The Covalent and Three-Dimensional Structure of Concanavalin A

(x-ray crystallography/sequence/2-Å resolution/binding sites/lectin)

GERALD M. EDELMAN, BRUCE A. CUNNINGHAM, GEORGE N. REEKE, JR., JOSEPH W. BECKER, MYRON J. WAXDAL, AND JOHN L. WANG

The Rockefeller University, New York, N.Y. 10021

Contributed by Gerald M. Edelman, July 5, 1972

ABSTRACT The tentative amino-acid sequence and three-dimensional structure of the lectin concanavalin A have been determined. The amino-acid sequence, which was determined chemically, contains 238 residues. The sequences of three short stretches were assigned on the basis of x-ray crystallographic data. Interpretation of an electron density map at 2-Å resolution indicates that the predominant structural element is extended polypeptide chain arranged in two anti-parallel pleated sheets or β structures. Residues not included in the β -structures are arranged in regions of random coil. One of the pleated sheets contributes extensively to the interactions among the monomers to form both dimers and tetramers. The positions at which Mn²⁺, Ca²⁺, and saccharide are bound to the protein, and the point of cleavage for the formation of the naturally occurring fragments A1 and A2, have been tentatively assigned. Both metal-binding sites are at least 20-Å removed from the position at which saccharides are bound. The saccharide-binding site is a deep pocket of approximately $6\text{\AA} \times 7.5\text{\AA} \times 18\text{\AA}$, the inner portion of which is occupied by hydrophobic residues.

Concanavalin A (Con A) is one of a group of plant proteins known as lectins (1), some of which have useful properties for studies of cell surfaces and cell division. Although the function of Con A in the jack bean (*Canavalia ensiformis*) is unknown, it can agglutinate various somatic (2, 3) and germ line cells (2), and it is mitogenic for lymphocytes (4). After suitable treatment, it can restore the growth pattern of virustransformed fibroblasts in tissue culture to that of normal cells (5). Various mammalian cells bind from 10⁶ to 10⁷ Con A molecules per cell, presumably via surface glycoprotein receptors. This binding can be inhibited by α -D-gluco- and α -D-mannopyranosides. Very little is known, however, of the mechanism by which Con A leads to agglutination, mitogenesis, or cell-surface alteration.

In order to understand the mode of action of Con A, we have undertaken structural and functional studies. We report here the tentative amino-acid sequence and an interpretation of an electron density map at 2-Å resolution. The results suggest useful models for analyzing the biological activities of this protein.

MATERIALS AND METHODS

The isolation of the intact subunit of Con A and the preparation of its CNBr fragments have been described (6, 7). The amino-acid sequence was determined by the dansyl-Edman procedure (8, 9), performed on peptides isolated after digestion of the CNBr fragments with trypsin, chymotrypsin, pepsin, thermolysin, and subtilisin. The details of these ex-

Abbreviations: Con A, concanavalin A; β -IPG, β -(o-iodophenyl)p-glucopyranoside. periments will be reported elsewhere (B. A. Cunningham, J. L. Wang, M. J. Waxdal, and G. M. Edelman, in preparation).

The 2-Å electron density map was prepared by the use of phases calculated by the method of multiple isomorphous replacement (10). The crystallographic data were acquired by standard photographic techniques, and integrated intensities were obtained with a digital densitometer (Optronics International, Chelmsford, Mass.). Data were collected to Bragg spacings of 2.0Å from crystals of the native protein and three heavy-metal derivatives: sodium mersalyl, Pb-(NO₃)₂, and K₂PtCl₄. Intensities within 2.8Å were also measured on three other derivatives: $UO_2(NO_3)_2$, β -(o-iodophenyl)-D-glucopyranoside (β -IPG), and a second K₂PtCl₄ derivative. Additional data have been collected on a $Sm(NO_3)_3$ derivative. A skeletal model of the protein was constructed from the electron density map by the use of Kendrew models and an optical comparator (11). A complete description of the data and procedures will be reported (G. N. Reeke, Jr., J. W. Becker, J. L. Wang, B. A. Cunningham, and G. M. Edelman, in preparation).

RESULTS AND DISCUSSION

Con A in solution consists largely of dimers at pH values less than 6 and of tetramers above pH 7(12). Electron density maps at low resolution reveal asymmetric units (molecular weight 25,500) forming ellipsoidal domes paired across a



Fig. 1. Schematic representation of the tetrameric structure of Con A viewed down the z axis. The proposed binding sites for transition metals, calcium, and saccharides are indicated by Mn, Ca, and C, respectively.

1	0
ALA-ASP-THR-ILE-VAL-ALA-VAL-GLU-LEU-ASP-THR-TYR-PRO-ASN-THR-ASP-ILE-GLY-ASP-PR	₹0-
30	0
SER-TYR-PRO-HIS-ILE-GLY-ILE-ASP-ILE-LYS-SER-VAL-ARG-SER-LYS-LYS-THR-ALA-LYS-TR	≀P-
41 50 66	0
ASN-MET-GLN-ASP-GLY-LYS-VAL-GLY-THR-ALA-HIS-ILE-ILE-TYR-ASN-SER-VAL-ASP-LYS-AR	:G-
61 70 LEU-SER-ALA-VAL-VAL-SER-TYR-PRO-ASX-ALA-ASX-ASX-ALA-THR-SER-VAL-SER-TYR-ASX-VA	0 .L-
81	0
ASX-LEU-ASX-ASX-VAL-LEU-PRO-GLN-TRP-VAL-ARG-VAL-GLY-LEU-SER-ALA-SER-THR-GLY-LE	U-
101 110 120	0
TYR-LYS-GLU-THR-ASN-THR-LEU-ILE-SER-PHE-SER-TRP-THR-SER-LYS-LEU-LYS-SER-ASX-SEI	R-
121 130 THR-HIS-GLX-THR-ASX-ALA-LEU-HIS-PHE-MET-PHE-ASN-GLN-PHE-SER-LYS-ASP-GLN-LYS-ASF	0 P-
141 150 LEU-ILE-LEU-GLN-GLY-ASP-ALA-THR-THR-GLY-THR-ASP-GLY-ASN-LEU-GLU-LEU-THR-ARG-VAL	0 L-
161 170 180)
SER-SER-ASX-(GLX, SER, PRO, GLY)-GLY-SER-SER-VAL-GLY-ARG-ALA-LEU-PHE-TYR-ALA-PRO-VAL	
181)
HIS-ILE-TRP-GLU-SER-SER-ALA-(GLX, ALA, SER, VAL)-PHE-GLU-ALA-THR-PHE-(THR,LEU, VAL)-ILE	I-
201 210)
LYS-SER-PRO-ASP-SER-HIS-PRO-ALA-ASP-GLY-ILE-ALA-PHE-PHE-ILE-SER-ASN-ILE-ASX-SER	!-
221 230 SER-ILE-PRO-SER-GLY-SER-THR-GLY-ARG-LEU-LEU-GLY-LEU-PHE-PRO-ASP-ALA-ASN	

FIG. 2. Tentative amino-acid sequence of Con A. Sequences assigned by interpretation of the electron density map are shown in parentheses.

2-fold axis parallel to the crystallographic c axis to form dimers. The dimers are, in turn, paired across points of D_2 symmetry to form tetramers of roughly tetrahedral shape (13, 14). Analysis of the location of iodine in β -IPG (15) suggested that the saccharide-binding site is located in each protomer, as shown in Fig. 1.

Solution studies have confirmed the molecular weight of the protomer and have also revealed that Con A exists in two forms, one containing an intact subunit and one with a subunit made up of two fragments, A_1 (molecular weight 13,000) and A_2 (molecular weight 11,000) (16). The fragments and intact subunits have been purified, and mixed crystals that are virtually indistinguishable from crystals of the intact subunits can be formed despite the fact that the solution properties of the two forms of Con A are somewhat different (6, 16).

The tentative amino-acid sequence of Con A is presented in Fig. 2. The intact subunit contains two methionines, and CNBr cleavage yields the expected three fragments: F_1 (residues 1-42), F_2 (residues 43-130), and F_3 (residues 131-238). There are three regions of the polypeptide chain (residues 164-167, 188-191, 197-199) in which the sequence has not been established chemically. The amino-acid residues in these three portions were tentatively identified by interpretation of the electron density map.

The studies at 2-Å resolution indicate that the Con A subunit is a globular protein of overall dimensions $42\text{\AA} \times 40\text{\AA} \times$ 39Å, measured roughly parallel to the crystallographic

[110], $[1\overline{10}]$, and [001] directions (Fig. 3). The predominant structural element is extended polypeptide chain, arranged in two anti-parallel pleated sheets or β -structures (Fig. 4). The first of these sheets contains 51 residues arranged in seven hydrogen-bonded chains (residues 132-125, 106-116, 199-190, 48-55, 66-59, and 73-78) and nine residues in short connecting loops. This sheet forms the back of the molecule (Figs. 3 and 4), including the back lining of the saccharidebinding cavity. The second sheet is roughly perpendicular to the first sheet, extending vertically from the back of the molecule to the front. It contains about 34 residues, arranged in six anti-parallel chains (residues 144-140, 173-177, 97-92, 209-215, 9-4, and 25-29). This second sheet divides the remainder of the molecule unequally into left and right randomly-coiled regions. The left-hand coil, consisting of residues 133-169, is arranged in three loosely organized turns. The right-hand coil includes the amino and carboxyl termini at the front of the molecule, the metal-binding region, and the front wall of the saccharide-binding cavity. At the far right of this coil is a single turn of approximately α -helical structure, the only such structure in the molecule.

The pleated sheet at the back of the molecule extends across the interface between the two subunits that comprise an ellipsoidal dimer. The main-chain nitrogen and carbonyl oxygen atoms of residues 125-132 are apparently hydrogen bonded to complementary atoms of the same residues in the protomer across the 2-fold axis parallel to c. The back of each dimer thus consists of a single anti-parallel pleated



FIG. 3. Skeletal model of the Con A protomer viewed down the z axis. [110] is horizontal, [110] is approximately vertical. The 2fold axis relating members of an ellipsoidal dimer (second protomer not shown) is roughly perpendicular to the illustration at the *bottom* center. The metal atoms are represented by the two white spheres at the top of the model. The β -IPG iodine-binding site is indicated by the sphere at the *lower right*.

sheet containing 14 hydrogen bonded chains, 7 from each monomer. Other interactions between halves of the ellipsoidal

dimer apparently involve residues 87-88, 136-140, and 177-180.



FIG. 4. Stereo drawing of the α -carbon backbone of the Con A protomer oriented as in Fig. 3. The positions shown for residues 34 and 35 are tentative. Computer-drawn stereo figure was made by the program OR-TEP of Dr. Carroll Johnson.

Ellipsoidal dimers interact primarily through side chains projecting from the large β -structure in the back of the molecule (Fig. 4) into the region of dimer-dimer contact (Fig. 1). Complementary side chains of both dimers project into this region. Although these side chains are predominantly hydrophilic, many hydrophobic groups are also present. Thus, the β -structure at the back of the molecule plays a major role in subunit interactions, with extension of the pleated sheet contributing to dimer formation and side-chain interactions to tetramer formation.

We have identified the site of the natural cleavage of Con A between residues 119–120. These residues are located on a loop that extends well away from the main body of the molecule (lower right, Fig. 4). This loop forms the connection between the first and second chains of the rear pleated sheet. It appears from the structure that, this loop would be readily accessible to the hypothesized enzymatic cleavage (16), and that the cleavage would have no major effect on the folding of the polypeptide chain. This conclusion is consistent with our failure to observe large structural differences between the two forms of Con A in 2.8-Å difference electron density maps (G. N. Reeke, Jr. and J. W. Becker, unpublished observations). In cleaved Con A, however, there may be a general loosening of the β -structure in this region, which is near the β -IPG-binding site.

We have also tentatively assigned the location of the Mn^{2+} and Ca^{2+} atoms in the Con A molecule. An intense, roughly spherical peak of electron density, which cannot be accounted for by the amino-acid sequence, is found at 0.69, 0.22, 0.24, the position suggested to be the transition metal-binding site of Con A by experiments on metal substitution (17). This peak is contained within a roughly octahedral coordination shell having a radius of approximately 2.7Å. Four of the six ligands are the side chains of Glu 8, Asp 10, Asp 19, and His 24. The other two are probably water molecules, one of which appears to be in a position to be hydrogen bonded to the carbonyl oxygen atom of Val 32. The sixth coordination position is at the bottom of a channel, apparently containing only solvent, that extends to the surface of the molecule.

At 0.74, 0.23, 0.28, approximately 5.3Å from this metal site, there is another strong peak that cannot be assigned to any part of the polypeptide chain. This peak is also surrounded by a roughly octahedral coordination shell with an approximate radius of 2.5Å. The apparent ligands are the carbonyl oxygen of Tyr 12, the side chains of Asn 14, Asp 10, and Asp 19, and two water molecules, which are possibly hydrogen bonded to the side chain of Asp 209 and the carbonyl oxygen of Arg 229. From these observations, we suggest that this second peak may represent the calcium-binding site of Con A.

The fact that the side chains of Asp 10 and Asp 19 are ligands of both of the proposed metals suggests an explanation for the observation (18) that transition metal binding is a necessary precursor of alkaline earth binding in this protein. Binding of Mn^{2+} to residues 8, 10, 19, and 24 may force the amino-terminal region of the polypeptide chain into the tightly folded configuration we observe, thus bringing residues 10, 12, 14, and 19 into the proper orientation with respect to residues 209 and 229 to form the calcium-binding site. If the acidic residues do indeed bind the metals, this would account for the observed dependence of metal binding on pH (18).

According to our hypothesis, the two metal atoms necessary for saccharide binding can be described as forming a binuclear complex consisting of two octahedra sharing a common edge. This complex is more than 20 Å removed from the saccharidebinding site; therefore, the observed influence of metal binding on the saccharide-binding ability of Con A must arise through a structural change in the saccharide-binding site induced by bound metal, rather than through direct participation of the metals or their ligands in saccharide binding.

At least one other peak of electron density might represent a metal-binding site. This potential site, at 0.35, 0.38, 0.29, is on the surface of the molecule, and is associated only with the side chain of Gln 88. This location is the site of substitution of several metal cations used in our search for heavyatom derivatives, particularly the Pb²⁺ atom in the Pb(NO₃)₂ derivative. Because the peak is associated with only one amino-acid residue, and is apparently unrelated to either the transition metal- or saccharide-binding sites, we assume that it represents a minor metal-binding site that is not necessary for carbohydrate binding.

We have previously investigated the interaction of Con A with various monosaccharides by both x-ray crystallography (15) and circular dichroism (19). The location of the saccharide-binding site of Con A was deduced from the position of the only peak in 2.8-Å difference projections of the β -IPG derivative of Con A (15). This compound is an inhibitor of both hemagglutination (G. R. Gunther and G. M. Edelman, unpublished observations) and precipitation of polysaccharides (20). Analysis of three-dimensional data on this derivative confirms the presence of a single molecule bound at the site previously identified. The iodine atom of β -IPG is located in a relatively narrow, but deep, pocket at the lower right of the molecule, as shown in Figs. 3 and 4, about 15Å from the surface of the protein. The pocket is from 3.5Å to 6Å wide, about 7.5Å high, and 18Å deep. The iodine atom is located at the inner end of the pocket, 6.5Å from the nearest backbone nitrogen or oxygen atom (the carbonyl oxygen of His 181). The iodine atom is surrounded by 12 side chains at distances of from 3.0Å to 7.5Å: Tyr 54, Leu 82, Leu 86, Val 90, Val 92, Trp 112, Ser 114, Val 180, Ile 182, Phe 192, Phe 213, and Ile 215. All of these residues except Ser 114 are hydrophobic. This observation is consistent with the previous hypothesis that there is a nonpolar region near the carbohydrate-binding site (20).

We were unable to locate the glucopyranoside ring in the β -IPG difference map, presumably due to the presence of the large iodine atom in the substituted sugar. However, from examination of the Con A model (Fig. 3) and consideration of the possible geometries of β -IPG, we find that the glucoside ring can best be accommodated at a site midway between the known position of the iodine atom and the surface of the protein. The available data indicate that Con A probably binds di- and oligosaccharides through their nonreducing terminal residues (21). In order to allow the binding of oligoand polysaccharides to Con A, the orientation of the nonreducing terminal residue must be reversed relative to its postulated orientation in β -IPG binding, so that C-1 is toward the surface of Con A and C-6 occupies a position close to the iodine site of β -IPG. Regardless of the orientation of the bound saccharide in the proposed binding site, the side chains of Tyr 54, Ser 56, Asx 83, Ser 114, and Ser 190 and the backbone carbonyl oxygen atoms of Ile 182 and Lys 115 would be available for use in carbohydrate binding.

The binding site of Con A for myoinositol (22) is 1.5Å from the iodine atom of β -IPG. This position is 15Å from

the molecular surface at the opening of the saccharidebinding cavity. Our proposed model for saccharide binding implies that this position is not involved in binding di-, oligo-, and polysaccharides to Con A, although it is near the saccharide-binding site. This interpretation is consistent with the fact that myoinositol is not an inhibitor of hemagglutination (J. L. Wang and G. M. Edelman, unpublished observations) or precipitation of dextran (21).

Con A is the first lectin and macromolecular mitogen whose structure has been determined. It is a subunit protein and a metalloprotein, and physical-chemical experiments interpreted in terms of the three-dimensional structure should provide insights into the stabilization of the protein by divalent ions, subunit equilibria, and cooperativity of saccharide binding and metal binding.

Above all, however, a knowledge of the structure of a macromolecule capable of interacting with cell surfaces should be helpful in studying the mechanism of mitogenesis and cell membrane alteration. The structural findings suggest two main hypotheses for the mechanism of action of Con A. (i) By virtue of its valence, Con A may function to crosslink mobile glycoprotein receptors into micropatches consisting of about 100–1000 receptor molecules. Formation of such patches may lead to an alteration in the activity of enzymes such as adenyl cyclase or phosphodiesterase. (ii) After binding to cell surface glycoproteins, Con A may interact with the cell membrane via portions of the molecular surface that alone would not bind. This secondary interaction may provide the necessary surface alteration to stimulate mitogenesis.

The valence hypothesis is supported by experiments on the inhibition of immunoglobulin receptor mobility by Con A (23) that suggest interaction between Con A receptors and immunoglobulin receptors. Moreover, experiments have shown that whereas free Con A stimulates thymus-derived (T) cells only, Con A cross-linked at surfaces stimulates bone marrow-derived (B) cells (24). These experiments, as well as similar studies on other lectins linked to beads (25), appear to rule out the possibility that endocytosed lectin molecules act primarily within the cell. Similar observations have been made on specific lymphoid cell populations isolated by fiber fractionation (26).

Knowledge of the sequence and three-dimensional structure of Con A and the chemical characterization of its cell receptors should help in the design of experiments to validate either the valency or the secondary receptor hypothesis. The structural results will be particularly valuable in correlating the modification of specific amino-acid residues in chemically derivatized Con A with corresponding alterations in biological activity.

This work was supported by a Teacher-Scholar Grant from the Camille and Henry Dreyfus Foundation and Grants AI-09921, AI-09999, AI-09273, and AM-14705 from the U.S. Public Health Service. A National Institutes of Health Postdoctoral Fellowship to J. W. B. is gratefully acknowledged. We are grateful to Professor Robert Langridge of Princeton University for making available to us his rotating drum scanner and to Professor Brian W. Matthews of the University of Oregon for making available to us his computer program for the digital scanner, which we modified for use in this project. We thank Miss Catherine Volin, Miss Iris Valters, Mrs. Martha O'Connor, Mrs. Helvi Hjelt, Miss Katherine Bergmann, and Miss Joan Beck for technical assistance.

- Boyd, W. C. (1963) "The lectins: their present status," Vox Sang. 8, 1-32.
- 2. Edelman, G. M. & Millette, C. F. (1971) "Molecular probes

of spermatozoan structures," Proc. Nat. Acad. Sci. USA 68, 2436-2440.

- 3. Inbar, M. & Sachs, L. (1969) "Interaction of the carbohydrate-binding protein concanavalin A with normal and transformed cells," *Proc. Nat. Acad. Sci. USA* 63, 1418-1425.
- Powell, A. E. & Leon, M. A. (1970) "Reversible interaction of human lymphocytes with the mitogen concanavalin A," *Exp. Cell Res.* 62, 315-325.
- 5. Burger, M. M. & Noonan, K. D. (1970) "Restoration of normal growth by covering of agglutinin sites on tumor cell surface," *Nature* 228, 512-515.
- Cunningham, B. A., Wang, J. L., Pflumm, M. N. & Edelman, G. M. (1972) "Isolation and proteolytic cleavage of the intact subunit of concanavalin A," *Biochemistry*, in press.
- Waxdal, M. J., Wang, J. L., Pflumm, M. N. & Edelman, G. M. (1971) "Isolation and order of the cyanogen bromide fragments of concanavalin A," *Biochemistry* 10, 3343-3347.
- 8. Cunningham, B. A., Gottlieb, P. D., Konigsberg, W. H. & Edelman, G. M. (1968) "The covalent structure of a human γ G-immunoglobulin. V. Partial amino-acid sequence of the light chain," *Biochemistry* 7, 1983-1995.
- 9. Gottlieb, P. D., Cunningham, B. A., Rutishauser, U. & Edelman, G. M. (1970) "The covalent structure of a human γ G-immunoglobulin. VI. Amino-acid sequence of the light chain," *Biochemistry* 9, 3155-3161.
- Blow, D. M. & Crick, F. H. (1959) "The treatment of errors in the isomorphous replacement method," Acta Crystallogr. 12, 794-802.
- Richards, F. M. (1968) "The matching of physical models to three-dimensional electron density maps: A simple optical device," J. Mol. Biol. 37, 225-230.
- Kalb, A. J. & Lustig, A. (1968) "The molecular weight of concanavalin A," Biochim. Biophys. Acta 168, 366-367.
- Reeke, G. N. Jr., Becker, J. W. & Quiocho, F. A. (1971) "The structure of concanavalin A at 4Å resolution," Cold Spring Harbor Symp. Quant. Biol. 36, 277-282.
- Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B. & Ainsworth, C. F. (1971) "X-ray crystallographic studies of concanavalin A," Cold Spring Harbor Symp. Quant. Biol. 36, 271-276.
- Becker, J. W., Reeke, G. N. Jr. & Edelman, G. M. (1971) "Location of the saccharide binding site of concanavalin A," J. Biol. Chem. 246, 6123-6125.
- Wang, J. L., Cunningham, B. A. & Edelman, G. M. (1971) "Unusual fragments in the subunit structure of concanavalin A," Proc. Nat. Acad. Sci. USA 68, 1130-1134.
 Weinzierl, J. & Kalb, A. J. (1971) "The transition metal-
- Weinzierl, J. & Kalb, A. J. (1971) "The transition metalbinding site of concanavalin at 2.8Å resolution," FEBS Lett., 18, 268-270.
- Kalb, A. J. & Levitzki, A. (1968) "Metal binding sites of concanavalin A and their role in the binding of α-methyl-Dglucopyranoside," *Biochem. J.* 109, 669-672.
- Pflumm, M. N., Wang, J. L. & Edelman, G. M. (1971) "Conformational changes in concanavalin A," J. Biol. Chem. 246, 4369-4370.
- Poretz, R. D. & Goldstein, I. J. (1971) "Protein-carbohydrate interaction. On the mode of binding of aromatic moieties to concanavalin A, the phytohemagglutinin of the jack bean," *Biochem. Pharmacol.* 20, 2727-2739.
- Goldstein, I. J., Hollerman, C. E. & Smith, E. E. (1965) "Protein-carbohydrate interaction II. Inhibition studies on the interaction of concanavalin A with polysaccharides," *Biochemistry* 4, 876-883.
- 22. Hardman, K. D. & Ainsworth, C. F. (1972) "Myo-inositol binding site of concanavalin A," *Nature New Biol.* 237, 54-55.
- Yahara, I. & Edelman, G. M. (1972) "Restriction of the mobility of lymphocyte immunoglobulin receptors by concanavalin A," Proc. Nat. Acad. Sci. USA 69, 608-612.
- Andersson, J., Edelman, G. M., Möller, G. & Sjöberg, O. (1972) "Activation of B lymphocytes by locally concentrated concanavalin A," *Eur. J. Immunol.*, 2, 233-235.
 Greaves, M. F. & Bauminger, S. (1972) "Activation of T
- Greaves, M. F. & Bauminger, S. (1972) "Activation of T and B lymphocytes by insoluble phytomitogens," Nature New Biol. 235, 67-70.
- Edelman, G. M., Rutishauser, U. & Millette, C. F. (1971) "Cell fractionation and arrangement on fibers, beads, and surfaces," Proc. Nat. Acad. Sci. USA 68, 2153-2157.