

## X-ray and primary structure of horse serum albumin (*Equus caballus*) at 0.27-nm resolution

Joseph X. HO<sup>1</sup>, Eugene W. HOLOWACHUK<sup>2</sup>, Elizabeth J. NORTON<sup>2</sup>, Pamela D. TWIGG<sup>1</sup> and Daniel C. CARTER<sup>1</sup>

<sup>1</sup> NASA, Space Science Laboratory, ES76 Biophysics Branch, Marshall Space Flight Center, Huntsville, USA

<sup>2</sup> Mary Imogene Bassett Hospital, Research Institute, Cooperstown, USA

(Received January 25, 1993) – EJB 93 0123/2

The amino-acid sequence and three-dimensional structure of equine serum albumin have been determined. The amino-acid sequence was deduced from cDNA isolated from equine liver. Comparisons of the primary structure of equine serum albumin with human serum albumin and bovine serum albumin reveal 76.1% and 73.9% sequence identity, respectively. The three-dimensional structure was determined crystallographically by the molecular-replacement method using molecular coordinates from the previously determined structure of human serum albumin, to a resolution of 0.27 nm. In accordance with the primary structure, the three-dimensional structures are highly conserved. There is a root-mean-square difference between  $\alpha$ -carbons of the two structures of 0.201 nm. The association constants ( $K_a$ ) for the binding of 2,3,5-triiodobenzoic acid were determined by ultra-filtration methods for equine and human serum albumins to be approximately  $10^4 \text{ M}^{-1}$  and  $10^5 \text{ M}^{-1}$ , respectively. Crystallographic studies of equine serum albumin reveal two binding sites for 2,3,5-triiodobenzoic acid identical with those previously reported for human serum albumin which are located within subdomains IIA and IIIA. Details and comparisons of the binding chemistry are discussed.

As the most abundant protein of the circulatory system albumin functions as a multipurpose transport molecule and is a principal contributor to colloid osmotic blood pressure. Its affinity for numerous small molecules is widely known. As extracellular proteins they are unusual in possessing a free sulphhydryl (Cys34), a large number of disulphides (17) and with rare exception, an absence of glycosylation sites. They are moderately large (65 kDa) proteins formed by a repeating domain structure which suggests that albumin is derived from a primordial gene duplication event involving a 22-kDa, 190-amino-acid proto-albumin [1]. Serum albumins are perhaps the most frequently studied and applied proteins in biochemistry largely because of their low cost and ready availability. For additional information see reviews by Peters [2] and Fehske et al. [3].

The sequences of a number of albumins have been determined including those of human serum albumin (HSA) [4–7], bovine serum albumin (BSA) [8] (Holowachuk E. W., unpublished results), ovine serum albumin (OSA) [9], porcine serum albumin (PSA) [10], rat serum albumin (RSA) [11], mouse serum albumin (MSA) [12], frog serum albumin [13], salmon serum albumin [14] and lamprey serum albumin [15]. In all of these albumins the characteristic repeating series of disulphides is strictly conserved leading to further expectation of a common topology. Indeed, for similar reasons one would expect that the other members of the albu-

min multigene family,  $\alpha$ -fetoprotein and vitamin-D-binding protein, will share a common three-dimensional structure with the albumins.

This work is part of a series of studies examining the structure, chemistry and evolution of serum albumin. Previously, the three-dimensional structure of HSA was determined revealing the three-domain, six-subdomain structure [16–18]. In addition, the crystallographic studies indicate that the principal binding regions of HSA reside within two of the six subdomains (IIA and IIIA). Here, we report the determination of both the primary and tertiary structures of equine serum albumin (ESA) based on the known atomic structure of human serum albumin [18]. Historically, ESA was among the first animal proteins crystallized. These crystals, originally reported by Gürber in 1894 [19], were formed from ammonium-sulphate precipitation and many years later crystallographically characterized by McClure and Craven [20]. The crystals utilized in this study are identical with those previously reported, but grown optimally under slightly different conditions.

As a family of proteins the serum albumins have been evolving rapidly. For example, lamprey and salmon serum albumins possess the characteristic repeating series of disulfide bridges with other albumins, but shows little other sequence identity. Despite the similarities of amino-acid sequence among mammalian albumins, there can be marked differences in the affinity among the albumins for various ligands. For example, canine serum albumin, unlike many of the albumins previously characterized, has no affinity for Cu or Ni, presumably due to the lack of His at position 3 [21]. Similarly, only the primate albumins have high affinity for heme [22]. HSA has an unusually high affinity for bili-

Correspondence to D. C. Carter, Chief, ES76 Biophysics, NASA, Marshall Space Flight Center, Huntsville, AL 35812, USA

Abbreviations. HSA, human serum albumin; BSA, bovine serum albumin; OSA, ovine serum albumin; PSA, porcine serum albumin; RSA, rat serum albumin; MSA, mouse serum albumin; rms, root-mean-square.

rubin which represents a major binding site distinct from several other albumins including ESA [23]. In addition, ESA is an exception among albumins studied so far in possessing no esterase activity associated with the highly conserved Tyr411 [24]. Originally, this esterase activity was discovered by Solene and Means [25] and results in the acetylation of Tyr411. A number of similar chemical reactions have been characterized including one involving the *p*-nitrophenyl esters of fatty acids [26]. Tyr411 of HSA is clearly the most reactive residue of the albumins studied [24]. Although very little of the ligand-binding chemistry has been conducted with ESA compared with the voluminous works on HSA and BSA, it is of interest to examine the comparative structure and chemistry of related albumins in order to gain further insight into the chemical basis for their extraordinary ligand-binding capabilities.

## MATERIALS AND METHODS

### Isolation and characterization of ESA cDNA

Horse liver biopsy samples, snap frozen in liquid N<sub>2</sub>, were obtained from Dr. M. Brunner (College of Veterinary Medicine, Cornell University, Ithaca, NY). Total cellular RNA was extracted into guanidine isothiocyanate and purified by centrifugation through CsCl [27] and poly(A)-rich RNA was purified by two rounds of affinity chromatography on oligo(dT)-cellulose [28]. First strand cDNA was synthesized from poly(A)-rich RNA, as previously described [29, 30] except for the use of oligo-(dT)-*Xho*I linker/primer (0.5 µg/µg mRNA), MLV reverse transcriptase, and the substitution of 5-methyl-dCTP for dCTP [31, 32]. Second-strand synthesis was initiated by the addition of RNase H, Klenow polymerase and dNTP [33]. The use of 5-methyl-dCTP in the first-strand reaction eliminates the need for site-specific methylation as hemimethylated cDNA is fully resistant to cleavage by methyl-sensitive restriction endonucleases such as *Eco*RI and *Xho*I. The double-stranded cDNA was flushed with T4 DNA polymerase and ligated to *Eco*RI adapters with T4 DNA ligase. The *Eco*RI adapter ends of the cDNA were phosphorylated with polynucleotide kinase and the cDNA digested with *Xho*I, generating cDNA with 5' *Eco*RI and 3' *Xho*I cohesive ends. The cDNA were size fractionated on sepharose-CL-4B columns and cDNA greater than 1000 bp were purified. The sized cDNA was ligated into the directional cDNA-cloning vector  $\lambda$  ZAP II [34] (Stratagene) previously cleaved with *Eco*RI and *Xho*I, and treated with alkaline phosphatase. The ligated DNA was packaged *in vitro* into infectious phage particles with Gigapack gold packaging extracts (Stratagene), and the library was plated on *Escherichia coli* K12 strain PLK-F' in order to determine the recombinant phage titer and to amplify the library.

The horse liver cDNA library was screened for albumin-specific cDNA by *in-situ* hybridization with a <sup>32</sup>P-labeled BSA cDNA probe of 830 bp, cleaved with *Eco*RI and *Nco*I from the 5' coding region of BSA cDNA (Holowachuk, E. W., unpublished results) in order to ensure the isolation of full-sized ESA cDNA clones as previously described [29, 30]. Full-size cDNA (≈2.2 kbp) were identified by amplification of the cDNA-cloning site of  $\lambda$  ZAP II and sizing the cDNA inserts on agarose gels [35].

Bacteriophage  $\lambda$  ZAP II carries the plasmid Bluescript SK (–) within which insert DNA are cloned, and which can be excised *in vivo* upon infection with M13 helper phage [34]. Bluescript SK (–) plasmids carrying ESA cDNA were

recovered from the recombinant  $\lambda$  ESA phage by infection with M13 R408 helper phage, as described [34]. The sequencing of the ESA cDNA was facilitated by the isolation of a series of 200–500-bp progressive unidirectional-deletion derivatives which extended across the cDNA subclones from an *Eco*RI site by treatment sequentially with *exo*III and Mung bean nuclease [36, 37]. Plasmid DNA minipreps were purified (Magic Minis, Promega Corp.) from detergent lysates prepared from overnight cultures for use as DNA-sequencing templates. Double-stranded nucleotide sequencing was performed by the dideoxynucleotide triphosphate method with modified T7 DNA polymerase and [<sup>32</sup>P]dATP [38]. All oligonucleotides were synthesized on a 40-nmol scale using the ABI model 391 DNA synthesizer. Nucleoside-3'-phosphoramidites, polystyrene supports and other synthesis reagents were from Applied Biosystems.

### Crystallization and data collection

Albumin used in these studies was obtained commercially from Sigma Chemical Company as recrystallized, essentially fatty-acid-free fraction V. Crystals of ESA were grown by the hanging-drop vapor-equilibration method. Solutions of 50–60 mg/ml ESA and 5 µl in volume were added to 5 µl precipitating solution of 55% saturated ammonium sulphate, 45% 0.05 M potassium phosphate, pH 7.0. This droplet was equilibrated to a 1-ml reservoir of the same 55% precipitating solution. Crystals are grown at 22°C and appear as hexagonal rods in 5–10 days. The crystals have Laue symmetry of P6/m and unit-cell constants of  $a = b = 9.66$  nm,  $c = 1.44$  nm,  $\alpha = \beta = 90^\circ$  and  $\gamma = 120.0^\circ$  and belong to the space group P6<sub>1</sub> or P6<sub>5</sub> as previously reported by McClure and Craven [20]. Co-crystals of ESA with the ligand 2,3,5-triiodobenzoic acid were produced in the same manner as above except with the addition of 4 mol 2,3,5-triiodobenzoic acid/mol ESA.

X-ray diffraction data were collected on a Siemens/Nicolet multi-wire and Rigaku RAXIS IIc area detectors operating on a Rigaku RU200 rotating anode using monochromated Cu K $\alpha$  radiation. Data were collected and merged from five crystals to give 176004 observations of 26408 unique reflections to 0.25-nm resolution with weighted Rmerge on intensity for  $I > 1.0 \sigma(I) = 10.21\%$ . After reducing the data to 0.25-nm resolution, data were truncated to 0.27 nm for completeness and reliability to give 19970  $> 1.0 \sigma(I)$  out of 20960 collected and 21031 theoretically possible (99.6% complete) for which 95% observed greater than  $1.0 \sigma(I)$ . Data for the triiodobenzoic acid co-crystalline complex were collected from a single crystal to give 47139 observations of 15094 unique reflections  $> 0.0 \sigma(I)$  to 0.30 nm (98% complete) of these data 12725  $> 1.0 \sigma(I)$  (82.4% complete) giving an Rmerge of 9.43%.

### Structure determination

The structure of ESA was determined by the molecular-replacement method [39] using the previously determined atomic coordinates of HSA [18] as a starting model. The orientation of the molecule was determined by the Crowther fast-rotation function [40] in the computer program package MERLOT [41] with good contrast as the highest peak in the rotation function. The solution was  $3.6\sigma$ , where  $\sigma$  is the square root of the variation of the rotation function for all possible orientations, with the next highest peak in the rotation function being 27% lower in peak height. This orienta-

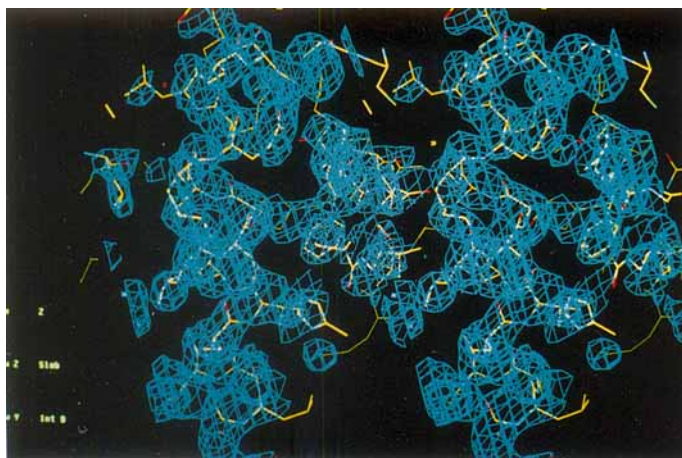


Fig. 1. Stereo view of the 2FO-FC electron density calculated at 0.27-nm resolution for the helical region Leu184–Cys 199. This figure was produced with the aid of the computer program FRODO [56].

tion was further fine adjusted using  $0.5^\circ$  grids by the Lattman rotation function [42]. The translation vector determination and proper space-group assignment were done in two different approaches. First, the Patterson translation function [43] was calculated on two different space groups,  $P6_1$  and  $P6_5$ . In the case of the space group  $P6_1$ , the two Harker sections gave consistent peaks with very good contrast (the second highest peaks were 52% and 40% lower, respectively), while no consistent solution could be found for the enantiomorphic space group  $P6_5$ . Secondly, correlation coefficient and  $R$ -factor searches were performed for both space groups. Again, in the case of the space group  $P6_1$ , the search produced the single solution consistent with the first approach, which gave the correlation coefficient 52.9% and  $R$ -factor 42.2% for 5299 reflections with the resolution range of 2.0–0.4 nm. In the case of the space group  $P6_5$ , the best correlation coefficient was 35.1% and the best  $R$ -factor 47.7%, and the corresponding translation vectors were inconsistent. After resolving the space-group ambiguity and placing the molecule in the unit cell as described, the model was subjected to 10 cycles of refinement as a rigid body giving the  $R$ -factor of 41.6% and correlation coefficient of 54.3% for the same data set.

The side chains of the amino acids of the molecule were then replaced by the newly determined amino-acid sequence of ESA with the aid of computer graphics. It was then put through a cycle of simulated annealing using program XPLOR [44]. This was followed by refitting the structure through a series of omit maps where 30–40 amino acids were deleted from the structure at a time and the remaining structure was refined by constrained-restrained refinement for several cycles using the program XPLOR or Hendrickson-Konnert least-squares program [45]. The refitted model was again subjected to the simulated annealing procedure. The current crystallographic  $R$ -factor is 23.8% for 19152 reflections with  $F > 2.0 \sigma(F)$  and resolution range from 1.0 nm to 0.27 nm. The root-mean-square deviation from ideality is 0.0026 nm for bond distance and  $5.3^\circ$  for bond angle. The general quality of the electron density is illustrated in Fig. 1.

Table 1. Data from ultrafiltration experiments comparing the binding of triiodobenzoic acid to human and equine serum albumins.  $[D_u]$ , molar concentration of free triiodobenzoic acid;  $[D_b]$ , molar concentration of bound triiodobenzoic acid;  $r$ , mol bound triiodobenzoic acid/mol serum albumin;  $A_{260}^{1\%}$  for triiodobenzoic acid, 144.2.

| Protein | $A_{260}$<br>of<br>filtrate | $[D_u]$                | $[D_b]$                | $r$     |
|---------|-----------------------------|------------------------|------------------------|---------|
|         |                             | M                      |                        | mol/mol |
| HSA     | 0.009                       | $1.168 \times 10^{-6}$ | $1.186 \times 10^{-4}$ | 0.538   |
|         | 0.012                       | $1.557 \times 10^{-6}$ | $2.380 \times 10^{-4}$ | 1.079   |
|         | 0.039                       | $5.060 \times 10^{-6}$ | $4.791 \times 10^{-4}$ | 2.149   |
|         | 0.285                       | $3.698 \times 10^{-5}$ | $6.817 \times 10^{-4}$ | 3.090   |
|         | 0.879                       | $1.140 \times 10^{-4}$ | $8.442 \times 10^{-4}$ | 3.827   |
| ESA     | 0.022                       | $2.854 \times 10^{-6}$ | $1.169 \times 10^{-4}$ | 0.530   |
|         | 0.020                       | $2.595 \times 10^{-6}$ | $2.369 \times 10^{-4}$ | 1.074   |
|         | 0.100                       | $1.298 \times 10^{-5}$ | $4.661 \times 10^{-4}$ | 2.113   |
|         | 0.290                       | $3.763 \times 10^{-5}$ | $6.810 \times 10^{-4}$ | 3.087   |
|         | 0.615                       | $7.980 \times 10^{-5}$ | $8.784 \times 10^{-4}$ | 3.982   |

### Ultrafiltration experiments with 2,3,5-triiodobenzoic acid

Ultrafiltration experiments were performed on both ESA and HSA using 2,3,5-triiodobenzoic acid (for a review see [46]). The albumins (essentially fatty-acid free) and triiodobenzoic acid were purchased from Sigma Chemical Co. Solutions were prepared in 4 ml and buffered with 0.05 M sodium phosphate, pH 7.4, using protein concentrations of 15 mg/ml (molar ratios of triiodobenzoic acid to ESA of approximately 0.5:1, 1:1, 2:1 and 3:1). All samples were equilibrated for 4 h at room temperature. Ultrafiltration was performed using Amicon Centriprep-10 Concentrators. The concentration of free triiodobenzoic acid in the filtrate was determined by measurement of absorbance at 260 nm using a Varian DMS 200 spectrophotometer. Data were analyzed and compared for the two species using a direct-linear-plot method [47] and Scatchard model [48]. Experimentally observed data are listed in Table 1.

## RESULTS AND DISCUSSION

### Primary structure

The ESA cDNA sequence (Fig. 2) indicated that the equine preproalbumin molecule is composed of 607 amino acids, including a hydrophobic signal sequence of 18 residues and a propeptide sequence of six basic amino acids. The mature ESA molecule comprises 583 amino acids, the same number of residues found in mature BSA (Holowachuk, E. W., unpublished results) and OSA [9], but one or two amino acids less than found in mature rat serum albumin (RSA) [11] or HSA [6, 7], respectively. ESA, like BSA and OSA, has a deletion of residues 116 and 585 found in HSA (Fig. 2). The ESA cDNA sequence was confirmed by nucleotide sequencing of at least two independent ESA cDNA clones. Based on the cDNA sequence, the protein sequence of mature ESA has a calculated molecular mass of 65709.

Overall, the mature ESA molecule exhibits the highest sequence identity with HSA (76.1%) and PSA (76.0%), followed by OSA (75.5%) and BSA (73.9%) (data not shown).

|                                                                                                        |      |  | Preproalbumin | -><- | Proalbumin | -><- | Mature albumin |  |
|--------------------------------------------------------------------------------------------------------|------|--|---------------|------|------------|------|----------------|--|
| ACAATGAAGTGGGTAACTTTGTTTCCCTTCTCTTTCTCTTCAGCTCTGCTTATTCAGAGGCGTGTTCGCTCGTGATACACACAAGAGTGAGATTGCTCAT   | 102  |  |               |      |            |      |                |  |
| M K W V T F V S L L F L F S S A Y S R G V L R R D T H K S E I A H                                      | 9    |  |               |      |            |      |                |  |
| CGGTTTAAATGATTGGGAGAAAAACATTTCAAGGCCGTGGTGGTGGCTTTTCTCAGTATCTCCAGCAGTGCCCATTTGAGGACCACGTAATAATTAGTG    | 204  |  |               |      |            |      |                |  |
| R F N D L G E K H F K A G L V L V A F S Q Y L Q Q C P F E D H V K L V                                  | 43   |  |               |      |            |      |                |  |
| AATGAAGTAACTGAATTTGCAAAAAAATGTGCTGCTGATGAGTCAGCTGAAAATTGTGACAAGTCACCTCATACTCTTTTGGAGATAAACTGTGTACAGTT  | 306  |  |               |      |            |      |                |  |
| N E V T E F A K C A A D E S A E N C D K S L H T L F G D K L C T V                                      | 77   |  |               |      |            |      |                |  |
| GCAACTCTTCGTCAACCTACGGTGAAGTGGCTGACTGCTGTGAAAAACAAGAACCTGAGAGAAATGAATGCTTCTGACACACAAGATGATCACCCGAAC    | 408  |  |               |      |            |      |                |  |
| A T L R A T Y G E L A D C C E K Q E P E R N E C F L T H K D D H P N                                    | 111  |  |               |      |            |      |                |  |
| CTCCCCAAGCTG---AAACCAGAACCGGACGCTCAGTGCCTGCTTCAGGAAGATCCAGACAAGTTTCTGGGAAAATACCTGTATGAAGTTGCCAGAAGA    | 507  |  |               |      |            |      |                |  |
| L P K L - K P E P D A Q C A A F Q E D P D K F L G K Y L Y E V A R R                                    | 144  |  |               |      |            |      |                |  |
| CATCCTTACTTTTACGGCCAGAACTACTTTTCCACGCTGAGGAATATAAAGCAGATTTTACAGAATGCTGTCAGCTGATGATAAGCTGGCTGTCTGATA    | 609  |  |               |      |            |      |                |  |
| H P Y F Y G P E L L F H A E E Y K A D F T E C C P A D D K L A C L I                                    | 178  |  |               |      |            |      |                |  |
| CCGAAGCTTGATGCTTTGAAGGAAAGAACTACTGCTTTTCATCTGCCAAGAGAGACTCAAGTGCTCCAGTTTCCAAAATTTGGGGAGAGAGCTGTCAAAGCA | 711  |  |               |      |            |      |                |  |
| P K L D A L K E R I L L S S A K E R L K C S S F Q N F G E R A V K A                                    | 212  |  |               |      |            |      |                |  |
| TGGTCAGTAGCTCGCCTGAGCCAGAAATTTCCCAAGGCTGACTTTGAGAGGTTTCCAAGATAGTGACAGATCTTACCAAAGTCCACAAGGAATGCTGCCAC  | 813  |  |               |      |            |      |                |  |
| W S V A R L S Q K F P K A D F A E V S K I V T D L T K V H K E C C H                                    | 246  |  |               |      |            |      |                |  |
| GGTGACCTGCTTGAATGTGCCGATGACAGGGCGGATCTTGCCAAGTACATATGTGAACATCAAGATTCAATCTCTGGCAAACGAAAGCATGCTGTGATAAA  | 915  |  |               |      |            |      |                |  |
| G D L L E C A D D R A D L A K Y I C E H Q D S I S G K L K A C C D K                                    | 280  |  |               |      |            |      |                |  |
| CCTTTGTTGCAAAAATCCCACTGCATTGCCGAGGTGAAAGAAGATGACCTGCCTAGTGACCTGCCTGCATTAGCTGCTGATTTTGCTGAAGATAAGGAAATC | 1017 |  |               |      |            |      |                |  |
| P L L Q K S H C I A E V K E D D L P S D L P A L A A D F A E D K E I                                    | 314  |  |               |      |            |      |                |  |
| TGCAAACTATGAAGATGCCAAGATGTCTTCTGGGCACGTTTGTATGAATACTCAAGAAGGCATCCTGATTACTCCGTCTCCTTCTGCTGTTGAGAATT     | 1119 |  |               |      |            |      |                |  |
| C K H Y K D A K D V F L G T F L Y E Y S R R H P D Y S V S L L L R I                                    | 348  |  |               |      |            |      |                |  |
| GCCAAGACATATGAAGCCACGCTGGAGAAGTGTGTGCGGAGGCCACCTCTCGTGTGTACCGCACAGTGTTTGTATGAGTTTACACCTCTTGTGGAGAG     | 1221 |  |               |      |            |      |                |  |
| A K T Y E A T L E K C C A E A D P P A C Y R T V F D Q F T P L V E E                                    | 382  |  |               |      |            |      |                |  |
| CCTAAGAGTTTAGTCAAGAAAACTGTGACCTTTTGAAGAGGTTGGAGAGTATGACTTTCAAAATGCGCTCATAGTTTCGTTACACCAAGAAAGCACCCAG   | 1323 |  |               |      |            |      |                |  |
| P K S L V K K N C D L F E E V G E Y D F Q N A L I V R Y T K K A P Q                                    | 416  |  |               |      |            |      |                |  |
| GTGTCAACTCCAACCTCTCGTGGAGATTGGAAGGACCTGGGGAAAGTGGGCAGCAGATGCTGTAACTTCTCTGAATCAGAAAGATTGCCCTGTTCTGAAAAC | 1425 |  |               |      |            |      |                |  |
| V S T P T L V E I G R T L G K V G S R C C K L P E S E R L P C S E N                                    | 450  |  |               |      |            |      |                |  |
| CATCTGGCCTTGGCCCTGAACCGGTTGTGTGTGTGATGAGAAGACACCAAGTGTGAGAAGATTACCAAATGCTGCACGGATTCTTGGCGGAAAGACGG     | 1527 |  |               |      |            |      |                |  |
| H L A L A L N R L C V L H E K T P V S E K I T K C C T D S L A E R R                                    | 484  |  |               |      |            |      |                |  |
| CCATGCTTTTCTGCTCTGGAACCTCGATGAAGGATATGTTCCCAAGAAATTTAAGGCTGAAACATTACCTTCCATGCAGATATATGCACACTTCTGGAAGAT | 1629 |  |               |      |            |      |                |  |
| P C F S A L E L D E G Y V P K E F K A E T F T F H A D I C T L P E D                                    | 518  |  |               |      |            |      |                |  |
| GAGAAACAAATCAAGAAACAATCTGCACTTGCTGAACTGGTGAACACAAGGCCAAGCAACAAAGAACTGAAACTGTCTGGGAAATTTCTCAGCC         | 1731 |  |               |      |            |      |                |  |
| E K Q I K K Q S A L A E L V K H K P K A T K E Q L K T V L G N F S A                                    | 552  |  |               |      |            |      |                |  |
| TTTGTAGCCAAGTGCTGGGACGTGAGGACAAAGAGGCTGCTTTGCTGAGGAGGGTCCAAAACCTTGTGCTTCGAGTCAACTTGCCCTTAGCC---TAAAAA  | 1830 |  |               |      |            |      |                |  |
| F V A K C C G R E D K E A C F A E E G P K L V A S S Q L A L A - *                                      | 583  |  |               |      |            |      |                |  |
| CAACGCAATCACGAGCATCTCAGCCTACCTGAGAATAAGAGAAAGAAAGGAAATGAAGACCCAGAGCTTATTCATCTGTTTTTATTTTCTGTTGGTGTA    | 1932 |  |               |      |            |      |                |  |
| AACCAACACTCTGTCTAAAAAACACAAATTTCTTTAAACATTTTGCTCTTTTCTCTGTGCTGCAATTAATAAAAAAATGAAAACAATCTAACATA        | 2030 |  |               |      |            |      |                |  |

Fig. 2. Nucleic-acid and amino-acid sequence of ESA compared to that of HSA (third line).

For comparative purposes and clarity, the sequence numbering of ESA will be normalized to that of HSA in the remainder of the text. The high level of amino-acid identity found between the mammalian serum albumins reflects the importance of the structure of the molecule for its function. The structural features of the binding cavities of serum albumin are important for the binding and delivery of the numerous ligands which interact with the molecule. For example, the importance of hydrophobic ligand-binding sites is seen by comparing the residues which form the HSA ligand-binding cavities in subdomains IIA and IIIA of HSA, previously identified [18]. Amino-acid identity of 12/16 (75%) and 12/17 (70.6%) is found between ESA and HSA when com-

paring the cavity-wall residues of subdomain IIA and IIIA. The amino-acid differences represent conservative changes which would be expected to maintain similar functional binding characteristics in the molecule: Phe201 for Leu, Val210 for Phe, Ser214 for Ala, and Ile233 for Leu in IIA; and Val387 for Ile, Ala414 for Val, Ile425 for Val, and Ala455 for Val in IIIA. Perhaps the most striking comparative observation relates to the highly conserved residues identified to be in close proximity with the bound ligand 2,3,5-triiodobenzoic acid (from 83% to 93% for IIA and IIIA, respectively), determined crystallographically for HSA [18] and ESA. Trp214 and Tyr411 are conserved as in other mammalian albumins. Consistent with the absence of fragmentation of

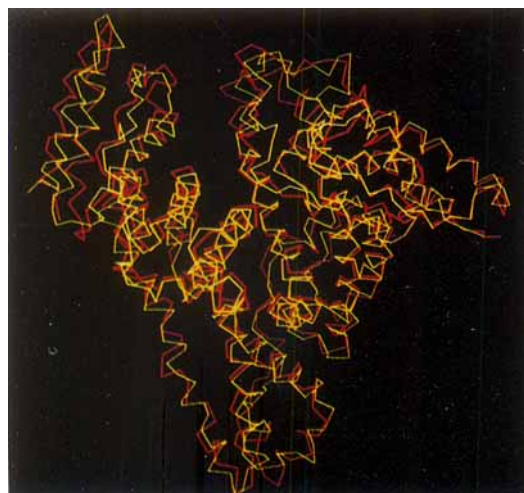


**Fig. 3.**  $\alpha$ -carbon backbones superimposed by the method of least-squares to illustrate structural similarity between HSA and ESA.

ESA with CNBr [49], there are no methionines in the mature ESA sequence (while there are as many six in HSA).

The structural features of the signal peptide and propeptide of serum albumins are important for the secretion of mature albumin [50, 51]. The 18-amino-acid signal peptide or prepeptide and 6-amino-acid propeptide sequences of pre-pro-ESA differ from the sequences previously reported for the serum albumins. The signal peptide or prepeptide of the prepro-ESA exhibits the highest identity with human and porcine signal peptides with only one amino acid difference, followed by bovine and sheep albumin with two residue differences, the chicken and rat with five and six amino-acid differences, respectively, and the non-mammalian serum albumin signal peptides which exhibit less than 44% identity. These amino-acid differences between the mammalian serum albumins represent conservative changes with no compromise of the hydrophobic nature of the signal peptide and no effect on secretory function. The propeptide of equine proalbumin, RGVLR, is unique and differs from the other known mammalian proalbumins with a Leu in place of Phe at position 4 of the propeptide sequence. This conservative change has no effect on the function and cleavage of the propeptide of pro-ESA [52, 53].

### Tertiary structure

The three-dimensional structures of equine and human serum albumin are highly similar (Fig. 3). Both structures have been refined to approximately equivalent resolutions, 0.28 nm for HSA and 0.27 nm for ESA. Least-squares superposition of the  $\alpha$ -carbons gave an average rms (root-mean-square) deviation of 0.201 nm (582 atom pairs) and a mean displacement of  $\alpha$ -carbons of 0.15 nm. Greater agreement exists when individual subdomains are compared. For example, the IIIA subdomains show an average rms of 0.127 nm (110 atom pairs). The greatest deviations between the two structures occur at the extended polypeptide intersubdomain connections and in the larger solvent-exposed loops between helices. Consistent with this observation, a region of high amino-acid substitution (115–125) occurs in the intersubdomain connection of domain I (Fig. 2). One of the two previously mentioned amino-acid deletions in ESA, BSA and

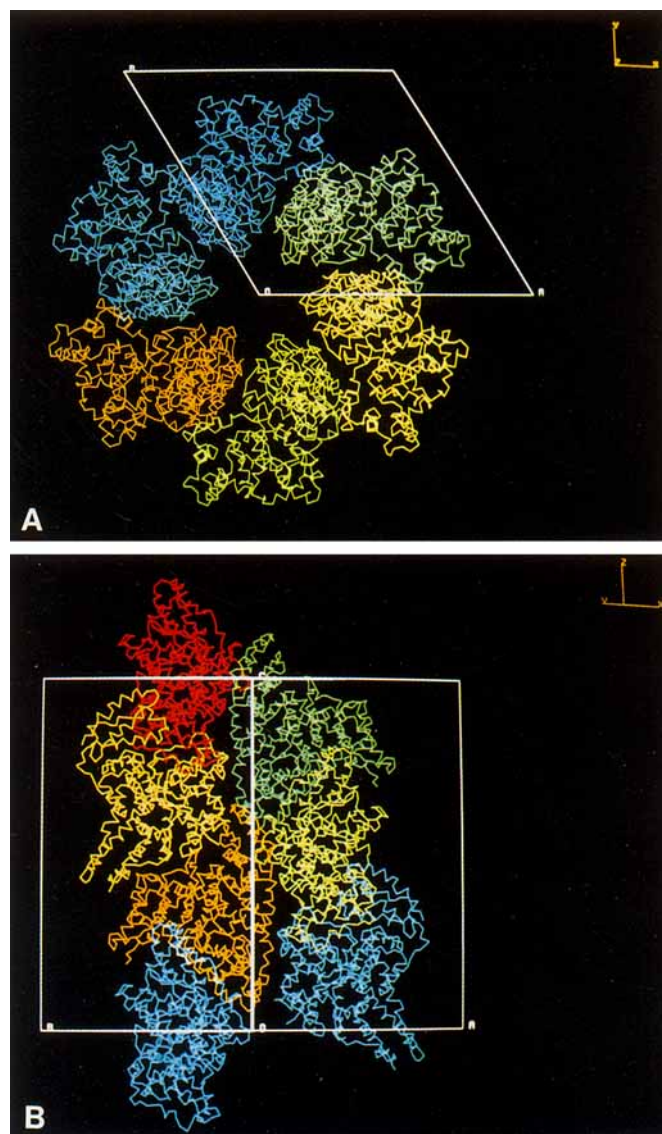


**Fig. 4.** Crystallographic unit cell packaging diagram for ESA. (A) Packing of the ESA molecules in the hexagonal space group  $P6_1$ , as projected down the crystallographic  $c$ -axis. Continuous solvent channels are centered on the six-fold screw axes which run parallel to the  $c$ -axis. Unit cell is outlined in white. (B) View of the unit cell packing perpendicular to the  $c$ -axis.

OSA (residue 116) is found in this same intersubdomain connection while the other represents the deletion of the terminal residue. As with HSA, the disulphide linkages are primarily gauche-gauche-gauche. The ten principal helices h1–h10 within each domain are conserved, however, some of these helices are shorter in length, e.g. h10(II) and h10(III). Consequently, ESA displays less overall helical content than HSA. As in the human albumin, the first two or three residues of the structure are disordered. This inferred mobility should provide the flexibility necessary for the chelation of Cu and Ni, since both the His3 and the N-terminal nitrogen are involved in metal coordination [54].

The solvent content of this crystal form is 43% and almost totally occurs as continuous solvent channels which run parallel to the  $c$  axis centered on the six-fold axes of symmetry (Fig. 4A). These solvent channels have cross-sections of approximately 4.0 nm. The remaining crystal structure is

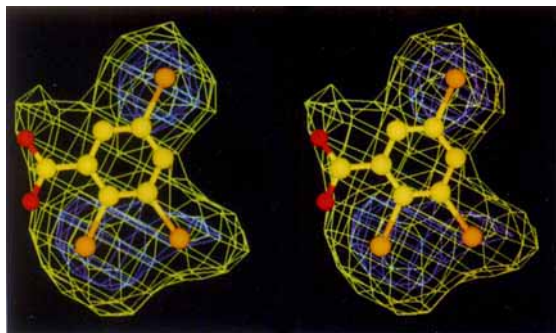


**Fig. 5.** Electron density of bound 2,3,5-triiodobenzoic acid at 0.30-nm resolution produced from  $(F_L - F_N)\alpha$  where  $F_L$  and  $F_N$  refer to measured diffraction data from the triiodobenzoic acid cocrystal complex and native data, respectively, and the phases,  $\alpha$ , are from the native model. Atoms: red, oxygen; yellow, carbon; orange, iodine. Yellow and blue contour levels at 2.0  $\sigma$  and 4.0  $\sigma$ , respectively.

formed by tight inter-molecular packing (Fig. 4B). Given the close interaction of the molecules in the crystal lattice, cocrystallization would thus seem the method of choice for crystallographic studies of ligand binding to ESA. It is interesting to note that despite a different amino-acid sequence, and markedly different crystallization media, both the tertiary and quaternary structures of ESA and HSA are virtually identical. These similarities were apparent beginning with the initial molecular-replacement studies which gave exceedingly clear solutions. The agreement between the structures is understandable since most of the differences in amino-acid sequence constitute changes which conserve charge, hydrophobicity and size. These observations suggest a conformational stability of serum albumin in solution (at pH 5–7.0) which has perhaps not been previously appreciated.

### Ligand-binding studies

Although very little ligand-binding chemistry has been conducted on ESA, it has been shown that the reactivity of ESA with *p*-nitrophenyl acetate is markedly less (essentially absent) than nine other albumins from vertebrate species [24]. This anomalous ESA esterase activity has been confirmed and further documented in a more recent study [55]. Earlier, Brown [51] provided a simple explanation for this difference in activity, namely that Tyr411 is replaced by Val. Brown's studies were based on the partial determination of the amino-acid sequence and are inconsistent with the sequence deduced from the cDNA which indicates Tyr411 is conserved (this work). One possible explanation is sequence polymorphism which is quite common in cattle, horses and sheep. In the structure of HSA, Tyr411 is in the IIIA binding pocket in close proximity to the carboxyl of many small aromatic compounds, consistent with its known chemistry [19]. Presumably it is the strong interaction of Arg410 with Tyr411 in HSA that is responsible for activating the hydroxyl for nucleophilic substitution. Given that Tyr411 is present in ESA, a structural explanation for the lack of esterase activity of ESA is not obvious, perhaps owing to the medium resolution of the two structures.

As previous crystallographic studies of 2,3,5-triiodobenzoic acid binding to HSA revealed the potential utility of this ligand for future research, studies were undertaken to determine its approximate binding constant to ESA and HSA.

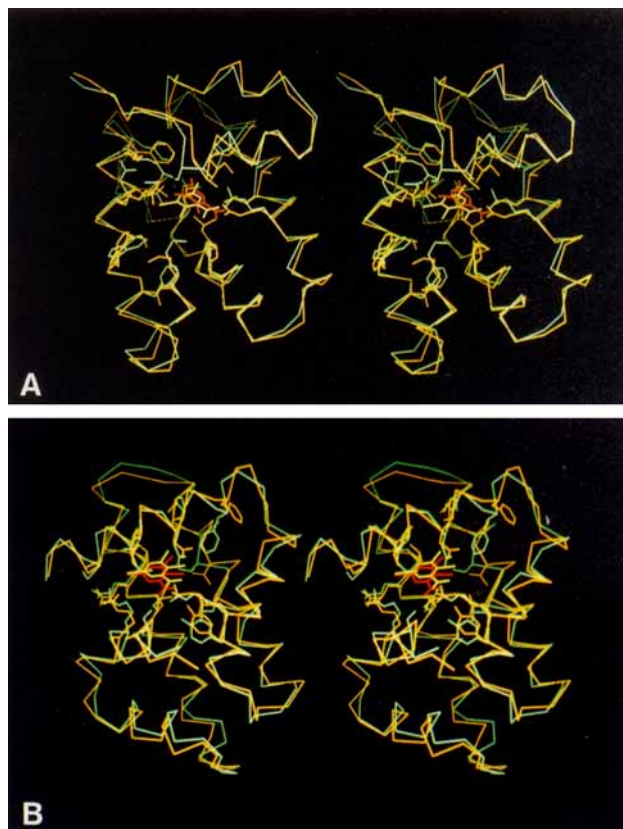


**Fig. 6.** Stereo view showing the similarity between the triiodobenzoic-acid-binding sites within IIA (A) and IIIA (B) where the subdomains have been superimposed by the method of least-squares. The protein components are shown in orange and green for ESA and HSA, respectively, and the positions of bound triiodobenzoic acid are shown in red and yellow for ESA and HSA, respectively. It should be noted that many of the side-chain conformations illustrated in Fig. 6 can only be determined with confidence at higher resolution. In addition, the complex itself has not been refined.

Although there is no exact agreement in the published literature between the various methods used for this purpose, we chose the direct-linear-plot method for convenience and future comparison purposes [47]. Measurements of the  $K_a$  by ultrafiltration experiments using this method produced estimates of  $2.2 \times 10^5 \text{ M}^{-1}$  and  $8.3 \times 10^4 \text{ M}^{-1}$  for HSA and ESA, respectively. These binding constants compare favorably with those obtained by others for similar molecules. In addition, the non-linearity of the Scatchard plot indicates that at least two binding sites for triiodobenzoic acid are present and suggests the possibility of a cooperative effect in both albumins (data not shown). Consequently more detailed studies should be undertaken to achieve a more complete understanding of this phenomenon.

The binding sites of triiodobenzoic acid were determined for ESA by direct observation through crystallographic analysis at 0.30-nm resolution. This is significantly higher resolution than the 0.40-nm resolution previously determined for triiodobenzoic-acid binding to HSA. As in the structure of HSA, there are two strong and distinct binding sites located within subdomains IIA and IIIA. Difference densities were unequivocal and clearly indicated the atomic positions of all three iodine atoms of each molecule of triiodobenzoic acid as the highest peaks ( $>6.5\sigma$ ) in the electron-density maps. Thus, the chemical structure and orientation of triodo-

| IIA    |        |        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|        | Y(149) | K(199) | F(211) | W(214) | S(215) | R(218) | L(219) | R(222) | F(223) |        |
| Equine | Y      | Y      | Y(V)w  | Y      | Y(A)w  | Y      | Y      | N(K)   | Y      |        |
| Human  | Y      | Y      | Yw     | Y      | Yw     | Yf     | Y      | Y      | Y      |        |
|        |        |        |        |        |        |        |        |        |        |        |
|        | L(234) | L(238) | H(242) | R(257) | L(260) | A(261) | I(264) | I(290) | A(291) | E(292) |
| Equine | Y(I)   | Y      | Y      | Y      | Yw     | Yw     | Y      | Y      | Y      | Yw     |
| Human  | Y      | Y      | Y      | Y      | Yw     | Yw     | Y      | Y      | Y      | Yw     |
|        |        |        |        |        |        |        |        |        |        |        |
| IIIA   |        |        |        |        |        |        |        |        |        |        |
|        | P(384) | L(387) | I(388) | N(391) | C(392) | F(395) | R(410) | Y(411) | L(430) |        |
| Equine | Y      | Y      | Y      | Y      | Yw     | Nw     | Y      | Y      | Y      |        |
| Human  | Y      | Y      | Y      | Y      | Yw     | Yw     | Y      | Y      | Y      |        |
|        |        |        |        |        |        |        |        |        |        |        |
|        | V(433) | C(438) | A(449) | E(450) | L(453) | R(485) | S(489) |        |        |        |
| Equine | Y      | Yw     | Y(S)   | Y      | Y      | Y      | Y      |        |        |        |
| Human  | Y      | Yw     | Y      | Y      | Y      | Y      | Y      |        |        |        |

**Fig. 7. Amino acids involved in ligand binding.** Amino acids which are involved in binding in ESA and HSA are marked Y for yes, N for no. Boldface Y indicates the amino acids are conserved in both ESA and HSA. Amino acids which are more remote but still form important sides of the binding region are designated w for wall and f for far, as additional identifiers.

benzoic acid in the binding pockets is unambiguous (Fig. 5). There are always concerns when ligands are diffused into crystals to determine binding loci, since potential binding sites may not be accessible because of restrictions in intermolecular contacts within the crystal structure. These concerns for triiodobenzoic acid binding to serum albumin can be eliminated since the ESA/triiodobenzoic acid complex was produced by co-crystallization with a triiodobenzoic acid to ESA molar ratio of 4: 1 and identical results were obtained with HSA under markedly different experimental conditions. As the triiodobenzoic acid structure has never been included in the model, this provides additional supporting evidence for the correctness of the structure. Overall, ESA and HSA exhibit similar but distinct binding chemistry of triiodobenzoic acid (Fig. 6). However, it is the similarities, not the differences which are striking. The major binding differences occur when comparing the two IIA subdomains. In IIA the carboxyl of ESA-bound triiodobenzoic acid is oriented primarily towards Arg257, but also shares this interaction with His242 (not with Lys199 as in HSA). This is the result of a shift of triiodobenzoic acid in ESA by a few tenths of a nm toward helix h6(II) (Fig. 6A). In HSA this carboxyl group is oriented midway between His242, Lys199 and Arg257 (Fig. 6A). The binding of triiodobenzoic acid in subdomain IIIA is quite similar for both ESA and HSA (Fig. 6B). Altogether, ESA shares 18 out of 19 equivalent residues in close proximity to the triiodobenzoic acid in IIA and 15 out of 16 for IIIA (Fig. 7). When one considers only the residues in immediate contact with the ligand (approximately 10–11 residues in each subdomain) eight or more in each subdomain are strictly conserved in all mammalian albumins (Carter, C. D. and Ho, J. X., unpublished results). Knowledge of these residues should provide extremely important information for guiding future studies of the underlying chemistry of albumin. Furthermore, these findings add convincing support to the notion that albumin's principal role in the circulatory system is in the transport, distribution and metabolism of various exogenous and endogenous ligands.

This research was supported by the Office of Space Science and Applications of the National Aeronautics and Space Administration

(NASA). J. X. H. and P. D. T. were supported under contract by NASA with the Universities Space Research Association.

## REFERENCES

- Brown, J. R. (1976) *Fed. Proc.* 35, 2141–2144.
- Peters, T., Jr., (1985) *Adv. Protein Chem.* 37, 161–245.
- Fehske, K. J., Müller, W. E. & Wollert, U. (1981) *Biochem. Pharmacol.* 30, 687–692.
- Behrens, P. Q., Spiekerman, A. M., Crabtree, G. R. & Long, G. L. (1975) *Fed. Proc.* 34, 591.
- Meloun, B., Moravsek, L. & Aikostka, V. (1975) *FEBS Lett.* 58, 134–137.
- Lawn, R. M., Adelman, J., Bock, S. C., Franke, A. E., Houck, C. M., Najarian, R. C., Seeburg, P. H. & Wion, K. L. (1981) *Nucleic Acids Res.* 9, 6103–6114.
- Dugaiczky, A., Law, S. W. & Dennison, O. E. (1982) *Proc. Natl. Acad. Sci. USA* 79, 71–75.
- Brown, J. R. (1975) *Fed. Proc.* 34, 591.
- Brown, W. M., Dziegielewska, K. M., Foreman, R. C. & Saunders, N. R. (1989) *Nucleic Acids Res.* 17, 10495.
- Weinstock, J. & Baldwin, G. S. (1988) *Nucleic Acids Res.* 16, 9045.
- Sargent, T. D., Yang, M. & Bonner, J. (1981) *Proc. Natl. Acad. Sci. USA* 78, 243–246.
- Minghetti, P. P., Law, S. W. & Dugaiczky, A. (1985) *Mol. Biol. Evol.* 2, 347–358.
- Haefliger, D. N., Moskaitis, J. E., Schoenberg, D. R. & Wahli, W. (1989) *J. Mol. Evol.* 29, 344–354.
- Byrnes, L. & Gannon, F. (1990) *DNA Cell Biol.* 9, 647–655.
- Gray, J. E. & Doolittle, R. F. (1992) *Protein Sci.* 1, 289–302.
- Carter, D. C., He, X. M., Munson, S. H., Twigg, P. D., Broom, M. B. & Miller, T. Y. (1989) *Science* 244, 1195–1198.
- Carter, D. C. & He, X. M. (1990) *Science* 249, 302–303.
- He, X. M. & Carter, D. C. (1992) *Nature* 358, 209–215.
- Gürber, A. (1894) *Sitzungsber. Phys. Med. Ges. Würzburg*, 143.
- McClure, R. J. & Craven, B. M. (1974) *J. Mol. Biol.* 83, 551–555.
- Dixon, J. W. & Sarkar, B. J. (1974) *J. Biol. Chem.* 249, 5872–5877.
- Adams, P. A. & Berman, M. C. (1980) *Biochem. J.* 191, 95–102.
- Leatherbarrow, R. J. & Dean, D. G. (1980) *Biochem. J.* 189, 27–34.

24. Elkarim, A. A. & Means, G. E. (1988) *Comp. Biochem. Physiol. B Comp. Biochem.* 91, 267–272.
25. Sollene, N. P. & Means, G. E. (1979) *Mol. Pharmacol.* 14, 754–757.
26. Koh, S.-N. & Means, G. E. (1979) *Arch. Biochem. Biophys.* 192, 73–79.
27. Chirgwin, J. M., Przbyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* 24, 5294–5299.
28. Aviv, H. & Leder, P. (1972) *Proc. Natl Acad. Sci. USA* 69, 1408–1412.
29. Holowachuk, E. W. (1985) *Immunogenetics* 22, 665–671.
30. Holowachuk, E. W., Greer, M. K. & Martin, D. R. (1987) *Nucleic Acids Res.* 15, 10551–10567.
31. Krawinkel, U. & Zobelein, R. (1986) *Nucleic Acids Res.* 14, 1913.
32. Han, J. H., Stratowa, C. & Rotter, W. J. (1987) *Biochemistry* 26, 1617–1625.
33. Gubler, U. & Hoffman, B. J. (1983) *Gene (Amst.)* 25, 263–269.
34. Short, J. M., Fernandez, J. M. & Huse, W. D. (1988) *Nucleic Acids Res.* 16, 7583–7600.
35. Mullis, K. B. & Faloona, F. A. (1987) *Methods Enzymol.* 155, 335–350.
36. Tibensky, D., Demars, R., Holowachuk, E. W. & Delovitch, T. L. (1989) *J. Immunol.* 143, 348–355.
37. Slatko, B., Heinrich, P., Nixon, B. T. & Voytas, D. (1989) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York.
38. Hattori, M. & Sakai, Y. (1986) *Anal. Biochem.* 152, 232–238.
39. Rossmann, M. G. & Blow, D. M. (1962) *Acta Crystallogr.* 15, 24–31.
40. Crowther, R. H. (1972) in *The molecular replacement method: a collection of papers on the use of non-crystallographic symmetry* (Rossmann, M. G., ed.) pp. 173–178, Gordon and Breach, New York.
41. Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr.* 21, 273–278.
42. Lattman, E. E. & Love, W. E. (1970) *Acta Crystallogr. B Struct. Crystallogr. Cryst. Chem.* 26, 1854–1857.
43. Crowther, R. A. & Blow, D. M. (1967) *Acta Crystallogr.* 23, 544–548.
44. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* 235, 458–460.
45. Hendrickson, W. A. (1985) *Methods Enzymol.* 115, 252–270.
46. Lindup, W. E. (1987) *Prog. Drug Metab.* 10, 141–185.
47. Eissenthal, R. & Cornish-Bowden, A. (1974) *Biochem. J.* 139, 715.
48. Scatchard, G. (1949) *Ann. NY Acad. Sci.* 51, 660.
49. Vötsh, W., Wagner, H. & Andere, F. A. (1980) *Comp. Biochem. Physiol. B Comp. Biochem.* 66, 87–91.
50. Patterson, J. E. & Geller, D. M. (1977) *Biochem. Biophys. Res. Commun.* 74, 1220–1226.
51. Macgillivray, R. T. A., Chung, D. W. & Davie, E. W. (1979) *Eur. J. Biochem.* 98, 477–485.
52. Brennan, S. O., Peach, R. J. & Bathurts, I. C. (1990) *J. Biol. Chem.* 265, 21494–21497.
53. Peters, T., Jr (1987) *Clin. Chem.* 33, 1317–1325.
54. Camerman, N., Camerman, A. & Sarkar, B. (1976) *Can. J. Chem.* 54, 1309–1316.
55. Panjehshahin, M. R., Yates, M. S. & Bowmer, C. S. (1992) *Biochem. Pharmacol.* 44, 873–879.
56. Carson, M. (1987) *J. Mol. Graphics* 5, 103–106.