand *c* of domain 1, a neutral N atom of His250 (His592) from strand *j*2 of domain 1 (immediately after the second interdomain link) and the guanidinium group of Arg121 (Arg460) on helix 5 in domain 2 which interacts with the bicarbonate. The iron-binding residues are at the ends of elements of secondary structure. The Fe—Fe distance of about 40 Å is similar to that found in HLT and RST. The connecting peptide comprising residues 333–344 and helix H12_C (675–683) plus the terminating residues 684–686 labelled Ct in Fig. 1 also lie between the two lobes (Table 1). Peptide bonds involving prolines are trans for all but Pro71; the *cis* conformation of this proline is also found in the RST and HLT structures. Pro71 lies in the loop connecting the third helix and fourth strand of the N-lobe and is near the surface of the molecule. Two residues Leu299 and Leu636 have well defined electron density, but unfavourable Ramachandran angles of (77°, −64°) and (60°, −58°), respectively. These residues are located in γ-turns in equivalent positions in the N- and C-lobes. Similar conformations at identical positions were observed in the RST and HLT structures (Smith, Anderson, Baker & Baker, 1994). This position is in the long loop connecting helix 9 and strand *k* which forms part of the interdomain contact. The structure of the connecting peptide between the two lobes (333–344) is non-helical as in the RST structure and in contrast to the helical structure found in HLT. HLT is the only structure that does not contain a proline in the connecting peptide. DOT contains one proline residue, HOT and RST two prolines, while HST and MLT contain three prolines.

The superpositions of the C-lobe on the N-lobe of DOT and the individual domains and lobes of HLT and

RST on to DOT show that the individual lobes and domains have very similar structures (Table 3). The structure of DOT is closer to that of RST than it is to HLT. The main difference between the structures of HLT, RST and DOT is in the relative orientations of the N- and C-lobes. The rotation angles of the relative orientation of 169° and 167° for DOT and RST are similar as might be expected from their similar functions. The precise relative orientation of the N- and C-lobes must be influenced both by the structure of the connecting peptide and by the residues in the interface between the two lobes.

4.2. *Interactions between domains*

Transferrins have a range of iron-binding affinities with lactoferrin having the highest affinity at lower pH values (Brock, 1985; Crichton, 1991). In addition the two iron-binding sites of a single trasferrin molecule have slightly different properties. These differences could be due to the precise interactions between the two domains as well as to the interactions between the two lobes. A number of hydrogen bonds and salt bridges are formed between domains N1 and N2 and C1 and C2, respectively. These interactions are very similar to those found in the HLT and RST structures. The OD2 atoms of Asp60 and Asp395 are iron ligands. In both lobes the OD1 of the Asp forms an interdomain hydrogen bond.

When the N- and C-lobes of DOT are superimposed, as described above, it can be seen that the positions and orientations of the iron-binding residues are very similar in the two lobes (Fig. 5). The only major difference in the vicinity of the iron is in the orientations of the equivalent lysines 301 and 638. In the N-lobe well defined and continuous density is observed through and between the side chains of Lys209 and Lys301 with a distance of only 2.62 Å between the two NZ atoms (Figs. 6*a* and 7*a*). This interaction, which is within 7 Å of the iron-binding site, is only possible if one or both of the lysine residues is uncharged. The NZ atoms of Lys209 and Lys301 form hydrogen bonds with Ser303 OG and Tyr191 OH, respectively. A similar interaction occurs in the N-lobe of the RST structure (Lys296···Lys206, Sarra, Garratt, Gorinsky, Jhoti & Lindley, 1990) and in the monoferric N-terminal half-molecule of HOT (Dewan, Mikami, Hirose & Sacchettini, 1993). However, no such close interaction occurs in the C-lobe, where the side chain of Lys638 curls away from that of Lys541 (Figs. 6*b* and 7*b*) or in the N-lobe of HLT where Lys209 is replaced by Arg210. In the C-lobe of DOT the environment of the lysines is similar to that in the N-lobe except for the replacement of Ser303 by Leu640. This replacement results in Lys638 forming salt bridges with Asp634 and Glu547 while Lys541 interacts with a water molecule close to the iron-binding site. The resulting difference between the environment of the iron ligands Tyr191 and Tyr524 may account for some of the differences in the affinity of the two lobes.

4.3. *Interactions between lobes*

The connecting peptide in DOT as in HLT and RST is outside the main area of contact between the two lobes (Fig. 1). Helices play the major role in the interactions between the two lobes. The terminal helix in the C-lobe (H12_C, residues 675–683, Table 1), which includes residues Val679 and Phe682, runs between the surfaces of the two lobes (Fig. 1). This helix (H12_C) is in contact with residues 309–314 (C-terminus of strand *k* and part of subsequent loop) in the N-lobe and two helices H10_N (315–320) in the N-lobe and H2_C (378–385) in the C-lobe (Fig. 8). Similar interactions occur in the RST and HLT structures. However, there are some clear differences in the interlobe contacts. In HLT the helix H12_C is in contact with the helix H11_N while in DOT this helix H12_C is in contact with helix H3_C and has fewer interactions with the N-lobe. The interactions between the two lobes in DOT include one salt bridge between the carboxyl group of Asp315 on helix H10_N and the side-chain amino group of Lys386 on helix H2_C and hydrogen bonds between the guanidinium group of Arg246 on strand *j*_N and the main-chain carbonyl groups of Asp685 and Lys686 and

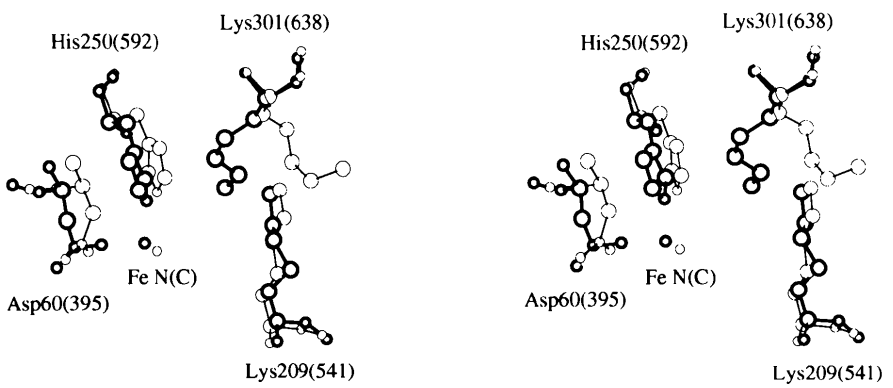

Fig. 5. Stereoview of some of the iron ligands plus the lysine residues in the vicinity after the superposition of the C-lobe on to the N-lobe using 70% of the Cα's. The N- and C-lobe residues are indicated by thick and thin lines respectively. The residue number in the C-lobe is given in brackets. The only major difference between the two lobes near to the Fe-binding site is in the orientation of the side chains of lysines 301 and 638.

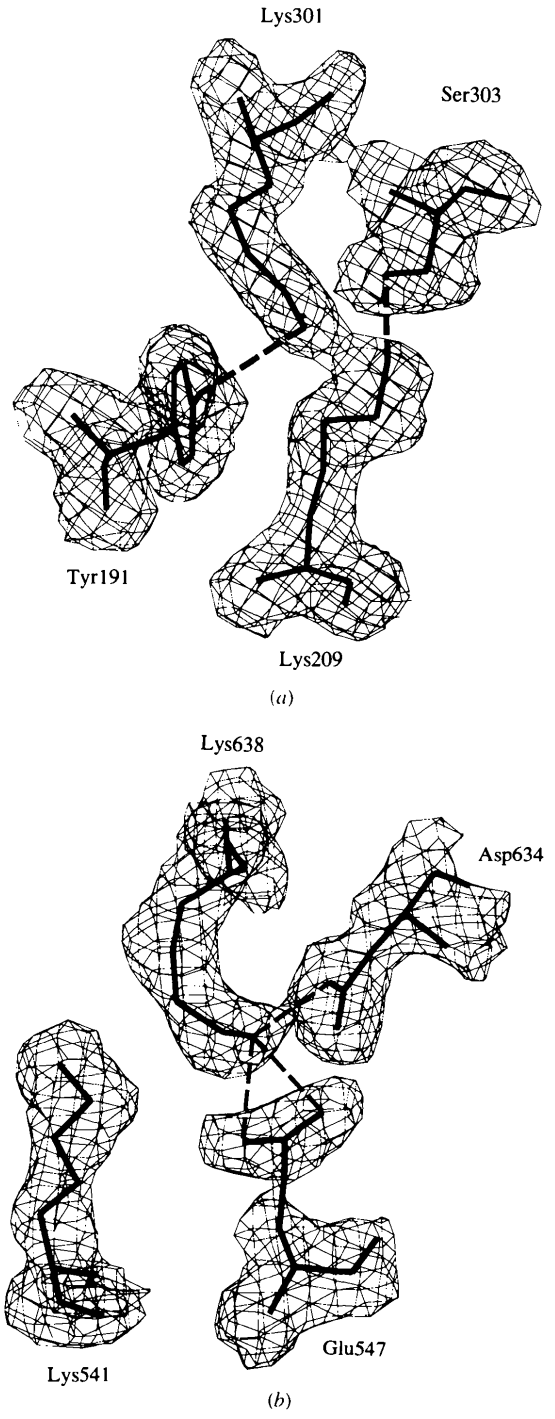

Fig. 6. The $2|F_o| - |F_c|$ electron-density map in the vicinity of the iron-binding site in the N-lobe and the C-lobe. Electron densities are contoured at 1.0σ level. (*a*) Strong density is present for the side chains of Lys301 and Lys209 in the N-lobe. The two NZ atoms are only 2.62 Å apart. This interaction will influence both the magnitude and the pH dependence of the affinity for Fe. (*b*) View of Lys541 and Lys638 in the C-lobe. The broken lines are the hydrogen bonds Lys209 NZ to Ser303 OG, and Lys301 NZ to Tyr191 OH in the N-lobe and Lys638 NZ to each of Asp634 OD1, Glu547 OE1 and Glu547 OE2 in the C-lobe.

between the carboxyl group of Asp685 and the main-chain amino group of Thr90. Hydrophobic interactions include Val679 and Phe682 on adjacent turns of helix H12_C packing into a hole formed by residues 308–311 and Leu313 (immediately preceding helix H10_N). The interactions of H12_C from the C-lobe and the terminating end residues of 684-686 with H10_N from the N-lobe contribute to the stability of the whole molecule. This explains the finding (Williams & Moreton, 1988) that while intact N-terminal and C-terminal half-molecule fragments of HOT will associate to form an N–C dimer, the N- and C-terminal fragments of HOT in which the residues 320–332 (carboxyl end of H10_N and all H11_N) and residues 683–686 (carboxyl

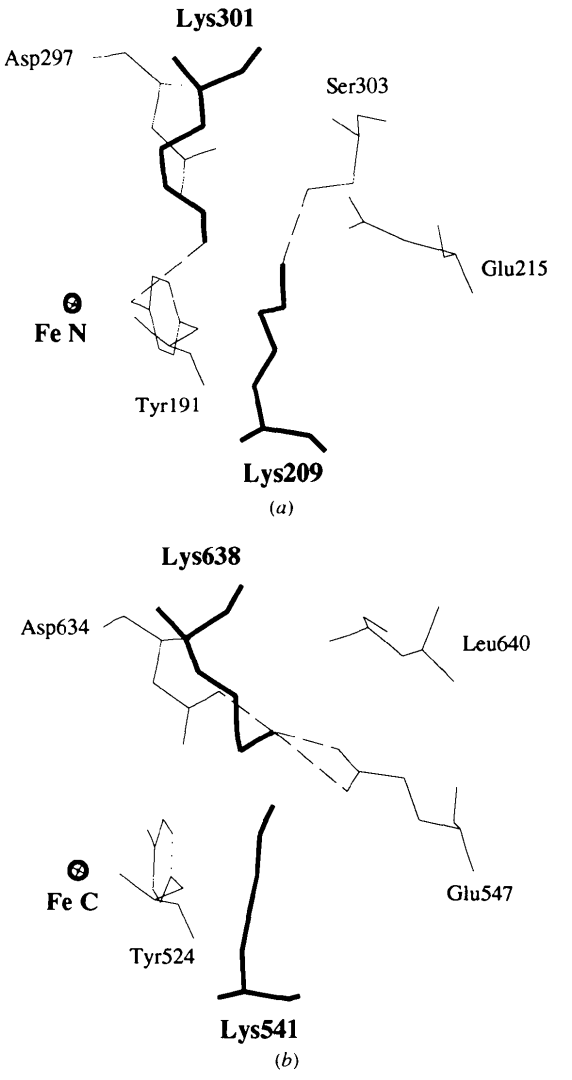

Fig. 7. (*a*) The environment of Lys301 and Lys209 in the N-lobe; all residues shown are conserved in DOT, HST, RST and HLT except for Lys209 in HLT which is replaced by Arg. (*b*) The environment of Lys541 and Lys638 in the C-lobe of DOT. The broken lines are the hydrogen bonds made by the lysine side chains.

end of H12_C) are removed by proteolytic digestion are unable to associate.

At the centre of this interface there is a close interaction between Phe322 on H11_N and the main chain of residue Lys386 and the conserved Gly387 at the end of helix H2_C (Figs. 8 and 9). Residue Lys386 is the lysine involved in an interlobe salt bridge with Asp315. Phe322 is replaced by Tyr317 in RST and by Ser322 in HLT, while Gly387 is conserved in all three structures. Superposition of the three structures indicates that this specific contact (H11_N against H2_C) is closer in LTF where a serine residue is present than in DOT and RST. This specific contact (H11_N against H2_C) with the presence of a large hydrophobic side chain in DOT and RST would explain why the relative orientations of the N- and C-lobes are more similar for DOT and RST than for DOT and HLT.*

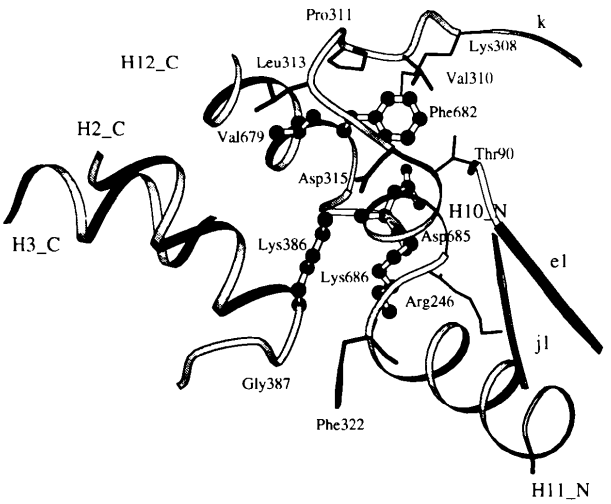

Fig. 8. Ribbon and ball-and-stick representation of the interface between the N- and C-lobes drawn with the program *MOLSCRIPT* (Kraulis, 1991). The C-lobe is to the left and the N-lobe to the right. Side chains from the N-lobe are represented by thick lines and those from the C-lobe are in ball-and-stick representation. Helix H12_C (675–686) forms part of the interlobe contact. A number of interactions are shown including the close contact between Phe322 on helix H11_N and the conserved Gly387 on helix H2_C. This close interaction together with a number of hydrogen bonds and salt bridges is important in determining the relative orientations of the two lobes.

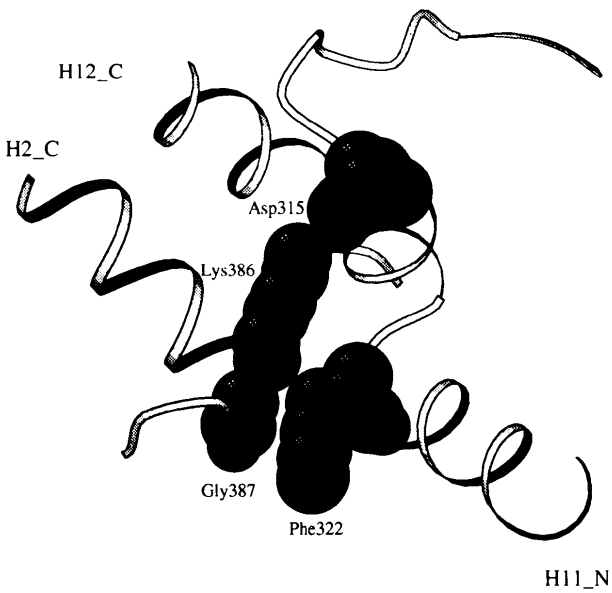

Fig. 9. Enlarged version of part of Fig. 8 showing the close contact between Gly387, the main chain of Lys386 and Phe322. Lys386 makes a salt bridge with Asp315. In lactoferrin, Phe322 is replaced by a serine residue and allows a different packing of the two lobes.

We thank E. N. Baker and his colleagues at Massey University, New Zealand, for providing the atomic coordinates of lactoferrin, P. F. Lindley and his colleagues at the Daresbury Laboratory and Birkbeck College, UK, for providing the atomic coordinates of serum transferrin and R. W. Evans and his colleagues at UMDS, Guy's Hospital, UK, for providing the amino-acid sequence of DOT. We are grateful to the SERC for funding and for synchrotron time at the SRS, Daresbury.

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1DOT, R1DOTSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: LI0214).

References

Anderson, B. F., Baker, H. M., Dodson, E. J., Norris, G. E., Rumball, S. V., Waters, J. M. & Baker, E. N. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 1768–1774.

Anderson, B. F., Baker, H. M., Norris, G. E., Rice, D. W. & Baker, E. N. (1989). *J. Mol. Biol.* **209**, 711–734.

Bailey, S., Evans, R. W., Garratt, R. C., Gorinsky, B., Hasnain, S., Horsburgh, C., Jhoti, H., Lindley, P. F., Mydin, A., Sarra, R. & Watson, L. (1988). *Biochemistry*, **27**, 5804–5812.

Baker, E. N., Baker, H. M., Smith, C. A., Stebbins, M. R., Kahn, M., Hellstrom, K. E. & Hellstrom, I. (1992). *FEBS Lett.* **298**, 215–219.

Banfield, D. K., Chow, B. K., Funk, W. D., Robertson, K. A., Umelas, T. M., Woodworth, R. C. & MacGillivray, R. T. A. (1991). *Biochim. Biophys. Acta*, **1089**, 262–265.

Bartfeld, N. S. & Law, J. H. (1990). *J. Biol. Chem.* **265**, 21684–21691.

Brock, J. H. (1985). *Metalloproteins*, Part II, edited by P. Harrison, pp. 183–262. London: Macmillan.

Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Crichton, R. R. (1991). *Chemistry and Biology of the Transferrins*. In *Inorganic Biochemistry of Iron Metabolism*, pp. 101–119, Ellis Horwood Series in Inorganic Chemistry. Chichester: Ellis Horwood.

Crowther, R. A. (1977). *The Molecular Replacement Method,* edited by M. E. Rossman, pp. 173–178. New York: Gordon and Breach.

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

Dewan, J. C., Mikami, B., Hirose, M. & Sacchettini, J. C. (1993). *Biochemistry,* **32**, 11963–11968.

Dorland, L., Haverkamp, J., Vliegenthart, J. F. G., Spik, G., Fournet, B. & Montreuil, J. (1979). *Eur. J. Biochem.* **100**, 569–574.

Evans, R. W., Aitken, A. & Patel, K. J. (1988). *FEBS Lett.* **238**, 39–42.

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

Heaphy, S. & Williams, J. (1982). *Biochem. J.* **205**, 611–617.

Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–349.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lattman, E. E., Nockolds, C. E., Kretsinger, R. H. & Love, W. E. (1971). *J. Mol. Biol.* **60**, 271–277.

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

MacGillivray, R. T. A., Mendez, E., Shewable, J. G., Sinha, C. K., Lineback-Zins, J. & Brew, K. (1983). *J. Biol. Chem.* **258**, 3543–3553.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Metz-Boutigue, M. H., Jollés, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J. & Joll, P. (1984). *Eur. J. Biochem.* **145**, 659–676.

Pentecost, B. T. & Teng, C. T. (1987). *J. Biol. Chem.* **262**, 10134–10139.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Rawas, A., Moreton, K., Muirhead, H. & Williams, J. (1989). *J. Mol. Biol.* **208**, 213–214.

Rose, T. M., Plowman, G. D., Teplow, D. B., Dreyer, W. J., Hellstrom, K. E. & Brown, J. P. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 1261–1265.

Sarra, R., Garratt, R., Gorinsky, B., Jhoti, H. & Lindley, P. (1990). *Acta Cryst.* B**46**, 763–771.

Smith, C. A., Anderson, B. F., Baker, H. M. & Baker, E. N. (1994). *Acta Cryst.* D**50**, 302–316.

Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. (1977). *Acta Cryst.* A**33**, 800–804.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Ward, K. B., Wishner, B. C., Lattmann, E. E. & Love, W. E. (1975). *J. Mol. Biol.* **98**, 161–177.

Williams, J. (1982). *Trends Biochem. Sci.* **7**, 394–397.

Williams, J., Elleman, T. C., Kingston, I. B., Wilkins, A. G. & Kuhn, K. A. (1982). *Eur. J. Biochem.* **122**, 297–303.

Williams, J. & Moreton, K. (1988). *Biochem. J.* **251**, 849–855.

Yang, F., Lum, J. B., McGill, J. R., Moore, C. M., Naylor, S. L., Van Bragt, P. H., Baldwin, W. D. & Bowman, B. H. (1984). *Proc. Natl Acad. Sci. USA,* **81**, 2752–2756.