Structure of Concanavalin A at 2.4-Å Resolution†

Karl D. Hardman* and Clinton F. Ainsworth

ABSTRACT: An electron density map of concanavalin A at 2.4-Å resolution has been produced by X-ray crystallographic methods with five heavy atom derivatives. The molecule is a tetramer with all subunits identical and each containing a single polypeptide chain of 231 amino acid residues. The course of the entire backbone has been traced and three different regions of β structure involve about 57% of the residues. One of these β-structure regions contains six strands of polypeptide chain and is related by a crystallographic twofold rotation axis to an additional six strands of the second subunit. These two regions are linked across the twofold axis by hydrogen bonds to form a continuous 12-strand β-pleated sheet. Concanavalin A contains no helix. The Mn^{2+} and Ca^{2+} ions are hexa- and pentacoordinated, respectively, are 4.3 Å apart, and occupy what is described as a double site, since two aspartyl residues are common ligands to both ions. The symmetry of the ligands bonded to the Mn^{2+} is very nearly octahedral. The double ion site is 23 Å from the carbohydrate binding site, which is a distinct cavity and contains one tyrosyl and perhaps two aspartyl residues.

Concanavalin A (Con A[1]) exhibits some unusual biological properties as a result of its ability to bind various carbohydrates. It agglutinates erythrocytes from certain animal species, starch granules, and some bacteria and yeasts (Sumner and Howell, 1936a), and has been shown to precipitate various glycogens, dextrans, mannans, glycoproteins, and blood group substances (Sumner and Howell, 1936b; So and Goldstein, 1968; Leon and Young, 1970; and Lloyd *et al.*, 1969). Carbohydrates with the minimum specificity for Con A binding contain hexose residues with the D-*arabino*-pyranoside configuration[2] at C-3, C-4, and C-5 (Goldstein *et al.*, 1965). Con A induces transformation of lymphocytes by reversibly binding to specific sites on the cell surface (Powell and Leon, 1970; Novogrodsky and Katchalski, 1971) and inhibits phagocytosis by polymorphonuclear leucocytes (Berlin, 1972). It also agglutinates embryonic tissue cells (Moscona, 1971) and various neoplastic cells in tissue cultures (Inbar and Sachs, 1969), whereas binding sites on the corresponding adult cells and normal cells appear to be masked. Studies by Inbar *et al.* (1971) indicate that the site for Con A on the cell surface membrane of hamster cells has two components, one which actually binds the Con A and the other which is responsible for the agglutination.

The purpose of our investigation is to determine the complete three-dimensional structure of Con A by X-ray crystallographic techniques and we report here features of the molecule, mainly the course of the polypeptide chain, subunit interactions, regions of β structure, the carbohydrate binding site, and the Mn^{2+} and Ca^{2+} sites as interpreted from our electron density map at 2.4-Å resolution.

Experimental Procedure

Con A, purified and crystallized as previously described (Hardman *et al.*, 1971a), was used to collect the lower resolution data. Crystals for higher resolution data were grown with protein whose preparation was modified as follows. After extraction from the bean meal and adsorption to Sephadex, the column was rinsed extensively with a buffer which was 0.01 M Tris–0.35 M NaCl, pH 7.4, and eluted with 10 mM methyl α-D-glucopyranoside in the Tris–NaCl buffer. The eluted protein was pressure dialyzed against the Tris–NaCl buffer, both to remove the glucopyranoside and to increase the protein concentration.

Some crystals were grown by dialysis in microdiffusion tubes inverted under a 2 M phosphate solution, pH 5.9. In the latter case, the protein was originally dissolved in an unbuffered 0.35 M NaCl solution, pH 5.9. The crystals were harvested after 11 days, at which time the mother liquor was about 1.2 M phosphate.

Heavy Atom Derivatives. The five heavy atom derivatives which were used for the calculation of the electron density map were all prepared by soaking native Con A crystals. The preparations of the $PtCl_4^{2-}$ and $PtCl_6^{2-}$ derivatives have been previously described (Hardman *et al.*, 1971a). The ClHgPhenol derivatives were obtained by soaking native crystals first in a 1 mM solution of the heavy metal for 8 days and then in a 4 mM solution for 19 days. The MeHgCl derivatives were prepared by soaking crystals for 1 or 2 days in a saturated solution of the heavy atom compound and, in order to prevent loss of the volatile MeHgCl during data collection, some of the saturated solution was added inside the quartz capillary in which the crystal was mounted. The $Hg(AcO)_2$ derivatives were made by 6–9 week soaks in 0.2 mM solutions. All heavy metal solutions were also 2 M in phosphate, pH 6.0.

Data Collection and Reduction. For calculation of the electron density map, three complete sets of native data (about 10,000 unique reflections per set, covering one octant of the sphere of reflection), two sets for the $PtCl_4^{2-}$ derivative and one set each for the other four derivatives, were collected to a resolution of 2.4 Å, except for the $PtCl_6^{2-}$ and ClHgPhenol derivatives which were collected to 2.58 and 3.25 Å, respectively. The total number of reflections used was approximately 90,000 which required nine native crystals, four $PtCl_4^{2-}$ crystals, and two crystals for each of the remaining derivatives. The method of data collection, correction, and reduction has previously been described (Hardman *et al.*, 1971a), except that anomalous scattering data were not col-

† From the Division of Biological and Medical Research, Argonne National Laboratory, Argonne, Illinois 60439. *Received August 7, 1972.* This work was supported by the U. S. Atomic Energy Commission.

[1] Abbreviations used are: Con A, concanavalin A; ClHgPhenol, *o*-chloromercuriphenol; MeHgCl, methylmercuric chloride; *F*, structure factor amplitudes.

[2] See Nomenclature of Carbohydrates (1971), *Biochemistry 10*, 3983.

lected beyond 6.0 Å. Data were collected in radial shells of 2θ with approximately 1000 reflections per shell, and the scaling of different crystals to form a complete data set was done by collection of an initial 6-Å data set on each crystal or by overlapping shells.

Determination and Refinement of Heavy Atom Positions and Phase Refinement. The heavy atom positions for the first two derivatives, $PtCl_4^{2-}$ and $PtCl_6^{2-}$, were located by examination of the Harker sections of three-dimensional difference Patterson maps as described previously (Hardman *et al.*, 1971a), followed by initial refinement with a modified ORFLS program (Busing *et al.*, 1962). Subsequent refinements of the heavy atom parameters and the protein phases were calculated by the method of alternate cycles of phase determinations and least-squares refinement with the Dickerson program (Dickerson *et al.*, 1968).

Protein phases determined from all available heavy atom data were used to locate the heavy atom positions for each new derivative. Initial least-squares refinement was done on each derivative separately before combining these data with previous derivatives for the next series of protein phase calculations. Preliminary protein phase calculations with the five heavy atom derivatives were made with 6.0- and 3.25-Å data. After all data to 2.4 Å were incorporated, six runs, with two refinement cycles each, were made.

Preparation of the Electron Density Map and Model Building. The 2.4-Å electron density map was calculated by the Fourier program of the X-ray 70 system of Stewart (1970). Electron densities were printed with Y and Z grid intervals of 0.416 and 0.50 Å, respectively (with a scale of 2.0 cm/Å) and were hand contoured directly on polyvinyl chloride sheets. The interval in the X coordinate (between sections) was 0.70 Å. Contours were drawn with intervals of 0.31 e/Å^3, beginning at the level 0.31 e/Å^3 above the average electron density for the crystal (0.40 e/Å^3). The electron density scale was set by averaging the integrated electron density peaks for Trp-40 and -109 and subtracting the average electron density of the crystal.

A skeletal model of Con A was built with the Kendrew type brass models (2 cm/Å) by fitting the model atoms into the electron density contours with an optical comparator (Richards, 1968). The optical comparator was modified upon the suggestion of Matthews *et al.* (1972b) so the mirror is parallel to the contoured sheets instead of at a 45° angle.

The amino acid sequences of four peptides have been published: the N-terminal peptide, Ala-Asp-Thr-Ile-Val-Ala-Val-Glu-Ile-Asp-Thr-Tyr-Pro (Edmundson *et al.*, 1971); the two Met peptides, Trp-Asp-Met-Gln-Asn-Gly-Lys (Dyckes, 1970; Waxdal *et al.*, 1971) and Met-Phe-Asx-Glx-Phe-Ser-Lys (Waxdal *et al.*, 1971); and the C-terminal peptide, Leu-Leu-Gly-Leu-Phe-Pro-Asp-Ala-Asn (Dyckes, 1970). With the exception of these residues, all other side chains have been tentatively identified by the best fit to the electron density map.

Results

Data Collection and Heavy Atom Refinement. The native Con A crystals are orthorhombic with unit cell dimensions of 63.15 ± 0.024, 86.91 ± 0.053, and 89.25 ± 0.042 Å for a, b, and c, respectively. The space group is $I222$, with eight asymmetric units per unit cell. The phenylmercuric acetate derivative, which was used for calculation of previous maps (Hardman *et al.*, 1971a), was not continued to higher resolution because of comparatively low occupancy and poor refinement. The ClHgPhenol derivative was not continued beyond 3.25 Å because the $Hg(AcO)_2$ derivative had very similar heavy atom sites and the data refined more satisfactorily. Final refinement parameters for all heavy atom derivatives used for calculation of the 2.4-Å electron density map are shown in Table I and the radial distribution of some of the parameters throughout reciprocal space where data were collected is shown in Table II. For the four main derivatives (all but ClHgPhenol) a total of 17 heavy atom sites were found, 12 of which were unique. No attempt was made to check the difference Patterson maps for the derivatives with five and six sites. The coordinates, occupancies, and temperature factors for all sites are shown in Table III. The determination of the correct enantiomorph of the protein molecule as done previously (Hardman *et al.*, 1971a) was confirmed by repeating the 6-Å protein phase calculations and including the additional heavy atom derivatives (MeHgCl, $Hg(OAc)_2$, ClHgPhenol). The axes and hkl indexes as originally chosen produced the wrong enantiomorph, so for subsequent calculations, h was set equal to $-h$ to make the correction.

TABLE I: Heavy Atom Refinement Parameters.

	Pt-Cl_4^{2-}	Pt-Cl_6^{2-}	MeHgCl	Hg-$(AcO)_2$	ClHg-Phenol[g]
R_K (all data)[a]	0.065	0.077	0.140	0.172	0.123
R_K (centric)	0.090	0.115	0.199	0.240	0.184
R (all data)[b]	0.537	0.524	0.599	0.623	0.633
R (centric)	0.460	0.476	0.616	0.667	0.654
K scale[c]	1.017	1.023	1.035	1.065	0.984
ΔB scale[d]	0.217	0.420	0.182	0.726	1.041
RMSE[e]	39.6	55.0	97.1	102.1	104.8
$\langle f_H \rangle$[f]	65.2	85.1	125.4	108.7	108.3
Figure of merit	all data = 0.63, centric = 0.648				

[a] $R_K = \Sigma|F_{PH(obsd)} - F_{PH(calcd)}|/\Sigma F_{PH(obsd)}$, the Kraut R factor (Kraut *et al.*, 1962) where F = structure factor amplitudes. The subscripts P, PH, H, (obsd), and (calcd) refer to terms for protein, protein plus heavy atom, heavy atom, observed, and calculated, respectively. [b] $R = \Sigma|F_{PH(obsd)} - F_{PH(calcd)}|/\Sigma|F_{PH(obsd)} - F_P|$, which for centric case is identical with R_{Cullis}. [c] K scale = scale factors between native and derivative data sets. [d] ΔB scale = difference temperature factor between native and derivative data sets. [e] RMSE = root-mean-square error of closure, in electrons. [f] $\langle f_H \rangle$ = root-mean-square heavy atom scattering, in electrons. [g] Data to 3.25 Å only.

Maximum absorption and decay corrections were most commonly between 15–30 and 10–15%, respectively. Agreement residuals, $R_F = \Sigma|F_{PH} - F_P|/\Sigma F_P$, for different native data sets ranged from 0.030 to 0.044 for data sets below 4.25 Å and from 0.042 to 0.049 for data sets between 4.25 and 2.4 Å.

Electron Density Map and Atomic Model. The contents of one asymmetric unit of the electron density map have been traced as a single polypeptide chain with 231 amino acid residues. Calculation of the molecular weight of the protein contained in one asymmetric unit, after adjusting the amino acid composition (Waxdal *et al.*, 1971; Edmundson *et al.*, 1971) for the best fit of a total of 231 residues, gives a value of 25,200. From this molecular weight, the unit cell volume, and the partial specific volume (0.73 cm^3/g, Hardman *et al.*, 1971a), the crystal volume contains 50% protein. The coordinates for the α-carbon positions, the Mn^{2+} and Ca^{2+} sites, and the

TABLE II: Radial Distribution of Refinement Parameters.

Av d spacing	Av sin θ	hkl No.	Av F_p	Figure of Merit	R (All Data) $PtCl_4^{2-}$	$PtCl_6^{2-}$	MeHgCl	$Hg(AcO)_2$	ClHgPhenol
10.1	0.0761	342	518	0.893	0.296	0.355	0.513	0.443	0.470
6.25	0.1234	582	388	0.835	0.330	0.316	0.477	0.438	0.524
4.86	0.1587	729	530	0.814	0.496	0.413	0.511	0.536	0.645
4.11	0.1875	836	474	0.760	0.572	0.544	0.667	0.627	0.711
3.63	0.2123	952	431	0.704	0.610	0.622	0.648	0.661	0.732
3.29	0.2346	1041	366	0.694	0.627	0.623	0.606	0.644	0.707
3.02	0.2549	1128	305	0.633	0.609	0.589	0.622	0.674	
2.81	0.2739	1228	248	0.599	0.642	0.622	0.609	0.693	
2.64	0.2917	1282	215	0.522	0.662	0.682	0.659	0.678	
2.48	0.3110	1828	190	0.407	0.708		0.643	0.729	

TABLE III: Heavy Atom Parameters.

Derivative	Site	Fractional Unit-Cell Coordinates X	Y	Z	A^a	B^b
$PtCl_4^{2-}$	1	0.0479	0.0055	0.0578	61.1	5.4
	2	0.0538	0.0077	0.1215	22.8	5.0
$PtCl_6^{2-}$	1	0.0496	0.0050	0.0576	78.0	5.8
	2	0.0549	0.0076	0.1226	23.4	5.1
	3	0.1550	0.2171	0.3142	24.8	6.8
MeHgCl	1	0.0535	−0.0020	0.0550	54.6	6.0
	2	0.2471	0.2852	0.4045	71.9	2.9
	3	0.3518	0.2830	0.3939	52.8	3.9
	4	0.0267	0.0071	0.1615	56.6	52.0
	5	0.3330	0.3165	0.4530	33.2	5.8
	6	0.2283	0.3712	0.3751	21.7	3.1
$Hg(AcO)_2$	1	0.0706	−0.0066	0.0179	35.9	4.2
	2	0.0452	0.0085	0.0585	46.8	4.3
	3	0.0249	0.0026	0.1460	38.7	22.5
	4	0.3280	0.2350	0.3949	67.8	3.6
	5	0.1161	0.3726	0.2572	35.7	3.5
	6	0.2811	0.2648	0.3390	31.0	1.7
ClHgPhenol	1	0.0417	0.0095	0.0574	61.1	8.7
	2	0.0636	−0.0059	0.0122	29.4	4.1
	3	0.0159	0.0071	0.1495	53.4	43.0
	4	0.3276	0.2358	0.3938	61.2	4.0
	5	0.1160	0.3727	0.2576	23.0	3.4

[a] A, occupancy, in electrons. [b] B, temperature factor.

inositol binding site as read from the skeletal model are given in Table IV and are plotted in Figure 1 by the ORTEP computer program (Johnson, 1965).

Beginning with the N-terminal Ala (Figure 1, front), which is well defined in the electron density map, the polypeptide chain leads through a strand[3] of sheet II (residues 3–11, see Figure 2) to the Mn^{2+} and Ca^{2+} sites (residues 8 and 10) where the chain turns out to the surface of the molecule. It loops into the solvent and back around the Mn^{2+} and Ca^{2+} (residues 14–24) and returns through sheet II (residues 23–30). After forming another loop which includes a Mn^{2+} ligand (Glx-34), another strand of sheet II (35–39) leads to Trp-40 and Met-42. A turn at the surface (43–45) continues into three strands (48–78) of sheet I. The chain forms a small loop (81–85), which contains *one* hydrogen bond of the α-helix type, and continues across the surface of the subunit (85–89) near the carbohydrate binding site (Hardman and Ainsworth, 1972). From there the chain goes through the center of the subunit as a strand of sheet II (90–97). It emerges on the surface to the right (toward positive Z), turns abruptly, and continues through sheet III (residues 103–106). Next are the two strands of sheet I nearest the twofold rotation axis (107–125), which include Met-122. Several loops (126–131, 143–145, and 151–160) are joined with two strands of sheet III (139–142 and 146–149) and two strands of sheet II (132–137 and 162–168), to form the upper right portion of the subunit. After a bend near the twofold rotation axis (168–169), the chain continues past the carbohydrate binding site (170–180) and loops down to form the last strand (182–191) in sheet I. After a slight bend the chain continues through a short strand of sheet III (191–194) on the right surface, forms a loop (195–201), and goes into sheet II (202–209). It bends sharply at the surface (left front, residues 210–211) and forms an extended section of chain across the extreme front of the subunit (212–220). At residue 221 the chain loops back immediately underneath itself into a hydrophobic region (222–227) and then emerges at the C terminus.

The three separate sheets of β structure are shown in Figure 2. Sheet I consists of six chains (62 residues) and forms the "back" of the subunit. The center of the subunit is formed by the seven strands of sheet II (51 residues). This region begins with residues 35–39, near the Mn^{2+} ion, and continues through the center to the front of the subunit, residues 132–137. Sheet III involves a total of 18 residues and contains five strands with an average of four residues each. The lower three strands form slight bends and continue into sheet I. The front strand of sheet III (residues 139–143) follows a sharp bend in the top section of sheet II (residues 132–137). The β structure of these three sheets involves about 131 residues, or 57% of the total residues in the polypeptide chain.

The portion of the map shown in Figure 3 is representative of the overall quality of the map. This region contains portions of the six strands of sheet I. The carbonyl oxygens are readily

[3] The term "strand" is used to refer to one section of extended polypeptide chain in a β-pleated sheet region and "sheet" refers to a single, continuous β-pleated sheet region, *i.e.*, of more than one strand.

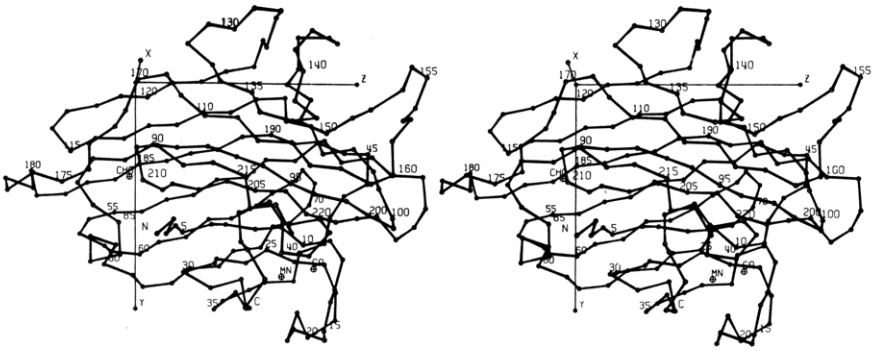

FIGURE 1: Stereogram of the α-carbon positions of the 231 amino acid residues in the Con A monomer. The positions of the Mn^{2+}, Ca^{2+} carbohydrate binding site, and the amino and carboxyl termini are labeled MN, CA, CHO, N, and C, respectively. The axes for all figures are of the same orientation, except Figure 5.

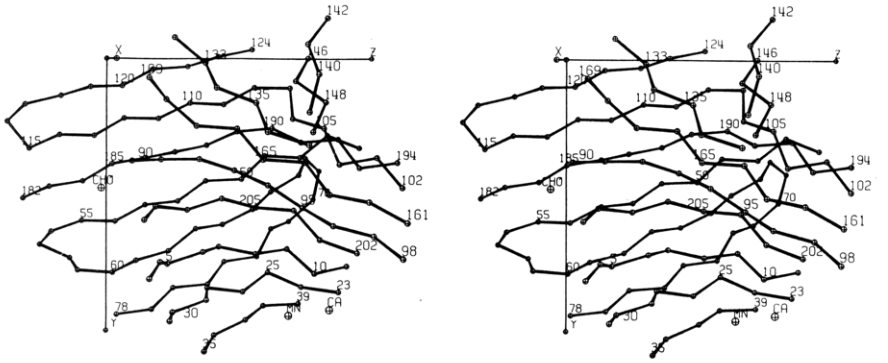

FIGURE 2: Stereogram of the α-carbon positions of the residues involved in the β structure. Three regions, sheets I, II, and III, contain six seven, and five strands, respectively, and account for 57% of the total residues.

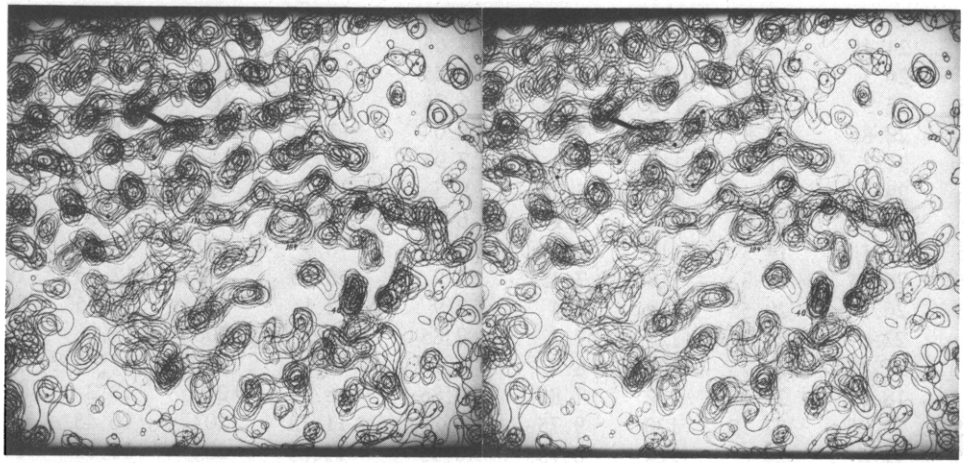

FIGURE 3: A portion of the 2.4-Å electron density map. This stereogram contains twelve serial sections, X = 0.044–0.167. Examples of "bumps" in the polypeptide chain indicating carbonyl oxygen positions are marked by solid circles in the two strands of β-structure sheet I nearest the X axis (column of dyad symbols toward upper-left corner). Examples of notable side chains are Trp-40 (ring nearly parallel to the X axis) and Phe-111 and -189 (rings nearly perpendicular to the X axis).

visible on the sides of the chains and in many cases hydrogen bonds between the carbonyl oxygens and amide hydrogens are indicated by continuous electron density contours. Many of the side chains were easily identifiable, with positive electron density contours enveloping all atoms. This includes many side chains on the surface of the molecule as well as on the interior. Examples of prominant side chains shown in Figure 3 are Trp-40 and Phe-111 and -189. The solvent boundaries are clearly visible. A *true* crystallographic twofold rotation axis (X axis) appears in Figure 3 around which two identical sections of chain interact to join the sheet I regions of two related subunits (see Discussion).

Discussion

Association of Subunits. From the densities of the crystals and solvent and the partial specific volume, the protein contents of the asymmetric unit have been previously calculated

TABLE IV: Concanavalin A α-Carbon Coordinates.

Residue	Coordinates (Å)			Residue	Coordinates (Å)			Residue	Coordinates (Å)		
	X	Y	Z		X	Y	Z		X	Y	Z
1	29.3	20.0	1.55	81	12.35	21.4	−2.75	161	27.2	14.4	25.65
2	26.5	18.35	3.55	82	14.35	20.35	−5.85	162	26.4	12.65	22.4
3	23.1	20.0	2.9	83	16.2	23.35	−5.85	163	24.55	12.2	19.25
4	19.7	19.05	4.2	84	17.4	22.5	−2.3	164	22.5	9.45	17.45
5	16.2	19.85	5.25	85	16.65	18.65	−2.45	165	22.0	9.25	13.6
6	15.3	18.85	8.85	86	19.05	16.7	−5.5	166	20.45	6.5	11.4
7	12.05	18.85	10.9	87	17.3	14.0	−7.35	167	19.95	6.3	7.6
8	13.3	19.4	14.45	88	19.4	11.3	−5.5	168	18.8	3.85	4.85
9	11.8	19.15	17.95	89	18.95	11.4	−1.75	169	21.23	1.75	3.0
10	13.7	21.35	20.35	90	19.15	9.6	1.55	170	19.3	2.1	−0.3
11	12.7	20.8	23.9	91	16.7	9.7	4.55	171	19.0	5.65	−1.7
12	15.55	23.05	24.9	92	17.4	9.6	8.2	172	15.7	7.2	−2.9
13	15.05	26.7	24.0	93	16.65	10.75	11.75	173	14.75	9.2	−6.0
14	18.5	28.0	22.8	94	18.85	11.95	14.65	174	12.25	12.2	−6.05
15	17.6	31.55	22.2	95	19.3	14.05	17.9	175	13.5	13.75	−9.35
16	21.4	32.25	21.55	96	22.25	15.35	20.15	176	13.25	12.65	−12.9
17	21.25	31.0	17.9	97	22.95	16.2	23.8	177	12.75	14.75	−16.05
18	17.65	32.6	17.75	98	25.8	17.85	25.7	178	9.35	13.15	−16.55
19	16.35	29.4	19.15	99	24.75	19.3	28.85	179	7.95	14.35	−13.15
20	12.9	32.65	21.05	100	21.25	18.15	29.6	180	4.9	12.0	−13.6
21	12.25	30.65	24.6	101	20.0	14.65	28.25	181	2.95	15.15	−13.0
22	10.45	27.3	25.05	102	17.05	12.45	28.3	182	3.7	14.95	−9.2
23	10.8	23.8	23.55	103	15.35	9.65	26.25	183	5.55	13.5	−6.34
24	9.4	23.5	19.95	104	15.65	10.3	22.4	184	4.5	13.05	−2.75
25	9.25	22.0	16.45	105	14.2	6.85	21.2	185	5.6	11.1	0.35
26	10.95	23.65	13.55	106	11.85	6.05	18.5	186	4.45	10.6	4.0
27	11.6	23.1	9.8	107	9.6	3.0	18.7	187	6.9	9.75	6.95
28	15.4	23.55	9.2	108	7.9	3.0	15.25	188	5.65	9.15	10.5
29	16.6	24.45	5.7	109	8.9	4.75	11.85	189	7.25	7.6	13.55
30	20.4	24.45	5.05	110	7.35	4.65	8.5	190	6.75	7.3	17.35
31	19.9	26.3	8.3	111	7.8	6.25	5.1	191	9.15	8.6	19.95
32	19.25	25.3	12.0	112	6.75	6.25	1.55	192	9.75	8.25	23.75
33	17.65	28.85	12.8	113	7.8	8.0	−1.65	193	11.45	11.05	25.55
34	14.25	27.1	11.9	114	7.4	7.95	−5.3	194	12.1	10.45	29.3
35	12.35	29.3	9.4	115	5.75	9.5	−8.5	195	14.45	12.8	31.15
36	9.05	28.45	10.85	116	6.1	6.65	−10.85	196	15.1	13.0	34.9
37	8.25	27.1	14.3	117	8.65	4.75	−8.9	197	15.8	16.85	35.05
38	5.35	26.25	16.7	118	7.45	3.9	−5.15	198	12.75	19.1	34.45
39	5.25	26.1	20.55	119	9.4	2.9	−2.1	199	14.75	20.65	31.65
40	5.7	22.4	21.2	120	7.7	2.85	1.3	200	15.4	17.8	29.1
41	4.5	21.25	24.75	121	8.6	1.1	4.45	201	17.75	18.1	26.15
42	6.5	18.0	25.05	122	7.85	0.85	8.2	202	18.2	18.55	23.5
43	5.35	15.05	27.15	123	9.75	−0.2	11.15	203	16.45	17.55	20.25
44	8.05	13.6	29.6	124	8.75	−0.95	14.85	204	15.35	14.8	17.9
45	8.0	9.75	29.7	125	11.1	−3.85	15.65	205	13.95	14.7	14.35
46	5.1	9.5	27.25	126	14.6	−2.45	15.6	206	15.1	13.65	10.75
47	5.3	8.15	23.7	127	16.5	−5.65	16.5	207	15.6	14.7	7.25
48	5.05	10.5	20.75	128	20.25	−5.8	16.5	208	17.55	14.1	3.95
49	4.75	10.35	16.95	129	22.25	−5.7	13.1	209	20.75	14.95	2.6
50	5.75	12.75	14.25	130	19.7	−3.7	11.05	210	23.45	13.95	0.2
51	4.9	13.0	10.5	131	21.95	−3.85	7.8	211	23.15	10.2	−0.5
52	5.7	15.05	7.4	132	19.55	−2.0	5.55	212	26.4	10.2	1.65
53	4.3	15.6	4.0	133	19.55	0.4	8.5	213	25.4	12.35	4.7
54	5.45	17.2	0.7	134	22.5	2.65	9.15	214	27.45	12.15	7.85
55	4.1	17.3	−2.85	135	23.35	4.0	12.7	215	27.5	13.65	11.45
56	5.7	18.2	−6.15	136	26.1	6.45	13.25	216	28.6	17.2	11.85
57	2.75	20.05	−7.4	137	27.2	7.5	16.75	217	31.45	16.85	14.3
58	1.9	21.65	−4.1	138	25.4	6.95	19.85	218	30.45	19.9	16.15
59	5.5	22.35	−3.3	139	23.2	4.85	17.7	219	26.75	18.95	16.45
60	5.0	22.7	0.45	140	22.3	1.4	18.8	220	26.35	18.3	20.05

TABLE IV (*Continued*)

Residue	Coordinates (Å)			Residue	Coordinates (Å)			Residue	Coordinates (Å)		
	X	*Y*	*Z*		*X*	*Y*	*Z*		*X*	*Y*	*Z*
61	6.9	21.1	2.85	141	19.95	−1.4	18.25	221	23.2	19.8	21.8
62	4.8	20.45	6.5	142	18.35	−3.9	20.55	222	21.6	22.55	19.65
63	4.75	18.1	9.45	143	16.9	−2.6	23.9	223	23.2	22.1	16.2
64	2.8	17.95	12.65	144	13.2	−3.9	23.6	224	21.6	18.7	16.6
65	3.65	16.3	15.95	145	13.05	−3.9	19.75	225	23.85	16.35	14.7
66	1.2	14.6	18.3	146	13.6	−0.05	19.7	226	22.4	17.55	11.35
67	0.5	12.85	21.6	147	16.15	2.15	17.9	227	24.15	20.9	11.35
68	−2.6	11.25	23.05	148	18.1	4.15	20.45	228	27.9	21.75	11.7
69	−5.2	12.95	25.25	149	20.05	6.85	18.75	229	28.7	24.95	13.5
70	−3.25	16.0	24.1	150	21.3	8.3	21.75	230	28.0	28.1	11.4
71	−3.55	18.55	21.3	151	23.75	6.75	24.3	231	24.15	28.4	12.35
72	−0.8	18.95	18.7	152	21.6	5.6	26.9	Mn	15.25	25.15	17.4
73	0.9	21.8	16.8	153	23.25	3.5	30.25	Ca	17.5	24.15	20.95
74	1.35	22.25	13.05	154	20.95	0.6	31.9	CHO	11.5	13.0	−1.05
75	3.8	24.25	10.9	155	18.35	1.45	34.45				
76	4.3	24.5	7.05	156	17.55	4.95	33.25				
77	6.4	26.35	4.55	157	20.15	6.7	31.1				
78	6.55	26.8	0.8	158	24.1	6.9	30.95				
79	9.75	25.2	−0.6	159	25.15	9.45	28.55				
80	10.0	23.85	−4.2	160	24.9	13.3	28.75				

to be 26,000–29,000 (Greer *et al.*, 1970; Hardman *et al.*, 1971a) and, considering the possible errors in these measurements, are in good agreement with the 25,200 reported here. Molecular weight estimates for Con A by ultracentrifugation have ranged from 50,000 to 100,000 (Sumner *et al.*, 1938; Agrawal and Goldstein, 1967; Kalb and Lustig, 1968). It has been concluded from electron density maps at lower resolution that at pH values below 6.0, two monomers interact across the *X* axis to form a dimer of mol wt 50,000 (Figure 4), and, at higher pH values, four monomers interact through a point of 222 symmetry to form pseudotetrahedral clusters (Hardman *et al.*, 1971a,b).

Different sets of interactions between subunits occur about each of the crystallographic axes. The most prominent set is around the *X* axis and is shown in Figures 4 and 5. The strand in sheet I nearest the *X* axis (residues 117–124, Figure 5) is hydrogen bonded to its symmetry-related mate in the typical antiparallel manner; hence, sheet I is multiplied by the twofold rotation axis into a continuous β-structure region of 12 chains. The other interactions which are involved in stabilizing the dimer are found just above the peptide section 117–124. Symmetry-related regions near residues Phe-168 and Tyr-169 interact across the *X* axis and section 125–132 of one monomer interacts with section 170–176 of the second (Figure 4).

A situation somewhat similar to that shown in Figure 5 occurs in the crystal structure of insulin where portions of the two different B chains within the dimer are hydrogen bonded together in a pleated sheet conformation around a twofold axis (Blundell *et al.*, 1971). In the insulin case, however, the twofold axis is noncrystallographic and the symmetry is only approximate. It is interesting to note that two Phe residues are present in each chain in both cases.

Association of the dimer around the *Y* and *Z* axes to another dimer produces the tetramer of four identical subunits. The β-structure sheets I of these four subunits are shown in Figure 6. The pronounced twist of the 12 strands of chain in each dimer is such that the two dimers are slightly wrapped around each other. If the dimer–tetramer association in solution does indeed involve these β-structure regions as indicated by the crystalline structure, then in the dimer the side chains between these pleated sheets must be on the surface, accessible to solvent, whereas in the tetramer they must be buried. Lys has been assigned to the residues on either side of the crystallographic *X* axis because of the excellent fit of the contours, although other possibilities are not excluded. This assignment is certainly acceptable for the dimer (pH below 6.0) where these two adjacent side chains are on the surface; however, it would be unusual to have four Lys side chains buried as would be necessary for the tetramer. This could be conceivable if the p*K* of this Lys were unusually low and an anion such as phosphate were also involved. Most of the other side chains between these β-structure regions are of intermediate size and appear to be Ser, Thr, or Asn, which could easily be exposed to solvent or hydrogen bonded on the interior of the tetramer.

Regions of contact between subunits other than those involved in the formation of the tetramer are toward the right front side of the subunit near residues 16, 222, and 198. The contacts between subunits in these regions are clearly fewer in number than those within the tetramer. The packing of the tetramers (mol wt ~100,000) within the crystalline lattice has been discussed previously (Hardman *et al.*, 1971a,b; Quiocho *et al.*, 1971; Reeke *et al.*, 1971).

The Polypeptide Chain. Studies of purification of Con A for the determination of the amino acid sequence have shown that various preparations contain a single polypeptide chain with a molecular weight near 26,000 and three fragments of this chain (Wang *et al.*, 1971; Edmundson *et al.*, 1971). The intact chain was obtained in 60–75% yields and the remainder was isolated as various fragments produced by splits in the

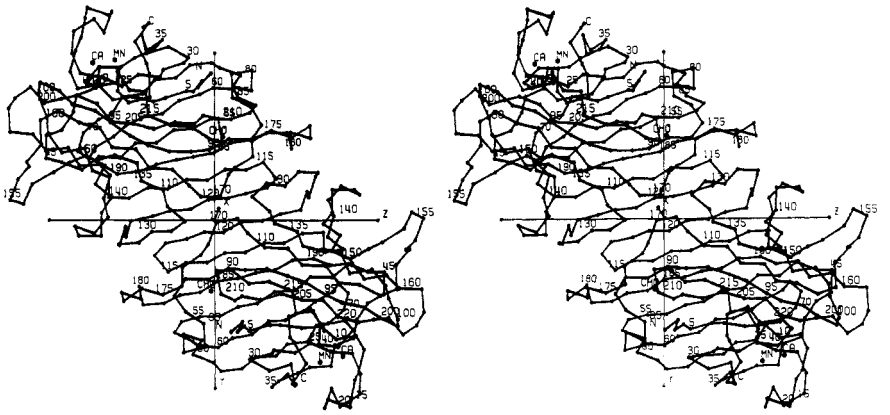

FIGURE 4: Stereogram of the α-carbon positions of two monomers bonded together across the X axis. The sheet I β-structure regions of these two subunits form a continuous 12-strand pleated sheet.

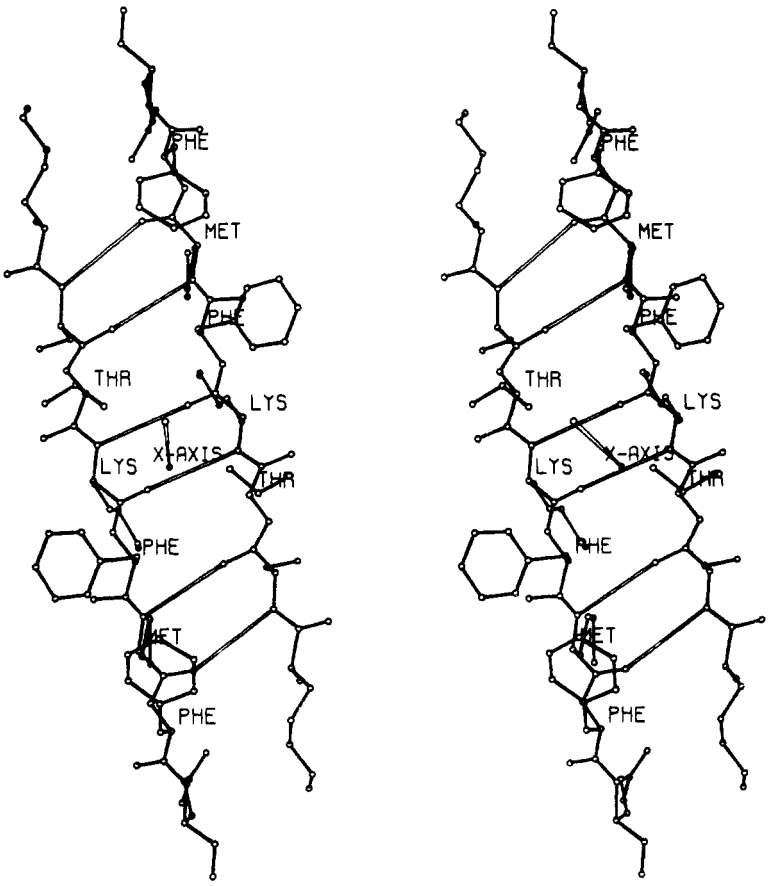

FIGURE 5: Stereogram of the strand of β-structure sheet I nearest the X axis and its symmetry-related equivalent. The X axis is a *true* crystallographic twofold rotation axis. Note this figure is rotated 90° around the X axis so that the Y axis is horizontal.

original chain between residues 107 and 130 (Wang *et al.*, 1971; Edmundson *et al.*, 1971). The preparation of protein used for growing crystals has been modified (see Experimental Procedure) to obtain preparations with the highest possible yields of intact chain. In the 2.4-Å electron density map, no discontinuities occur in the polypeptide chain in this general region (residues 100–135); thus the subunit contained in one asymmetric unit is a single polypeptide chain. Two points exist, however, where the electron density of the backbone drops below the average electron density. One of these is where the chain loops out in solution (between residues 155 and 156) where low electron density might be expected because the chain is not constrained. The other is in the center of sheet II (between residues 205 and 206); however, the general course that the chain must follow is clear.

The amino acid sequence information which is available was used to confirm the identification of the N terminus and the C terminus. The electron density contours for the N-terminal region fit the first 13 residues of the sequence (Edmundson *et al.*, 1971), including the N-terminal Ala. Nine residues from the C terminus, the chain turns abruptly underneath itself (Figure 1, front) and every atom in the hydrophobic residues

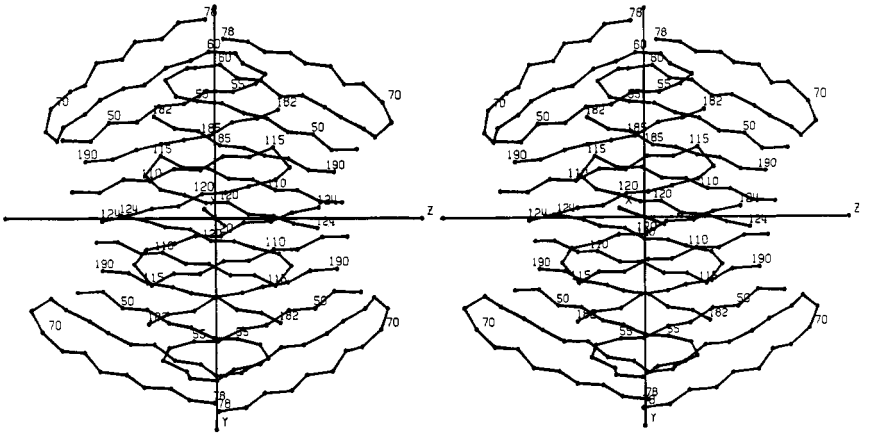

FIGURE 6: Stereogram of the four sheet I regions of the tetramer.

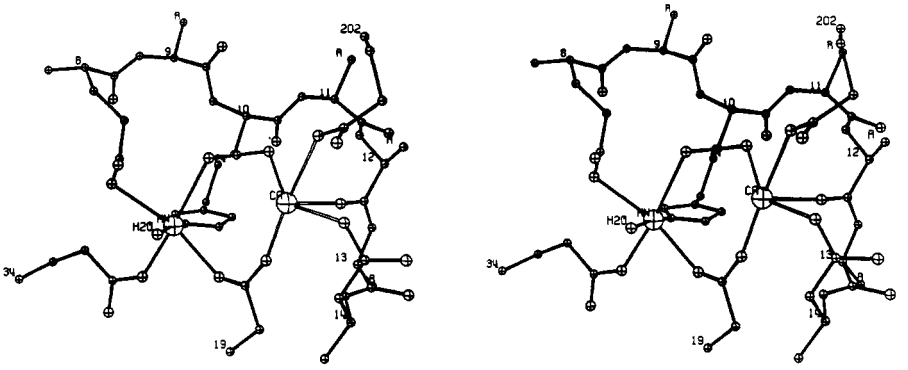

FIGURE 7: Stereogram of the Mn^{2+} and Ca^{2+} double ion site. The amino acid residues are labeled at the α-carbon position. The Mn^{2+} and Ca^{2+} ions are hexa- and pentacoordinated, respectively, and both carboxyl groups of residues 10 and 19 donate oxygen ligands to both ions.

in this sequence fits within positive electron density contours. Electrostatic interactions between the carboxyl groups of the C-terminal Asp and Arg-33 stabilize this terminus. The total amino acid residues in the polypeptide chain as counted from the electron density map, 231, is expected to be within about two residues since both ends are well defined and there are two loops projecting into solvent regions where the chain might be shortened or lengthened by one residue.

The positions of the two Met residues were first indicated by the heavy atom sites for the Pt derivatives (Dickerson *et al.*, 1969). $PtCl_4^{2-}$ produced two sites (see Table III), one on either side of the side chain at residue 122, and $PtCl_6^{2-}$ had an additional site at residue 42. The contours of both side chains are well defined and are easily fit with Met side chains. These positions and the direction of the chain at each point were confirmed by fitting the two Met peptides (Dyckes, 1970; Waxdal *et al.*, 1971). The positions of these residues in the sequence are in close agreement with the positions predicted from the molecular weights of the three CNBr fragments. The calculated positions are 44 and 129.

Manganese and Calcium. The Mn^{2+} and Ca^{2+} ions are 4.3 Å apart and occupy what can be considered to be a double site, since Asp-10 and Asp-19 each contribute both carboxyl oxygens as ligands, one to each ion (Figure 7). The Mn^{2+} ion is the highest electron density in the map and corresponds to the site where Cd^{2+} binds to the apoprotein (Weinzierl and Kalb, 1971). From the N terminus the chain runs through the center of the subunit (through sheet II) and wraps around the Mn^{2+} and Ca^{2+} supplying the carboxyl of Glu-8 to Mn^{2+}, the carboxyl of Asp-10 to *both* Mn^{2+} and Ca^{2+}, the carbonyl oxygen of Tyr-12 to Ca^{2+}, and the side chain of Asx-14 to Ca^{2+}. After a small loop out into solution the chain returns underneath the ions and contributes the side chain of Asp-19 to both the Ca^{2+} and Mn^{2+}. His-24 contributes N-3 of the imidazole ring to Mn^{2+} and N-1 appears hydrogen bonded to the carbonyl oxygen of residue 22. After the chain returns through sheet II, Glx-34 contributes its side chain to Mn^{2+}. The symmetry of Mn^{2+} is nearly octahedral with ligands of one nitrogen from imidazole, three oxygens from one Glu and two Asp, and an oxygen or perhaps nitrogen from a Glx side chain. A weak region of electron density appears at the sixth coordinate position for the Mn^{2+} and probably results from partial occupancy of a water molecule. Note that all but one ligand for *both ions* are found in the first 34 amino acids of the sequence. The fifth side chain close to Ca^{2+} is from residue 202 and is probably Glu (Figure 7). However, the possibility cannot be excluded that it may be Lys, which could assist in neutralizing the large cluster of negatively charged acid groups, as found in thermolysin (Matthews *et al.*, 1972a).

It is obvious from the course of the polypeptide chain in this region that Mn^{2+} and Ca^{2+} greatly stabilize the conformation of the subunit. In the absence of these ions the negatively charged side chains clustered at this site would repel one another, forcing this region of the chain out into the solvent. It is not surprising therefore that the apoprotein cannot be crystallized in the same space group (Jack *et al.*, 1971).

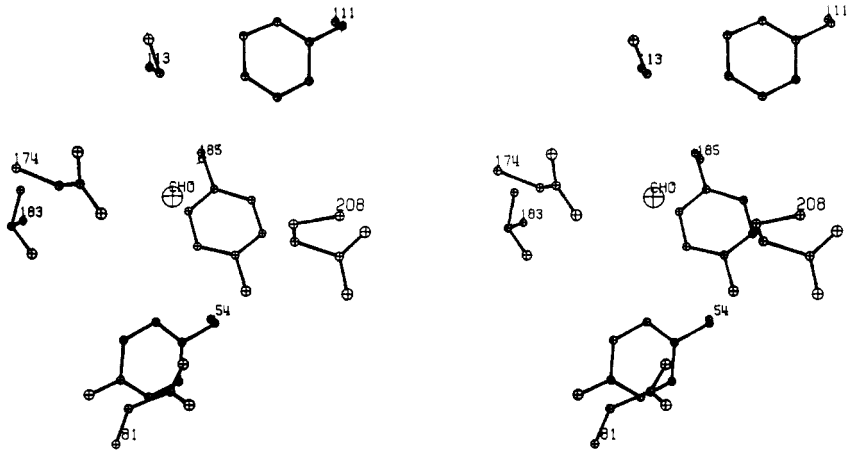

FIGURE 8: Stereogram of the amino acid residues of native Con A in the carbohydrate binding cavity. CHO marks the center of the electron density peak for inositol found previously in the inositol–Con A complex.

The apoprotein must first bind Mn^{2+}, or some other specific transition metal ion, *e.g.*, Ni^{2+}, before it is capable of binding Ca^{2+}. The association constants for Ni^{2+} and Ca^{2+} have been found to be 1.3×10^5 and 3.0×10^3 l./mol, respectively (Kalb and Levitzki, 1968). The transition metal ion therefore must stabilize the conformation of the groups in this region to the extent that at least some of the groups coordinating the more weakly bound Ca^{2+} are properly positioned. This is easily explainable by examination of the Mn^{2+}–Ca^{2+} double site. Upon binding Mn^{2+}, which is located closer to the interior of the subunit (Figure 1), the Asp-10 and Asp-19 side chains would be properly positioned to initiate Ca^{2+} binding (Figure 7).

The Ca^{2+} site of Con A differs from any of the four Ca^{2+} sites of thermolysin (Colman *et al.*, 1972; Matthews *et al.*, 1972b) in that soaking crystals of native Con A in solutions containing elements of the lanthanide series produces no changes in electron density at the Ca^{2+} site. In thermolysin two Ca^{2+} ions are bound 3.8 Å apart in a double site, one of which can be replaced by lanthanides (Matthews *et al.*, 1972a). The failure of lanthanide replacement of Ca^{2+} in Con A may result because the Ca^{2+} is less accessible to solvent than in thermolysin or more probably because of the influence of the more tightly bound Mn^{2+}.

Both a transition metal ion and Ca^{2+} must be bound before Con A is capable of binding carbohydrate (Kalb and Levitzki, 1968), even though the carbohydrate binding site and the Mn^{2+}–Ca^{2+} double site are approximately 23 Å apart. It is therefore speculated that effects of the distortions of the conformation in the Mn^{2+}–Ca^{2+} region are transmitted through the center of the subunit by one or more chains in sheet II to the carbohydrate binding site. Another related question which may be raised is what effect, if any, would removal of Mn^{2+} and/or Ca^{2+} have on the subunit associations, which involve sheet I, to which neither the Mn^{2+} or the Ca^{2+} is directly bound.

The Carbohydrate Binding Site. A preliminary discussion of the carbohydrate binding site as indicated by the binding of *myo*-inositol has been reported (Hardman and Ainsworth, 1972). The numbering of the residues has been revised in this article because of errors in tracing the backbone in lower resolution maps of native Con A. The structure in the binding site region is dominated by polypeptide chains of the β-structure sheets I and II (Figures 1 and 2) which spread apart forming a "v"-shaped framework. This region is then completed with two additional sections of chain, 80–90 and 170–180. Figure 8 shows the side-chain positions of prominent residues in the *native* Con A structure.[4] Only the center of the difference electron density peak found for the inositol molecule (Hardman and Ainsworth, 1972) is marked. Side chains in the immediate vicinity are Tyr-54 and -185, Asx-81 and -174, Thr-183, Ser-113, Glx-208, and Phe-111 (Figure 8). The center of the inositol ring is 8–10 Å from the surface of the subunit and the largest opening to the cavity is delineated by residues 56, 82, 86, 175, and 182. The size of the cavity suggests that in the absence of bound carbohydrate, solvent molecules occupy this region.

The pairs of negative and positive peaks shown in the difference electron density map (Hardman and Ainsworth, 1972), peaks C–D and E–F, indicate movement of Asx-174 and Tyr-185, respectively. The *native* map indicates that there is a hydrogen bond between Asx-81 and Tyr-185; however, upon binding inositol, Tyr-185 moves making these side chains available for bonding to the carbohydrate. Tyrosine acetylation (Doyle and Roholt, 1968) and titration studies (Hassing *et al.*, 1971) had previously implicated involvement of tyrosyl and carboxyl groups, respectively, in carbohydrate binding.

Comparative studies of carbohydrate binding to Con A have show the minimum configuration required is 2-deoxy-1,5-anhydro-D-*arabino*-hexitol, or a carbohydrate with hydroxyl positions which mimic this steric arrangement (Goldstein *et al.*, 1965; Poretz and Goldstein, 1970), and α-D-mannopyranose is the preferred monohexose configuration (Goldstein *et al.*, 1965). When *myo*-inositol is oriented with its only axial hydroxyl (C-4) equivalent to the axial C-2 hydroxyl of α-D-mannopyranose, the structures are very similar. If a model of α-D-mannopyranose is placed in the binding site of the *native* Con A skeletal model with the C-1 hydroxyl toward the cavity opening, the hydroxyls point towards oxygens of the side chains shown in Figure 8.

Acknowledgments

We thank Marianne Schiffer, Mical Wood, Dayle Sly, and Allen Edmundson for their discussions, and Frank Williamson,

[4] In comparison to the α-carbon stereogram previously published (Figure 3 of Hardman and Ainsworth, 1972), Figure 8 is rotated clockwise 90°.

Jack Williams, Jeanne Blomquist, and Rowland Girling for computer programming advice.

Added in Proof

Since this paper was submitted, an article (Edelman *et al.*, 1972) discussing the tentative amino acid sequence and three-dimensional structure of concanavalin A has appeared. According to this sequence, a loop of 7 amino acids should be inserted between our residues 115 and 116. In our electron density map there are contours in this region which are continuous and can accommodate these 7 residues with no break in the chain. However, these contours are much weaker than the usual chain density. This and other differences are to be discussed in a subsequent article.

References

Agrawal, B. B. L., and Goldstein, I. J. (1967), *Biochim. Biophys. Acta 133*, 367.

Berlin, R. D. (1972), *Nature (London), New Biol. 235*, 44.

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.

Busing, W. R., Martin, K. D., and Levy, H. A. (1962), *Oak Ridge Tech. Manual*, 305.

Colman, P. M., Weaver, L. H., and Matthews, B. W. (1972), *Biochem. Biophys. Res. Commun. 46*, 1999.

Dickerson, R. E., Eisenberg, D., Varnum, J., and Kopka, M. L. (1969), *J. Mol. Biol. 45*, 77.

Dickerson, R. E., Weinzierl, J. E., and Palmer, R. A. (1968), *Acta Crystallogr., Sect. B 24*, 997.

Doyle, R. J., and Roholt, O. A. (1968), *Life Sci. 7*, 841.

Dyckes, D. F. (1970), Ph.D. Thesis, Case Western Reserve University.

Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. (1972), *Proc. Nat. Acad. Sci. U. S. 69*, 2580.

Edmundson, A. B., Ely, K. R., Sly, D. A., Westholm, F. A., Powers, D. A., and Liener, I. E. (1971), *Biochemistry 10*, 3554.

Goldstein, I. J., Hollerman, C. E., and Smith, E. E. (1965), *Biochemistry 4*, 876.

Greer, J., Kaufman, H. W., and Kalb, A. J. (1970), *J. Mol. Biol. 48*, 365.

Hardman, K. D., and Ainsworth, C. F. (1972), *Nature (London), New Biol. 237*, 54.

Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B., and Ainsworth, C. F. (1971a), *Proc. Nat. Acad. Sci. U. S. 68*, 1393.

Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B., and Ainsworth, C. F. (1971b), *Cold Spring Harbor Symp. Quant. Biol. 36*, 271.

Hassing, G. S., Goldstein, I. J., and Marini, M. (1971), *Biochim. Biophys. Acta 243*, 90.

Inbar, M., Ben-Bassat, H., and Sachs, L. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 2748.

Inbar, M., and Sachs, L. (1969), *Proc. Nat. Acad. Sci. U. S. 63*, 1418.

Jack, A., Weinzierl, J., and Kalb, A. J. (1971), *J. Mol. Biol. 58*, 389.

Johnson, C. K. (1965), *Oak Ridge Tech. Manual*, 3794.

Kalb, A. J., and Levitzki, A. (1968), *Biochem. J. 109*, 669.

Kalb, A. J., and Lustig, A. (1968), *Biochim. Biophys. Acta 168*, 366.

Kraut, J., Sieker, L. C., High, D. F., and Freer, S. T. (1962), *Proc. Nat. Acad. Sci. U. S. 48*, 1417.

Leon, M. A., and Young, N. M. (1970), *J. Immunol. 104*, 1556.

Lloyd, K. O., Kabat, E. A., and Beychok, S. (1969), *J. Immunol. 102*, 1354.

Matthews, B. W., Colman, P. M., Jansonius, J. N., Titani, K., Walsh, K. A., and Neurath, H. (1972a), *Nature (London), New Biol. 238*, 41.

Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., and Dupourque, D. (1972b), *Nature (London), New Biol. 238*, 37.

Moscona, A. A. (1971), *Science 171*, 905.

Novogrodsky, A., and Katchalski, E. (1971), *Biochim. Biophys. Acta 228*, 579.

Poretz, R. D., and Goldstein, I. J. (1970), *Biochemistry 9*, 2890.

Powell, A. E., and Leon, M. A. (1970), *Exp. Cell Res. 62*, 315.

Quiocho, F. A., Reeke, G. N., Jr., Becker, J. W., Lipscomb, W. N., and Edelman, G. M. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 1853.

Reeke, G. N., Becker, J. W., and Quiocho, F. A. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36*, 277.

Richards, F. M. (1968), *J. Mol. Biol. 37*, 225.

So, L. L., and Goldstein, I. J. (1968), *J. Biol. Chem. 243*, 2003.

Stewart, J. M. (1970), TR-70-58 (NSG-398), Computer Science Center, University of Maryland.

Sumner, J. B., Gralen, N., and Eriksson-Quensel, I. B. (1938), *J. Biol. Chem. 125*, 45.

Sumner, J. B., and Howell, S. F. (1936a), *J. Bacteriol. 32*, 227.

Sumner, J. B., and Howell, S. F. (1936b), *J. Biol. Chem. 115*, 583.

Wang, J. L., Cunningham, B. A., and Edelman, G. M. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 1130.

Waxdal, M. J., Wang, J. L., Pflumm, M. N., and Edelman, G. M. (1971), *Biochemistry 10*, 3343.

Weinzierl, J., and Kalb, A. J. (1971), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 18*, 268.