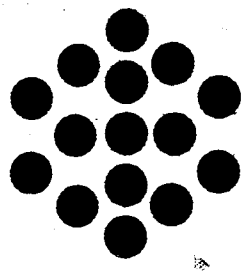

FEBS Federation of European Biochemical Societies

11th Meeting Copenhagen 1977

Volume 50
Colloquium B9

Albumin: Structure, Biosynthesis, Function

Editors
Theodore Peters,
Cooperstown, New York
Ingvar Sjöholm, *Uppsala*

Pergamon Press

FEBS Federation of European
Biochemical Societies
11th Meeting Copenhagen 1977

VOLUME 50
Colloquium B9

ALBUMIN
STRUCTURE, BIOSYNTHESIS, FUNCTION

Editors
THEODORE PETERS, Cooperstown
INGVAR SJÖHOLM, Uppsala

PERGAMON PRESS
OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada Ltd., 75 The East Mall, Toronto, Ontario, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., 19a Boundary Street, Rushcutters Bay, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Pferdstrasse 1, Federal Republic of Germany

First edition 1978

British Library Cataloguing in Publication Data

Federation of European Biochemical Societies.
Meeting, 11th, Copenhagen, 1977
Albumin. — (Publications ; vol.50)
1. Albumin — Congresses
I. Title II. Schambye, Per III. Peters, Theodore IV. Sjöholm, Ingvar
574.1'9245 QP99.3.A4 77-30609

ISBN 0-08-022631-0
ISBN 0-08-021527-0 Set of 9 vols

In order to make this volume available as economically and as rapidly as possible the authors' typescripts have been reproduced in their original forms. This method unfortunately has its typographical limitations but it is hoped that they in no way distract the reader.

Printed in Great Britain by William Clowes & Sons Limited
London, Beccles and Colchester

GENERAL INTRODUCTION TO THE PROCEEDINGS

The 11th FEBS Meeting, Copenhagen 1977, was attended by more than 2500 biochemists and their associates. More than 1300 posters, which attracted many spectators and discussants, and about 220 lectures constituted the back-bone of the Meeting. It proved possible to run specialist-sessions on grand topics in five days' symposia as well as colloquia-sessions treating more limited problems.

We hope that the lectures from all six symposia and three of the colloquia published in the Proceedings volumes will be as supportive to our science as they were to the substance of the Meeting.

We are grateful for all the cooperative efforts, in spite of the fact that the work had to be done against deadlines, and also for the support from the Publisher.

Per Schambye

Secretary-General
Professor of Biochemistry
Odense University

CONTENTS

STRUCTURE AND EVOLUTION OF SERUM ALBUMIN

James R. Brown
Clayton Foundation Biochemical Institute, University of Texas, Austin, Texas 78712, U.S.A.

ABSTRACT

The amino acid sequence and disulfide bridges of serum albumin reveal a pattern of loops and connecting segments between loops that indicate how the molecule evolved from a protein of about 77 residues by four fold tandem gene duplication and one deletion, to give a protein of about 580 residues. From the difference in sequence between duplicated structures and the rate of evolution of albumin, the age of the first duplication is estimated at about one billion years, and the fourth at about 700 million years. The pattern also suggests that albumin consists of three major repeat units or domains and that each domain consists of two subdomains. From the helix content, the location of proline residues, and restrictions imposed by the disulfide bridges, the subdomain is proposed as consisting of a trough like structure formed by three alpha-helical rods, each about 20 residues long. The outer helices, "X" and "Z" are attached to the middle "Y" helix by disulfide bridges at the Cys-Cys sequence at the end of the "Y" helix. The helices are approximately parallel and in contact along the sides where the disulfide bridges are located. Inside the trough is predominantly nonpolar, whereas the outside is mostly polar. The domain is proposed as consisting of a hydrophobic face to face arrangement of two subdomains. The hydrophobic ligand binding site is thus explained as consisting of the hole or slot between the paired subdomains. Similarity in three dimensional structure and amino acid sequence between the "X" and "Y" helices of albumin and the "G" and "H" helices of myoglobin or hemoglobin suggest that the primordial albumin gene may have come from a gene fragment of a primitive globin.

INTRODUCTION

Over 17 years ago Joe Foster (1) proposed that albumin contained four similar subunits on a single polypeptide chain. This was perhaps the first clear proposal of a protein consisting of domains. Foster's model was based on physicochemical observations, especially the N-F transformation. More recently Peters (2) proposed a four domain model of albumin, based especially on large fragments which retained structural and functional properties. Although our (3) preliminary sequence analysis of bovine serum albumin suggested (wrongly) that albumin consisted of a duplicated sequence of about 300 residues, we subsequently found little evidence of repeated sequence or structure, even when about 80% of the BSA sequence had been determined (4). However, when the sequence was effectively complete, a pattern in spacing and bridging of Cys residues revealed that the protein contained nine double loop structures in a linear arrangement along the sequence, suggesting that the protein consisted of a series of globular subunits or domains (5). After solving an error in sequence due to faulty overlapping, it was possible to deduce many details of

the molecular evolution and structural organization of serum albumin (6, 7, 8, 9, 10). Fragments of albumin provide independent evidence of a multidomain structure, and these studies, recently reviewed by Peters (11), beautifully support and extend structural details revealed by the sequence. This report will describe the evolution and structure of albumin, mainly as deduced from the amino acid sequence and disulfide bridges.

AMINO ACID SEQUENCE

The complete amino acid sequence of bovine serum albumin is given in Fig. 1. This sequence differs in several places from our first reported sequence (6) based on new evidence, especially from analysis of peptides of BSA produced by digestion with S. Aureus protease which cleaves at Glu and occasionally at Asp (12, 13). The sequence 164-170 has been corrected for Gln and Glu locations and location of Ala at 169 and it now agrees with the sequence determined by Spencer (14). The sequence 264-267 contains two amides and our preliminary results indicate that the sequence is Asp-Asn-Gln-Asp. The sequence 277-280 has been reinvestigated and our preliminary results locate Cys at 277 rather than at 280. Our error was probably due to lack of extraction of the polar cysteic acid thiazolinone which remained with the peptide after subsequent Edman cycles and was detected as present by subtractive analysis. We had the same problem in sequencing human albumin at this place so that our HSA sequence (7) should be revised here to agree with that of Meloun and coworkers (15). The sequence 388-392 contains three amides and the correct sequence is probably Gln-Asn-Cys-Asn-Glu. An additional space for a missing residue has been added to 404 to conform to the homologous HSA sequence, so that residues 405-582 will have one number higher than in our previous sequence. We had high hopes of finding the missing four residues in a peptide from the S. Aureus protease digest, but we were without success. Finally, the sequence at 468-469 has been corrected. This sequence is in complete agreement with the sequences of some cyanogen bromide peptides, 1-87 and 547-582 determined by King and Spencer (16) and 88-183 by Spencer (14).

EVOLUTION

The sequence and disulfide bridges as shown in Fig. 1 reveal a pattern of loops and connecting segments that repeats three times. The three major repeat units, consisting of two large double loops separated by a small double loop, correspond approximately to residues 1-190, 191-382, and 383-582. One exception to the pattern is the first loop which is missing Cys residues that should close the large loop at 8 and 54. This bridge was probably lost during evolution. The sequence in Fig. 1 has been arranged so that the repeat units, for the most part, are vertically aligned, thus allowing homologous sequences to be compared. Although the repeat units appear very similar due to the almost invariant spacing of Cys residues, the sequences have diverged greatly, leaving only about 18 to 25% identity between repeat units (8). However, there is little doubt that these repeat units are the result of tandem gene duplication. Recently McLachlan and Walker (17) examined the repeat units of human albumin for evidence of homology by statistical analysis and found the similarity highly significant, with a chance of less than 10^{-11} that the similarity could have occured by chance. From the difference in sequence between repeat units and the rate of change of albumin with time, as determined by the difference between human and bovine albumin, Brown (8) has estimated the time of gene duplication of the major repeat units to be about 700 million years ago. The large loops within each repeat unit are obviously similar and

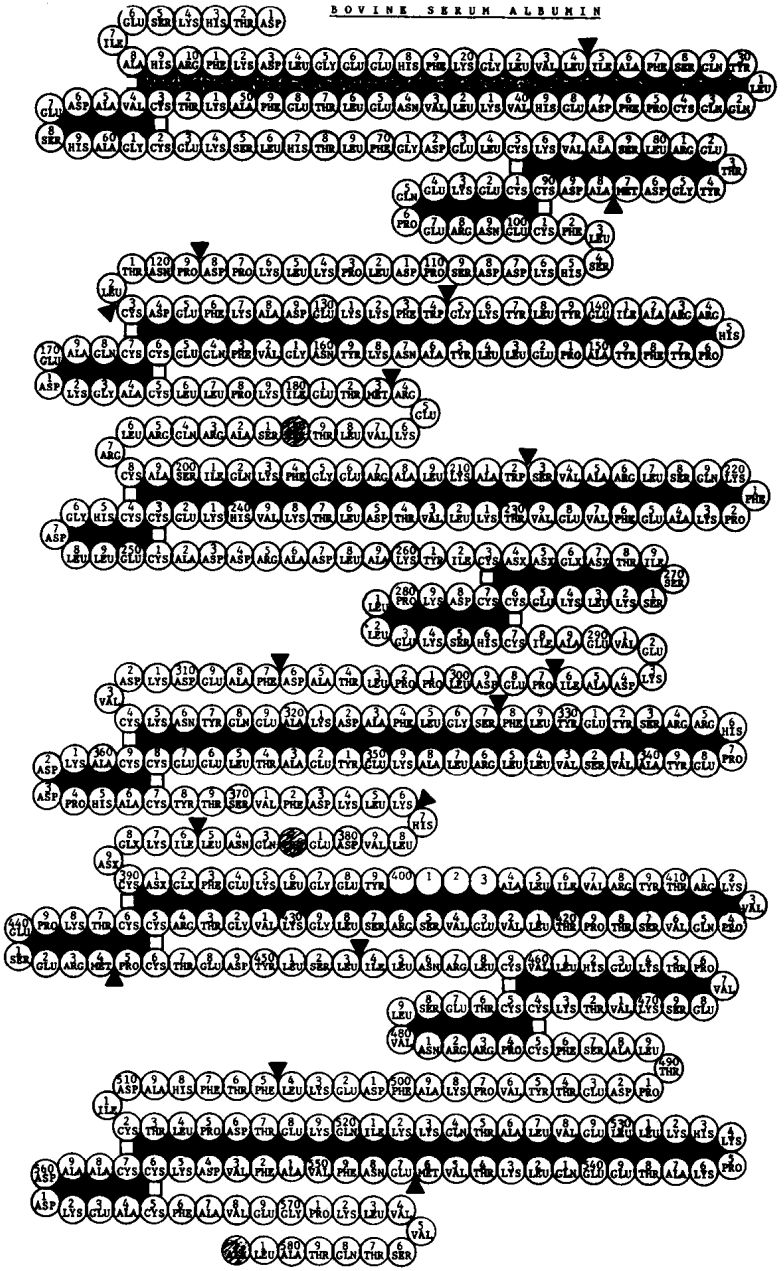

Figure 1. Amino acid sequence and disulfide bridges of bovine serum albumin.

comparison indicates about 20% identity (8). Also, the small loops have the same double Cys bridge arrangement as the large double loops and also some sequence similarity (8). It therefore seems plausible that even farther back in time (one billion years) albumin arose by a series of gene duplications from a primordial protein of about 77 residues, consisting of one large double loop and some sequence on either side (8). This overall scheme is given in Fig. 2 taken from Brown (10). The first step in Fig. 2 indicating ancestry with myoglobin or hemoglobin is based on consideration of three dimensional models and will be discussed later. Duplication followed by half duplication gives a protein with three large loops. A deletion of part of the middle large loop and part of the connecting peptide to the loop above produced the large-small-large double loop pattern of the major repeat unit. Two additional gene duplications tripled the structure to the size of vertebrate albumin. The recent analysis of McLachlan and Walker (17) agrees with this scheme except for the first step where they find insufficient sequence similarity between albumin and myoglobin to indicate homology. .

DOMAINS AND SUBDOMAINS

Because the repeated loop structures discussed above arose by gene duplication, it is reasonable to propose that each duplicated unit has a similar tertiary structure and consists of a somewhat independent globular subunit or domain. My first proposal was that the protein contained nine domains corresponding to the six large and three small double loops (5). However, the major repeat unit consisting of one small and two large double loops suggests the possibility of further organization within that unit. I therefore decided to call the major repeat unit a domain (6) and the structure organized around a large loop a subdomain (8, 10). The small double loop could also be considered a subdomain, however, three dimensional models which I will describe later suggest that after the deletion that produced the small loop, it became integrated with the subdomain above it (10), so that the domain is proposed as containing two subdomains rather than three. The domains and subdomains are defined as architectural features of protein structure and do not necessarily correspond to the same sequence boundaries as the repeat units which resulted from gene duplication. This distinction is important, but it is convenient and common practice to refer to the homology units as domains or subdomains. Fragments of albumin, especially those produced by limited proteolysis, provide ample independent evidence for a multidomain model of albumin structure. Some fragments correspond to complete single domains or subdomains. Such fragments generally retain helicity and functional properties such as ligand binding. Although pertinent to this discussion, space does not permit a discussion of these studies. The reader is referred to the excellent review by Peters (11).

THREE DIMENSIONAL MODELS

The repetitive nature of the albumin structure, the high helix content of 54-68% (18, 19), and restrictions imposed by the double cystine bridges greatly simplify consideration of three dimensional models. Thus, if the structure of a large loop could be deduced, a large part of the total structure would be solved. The first key observation that I made was that an invariant proline residue was strategically located at the tips of each large loop (i.e. at 146, 222, 337, 414, and 530), whereas few proline residues occurred elsewhere in the large loops. This suggested that the large loops consist of two helices starting at the disulfide bridge and continuing for about 20 residues down each helix to the end of the loop where a hairpin turn at the invariant proline at the end of the loop connects the two helices. Since the helices are connected

at one end by a disulfide bridge and at the other by the peptide chain, they should be approximately parallel and in contact along the edges where the disulfide bridge is located. The overall length of the loop including the helices and turn arounds at each end would be close to the short axis of albumin of about forty Angstroms. The next observation is that the segment from the large loop to the following small loop, plus part of the small loop (i.e., 62-82, 251-270, and 446-466) is devoid of proline and the correct length for a third helix, which would start at the disulfide bridge and run parallel and in contact with the helix above it along the edge where the two helices are

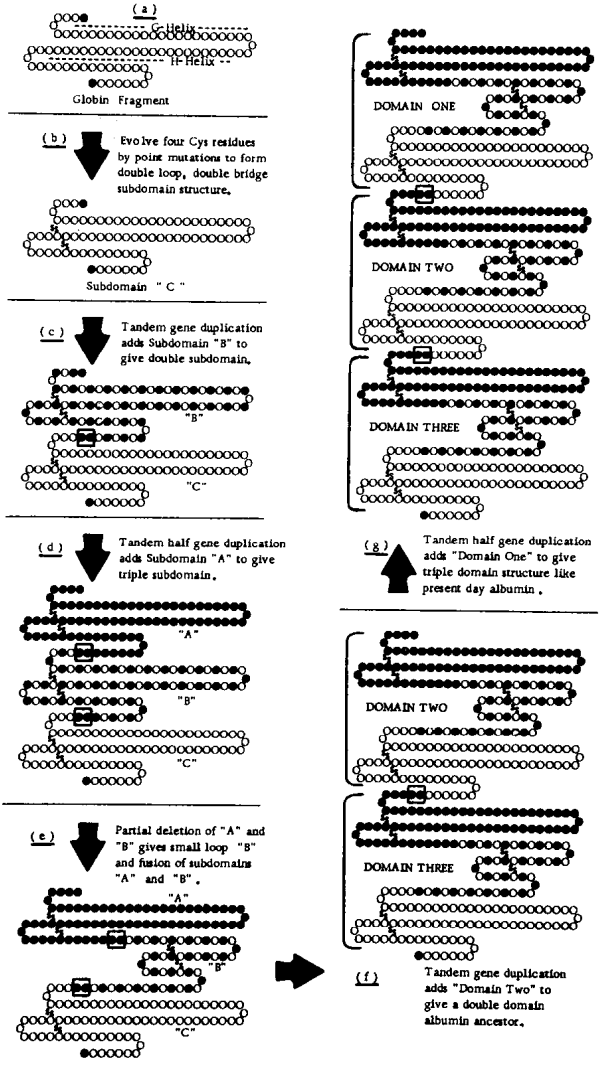

Fig. 2. Evolution of serum albumin.

attached by the disulfide bridge. It is this helix which integrates the small with the large double loop above. Because of the rotation of 100 degrees along the helix axis between the disulfide bridges of the Cys-Cys sequence, the three helices should form a trough-like structure with an inside and outside. This essentially corresponds to my proposed model of an albumin subdomain (8, 10). A similar structure can be made within each domain by adding a third helix from the segment following the lower loop in each domain (i.e., segments 175-194, 367-386, and 565-582). These sequences appear somewhat less favorable because of Pro residues at 178, 382, and 571, but models show a favorable location of polar and nonpolar residues. A model of a subdomain is given in Fig. 3, corresponding to the sequence 313-388. In the view of the outside in Fig. 3 one can see many polar residues except at the contacts between helices where mainly nonpolar residues occur. The view inside the trough shows a preponderance of hydrophobic residues. I should emphasize that this polar-nonpolar distribution played no role in determining the structures of the model. Fig. 4 and Table 1, taken from Brown (10) give a detailed description of the three dimensional structures of the six subdomains of bovine serum albumin. The next step in model building is fairly obvious. If a pair of adjacent subdomains are arranged with their hydrophobic faces together, then closed cylindrical structures will result with a polar outside and a hydrophobic hole or slot between the subdomains, which should correspond to the binding site for fatty acids and other ligands. An important feature of this model of the binding site is that the tips of the large loops, where clusters of basic residues occur (e.g., Arg-Arg-His, 334-336), are located at the entrance to the hydrophobic binding hole. Thus, when a fatty acid inserts its hydrocarbon shaft into the hole, an electrostatic bond can form between its carboxyl group and the plus charged basic groups. Also in support of this model is the special reactivity of basic residues at the ends of loops observed by affinity labeling. Gambhir et. al. (2) find that bromoacetyl tryptophan labels His-145. Andersson et. al. (21) have labeled Lys-220 with fluorescene isothiocyanate. We have labeled the analogous Lys-412 in domain three with trinitrobenzene sulfonic acid.

STRUCTURAL ORGANIZATION

The details of structural organization of bovine serum albumin are given in Fig. 5. Every 10th residue is underlined. The arrangement of the sequence of the middle domain in Fig. 5 is rotated 180 degrees relative to the others or its arrangement in Fig. 1. There are several alternative ways that subdomains can be paired to form a domain. As in Fig. 5, the subdomains could be paired with the tips of the loops at the same end. I refer to this as the parallel arrangement. The antiparallel arrangement would be to rotate one of the subdomains 180 degrees so that the tips of its loops and the disulfide bridges of the other subdomain are at the same end. This arrangement has the attractive feature of pseudo-symmetry, so that a similar binding hole would occur at each end. The third possibility would be to pair adjacent subdomains between domains, rather than within domains as shown in Fig. 5. In this arrangement subdomains "1C" and "2A-B" would be paired as well as "2C" with "3AB". Only the antiparallel arrangement would be possible in this case. Subdomains "1A-B" and "3C" would be unpaired and perhaps nonfunctional. The three "hinge" segments shown in Fig. 5 are interesting sequences because they differ from the rest of the molecule by their high content of proline. They are also highly polar and probably correspond to flexible unorganized connecting segments between subdomains.

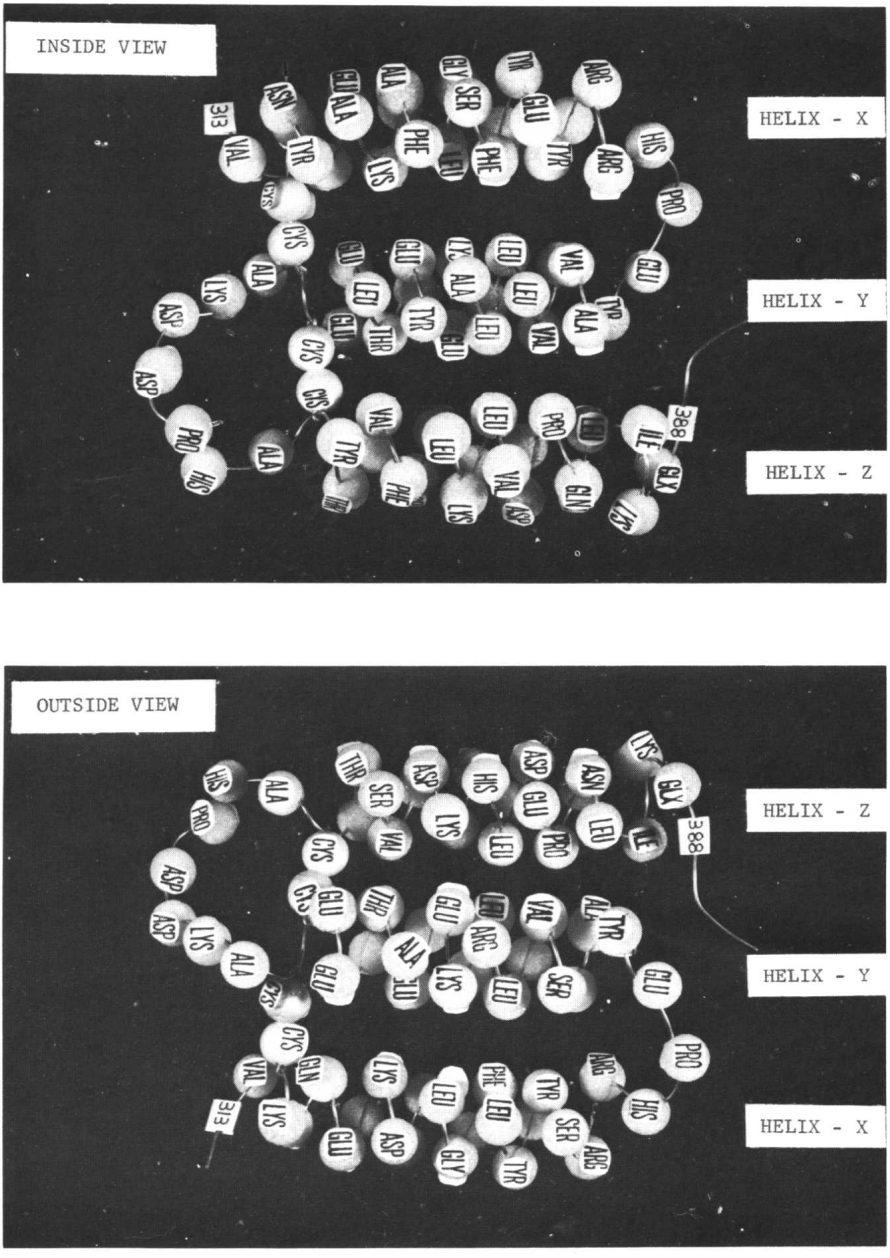

Figure 3. Subdomain models.

GLOBIN-ALBUMIN RELATIONSHIP

While considering the pair of helices in my model of the large loop it occurred to me that the structures were similar to "G" and "H" helices of sperm whale myoglobin. The helices are about the same length as the "X" and "Y" type helices proposed for albumin. They are approximately parallel and are connected by a hairpin turn in the sequence as in albumin. This similarity then became very exciting when I noticed that the sequence of the hairpin turn was Arg-

TABLE 1 Helices of Bovine Serum Albumin

HELIX "X"									HELIX "Y"									HELIX "Z"							
Degrees	1A	1C	2A	2C	3A	3C	HbG	MyG	Residue	1A	1C	2A	2C	3A	3C	HbH	MyH	Degrees	1AB	1C	2AB	2C	3AB	3C
Rotation	8	123	198	314	390	512	91	97	Number	54	167	244	359	436	557	136	142	Rotation	62	175	251	367	446	565
1001	ALA	CYS	CYS	CYS	CYS	CYS	LEU	HIS	1	VAL	CYS	CYS	CYS	CYS	CYS	LEU	ILE	01	CYS	CYS	CYS	CYS	CYS	CYS
200	HIS	ASP	ALA	LYS	ASX	THR	ARG	LYS	2	CYS	CYS	CYS	CYS	CYS	CYS	VAL	ASP	1001	GLU	LEU	ALA	TYR	THR	PHE
300	ARG	GLU	SER	ASN	GLX	LEU	VAL	ILE	3	THR	GLU	GLU	GLU	ARG	LYS	THR	LYS	200	LYS	LEU	ASP	THR	GLU	ALA
401	PHE	PHE	ILE	TYR	PHE	PRO	ASP	PRO	4	LYS	GLN	LYS	GLU	THR	ASP	SER	ARG	300	SER	PRO	ASP	SER	ASP	VAL
140	LYS	LYS	GLN	GLN	GLU	ASP	PRO	ILE	5	ALA	PHE	HIS	LEU	GLY	VAL	VAL	PHE	401	LEU	LYS	ARG	VAL	TYR	GLU
240	ASP	ALA	LYS	GLU	LYS	THR	VAL	LYS	6	PHE	VAL	VAL	THR	VAL	PHE	SER	LEU	140	HIS	ILE	ALA	PHE	LEU	GLY
-201	LEU	ASP	PHE	ALA	LEU	GLU	ASN	TYR	7	GLU	GLY	LYS	ALA	LYS	ALA	ALA	GLU	240	THR	ASP	GLU	ASP	SER	PRO
801	GLY	GLU	GLY	LYS	GLY	LYS	PHE	LEU	8	THR	ASN	THR	GLU	GLY	VAL	LEU	LEU	-201	LEU	THR	LEU	LYS	LEU	LYS
180	GLU	LYS	GLU	ASP	GLU	GLN	LYS	GLU	9	LEU	TYR	LEU	TYR	LEU	PHE	PHE	ALA	801	PHE	MET	ALA	LEU	ILE	LEU
280	GLU	LYS	ARG	ALA	TYR	ILE	LEU	PHE	10	GLU	LYS	ASP	GLU	SER	ASN	LYS	LYS	180	GLY	ARG	LYS	LYS	LEU	VAL
201	HIS	PHE	ALA	PHE	LYS	LYS	LEU	ILE	11	ASN	ASN	THR	LYS	ARG	GLU	ASP	ASN	280	ASP	GLU	TYR	HIS	ASN	VAL
1201	PHE	TRP	LEU	LEU	PHE	LYS	SER	SER	12	VAL	ALA	VAL	ALA	SER	MET	LEU	MET	201	GLU	LYS	ILE	LEU	ARG	SER
220	LYS	GLY	LYS	GLY	GLN	GLN	HIS	GLU	13	LEU	TYR	LEU	LEU	VAL	VAL	SER	ALA	1201	LEU	VAL	CYS	VAL	LEU	THR
320	GLY	LYS	ALA	SER	ASN	THR	CYS	ALA	14	LYS	LEU	LYS	ARG	GLU	THR	ALA	GLY	220	CYS	LEU	ASX	ASP	CYS	GLN
601	LEU	TYR	TRP	PHE	ALA	ALA	LEU	ILE	15	VAL	LEU	THR	LEU	VAL	LYS	HIS	GLN	320	LYS	THR	ASX	GLU	VAL	THR
160	VAL	LEU	SER	LEU	LEU	LEU	LEU	ILE	16	HIS		VAL	LEU	LEU	LEU	VAL	ALA	601	VAL	SER	GLX	PRO	LEU	ALA
260	LEU	TYR	VAL	TYR	ILE	VAL	VAL	HIS	17	GLU	GLU	GLU	VAL	THR	GLN	ALA	ASP	160	ALA	SER	ASX	GLN	HIS	LEU
01	ILE	GLU	ALA	GLU	VAL	GLU	THR	VAL	18	ASP	PRO	VAL	SER	PRO	GLU	PRO	ALA	260	SER	ALA	THR	ASN	GLU	ALA
	ALA		ARG		ARG					PHE	ALA	PHE	VAL	THR	GLU	THR	GLY	01	LEU	ARG	ILE	LEU	LYS	
1001	PHE	ILE	LEU	TYR	TYR	LEU	LEU	LEU	19	PRO	TYR	GLU	ALA	SER	THR	PHE	PHE	1001	ARG	GLN	SER	ILE	THR	
200	SER	ALA	SER	SER	THR	LEU	ALA	HIS	20	CYS	PHE	ALA	THR	VAL	ALA	GLU	ASN	200	GLU	ARG	SER	LYS	PRO	
300	GLN	ARG	GLN	ARG	ARG	LYS	ALA	SER	21	GLN	TYR	LYS	GLU	GLN	LYS	ALA	GLY	300	THR	LEU	LYS	GLX	VAL	
	TYR	ARG	LYS	ARG	LYS	HIS	HIS	ARG		GLN	PRO	PRO	PRO	PRO	PRO	PRO	PRO		TYR	ARG	LEU	ASN	GLU	
	LEU	HIS	PHE	HIS	VAL	LYS	LEU	HIS																
	GLN	PRO	PRO	PRO	PRO	PRO	PRO	PRO																
	32	146	222	337	414	535	114	120		32	146	222	337	414	535	114	120		84	197	273	389	468	582

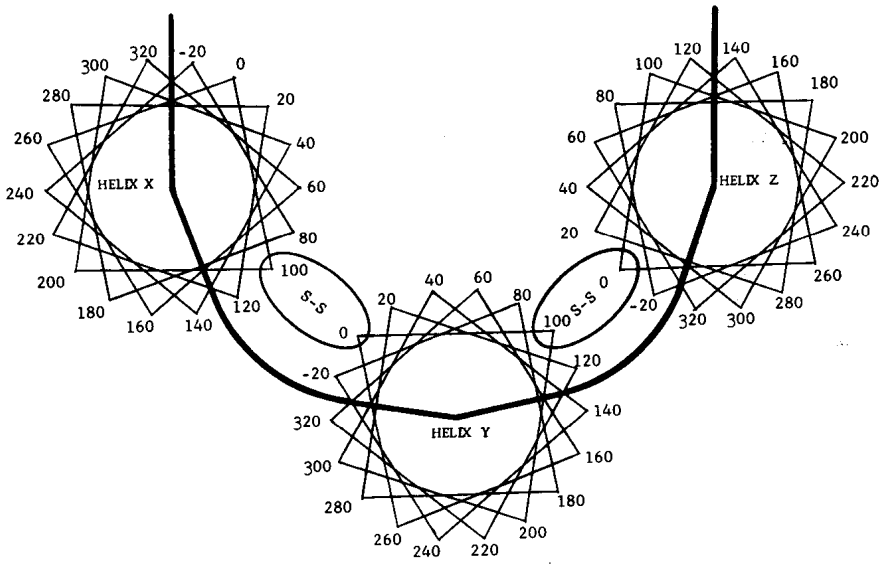

Figure 4. Subdomain structure of serum albumin.

His-Pro, exactly the same as in two analogous hairpin turns in albumin, i.e., 143-146 and 135-137). I then found that if I aligned (at this critical Pro residue) the sequence of albumin loops with the "G-H" regions of sperm whale myoglobin or human alpha hemoglobin, the sequences of albumin were as similar to hemoglobin or myoglobin as the globins were to each other (8). Table 1 shows a detailed comparison of the "G" and "H" globin helices with albumin. Furthermore, when the myoglobin or hemoglobin sequences are constructed in an albumin loop model, the three-dimensional distribution of the residues is

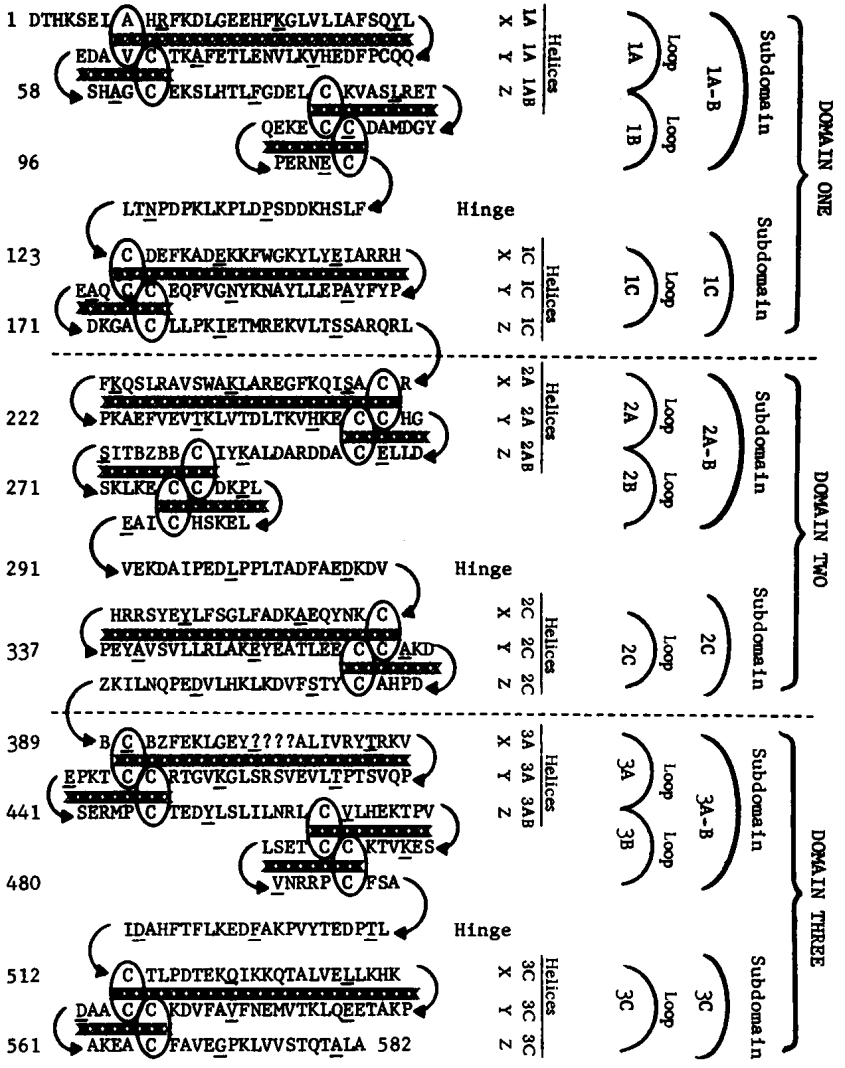

Figure 5. Structural organization of bovine serum albumin.

almost identical with the globin structure (10). Taken together these observations lend plausibility to the model of albumin and suggest the possibility of a common ancestry with a primitive globin as indicated in Fig. 2.

REFERENCES

(1) Foster, J. F. Plasma albumin in The Plasma Proteins, F. W. Putnam, Ed. Academic Press, New York, 1, 179-223 (1960).

(2) Peters, T., Jr. Serum albumin. Advan. Clin. Chem. 13, 37 (1970).

(3) Brown, J. R., T. Low, P. Behrens, and M. Parker, Structural studies on bovine serum albumin, Fed. Proc. 28, 662 (1969).

(4) Brown, J. R., T. Low, P. Behrens, P. Sepulveda, M. Parker, and E. Blakeney, Amino acid sequence of bovine and porcine serum albumin, Fed. Proc. 30, 1241 (1971).

(5) Brown, J. R. Structure of serum albumin: disulfide bridges, Fed. Proc. 33, 1389 (1974).

(6) Brown, J. R., Structure of bovine serum albumin, Fed. Proc. 34, 591 (1975).

(7) Behrens, P. Q., A. M. Spiekerman, and J. R. Brown, Structure of human serum albumin, Fed. Proc. 34, 591 (1975).

(8) Brown, J. R. Structural origins of mammalian albumin, Fed. Proc. 35, 2141 (1976).

(9) Brown, J. R. Evolution of serum albumin, Fed. Proc. 35, 1621 (1976).

(10) Brown, J. R. Serum albumin: amino acid sequence, in Albumin Structure Function and Uses, V. M. Rosenoer, M. Oratz and M. A. Rothschild, Eds., Pergamon Press, Elmsford, N.Y. and Oxford, 27-51, (1977).

(11) Peters, T., Jr. Serum albumin, in The Plasma Proteins, second edition, F. W. Putnam, Ed., Academic Press, New York, 1, 133-181 (1975).

(12) Houmard, J., and G. R. Drapeau, Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds, Proc. Nat. Acad. Sci. 69, 3506 (1972).

(13) Behrens, P. Q. and J. R. Brown, S. Aureus V8 protease specificity on human serum albumin, Fed. Proc. 35, 1474 (1976).

(14) Spencer, E. M., Amino acid sequence from the alanyl peptide from cyanogen bromide cleavage of bovine plasma albumin, Arch. Biochem. Biophys. 165, 80 (1974).

(15) Meloun, B., L. Moravek and V. Kostka, Complete amino acid sequence of human serum albumin, FEBS Letters 58, 134 (1975).

(16) King, T. P. and E. M. Spencer, Amino acid sequence of the amino and carboxyl terminal cyanogen bromide peptides of bovine plasma albumin, Arch. Biochem. Biophys. 153, 627 (1972).

(17) McLachlan, A. D. and J. E. Walker, Evolution of serum albumin, J. Mol. Biol. 112, 543 (1977).

(18) Sogami, M., and J. F. Foster, Isomerization reactions of charcoal deffated bovine plasma albumin. The N-F transition and acid expansion, Biochem. 7, 2172 (1968).

(19) Reed, R. G., R. C. Feldhoff, O. L. Clute, and T. Peters, Fragments of bovine serum albumin produced by limited proteolysis. Conformation and ligand binding, Biochem. 14, 4578 (1975).

(20) Gambhir, K. K., R. A. McMenamy, and F. Watson, Positions in human serum albumin which involve the indol binding site. Sequence of 107-residue fragment, J. Biol. Chem. 250, 6711-6719 (1975).

(21) Andersson, L. O., A. Rehnstrom, and D. L. Eaker, Studies on nonspecific binding. The nature of the binding of fluorescein to bovine serum albumin, Eur. J. Biochem. 20, 371-380 (1971).

Supported in part by National Institutes of Health Grant HD 02899.

SERUM ALBUMIN: CONFORMATION AND ACTIVE SITES

Theodore Peters, Jr. and Roberta G. Reed
The Mary Imogene Bassett Hospital,
Cooperstown, New York, 13326, U.S.A.

INTRODUCTION

Albumin is a versatile, rugged protein with functions so numerous that it is difficult to name one as primary. As the chief extracellular circulating protein, albumin is important in maintaining the circulation; as the solubilizing agent for long chain fatty acids, it is essential to the metabolism of lipids. Albumin is life-saving through its binding of toxins and through its binding of bilirubin in the jaundiced newborn. Its breakdown provides amino acids for nutrition of peripheral tissues.

Clinicians and protein chemists both find uses for albumin. The concentration of albumin in the blood plasma is a measure of circulatory integrity, of liver and kidney diseases and of malnutrition. Over 10,000 kg of purified albumin are administered annually in the United States alone to patients with circulatory failure or with mere albumin depletion. In the laboratory, albumin is employed to protect other macromolecules, usually enzymes or antigens, from denaturation or absorption to the walls of containers. Albumin acts as a reservoir for fatty acids to provide nutrition for cells in culture, and sequesters lipophilic compounds which might harm the oxidative process in isolated mitochondria. It has been used by the protein chemist as a model for almost every chemical reaction or physical measurement attempted with proteins, and by the immunochemist as a model antigen and as a carrier for haptenic molecules. Albumin has an important role as a reference standard in the assay of proteins, since it is one of the few readily available polypeptides which contain no carbohydrate. The U.S. National Bureau of Standards now has available a Standard Reference Material (No. 927) consisting of a 70 g/liter solution of pure bovine albumin, which should prove useful in improving agreement among laboratories engaged in measurements of proteins.

The early methods of isolation of albumin took advantage of its high solubility, the temperature sensitivity of its solubility, its resistance to organic solvents and its low isoelectric point. These principles are still employed in commercial-scale preparations, which often precipitate a "Cohn Fraction V" in 40% ethanol at -5° at pH 5: a refinement has been the use of polyethylene glycol as precipitant. Fraction V contains 3 to 4% of globulins. In laboratory preparations, initial purification with trichloroacetic acid:alcohol followed by ion exchange chromatography and molecular exclusion chromatography provides a purer monomeric albumin than can be achieved on a commercial scale.

Heterogeneity of albumin preparations arises from many causes (1,2). Impurities such as α-globulins, insulin and enzymes (proteases, amylase, placental alkaline phosphatase) occur in crude preparations. To avoid these impurities the time

seems at hand when use of Fraction V for experimental studies should be avoided in favor of crystallized or chromatographically purified albumin. Genetic variants, observed as bisalbuminemia, occur with a frequency of about 1:1000 in Western populations and so would be only a rare cause of heterogeneity. Detection of an additional aminoterminal arginine residue in 5-10% of commercial bovine albumin by Leibowitz and Soffer (3) suggests incomplete cleavage of the hexapeptide from proalbumin or, as those authors suggest, transfer of arginine from charged tRNA-Arg to albumin. The free sulfhydryl group of Cys_{34} provides another source of heterogeneity, occurring as a mixed disulfide with cysteine or glutathione or in higher oxidation states such as sulfonic acid (4).

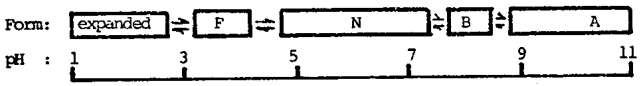

Fig. 1. Isomeric forms of albumin observed at different pH values.

The most-studied cause of heterogeneity is isomerization (Fig. 1). The isomerization ranges from simple, reversible expansion to irreversible disulfide rearrangement, and has been investigated particularly by Aoki and others in the laboratory of the late Joseph F. Foster. The normal, or N, form (Fig. 1) predominates at pH 5-7. Below pH 4 appears the F, for faster-migrating, form which is somewhat expanded; below pH 3 the molecule expands further to expose most tyrosines and other hydrophobic residues to the solvent. Near pH 8, particularly in the presence of calcium ions, a different expansion, with increased accessibility of hydrogen atoms for exchange, increased mobility of the thiol group and a slight loss of helix, produces the B form (5).

The above changes are readily reversible. Above pH 8 a slow transformation to the A form occurs (6,7). The A form is detectable through its slightly slower migration than the N form on electrophoresis (pI of 5.45 compared to 5.24). Disulfide bond rearrangement is involved since the A transformation is accelerated by small amounts of thiol compounds and suppressed by alkylating agents such as iodoacetamide. The change to the A form becomes irreversible near pH 10. Wallevik (8) has shown that the A form disappears faster than the normal or N form in vivo and suggests that it may be an intermediate in albumin degradation.

STRUCTURE AND CONFORMATION

Elucidation of the entire amino acid sequence of bovine and human albumins by Dr. Brown and his colleagues (9,10) was a monumental piece of work, as was the independent determination of the sequence of human albumin by Meloun's group (11). Both teams noted the regular occurrence of adjacent Cys-Cys sequences which are the basis of the repeating loop structure. Brown proposed a linear 9-loop arrangement which Meloun's group confirmed through studies of cyanogen bromide-cleaved peptides (12) and other laboratories, including our own, confirmed by identification of proteolytic fragments obtained without cleavage of disulfide bridges. The latter included loops 3, 6, 8 and 9 of bovine albumin and loops 3 and 6 of human albumin, as well as domain II of human albumin and domain III of bovine albumin (13,14). (In this article we number the loops sequentially from 1 to 9 rather than by domains; thus, loops 1-3, 4-6 and 7-9 are domains I, II and III, respectively.)

Agreement between Meloun's and Brown's sequences for human albumin was within 97% of the residues excluding amide assignments or 93% overall, a tribute to the thoroughness of both groups. From the sequence of Meloun et al. (11) one