1. von Kiedrowski, G. Angew. *Chem. int. Ed. Engl.* **25,** 932–935 (1986).
2. von Kiedrowski, G., Wlotzka, B. & Helbing, J. *Angew. Chem. int. Ed. Engl.* **28,** 1235–1237 (1989).
3. von Kiedrowski, G., Wlotzka, B., Helbing, J. Matzen, M. & Jordan, S. *Angew. Chem. int. Ed. Engl.* **30,** 423–426 (1991).
4. Tjivikua, T., Ballester, P. & Rebek Jr, J. *J. Am. chem. Soc.* **112,** 1249–1250 (1990).
5. Rotello, V., Hong, J.-I. & Rebek Jr, J. *J. Am. chem. Soc.* **113,** 9422–9423 (1991).
6. Nowick, J. S., Feng, Q., Tjivikua, T., Ballester, P. & Rebek Jr, J. *J. Am. chem. Soc.* **113,** 8831–8839 (1991).
7. Rebek Jr, J. *Chem. Ind.* **1992(3),** 171–174 (1992).
8. Hong, J.-I., Feng, Q., Rotello, V. & Rebek Jr, J. *Science* **255,** 848–850 (1992).
9. Feng, Q., Park, T. K. & Rebek Jr, J. *Science* (in the press).
10. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. *Proc. Natn. Acad. Sci. USA* **84,** 4398–4402 (1987).
11. Joyce, G. F. in *Cold Spring Harbor Symp. Quantitative Biology*, Vol. LII, 41–51 (Cold Spring Harbor Press, New York, 1987).
12. Mizuno, T. & Weiss, A. H. in *Advances in Carbohydrate Chemistry and Biochemistry* (eds Tipson, R. S. & Horton, D.) 173–227 (Academic, New York, 1974).
13. Bonner, W. A. *Origins Life* **21,** 59–111 (1991).
14. Jacques, J., Collet, A. & Wilen, S. H. *Enantiomers, Racemates, and Resolutions* (Krieger, Malabar, Florida, 1991).
15. Kondepudi, D. K., Kaufman, R. J. & Singh, N. *Science* **250,** 975–976 (1990).
16. Morowitz, J. J., Heinz, B. & Deamer, D. W. *Origins Life* **18,** 281–287 (1988).
17. Luisi, P. L. & Varela, F. J. *Origins Life* **19,** 633–643 (1989).
18. Bachmann, P. A., Walde, P., Luisi, P. L. & Lang, J. *J. Am. chem. Soc.* **112,** 8200–8201 (1990).
19. Bachmann, P. A., Luisi, P. L. & Lang, J. *Nature* **357,** 57–59 (1992).
20. Spiegelman, S. *Q. Rev. Biophys.* **4,** 213–253 (1971).
21. Biebricher, C. K. in *Evolutionary Biology*, Vol. 16 (eds Hechet, M. K., Wallace, B. & Prance, G. T.) (Plenum, New York, 1983).
22. von Kiedrowski, G. in *40 Jahre Fonds der Chemischen Industrie 1950–1990*, 197–218 (Verband der Chemischen Industrie, Frankfurt, 1990).
23. von Kiedrowski, G. *et al. Nachr. Chem. Tech. Lab.* (in the press).
24. Doudna, J. A., Couture, S. & Szostak, J. W. *Science* **251,** 1605–1610 (1991).
25. Terfort, A. & von Kiedrowski, G. *Angew. Chem. int. Ed. Engl.* **31** (in the press).
26. Inoue, T. & Orgel, L. E. *J. Am. chem. Soc.* **103,** 7666–7667 (1981).
27. Inoue, T. & Orgel, L. E. *Science* **219,** 859–862 (1983).
28. Wu, T. & Orgel, L. E. *J. Am. chem. Soc.* **114** (in the press).
29. Acevedo, O. L. & Orgel, L. E. *J. molec. Biol.* **197,** 187–193 (1987).
30. Kang, CH., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. *Nature* **356,** 126–131 (1992).
31. Challa, G. & Tan, Y. Y. *Pure appl. Chem.* **53,** 627–641 (1981).
32. van de Grampel, H. T., Tuin, G., Tan, Y. Y. & Challa, G. *Macromolecules* **25,** 1049–1056 (1992).
33. Ballard, D. G. H. & Bamford, C. H. *Proc. R. Soc.* **A236,** 384 (1956).
34. Bamford, C. H. in *Reactions on Polymers, Proc. NATO Adv. Study Inst.* (ed. Moore, J. A.) 54–60 (Reidel, Dordrecht, 1973).
35. Shapiro, R. *Origins. A Skeptic's Guide to the Creation of Life on Earth* (Summit, New York, 1986).
36. Müller, D. *et al. Helv. chim. Acta* **73,** 1410–1468 (1990).
37. Joyce, G. F. *et al. Nature* **310,** 602–604 (1984).
38. Hill Jr, A. R., Nord, L. D., Orgel, L. E. & Robins, R. K. *J. molec. Evol.* **28,** 170–171 (1989).
39. Visscher, J. & Schwartz, A. W. *J. molec. Evol.* **26,** 291–293 (1987).
40. Harada, K. & Orgel, L. E. *Origins Life* **20,** 151–160 (1990).
41. Harada, K. & Orgel, L. E. *J. molec. Evol.* **32,** 358–359 (1991).
42. Brack, A., Ehler, K. W. & Orgel, L. E. *J. molec. Evol.* **8,** 307–310 (1976).
43. Joyce, G. F. & Orgel, L. E. *J. molec. Biol.* **188,** 433–441 (1986).
44. Joyce, G. F., Inoue, T. & Orgel, L. E. *J. molec. Biol.* **176,** 279–306 (1984).
45. Eigen, M. *Naturwissenschaften* **58,** 465–532 (1971).
46. Kaufmann, S. A. *J. theor. Biol.* **119,** 1 (1986).
47. Wächtershäuser, G. *Microbiol. Rev.* **52,** 452–484 (1988).
48. De Duve, C. *Blueprint for a Cell: The Nature and Origin of Life* (Patterson, Burlington, North Carolina, 1991).
49. von Kiedrowski, G. *et al. Origins Life* **22** (in the press).
50. Cairns-Smith, A. G. *Genetic Takeover and the Mineral Origins of Life* (Cambridge Univ. Press, 1982).
51. Cairns-Smith, A. G. & Davies, C. J. in *Encyclopaedia of Ignorance* (eds Duncan, R. & Weston-Smith, M.) 391–403 (Pergamon, Oxford, 1977).
52. Weber, A. L. *Origins Life* **17,** 107–119 (1987).
53. Orgel, L. E. *J. molec. Biol.* **38,** 381–393 (1968).
54. Weber, A. L. *Origins Life* **19,** 179–186 (1989).
55. Nelsestuen, G. L. *J. molec. Evol.* **15,** 59–72 (1980).
56. Schwartz, A. W. & Orgel, L. E. *Science* **228,** 585–587 (1985).
57. Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. *J. Am. chem. Soc.* **114,** 1895–1897 (1992).
58. Bosscher, F., Ten Brinker, G. & Challa, G. *Macromolecules* **15,** 1442–1444 (1982).

ACKNOWLEDGEMENTS. I thank A. Weber for drawing my attention to literature on template-directed vinyl polymerizations, A. Hill for drawing the figures and S. Bailey for manuscript preparation. This work was supported by NASA and NSCORT NASA. Figure 6 is reprinted with permission from ref. 58.

Atomic structure and chemistry of human serum albumin

Xiao Min He & Daniel C. Carter*

The Space Science Laboratory, ES76 Biophysics, Marshall Space Flight Center, Huntsville, Alabama 35812, USA

The three-dimensional structure of human serum albumin has been determined crystallographically to a resolution of 2.8 Å. It comprises three homologous domains that assemble to form a heart-shaped molecule. Each domain is a product of two subdomains that possess common structural motifs. The principal regions of ligand binding to human serum albumin are located in hydrophobic cavities in subdomains IIA and IIIA, which exhibit similar chemistry. The structure explains numerous physical phenomena and should provide insight into future pharmacokinetic and genetically engineered therapeutic applications of serum albumin.

THE serum albumins belong to a multigene family of proteins that includes α-fetoprotein (AFP) and human group-specific component (Gc) or vitamin D-binding protein. They are relatively large multi-domain proteins which, as the major soluble protein constituents of the circulatory system, have many physiological functions. The albumins contribute significantly to colloid osmotic blood pressure and aid in the transport, distribution and metabolism of many endogenous and exogenous ligands. These ligands represent a spectrum of chemically diverse molecules, including fatty acids, amino acids (notably tryptophan and cysteine), steroids, metals such as calcium, copper and zinc, and numerous pharmaceuticals. They are implicated in the facilitated transfer of many ligands across organ–circulatory interfaces such as in the liver, intestine, kidney and brain[1], and evidence suggests the existence of an albumin cell surface receptor[2]. In addition to blood plasma, serum albumins are also found in tissues and bodily secretions throughout the body; the extravascular protein comprises 60% of the total albumin. Unlike AFP or Gc, albumins are not glycosylated and play no role in immunosuppression. Human serum albumin (HSA), a protein of M_r 65K, consists of 585 amino acids. Its amino-acid sequence contains a total of 17 disulphide bridges, one free thiol (Cys 34), and a single tryptophan (Trp 214). The disulphides are positioned in a repeating series of nine loop–link–loop structures centred around eight sequential Cys–Cys pairs. A total of 61% of the amino-acid sequences are conserved among the known

* To whom correspondence should be addressed.

sequences of bovine[3], rat[4] and human serum albumins[5]. More recently, several additional sequences have been determined including sheep[6], frog[7], salmon[8], mouse[9], pig[10] and sea lamprey[11]. Sequences are also known for α-AFP[12-14], and Gc proteins[15-17] of human, rat and mouse origin. Most of these proteins share high sequence homology and all of them share the characteristic repeating series of disulphide bridges. All members of the albumin multigene family for which sequences have been determined have internal sequence homology (from two- to sevenfold), suggesting the proteins evolved from a common ancestral protein of about 190 amino acids[18]. This internal structural homology has been previously verified by low-resolution crystallographic studies with HSA[19,20]. (For reviews see refs 21-24.)

In this paper, we report the complete atomic model of two crystal forms of human serum albumin as determined crystallographically at 3.1 Å for the wild-type HSA and 2.8 Å for a recombinant form expressed in yeast (rHSA)[25]. Additionally, the characteristic binding locations and chemistry for a selection of representative biological and pharmaceutical ligands are presented.

Structure determination

Crystals of HSA used in this study were obtained from solutions of defatted albumin in polyethylene glycol (PEG) with an M_r of 400 at neutral pH as previously described[19]. These crystals are unusual because: (1) they grow in the rarely observed space group $P42_12$; (2) they have continuous solvent channels centred on the 2-fold axes parallel to the crystallographic c-axis which have a cross-section of ~90 Å; and (3) the crystals have a very high solvent content of 78%. The crystals are weak scatterers of X-rays and although selected crystals may show diffraction to d-spacings less than 3.0 Å, crystallization and X-ray diffraction experiments indicate a useful limit of 3.1 Å. Crystals of recombinant human serum albumin (rHSA) were obtained from solutions of PEG at neutral pH, belong to the space group $P2_1$, and diffract X-rays to d-spacings of <2.4 Å. The structures of HSA and rHSA were determined by isomorphous and molecular replacement methods, respectively, as outlined in Table 1 and Fig. 1.

Molecular configuration

The topology of HSA is created by a repeating series of six helical subdomains[20]. Consequently, the serum albumin multigene family represents, taxonomically, a distinct structural class of proteins. These six subdomains assemble to form a heart-shaped molecule identical in size and shape to the low-resolution electron micrographs of the 39% homologous human and bovine α-AFP[26] (Fig. 2). The shape is markedly asymmetric, and can be approximated to a solid equilateral triangle with sides of ~80 Å and average depth of ~30 Å. Consistent with the internal amino-acid sequence homology, there are three structurally homologous domains that repeat in the molecule (denoted I, II and III). Each domain is formed by two smaller subdomains, A and B, as previously described at low resolution[19,20]. Altogether, roughly 67% of HSA is helical, with the remainder in turns and extended polypeptide. There are 10 principal helices in each domain (h1-h10) (Fig. 3*a*). The topology of a typical domain is further illustrated in Fig. 3*b*. Subdomain A and B share a common motif that includes h1, h2, h3 and h4 for subdomain A, and h7, h8, h9 and h10 for subdomain B. One exception is that the disulphide bridge connecting h1 and h3 does not exist in subdomain IA. The r.m.s. distances between the corresponding α-carbons of the common motifs of subdomains A and B were determined by least-squares fitting to be 2.47 Å, 2.53 Å and 2.60 Å (57 atom pairs) for domains I, II and III, respectively. In addition to this common motif, subdomain A is supplemented by two additional short antiparallel helices, h5 and h6, which are tied together by a pair of disulphide bridges forming a smaller disulphide double loop (loops 2, 5 and 8, described in Fig. 3). Together, these additions form a virtually continuous helical globin-like fold for the A subdomains which is extensively crosslinked by a total of four interhelical disulphide bridges. The B subdomains supplement the helical motif with an N-terminal portion of extended polypeptide, which creates a folding topology closely resembling a simple up-down helical bundle. The intradomain connections between subdomains IA-IB, IIA-IIB and IIIA-IIIB consist of extended polypeptide from residues Lys 106 to Glu 119, Glu 292 to Val 315, and Glu 492 to Ala 511 in domains I, II and III, respectively. In contrast, the interdomain connections represent a helical continuation from the C-terminal portion of IB and IIB helices to the N-terminal helices of IIA and IIIA (helices: h10(I)-h1(II) and h10(II)-h1(III)), respectively. In the case of the IB-IIA connection this results in the largest helix in the structure (h10(I)-h1(II)) which contains 31 residues.

The disulphide pairings in the primary structure of HSA occur as predicted by Brown[6]. In the three-dimensional structure, the 17 cystines form disulphide linkages which occur primarily between α-helices, often distorting the local helical conformation. It is notable that the occurrence of interhelical disulphides has been rarely observed in protein structure based on a cursory survey of the Brookhaven Protein Data Bank. As predicted by Raman spectroscopic studies[27], the conformations of the thioether linkages are predominantly gauche-gauche-gauche, and typical $C\beta_1$-S_1-S_2-$C\beta_2$ torsion angles cluster around ±80 degrees.

Ligand binding

It became evident from early heavy-atom derivative screens and from preliminary binding studies that the principal binding regions on HSA were located in subdomains IIA and IIIA. As a result, the relative binding locations have now been determined crystallographically for several ligands at low resolutions (Table 2). The binding cavity in IIIA is the most active and accommodating on HSA, many ligands were found to bind preferentially there, for example, digitoxin, ibuprofen and tryptophan. Aspirin and iodinated aspirin analogues show nearly equal distributions between binding sites located in IIA and IIIA. Warfarin occupies a single site in IIA. These observations agree with the predicted locations based on competitive inhibition and spectroscopic studies. Residues Trp 214, Lys 199 and Tyr 411 have been implicated in the binding process by several studies[28-32], and each is located strategically in the IIA or IIIA hydrophobic pockets.

For brevity, only the binding chemistry of ligand 2,3,5-triiodobenzoic acid (TIB) is discussed in detail. TIB, as with many small aromatic carboxylic acids, is bound more or less equally in both IIA and IIIA. Additionally, this ligand serves to illustrate the differences and similarities of the protein-ligand interactions between the two subdomains. In subdomain IIA (Fig. 4*a*) TIB is bound in a hydrophobic crevice in the cavity. The aromatic ring forms hydrophobic interactions with residues Leu 219, Phe 223, Leu 234, Leu 238, Leu 260, Ala 261, Ile 264, Ile 290, Ala 291 and the hydrocarbon chain of Glu 292. The carboxylate group interacts with Arg 257, Arg 222 and Lys 199, and is within ~4.0 Å of His 242. The distribution of hydrophobic and hydrophilic residues in the binding crevice is distinctly asymmetric (Fig. 4*a*). In the IIIA binding pocket (Fig. 4*b*) the hydrophobic portion of the aromatic ring of TIB is packed against Pro 384, Leu 387, Ile 388, Phe 395, Leu 407, Leu 430, Val 433, Ala 449, Leu 453 and the hydrocarbon chains of Arg 485 and Glu 450. The carboxylate interacts primarily with Arg 410, and is within 4.0 Å of the oxygen of Tyr 411.

Although the chemistry of ligand binding of IIIA is analogous to that of IIA, there are marked differences in the location of bound ligand. In subdomain IIA, TIB is located in that part of the cavity created by the smaller disulphide double loop (h5(II) and h6(II)); whereas in IIIA, it is located much closer to h1(III). When salicylic acid derivatives or other small aromatic

TABLE 1 Heavy-atom derivatives, phasing methods and statistics

Data set		Atom	X	Y	Z	O_c	R_c	R_f	R_k	N	$f_{h/e}$	D
PIP	1	Pt	0.3878	0.0797	0.1498	1.07	0.59	0.22	0.16	6778	1.31	4.0
		Pt	0.0258	0.3863	0.2275	0.37						
		Pt	0.2298	0.4331	0.3037	0.77						
		Pt	0.2940	0.0673	0.0803	0.79						
	2						0.53	0.17	0.11	3645	1.72	4.0
	3						0.59	0.18	0.11	2641	1.66	5.6
HgI_2	1	Hg	0.1568	0.3595	0.0660	0.38	0.62	0.21	0.16	4565	1.25	4.0
		Hg	0.1857	0.2778	0.0615	1.11						
		Hg	0.0873	0.3761	0.1397	0.78						
		Hg	0.2443	0.2172	0.1718	0.30						
	2						0.65	0.20	0.13	4334	1.22	4.0
	3						0.73	0.18	0.12	1557	1.15	5.0
IS	1	I	0.0797	0.3778	0.1378	0.73	0.64	0.18	0.12	3876	1.47	4.0
		I	0.1811	0.2695	0.0751	0.71						
	2						0.66	0.20	0.15	4043	1.08	4.0
	3						0.69	0.20	0.15	4108	1.17	4.0
	4						0.66	0.19	0.14	6334	1.27	4.0
	5						0.68	0.18	0.14	6715	1.09	4.0
DIS	1	I	0.1820	0.2716	0.0720	0.77	0.67	0.20	0.14	4734	1.29	4.0
		I	0.0819	0.3756	0.1289	0.91						
TAM	1	Hg	0.0556	0.4554	0.2112	0.64	0.69	0.13	0.11	6334	1.27	4.0
		Hg	0.2088	0.4376	0.4413	0.57						
	2						0.74	0.17	0.12	4277	0.52	4.0
	3						0.73	0.17	0.10	6277	0.66	4.0
TIB	1	I	0.1827	0.2727	0.0761	1.32	0.61	0.22	0.16	5431	1.36	4.0
		I	0.0830	0.3753	0.1248	1.37						
$HgCl_2$	1	Hg	0.0578	0.4540	0.2217	1.04	0.49	0.26	0.14	2201	1.89	6.8
		Hg	0.3429	0.1026	0.3494	0.40						
		Hg	0.1547	0.4589	0.2035	0.49						
$RuCl_2$	1	Ru	0.0784	0.3732	0.1334	0.18	0.75	0.17	0.15	2490	0.47	6.0
		Ru	0.1003	0.3125	0.2495	0.36						
		Ru	0.1916	0.2999	0.3024	0.27						
K_2PtCl_6		Pt	0.1039	0.3686	0.1315	0.57	0.74	0.11	0.10	5883	0.58	4.0
		Pt	0.1535	0.2959	0.3892	0.10						
MER	1	Hg	0.2499	0.2159	0.1443	0.67	0.69	0.11	0.12	7778	0.91	4.0
		Hg	0.1665	0.3367	0.3718	0.73						
	2						0.68	0.11	0.12	3200	0.88	4.0

The crystal structure of HSA (space group $P42_12$; unit cell constants: $a=b=186.5(5)$ Å, $c=81.0(5)$ Å) was phased to 4.0 Å resolution using multiple isomorphous replacement (MIR) data collected from 12 heavy-atom derivatives tabulated above. To prepare heavy-atom–HSA complexes, crystals of HSA were placed in a stabilizing solution of 45% PEG 400 (pH 6.8–7.0) which contained about 0.5 to 5 mM of the compound of interest and allowed to stand for 24 to 72 h. X-ray diffraction data were collected on a Siemens multiwire area detector using a rotating anode source operating at 40 kV and from 70 to 90 mA. Data were reduced using the Xengen program package operating on a microVAX III[42]. Diffraction data used for the isomorphous replacement studies, and initial model construction and refinement were collected from three crystals and merged to give $R_{merge}=11.6\%$ based on intensity for 225,950 observations of 24,249 unique reflections (17,575 unique reflections $>1\sigma F$). Native data used in the final refinement were collected from a selected crystal grown in microgravity (STS-42), $R_{merge}=9.25\%$ based on 49,113 observations of 27,096 unique reflections (21,622 unique reflections $>1\sigma F$). These data were used in the current refinement of HSA. Difference maps ($F_{PH,obs}-F_{P,calc}$) were calculated to identify the binding loci using phases obtained from MIR and solvent flattening as previously described[19]. Because of problems with the radiation stability of the heavy-atom complexes, several partial data sets were required to complete the data on a number of derivatives. Each data set was refined separately: first by least-squares against the centric data, then by an iterative series of lack-of-closure refinements[43] (only one set of refined coordinates are shown in the table for each derivative). This resulted in the combination of 21 individual data sets that were selected from >100 low-to-medium resolution data sets collected with a multiwire area detector. These data produced a mean figure of merit = 66.8 for 9,253 paired reflections greater than 5σ excluding anomalous data and 0.71 including anomalous data from eight of the derivatives. In many cases it was not possible to distinguish the individual heavy atoms in some complexes, such as TAM, PIP; consequently, the complexes were treated as individual heavy atoms (usually with larger B-factors) in the refinement. Solvent flattening was used to improve further the MIR phases[44]. From the resulting electron density, the topology of the structure could be unambiguously traced, and the positions of all 17 disulphide bridges were clearly recognizable. At this time, a polyalanine model (including the 17 disulphides) of HSA was constructed from the electron density using computer graphics[45], and refined by simulated annealing (SA)[46] to $R=0.29$ at 3.5 Å. After an extensive series of unbiased omit maps in which 40 residues were deleted at one time, the complete amino-acid sequence, with the exception of residues 1–3, was incorporated and refined to 3.2 Å against the native data (Fig. 1*a*). The current crystallographic *R*-factor is 24.5% for 15,594 reflections from 6.0 to 3.1 Å with $F>2\sigma$. Current r.m.s. deviations of the model from ideality are 0.028 Å for bond lengths and 5.3 degrees for bond angles. Crystals of rHSA belong to the monoclinic space group $P2_1$ with cell constants: $a=58.9$ Å, $b=38.3$ Å, $c=60.7$ Å and $\beta=101.9$ degrees. In contrast to the tetragonal crystal form above, these crystals have solvent contents of ~33%. Diffraction data were collected from a single crystal and merged to give $R_{merge}=8.0\%$ based on intensity for 14,772 unique data of 51,680 observations to 2.27 Å (97% complete to 2.8 Å). The atomic model of tetragonal HSA determined as above, was used to solve the structure of rHSA by the method of molecular replacement[47]. Rotation function [48,49] and translation studies[50] using the program package MERLOT[51] yielded single high-contrast solutions. The preliminary molecular replacement model produced an *R*-factor of 44% and a correlation coefficient (cc) of 58% without refinement for 4,979 reflections with structure factors greater than 3σ in the resolution range of 8.0 Å to 4.0 Å. The model was subjected to rigid-body refinement (in six segments) and then was refined by SA to give an *R*-factor of 24.5% for 11,885 reflections with $F>1\sigma$ in the resolution range of 10.0 Å–3.0 Å. This model was refit in an analogous manner as described for HSA using omit maps (Fig. 1*b*) and subjected to further SA refinement to give $R=23.8\%$ for 10,672 reflections from 6.0 to 2.8 Å with $F>2\sigma$. R.m.s. deviations of the model from ideality were 0.023 Å for bond lengths and 4.9 degrees for bond angles.

O_c, Relative occupancy.

R_c, Centric *r*-factor (Cullis), $\sum||F_{PH,obs}\pm F_{P,obs}|-F_{H,calc}|/\sum|F_{PH,obs}\pm F_{P,obs}|$.

R_f, $\sum|F_{PH}-F_P|/\sum F_P$.

R_k, Kraut *R*-factor $\sum||F_{PH,obs}|-|F_{PH,calc}||/\sum|F_{PH,obs}|$.

N, Number of unique paired reflections $>5\sigma$.

$f_{h/e}$, Phasing power, $[\sum|F_{H,calc}|^2/\sum(|F_{PH,obs}|-|F_{PH,calc}|)^2]^{1/2}$.

MER, Mersalyl acid, 2-(*N*-(3-hydroxymercuri-2-methoxypropyl)carbanoyl)phenoxyacetic acid.

cc, $[\langle F_oF_c\rangle-\langle F_o\rangle\langle F_c\rangle]/\{[\langle F_o^2\rangle-\langle F_o\rangle^2]^{1/2}[\langle F_c^2\rangle-\langle F_c\rangle^2]^{1/2}\}$.

D, Resolution of data in Å.

PIP, Di-μ-iodobis(ethylenediamine)-di-platinum(II) nitrate.

IS, 5-Iodosalicylic acid.

DIS, 3,5-Diiodosalicylic acid.

TAM, Tetrakis(acetoxymercuri)methane.

TIB, 2,3,5-Triiodobenzoic acid.

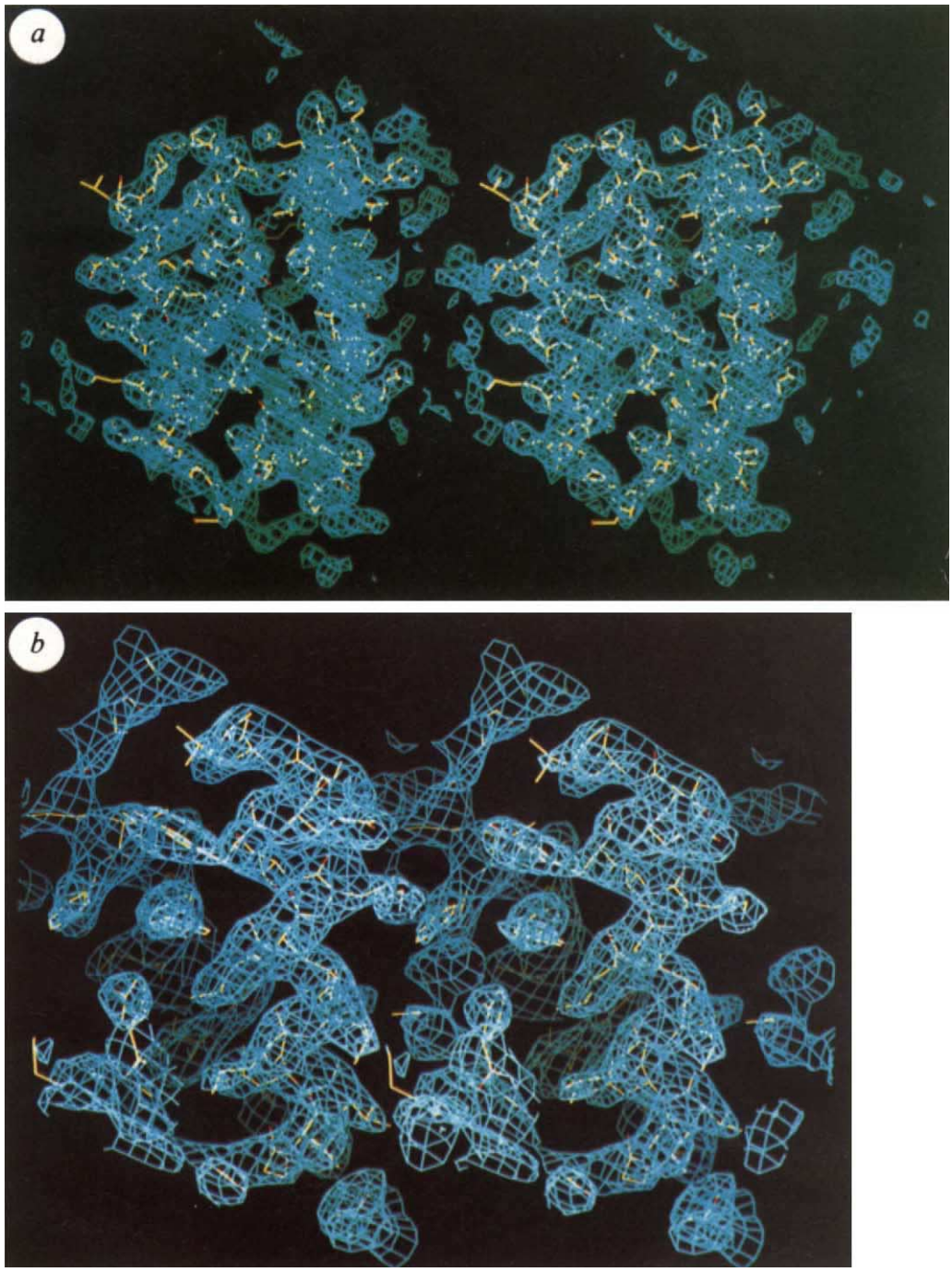


FIG. 1 *a*, Omit map produced from the deletion of a helical motif (residues 196–272) from the model ($\frac{1}{7}$ of the structure). The difference map is shown at 3.2 Å with a 1.8σ contour level (σ is the square root of the map variance). *b*, $2F_o - F_c$ electron density of the helical region from residues 344–355 of rHSA at 2.8 Å.

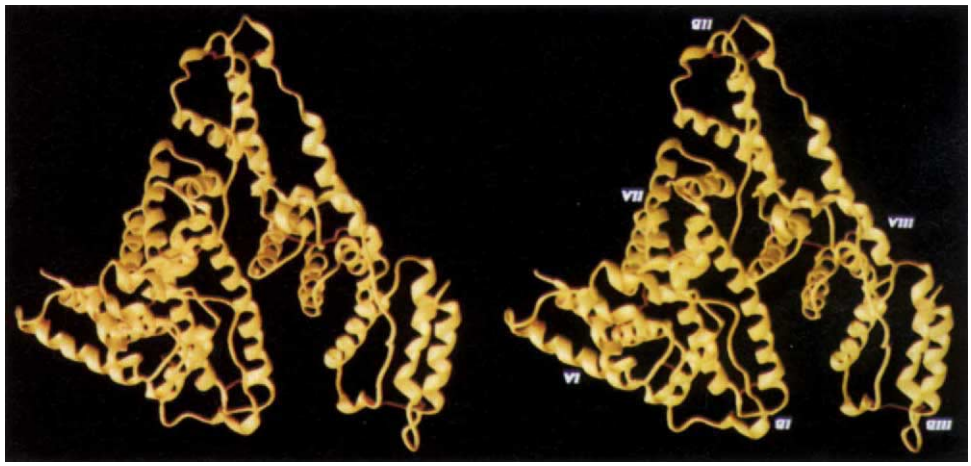


FIG. 2 Stereo view of human serum albumin illustrating the overall topology and secondary structure. The positions of the 17 disulphides and the side chain of Cys 34 are shown in red. Structurally, HSA consists of 28 helices which range in size from 5 to 31 amino acids in length, and can be grouped into 10 principal helices within each domain. This figure was produced with the program Ribbons[52].

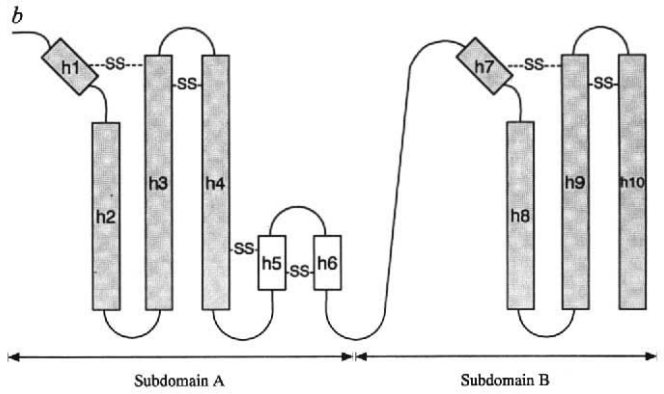
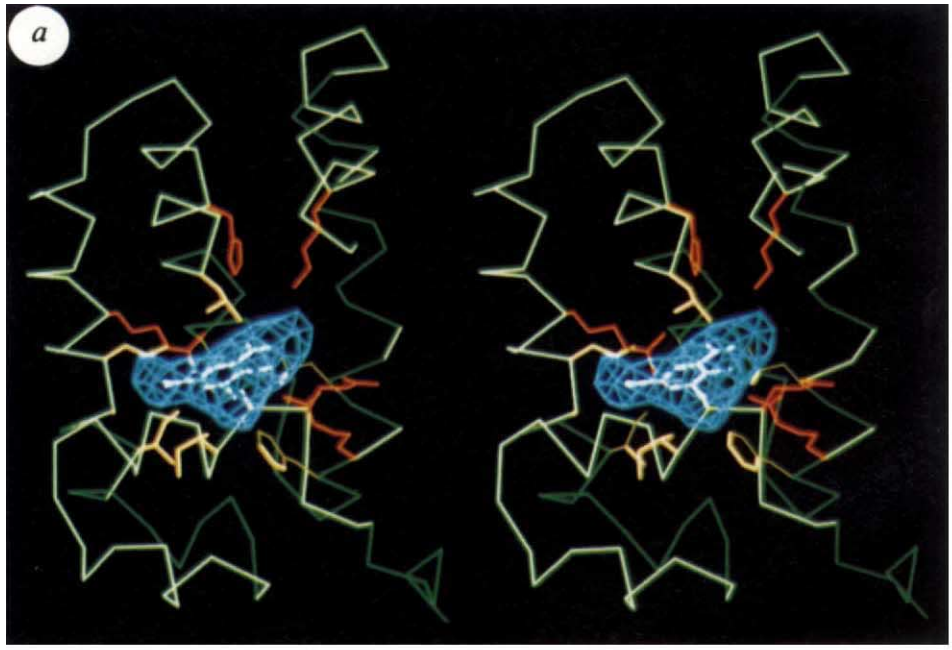


FIG. 3 *a*, Ribbon diagram of domain II. Disulphides are shown in red. The 10 principal helices in each domain are labelled as h1 to h10. *b*, Topological illustration of a typical domain in HSA. The shaded areas represent the common structural motif shared by all six subdomains. This motif is characterized by a short helix (or a short stretch of helix for subdomains IIA and IIIA) followed by three consecutive antiparallel long helices tied together by a pair of disulphide bridges. The principal portion of each subdomain is formed by one of the six larger disulphide double loops in the amino acid sequence, often referred to in the literature as loops 1, 3, 4, 6, 7 and 9 or 1A, 1C, 2A, 2C, 3A and 3C[22]. Three of these loops, 1, 4 and 7 are each supplemented by an additional, but smaller disulphide double loop (2, 5 and 8, respectively). Consequently the combinations of loops 1 and 2, 4 and 5, 7 and 8 represent the three A subdomains (IA, IIA and IIIA) and loops 3, 6 and 9 the B subdomains (IB, IIB and IIIB). The two largest helices in the structure occur at the interdomain connections (IB–IIA and IIB–IIIA) and represent the merged C-terminal and N-terminal helices of both domains. Thus the total number of helices is 28 rather than 30.



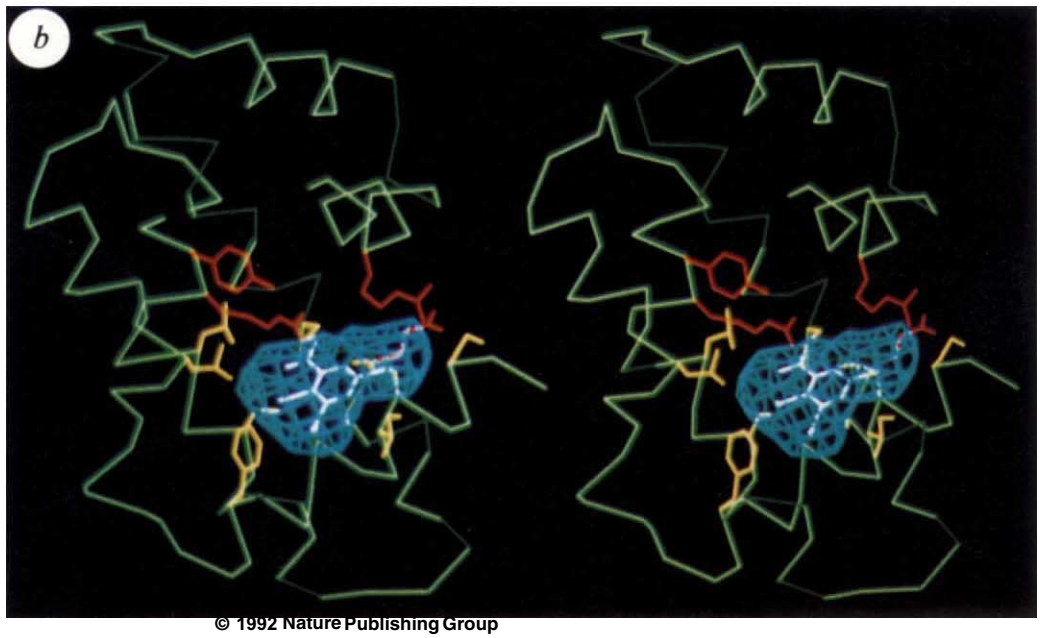


FIG. 4 *a*, Stereo view of subdomain IIA illustrating the chemistry of ligand binding. α-Carbons are shown in green, hydrophobic residues in yellow and hydrophilic residues in red (see text). The difference density $(F_d - F_n)_\alpha$ (where F_d represents the measured structure factors for the TIB complex, F_n is the native data, and the phases, α, are from the model) for TIB is shown contoured at 4σ. The N and C termini are located in the lower and upper left-hand sides of the figure, respectively. Lys 199 and His 242 are located at the bottom of the figure. *b*, Stereo view of domain IIIA illustrating the chemistry of TIB binding. Colour scheme and contour of electron density are the same as in *a*. The N and C termini are located in the lower and upper right-hand side of the figure, respectively. Arg 410 and Tyr 411 are shown in red at the upper left side of the TIB binding site. The binding pockets are accessed through openings of ~10 Å in diameter between helices h1 and h2. Except for three to four residues in the cavities and those surrounding the opening, the pockets are lined with hydrophobic residues. For IIA, these hydrophobic residues are Leu 203, Phe 211, Trp 214, Ala 215, Leu 219, Phe 223, Leu 234, Val 235, Leu 238, Val 241, Leu 260, Ala 261, Ile 264, Ile 271, Leu 275 and Ile 290. The principal hydrophobic residues in IIIA are Pro 384, Leu 387, Ile 388, Phe 395, Leu 407, Val 415, Val 418, Leu 423, Val 426, Leu 430, Val 433, Leu 453, Val 456, Leu 457, Leu 460, Val 473 and Phe 488. These hydrophobic residues form a focal point around Lys 199 and His 242 in IIA and Tyr 411 and Arg 410 in IIIA. These figures were produced using TOM/FRODO[53].

TABLE 2 Ligand binding locations to HSA

Ligand	D	N	R_f	Observed location
Aspirin	4.0	7362	0.11	IIA IIIA
Warfarin	5.0	2555	0.167	IIA
Diazepam	6.8	2075	0.118	IIIA
Digitoxin	5.0	3751	0.137	IIIA
Clofibrate	6.0	2175	0.138	IIIA
Ibuprofen	6.0	2402	0.215	IIIA
AZT	4.0	7548	0.124	IIIA
IS	4.0	6334	0.19	IIA IIIA
DIS	4.0	4734	0.20	IIA IIIA
TIB	4.0	5431	0.12	IIA IIIA

Ligand–HSA complexes and X-ray diffraction data were obtained in a manner as previously described in Table 1. The observed locations refer to the primary binding sites.

D, Resolution or d-spacing in Å.

N, Number of paired unique reflections with $F > 5\sigma$.

R_f, $\Sigma|F_{PH} - F_P|/\Sigma F_P$.

AZT, 3′-Azido-3′-deoxythymidine.

IS, 5-Iodosalicylic acid.

DIS, 3,5-Diiodosalicylic acid.

TIB, 2,3,5-Triiodobenzoic acid.

compounds are compared with TIB, there are differences in the orientation of the aromatic rings and in the details of the interactions of the carboxyl groups depending on the substitutions in the aromatic rings. But in general we found that all of the compounds listed in Table 2 were bound in the same locations within the respective subdomains with similar chemistry.

Discussion

The structures of HSA and rHSA have been determined crystallographically. The higher diffracting monoclinic rHSA crystal is virtually identical with the tetragonal HSA structure. Recently, crystals of the monoclinic form of comparable quality have been produced with HSA, which indicates that the observed structural differences may be attributed to crystal packing forces. These crystals should provide an avenue for future high-resolution refinement of the structure and its complexes. Subsequent discussion, however, will be limited to the crystallographic structure of tetragonal HSA, which explains numerous phenomena regarding the structure and chemistry of serum albumin. For example, the crystal structure explains why ligands bound to domain III affect conformational changes, as well as binding affinities, in domain II, because the binding subdomains share a common interface. Trp 214, conserved in mammalian albumins, plays an important structural role in the formation of the IIA binding site by limiting the solvent accessibility, and it participates in an additional hydrophobic packing interaction between the IIA and IIIA interface. Both Lys 199, which can be acetylated by aspirin[33,54], and Tyr 411, which can be esterified by *p*-nitrophenyl acetate[30–32], are located in the primary binding pockets close to bound ligands. It is now understandable why proteolytic cleavage of HSA would produce two halves of the molecule that can reassociate in solution, thereby restoring the binding properties of the intact albumin[34]. In addition, the amino-acid sequences of serum albumin recognized by the proteases, trypsin and pepsin, lie on the surface of the molecule in flexible solvent accessible inter-subdomain connections. Similarly, a variety of rare, naturally occurring single-site point mutations of HSA identified by anomalous electrophoretic migration are located on the surface of the molecule and exposed to solvent[35–37].

Ligand binding to albumin has been studied for over 50 years. Generally, serum albumin has a greater affinity for small, negatively charged hydrophobic molecules. In this regard it is notable that serum albumin is the principal carrier of fatty acids in the blood which are otherwise insoluble. On the basis of studies with proteolytic fragments of both human and bovine serum albumin[38], the principal binding site for long-chain fatty-acids has been shown to occupy the third domain. In a similar manner, the high-affinity site for bilirubin has been isolated in domain II[39]. Thus, one can infer that the primary long-chain fatty-acid and bilirubin-binding sites reside within IIIA and IIA, respectively. Cys 34, the single free thiol of HSA, is the location of bound cysteine and glutathione[22], and also forms complexes with various mercurial and gold compounds, such as the anti-arthritic auranofin[40]. In the structure, Cys 34 is partially protected from the solvent, located in a turn between helices h2(I) and h3(I). The strong binding of other metals, such as Cu(II) and Ni(II), occurs only in albumins with a histidine at residue 3[41]. Unfortunately, in both crystal structures, residues 1–3 were not observed in the electron density, which suggests that this is a flexible region of the molecule.

The absence of observed ligand binding by homologous subdomain IA can be explained by the structural differences between this subdomain and subdomains IIA and IIIA. In IA there is a region of extended poplypeptide after Cys 62, which allows helix h4(I) to adjust its packing interaction with h3(I) (which is near zero degrees in equivalent helix pairs, such as h3(II) and h4(II)), thereby effectively eliminating the potential binding pocket.

Serum albumin has long been used as a model protein and served numerous applications in both industrial processes and academic research areas. Current advances in recombinant technologies, coupled with a more complete understanding of albumin structure and function, should provide for a greater abundance of future applications. Accordingly, with the wealth of drug and ligand binding data, researchers should now be in a position to understand and predict ligand displacement interactions for a variety of endogenous and exogenous ligands. In this regard, IS, DIS and TIB should prove to be valuable tools in future ligand/drug displacement studies. Moreover, the structure should provide a basis for modifications of the protein to carry therapeutic or diagnostic agents, and create smaller proteins with enhanced binding activities for a variety of applications. □

Received 31 March; accepted 11 June 1992.

1. Pardridge, W. M. *Am. J. Physiol.* **252,** 157–164 (1987).
2. Schnitzer, J. E., Carley, W. W. & Palade, G. E. *Proc. natn. Acad. Sci. U.S.A.* **85,** 6773–6777 (1988).
3. Brown, J. R. & Shockley, P. in *Lipid–Protein Interactions* Vol. 1 (eds Jost, P. & Griffith, O. H.) 25–68 (Wiley, New York, 1982).
4. Sargent, T. D., Yang, M. & Bonner, J. *Proc. natn. Acad. Sci. U.S.A.* **78,** 243–246 (1981).
5. Dugiaczyk, A., Law, S. W. & Dennison, O. E. *Proc. natn. Acad. Sci. U.S.A.* **79,** 71–75 (1982).
6. Brown, W. M., Dziegielewska, K. M., Foreman, R. C. & Saunders, N. R. *Nucleic Acids Res.* **17,** 10495 (1989).
7. Moskaitis, J. E., Sargent, T. D., Smith, L. H. Jr, Pastori, R. L. & Schoenberg, D. R. *Molec. Endocr.* **3,** 464–473 (1989).
8. Byrnes, L. & Gannon, F. *DNA Cell Biol.* **9,** 647–655 (1990).
9. Minghetti, P. P., Law, S. W. & Dugaiczyk, A. *Molec. biol. Evol.* **2,** 347–358 (1985).
10. Weinstock, J. & Baldwin, G. S. *Nucleic Acids Res.* **16,** 9045 (1988).
11. Gray, J. E. & Doolittle, R. F. *Protein Sci.* **1,** 289–302 (1992).
12. Moringa, T., Sakai, M., Wegmann, T. G. & Tamaoki, T. *Proc. natn. Acad. Sci. U.S.A.* **80,** 4604–4608 (1983).
13. Jagdodzinski, L. L., Sargent, T. D., Yang, M., Glackin, C. & Bonner, J. *Proc. natn. Acad. Sci. U.S.A.* **78,** 3521–3525 (1981).
14. Gorin, M. B. & Hoffman, B. J. *Proc. natn. Acad. Sci. U.S.A.* **77,** 1351–1355 (1980).
15. Schoentgen, F. Metz-Boutigue, M.-H., Jolles, J., Constants, J. & Lolles, P. *Biochim. biophys. Acta* **871,** 189–198.
16. Cooke, N. E. & David, E. V. *J. clin. Invest.* **76,** 2420–2424 (1985).
17. Yang, F. *et al. Genomics* **7,** 509–516 (1990).
18. Brown, J. R. *Fed. Proc.* **35,** 2141–2144 (1976).
19. Carter, D. C. *et al. Science* **244,** 1195–1198 (1989).
20. Carter, D. C. & He, X. M. *Science* **249,** 302–303 (1990).
21. Putnam, F. W. *The Plasma Proteins* 2nd edn Vol. 4 (Academic, London, 1984).
22. Peters, T. Jr *Adv. Protein Chem.* **37,** 161–245 (1985).
23. Fehske, K. J., Muller, W. E. & Wollert, U. *Biochem. Pharmac.* **30,** 687–692 (1981).
24. Kragh-Hannsen, U. *Pharmac. Rev.* **33,** 17–53 (1981).
25. Quirk, A. V. *et al. Biotechnol. appl. Biochem.* **11,** 273–287 (1989).
26. Luft, A. J. & Lorscheider, F. L. *Biochemistry* **22,** 5978–5980 (1983).
27. Aoki, K., Sato, K., Nagoaka, S., Kamada, M. & Hiramatsu, K. *Biochim. biophys. Acta* **328,** 323–333 (1973).
28. Sudlow, G., Birkett, D. J. & Wade, D. N. *Molec. Pharmac.* **11,** 824–832 (1975).
29. Sudlow, G., Birkett, D. J. & Wade, D. N. *Molec. Pharmac.* **12,** 1052–1061 (1977).
30. Sollene, N. P. & Means, G. E. *Molec. Pharmac.* **14,** 754–757 (1979).

31. Ozeki, Y., Kurono, Y., Yotsuyanagi, T. & Ikeda, K. *Pharmac. Bull.* **28,** 535-540 (1980).
32. Kurono, Y., Ozeki, Y., Yamada, H., Takeuchi, T. & Ikeda, K. *Chem. Pharmac. Bull.* **35,** 734-739 (1987).
33. Hagag, N., Birnbaum, E. R. & Darnall, D. W. *Biochemistry* **22,** 2420-2427 (1983).
34. Feldoff, R. C. & Ledden, D. J. *Bioch. biophys. Res. Commun.* **114,** 20-27 (1983).
35. Takahashi, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 8001-8005 (1987).
36. Arai, K., Ishioka, N., Huss, K., Madison, J. & Putnam, F. W. *Proc. natn. Acad. Sci. U.S.A.* **86,** 434-438 (1989).
37. Galliano, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **87,** 8721-8725 (1990).
38. King, T. P. *Arch. biochem. Biophys.* **156,** 509-520 (1973).
39. Bos, O. J. M., Remijn, J. P. M., Fischer, J. E., Witling, J. & Janssen, L. H. M. *Biochem. Pharmac.* **37,** 3905-3909 (1988).
40. Pedersen, S. M. *Biochem. Pharmac.* **35,** 2661-2666 (1987).
41. Dixon, J. W. & Sarkar, B. *J. biol. Chem.* **249,** 5872-5877 (1974).
42. Howard, A. J. *et al. J. appl. Crystallogr.* **20,** 383-387 (1987).
43. Furey, W. & Swaminathan, S. *Am. crystallogr. Assoc. Mtg. Abstr. Ser 2* **18,** 73 (1990).
44. Wang, B. C. *Meth. Enzym.* **115,** 90-112 (1985).
45. Jones, T. A. *J. appl. Crystallogr.* **20,** 383 (1987).
46. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235,** 458-460 (1987).
47. Rossmann, M. G. & Blow, D. M. *Acta Crystallogr.* **15,** 24-31 (1962).
48. Crowther, R. H. in *The Molecular Replacement Method* (ed. Rossmann, M. G.) 173-178 (Gordon & Breach, New York, 1972).
49. Lattman, E. E. & Love, W. E. *Acta Crystallogr.* B**26,** 1854-1857 (1970).
50. Crowther, R. A. & Blow, D. M. *Acta Crystallogr.* **23,** 544-548 (1967).
51. Fitzgerald, P. M. D. *J. appl. Crystallogr.* **21,** 273-278 (1988).
52. Carson, M. *J. molec. Graphics* **5,** 103-106 (1987).
53. Cambillau, C. & Horjales, E. *J. molec. Graphics* **5,** 174 (1987).
54. Walker, J. E. *FEBS Lett.* **66,** 173-175 (1976).

ACKNOWLEDGEMENTS. We thank T. Peters Jr for comments on the manuscript, B. Chang, P. Twigg, E. Casale, Z. Krishnasami, K. Keeling, B. Barnes and J. Walraven for assistance with growing the crystals, D. Donovan, G. Roberts and D. Hecht for assistance, E. Abola for the cursory survey of disulphides in the Brookhaven Protein Data Bank, W. Furey for providing the computer program package PHASES, the crew of the First International Microgravity Laboratory (IML-1, STS-42) for their help with the microgravity crystallization experiments and R. S. Snyder for encouragement and support. Recombinant serum albumin was generously provided by Delta Biotechnologies in Nottingham, UK, diffraction data from the tetragonal form of HSA crystals were collected at several facilities and we thank R. Sweet, Y. Satow, Z. Otwinowski, D. Yang and Jan Troup for help and access to the Brookhaven National Synchrotron Light Source, the Photon Factory, Yale University, McMaster University and the Molecular Structure Corporation. This research was supported by a grant from the Office of Space Science and Applications of the National Aeronautics and Space Administration (NASA) to D.C.C. X.-M.H. is supported by NASA under a contract with the Universities Space Research Association.

LETTERS TO NATURE

A double-sided radio jet from the compact Galactic Centre annihilator 1E1740.7−2942

I. F. Mirabel*, L. F. Rodríguez†, B. Cordier*, J. Paul* & F. Lebrun*

* Service d'Astrophysique, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette, France
† Instituto de Astronomía, UNAM, Apartado Postal 70-264, 04510 México, DF, Mexico

RECENT observations[1,2] with the γ-ray telescope SIGMA, on the GRANAT satellite, indicated that the hard X-ray source 1E1740.7−2942 may be the source of the strongest outbursts of 511-keV electron–positron annihilation radiation from the Galactic Centre region[3]. We have observed this source using the Very Large Array, and find that its radio structure is that of a double-sided jet emanating from a compact and variable core. The changes in flux density and spectral index of the core are correlated with variations in the hard X-ray output. The jets are symmetrical about the core, and end in edge-brightened radio lobes; they are probably a result of synchrotron emission of electrons and positrons from the compact core. Our observation suggest that 1E1740.7−2942 is a 'microquasar' stellar remnant near the Galactic Centre, which ejects positrons that travel more than a parsec before slowing and annihilating in the interstellar gas.

After the first observations with SIGMA, we initiated radio monitoring, at several wavelengths, of 1E1740.7−2942, the dominant hard X-ray source (≥30 keV) near the Galactic Centre[4]. The hard X-ray spectrum in its normal state resembles that of Cygnus X-1, one of the best candidates for an accreting black hole of stellar mass. During 13-14 October 1990, SIGMA detected a burst in the 300-600-keV energy range which has been interpreted as annihilation of positrons in a hot medium of temperature ~40 keV; this is consistent with the temperature of the accretion disk derived from the X-ray continuum spectrum[2]. Subsequently it was proposed[5,6] that in addition this high-energy source injects positrons into a molecular cloud where they slow down and annihilate to produce the narrow component of the 511-keV line emission.

Our radio observations were done with the Very Large Array (VLA) at New Mexico, USA, during 10 periods of observation as part of a continuing monitoring programme coordinated with hard X-ray observations from space with GRANAT. The epoch of the VLA observations of 1E1740.7−2942, the array configuration, and the wavelengths observed are listed in Table 1. We have calibrated and analysed the VLA archive data of 1E1740.7−2942 for October 1988 and March 1989. For all observations we have used a bandwidth of 100 MHz. The absolute amplitude calibrator was 1328+307 and the phase calibrator 1748−253, whose 'bootstrapped' flux densities at 20, 6 and 3.6 cm were in the ranges of 1.13-1.17, 0.47-0.50 and 0.27-0.28 Jy, respectively.

After our first period of observation (August-September 1991) we noted a change of a factor of four in the flux density of a compact radio source seen projected inside the 12″ error circle[7] of 1E1740.7−2942. The position of this unresolved source is: $\alpha(1950) = 17\text{ h }40\text{ min }43.01\text{ s}$, $\delta(1950) = -29°43'25.5''$. By analysing source counts[8] in radio surveys, we estimate that the probability is about 0.3% that a radio source with a flux density of 0.4 mJy is an unrelated background radio source in the 12″ error circle of the X-ray source.

In Fig. 1 we compare the time variations of this radio compact source at wavelength 6 cm with the changes observed in 1E1740.7−2942 by SIGMA[9] at hard X-ray wavelengths (1990-92) and by HEXE on the Kvant module of the MIR station[10] (20-21 March 1989). Because no significant day-to-day variations of the radio flux were detected, multiple observations during intervals smaller than a week were averaged and are represented in Fig. 1 by a single point. Variations by factors of

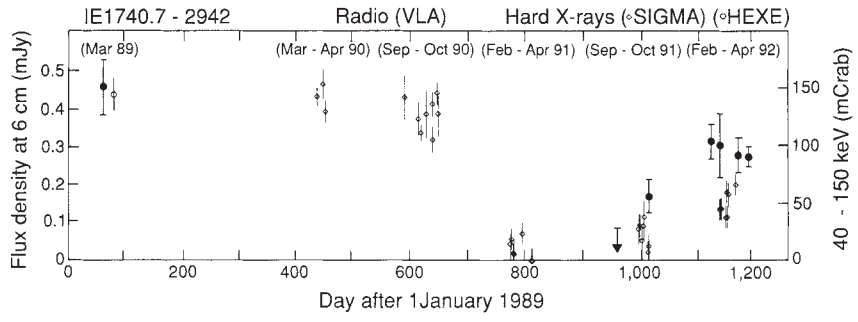


FIG. 1 VLA radio flux densities (●) at 6 cm wavelength of the radio core component, and hard X-rays measured by SIGMA (◇) and by HEXE on MIR (○) between March 1989 and April 1992.