Proc. Nat. Acad. Sci. USA
Vol. 68, No. 8, pp. 1853–1857, August 1971

Structure of Concanavalin A at 4 Å Resolution

(x-ray diffraction/heavy-atom derivatives/lectins)

FLORANTE A. QUIOCHO*, GEORGE N. REEKE, JR.†, JOSEPH W. BECKER†,
WILLIAM N. LIPSCOMB*, AND GERALD M. EDELMAN†

* Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138; and
† The Rockefeller University, New York, N.Y., 10021

Contributed by William N. Lipscomb and Gerald M. Edelman, May 17, 1971

ABSTRACT **Concanavalin A, a phytohemagglutinin isolated from the jack bean, crystallizes at pH 6.8 in the orthorhombic space group I222 with a = 89.9, b = 87.2, and c = 63.1 Å. We have analyzed x-ray diffraction intensity data to 4 Å resolution on native concanavalin A and five heavy-metal derivatives: lead, mersalyl, chloroplatinate, uranyl, and *o*-mercuri-*p*-nitrophenol. Heavy-atom positions, occupancies, and isotropic thermal parameters have been refined by least-squares methods.**

The electron density maps clearly show the molecular shape and the packing of the concanavalin A molecules. The asymmetric unit (mol wt 27,000) forms an elliptical dome or "gumdrop" with a base of approximately 46 × 26 Å and a height of 42 Å. The subunits are paired across 2-fold axes parallel to the c-axis to form dimers. The dimers are in turn paired across points of D_2 symmetry to form tetramers of roughly tetrahedral shape. Each unit has a depression located on the surface which could be the site of saccharide binding. In many regions we have been able to trace the course of the polypeptide chain.

Concanavalin A (Con A), a protein obtained from the jack bean, *Canavalia ensiformis*, belongs to a class of compounds known as lectins—plant proteins capable of agglutinating various types of cells and inducing cellular transformation and mitosis (1, 2). Con A is bound in different amounts by normal, trypsinized, and transformed cells (3, 4), inhibits the growth of mouse myelomas *in vivo* (5), and, after treatment with trypsin, restores normal growth patterns to transformed fibroblasts (6).

Con A binds specifically to α-D-mannopyranosyl, α-D-glucopyranosyl, and β-D-fructofuranosyl residues in polysaccharides such as mannan and dextran, as well as to mono- and oligosaccharides bearing the same groups (7–12). Native Con A contains Mn^{2+} and Ca^{2+}, which are necessary for saccharide-binding activity (13, 14). Sedimentation equilibrium (15) and equilibrium dialysis studies (14, 16) indicate that the molecule consists of subunits of molecular weight 3×10^4, each of which contains binding sites for both metals and sugar, and that each molecule contains two subunits at pH values below 6 and four at pH values above 7. The existence of several saccharide-binding sites could account for the ability of Con A to agglutinate erythrocytes and to form precipitates with polysaccharides and glycoproteins by lattice formation.

Crystallization of Con A was first reported by Sumner in 1919 (17). Crystals were also prepared by Nakamura and Sozuno (18), and preliminary crystallographic studies have been reported by Greer *et al.* (19) and by Hardman *et al.* (20).

We have prepared crystals of native Con A and of several isomorphous heavy-metal derivatives and have computed an electron density map of the protein at a nominal resolution of 4 Å.

Abbreviations: Con A, concanavalin A; HMN, *o*-hydroxymercuri-*p*-nitrophenol.

MATERIALS AND METHODS

Con A was prepared by the methods of Sumner (17) and Sumner and Howell (7). All subsequent experiments were performed at 4°C. Crystals were prepared by dialysis of filtered stock solution against 0.1 M Na_2SO_4–0.01 M sodium maleate, pH 6.8. Crystals appeared within a few days and were allowed to grow for 10–14 days. The crystals are bisphenoids with rounded edges, having dimensions approximately 0.5 × 0.7 × 1.0 mm.

Diffraction symmetry and systematic absences observed on precession photographs indicate either of the body-centered orthorhombic space groups, I222 or $I2_12_12_1$. The positions of axial peaks in the heavy-metal difference Patterson syntheses (see below) require the space group to be I222. Unit-cell dimensions were obtained by least-squares analysis of orientation parameters for 12 high-angle reflections measured on the Picker F A C S-I diffractometer. They were found to be a = 89.9 ± 0.1 Å, b = 87.2 ± 0.1 Å, c = 63.1 ± 0.2 Å. These dimensions, the crystal density of 1.20 g/ml obtained by Greer *et al.* (19), and the partial specific volume of 0.73

TABLE 1. *Isomorphous derivatives*

Name	Reagent	Soaking time (days)	No. of crystals used	No. of overlaps	R-factor*
Native	—	—	2	2515	0.036
Lead	1 mM $Pb(NO_3)_2$	2	1	2087	0.027
Platinum	1 mM K_2PtCl_4	3†	2	264	0.065
Mersalyl	0.1 mM sodium mersalyl	2	1	4918	0.039
Uranyl	1 mM $UO_2(NO_3)_2$	5	1	1566	0.042
HMN	1 mM *o*-hydroxymercuri-*p*-nitrophenol	30	1	400	0.037

* $R = \Sigma|\bar{I}(\mathbf{H}) - I(\mathbf{H})|/\Sigma|\bar{I}(\mathbf{H})|$ for the averaging of symmetry-equivalent reflections, where $\bar{I}$ is the average intensity of reflection **H** and I is any measurement of reflection **H**.

† After soaking in K_2PtCl_4, the platinum-derivative crystals were washed for 9 days in three changes of crystallization buffer.

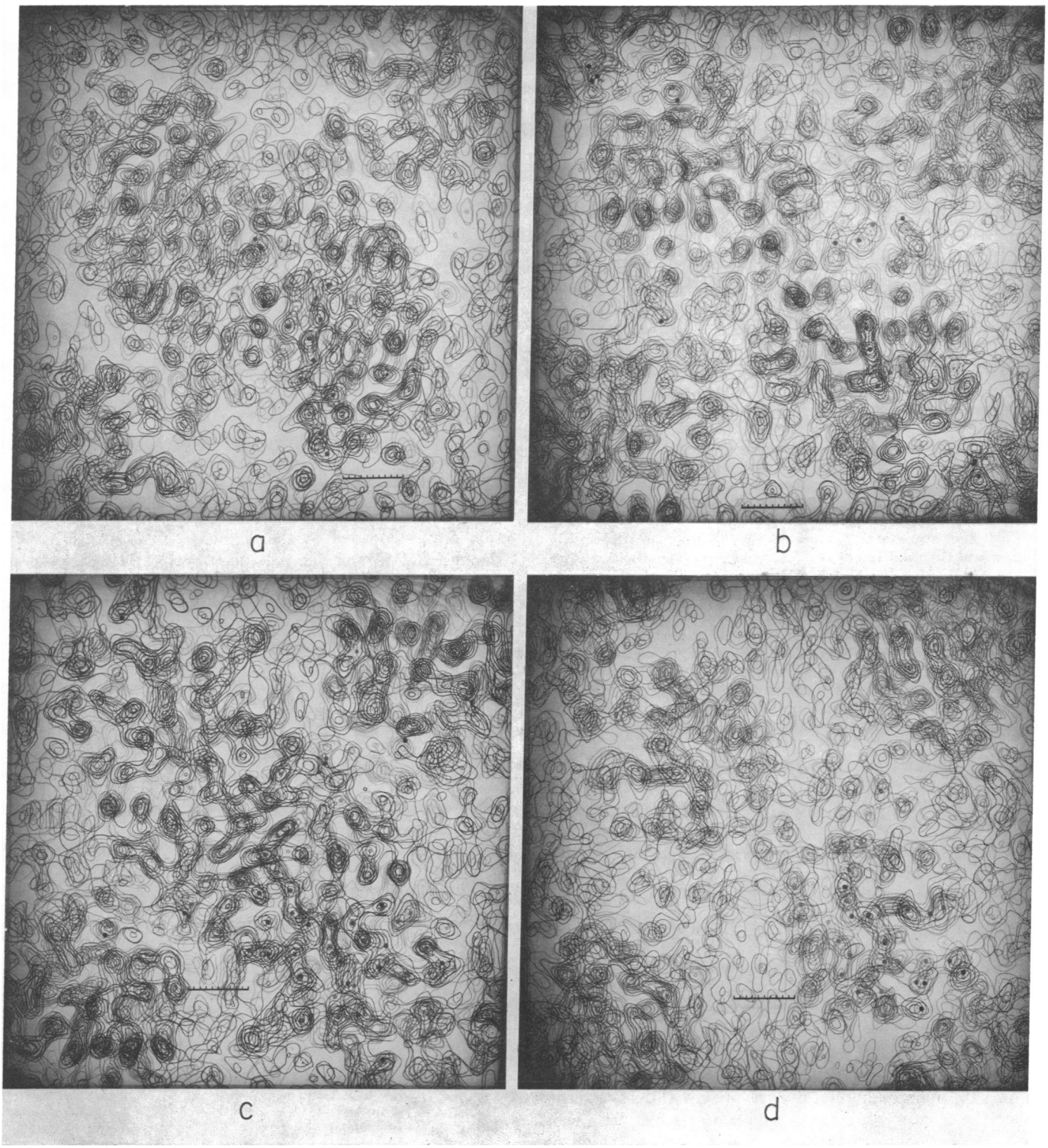

Fig. 1. The complete 4-Å Con A electron-density map (four asymmetric units) in sets of superposed sections viewed down the z axis. The origin is at the upper left corner, with x vertical ($0 \leqslant x \leqslant 1$) and y horizontal ($0 \leqslant y \leqslant 1$). Each scale division = 1 Å. The 2-fold dimer axis is normal to the page at the center of each view. (*a*) $z = 0.00$ to 0.12. The low-density boundary around the central molecule is clearly visible, along with portions of other molecules at the corners of the figures. Many stretches of polypeptide chain can be seen, for example, diagonally across the lower right of the figure. (*b*) $z = 0.12$ to 0.24. The possible saccharide binding sites are seen as regions of low density which nearly bisect the dimers across the two-fold axis. (*c*) $z = 0.24$ to 0.36. At this level the cavities have been completed by upper walls of polypeptide chain. The region of monomer–monomer interaction is visible as a narrow strip of low density passing diagonally from lower left to upper right across the center of the figure. (*d*) $z = 0.36$ to 0.48. At this level there is very little monomer–monomer interaction, and the subunits are seen as well-resolved units.

ml/g obtained by Sumner *et al.* (21) can be used to obtain a value of 2.55×10^4 for the weight of protein per asymmetric unit. This value corresponds to the equivalent binding weight of Con A for metal ions and saccharides (14, 16). The diffraction pattern from these crystals extends beyond spacings of 2.0 Å.

TABLE 2. *Criteria of refinement*

Derivative	Kraut R factor*	Root-mean-square lack of closure vs. $\sin^2\theta/\lambda^2$ 0.0140 to 0.0155	0.0105 to 0.0140	0.0070 to 0.0105	0.0035 to 0.0070	0.0 to 0.0035	All data
Uranyl	0.111	99	91	73	96	199	107
HMN†	0.077	69	71	65	42	103	69
Lead	0.039	44	40	30	27	31	35
Platinum	0.081	74	63	55	64	162	81
Mersalyl	0.068	56	52	41	48	123	63
Figure of merit		0.71	0.76	0.83	0.83	0.81	0.79

* $R_K = [\Sigma||F_D| - |\mathbf{F}_N + \mathbf{f}_c||]/\Sigma|F_D|$.
† *o*-Hydroxymercuri-*p*-nitrophenol.

Heavy-metal derivatives were prepared by placing crystals of native Con A in solutions made up by dissolving heavy-metal compounds in the crystallization buffer. The derivatives and conditions for their preparation are presented in Table 1. For the $PtCl_4^{2-}$ derivative, we found that the quality of the difference Patterson maps could be improved considerably by washing the crystals in platinum-free crystallization buffer after the heavy atom was introduced. The apparently irreversible binding is similar to that observed by Wyckoff *et al.* (22) for the dichloroethylenediamine platinum (II) derivative of ribonuclease S, in which a bond was formed between platinum and a methionyl sulfur.

X-ray diffraction intensities were measured on a Picker FACS-I four-circle diffractometer using CuK_α radiation filtered through nickel foil. The omega-scan method was used, with a scan width of 0.80–1.00° and background counting for a fixed time interval before and after the peak scan. The temperature was maintained at 4°C by means of a cold air stream, and the path between crystal and detector was evacuated by use of a beam tunnel device which could be kept at 1–2 torr (W. H. Haugen, unpublished apparatus). Crystal quality was monitored by observing three standard high-angle reflections every 100–150 reflections and rejecting a crystal when the net intensity of any one of them fell below 90% of its original value. The crystals were exceptionally hardy, with an average lifetime of 6–8 days in the x-ray beam, which allowed a complete data set (about 2200 reflections to 4 Å resolution) to be collected for each derivative from only one or two crystals. Reflections were corrected for background, Lorentz, and polarization effects, and a semiempirical absorption correction was applied (23). Symmetry-equivalent reflections were averaged; the R-factors for agreement between symmetry-equivalent data are presented in Table 1.

LOCATION OF HEAVY ATOMS

One lead, one mersalyl, and two platinum sites were located in Harker sections of the "derivative minus native" difference Patterson functions. The difference Patterson map for lead was exceptionally clean and unambiguous in interpretation. The platinum and mersalyl maps contained concentrations of extra (unexplained) density, particularly along the axes, which could be reduced by omitting the data with $d > 10$ Å from the calculations.

Heavy-atom positions were related to a common origin by use of the Rossmann function (24). These assignments were checked by examination of difference Fourier syntheses for

TABLE 3. *Final heavy-atom parameters**

Site	*x*	*y*	*z*	*B* (Å)²	Fractional occupancy
		Uranyl derivative			
1	0.1420 (20)	0.1418 (21)	0.2008 (30)	30.0†	0.116 (12)
2	0.0955 (10)	0.2396 (11)	0.3211 (15)	30.0†	0.220 (12)
3	0.0000†	0.9030 (15)	0.5000†	30.0†	0.232 (18)
		HMN derivative			
1	0.4753 (5)	0.4506 (6)	0.0054 (7)	30.0†	0.274 (8)
2	0.1300 (3)	0.1269 (3)	0.2363 (5)	30.0†	0.437 (8)
3	0.0593 (4)	−0.0071 (5)	0.4562 (6)	30.0†	0.345 (9)
		Lead derivative			
1	0.1493 (2)	0.1314 (3)	0.2138 (4)	33.1 (2.5)	0.319 (7)
		Platinum derivative			
1	0.0609 (2)	−0.0033 (3)	0.4468 (3)	43.1 (3.4)	0.880 (23)
2	0.1836 (3)	0.2841 (3)	0.1514 (4)	40.6 (4.2)	0.636 (21)
		Mersalyl derivative			
1	0.0536 (4)	−0.0016 (4)	0.4601 (5)	39.9 (4.9)	0.428 (16)

* Standard deviations, in units of the last decimal place, are given in parentheses.
† Parameters that were not refined.

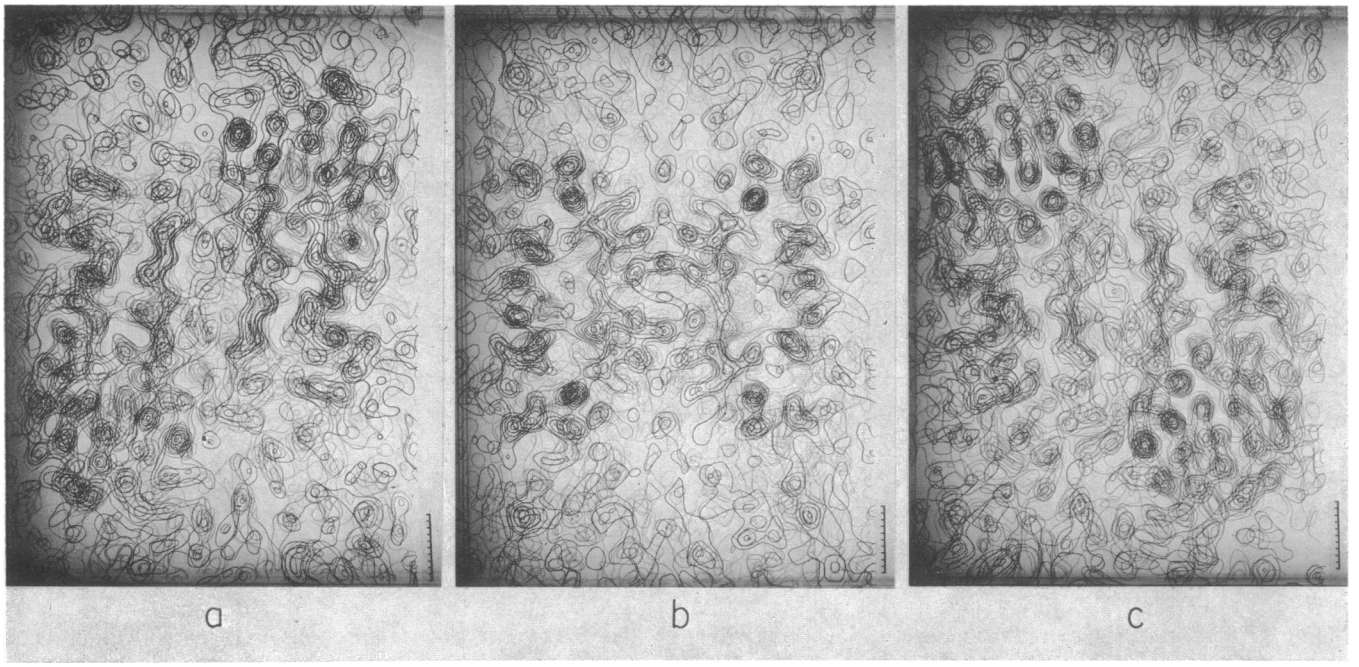

FIG. 2. Portion of the Con A map as viewed down the y axis with x vertical ($0 \leqslant x \leqslant 1$) and z horizontal ($-1/2 \leqslant z \leqslant +1/2$). Fig. 1 includes the right-hand half of each part of Fig. 2, as viewed from the right in Fig. 2. The plane $z = 0$ is vertical and passes through the center of each photograph. Each scale division = 1 Å. The tetramers are seen as pairs of dimers arranged one to the left and one to the right about a point of D_2 symmetry at the center of *b*. The tetrahedral shape of the tetramers can be visualized by imagining the parts of the figure to be stacked on top of one another. (*a*) $y = 0.39$ to 0.48. Regions of possible β structure are visible as parallel polypeptide chains. (*b*) $y = 0.48$ to 0.57, (*c*) $y = 0.57$ to 0.66.

the mersalyl and platinum derivatives using single isomorphous replacement (SIR) phases from the lead derivative, and of each derivative using multiple isomorphous (MIR) phases from the other two. The lead MIR map was quite clean, the height of the second largest peak being 42% that of the main site. In contrast, the other maps contained many additional maxima: the mersalyl map had six additional maxima greater than half the height of the main peak, and the platinum map had 39. None of these peaks was consistent with the difference Patterson functions or the Rossmann functions, so they were ignored.

The uranyl and *o*-hydroxymercuri-*p*-nitrophenol (HMN) derivatives were added to the data sets after the first three derivatives had been refined. Difference Patterson functions were not calculated for these derivatives; instead, sites were located from difference Fourier syntheses calculated with refined phases from the first three derivatives. The general correctness of these phases was indicated by the fact that they had already given a reasonable electron density map of Con A.

Refinement of heavy-atom parameters

Protein phases were calculated by the method of Blow and Crick (25). Integrations were performed at 5° intervals. Cycles of least-squares refinement and phasing were alternated in the usual manner. The quantity minimized for each derivative was $R = \Sigma_{\mathbf{H}} w_{\mathbf{H}} (kF_D(\mathbf{H})^2 - |\mathbf{F}_N(\mathbf{H}) + \mathbf{f}_c(\mathbf{H})|^2)^2$, where $\mathbf{H}$ is a scattering vector; $w_{\mathbf{H}}$ is a weight (see below); k is the scale factor; $F_D(\mathbf{H})$ is the observed derivative structure factor for reflection $\mathbf{H}$; $\mathbf{F}_N(\mathbf{H})$ is the phased native structure factor for reflection $\mathbf{H}$; and $\mathbf{f}_c(\mathbf{H})$ is the calculated heavy-atom contribution to $F_D(\mathbf{H})$.

Estimated standard errors were assigned to the reflections during initial processing based on counting statistics plus a constant fraction of the net intensity. The final standard error assigned to each reflection was either the error estimated from counting statistics or the standard deviation in averaging symmetry-equivalent reflections, whichever quantity was larger. The least-squares weights were the reciprocals of the squares of the estimated standard errors.

Initial values for the lack-of-closure errors (E) were estimated from the values found for the lead SIR calculation. Lack-of-closure errors and R-factors were tabulated during refinement for five equal ranges of $\sin^2\theta/\lambda^2$ (Table 2).

Scale factors for each derivative were initially estimated from the average ratios of the observed structure factors (native:derivative), with allowance for heavy-atom scattering. The absolute scale assigned to the native data was arbitrary but was estimated from the Wilson plot of the native data assuming the presence of 9750 carbon atoms/unit cell. The values reported for the populations of the heavy-atom sites are of course dependent on this choice of absolute scale.

The Pb, Pt, and Hg derivatives were first refined for six cycles in which positional and population parameters alone were refined. During these cycles various weighting schemes were tested. The exact choice of weighting scheme did not affect the refinement strongly. Eight more cycles were then performed, during which isotropic thermal parameters for each site were also allowed to vary. At this point the overall figure of merit was 0.734 and a native electron-density map was computed. This map displayed some signs of phasing errors, such as negative regions around the heavy-atom sites. Therefore the uranyl and HMN derivatives were added to the refinement. First, four cycles of refinement were performed in which the parameters of the original three derivatives were held constant. Then five cycles were performed in which parameters of all derivatives were allowed to vary. During

these cycles, the isotropic thermal parameters of several of the new sites refined to unreasonably large positive or negative values. The thermal parameters of uranyl and HMN were then set to 60.0 Å^2 (approximately the value of the average of the Pb, Pt, and Hg B's at that stage), and an additional eight cycles were performed. At this point convergence had been reached, but it was felt that other efforts should be made to achieve reasonable values for the uranyl and HMN thermal parameters. Therefore four cycles were performed in which reflections with $d \geqslant 12.5$ Å were omitted from the calculations. This resulted in a slight improvement in the figure of merit. The omission of the inner reflections also led to much more reasonable values of the Pb, Pt, and Hg thermal parameters (about 30 Å^2), so an additional four cycles were performed in which the uranyl and HMN thermal parameters were set to 30.0 Å^2 and positional, population, and scale parameters were refined with the inner reflections omitted. After convergence was reached, further attempts to refine the uranyl and HMN thermal parameters again failed, so refinement was terminated. A final cycle of phase determination was performed in order to phase the omitted inner reflections. The final overall figure of merit is 0.791. The average lack-of-closure errors, observed structure factors, Kraut R-factors, and figures of merit for the five ranges of $\sin^2\theta/\lambda^2$ are given in Table 2. The final heavy-atom parameters are listed in Table 3. Because of the difficulty with refinement of the thermal parameters and the small range of $\sin^2\theta/\lambda^2$ represented, no attempt was made to refine anisotropic thermal parameters.

Early in the refinement process, Wilson plots were made of each of the proteins. When plotted on a common set of axes, these curves showed that all of the derivatives were quite similar to the native protein in radial intensity distribution, except for the platinum derivative, which fell off more rapidly than the others with increasing scattering angle. All the plots showed strong peaks at 10 and 4.5 Å. To assess the effect of differing angular intensity distributions of the derivatives on the course of refinement, we divided the data sets into 12 ranges of $\sin^2\theta/\lambda^2$ and multiplied by a separate scale factor in each spherical shell so as to bring them onto the same scale as the native data. Least-squares refinement of heavy-metal parameters using the rescaled data did not give significantly different results from those obtained by the ordinary refinement.

ELECTRON-DENSITY MAPS AND DISCUSSION

Electron-density Fourier maps of the native protein were computed for the region $0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1, 0 \leqslant z \leqslant 0.5$ (four asymmetric units) at a scale 1 cm = 1 Å in sections at intervals of 0.02 down the c axis, for a similar region in sections down the b axis, and at a scale of 2 cm = 1 Å for a region encompassing one protein subunit.

The molecular boundary is clearly discernible in these maps (Fig. 1). Each subunit forms a roughly half-ellipsoidal volume or "gumdrop" whose altitude is parallel to the 110 direction, with a base of 46 × 26 Å, and a height of roughly 42 Å. The bases of pairs of two such subunits are in close contact (van der Waals distance) over an area of about 1200 Å^2 across crystallographic twofold axes parallel to the c axis. The dimers are thus shaped like elongated ellipsoids. Pairs of dimers interact at a greater distance across points of D_2 symmetry, for example at ($^1/_2$, $^1/_2$, 0), to form roughly tetrahedral tetramers (Fig. 2). We infer that the tightly-bound dimers represent the molecules of mol wt 55,000 found at acidic pH values.

The molecular structure of the subunits is not yet clear at this resolution. Large portions of the polypeptide chain can be clearly traced; however, an unambiguous trace of the chain from one end to the other is not yet possible. Sites of binding of the metals in the native protein and of the two methionyl residues might be expected to give large peaks, but these cannot be located by straightforward analysis of peak heights. The molecular surface does contain one large depression on each subunit, which could possibly be a binding site for saccharides.

We thank Dr. Fred Dodge and members of the Mathematical Sciences Department of IBM Corp. for help with computational aspects of the problem. This research was supported by USPHS grants GM-06920 and AI-09921 from the National Institutes of Health. A National Institutes of Health Postdoctoral Fellowship to Joseph W. Becker is gratefully acknowledged.

1. Boyd, W. C., *Vox Sang.*, **8**, 1 (1963).
2. Rieke, W. O., ed., *Proceedings of the Third Annual Leukocyte Culture Conference* (Appleton-Century Crofts, New York, 1969).
3. Inbar, M., and L. Sachs, *Nature*, **223**, 710 (1969).
4. Inbar, M., and L. Sachs, *Proc. Nat. Acad. Sci. USA*, **63**, 1418 (1969).
5. Shoham, J., M. Inbar, and L. Sachs, *Nature*, **227**, 1244 (1970).
6. Burger, M. M., and K. D. Noonan, *Nature*, **228**, 512 (1970).
7. Sumner, J. B., and S. F. Howell, *J. Bacteriology*, **32**, 227 (1936).
8. Goldstein, I. J., C. E. Hollerman, and J. M. Merrick, *Biochim. Biophys. Acta*, **97**, 68 (1965).
9. Goldstein, I. J., C. E. Hollerman, and E. E. Smith, *Biochemistry*, **4**, 876 (1965).
10. Goldstein, I., and L. L. So, *Arch. Biochem. Biophys.*, **111**, 407 (1965).
11. So, L. L., and I. J. Goldstein, *J. Biol. Chem.*, **242**, 1617 (1967).
12. So, L. L., and I. J. Goldstein, *J. Immunol.*, **99**, 158 (1967).
13. Sumner, J. B., and S. F. Howell, *J. Biol. Chem.*, **115**, 583 (1936).
14. Yariv, J., A. J. Kalb, and A. Levitski, *Biochim. Biophys. Acta*, **165**, 303 (1968).
15. Kalb, A. J., and A. Lustig, *Biochim. Biophys. Acta*, **168**, 366 (1968).
16. Kalb, A. J., and A. Levitski, *Biochem. J.*, **109**, 669 (1968).
17. Sumner, J. B., *J. Biol. Chem.*, **37**, 137 (1919).
18. Nakamura, S., and R. Sozuno, *Arch. Biochem. Biophys.*, **111**, 499 (1965).
19. Greer, J., H. W. Kaufman, and A. J. Kalb, *J. Mol. Biol.*, **48**, 365 (1970).
20. Hardman, K. D., M. K. Wood, M. Schiffer, A. B. Edmundson, C. F. Ainsworth, M. E. Hook, and K. R. Ely, *American Crystallographic Association Meeting, August 1970*, Paper F-9.
21. Sumner, J. B., N. Gralén, and I.-B. Ericksson-Quensel, *J. Biol. Chem.*, **125**, 45 (1938).
22. Wyckoff, H. W., K. D. Hardman, N. M. Allewell, T. Inagami, D. Tsernoglou, L. N. Johnson, and F. M. Richards, *J. Biol. Chem.*, **242**, 3749 (1967).
23. North, A. C. T., D. C. Phillips, and F. S. Mathews, *Acta Crystallogr.*, **A24**, 351 (1968).
24. Rossmann, M. G., *Acta Crystallogr.*, **13**, 221 (1960).
25. Blow, D. M., and F. H. C. Crick, *Acta Crystallogr.*, **12**, 794 (1959).