THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 248, No. 9, Issue of May 10, pp. 3313–3326, 1973
Printed in U.S.A.

Carp Muscle Calcium-binding Protein

II. STRUCTURE DETERMINATION AND GENERAL DESCRIPTION*

(Received for publication, October 30, 1972)

ROBERT H. KRETSINGER‡ AND CLIVE E. NOCKOLDS§
From the Department of Biology, University of Virginia, Charlottesville, Virginia 22901

SUMMARY

The structure of crystalline carp muscle calcium-binding protein (parvalbumin) has been determined by x-ray diffraction techniques to nominal 1.85-A resolution. Isomorphous and anomalous scattering data were measured for three heavy atom derivatives, 3-chloromercuri-2-methoxypropyl urea, mercury bromide, and ethyl mercury chloride, to 2.0-A resolution using precession photography. As described in Paper III in this series the 2.0-A phases were refined and the 2.0- to 1.85-A phases were determined by use of the tangent formula.

The electron density map is interpreted in terms of the 108 amino acid sequence described in Paper I in this series. A calcium ion is bound in the loop between helix C and helix D and a second calcium is bound in the EF loop. The entire CD region is related to helix E, the EF loop, and the terminal helix F by an approximate intramolecular 2-fold axis. Although it does not bind calcium the AB region has a structure similar to the CD and EF regions and appears to have resulted from a gene triplication.

The molecule is generally spherical with a well defined hydrophobic core, one-seventh of its total volume, composed of side chains of phenylalanine, isoleucine, leucine, and valine. All of the polar side chains are at the surface except those associated with calcium binding and with an invariant internal salt bridge between arginine-75 and glutamic acid-81.

In the first paper of this series Coffee and Bradshaw (1) describe the determination of the amino acid sequence of carp muscle calcium-binding protein component B. They also discuss the general characteristics of the molecule. In this second paper we describe the solution of the three-dimensional structure by means of x-ray diffraction. The structure is discussed in terms of its assumed role in muscle contraction and in terms of its apparent evolution via gene triplication (2). In the final paper by Hendrickson and Karle (3) the tangent formula refinement and extension of phases is described.

EXPERIMENTAL PROCEDURES

Preparation of Muscle Calcium-binding Protein—All of the protein used in these studies came from one preparation made in the summer of 1968 (4). The steps through Sephadex G-75 chromatography were based on the procedure of Pechère and Focant (5). Fillets, free of skin, from 500-g carp (*Cyprinus carpio*) were homogenized. The homogenate was centrifuged 30 min in a Sorvall GS-3 rotor at 9,000 rpm (9,300 to 14,700 × *g*). The lipid film on the surface was aspirated off. The supernatant was dialyzed in preboiled Visking tubing (40/32 inches) for 3 days against five changes of water. The dialysate was centrifuged and the substantial pink precipitate discarded. Although muscle calcium-binding protein is found only in "white" muscle (6) we did not dissect away the superficial median strip of "red" muscle rich in myoglobin. Solid ammonium sulfate, 530 g per liter original volume, was slowly added with stirring over 24 hours to the solution of albumins. The red precipitate was discarded and the supernatant was brought to saturation with ammonium sulfate and stirred for 24 hours. The pH was maintained between 6.0 and 8.0 with phosphate. All of these procedures were performed in a 4° cold room. The 75 to 100% saturated cut was dissolved and dialyzed against Tris-HCl buffer, pH 7.3, 5 mM. The low (1.4 S) and high (7.05 S) molecular weight albumins were well resolved on Sephadex G-75 chromatography.

Hamoir *et al.* (6, 7) identified by moving boundary electrophoresis three main low molecular weight components in carp myogen. We separated these components by DE-32 diethylaminoethyl cellulose chromatography at pH 7.0, in 2 mM phosphate, with a 0.0 to 0.5 M KCl linear gradient. The three components can be resolved and identified by electrophoresis on acrylamide gels (4). Recently Pechère *et al.* (8) have described a DE-11 preparation using a piperazine buffer and a NaCl elution gradient. We find this method to give higher resolution and to be more reproducible. They found that a single carp contained five components having the following isoelectric points: *5a*, 4.47; *5b*, 4.37; *3*, 4.25; *2*, 3.95; and *1*, <3.9 (8). Gerday and Bhushana Rao (9) also isolated the first four and have called them IVa, IVb, III, and II. To be consistent with our original

* This research has been supported by National Science Foundation (GB-29306) and by National Institutes of Health (1 R01 AM-16253) grants. We thank Dorothea Dangelat and Marianne Dannbeck for technical assistance and Robert Tufty for computer programming.

‡ United States Public Health Service Research Career Development Awardee, AM-16253.

§ Present address, Electron Microscopy Unit, University of Sydney, Sydney, Australia 2006.

(4) notation we call *5a* and *5b* (which are not resolvable on acrylamide gel electrophoresis) A, *3* is B, and *2* is C.

Jebsen and Hamoir (10) found that, per 100 g of fresh muscle, the total protein content is about 16 g and that the content of muscle albumin (myogen) is 2.0 g for plaice and 2.7 g for carp. Focant and Pechère (11) explored the phylogenetic distribution of the low molecular weight components, as defined by Sephadex G-75 chromatography. Reptiles, birds, and mammals have none, while the content of the low molecular weight component, which consists solely of muscle calcium-binding protein or "parvalbumins" (12), ranges from 15% in amphibia to 28% in teleost fish. Our yield for the sum of the three fractions was somewhat less than the predicted 700 mg/100 g of fresh muscle. This is hardly surprising, since we were exploring preparative techniques.

Crystallization and Heavy Atom Derivatives—We initially chose to work with Component B because it gave the best crystals (4, 7). We were never able to obtain large crystals of Components A or C. Both lyophilized protein and frozen solutions were used. Subsequently we have failed to grow usable crystals from other fish, pike, hake and *Tilapia*, and from frog. We took no special precautions either to add or to remove calcium. Only after the data collection was completed did we realize that the protein binds calcium (8). Flame absorption spectrophotometry[1] performed on some of the crystals actually used for data collection gave a calcium to protein molar ratio of 0.75 (13). However, due to the very small amount of material actually available, this result could easily be in error by a factor of two. As will be discussed by Hendrickson and Karle (3), we feel that the crystals used for data collection had high calcium occupancy at both sites.

Although a wide variety of crystallization conditions were explored, we succeeded in growing crystals only from solutions of 70 to 85% saturated ammonium sulfate at pH 5 to 8. Hoping that the various heavy atom compounds would be more soluble in phosphate, we transferred the crystals through several stages to 4.0 M phosphate, pH 6.8. We never grew usable crystals directly from phosphate. As it turned out, the derivatives we finally chose to use would probably have bound as well in ammonium sulfate. The transfer to phosphate caused a slight change in unit cell dimensions, space group C2,

Sulfate:		Phosphate:	
a	= 28.7 A	a	= 28.2 A
b	= 61.1 A	b	= 61.0 A
c	= 54.5 A	c	= 54.3 A
β	= 94° 37′	β	= 95° 10′

as cited in Reference 4. Subsequently we found that the angle, β, varied from 94° 50′ to 95° 10′ in the phosphate-soaked crystals. Standardizing the transfer conditions reduced this variability. In our calculations we have used β = 95° 0′. None of the heavy atom derivatives used changed the unit cell dimensions by over 0.4%. Crystallization conditions were tested and crystals for use were grown by the microdiffusion technique (14). Possible heavy atom derivatives were screened by soaking four to six crystals in 0.5 ml of 4.0 M phosphate containing the heavy atom compound in concentrations of 0.5 to 5.0 mM. Crystals were soaked 10 to 30 days at room temperature. They were briefly washed in phosphate (free of heavy atoms) before being mounted in quartz capillaries and photographed.

In order to assure constant heavy atom occupancy for all of the crystals used in recording three-dimensional data for each

[1] Performed by J. Turnipseed, Chemistry Department, University of Virginia.

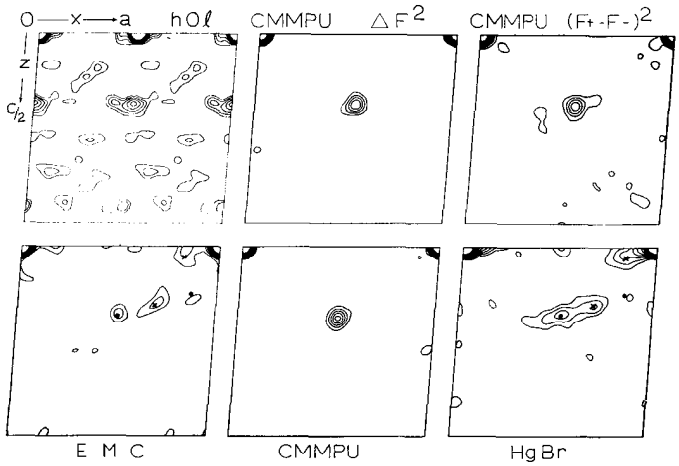

FIG. 1. All difference Patterson syntheses are calculated with 2.0-A resolution data and are contoured at arbitrarily chosen, but equal, intervals. The *map* in the *upper left* is a projection difference Patterson for chloromercuri-2-methoxypropyl urea (*CMMPU*) calculated with h0l coefficients $[F(\text{nat}) - F(\text{CMMPU})]^2$ from the complete set of hkl data. The *map* in the *middle* of the *upper row* is the Harker section, $y = 0$, calculated with complete three-dimensional data $[F(\text{nat}) - F(\text{CMMPU})]^2$. The h0l projection has a higher relative background. It shows translational symmetry due to the c face centering. The Harker section on the *upper right* illustrates the information content of the Bijovet pairs $[F + (\text{CMMPU}) - F - (\text{CMMPU})]^2$. The *three* Harker sections on the *bottom row* (C_2H_5HgCl (*EMC*), CMMPU, and HgBr) were calculated using as coefficients.

$$F^2(\text{nat}) + F^2(\text{der}) - 2 \cdot F(\text{nat}) \cdot F(\text{der}) \cdot$$
$$[1 - K(F + (\text{der}) - F - (\text{der}))/2 \cdot F(\text{nat})]^{1/2}$$

(Ref. 36) with $K = 5.0$. This formula combines heavy atom difference and anomalous dispersion information. The *circles* mark the positions of mercury self-vectors as calculated from least squares refinement coordinates. The *crosses* mark the positions of major site to minor site cross-vectors. These appear in the Harker section because, by coincidence, the two sites differ by only 0.8 A in y.

of the three derivatives, all the crystals for that particular derivative were soaked together in the same vial. The conditions were: C_2H_5HgCl, 0.8 mM, 4 days; $HgBr_2$, 1.2 mM, 4 days; and chloromercuri-2-methoxypropyl urea, 1.5 mM, 3 days. These conditions had been judged to give maximum peak background ratios in the difference Patterson maps (Fig. 1).

Initially we processed data from only those films showing visible intensity changes relative to native films. Subsequently difference Patterson and difference Fourier syntheses were calculated for less promising derivatives. As summarized in Table I most of the compounds either did not bind to the protein or else bound to the sulfur atom of cysteine-18. Several of the compounds which did not show a significant reaction with the protein were soluble to less than 0.1 mM in phosphate and are indicated in Table I by ($<10^{-4}$). A third group of compounds either distorted the unit cell or bound at several sites with very low occupancy.

We are now testing the ability of various lanthanide series compounds to replace calcium. The results will be described later.

Data Processing—Since most of the procedures involved have been well described and are becoming routine, we will emphasize only those aspects which are somewhat unique to the present problem. Much of this information is summarized in Tables II and III. The data-processing procedures and evaluations are

TABLE I

Heavy atom compounds

—SH site(s)	Disorder or weak	No reaction
ClHg OCH$_3$ | | CH$_2$—CH—CH—NHCONH$_2$	K_2HgI_4	$KAuCl_4$
$HgBr_2$	$PbBr_2$	$NaWO_4$
C_2H_5HgCl	K_2PtCl_4	$ClHgC_6H_4OH$
CH_3HgCl	$ThCl_4$	$Pb(CH_3CO_2)_2$
$C_6H_5HgNO_3$	$AgNO_3$	$Pb(NO_3)_2$
$Hg(C_2H_5)_2$	$UO_2(CH_3CO_2)_2$	K_2PtI_4
$Hg(CH_3CO_2)_2$	*o*-$(OHHg)C_6H_4CO_2$	H_2PtCl_6
$C_6H_5HgO_2CCH_3$	*p*-$ClHgC_6H_4CH_3$	$IrCl_3$
p-$ClHgC_6H_4CO_2H$	*p*-$ClHgC_6H_4CH_2$	$UO_2(NO_3)_2$
	$(NO_2C_6H_3(CHO)O)_3Eu$	$Ba(CH_3CO_2)_2$
p-$CH_3CO_2HgC_6H_4NH_2$		CH_3HgCl ($<10^{-4}$)
o-$Hg(HOC_6H_4CO_2)_2$		$Hg(C_6H_5)_2$ ($<10^{-4}$)
		CH_3HgI ($<10^{-4}$)
o-$CH_3CH_2HgSC_6H_4CO_2Na$ (thimerosal)		$CH_3C_6H_4HgCl$ ($<10^{-4}$)
o-$NaCO_2OC_6H_4CONHCH_2CH(OCH_2)CH_2HgOH$ (mersalyl)		$NH_2C_2H_4NH_2PtCl_2$ ($<10^{-4}$)

TABLE II

Summary of data processing and heavy atom parameters

Film packs		*R* merge		*x*	*y*	*z*	W (80 electrons)	B
Native	56	0.067						
CMMPU[a]	25	0.100	Hg	0.2570	0.0	0.0931	0.490	12.0
	34	0.112						
EMC	30	0.102	Hg_1	0.2546	0.0	0.0921	0.250	9.0
			Hg_2	0.4354	−0.0146	0.0613	0.130	13.0
HgBr	32	0.103	Hg_1	0.2608	0.0	0.0917	0.346	8.0
			Br_1	0.1817	0.0154	0.0988	0.150	28.0
			Hg_2	0.4363	−0.0058	0.0650	0.113	20.0
			Br_2	0.5000	0.0121	0.0600	0.050	18.0

[a] CMMPU, chloromercuri-2-methoxypropyl urea; EMC, C_2H_5HgCl.

summarized for the native data, chloromercuri-2-methoxypropyl urea, C_2H_5HgCl, and $HgBr_2$ (bound as —HgBr). About 30 zones or packs of one-zone precession photographs are required to record essentially all of the 2.0-A resolution data, using nickel-filtered CuKα radiation at crystal to film distance of 6 cm. Fifty-six native planes were processed in checking reproducibility of the densitometry procedures and in obtaining intense enough photographs to record 1.85-A data. For chloromercuri-2-methoxypropyl urea, 34 all-zone (no layer line screen) film packs were processed and compared with the 25 one-zone photographs in evaluating the all-zone method.

To a great extent we followed the procedures described by Xuong and Freer (15) for treating the all-zone data. Crystals, (0.3 mm)^3, were mounted with the *b* axis (61.0 A) parallel to the rotation or spindle axis of the camera to ±2.5′. Six- to eight-hour photographs were taken at 6° intervals on the spindle axis using a 0.5-mm collimator and tube settings of 40 kv, 16 ma. The precession angle was 3° 30′; data beyond 3° 10′ were not processed because of the uncertainty in assigning the Lorentz correction. This procedure recorded a set of data in reciprocal space as a torus whose inner radius is zero, or in fact slightly negative, and whose outer radius is $2/\lambda CuK\alpha = 1.30$ A^{-1}. At a cassette advance setting of d = 0 mm only data to about 1.9-A resolution fall on the film. We processed data to 2.1-A resolution. With 3° 10′ data at 6° intervals, even on the torus equator there is a 20′ overlap of data between film sets; reflections away from the torus equator may be recorded on five or more different photographs. Since all photographs were taken about the *b* axis, Bijvoet pairs occur on the same film and are related by mirror symmetry. This symmetry is a great aid in checking crystal alignment, reflection indexing, and intensity measurement. There is a cone-shaped volume of reciprocal space not recorded in the torus of data. It consists of 5% of the volume of the 0.5-A^{-1} sphere of 2.0-A data. Most of this volume can be recorded in eight *b* axis, zero level photographs.

Although we have not yet completed an analysis of this no-screen method, our impressions seem to correspond to those of other users. Exposure times are reduced by a factor of six to ten; however, one requires a better crystal than those often used in one-zone work. Data processing is more expensive and manual checking of questionable reflections is much more tedious. The internal consistency of measured intensities is slightly poorer; however, this may be attributed to an inadequate computer algorithm. The intensities were measured using a rotating drum densitometer at 0.2-mm raster with magnetic tape output. Integrated intensities were determined using a modification of an algorithm previously described (15).

The data from each film pack, consisting of from two to four films in both one-zone and all-zone procedures, were corrected for Lorentz and polarization factors and scaled internally. Data from different film packs were scaled together in four equally populated shells of sin θ/λ (17). Using the scale factor from these four zones one could calculate a relative temperature factor in the expression exp-B $(\sin\theta/\lambda)^2$ so that all of the data had the same average intensity distribution as a function of sin θ/λ. The residual for the merging of all data into one set is defined: R(merge) = Σ(all reflections) Σ(all packs) $|\bar{I}$(average) − I(pack) $|$ /Σ(all reflections) Σ(all packs) $\bar{I}$(pack). The residuals for the one-zone and all-zone data of chloromercuri-2-methoxypropyl urea are presented separately; the $\bar{I}$(average) were determined for all 59 packs. In the native data the Bijvoet pairs within each pack were averaged before merging, whereas in the derivative data they were carried separately. This accounts to some extent for the lower R value of the native data. Native

TABLE III

Summary of phase calculations

As defined in the text ϵ_{12} is the estimated isomorphic error. $A\pm$ is the average error in the Bijvoet difference per film pack. $\epsilon\pm$ is the estimated average anomalous dispersion error per reflection. ϵ_j is the average lack of closure calculated in the phase determination procedure. D_j is the average calculated anomalous dispersion error. The numbers beneath the name of the derivatives indicate, respectively, the total number of reflections used in the phase calculations and the number for which the Bijvoet difference was used.

Sin θ/λ		0.000	0.126	0.179	0.219	0.25
res (A)		∞	4.0	2.8	2.3	2.0
CMMPU[a]	ϵ_{12}	33.0		24.7		
4640	ϵ_j	27.5	25.0	23.7	25.7	
	$\bar{\epsilon}_j/\bar{f}$	0.098	0.108	0.137	0.181	
3220	$A\pm$	11.0	12.4	14.6	15.5	
	$\epsilon\pm$	5.2	6.6	7.6	7.4	
	D_j	5.3	6.3	6.9	9.1	
EMC	ϵ_{12}	38.2		27.9		
4085	ϵ_j	24.8	23.9	21.2	23.5	
	$\bar{\epsilon}_j/\bar{f}$	0.088	0.104	0.124	0.150	
2405	$A\pm$	13.7	17.9	18.9	19.3	
	$\epsilon\pm$	9.5	10.8	11.4	12.4	
	D_j	8.8	11.6	11.4	10.0	
HgBr	ϵ_{12}	33.2		21.8		
3565	ϵ_j	23.4	22.7	20.5	22.0	
	$\bar{\epsilon}_j/\bar{f}$	0.083	0.098	0.116	0.153	
2292	$A\pm$	15.8	21.0	20.3	19.4	
	$\epsilon\pm$	9.1	12.1	14.4	16.9	
	D_j	10.3	11.9	12.2	11.2	
Figures-of-merit		0.887	0.817	0.748	0.612	
No. of native reflections		795	1378	1496	1386	

[a] CMMPU, chloromercuri-2-methoxypropyl urea; EMC, C_2H_5-HgCl.

data were put on an absolute scale by standard Wilson statistics. Derivative data were scaled to the native data by applying relative scale factors and temperature coefficients.

Phase Determination—Initially we calculated h0l (centric projection, $\alpha = 0$ or π) difference Patterson syntheses for each heavy atom derivative showing large intensity differences. Several of the derivatives gave similar difference Pattersons. We explored binding conditions for chloromercuri-2-methoxypropyl urea and collected three-dimensional data for it because it had given the "cleanest" appearing projection Patterson (Fig. 1). We calculated the three-dimensional difference Patterson and in particular the Harker section at $y = 0$. In the Harker section one should see only those vectors between heavy atoms at the same "height" or y coordinate. The difference Patterson functions in Fig: 1 illustrate the improvement in going from projection to three-dimensional data.

All of the other compounds, Table I, which reacted with the protein bound at the chloromercuri-2-methoxypropyl urea (major) site. Two of them, HgBr and C_2H_5HgCl also bound at a nearby minor site. Subsequently we have interpreted the major site as a rather large pocket on one side of the sulfur atom of cysteine-18, while the minor site is a smaller pocket on the opposite side of the sulfur atom. The chloromercuri-2-methoxypropyl urea molecule is apparently too large to fit into the smaller pocket. As judged by the mercury-sulfur distances there must have been a movement of the sulfur atom of some 0.5 to 1.0 A upon bonding. The heavy atom occupancies are low enough, Table II, so that one need not postulate simultaneous binding of a mercury at both sites of an individual protein.

The Harker sections of HgBr and C_2H_5HgCl (Fig. 1) illustrate an unfortunate coincidence. Not only are the major site and minor site self-vectors seen, but also the pair of major-minor cross-vectors. The major and minor sites differ by only 0.8 A in y. This means that the 2-fold phase ambiguity associated with having only 1 heavy atom derivative (the major site chloromercuri-2-methoxypropyl urea) is not generally resolved by the second derivative (the minor site HgBr or C_2H_5HgCl). We, therefore, had to rely on the anomalous dispersion effect to resolve the phase ambiguity. The information content of the Bijvoet pairs is illustrated in the anomalous dispersion Patterson of Fig. 1.

Heavy atom coordinates were determined by least squares refinement with Patterson-determined coordinates as starting values for both major and minor site mercury atoms. The y coordinates of the major site for all three derivatives are assumed to be 0. The sense of the y coordinate for the minor mercury site in both HgBr and C_2H_5HgCl was determined by refining both choices against the experimentally observed anomalous dispersion differences. Both major and minor bromine sites were located in a difference, difference Fourier synthesis and subsequently refined. In Table II these coordinates are listed in fractions of the unit cell. The occupancy, W, is given relative to 80 electrons for both mercury and bromine.

Phases were calculated by the method of Blow and Crick (18) as described by Wyckoff *et al.* (19). In this method an estimate of input error is required to determine the phase probability distribution in the complex plane. As is routinely done, we estimate the three-dimensional isomorphic error, ϵ_{12} in Wyckoff's notation, from the residual errors in the centric zone least square refinement of heavy atom parameters, $\epsilon_{12} = \Sigma(\text{all reflections}) \mid F(\text{obs}) - F(\text{calc}) \mid /n(\text{all reflections})$, where F is the structure factor amplitude as observed experimentally and as calculated. The phase determination calculation yields a final estimate of lack of closure error. Our estimate of isomorphic error corresponded rather well with the calculated lack of closure, ϵ_j (see Table III).

In estimating the anomalous dispersion errors, $\epsilon\pm$, we took advantage of the considerable amount of redundancy in our data. We first determined population statistics by analyzing all those cases in which a Bijvoet pair of reflections had been observed on four or more film packs. The average error in the Bijvoet difference, $\Delta F \pm = F_+ - F_-$, per film pack is

$$A\pm = \sum_{i=1}^{n} \mid (\Delta F\pm)_i - \overline{\Delta F}\pm \mid /(n(n-1))^{1/2}$$

where n is the number of observations and

$$\overline{\Delta F}\pm = \sum_{i=1}^{n} (\Delta F\pm)_i/n.$$

The distribution of this average error was recorded as a function of $\sin\theta/\lambda$; it did not show any variation as a function of $\bar{F}$. For input to the phasing calculation, the reflections were divided into two classes. For those reflections where the Bijvoet pair had been observed six or more times, the anomalous dispersion error was calculated directly as $\epsilon\pm = A\pm/(n)^{1/2}$. For the remainder of the reflections we set $\epsilon\pm = A\pm(\sin\theta/\lambda)/(n)^{1/2}$,

where $A\pm(\sin\theta/\lambda)$ is the value of the average error derived from the $\sin\theta/\lambda$ distribution. For all three derivatives the estimated anomalous dispersion error per reflection, $\epsilon\pm$, agrees rather well with the calculated error, $D_j = |\Delta\bar{F}\pm - (F + j - F - j)|$. The calculation of the anomalous dispersion difference, $(F + j - F - j)$, is thoroughly discussed by Wyckoff *et al.* (19). The ratio of the imaginary and real components of the mercury atomic scattering factor was used as $\kappa = 0.12$. The average error per reflection, $\epsilon\pm$, is significantly lower than the average error per reading, $A\pm$, because so many Bijvoet pairs were recorded on different film packs.

Within the 2-A sphere there are 6725 lattice points in the quadrant, with $h \geqslant 0$, $k \geqslant 0$, of which 306 are centric reflections, *i.e.* h0l. Of these, 5055 (226 of which were centric) reflections were assigned phases and used in the electron density map calculation. This means that both the native and at least one derivative were experimentally observed as non-zero. Although we judged chloromercuri-2-methoxypropyl urea to be the best derivative, the ratio of the average lack of closure to average structure factor amplitude, $\epsilon_j/\bar{F}$, is actually higher than for other derivatives. This is due partly to the fact that the chloromercuri-2-methoxypropyl urea films are more intense and include more weak reflections.

The distribution of calculated phases angles for the native protein is essentially flat between 30 and 150° (as well as 210–330°) with an average of 51.6 phases per 5° interval. The distribution then rises toward 0° and peaks sharply with 311 reflections (excluding h0l reflections) in the $-2\frac{1}{2}°$ to $+2\frac{1}{2}°$ interval. Of the 4829 noncentric reflections there are 773 more near 0 or 180° than predicted by a random distribution of relative phase angle (see also Fig. 2 in Paper III (3)).

Electron Density Map Calculations and General Characteristics—The maps are calculated at intervals of 0.705 A in X and 0.762 A in Y and in layers 0.676 A in Z, that is perpendicular to the c face of the unit cell. The map was contoured on paper with a Calcomp plotter at 2 cm = 1 A and traced onto Perspex sheets. These were mounted vertically and viewed through a half-silvered mirror (20) so that one could visually superimpose a Kendrew skeletal model on the electron density.

The first map was calculated in May 1971 (13) with the 2.0-A data from chloromercuri-2-methoxypropyl urea and with two-thirds of the 2.0-A data from $HgBr_2$ and C_2H_5HgCl. The general course of the main chain and about one-third of the side groups were correctly interpreted during the few weeks before Coffee and Bradshaw had the sequence data for the tryptic peptides. As discussed in their paper (1) our tentative interpretation of the electron density map allowed them to assign the relative order of most of the peptides. Later in the summer of 1971 we calculated an improved map with the data summarized in Tables II and III. In general this map was easily interpreted; however, several regions, particularly the CD calcium-binding loop, could not be interpreted until the tangent formula map[2] was calculated by Hendrickson and Karle (3). Consistent with the objective evaluation presented in Paper III (3) it is our subjective impression that the tangent formula map is more readily interpreted. The electron density protrusions of carbonyl groups are enhanced and in 95 instances appear as a distinct knob of electron density. Five of the phenylalanine benzene rings show a dimpling at their center; whereas only one did in the previous map (see Fig. 8 in Paper III (3)). The density at sulfur and calcium positions is more distinct. The continuity of density is strengthened at some weak places along the chain. Spurious joins between groups in van der Waals contact are diminished. We initially built the molecular model using only those chemical constraints implicit in the wire model, *i.e.* canonical bond lengths and angles. Subsequently we examined the model in terms of van der Waals contacts, hydrogen bonds, and ϕ,ψ angles. In a few instances the model was altered to relieve bad contacts but in general we tried to avoid incorporating preconceived notions.

At only two regions in the molecule is there any significant uncertainty in the interpretation. The loop preceding helix A has low electron density, particularly residues alanine-3 and glycine-4. The exact course of the chain is not well defined; as can be seen in Table V the ϕ,ψ values of residues 1, 2, 3, and 5 are outside normally observed ranges. We are therefore not certain of the bond direction between the α carbon and the carbonyl carbon of phenylalanine-2. The positions of the side chain of phenylalanine-2 at the surface and the N-acetylated alanine-1, tucked into the interior of the protein, could conceivably be interchanged. However the present interpretation is reinforced by the finding of a tyrosine at position 2 in carp component C.[3] The additional hydroxyl group is readily accommodated in the present structure; it would make unacceptable van der Waals contacts if the phenylalanine side chain were built into the interior.

The other uncertainty concerns an ellipsoidal mass ($2 \times 2 \times 3$ A) of electron density approximately at the position of the side chain of leucine-35. There seems little doubt of the correctness of glycine-34 or of lysine-38. Threonine-36 and serine-37 are in a β bend (21). Coffee and Bradshaw[4] found no covalently attached phosphate in the molecule. Neutron activation analyses[5] ruled out the presence of mercury, iron, manganese, or magnesium. Electron microprobe analyses[6] of the crystal reveal no atoms of atomic number greater than 20. In hake muscle calcium-binding protein there is neither attached phosphate nor metal ion other than calcium (22). As will be described later, there is an intramolecular 2-fold axis relating the CD and the EF regions of the molecule (see Table IV for direction cosines). It is suggestive that this 2-fold rotation places the mass of electron density only 1.3 A from the major heavy atom site. Further, even though it is not in a special position, its y coordinate is very near 0 ($x = 0.2713$, $y = -0.0025$, $z = 0.2775$).

DESCRIPTION OF MOLECULE

General Shape—The molecule has the approximate shape of a prolate ellipsoid of revolution (Fig. 2). The course of the main chain is best visualized in terms of the six helices A, B, C, D, E, and F which have been interpreted as being derived from a gene triplication (2). Helix C, the CD loop, and helix D are related to the EF region by an approximate 2-fold axis (Table V and Fig. 3), which roughly coincides with the long axis of the ellipsoid. The over-all configuration of the EF region is remarkably similar to a right hand with thumb and forefinger extended at approximate right angle and the remaining three fingers clenched

[2] Structure factors, phases, and atomic coordinates have been deposited with Protein Data Bank, Department of Chemistry, Brookhaven National Laboratory, Upton, N. Y. 11973, and University Chemical Laboratory, Cambridge CB2 1EW, England.

[3] C. J. Coffee, R. H. Kretsinger, and R. A. Bradshaw, manuscript in preparation.

[4] C. J. Coffee and R. A. Bradshaw, personal communication.

[5] Performed by R. Allen, Chemistry Department, University of Virginia.

[6] Performed by F. W. Fraser and E. J. Brooks, Central Materials Research Activity, United States Naval Research Laboratory.

(Fig. 4). The thumb points toward the COOH terminus of helix F. The forefinger points along helix E in the NH_2-terminal direction. The clenched fingers trace the course of the EF loop about the calcium ion. The two right hands representing the EF and CD regions are related by a 2-fold axis. The D thumb is tilted outward representing the bend in helix D at residue 65.

Again in Fig. 5 one views the entire molecule looking down the 2-fold axis. The AB loop does not bind calcium, as do the CD and EF loops, and is 2 residues shorter. It covers the end of helix E and the arginine-75 to glutamic acid-81 internal salt bridge. The AB region can be represented by a third right hand covering the top of the molecule. The thumb and forefinger must be drawn together almost parallel with one another because the AB loop turns out from the surface of the molecule, while the CD and EF loops turn in. The palms of all three right hands face the interior of the molecule.

Solvent Interactions and Core Formation—The ellipsoid volume is 16,900 A^3. The molecule has 812 atoms excluding hydrogen atoms, or 20.8 A^3 per atom. If the ellipsoid had a surface free of convolutions, one could estimate that a shell 2.7 A thick would contain those atoms, plus associated protons, exposed to the solvent. The "inside" of this ellipsoid would then have a volume of 9700 A^3. That is 51.6% of the atoms would be inside or inaccessible to solvent. In Table IV each atom is designated I (internal) or S (surface) as judged by the criterion of whether a water molecule, a sphere of van der Waals radius 1.7 A, could make contact with the van der Waals surface about each atom, or associated hydrogen atom. Of all 812 atoms, 43.2% are inside. Even for this smooth protein, seemingly free of clefts or grooves characteristic of enzymes, more of the atoms are exposed to the surface than predicted by the ellipsoid model.

A general examination of solvent accessibility, as well as a subsequent consideration of hydrogen bonding, gives some insight into the forces stabilizing the protein. Of the 324 main chain atoms (nitrogen, carbonyl carbon, and oxygen) which carry a partial charge 54.3% are internal (Table VI). It is not valid to say simply that nominally hydrophilic atoms are found at the surface.

In contrast, if one considers the partially charged atoms of the side chains, only 17.9% are internal. These partially charged atoms consist of all side chain nitrogen and oxygen atoms as well as the terminal carbon atoms of carboxyl and carbamyl groups. All 10 serine and threonine hydroxyl groups and all 13 lysine NH_2 groups are at the surface. Eight of the nineteen partially charged groups which are internal are γ carbon atoms of aspartic acid and asparagine or δ carbon atoms of glutamic acid and glutamine. Excepting glutamic acid-81, at least 1 of the oxygen atoms bonded to these γ or δ carbon atoms is at the surface. Six of the ten carboxyl oxygens which coordinate the 2 calcium ions are internal. Finally the 2 carboxyl oxygens of glutamic acid-81 and 3 of the 4 atoms of the arginine-75 guanidino group are internal and hydrogen-bonded together. All of the side chain dipoles are at the surface except those associated with the calcium ions or the arginine-75 to glutamic acid-81 salt bridge. The tendency of these dipoles to be at the surface of the molecule is strong enough to bring 89.5% of the other side chain carbon atoms to the surface. For instance, all of the 52 β through ϵ carbon atoms of the 13 lysine chains are at the surface except for 1 β carbon.

Eighteen of the twenty alanine β carbon atoms, as well as the cysteine β carbon, and sulfur are exposed to solvent. The COOH-terminal oxygen atoms of alanine-108 are at the surface. The acetyl group, APO, of alanine-1 is buried.

Even of the strongly hydrophobic side chain atoms, only 69.5% are internal. Hence the generalization that neutral atoms or even neutral side chains are found on the inside is not so valid as is the statement that side groups with dipoles are found at the surface. Nonetheless this protein has a well defined, strongly hydrophobic core consisting of all, or all except 1, of the side chain carbon atoms of 7 phenylalanine, 4 isoleucine, 5 leucine, and 3 valine residues plus half of the side chains of phenylalanine-47, leucine-86, and leucine-105 (Table V and Fig. 2). The volume of this core (2370 A^3) is about one-seventh that of the entire molecule. It is composed of 115 carbon (plus bonded hydrogen) atoms or 20.6 A^3 per atom. The main chain encloses this core; it does not pass through it.

The structural similarity of the AB, CD, and EF regions is seen in the composition of the core. Each of the six loops, either at its beginning or end, has one core group. The homology is strongest in the AB (phenylalanine-24) and CD (isoleucine-97) loops. In general at each turn of each of the six helices there are one or two groups contributed to the core from the inner aspect of that helix. Isoleucine-11 of helix A and the homologous valine-43 of helix C are both part of the core. Interestingly the homologous position 82 in helix E is threonine. Even though valine-43 and threonine-82 are related by the 2-fold axis and have exactly the same staggered configurations relative to the main chain, the hydroxyl group of threonine-82 is just accessible to the solvent; the corresponding carbon of valine-43 is not, because helix D and helix E are 0.7 A further apart than are helices C and F at the homologous level. The inner halves of phenylalanine-47 and leucine-86 form the surface of the core. The homologous leucine-15 is also half-internal, but helix A has a slightly different orientation relative to helix B, compared to CD and EF; so that leucine-15 is barely continuous with the core. The homologous positions in helices A and C are alanine-14 and alanine-46. These are the only 2 internal alanines of the 20 present. Particularly at alanine-46 there is space in the core for a larger group; in the hake protein position 46 is valine (23). As defined by ϕ,ψ angles helix C continues through position 50, which is a core isoleucine. Glycine-89, turns in such a way to terminate the helix; however, the amide hydrogen bond of aspartic acid-90 indicates that the same helical course in helix E is continued as far as it is in helix C. The β position of glycine-89 is oriented so that only an alanine could substitute for it without causing significant shifts of the main chain. The β-γ bond of cysteine-18 does point to the inside of the molecule. There is just enough room for water to approach the sulfur from

TABLE IV (*Table IV on pages* 3319–3321)

The atomic coordinates were measured from the 2 cm per A wire model. Refinement[7] against observed bond lengths and angles using the program of R. Diamond (35) produced an average shift of 0.15 A per atom. In the second column the letter S indicates that the atom is on the surface or accessible to a solvent water molecule of van der Waals radius 1.7 A. An interior or inaccessible atom is indicated by I. The coordinates are listed in A·100 with the origin on a crystallographic 2-fold axis. The height of the mercury atom in chloromercuri-2-methoxypropyl urea defines $y = 0$; x coincides with crystallographic a and z with c^*. The intramolecular 2-fold axis intersects the $y = 0$ plane at $x = 1.82$ A, $z = 9.28$ A, with direction cosines: $\alpha = 0.258$, $\beta = 0.812$ $\gamma = -.524$.

[7] P. Moews performed the refinement calculations.

Table IV--continued

ATOM		X	Y	Z
	ACETYL	0		
CA	I	-255	-57	1837
CC	I	-238	92	1811
O	I	-330	158	1760
N	I	-122	143	1847
	ALA	1		
CA	S	-94	286	1828
CO	I	-147	364	1947
O	I	-163	310	2058
CB	I	-169	338	1704
N	I	-174	491	1923
	PHE	2		
CD2	S	-527	381	2153
CE2	S	-661	359	2105
CZ	S	-710	428	2001
CE1	S	-631	516	1935
CD1	S	-497	535	1979
CG	S	-448	471	2087
CB	S	-305	495	2127
CA	S	-225	578	2029
CO	S	-109	649	2099
O	I	-43	736	2040
N	S	-85	610	2223
	ALA	3		
CB	S	159	639	2242
CA	S	22	671	2301
CO	S	3	822	2285
O	S	-102	876	2318
N	S	106	887	2235
	GLY	4		
CA	S	101	1032	2215
CC	S	63	1031	2067
O	S	100	1122	1990
N	I	-10	929	2028
	VAL	5		
CG2	S	-224	1092	1936
CG1	S	-167	1044	1694
CP	S	-118	1045	1841
CA	I	-53	916	1888
CO	I	70	881	1808
O	S	183	913	1848
N	I	48	817	1695
	LEU	6		
CD2	S	262	485	1631
CD1	I	71	416	1490
CG	I	122	519	1588
CB	I	118	656	1523
CA	I	159	777	1608
CC	I	196	897	1520
O	S	283	977	1555
N	I	128	907	1407
	ASN	7		
NOD2	S	386	1163	1330
NOD1	S	446	1110	1120
CG	I	375	1085	1215
CB	I	268	979	1218
CA	S	154	1017	1313
CO	I	21	1040	1241
O	I	12	1026	1118
N	S	-79	1073	1319
	ASP	8		
OD2	S	-381	1047	1454
OD1	S	-446	1238	1507
CG	S	-374	1166	1441
CB	S	-278	1215	1335
CA	I	-212	1097	1262
CC	I	-209	1126	1112
O	S	-283	1065	1034
N	S	-124	1219	1074
	ALA	9		
CB	S	9	1349	911
CA	S	-111	1257	933
CO	S	-82	1130	854
O	S	-137	1108	745
N	I	2	1047	911
	ASP	10		
OD2	S	383	941	927
OD1	S	298	972	738
CG	I	290	929	852
CB	I	167	866	912
CA	I	39	921	847
CC	I	-72	818	856
O	I	-94	739	763
N	I	-142	820	967
	ILE	11		
CD1	I	-211	668	1303
CG1	I	-266	796	1237
CB	I	-344	770	1105
CG2	S	-447	657	1109
CA	I	-252	725	989
CO	I	-340	712	864
O	I	-344	606	801
N	S	-408	819	831
	ALA	12		
CB	S	-545	962	680
CA	I	-496	820	713
CC	S	-409	778	596
O	I	-455	706	505
N	S	-285	822	598
	ALA	13		
CB	S	-52	842	530
CA	S	-191	789	491
CO	I	-177	638	474
O	S	-190	585	363
N	I	-149	571	585
	ALA	14		
CB	I	-101	370	722
CA	I	-132	426	582
CO	I	-268	369	543
O	I	-291	333	427
N	I	-356	361	641
	LEU	15		
CD2	S	-729	554	800
CD1	S	-730	497	560
CG	I	-646	505	685
CB	I	-590	368	716
CA	I	-490	307	617
CC	I	-541	334	475
O	S	-606	249	413
N	I	-510	453	425
	GLU	16		
OE2	S	-433	653	-39
OE1	S	-454	868	0
CD	S	-476	752	18
CG	S	-573	702	132
CB	S	-504	634	252
CA	S	-553	492	290
CO	I	-516	416	163
O	S	-593	413	65
N	I	-400	355	166
	ALA	17		
CB	S	-201	264	64
CA	I	-352	279	50
CO	S	-409	137	41
O	S	-388	67	-57
N	I	-479	97	145
	CYH	18		
SG	S	-469	5	418
CB	S	-494	-109	274
CA	I	-539	-36	150
CC	I	-689	-14	148
O	I	-767	-103	185
N	I	-728	104	104
	LYS	19		
NZ	S	-1015	629	-194
CE	S	-998	557	-65
CD	S	-1067	421	-69
CG	S	-1026	334	48
CB	S	-892	270	24
CA	I	-870	138	98
CO	I	-947	27	27
O	S	-1063	-0	59
N	S	-881	-34	-68
	ALA	20		
CB	S	-870	-144	-279
CA	S	-943	-142	-144
CO	I	-936	-285	-90
O	I	-843	-320	-15
N	S	-1033	-365	-128
	ALA	21		
CB	S	-1170	-565	-133
CA	I	-1038	-505	-82
CO	I	-928	-594	-136
O	S	-918	-615	-258
N	I	-846	-644	-45
	ASP	22		
CA	S	-735	-731	-84
OD2	S	-850	-966	14
OD1	S	-990	-974	-140
CG	S	-885	-933	-95
CB	S	-788	-846	-171
CO	S	-643	-632	-156
O	S	-631	-634	-279
N	S	-579	-547	-77
	SER	23		
OG	S	-631	-252	-111
CB	S	-565	-337	-204
CA	S	-488	-447	-134
CC	S	-405	-391	-18
O	S	-338	-288	-31
N	S	-412	-460	94
	PHE	24		
CD2	I	-368	-356	568
CE2	I	-295	-383	689
CZ	I	-224	-496	704
CE1	I	-216	-584	602
CD1	S	-284	-556	480
CG	I	-362	-446	465
CB	I	-430	-422	334
CA	S	-338	-417	214
CC	S	-215	-501	244
O	S	-218	-624	234
N	S	-108	-433	279
	ASN	25		
NOD2	S	303	-519	358
NCD1	S	316	-600	149
CG	S	252	-537	230
CB	S	121	-469	202
CA	S	17	-501	311
CC	I	60	-453	449
O	S	86	-334	469
N	S	66	-546	543
	HIS	26		
CD2	S	329	-764	731
NE2	S	360	-881	671
CE1	S	251	-939	627
ND1	S	148	-861	652
CG	I	195	-749	719
CB	I	103	-640	765
CA	I	105	-513	680
CC	I	243	-448	684
O	I	262	-341	742
N	S	338	-515	621
	LYS	27		
NZ	S	851	-941	468
CE	S	803	-843	567
CD	S	732	-728	498
CG	S	654	-644	598
CB	S	562	-548	526
CA	S	475	-464	617
CO	I	480	-319	569
O	S	568	-241	609
N	I	385	-285	484
	ALA	28		
CB	S	297	-150	298
CA	S	378	-149	429
CO	S	302	-59	526
O	S	328	61	534
N	S	208	-119	597
	PHE	29		
CD2	S	-196	-16	737
CE2	I	-301	54	801
CZ	I	-304	70	935
CE1	I	-201	23	1011
CD1	I	-94	-43	947
CG	I	-93	-66	813
CB	I	23	-135	753
CA	S	128	-43	693
CC	I	219	14	800
O	I	213	134	831
N	I	393	-70	856
	PHE	30		
CD2	I	269	-234	1172
CE2	I	181	-342	1210
CZ	I	181	-458	1144
CE1	I	269	-479	1044
CD1	I	360	-374	1009
CG	I	357	-254	1069
CB	S	456	-149	1029
CA	I	396	-27	961
CO	I	503	61	899
O	I	575	133	970
N	I	513	55	768
	ALA	31		
CB	S	627	88	550
CA	I	612	136	695
CO	S	556	277	687
O	S	627	376	710
N	I	428	285	654
	LYS	32		
NZ	S	506	632	192
CE	S	370	614	247
CD	S	369	508	356
CG	S	242	513	439
CB	S	243	408	548
CA	S	362	415	642
CC	I	317	458	782
O	I	355	566	830
N	I	238	373	844
	VAL	33		
CG2	I	-29	299	1051
CG1	I	183	222	1164
CB	I	118	280	1036
CA	I	187	403	979
CO	I	303	447	1068
O	I	283	487	1183
N	I	422	441	1013
	GLY	34		
CA	S	542	481	1087
CC	I	570	421	1226
O	S	576	493	1326
N	I	586	290	1229
	LEU	35		
CA	I	613	220	1355
CD2	S	465	108	1628
CD1	I	278	65	1472
CG	I	405	144	1494
CB	I	502	116	1383
CO	I	749	154	1343
O	I	837	172	1429
N	I	766	79	1236
	THR	36		
CG2	S	1057	-32	1027
OG1	S	820	-63	1002
CB	S	921	18	1061
CA	I	892	9	1213
CO	I	998	77	1300
O	S	1062	13	1384
N	S	1013	207	1279
	SER	37		
OG	S	965	474	1319
CB	S	1100	431	1324
CA	S	1112	284	1356
CC	I	1101	270	1508
O	S	1200	286	1581
N	T	982	239	1554

Table IV—continued

	LYS	38			PHE	47			GLY	56			LYS	64	
NZ	S	758	659	1533	CD2	S	-17	-879	2315	CA	S	-329	-1350	2561	NZ	S	-1465	-1267	958
CE	S	743	577	1656	CE2	S	-54	-1018	2313	CO	I	-281	-1375	2419	CE	S	-1404	-1134	938
CD	S	803	440	1636	CZ	I	-59	-1087	2197	O	S	-160	-1383	2393	CD	S	-1312	-1099	1054
CG	S	729	334	1717	CE1	I	-23	-1027	2081	N	S	-376	-1387	2328	CG	S	-1311	-950	1082
CB	S	810	206	1728	CD1	I	17	-891	2083		PHE	57			CB	S	-1206	-914	1185
CA	S	958	223	1698	CG	S	17	-818	2197	CD2	S	-182	-1635	2329	CA	S	-1073	-867	1127
CO	S	1026	106	1768	CB	S	63	-675	2192	CE2	S	-62	-1707	2360	CO	I	-1083	-715	1113
O	S	1063	7	1705	CA	I	-41	-577	2145	CZ	S	1	-1779	2267	O	I	-1097	-662	1002
N	S	1043	119	1899	CO	I	-163	-574	2236	CE1	S	-45	-1780	2139	N	S	-1075	-648	1226
	SER	39			O	I	-271	-623	2198	CD1	S	-161	-1705	2107		LEU	65		
OG	S	1097	111	2201	N	S	-145	-517	2353	CG	I	-230	-1638	2201	CD2	S	-1381	-642	1321
CB	S	1183	74	2095		ALA	48			CB	S	-351	-1560	2160	CD1	S	-1363	-455	1481
CA	S	1107	14	1978	CB	S	-215	-410	2561	CA	I	-344	-1412	2187	CG	S	-1287	-543	1385
CO	I	1026	-97	2043	CA	S	-254	-508	2450	CC	I	-444	-1331	2104	CB	I	-1221	-457	1280
O	I	903	-89	2052	CC	I	-380	-452	2385	O	I	-552	-1297	2152	CA	I	-1084	-502	1227
N	S	1095	-200	2086	O	S	-492	-496	2414	N	I	-405	-1300	1982	CO	S	-974	-439	1313
	ALA	40			N	I	-361	-354	2298		ILE	58			O	S	-1000	-348	1393
CB	S	1135	-405	2214		ILE	49			CA	I	-492	-1222	1892	N	I	-853	-488	1296
CA	S	1029	-316	2149	CD1	S	-303	37	2271	CD1	I	-256	-917	1856		PHE	66		
CO	S	937	-271	2262	CG1	S	-437	-77	2257	CG1	I	-394	-983	1880	CD2	I	-381	-493	1368
O	I	828	-326	2280	CB	I	-431	-158	2161	CB	I	-411	-1120	1809	CE2	I	-266	-441	1300
N	S	983	-173	2337	CG2	S	-525	-153	2040	CG2	I	-489	-1118	1677	CZ	I	-274	-395	1173
	ASP	41			CA	S	-473	-291	2228	CO	I	-563	-1319	1798	CE1	I	-394	-391	1111
OD2	S	1153	-99	2525	CO	I	-529	-383	2120	O	I	-513	-1348	1689	CD1	I	-510	-439	1179
OD1	S	1164	98	2589	C	S	-651	-400	2108	N	I	-677	-1366	1841	CG	I	-502	-491	1304
CG	S	1102	7	2537	N	I	-440	-440	2043		GLU	59			CB	I	-628	-538	1371
CB	S	959	16	2491		ILE	50			CE1	S	-872	-1361	2082	CA	S	-739	-436	1373
CA	S	905	-121	2449	CD1	I	-177	-503	1685	OE2	S	-810	-1542	2190	CO	I	-701	-304	1307
CO	S	755	-107	2424	CG1	I	-269	-463	1802	CD	S	-848	-1477	2093	O	I	-603	-239	1347
O	S	675	-189	2470	CB	I	-359	-579	1854	CG	S	-864	-1576	1972	N	I	-779	-265	1208
N	I	720	-3	2351	CG2	I	-426	-666	1747	CB	S	-884	-1508	1834		LEU	67		
	ASP	42			CA	S	-480	-530	1934	CA	S	-755	-1461	1761	CD2	I	-560	-133	904
OD2	S	762	247	2334	CO	I	-550	-647	2003	CO	I	-805	-1412	1624	CD1	S	-742	-204	752
OD1	S	595	362	2382	O	S	-642	-708	1947	O	I	-805	-1291	1595	CG	I	-689	-210	893
CG	S	644	266	2327	N	I	-506	-677	2124	N	I	-846	-1506	1542	CB	I	-792	-154	988
CB	S	565	160	2253		ASP	51				GLU	60			CA	I	-754	-139	1136
CA	S	579	22	2320	OD1	I	-602	-1028	2344	OE2	S	-973	-1897	1502	CC	I	-837	-35	1209
CO	S	519	-84	2230	OD2	S	-415	-1016	2435	OE1	S	-844	-1736	1579	O	I	-799	82	1217
O	S	396	-101	2223	CG	S	-501	-965	2369	CD	S	-938	-1781	1520	N	I	-951	-79	1269
N	I	605	-155	2162	CB	S	-469	-828	2316	CG	S	-1043	-1687	1449		GLN	68		
	VAL	43			CA	I	-564	-788	2201	CB	S	-984	-1586	1348	CA	I	-1041	10	1331
CG2	I	756	-278	1916	CC	I	-701	-750	2255	CA	S	-896	-1473	1407	NOE2	S	-1396	-216	1395
CG1	I	648	-493	1998	O	S	-711	-690	2363	CO	S	-980	-1346	1405	NOE1	S	-1436	-230	1174
CB	I	680	-344	2028	N	S	-804	-785	2181	O	S	-935	-1240	1360	CC	S	-1365	-193	1273
CA	I	561	-262	2071		GLN	52			N	S	-1102	-1359	1454	CG	I	-1239	-110	1244
CO	I	457	-345	2147	NOE2	S	-1367	-960	2104		ASP	61			CB	S	-1156	-65	1364
O	I	338	-344	2112	NOE1	S	-1174	-1044	2033	OD2	S	-1231	-1431	1704	CO	S	-970	62	1457
N	I	504	-415	2248	CD	S	-1240	-961	2101	OD1	S	-1246	-1240	1786	O	S	-1019	154	1523
	LYS	44			CG	S	-1164	-853	2178	CG	S	-1257	-1315	1690	N	I	-857	2	1488
NZ	S	928	-585	2605	CB	S	-1035	-815	2121	CB	S	-1309	-1275	1555		ASN	69		
CE	S	794	-527	2588	CA	S	-940	-754	2221	CA	S	-1193	-1245	1457	ND2	S	-745	-295	1588
CD	S	689	-635	2568	CO	I	-973	-805	2362	CC	S	-1119	-1118	1498	ND1	S	-799	-212	1789
CG	S	615	-618	2436	O	S	-976	-727	2458	O	S	-1163	-1006	1468	CG	I	-743	-194	1683
CB	S	483	-547	2457	N	I	-996	-934	2373	N	I	-1008	-1136	1566	CB	I	-676	-66	1640
CA	S	415	-500	2329		ASP	53				GLU	62			CA	S	-780	42	1606
CO	S	289	-421	2363	OD1	I	-952	-1150	2283	OE2	I	-718	-1123	2008	CO	I	-711	176	1582
O	S	177	-469	2343	OD2	S	-1081	-1307	2330	OE1	S	-933	-1094	1981	O	I	-669	243	1677
N	S	311	-302	2415	CG	S	-1038	-1198	2355	CD	I	-819	-1081	1952	N	I	-700	213	1456
	LYS	45			CB	S	-1090	-1135	2480	CG	S	-773	-999	1825		PHE	70		
NZ	S	27	270	2684	CA	S	-1029	-995	2502	CB	I	-802	-1068	1690	CD2	I	-509	70	1273
CE	S	73	189	2570	CC	S	-910	-1009	2597	CA	S	-928	-1022	1612	CE2	I	-427	-45	1290
CD	S	215	142	2592	O	S	-926	-1038	2717	CC	I	-876	-929	1502	CZ	I	-302	-35	1339
CG	S	222	-9	2601	N	I	-792	-987	2543	O	S	-910	-810	1499	CE1	I	-254	85	1378
CB	S	242	-70	2464		LYS	54			N	I	-796	-984	1414	CD1	I	-337	199	1365
CA	S	199	-216	2453	NZ	S	-625	-456	2605		LEU	63			CG	I	-460	192	1311
CO	I	83	-224	2353	CE	S	-676	-544	2713	CD2	I	-408	-965	1265	CB	I	-543	316	1301
O	S	-30	-252	2390	CD	S	-636	-688	2689	CD1	I	-501	-1190	1217	CA	I	-635	339	1419
N	S	115	-198	2227	CG	S	-740	-785	2745	CG	I	-528	-1056	1281	CO	I	-739	446	1386
	ALA	46			CB	S	-683	-925	2756	CB	I	-650	-994	1216	O	I	-707	565	1377
CB	S	84	-187	1985	CA	S	-669	-997	2623	CA	I	-739	-906	1304	N	I	-861	401	1367
CA	I	14	-202	2121	CO	S	-646	-1147	2644	CC	I	-849	-844	1217		LYS	71		
CO	I	-58	-336	2117	O	S	-564	-1188	2727	O	I	-832	-735	1161	NZ	S	-1099	979	1112
O	I	-178	-343	2089	N	I	-719	-1226	2567	N	S	-960	-915	1208	CE	S	-1017	864	1153
N	I	15	-441	2146		SER	55								CD	S	-1040	746	1060
					OG	S	-675	-1412	2339						CG	S	-1077	620	1138
					CB	S	-758	-1438	2452						CB	S	-953	553	1195
					CA	S	-706	-1372	2577						CA	S	-970	493	1334
					CO	S	-560	-1409	2600						CO	I	-1098	410	1344
					O	S	-530	-1511	2663						O	I	-1144	352	1245
					N	I	-472	-1326	2548						N	S	-1153	405	1464

Table IV—continued

	ALA	72		
CB	S	-1351	405	1596
CA	S	-1275	328	1488
CO	S	-1373	310	1372
O	S	-1455	216	1373
N	S	-1364	398	1275
	ASP	73		
CA	S	-1452	392	1158
OD2	S	-1309	527	966
OD1	S	-1472	632	889
CG	S	-1421	566	978
CB	S	-1487	534	1110
CO	I	-1397	315	1039
O	S	-1470	279	945
N	I	-1268	290	1044
	ALA	74		
CB	S	-1064	173	985
CA	S	-1202	217	935
CO	I	-1276	90	897
O	S	-1345	28	980
N	S	-1262	52	771
	ARG	75		
NH2	I	-914	-397	289
NH1	S	-1132	-432	353
CZ	I	-1031	-349	326
NE	I	-1046	-215	334
CD	S	-1171	-152	375
CG	S	-1173	-126	525
CB	S	-1313	-82	569
CA	S	-1329	-68	721
CO	I	-1283	-201	781
O	I	-1203	-205	875
N	S	-1335	-309	724
	ALA	76		
CB	S	-1430	-519	799
CA	S	-1299	-443	772
CO	I	-1222	-526	671
O	S	-1247	-519	559
N	I	-1128	-604	723
	LEU	77		
CD2	I	-731	-588	659
CD1	I	-749	-689	884
CG	I	-826	-632	767
CB	I	-922	-737	714
CA	I	-1046	-688	637
CO	I	-1126	-808	587
O	S	-1219	-856	654
N	I	-1090	-856	469
	THR	78		
CG2	S	-1228	-785	252
OG1	S	-1082	-965	199
CB	S	-1197	-931	265
CA	S	-1159	-971	410
CO	S	-1100	-1111	398
O	S	-986	-1135	441
N	S	-1177	-1200	340
	ASP	79		
OD2	S	-1158	-1625	195
OD1	S	-1349	-1580	265
CG	S	-1240	-1543	224
CB	S	-1201	-1400	199
CA	S	-1133	-1339	323
CO	I	-981	-1335	310
O	I	-908	-1361	408
N	S	-935	-1304	191
	GLY	80		
CA	S	-791	-1297	165
CO	S	-684	-1235	254
O	S	-596	-1305	305
N	I	-694	-1105	272
	GLU	81		
OE2	I	-818	-690	374
OE1	I	-727	-567	215
CD	I	-739	-665	282
CG	S	-648	-791	262
CB	I	-643	-887	383
CA	S	-597	-1033	355
CO	I	-570	-1096	492
O	S	-455	-1108	535
N	I	-677	-1135	558
	THR	82		
OG1	I	-875	-1102	727
CG2	S	-811	-1274	883
CB	S	-808	-1227	740
CA	I	-666	-1197	691
CO	I	-580	-1323	685
O	I	-480	-1335	757
N	S	-621	-1415	600
	LYS	83		
NZ	S	-948	-1858	213
CE	S	-917	-1781	334
CD	S	-775	-1727	330
CG	S	-728	-1681	468
CB	S	-592	-1615	459
CA	S	-549	-1541	585
CO	S	-400	-1510	580
O	S	-322	-1558	664
N	S	-362	-1430	482
	THR	84		
CG2	S	-78	-1287	284
OG1	S	-310	-1342	254
CB	S	-214	-1293	347
CA	S	-221	-1392	465
CO	I	-154	-1332	588
O	S	-49	-1379	633
N	S	-216	-1229	641
	PHE	85		
CD2	I	-285	-867	968
CE2	I	-224	-769	1053
CZ	I	-91	-766	1073
CE1	I	-12	-854	1008
CD1	I	-71	-949	920
CG	I	-205	-957	903
CB	I	-261	-1058	808
CA	I	-161	-1161	759
CO	I	-128	-1259	872
O	I	-12	-1272	913
N	I	-231	-1328	919
	LEU	86		
CD2	S	-496	-1610	1214
CD1	I	-275	-1528	1286
CG	S	-352	-1592	1174
CB	S	-342	-1506	1049
CA	I	-213	-1425	1028
CO	I	-101	-1527	1010
O	I	-25	-1555	1104
N	I	-92	-1580	890
	LYS	87		
NZ	S	-355	-1936	417
CE	S	-329	-1887	555
CD	S	-180	-1873	581
CG	S	-152	-1821	721
CB	S	-18	-1750	727
CA	S	9	-1680	859
CO	S	157	-1642	846
O	S	246	-1719	884
N	S	180	-1524	792
	ALA	88		
CB	S	312	-1328	728
CA	S	317	-1475	772
CO	S	396	-1478	904
O	S	518	-1494	904
N	I	325	-1461	1013
	GLY	89		
CA	S	388	-1461	1146
CO	I	333	-1556	1252
O	S	371	-1548	1370
N	I	245	-1645	1209
	ASP	90		
OD2	I	35	-1900	1448
OD1	S	-128	-1931	1323
CG	I	-25	-1872	1348
CB	S	39	-1768	1259
CA	I	185	-1742	1301
CO	I	276	-1863	1290
O	S	230	-1974	1259
N	S	403	-1841	1317
	SER	91		
OG	S	676	-1785	1352
CB	S	625	-1913	1388
CA	S	501	-1949	1309
CO	S	452	-2084	1362
O	S	429	-2178	1284
N	S	436	-2092	1492
	ASP	92		
OD2	I	224	-2030	1815
OD1	S	205	-2012	1607
CG	I	253	-2068	1705
CB	S	343	-2187	1700
CA	S	389	-2216	1556
CO	S	276	-2280	1476
O	S	245	-2399	1494
N	I	215	-2201	1389
	GLY	93		
CA	S	104	-2250	1307
CO	S	-9	-2314	1386
O	S	-38	-2433	1371
N	I	-71	-2233	1469
	ASP	94		
OD1	I	-101	-2042	1669
OD2	S	-69	-2127	1857
CG	I	-100	-2139	1742
CB	S	-142	-2278	1701
CA	S	-182	-2282	1551
CO	S	-305	-2196	1523
O	S	-413	-2220	1578
N	S	-286	-2096	1438
	GLY	95		
CA	S	-395	-2005	1402
CO	S	-402	-1908	1520
O	S	-511	-1876	1568
N	I	-286	-1864	1564
	LYS	96		
NZ	S	-307	-2054	2163
CE	S	-301	-1981	2034
CD	S	-435	-1917	2002
CG	S	-439	-1837	1872
CB	I	-292	-1843	1810
CA	I	-279	-1770	1677
CO	I	-149	-1690	1681
O	I	-41	-1744	1658
N	I	-162	-1562	1709
	ILE	97		
CD1	I	-144	-1308	1370
CG1	I	-100	-1419	1469
CB	I	-49	-1365	1605
CG2	I	93	-1312	1608
CA	I	-45	-1474	1716
CO	I	-45	-1408	1853
O	I	-149	-1364	1902
N	I	71	-1401	1913
	GLY	98		
CA	S	85	-1340	2046
CO	I	212	-1257	2060
O	I	275	-1220	1961
N	S	248	-1229	2184
	VAL	99		
CG2	S	442	-1351	2342
CG1	S	384	-1141	2470
CB	S	437	-1201	2338
CA	I	368	-1149	2212
CO	I	478	-1139	2106
O	I	467	-1062	2010
N	S	582	-1217	2126
	ASP	100		
OD2	S	812	-1470	2249
OD1	S	805	-1266	2294
CG	S	804	-1356	2212
CB	S	787	-1336	2064
CA	S	694	-1217	2032
CO	S	663	-1224	1883
O	S	721	-1149	1802
N	S	573	-1313	1848
	GLU	101		
OE1	S	210	-1632	1675
OE2	S	366	-1786	1665
CD	S	322	-1672	1667
CG	S	444	-1573	1657
CB	I	413	-1426	1691
CA	S	534	-1330	1707
CO	I	496	-1198	1641
O	S	524	-1174	1523
N	I	431	-1113	1718
	PHE	102		
CD2	I	126	-805	1605
CE2	I	71	-681	1557
CZ	I	102	-564	1614
CE1	I	193	-561	1716
CD1	I	251	-682	1761
CG	I	216	-802	1708
CB	I	283	-926	1759
CA	I	388	-983	1667
CO	I	511	-892	1657
O	I	545	-842	1549
N	I	577	-874	1771
	THR	103		
CG2	S	893	-737	1927
OG1	S	670	-777	2007
CB	S	766	-816	1911
CA	I	697	-789	1776
CO	I	788	-818	1657
O	I	828	-726	1584
N	S	819	-945	1640
	ALA	104		
CB	S	912	-1139	1536
CA	S	906	-986	1529
CO	S	862	-950	1387
O	S	944	-910	1303
N	S	734	-966	1362
	LEU	105		
CD2	I	303	-976	1176
CD1	S	452	-997	980
CG	I	440	-947	1122
CB	S	546	-1012	1208
CA	S	678	-936	1229
CO	I	656	-786	1212
O	S	667	-731	1101
N	I	625	-722	1322
	VAL	106		
CG2	I	380	-557	1439
CG1	I	566	-396	1500
CB	I	529	-536	1448
CA	I	600	-577	1320
CO	I	740	-518	1304
O	S	765	-438	1213
N	I	828	-557	1394
	LYS	107		
NZ	S	1153	-130	1536
CE	S	1088	-261	1518
CD	S	1142	-363	1618
CG	S	1049	-483	1632
CB	S	1048	-566	1505
CA	I	966	-508	1391
CO	I	1027	-545	1256
O	S	1132	-491	1217
N	I	962	-637	1188
	ALA	108		
CB	S	1067	-563	980
CA	S	1011	-683	1058
CO	S	1126	-781	1075
O	S	1193	-817	978
	CALCIUM			
CACO		-717	-1205	2237
CAEF		111	-1870	1714

both sides of the main chain. The heavy atom derivatives, HgBr and C_2H_5HgCl can attach to the sulfur atom from either side, whereas chloromercuri-2-methyoxypropyl urea is apparently too large to bind at the inner site.

The three helices, B, D, and F, also show homologous contributions to the core. At the first inner turn of the helices there are phenylalanine-29, leucine-63, and phenylalanine-102. As was the case with leucine-15, the A and the B helices are related slightly differently from helices C, D, and helices E, F so that phenylalanine-30 is also in the core while lysine-64 and threonine-103 are not. In fact leucine-15 twists away from the core in the same sense and in the same region that phenylalanine-30 twists in toward the core. Phenylalanine-66 and leucine-105 are homologous and internal, as are valine-33, leucine-67, and valine-106. Finally helix B crosses the end of helix F with its phenylalanine-30 in such a way that helix F could continue at most 1 small residue at position 109 in helical configuration. In fact the pike protein has alanine at position 109 (24). The terminus of helix D is not so physically blocked and can complete the last turn, contributing phenylalanine-70 to the core. In summary, even though there are many hydrophobic atoms at the surface, even of residues like phenylalanine, leucine, isoleucine, and valine, there is a large, well defined hydrocarbon core.

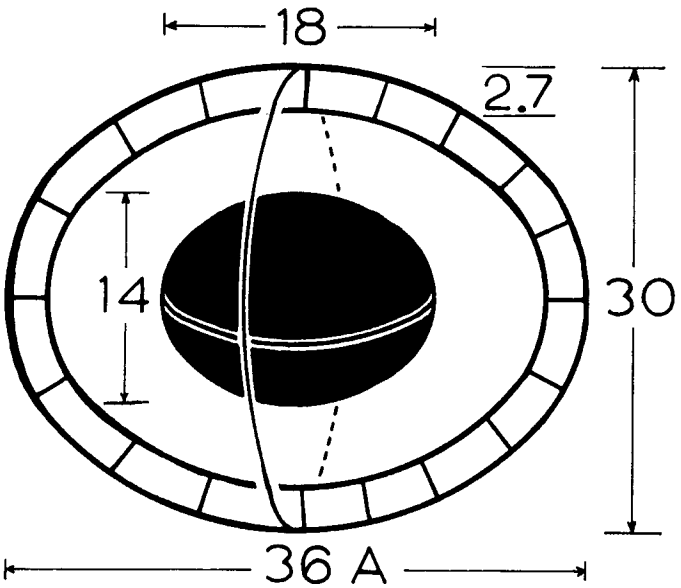

FIG. 2. The muscle calcium-binding protein molecule has the general shape of a prolate ellipsoid of revolution. The intramolecular approximate 2-fold axis corresponds to the long axis of the ellipsoid. The shell, 2.7 A thick, contains those atoms, exclusive of hydrogen, which would be exposed to the solvent if there were no surface indentations. There is an oblate ellipsoid hydrocarbon core consisting of side chains of phenylalanine, leucine, isoleucine, and valine. The volume of the entire molecule is 16,900 A^3; of the molecule inside the 2.7-A shell, 9700 A^3; and of the core, 2400 A^3.

TABLE V

Summary of structural parameters and interactions

Sequence numbers and residues (A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; N, asparagine; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine) are listed in three columns whose alignment corresponds to the three homologous regions of the protein. Question marks after residues 1, 2, 3, 4, 5 and 36 indicate uncertainty as to their orientation. Regions enclosed by solid lines define helices A through F. Dashed lines indicate ambiguity in defining the termini of helices

Table V—continued

B and E. Column 5, headed BN, indicates receivers of hydrogen bonds from backbone nitrogen atoms. A number indicates which backbone oxygen receives a hydrogen bond; a blank indicates no bond formed. The column headed Side summarizes side chain interactions. For both BN and Side s means the proton available for hydrogen bonding is accessible to solvent. The following symbols preceded by numbers indicate hydrogen bond receivers: CO, backbone oxygen; O_2, carboxylate oxygen; NO, carbamyl group oxygen; and OH, hydroxyl group oxygen. The symbols ■ and ▲ refer to hydrophobic side groups contributing wholly or partially to the core. Finally X, Y, and Z assign octahedral vertices to the calcium ligands. The axis is followed by the calcium-oxygen bond distance.

	φ	ψ	BN	Side
1 A	85	-149?		
2 F	93	-107?		
3 A	48	-123?		s
4 G	-97	31?		s
5 V	71	155?	2	
6 L	-83	-82		■
7 N	-151	62		ss
8 D	-23	-55		s
9 A	-55	-42		s
10 D	-55	-42	7	
11 I	-41	-67	7	■
12 A	-55	-37	8	
13 A	-54	-55	9	
14 A	-67	-80	10	
15 L	-31	-39	11	
16 E	-59	-26	12	
17 A	-80	-12	14	
18 C	-117	11	15	
19 K	-50	-32		sss
20 A	-89	146		s
21 A	-62	112		s
22 D	65	75	$81O_2$	
23 S	-165	12	$81O_2$	20CO
24 F	-105	132		s ■
25 N	-122	115		s 22CO s
26 H	-51	-48		s ss
27 K	-52	-25		s sss
28 A	-83	-31	25NO	
29 F	-61	-48	25	■
30 F	-70	-16	26	■
31 A	-79	-44		
32 K	-84	-65	28	sss
33 V	-46	-6	29	■
34 G	50	63		
35 L	-112	-58	30	■
36 T	-18	-57?		s
37 S	-56	-27		s s
38 K	-66	162	35	sss
39 S	-97	165		s s
40 A	-52	-39		s
41 D	-41	-77		s
42 D	-70	-24		
43 V	-41	-70	39	■
44 K	-43	-51	40	sss
45 K	-34	-50		s sss
46 A	-59	-39	42	
47 F	-53	-44	43	◣
48 A	-48	-39	44	
49 I	-70	-50	45	
50 I	-58	-37	46	■
51 D	-71	86	47	X2.4
52 Q	-48	-77		s $62O_2$ s
53 D	-67	-15	$53O_2$	Y2.5
54 K	76	18	51	48 51CO s
55 S	-43	-28		s Z2.3
56 G	113	-7	$51O_2$	
57 F	-142	155	55OH	-Y2.1
58 I	-99	89	97	■
59 E	-55	164		-X2.7
60 E	-47	-69	$60O_2$	
61 D	-41	-15		s
62 E	-56	-75		-Z2.4
63 L	-58	-32	60	■
64 K	-28	-57	61	sss
65 L	-140	48		s
66 F	-78	-2	63	■
67 L	-86	-39		■
68 Q	-67	-8	65	ss
69 N	-73	-18		ss
70 F	-102	-14		■
71 K	-166	92		sss
72 A	-24	-26		s
73 D	-93	-16		s
74 A	-47	152	71	
75 R	-69	-179	s	18CO 18CO 81 O_2 s s
76 A	-111	141		s
77 L	-69	154	64	■
78 T	-108	180	$81O_2$	s
79 D	-30	-79		s
80 G	-53	-60		s
81 E	-50	-40	220_2	
82 T	-65	-52	78	s
83 K	-48	-59	79	sss
84 T	-54	-53	80	s
85 F	-46	-64	81	■
86 L	-48	-35	82	◣
87 K	-72	-35	83	sss
88 A	-45	-28	84	
89 G	-127	5		
90 D	-93	61	87	X2.7
91 S	-39	-72		s s
92 D	-46	-16	$101O_2$	Y2.0 2.2
93 G	54	69	90	
94 D	-123	-1		Z2.8
95 G	79	45		s
96 K	-160	145		sss-Y2.1
97 I	-131	136		■
98 G	-133	-164		
99 V	-31	-93		s
100 D	-47	-46		s
101 E	-52	-45		s -Z2.5
102 F	-75	-57	98	■
103 T	-47	-46	99	s
104 A	-66	-44	100	
105 L	-78	-35	101	◣
106 V	-74	-60	102	■
107 K	-52	-18	103	36 38COs
108 A	78			

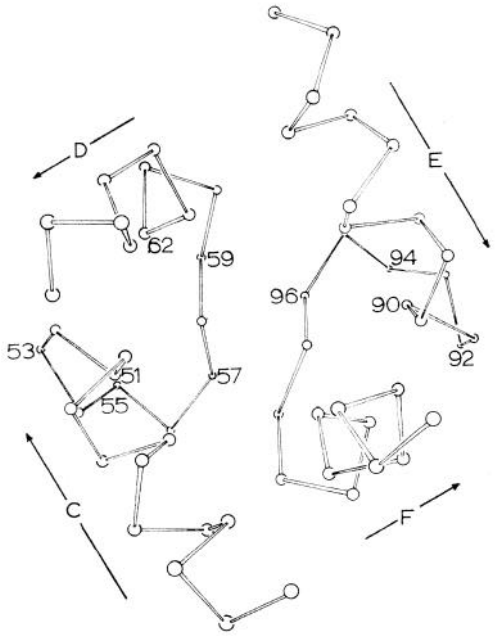

FIG. 3. Helix C, the CD calcium binding loop, and helix D are related to the EF regions by an approximate 2-fold axis.

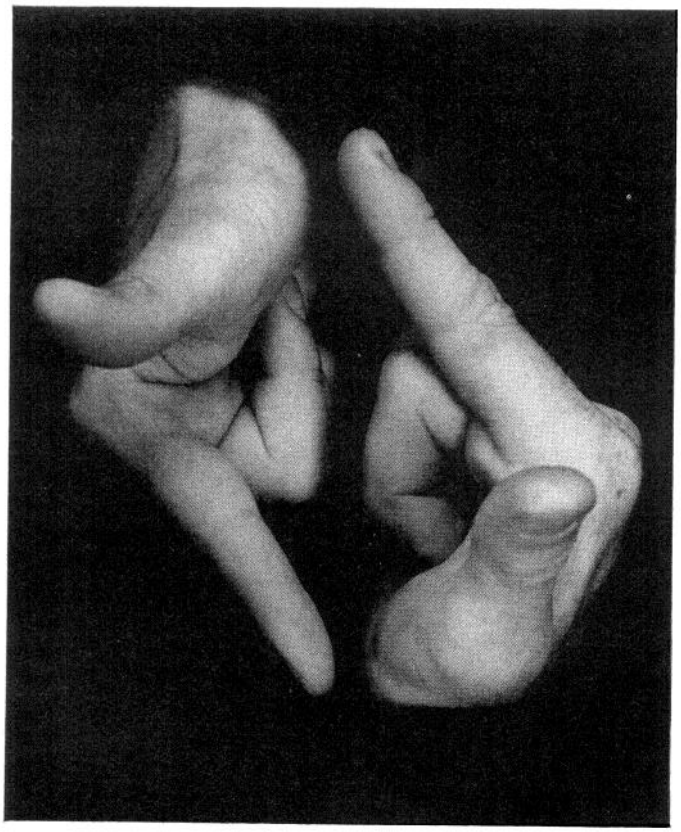

FIG. 4. The CD and EF regions are symbolized by a pair of right hands. Helix C (and helix E) runs from the tip to the base of the forefinger. The flexed middle finger corresponds to the CD (and the EF) calcium binding loop. Helix D (and helix F) runs to the end of the thumb. The thumb for helix D is tilted outward, representing the kink in this helix at leucine-65.

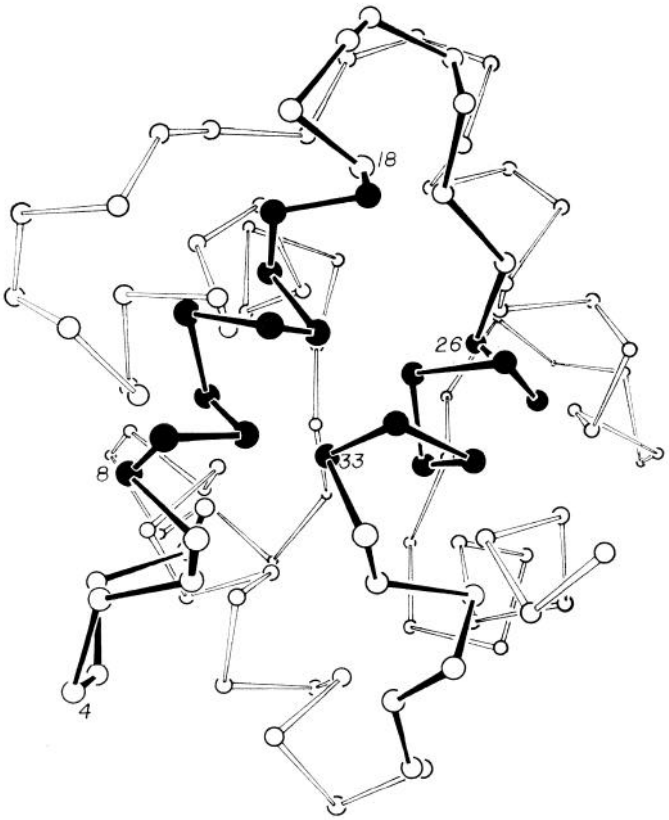

FIG. 5. The pre-A bend, A helix, AB bend, and B helix are indicated by *solid lines* between α carbons. The methyl carbon of the *N*-acetyl group is drawn as an α carbon to illustrate that it is tucked back into the interior of the protein. If the AB region is symbolized by a right hand the forefinger and thumb must be drawn together almost parallel.

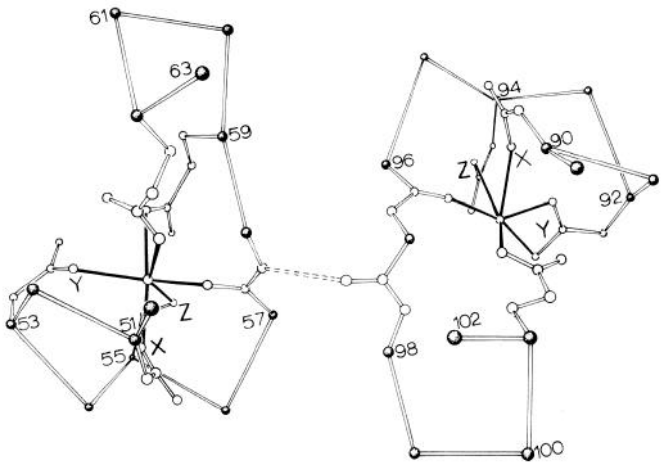

FIG. 6. The CD and EF calcium binding loops are viewed, as in Figs. 3, 4, and 5, down the 2-fold axis. α carbons are indicated by *stippling*. The individual side chain atoms which coordinate the 2 calcium ions are drawn. The main chain carbonyl oxygen atoms of phenylalanine-57 and of lysine-96 coordinate the CD and EF calcium ions. There is a hydrogen bond from the peptide nitrogen of isoleucine-58 to the carbonyl oxygen of isoleucine-97.

Analysis of Hydrogen Bonding—Due primarily to the fact that oxygen atoms can accept 2 protons in hydrogen bonds, whereas hydroxyl oxygen and peptide nitrogen atoms can donate only 1 proton to hydrogen bonds, most proteins have many more potential hydrogen bond acceptors than donors. In muscle calcium-

Table VI

Solvent Exposure

atom	sol	int	Σ
BN	44	64	108
BA	67	41	108
BP	37	71	108
BO	67	41	108
dipole	87	19	106
KRH^{DNS}_{EQT}	94	11	105
FILV	43	98	141
A	18	2	20
C	2	0	2
O-108	1	0	1
APO	0	3	3
Ca	1	1	2
Σ	461	351	812

TABLE VII

Hydrogen Bond

Donors and Receivers

	BO	side	sol	no	Σ
BN	48	10	29	21	108
K	4		35		39
R	2	1	2		5
H			2		2
N	1		5		6
S	1		4		5
Q		1	3		4
T	—	—	5	—	5
	56	12	86	21	174

Backbone atoms are BN, BA, BP and BO. "dipole" designates all side chain nitrogen and oxygen atoms as well as the terminal carbons of carboxyl and carbamyl groups. "KRHDNSEQT" refers to the neutral carbon atoms of those side chains carrying a dipole. "FILV" refer to hydrophobic side chains; A, alanine; C cysteine. "O-108" is a C-terminal oxygen atom; "APO" is the N-acetyl group.

H-bond donors are backbone nitrogen atoms and side chains of K,R,H,N,S,Q and T. Receivers are oxygen atoms of the backbone, side chain and solvent water, and for some internal BN's none.

binding protein the acceptor, donor ratio is 1.88 (329:174). Obviously in a structure determined to only 1.85-A nominal resolution and not yet refined, there will be an uncertainty in atomic coordinates of some 0.4 A. Nonetheless, the generalizations drawn from these detailed analyses of solvent accessibility and hydrogen-bonding are certainly valid. In particular, we have assumed hydrogen bonds if the nitrogen to oxygen distance is less than 3.4 A (standard value 2.9) and if the donor, assumed proton position, acceptor angle is greater than 145°.

All of the side chain proton donors, including the weakly hydrogen-bonding cysteine sulfhydryl group, which is not counted in the 174 donors, are at the surface, except for arginine-75 (Table VII). Near many of these donors as well as hydrogen

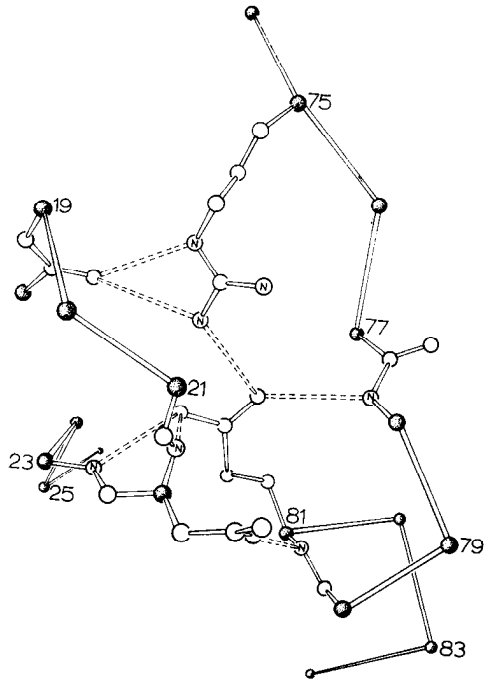

FIG. 7. Glutamic acid-81 is near the NH_2 terminus of helix E; its carboxyl group is shielded from the solvent by the AB loop. Its charge is balanced by the partially buried guanidinium group of arginine-75.

bond receivers there is electron density tentatively identified as water. A description of surface water is not included in this report. There is no water seen inside the molecule, nor are there any internal cavities large enough to accept water.

The hydrogen-bonding in the arginine-75 to glutamic acid-81 region is intricate (Fig. 7). One carboxyl oxygen receives hydrogen bonds from peptide nitrogens of aspartic acid-22 and serine-23. The other oxygen receives protons from the inside η-nitrogen of arginine-75 and from the peptide nitrogen of threonine-78. The 2nd proton of the inside η-nitrogen as well as the proton of the ϵ-nitrogen are bonded to the carbonyl oxygen of cysteine-18. Finally, the peptide nitrogen of glutamic acid-81 is hydrogen-bonded to a carboxyl oxygen of aspartic acid-22. Seven hydrogen bonds are formed in payment for bringing the carboxyl-guanidine dipole inside the protein.

In addition to hydrogen bonds of peptide nitrogens from aspartic acid-22, serine-23, and threonine-78 to the carboxyl oxygens of glutamic acid-81, there are seven peptide hydrogen bonds to side groups. In only four instances, in addition to arginine-75, do there appear to be intraprotein hydrogen bonds involving side chain donors. The hydroxyl proton of serine-23 bonds to the carbonyl oxygen of alanine-20 in the AB loop where one might have expected a β bend. The peptide proton of glutamic acid-52 bonds to the carboxyl oxygen of glutamic acid-62 which does not coordinate to calcium. Lysine-54 bonds to the carbonyl oxygen of alanine-48 and possibly to the carbonyl oxygen of aspartic acid-51 as well. Lysine-107 is in good position to bond the carbonyl oxygens of both threonine-36 and lysine-38.

Of the 108 main chain amide protons in Table VII, 48 bond to main chain carbonyl oxygen atoms. Ten bond to side chain oxygens. Twenty-nine are exposed to solvent. Twenty-one are internal and appear to have no hydrogen-bond receiver. Table V lists the principal interactions of each amino acid.

Their alignment comes from the interpretation of gene triplication (2).

Helices, Sheets, and Bends—Each of the six helices shows a significant deviation from the canonical α helix. This ideal α helix has ϕ and ψ values near $-60°$ and $-40°$. The amide nitrogen of residue n forms a linear hydrogen bond to the carbonyl oxygen of residue n-4. In a 3_{10} helix the amide nitrogen of residue n bonds back to the carbonyl oxygen n-3. For example, helix A both begins and ends with a 3_{10} configuration. As seen in Table V, "N" of "C-18" bonds to "CO" of residue 15. Helix D has only 3_{10} hydrogen-bonding, 63 to 60, 64 to 61, 66 to 63, and 68 to 65. The helix bends at leucine-65, as exemplified by $\phi = -140°$, $\psi = 48°$. At the terminus of helix E glycine-89 ($\phi = -127°$, $\psi = 5°$) twists so as to formally end the helix, but then aspartic acid-90 forms a 3_{10}-type hydrogen bond back to lysine-87, making helix E as long as the homologous helix C.

In addition to these examples in helices D and E, alanine-31 and lysine-45 in the middle of helices B and C appear not to form hydrogen bonds. As can be seen the ϕ,ψ values of many nominally helical residues differ significantly from the canonical $-60°$, $-40°$. In view of these variations in hydrogen-bonding patterns and in ϕ,ψ values, it is often arbitrary as to where one assigns the helix termini. For example, aspartic acid-51 at the end of the C helix bonds back to phenylalanine-47. The ϕ value is $-71°$ whereas ψ is 86°. In contrast cysteine-18 bonds back to leucine-15. The ϕ value is $-117°$ and ψ is 11°. Helix F has a regular course; however, its 1st residue, valine-99 is 25° away from the accepted ϕ,ψ range. The COOH-terminal alanine-108 turns away from a hydrogen-bonding direction. A glycine or alanine could be added as residue 109, thereby obliging alanine-108 to bond to alanine-104.

There are two single antiparallel β-pleated sheet configuration hydrogen bonds: isoleucine-97 to its homologue, isoleucine-58, and leucine-77 to lysine-64. In both of these regions the β-bonding pattern could be extended to three amide groups with only minor changes in the configuration of the protein.

Both the CD and the EF loops contain type I β bends (21) as indicated by the hydrogen bond at lysine-54 to -51 and at its homologue lysine-93 to -90. In the homologous region of the AB loop the amide proton of aspartic acid-22, as well as that of serine-23, are bonded to the internal carboxyl oxygen of glutamic acid-81. Although each of the three loops, pre A, BC, and DE, have one type I β bend, these bends do not occur at homologous sites. This is not surprising, since each of the loops have different environments.

Calcium Coordination—The CD calcium is coordinated by 6 oxygen atoms in an octahedral arrangement (Fig. 6). Each of these ligands in the CD loop is indicated in Table V by $\pm$X or $\pm$Y or $\pm$Z to define a local coordinate system. The ligands of the homologous EF loop can also be visualized as occupying the corners of an octahedron. The approximate 2-fold axis which relates the CD and EF regions also relates the two octahedra. However the EF octahedron is distorted in that there is no ligand in the $-$X direction; residue 98 is glycine. Nonetheless, the EF calcium ion is six-coordinate; aspartic acid-92 bonds to the calcium with both of its carboxyl oxygen atoms. The EF calcium ion appears to be accessible to the solvent in the $-$X direction; however, there does not appear to be a bound water in the electron density map. The CD calcium is not exposed to solvent.

In smaller organic and inorganic compounds calcium coordination number and geometry are variable. Calcium is coordinated by oxygen as opposed to nitrogen ligands. In $CaCl_2$·glycylglycylglycine·$3H_2O$, Van Der Helm and Willoughby (25) found calcium to be seven-coordinate with oxygen-calcium bond distances ranging from 2.296 to 2.503 A, with an average distance of 2.390 A for the 2 water, 2 carbonyl, and 3 carboxyl oxygen atoms. The oxygen-calcium bond distances in muscle calcium-binding protein range from 2.01 to 2.78 A, as indicated in Table V, with an average for the 12 distances of 2.40 A. Considering the nominal resolution of this structure determination and the fact that the structure has not yet been refined, the bonding distance of 2.0 A cannot be considered significantly different from 2.4 A. At each site there are four carboxylate groups with no lysine or arginine residue near enough to make formal electrical neutrality.

Calcium is bound by two other proteins whose structures are known. In Staphylococcal nuclease (26) "the calcium ion is coordinated by an approximately square array of carboxylate groups with the distance to aspartic acid-19 being somewhat longer than that to the others in the array." A peptide carbonyl oxygen is in the primary coordination sphere in displaced octahedral geometry. Several other oxygen atoms, including one from water, form a secondary coordination shell. In thermolysin (27) there are 4 calcium ions. Two are at a double site with calcium-calcium distance 3.8 A. "This pair of ions is surrounded by a cluster of ordered water molecules and backbone carbonyl groups" and five carboxylate groups. One of the ions is not accessible to solvent. The coordination of both "may be octahedral." "A third presumed calcium site is situated in an exposed region" near aspartic acid-57. Possibly a fourth calcium interacts with an exposed loop in the region 196 to 199.

The important point is that even though the inner coordination shell about the calcium ion tends to be octahedral, the backbone polypeptide configuration varies widely among the sites in these three proteins. It seems highly improbable that the very similar configurations of the CD and EF regions result from convergent evolution. Their structural similarity results from their homology.

In most metal-binding proteins other than muscle calcium-binding protein the metal-coordinating residues come from distant parts of the sequence: in carboxypeptidase: zinc, histidine-69, glutamic acid-72, and histidine-196, (28, 29); in insulin: zinc, histidine-B10 (trimer) (30); in carbonic anhydrase: zinc, histidine-93, histidine-95, and histidine-117 (31); in staphylococcal nuclease: calcium, aspartic acid-19, aspartic acid-21, aspartic acid-40, threonine-41, and glutamic acid-43 (26); in thermolysin: zinc, histidine-146, glutamic acid-166, calcium-calcium, aspartic acid-138, glutamic acid-177, aspartic acid-185, glutamic acid-190, and aspartic acid-191; in rubredoxin: iron, cysteine-5, -10, -38, -41; in high potential iron-sulfur protein: iron, cysteine-43, -46, -63, -77 (32). In contrast, ferredoxin coordinates 2 iron atoms with closely spaced cysteine sulfhydryl groups 8, 11, 14, 18, and 35, 38, 41, 45. The similarity to muscle calcium-binding protein is incomplete in that cysteine-18 coordinates the second iron while cysteine-41 coordinates the first. However, in ferredoxin, as in muscle calcium-binding protein, the two metal-binding regions are related by an approximate 2-fold axis.[8] It may well be that the linear proximity of the calcium coordination ligands in this protein increased the probability that a duplication in the gene coding for one section would meet with evolutionary success.

Summary of Structural Characteristics—Although denaturation

[8] L. Jensen, personal communication.

conditions for muscle calcium-binding protein have not yet been fully explored, it does appear to be a rather stable protein. The calcium form is reported to be stable in 8 M urea (7, 33). In the presence of calcium the protein is soluble up to 80 mg per ml; while in its absence it is much less soluble. We have failed to grow crystals of the calcium-free form. However, we have grown crystals of the carp protein B after adding calcium to previously EGTA-treated protein, even after it stood at room temperature several days. Konosu *et al.* (7) observed a low intrinsic viscosity of 2.3 ml per g and suggested that the protein "approximates more closely to the ideal case of the spherical protein molecule."

We have noted the apparent loss, relative to the denatured state, of 21 main chain hydrogen bonds. This inferred instability is apparently fully compensated by the formation of the large hydrocarbon core and by the coordination of the 2 calcium ions. There are from 57 to 64, depending on the definition of helix termini, residues in six helices. They too appear to be stabilizing configurations, even though each of the helices has significant deviations from the ideal α helix.

POSSIBLE FUNCTIONAL MECHANISMS

The function of muscle calcium-binding protein remains unknown. Attempts to detect any catalytic activity have been unsuccessful (5). An examination of the surface reveals no pits or grooves characteristic of the enzymes of known structure. It seems reasonable then to explore the idea that the protein functions by mediating the concentration of calcium ions in muscle.

The fact that the CD and EF calcium-binding loops are related by an intraprotein 2-fold axis suggests that one consider the possibility of cooperative binding between the two calcium sites. Benzonana *et al.* (34) studied calcium binding and state that "a closer inspection of" the Scatchard plots "suggests a downward curvature of the binding curves at low calcium concentration, an indication, generally of cooperativity between sites." If this suggestion of cooperativity is confirmed by studies at lower calcium ion concentration, one might have an explanation for the glycine at position 98. That vertex, "−X," of the EF octahedron is open to the solvent. It may well be that the EF calcium is the first to dissociate. Even though the two sites are 11.9 A apart, there is a direct link between them via the isoleucine-58 to isoleucine-97 hydrogen bonds (Fig. 6). The peptide nitrogen of valine-99 is only 4.8 A from the carbonyl oxygen of glycine-56. And the peptide nitrogen of isoleucine-97 is 4.4 A from the carbonyl oxygen of isoleucine-58. Hence the single hydrogen bond in β-antiparallel sheet configuration seen in the calcium-bound structure could easily be extended to three with a backbone movement of some 1.5 A.

There is enough sequence and structure similarity between the AB and the CD-EF regions to suggest a third homologous region in the protein (2). This interpretation demands a 2-residue deletion in the AB loop (Table V). Further, there is not a 3rd calcium ion bound in the AB loop. Instead, at the place where a calcium would be, by homology with CD or EF, is found the invariant internal salt bridge, arginine-75 to glutamic acid-81 (Fig. 7). The suggested cooperativity between the CD and EF regions could reasonably be expected to extend to this AB loop. The D helix is uniquely distorted, particularly at leucine-65. The second β-antiparallel sheet-type hydrogen bond links the peptide nitrogen of leucine-77 (near the salt bridge) to the carbonyl oxygen of lysine-64 (in the middle of the distorted D helix). The release of calcium could change the configuration of the AB loop and expose glutamic acid-81 and arginine-75 to the solvent. In this new state the calcium-binding protein would have a different affinity for another muscle protein.

REFERENCES

1. Coffee, C. J., and Bradshaw, R. A. (1973) *J. Biol. Chem.* **248,** 3305
2. Kretsinger, R. H. (1972) *Nature* **240,** 85
3. Hendrickson, W. A., and Karle, J. (1973) *J. Biol. Chem.* **248,** 3327
4. Kretsinger, R. H., Dangelat, D., and Bryan, R. F. (1971) *J. Mol. Biol.* **59,** 312
5. Pechère, J.-F., and Focant, B. (1965) *Biochem. J.* **96,** 113
6. Hamoir, G., and Konosu, S. (1965) *Biochem. J.* **96,** 85
7. Konosu, S., Hamoir, G., and Pechère, J.-F. (1965) *Biochem. J.* **96,** 98
8. Pechère, J.-F., Demaille, J., and Capony, J.-P. (1971) *Biochim. Biophys. Acta* **236,** 391
9. Gerday, C., and Bushana Rao, K. S. P. (1970) *Comp. Biochem. Physiol.* **36,** 229
10. Jebsen, J. W., and Hamoir, G. (1958) *Acta Chem. Scand.* **12,** 1851
11. Focant, B., and Pechère, J.-F. (1965) *Arch. Int. Physiol. Biochem.* **73,** 334
12. Pechère, J.-F. (1968) *Comp. Biochem. Physiol.* **24,** 289
13. Kretsinger, R. H., Nockolds, C. E., Coffee, C. J., and Bradshaw, R. A. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 217
14. Zeppezauer, M., Eklund, H., and Zeppezauer, E. S. (1968) *Arch. Biochem. Biophys.* **126,** 564
15. Xuong, N., and Freer, S. T. (1971) *Acta Crystallogr. Sect. B* **27,** 2380
16. Nockolds, C. E., and Kretsinger, R. H. (1970) *J. Sci. Instrum.* **3,** 842
17. Rae, A. D. (1965) *Acta Crystallogr. Sect. B* **19,** 683
18. Blow, D. M., and Crick, F. H. C. (1959) *Acta Crystallogr.* **12,** 794
19. Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., and Richards, F. M. (1970) *J. Biol. Chem.* **245,** 305
20. Richards, F. M. (1968) *J. Mol. Biol.* **37,** 225
21. Venkatchalam, C. M. (1968) *Biopolymers* **6,** 1425
22. Pechère, J.-F., Capony, J.-P., and Ryden, L. (1971) *Eur. J. Biochem.* **23,** 421
23. Pechère, J.-F., Capony, J.-P., Ryden, L., and Demaille, J. (1971) *Biochem. Biophys. Res. Commun.* **43,** 1106
24. Frankenne, F., Joassin, L., Closset, J., and Gerday, C. (1971) *Arch. Int. Physiol. Biochem.* **79,** 15
25. Van Der Helm, D., and Willoughby, T. V. (1969) *Acta Crystallogr. Sect. B* **25,** 2317
26. Cotton, F. A., Bier, C. J., Day, V. W., Hazen, E. E., and Larson, S. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 243
27. Matthews, B. W., Colman, P. M., Jansonius, J. N., Titani, K., Walsh, K. A., 2nd, Neurath, H. (1972) *Nature* **238,** 41
28. Lipscomb, W. N., Hartsuck, J. A., Quiocho, F. A., and Reeke, G. N. (1969) *Proc. Nat. Acad. Sci. U. S. A.* **64,** 28
29. Bradshaw, R. A., Ericsson, L. H., Walsh, K. A., and Neurath, H. (1969) *Proc. Nat. Acad. Sci. U. S. A.* **63,** 1389
30. Blundell, T. L., Cutfield, J. F., Dodson, E. J., Dodson, G. G., Hodgkin, D. C., and Mercola, D. A. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 233
31. Kannan, K. K., Liljas, A., Waara, I., Bergsten, P. C., Lovgen, S., Standberg, B., Bengtsson, J., Carlbom, U., Fridberg, K., Jarup, L., and Petef, M. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 221
32. Watenpaugh, K. D., Sieker, L. C., Herriott, J. R., and Jensen, L. H. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36,** 359
33. Piront, A., Hamoir, G., and Crokaert, R. (1968) *Arch. Int. Physiol. Biochem.* **76,** 1
34. Benzonana, G., Capony, J.-P., and Pechère, J.-F. (1972) *Biochim. Biophys. Acta* **278,** 110
35. Diamond, R. (1966) *Acta Crystallogr. Sect. B* **21,** 253
36. Matthews, B. W. (1966) *Acta Crystallogr. Sect. B* **20,** 230