Acta Cryst. (1990). B46, 763–771

High-Resolution X-ray Studies on Rabbit Serum Transferrin: Preliminary Structure Analysis of the N-Terminal Half-Molecule at 2·3 Å Resolution

BY ROBERT SARRA,* RICHARD GARRATT, BEATRICE GORINSKY, HARREN JHOTI AND PETER LINDLEY

Department of Crystallography, Birkbeck College, University of London, Malet St, London WC1E 7HX, England

(*Received* 16 *February* 1990; *accepted* 4 *June* 1990)

Abstract

An iron-containing half-molecule fragment ($M_r \simeq$ 39 000) of rabbit serum transferrin has been crystallized using the hanging-drop vapour-diffusion technique. The crystals belong to the space group $P3_121$ with cell parameters $a = 67{\cdot}06$ (3), $c = 138{\cdot}33$ (11) Å. Partial amino-acid sequence analysis has shown that this fragment corresponds to the N-terminal lobe of the intact molecule. Intensity data, to a resolution of 2·3 Å, were collected photographically using the oscillation method. The structure was determined by molecular replacement using the coordinates of the N-terminal lobe of the undigested parent molecule as the search model. A preliminary least-squares refinement of the molecular-replacement solution resulted in an R factor of 28% for all data between 7·0 and 2·3 Å. The refinement was performed initially with both the iron atom and the carbonate anion excluded from the model. The resulting $|F_o| - |F_c|$ difference Fourier map was readily interpreted in the vicinity of the iron-binding site in terms of a carbonate anion directly bound to the iron in a bidentate fashion. The current R factor is 22·5% for all data between 7·0 and 2·3 Å. The carbonate anion with the four protein ligands (Asp63, Tyr95, Tyr188 and His249) forms a distorted octahedral arrangement around the Fe^{3+} cation. The anion appears to be 'locked' in position by the formation of hydrogen bonds to residues at the N-terminus of helix 5; these involve the side chains of Thr120 and Arg124 and the main-chain nitrogens of Ala126 and Gly127. The importance of this loop at the N-terminus of helix 5 in stabilizing the Fe^{3+} transferrin complex is further emphasized by the involvement of Ser125 in hydrogen bonding to the distal oxygen of Asp63; this is also an interdomain hydrogen bond. Water molecules have also been identified, some in close proximity to the iron-binding site, and in particular to the side chain of Arg124.

* Author to whom correspondence should be addressed.

0108-7681/90/060763-09$03.00

Introduction

Transferrins are a family of iron-binding glycoproteins present in a variety of physiological compartments (for recent reviews see Harris & Aisen, 1989; Aisen, 1989; Huebers & Finch, 1987; Brock, 1985). The crystal structures of two functionally distinct transferrins have recently been determined, rabbit serum transferrin at 3·3 Å (Bailey, Evans, Garratt, Gorinsky, Hasnain, Horsburgh, Jhoti, Lindley, Mydin, Sarra & Watson, 1988) and human lactoferrin at 2·8 Å (Anderson, Baker, Dodson, Norris, Rumball, Waters & Baker, 1987; Anderson, Baker, Norris, Rice & Baker, 1989). Both these

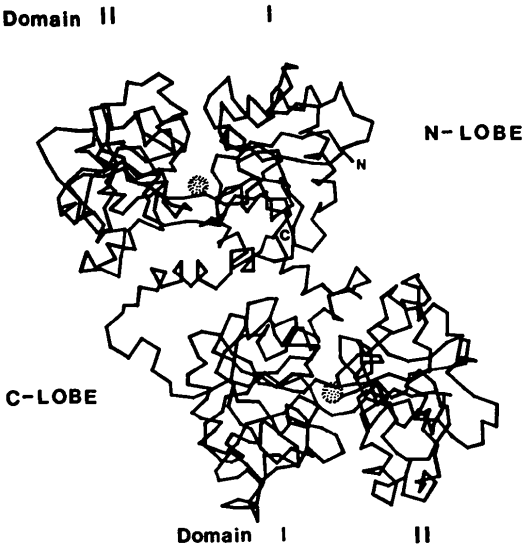

Fig. 1. The overall organization of the rabbit serum transferrin molecule is shown here as a Cα chain trace. The structure comprises two structurally homologous lobes (*ca* 330 residues each), resulting from an ancient gene-duplication event. The lobes are further divided into two dissimilar domains, which form a cleft inside which the iron-binding sites reside (the iron atoms are drawn as spheres). Also indicated are the N and C termini of the molecule.

proteins have very similar conformations, each consisting of two structurally homologous iron-binding lobes as shown in Fig. 1. Each lobe is further divided into two domains with the iron-binding site situated at the interdomain cleft. Characteristic of transferrin chemistry is the requirement for a synergistic anion (physiologically carbonate or bicarbonate) in the binding of Fe^{3+}. This was first demonstrated by Schade, Reinhart & Levy (1949) and since then many studies have been undertaken in order to determine the precise nature for this requirement (Harris & Aisen, 1989; and references therein). Recent work on the crystal structure of human lactoferrin (Anderson *et al.*, 1989) has shown that it is the carbonate form of the anion which is involved and that this anion is bidentately bound to the iron. The various transferrins exhibit a range of iron-binding affinities with lactoferrin having the most acid-stable binding sites. However, it is not clear whether these differences are directly associated with differences in the iron-binding sites, in particular the mode of binding of the anion, and/or differences occurring at remoter parts of the structure, for example, within the interdomain cleft. The elucidation of the three-dimensional structures of different species of transferrin, to high resolution, should provide a structural basis which will enable some of these questions to be resolved.

Crystals of rabbit serum transferrin provide X-ray diffraction data to, at best, 3·0 Å. The crystals used in structure determination to 3·3 Å (Bailey *et al.*, 1988) were grown using the batch method (Al-Hilal, Baker, Carlisle, Gorinsky, Horsburgh, Lindley, Moss, Schneider & Stimpson, 1976) and belong to space group $P4_32_12$ with $a = 127{\cdot}4\,(3)$, $c = 145{\cdot}3\,(4)$ Å; the solvent content is high, 68%(*v*/*v*). Numerous attempts to alter the crystallization conditions in order to improve the quality of the diffraction patterns have met with only partial success. Specifically, the tetragonal crystals described above have also been grown using the hanging-drop vapour-diffusion technique from a solution, 4%(*w*/*v*) protein, 12%(*w*/*v*) polyethylene glycol 6000, 50 m*M* disodium piperazine-*N*,*N*′-bis(2-ethanesulfonate) buffer adjusted to pH 5·75 at 277 K. Again the diffracting power of these crystals is limited to about 3·0 Å. Under similar conditions, but with 16% polyethylene glycol 6000, orthorhombic crystals were obtained. These appeared as needles with typical dimensions of 0·1 × 0·1 × 0·5 mm belonging to the space group $P2_12_12_1$ with cell dimensions $a = 63{\cdot}1\,(2)$, $b = 101{\cdot}6\,(2)$, $c = 131{\cdot}6\,(2)$ Å and solvent content in the range 50–55%(*v*/*v*). These crystals diffract to give a significant increase in resolution, attributable possibly to the lower solvent content, but nevertheless at 2·8 Å this is still in the medium-resolution range.

Table 1. *Alignment of the N-terminal sequences of both the N-lobe and C-lobe of rabbit serum transferrin, using single letter codes for amino-acid residues*

The fragment sequences clearly match that of the N-lobe.

	^{1}											^{12}
N-lobe*	V	T	E	K	T	V	R	W	C	A	V	N
Fragment *A*	V	T	E	K	T	V	R	W	C	A	V	N
Fragment *B*	V	T	E	K	T	V	R	?	C	A	V	N
	^{337}											^{348}
C-lobe*	D	E	C	K	A	V	K	W	C	A	L	S

* Sequence derived from cDNA (R. MacGillvray, personal communication).

However, it has been shown by many workers that transferrins can be proteolytically cleaved to yield half-molecular-weight fragments corresponding to individual lobes, each containing a single iron-binding site (Harris & Aisen, 1989; and references therein). In some cases further cleavage can occur to produce stable quarter-sized fragments which can also bind iron (Evans & Madden, 1984; Legrand, Mazurier, Metz-Boutigue, Jolles, Jolles, Montreuil & Spik, 1984). It was therefore thought that crystallization of these molecular fragments might lead to better ordered, more closely packed crystals which would, in turn, give higher-resolution diffraction patterns and therefore finer details of the protein structure. In the case of the half-molecules it was assumed that the iron-binding site would be representative of that found in the uncleaved protein.

Fragments corresponding to both the N- and C-terminal lobes of rabbit serum transferrin can be prepared as described by Heaphy & Williams (1982). These authors showed that, starting from fully saturated diferric transferrin, it is possible to produce a fragment corresponding to the N-lobe of the molecule using subtilisin in *solution*. This fragment however, suffered a number of internal cleavages which could introduce excessive heterogeneity in the molecule and be detrimental in crystallization attempts. Keung, Azari & Phillips (1982) showed that for chicken ovotransferrin the number of internal cleavages could be significantly reduced if the subtilisin is first *immobilized* onto a resin. In the light of these findings it was decided to cleave rabbit serum transferrin using immobilized subtilisin Carlsberg. This procedure resulted in an iron-containing non-internally cleaved fragment of approximately 39 000 molecular weight. However, on further purification using anion-exchange chromatography (Mono Q, Pharmacia FPLC system) this fragment fraction could be resolved into two major components, fragments *A* and *B* (R. Sarra, unpublished work). Both fragments contain iron, $\lambda_{max} = 465$ nm, and are indistinguishable in terms of molecular weight. The

difference in charge between these two fragments may be a result of differential cleavage at the C-terminal end of the fragment resulting in the loss of small peptides. Partial amino-acid sequence analyses confirm that both these fragments are derived from the N-lobe of the intact molecule (Table 1); the C-lobe is completely destroyed by this process (*cf.* proteolysis in solution; Heaphy & Williams, 1982). Both these fragments have been crystallized (Sarra & Lindley, 1986) and diffract to at least 1·8 Å on an Elliott GX20 rotating-anode source.

Here we report a preliminary investigation on the three-dimensional structure of fragment *A* at 2·3 Å resolution.

Experimental

All computer programs used in this study, unless otherwise stated, were obtained from the Science and Engineering Research Council program suite for protein crystallography (*CCP*4), Daresbury Laboratory, England. Crystals of fragment *A* were grown using the hanging-drop vapour-diffusion technique from a solution of 4%(*w*/*v*) protein, 25%(*w*/*v*) polyethylene glycol 6000, 50 mM disodium piperazine-*N*′*N*-bis(2-ethanesulfonate) adjusted to pH 7·0 and kept at a constant temperature of 279 K. The crystals belong to the space group $P3_121$ with cell dimensions $a = 67{\cdot}06\ (3)$, $c = 138{\cdot}33\ (11)$ Å. Crystals of fragment *B* were also grown using similar conditions but at pH 6·0. Precession photographs indicate that, to at least 6·0 Å resolution, the diffraction patterns of both these fragment crystals are identical. This is consistent with the sequence information that they are both derived from the N-terminal lobe of the intact molecule. The density of these crystals, $D_m = 1{\cdot}18\ (6)$ g cm^{-3}, measured by flotation on layers of appropriate bromobenzene/toluene mixtures, indicates that there is one molecule in the asymmetric unit assuming a solvent content in the range 40–45%(*v*/*v*). At temperatures of 279 K the crystals are stable for at least 130 h in an X-ray beam generated by an Elliott GX20 rotating anode (40 mA, 40 kV) using Cu $K\alpha$ radiation. They are well ordered and diffract to 1·8 Å resolution. At ambient temperatures the crystals are relatively unstable and cease to diffract after approximately 20 h of exposure to X-rays.

Data collection

X-ray intensity data to a resolution of 2·3 Å were collected photographically using an Arndt–Wonacott oscillation camera (Arndt & Wonacott, 1977) with Ni-filtered Cu $K\alpha$ radiation ($\lambda = 1{\cdot}5418$ Å) and a crystal-to-film distance of 69·5 mm. A single crystal of fragment *A* maintained at 279 K was sufficient to provide an almost complete data set as shown in Table 2. The crystal, dimensions 0·5 × 0·5 × 1·0 mm, was mounted with its *c* axis parallel to the camera rotation axis, and its *a* axis parallel to the X-ray beam. Using this setting it is possible to collect 97% of the data after a total rotation of only 30° (see, for example, Munshi & Murthy, 1986). A total of 32° of data was collected using 2° oscillation steps with an exposure of 15000 s deg^{-1}; three films per film pack, Kodak no-screen type NS-59T, were used in order to improve the dynamic range.

Table 2. *Three-dimensional scaling statistics and proportion of data collected as a function of resolution*

(*a*) Three-dimensional scaling statistics

D_{min} (Å)	R_{merge} (%)	$\langle I\rangle$	$\sigma(I)$	N_{ind}	N_{hkl}	$I > 3\sigma(I)$ (%)
7·09	2·4	564	18·5	555	45	97·1
5·08	3·4	422	21·3	974	185	98·4
4·17	3·6	687	42·5	1212	332	98·3
3·62	3·8	557	33·4	1414	483	97·4
3·24	4·5	416	27·9	1576	602	95·6
2·96	5·8	233	20·7	1688	729	91·2
2·75	7·5	143	16·7	1834	841	86·4
2·57	10·1	97	15·8	1939	961	81·5
2·42	13·0	79	17·9	2090	1042	73·8
2·30	17·2	65	19·1	2133	1061	66·0
Overall	6·1	220		15395	6287	85·8

(*b*) Proportion of data collected

D_{min} (Å)	N_{ref} (theoretical)	N_{ref} (actual)	%
6·50	848	727	85·7
4·60	1405	1342	95·5
3·76	1759	1666	94·7
3·25	2090	1965	94·0
2·91	2292	2108	92·0
2·66	2568	2372	92·4
2·46	2780	2563	92·2
2·30	2945	2652	90·1
Overall for 26° of data			92·3

Notes: N_{ind} is the number of independent reflections measured; N_{hkl} is the number of reflections that were measured more then once, excluding Friedel-related pairs; R_{merge} is given by $\Sigma(I - \langle I\rangle)/\Sigma I$; N_{ref} is the number of reflections.

The photographic film data were processed using the *MOSFLM* suite of programs based on the Cambridge program *MOSCO* (Nyborg & Wonacott, 1977), and extensively re-written by A. J. Wonacott and colleagues at Imperial College. The films were digitized on a Joyce–Loebl Scandig 3 microdensitometer (Imperial College) using a raster step size of 50 μm and an optical-density range of 0·0 to 2·0 units. Profile-fitted intensities, in addition to the traditional integrated intensities, were also measured (Leslie, 1987). Films within film packs were scaled together using the program *ABSCALE*, where Lorentz and polarization corrections were also applied. Out of the 32° of data collected, only 26° (the first 13 film packs) were deemed useful; the reflections on the later films showed obvious signs of crystal deterioration. Scaling and merging of the symmetry-related reflections was performed using the Fox & Holmes (1966) algorithm (programs *ROTAVATA* and *AGROVATA*), where profile-fitted intensities were

used in preference to integrated intensities if $I < \langle I \rangle$. Details of the data-processing statistics are given in Table 2. No absorption corrections were applied.

Rotation function

The search model, comprising the N-terminal coordinates of the intact molecule (residues 1 to 330; rabbit serum transferrin numbering), was put into a *P*1 orthogonal cell with dimensions $a = 86$, $b = 94$, $c = 92$ Å. Structure factors for the model were calculated using the program *GENSFC* to a resolution equal to that of the search model, 3·3 Å. The calculated and observed structure factors were normalized to $E(\mathbf{h})$ values using the program *ECALC*, where $E^2(\mathbf{h})| = |F(\mathbf{h})|^2/\langle|F(\mathbf{h})|^2\rangle$, and where average intensities are obtained in shells; no temperature factor was employed. It has been shown that normalized structure factors may be beneficial in molecular-replacement analyses (Tickle, 1985; White, Driessen, Slingsby, Moss, Turnell & Lindley, 1988), leading to a reduction in the background noise and allowing larger resolution ranges to be investigated.

Reciprocal-space cross-rotation searches were performed using the 60 Bessel function version of the program *ALMN*, based on Crowther's (1972) fast-rotation function as modified by E. Dodson. Initial searches were performed using a coarse 5·0° step size in all three Eulerian angles (α, β, γ), followed by a finer 2·5° search once a peak had been found. The resolution range 20–3·3 Å was used with an origin cutoff in the Patterson synthesis of 6·0 Å; Patterson radii of 30–35 Å were equally effective. In this study the Patterson synthesis of the *P*1 cell was rotated against that of a $P3_121$ cell (or the enantiomorph), giving rise to the rotation-function space group $Pbn2_1$ (Moss, 1985), with dimensions for the primitive Eulerian cell, $\alpha = 120$, $\beta = 360$ and $\gamma = 360°$. Maps were calculated over a single asymmetric unit, defined by the angular ranges, $\alpha = 0$–120, $\beta = 0$–90 and $\gamma = 0$–$360°$.

Translation function

The model coordinates after being transformed by the rotation angles previously determined in the rotation function, were put into the unit cell of the target molecule with *P*1 symmetry. Structure factors were calculated again to a resolution of 3·3 Å. The partial structure factors for each general equivalent position, in the space group *P*321 (true space group with translational components omitted), were calculated using the programs *PREPARE* and *COLLATE*. The translation function, with self vectors subtracted, was performed using the program *TFSGEN* (Tickle, 1985). This produces Fourier coefficients for a modified Crowther & Blow T_2 function (Crowther & Blow, 1967), from which three-dimensional translation function maps can be calculated using fast Fourier methods (Ten Eyck, 1973). At this point, owing to the ambiguity in the space group, translational components for the general equivalent positions were specified and Fourier coefficients calculated for both possible space groups, $P3_121$ and $P3_221$. The fact that the origin may be arbitrarily defined in the *z* direction at either $z = 0$ or $z = \frac{1}{2}$ meant that it was necessary to search over only half the unit cell along *z*. As with the rotation function normalized structure factors were used (these were calculated internally by *TFSGEN*).

Preliminary least-squares refinement

Reciprocal-space refinement was performed using the in-house least-squares program *RESTRAIN* (Moss & Morphew, 1982; Haneef, Moss, Stanford & Borkakoti, 1985). The protocol followed was intended to alleviate gross errors in the model, namely those resulting from inaccuracies in the molecular-replacement parameters, and any relative domain movements caused by the proteolysis and/or crystal-packing forces. With this in mind the resolution ranges used in the refinement were chosen in order to increase the radius of convergence as much as possible. Constrained refinement, treating the model as a single rigid body, was used in the initial cycles to refine the six orientational and translational parameters. The two domains were then refined independently by dividing the model into three rigid bodies. The selection of these three segments was based upon the polypeptide fold of the molecule; residues 1–84 and 243–330 fold to form domain I, and residues 85–242 form domain II. Extensive use of constrained/restrained refinement was then employed, whereby the program allows atoms within each 'rigid body' to shift in a correlated manner. It was also hoped that any shifts in secondary structural units, if required, would be accommodated without the need for defining these units explicitly as independent rigid bodies. Corrections to the low-angle reflections due to disordered solvent scattering were not made, other than applying a low-angle cutoff of 7·0 Å. The refinement was concluded by a number of cycles using restraints only, followed by individual atomic isotropic temperature-factor refinement. Throughout the refinement all reflections were included and given equal weight. At this point it should be emphasized that the refinement was performed in the absence of both the anion and the Fe^{3+} cation, in the hope that density for these atoms would be observed in an $|F_o| - |F_c|$ difference Fourier map. At this time, the structure of the diferric rabbit serum transferrin at 3·3 Å resolution did not allow a confident assignment of the anion location.

Electron density maps were calculated using weighted Fourier coefficients, as described by Read (1986) and employing the program *SIMWT* (I. J. Tickle, unpublished work).

Results

Rotation function

Initially the search model consisted of the N-terminal lobe of the diferric protein which had been subjected to restrained least-squares refinement (*RESTRAIN*, Moss & Morphew, 1982) to give an R factor of 34·4% for data with $I > 2\sigma(I)$ between 8·0 and 3·3 Å. This model, although incomplete since it lacked some 30% of the side chains and contained some poorly defined loop regions, was sufficient to give the expected origin peak when its calculated Patterson synthesis was rotated against an equivalent observed synthesis for the diferric serum transferrin structure. However, it was not precise enough to give a clear solution to the cross-rotation function for the N-terminal fragment data. Further refinement of the intact differic molecule using a single pass of the 'simulated annealing' refinement technique (*XPLOR*, Brünger, 1988; Weiss & Brünger, 1989) enabled nearly all the side chains to be incorporated and gave an R factor of 26·0% (Jhoti, 1989). The use of the N-terminal lobe from this refinement as an improved search model resulted in an unambiguous solution to the cross-rotation function. A subsequent post-mortem revealed that the r.m.s. difference between the two search models was 2·4 Å and highlighted the requirement for correct search models for success with molecular-replacement techniques. The value of 2·4 Å is also well beyond the accepted radius of convergence for traditional least-squares techniques and although extensive manual rebuilding and further restrained refinement would probably have achieved a similar improved model, there is no doubt that in this case the use of the 'simulated-annealing' technique led to a considerable saving in computing and computer-graphics resources. The choice of cell dimensions for the search model is an important consideration in rotation-function calculations. Cross-vectors arising from neighbouring molecules should be excluded as much as possible in the Patterson synthesis, at least within the integrated volume, without making the unit cell excessively large. The dimensions of the search model, 44 × 52 × 50 Å, were determined simply by finding the limits of the orthogonal coordinates along X, Y and Z. To these values a further 42 Å was added, corresponding to 80% of the geometrical mean diameter, 53 Å, obtained by fitting a prolate ellipsoid to the coordinates using the method of Taylor, Thornton & Turnell (1983). The final dimensions for the hypothetical $P1$ cell were therefore 86 × 94 × 92 Å.

In order to estimate the optimum values for the parameters used in the rotation function, a series of cross-rotation functions were performed (model against differic transferrin data). These gave a clear solution with the best results using the resolution range 20–3·3 Å, an origin cut-off in the Patterson synthesis of 6·0 Å and a radius of integration of 35 Å. The cross-rotation function (model against fragment data) performed with these values yielded an unambiguous solution with the best results obtained using outer Patterson radii in the range 30–35 Å (slightly larger than the calculated geometrical mean radius of 26·5 Å; see Lifchitz, 1983). The final solution was obtained by performing searches in 2·5° steps in all three Eulerian angles, resulting in the angles, $\alpha = 40$, $\beta = 80$ and $\gamma = 82{\cdot}5°$. Fig. 2 shows the results of the rotation search in which the height of the solution peak was 6·7 × the r.m.s. density; the second highest peak was only 3·7 × r.m.s. These rotation angles were then applied to the model coordinates in preparation for the translation function.

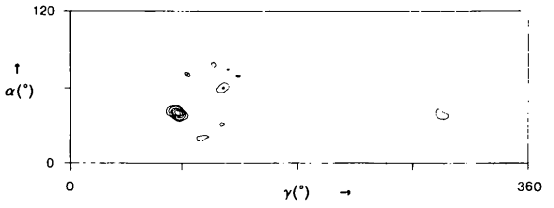

Fig. 2. Section, $\beta = 80{\cdot}0°$, of the cross-rotation function map (model against fragment data). For this particular map the parameters used were: resolution range 20–3·3 Å, Patterson vectors between 6·0–35 Å, step size of 2·5°, normalized structure factors used as coefficients. Contour levels are 2·5, 3·0, 4·0, 5·0 and 6·0 × r.m.s. density.

Positioning the model in the unknown unit cell

The position of the rotated model in the unknown unit cell was determined by the program *TFSGEN*, using data to a resolution of 3·3 Å. The computation of three-dimensional translation functions in this way greatly facilitates the determination of the translation vector, which can be obtained directly without the need to solve various Harker sections. This is of particular benefit when working with non-polar space groups containing high symmetry, as is the case in this study. Searches were performed in both possible space groups ($P3_121$ and $P3_221$) over half the unit cell in Z. The grid steps used along x, y and z were 0·65, 0·65 and 0·58 Å respectively. Only in the space group $P3_121$ was a single peak observed as shown in Fig. 3; this peak was 21 × r.m.s. density whereas the second highest peak was only 11·7 × r.m.s. For space group $P3_221$ the first and second highest peaks were 7·7 and 7·5 × r.m.s. density

respectively. The resulting translation vector (T_x = 14·2, T_y = 66·4 and T_z = 55·3 Å) was orthogonalized (T'_x = −19·0, T'_y = 57·5 and T'_z = 55·3 Å) and applied to the model coordinates. The molecular packing of this model in the unknown unit cell was checked using *FRODO* (Jones, 1978) implemented on an Evans and Sutherland Picture System 300, and showed no unacceptable close contacts.

Refinement of the molecular-replacement solution

The course followed in the refinement is summarized in Table 3. The refinement of the six rigid-body parameters at 10·0–3·5 Å resolution resulted in an R factor of 42%, suggesting that the molecular-replacement solution was authentic and further refinement would be worthwhile. Extensive use of constrained/restrained refinement followed this, treating the model as two independent domains, and extending the resolution limits gradually. This procedure resulted in an R factor of 33%, and a correlation coefficient of 0·81. The atomic positions were then refined using restraints only and the refinement converged rapidly with marginal improvement in the model, indicating that satisfactory convergence had been reached during the previous steps. The resolution of the data enabled the refinement of individual atomic temperature factors (U_{iso}), which improved the R factor to 28%; up to this point an overall temperature factor had been refined. The mean U_{iso} obtained was 0·25 Å^2 with 11% of the atoms (277 out of 2553) having values greater than 0·50 Å^2. The total r.m.s. shift from the starting coordinates was 1·05 Å, resulting in shifts of 4·4, 0·9 and −0·3° in α, β and γ; 0·7, −1·7 and 0·3 Å in T'_x, T'_y and T'_z. The total r.m.s. deviation in bond lengths from ideality was 0·017.

The overall fit of the refined model to the electron density (both $2|F_o| - |F_o|$ and $|F_o| - |F_o|$), was exceedingly good, but there were three particularly poor regions which corresponded to parts of the model with high isotropic temperature factors. The first two regions, residues 1–5 at the N terminus and residues 164–168, a surface loop within disulfide bridge 11, were ill-defined in the intact diferric rabbit serum transferrin structure used as the search model; the third region, residues 305–330, occurs at the C-terminus of the fragment. In the intact molecule, residues 305–331 form a well-ordered region containing helices 10 and 11 and anchored to domain II of the N-terminal lobe by disulfide bridge 10; this part of the molecule forms the main contact surface with helices 2 and 13 of the C-terminal lobe. The N-terminal fragment does not possess a C-terminal lobe and although this region is involved in forming intermolecular contacts, these are rather different to those found in the intact molecule, and may therefore result in a significantly different conformation. A further complicating factor, and possibly more significant, arises from the nature of the subtilisin cleavage. At the preparative stage it was hoped that the sole cleavage site would lie within the connecting peptide, residues 332–338, but the precise location is, as yet, unknown. The presence of Leu313 and Tyr314, an amino-acid doublet known to be very susceptible to cleavage by subtilisin Carlsberg (Morihara & Tsuzuki, 1969), presents a further option which may also lead to a radical alteration in the conformation of this region. At present, the $|F_o| - |F_o|$ electron density map does not present an unequivocal interpretation for this region, although the presence of small stretches of connected density support the suggestion of an alternative conformation; further studies are in progress.

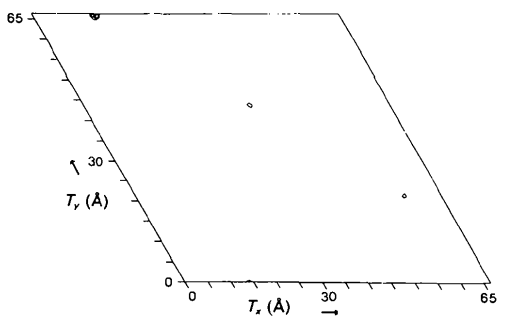

Fig. 3. Section, T_z = 55·3 Å, of the three-dimensional translation function map. E values to 3·3 Å used. Contour levels are 7·3, 12·7, 18·2 and 20 × r.m.s. density.

Table 3. *Stages followed during the least-squares refinement of the molecular-replacement solution*

The current R factor and correlation coefficient, after four rebuilds, are 22·5% and 0·92 respectively.

Step	R factor* (%)	Correlation coefficient†	Resolution range (Å)	Cycles	No. of reflections	Comments
0	46	0·40	10·0–5·0	–	1429	Without refinement
1	46	0·48	10·0–5·0	5	1429	Constrained rigid-body
2	43	0·53	10·0–4·5	6	2006	As step 1
3	42	0·51	10·0–3·5	3	4376	As step 1
4	41	0·53	10·0–3·5	5	4376	Constrained, three rigid bodies
5	25	0·82	10·0–3·5	8	4376	Constrained/ restrained, three rigid bodies
6	29	0·77	10·0–3·0	5	6929	As step 5
7	31	0·77	10·0–2·7	5	9410	As step 5
8	29	0·81	7·0–2·7	7	9014	As step 5
9	32	0·81	7·0–2·5	7	11342	As step 5
10	33	0·81	7·0–2·3	3	14464	As step 5
11	33	0·82	7·0–2·3	2	14464	Restraints only
12	28	0·86	7·0–2·3	5	14464	Restraints with atomic isotropic thermal parameters

* The R factor is defined as $\Sigma||F_o| - G|F_c||/\Sigma|F_o|$.

† The correlation coefficient is defined as $\{\Sigma(|F_o|G|F_c|) - \Sigma|F_o|\Sigma G|F_c|\}/\{[\Sigma|F_o|^2 - (\Sigma|F_o|)^2][\Sigma(G|F_c|)^2 - (\Sigma G|F_c|)^2]\}^{1/2}$. $G = 1/k$, where k is a scale factor applied to the observed structure factors. G was refined in each cycle coupled to an overall temperature factor, U_{iso}.

In the refinement no allowance was made for either the ferric ion or the synergistic anion, but the $|F_o| - |F_c|$ map calculated at the end of the refinement clearly showed a region of electron density, Fig. 4, which can readily be interpreted in terms of the metal ion and the anion bound to it directly in a bidentate fashion. The anion with the four protein ligands (Asp63, Tyr95, Tyr188 and His249; rabbit serum transferrin numbering) forms a distorted octahedral arrangement around the Fe^{3+} cation. In turn, the anion is involved in an elegant network of hydrogen bonds with residues at the N-terminus of helix number 5 as shown in Fig. 5. The atoms involved, OG1 of Thr120, NH2 of Arg124 and the main-chain nitrogens of Ala126 and Gly127, are all hydrogen donors and lie approximately in the anion plane. The importance of this region of the molecule in stabilizing the Fe^{3+}-transferrin complex is further emphasized by the involvement of Ser125 in hydrogen bonding to the distal oxygen of Asp63; this is also an interdomain hydrogen bond. This situation is very similar to that observed in human lactoferrin (Anderson *et al.*, 1989).

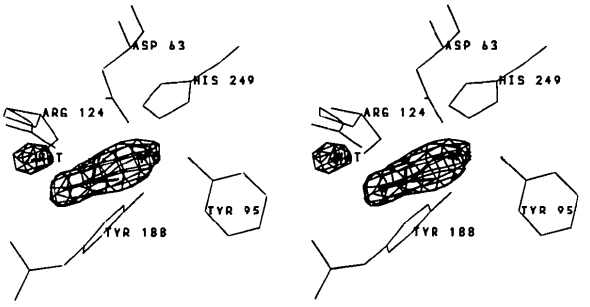

Fig. 4. Stereo diagram showing the $|F_o| - |F_c|$ map in the iron-binding site. The diagram also shows Arg124 and the four protein residues involved in iron binding. The density allows unambiguous positioning of the iron and carbonate atoms. Also shown is spherical density corresponding to a water molecule close to the guanidinium group of the arginine side chain. The contour level is 3·5 × r.m.s. density.

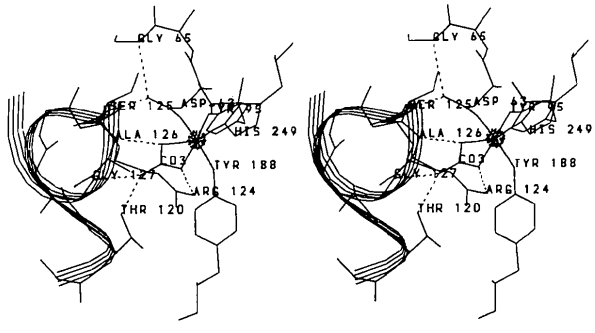

Fig. 5. Stereo diagram of the iron-binding site. Hydrogen bonds are represented as dashed lines. The iron atom is shown here as a sphere, with bonds to the ligands also drawn in. The N-terminus of helix 5 is highlighted by a ribbon trace.

From the $|F_o| - |F_c|$ map it has also been possible to identify a number of water molecules in the interdomain cleft, some in close proximity to the iron-binding site. In particular there are water molecules adjacent to the guanidinium moiety of Arg124.

Discussion

The fragment structure appears not to have been significantly perturbed during the proteolysis. This is indicated by the ease with which a solution was found in the molecular replacement, and the attainment of a relatively good *R* factor and correlation coefficient, 28 and 86% respectively, after the constrained/restrained least-squares refinement (with small radius of convergence) shown in Table 3 and with no rebuilding of the model. The lack of perturbation is not surprising if a lobe is considered as a stable structural unit which can fold in an independent manner. Recently an iron-binding fragment corresponding to the N-lobe of human serum transferrin has been produced using recombinant techniques (Funk, MacGillvray, Mason, Brown & Woodworth, 1989).

This high-resolution structure has enabled certain structural questions to be answered unambiguously and, in particular, the mode of binding of the anion. The anion binds directly to Fe^{3+} providing a further two non-protein ligands. The environment and hydrogen bonding of the anion strongly indicate that it is carbonate and not bicarbonate which is involved in the coordination. This is in accordance with spectroscopic studies and recent work on the three-dimensional structure of human lactoferrin (Anderson *et al.*, 1989; and references therein). In lactoferrin the carbonate is also 'locked' in position by hydrogen bonding to residues identical to those described here, namely those at the N-terminus of helix 5. In addition these residues are conserved in all transferrins sequenced to date (see, for example, Metz-Boutigue, Jolles, Mazurier, Schoentgen, Legrand, Spik, Montreuil & Jolles, 1984; Anderson *et al.*, 1989) and the attachment of the carbonate anion in this way may be a general feature in all members of the transferrin family. A notable exception is the C-lobe of melanotransferrin, a membrane bound protein (Rose, Plowman, Teplow, Dreyer, Hellstrom & Brown, 1986). In this protein Thr120 and Arg124 (which provide hydrogen bonds to the anion through their side chains) have been replaced by Ala and Ser respectively; also Asp63 is substituted by a Ser. It is not specifically known whether this extensively modified site maintains the ability for normal iron and carbonate binding, but this seems unlikely, although it may conceivably be involved in binding to hitherto unknown species. It is intriguing

to note that the three amino-acid substitutions found in the C-lobe of this protein, involve directly or indirectly residues at the N-terminus of helix 5. The importance of this helix is emphasized by the hydrogen bond formed between the main-chain nitrogen of Ser125 and the distal oxygen of the carboxylate of Asp63; this is also an interdomain hydrogen bond.

Water molecules have been positively identified within the interdomain cleft and partially filling this cavity. In particular, there are water molecules in close proximity to the iron-binding site hydrogen bonded to the side chain of Arg124. Spectroscopic evidence, in particular, EXAFS experiments on Fe^{3+}-chicken ovotransferrin (Hasnain, Evans, Garratt & Lindley, 1987) indicate a loss of a ligand when water is removed by freeze-drying. There is no evidence for direct coordination of water, and it is difficult to envisage a process where freeze-drying could remove a water molecule tightly bound to a Fe^{3+}, but it is conceivable that conformational changes resulting from the loss of water around Arg124 and/or in the cleft region in general might result in disruption of the hydrogen-bonding network. In turn, this could cause rotation of the anion leading to a decrease in the coordination at the ferric ion. This conformationally induced reduction in coordination might be an important step in iron release and uptake *in vivo*.

Further refinement of this structure is in progress, the current R factor is 22·5% with a correlation coefficient of 0·92; the results of this refinement will be published elsewhere.* Structural studies on other half-molecule transferrins including apo-fragments are being pursued. A comparison of their conformations may give some insight into the mechanism of iron uptake and release; the apo form of the C-terminal fragment of chicken ovotransferrin has been crystallized.

We gratefully acknowledge Dr E. N. Baker and colleagues, Massey University, New Zealand, with whom we have been able to exchange and discuss many ideas regarding structural and functional relationships in transferrins. We are also grateful for the help of Dr A. Wonacott and colleagues, Imperial College, University of London, England, with regard to the scanning and processing of the diffraction data, and of Dr J. B. Findlay, University of Leeds, England, for use of the SERC amino-acid sequencing facility. Many colleagues in the Department of Crystallography, Birkbeck College, London, England (Drs I. J. Tickle and H. Driessen), UMDS Division of Biochemistry, Guy's Hospital, London, England (Dr R. W. Evans), and SERC Daresbury Laboratory, Warrington, England (Dr S S. Hasnain), have contributed to the overall success of the structure analysis. This research was supported, in part, by the Science and Engineering Research Council and the Wellcome Trust.

* Atomic coordinates at the present stage of refinement have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1TFD), and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37036 (as microfiche). Free copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

References

AISEN, P. (1989). *Physical Bioinorganic Chemistry*, Vol. 5, *Iron Carriers and Iron Proteins*, edited by T. M. LOEHR, pp. 353–371. New York: VCH Publishers.

AL-HILAL, D., BAKER, E., CARLISLE, C. H., GORINSKY, B., HORSBURGH, R. C., LINDLEY, P. F., MOSS, D. S., SCHNEIDER, H. & STIMPSON, R. (1976). *J. Mol. Biol.* **108**, 255–257.

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.

ARNDT, U. W. & WONACOTT, A. L. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

BAILEY, S., EVANS, R. W., GARRATT, R. C., GORINSKY, B., HASNAIN, S., HORSBURGH, C., JHOTI, H., LINDLEY, P. F., MYDIN, A., SARRA, R. & WATSON, J. L. (1988). *Biochemistry*, **27**, 5804–5812.

BROCK, J. H. (1985). *Metalloproteins*, Part 2, edited by P. HARRISON, pp. 183–262. London: Macmillan.

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

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

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

EVANS, R. W. & MADDEN, A. D. (1984). *Biochem. Soc. Trans.* **12**, 661–662.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

FUNK, W. D., MACGILLVRAY, R. T. A., MASON, A. B., BROWN, S. A. & WOODWORTH, R. C. (1989). *Biochemistry*. In the press.

HANEEF, I., MOSS, D. S., STANFORD, M. J. & BORKAKOTI, N. (1985). *Acta Cryst.* A**41**, 426–433.

HARRIS, D. C. & AISEN, P. (1989). *Physical Bioinorganic Chemistry*, Vol. 5, *Iron Carriers and Iron Proteins*, edited by T. M. LOEHR, pp. 239–351. New York: VCH Publishers.

HASNAIN, S. S., EVANS, R. W., GARRATT, R. C. & LINDLEY, P. F. (1987). *Biochem. J.* **247**, 369–375.

HEAPHY, S. & WILLIAMS, J. (1982). *Biochem. J.* **205**, 611–617.

HUEBERS, H. A. & FINCH, C. A. (1987). *Physiol. Rev.* **67**, 520–582.

JHOTI, H. (1989). *Molecular Simulation and Protein Crystallography*, edited by J. GOODFELLOW, K. HENDRICK & R. HUBBARD, pp. 42–53. Daresbury: Science and Engineering Research Council.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 265–272.

KEUNG, W.-M., AZARI, P. & PHILLIPS, J. L. (1982). *J. Biol. Chem.* **257**, 1177–1183.

LEGRAND, D., MAZURIER, J., METZ-BOUTIGUE, M.-H., JOLLES, J., JOLLES, P., MONTREUIL, J. & SPIK, G. (1984). *Biochim. Biophys. Acta*, **787**, 90–96.

LESLIE, A. G. W. (1987). *Computational Aspects of Protein Crystal Data Analysis*, edited by J. R. HELLIWELL, P. A. MACHIN & M. Z. PAPIZ, pp. 39–50. Daresbury: Science and Engineering Research Council.

LIFCHITZ, A. (1983). *Acta Cryst.* A**39**, 130–139.

METZ-BOUTIGUE, M.-H., JOLLES, J., MAZURIER, J., SCHOENTGEN, D., LEGRAND, D., SPIK, G., MONTREUIL, J. & JOLLES, P. (1984). *Eur. J. Biochem.* **145**, 659–676.
MORIHARA, K. & TSUZUKI, H. (1969). *Arch. Biochem. Biophys.* **129**, 620–634.
MOSS, D. S. (1985). *Acta Cryst.* A**41**, 470–475.
MOSS, D. S. & MORPHEW, A. J. (1982). *Comput. Chem.* **6**, 1–3.
MUNSHI, S. K. & MURPHY, M. R. N. (1986). *J. Appl. Cryst.* **19**, 61–61.
NYBORG, J. & WONACOTT, A. J. (1977). *The Rotation Method in Crystallography*, edited by U. W. ARNDT & A. J. WONACOTT, pp. 139–152. Amsterdam: North-Holland.
READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
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. & LINDLEY, P. F. (1986). *J. Mol. Biol.* **188**, 727–728.
SCHADE, A. L., REINHART, R. W. & LEVY, H. (1949). *Arch. Biochem. Biophys.* **20**, 170–172.
TAYLOR, W. R., THORNTON, J. M. & TURNELL, W. G. (1983). *J. Mol. Graphics*, **1**, 5–8.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
TICKLE, I. J. (1985). *Molecular Replacement*, edited by P. A. MACHIN, pp. 22–26. Daresbury: Science and Engineering Research Council.
WEISS, W. I. & BRÜNGER, A. T. (1989). *Molecular Simulation and Protein Crystallography*, edited by J. GOODFELLOW, K. HENRICK & R. HUBBARD, pp. 16–28. Daresbury: Science and Engineering Research Council.
WHITE, H. E., DRIESSEN, H. P. C., SLINGSBY, C., MOSS, D. S., TURNELL, W. G. & LINDLEY, P. F. (1988). *Acta Cryst.* B**44**, 172–178.

Acta Cryst. (1990). B**46**, 771–780

Complexation with Diol Host Compounds. 5. Structures and Thermal Analyses of Inclusion Compounds of *trans*-9,10-Dihydroxy-9,10-diphenyl-9,10-dihydroanthracene with 2-Butanone, 4-Vinylpyridine, 4-Methylpyridine and 2-Methylpyridine

BY DIANNE R. BOND, MINO R. CAIRA, GRANT A. HARVEY AND LUIGI R. NASSIMBENI

Department of Chemistry, University of Cape Town, Rondebosch 7700, *South Africa*

AND FUMIO TODA

Department of Industrial Chemistry, Faculty of Engineering, Ehime University, Matsuyama 790, *Japan*

(*Received* 23 *April* 1990; *accepted* 13 *June* 1990)

Abstract

The structures of four inclusion compounds of *trans*-9,10-dihydroxy-9,10-diphenyl-9,10-dihydroanthracene with 2-butanone (1), 4-vinylpyridine (2), 4-methylpyridine (3) and 2-methylpyridine (4) have been determined by single-crystal X-ray diffraction and their thermal decomposition investigated by X-ray powder diffractometry, thermogravimetry and differential scanning calorimetry. Crystal data are: (1), $C_{26}H_{20}O_2.C_4H_8O$, $M_r = 436{\cdot}55$, triclinic, $P\bar{1}$, $a = 7{\cdot}924$ (1), $b = 8{\cdot}827$ (2), $c = 9{\cdot}050$ (1) Å, $\alpha = 109{\cdot}93$ (2), $\beta = 97{\cdot}31$ (1), $\gamma = 97{\cdot}65$ (1)°, $V = 579{\cdot}5$ (2) Å^3, $Z = 1$, $D_m = 1{\cdot}248$ (4), $D_x = 1{\cdot}251$ Mg m^{-3}, Mo $K\alpha$, $\lambda = 0{\cdot}7107$ Å, $\mu = 0{\cdot}074$ mm^{-1}, $F(000) = 232$, $T = 294$ K, final $R = 0{\cdot}059$ ($wR = 0{\cdot}077$) for 1690 independent reflections; (2), $C_{26}H_{20}O_2.2C_7H_7N$, $M_r = 574{\cdot}72$, monoclinic, $P2_1/c$, $a = 10{\cdot}045$ (2), $b = 18{\cdot}630$ (3), $c = 8{\cdot}419$ (4) Å, $\beta = 96{\cdot}29$ (3)°, $V = 1566{\cdot}0$ (9) Å^3, $Z = 2$, $D_m = 1{\cdot}208$ (3), $D_x = 1{\cdot}219$ Mg m^{-3}, Mo $K\alpha$, $\lambda = 0{\cdot}7107$ Å, $\mu = 0{\cdot}070$ mm^{-1}, $F(000) = 608$, $T = 294$ K, final $R = 0{\cdot}054$ ($wR = 0{\cdot}066$) for 1655 independent reflections; (3), $C_{26}H_{20}O_2.2C_6H_7N$, $M_r = 550{\cdot}67$, triclinic, $P\bar{1}$, $a = 8{\cdot}371$ (5), $b = 9{\cdot}599$ (2), $c = 10{\cdot}182$ (2) Å, $\alpha = 71{\cdot}12$ (1), $\beta = 82{\cdot}05$ (3), $\gamma = 74{\cdot}45$ (3)°, $V = 744{\cdot}6$ (5) Å^3, $Z = 1$, $D_m = 1{\cdot}226$ (3), $D_x = 1{\cdot}228$ Mg m^{-3}, Mo $K\alpha$, $\lambda = 0{\cdot}7107$ Å, $\mu = 0{\cdot}070$ mm^{-1}, $F(000) = 292$, $T = 294$ K, final $R = 0{\cdot}041$ ($wR = 0{\cdot}040$) for 1932 independent reflections; (4), $C_{26}H_{20}O_2.C_6H_7N$, $M_r = 457{\cdot}57$, triclinic, $P\bar{1}$, $a = 9{\cdot}337$ (2), $b = 10{\cdot}407$ (3), $c = 13{\cdot}470$ (9) Å, $\alpha = 70{\cdot}75$ (3), $\beta = 87{\cdot}65$ (3), $\gamma = 73{\cdot}58$ (2)°, $V = 1183{\cdot}2$ (9) Å^3, $Z = 2$, $D_m = 1{\cdot}271$ (4), $D_x = 1{\cdot}284$ Mg m^{-3}, Mo $K\alpha$, $\lambda = 0{\cdot}7107$ Å, $\mu = 0{\cdot}074$ mm^{-1}, $F(000) = 484$, $T = 294$ K, final $R = 0{\cdot}060$ ($wR = 0{\cdot}065$) for 2530 independent reflections. For compounds (2)–(4), a qualitative inverse correlation between the O···N hydrogen-bond distances and ΔH of the guest-release reaction is observed.

Introduction

The ability of the host *trans*-9,10-dihydroxy-9,10-diphenyl-9,10-dihydroanthracene (*H*) to form inclusion compounds with a variety of guest molecules (Toda, 1987; Tanaka, Toda & Mak, 1984; Toda,

0108-7681/90/060771-10$03.00