

Crystal Structure of Glucoamylase from *Aspergillus awamori* var. *X100* to 2.2-Å Resolution*

(Received for publication, May 20, 1992)

Alexander Aleshin†, Alexander Golubev‡, Leonid M. Firsov‡, and Richard B. Honzatko§¶

From the ‡Department of Molecular Biology, Leningrad Nuclear Physics Institute, St. Petersburg 188350, Russia and the

§Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50011

The crystal structure of a catalytically active fragment of glucoamylase-I from *Aspergillus awamori* var. *X100* has been determined to a resolution of 2.2 Å. Twelve of its 13 α -helices are arranged into an " α / α -barrel." An inner core of six mutually parallel α -helices are connected to each other through a peripheral set of six α -helices. The peripheral helices are parallel to each other, but approximately antiparallel to the inner core of α -helices. The putative active site lies in the packing void of the inner set of helices. The last 30 residues of the enzyme comprise a separate domain containing 10 sites of *O*-glycosylation. Each instance of *O*-glycosylation involves a serine or threonine side chain linked to the α -anomer of a single mannosyl residue. The *O*-glycosylated domain is in an extended conformation, wrapping around the "waist" of the α / α -barrel. Two additional sites of *N*-glycosylation contribute well ordered glycosyl chains that lie in proximity to the belt of *O*-glycosylation. The model developed for glucoamylase is a rare and valuable structural example of a glycoprotein and an exo-acting amylolytic enzyme.

Glucoamylase (α 1 $\rightarrow$ 4-glucan glucohydrolase, EC 3.2.1.3) catalyzes the release of β -glucose from the nonreducing end of maltodextrin chains (1–3). Along with α -amylases, glucoamylases from various species of *Aspergillus* are used extensively in the industrial hydrolysis of starch (1). However in contrast to α -amylases, which are endo-acting (4) (releasing a series of α -maltodextrins from the inner part of a starch or maltodextrin molecule), glucoamylases are exo-acting. Glucoamylases most readily hydrolyze α 1 $\rightarrow$ 4 glucosidic bonds. In addition, glucoamylase can cleave α 1 $\rightarrow$ 6 linkages, but at rates 30- to 50-fold slower than those for α 1 $\rightarrow$ 4 linkages (5). The primary industrial use of glucoamylase is in the production of glucose syrup, which in turn serves as a feedstock for biological fermentations in the production of ethanol or in the production of high fructose sweeteners (1).

* This work was supported by Grant DMB-9018205 from the National Science Foundation, Grant 2-S07-RR07034-24 from the National Institutes of Health, and the Iowa State University Biotechnology Program. This is Journal Paper J-14784 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa; project 3040. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ The atomic coordinates and structure factors (code 1GLY) have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY.

¶ To whom correspondence should be addressed: Dept. of Biochemistry and Biophysics, 1210 Molecular Biology Bldg., Iowa State University, Ames, IA 50011.

For specific species of *Aspergillus*, which produce two to three major variants of glucoamylase, studies in molecular biology have revealed only one gene (6–8). The cloned genes of *Aspergillus niger*, *Aspergillus awamori* and *Aspergillus awamori* var. *kawachi*, each code for a N-terminal signal peptide along with a parent protein of 615 or 616 amino acid residues. After cleavage of the signal peptide and post-translational glycosylation, the resulting protein (called glucoamylase-I) appears in the extracellular medium. Other variants of glucoamylase (usually called glucoamylase-II in the literature) are of lower molecular weight and arise by limited proteolysis of the parent protein. All major variants of glucoamylase are indistinguishable in terms of their kinetics of hydrolysis of soluble maltooligosaccharides. However, only glucoamylase-I can hydrolyze raw starch granules (9, 10).

Glucoamylase-I has three functional domains (11): 1) a catalytic domain (residues 1–440), 2) an *O*-glycosylated or linker domain (residues 441–512), and 3) a starch binding domain (residues 513–616). We report here the 2.2-Å crystal structure of a proteolyzed form of glucoamylase-I from *A. awamori* var. *X100* (12), which contains the complete catalytic domain plus the N-terminal half of the *O*-glycosylated domain (residues 1–471). Hereafter we shall refer to this proteolyzed form of the enzyme as glucoamylase-II(471).

EXPERIMENTAL PROCEDURES

Glucoamylase-II(471) was prepared from *A. awamori* var. *X100* as described elsewhere (12). Growth conditions for the fungus were adjusted so that glucoamylase-I (*M*_r 85,000 by sodium dodecyl sulfate-polyacrylamide gel electrophoresis) was reduced in size by the action of fungal acid proteases to a mixture of glucoamylases of an average *M*_r of 72,000. The mixture of proteins migrated as two partially resolved bands on sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

In spite of the apparent heterogeneity of the material, single crystals of glucoamylase-II(471) could be grown reproducibly by vapor diffusion. A solution consisting of 15 mg/ml protein, 50 mM potassium phosphate (pH 5.95) and 13% (w/v) polyethylene glycol 6000 was equilibrated against a solution of 50 mM phosphate (pH 5.95) and 20% (w/v) polyethylene glycol 6000. Crystals (space group P2₁2₁1, *a* = 116.7, *b* = 104.3, and *c* = 48.49) appeared in 1–2 weeks and grew slowly to a size of 0.2–0.3 mm. Crystals of glucoamylase-II(471) were stable indefinitely in 50 mM potassium phosphate (pH 5.95) and 28% (w/v) polyethylene glycol 6000.

Heavy atom derivatives were obtained by soaking crystals of glucoamylase-II(471) in solutions of UO₂(acetate)₂, K₃UO₂F₆, and Pt(ethyldiamine)Cl₂. Potential derivatives were screened for significant changes in intensities by precession photography. In the preparation of the uranyl derivatives, crystals were transferred to 100 mM sodium acetate (pH 5.35) and 28% (w/v) polyethylene glycol 6000. The concentration of K₃UO₂F₆ was 40 mM and that of UO₂(acetate)₂ was 10 mM. Soaking times for each of the uranyl derivatives was 3 days. In the preparation of the platinum derivative crystals were soaked in 50 mM phosphate (pH 5.95) 28% (w/v) polyethylene glycol 6000 and 2.5 mM Pt(ethyldiamine)Cl₂ for a period of 7 days.

Data collection was done in the laboratory of Dr. Arthur Arnone,

Dept. of Biochemistry, University of Iowa, on a Hamlin area detector and at Siemens Analytical Inc., Madison, WI, on a Siemens detector (Table I). The R_{sym} value for data from the Hamlin system is high in shells beyond 2.8-Å resolution, especially in comparison to data collected on the Siemens system. In addition, the ratio of $\langle I \rangle / \langle \sigma(I) \rangle$ for data collected on the Hamlin system falls off more rapidly as a function of resolution than for data collected on the Siemens system. The difference in the quality of data from the two detectors probably stems from the collimation of the incident x-ray beam. The Hamlin system at the University of Iowa was fitted with a 0.5-mm collimator (the smallest available in the laboratory), which was not an optimal size for crystals of 0.2–0.3 mm. Data collected with the Siemens detector employed crystals measuring 0.3 mm and a 0.3-mm collimator. However, the redundancy of the data from the Hamlin system is high, and, as a consequence, the two native data sets merged with an R_{sym} of 4.4% to 2.2-Å resolution (Table I). Data sets for the heavy atom derivatives, although collected to a resolution of approximately 2.2 Å, were truncated to a resolution of 3.0 Å, thus avoiding complications due to the systematic differences in data obtained using two different collimator sizes. The greatest decay at high resolution exhibited by any data set was 15%.

All data sets were converted to absolute scale by using a Wilson plot (13). Native data from the Siemens system were scaled against native data from the Hamlin system using a local scaling procedure (14). The two native data sets were merged and then averaged. All derivative data sets were scaled, in turn, against the combined native data set using the same procedure of local scaling.

Isomorphous and anomalous difference Patterson maps were analyzed independently for each of the three derivatives. One major site was common to all three derivatives, thus immediately permitting the choice of the same origin for the unit cell for all three derivatives. Major sites were derived by inspection of difference Patterson maps, whereas sites corresponding to heavy atoms in low occupancy were located in difference Fourier maps and then checked against the original difference Patterson maps. The correct choice of hand for the heavy atom positions rested upon the improvement in phasing statistics by the inclusion of anomalous scattering data for the two uranyl derivatives. Anomalous scattering data were not used in the case of the platinum derivative.

The program PHARE was used in the refinement of heavy atom parameters and in the calculation of phase probabilities. Initially, parameters for each derivative were refined independently using data from centric zones. Additional refinement of heavy atom parameters then employed all data to a resolution of 3.0 Å. Anomalous scattering differences for the two uranyl derivatives were included in the final stages of refinement as were anisotropic thermal parameters for heavy atoms with occupancies higher than 0.25.

TABLE I
Data collection statistics

Data set	Number of measurements	Unique reflections	Unique reflections with $F > 2\sigma(F)$	Total possible reflections	R_{sym}^a	Resolution
						Å
Native-1 ^a	176,220	34,163	33,290	35,431	0.094	2.1
Native-2 ^b	57,095	28,373	28,128	35,431	0.058	2.1
Native-3 ^c		34,296	33,871	35,431	0.044	2.1
Uranyl-1 ^d	175,707	22,584	22,495	22,935	0.059	3.0
Uranyl-2 ^e	105,493	22,784	22,542	22,935	0.070	3.0
Pt ^f	49,140	22,720	22,508	22,935	0.049	3.0

^a Data collected on a Hamlin area detector in the laboratory of Dr. A. Arnone, Dept. of Biochemistry, University of Iowa, Iowa City, IA.

^b Data collected on a Siemens area detector a Siemens Analytical Instruments, Inc., Madison, WI.

^c Data sets Native-1 and Native-2 merged and averaged with weights $(1/\sigma(F))^2$.

^d $\text{UO}_2(\text{acetate})_2$ derivative. Friedel pairs collected on the Hamlin area detector.

^e $\text{K}_3\text{UO}_2\text{F}_6$ derivative. Friedel pairs collected on the Hamlin area detector.

^f Pt(ethyldiamine) Cl_2 derivative. Friedel pairs collected on the Siemens detector.

^g $R_{\text{sym}} = \sum_i \sum_j |I_{ij} - \langle I_j \rangle| / \sum_j \sum_i I_{ij}$, where the index i runs over multiple observations of the same intensity, and the index j runs over all crystallographically unique intensities.

An electron density map, calculated with "best" Fourier coefficients (mean figure-of-merit of 0.80 to a resolution of 3.0 Å), was drawn on page-sized transparencies, in order to facilitate the tracing of the polypeptide chain and the assignment of α -carbons to residues with large side chains (tryptophans and tyrosines). Although the amino acid sequence for glucoamylase-I from *A. awamori* var. *X100* is unavailable, the first 200 bases of a partial DNA sequence of the gene coding for glucoamylase-II(471) are identical to the glucoamylase-I gene from *A. awamori* var. *kawachi*.¹ On the strength of this strong similarity, we have used the sequence of glucoamylase-I from *A. awamori* var. *kawachi* to model the electron density. We found that each of five positions of disagreement between the sequence of glucoamylase-I from *A. awamori* var. *kawachi* and the electron density could be resolved by choosing the corresponding amino acid from the sequence of glucoamylase-I from *A. niger* (6, 11). Close agreement between the electron density of glucoamylase-II(471) and the sequences of glucoamylase from *A. awamori* var. *kawachi* and *A. niger* is expected, as all of the known sequences of glucoamylase-I from species of *Aspergillus* are identical in 95% of their positions (6–8).

A preliminary model of glucoamylase-II(471) was built using the program TOM (15) and a Silicon Graphic 4D25. The high mannose glycosyl chain (16) was used as a guide in modeling the *N*-glycosyl chains of glucoamylase, whereas α -D-mannose was used to represent the density of *O*-glycosyl residues, in accordance with the findings of Gunnarsson *et al.* (17). Standard groups for the glycosyl residues (α -D-mannose and *N*-acetyl-D-glucosamine) were based on the coordinates of monosaccharides obtained by x-ray diffraction studies (18, 19). The glycosidic bond angles were set initially to 115°.

The preliminary model was subjected to 37 cycles of restrained least squares refinement using the program PROLSQ (20). Standard groups were developed for serine-mannose, threonine-mannose, and asparagine-glycosyl chains. All glycosyl residues were restrained to a C₁-chair conformation by specification of all two-, three-, and four-center distance restraints. Four-center distance restraints involving the 6-OH atoms of each glycosyl residue and the atoms of the *N*-acetyl group of *N*-acetyl glucosamine, however, were omitted, as well as four-center distance restraints that influence the dihedral angles associated with the oxygen atoms of glycosidic bonds. The preliminary model (x , y , and z coordinates for each atom and an overall thermal parameter) was refined initially against data to 3.0-Å resolution, then 2.5-Å resolution, and finally 2.2-Å resolution. The last cycles of refinement employed x , y , and z coordinates and a thermal parameter for each atom. The refinement was interrupted at several stages, in order to rebuild side chains that deviated significantly from a staggered conformation or that had close-nonbonded interactions with nearby atoms. Manual refitting of the model employed difference Fourier maps based on coefficients $(2^*F_{\text{obs}} - F_{\text{calc}})\exp(i\alpha_{\text{calc}})$ and $(F_{\text{obs}} - F_{\text{calc}})\exp(i\alpha_{\text{calc}})$.

RESULTS AND DISCUSSION

The model for glucoamylase-II(471) has been deposited with the Protein Data Bank, Brookhaven National Laboratory. For the purposes of this and subsequent communications, we have adopted the amino acid numbering of glucoamylase from *A. niger* (6, 11). Thus, the sequence numbers given here for residues beyond 101 are larger by one unit, than the sequence numbering reported by Hayashida *et al.* (8) for glucoamylase-I from *A. awamori* var. *kawachi*. We have adopted the sequence numbering of glucoamylase from *A. niger*, simply because it facilitates the comparison of the present structure to the literature.

Results of the heavy atom refinement and phasing by multiple isomorphous replacement appear in Tables II and III. Fourteen of the 15 heavy atom sites comprising the three derivatives are in proximity to side chains that have binding affinity for the heavy atom. Sites 1 of the uranyl acetate, uranyl fluoride, and platinum derivatives are all near each other and in proximity to Asp⁴⁰⁶ and Asp⁴⁰⁸. Sites 2 of uranyl acetate and uranyl fluoride are the same, being near Glu²⁸⁰, and sites 3 of the two uranyl derivatives are also close to each other, in proximity to Glu⁴³⁹ and to the 2-OH of the *O*-linked

¹ L. Firsov, unpublished results.

TABLE II
Refined heavy atom parameters

Heavy atom site ^a	Coordinates ^b			Occupancy, ^c real/anomalous	Thermal parameters ^d					
	x	y	z		β_1	β_2	β_3	β_4	β_5	β_6
Uranyl-1										
1	−0.1901	0.1398	0.0696	0.66/0.73	0.00076	−0.00035	0.00005	0.00124	0.00133	0.00296
2	−0.4886	0.1701	0.1116	0.27/0.31	0.00053	0.00006	0.00122	0.00110	0.00005	0.00352
3	−0.0343	0.3680	0.3651	0.13/0.14	23					
4	−0.1725	0.3254	0.1426	0.31/0.34	0.00008	−0.00014	0.00164	0.00080	0.00254	0.00300
5	−0.4864	0.2727	0.0553	0.06/0.06	21					
6	−0.4715	0.1892	0.1860	0.09/0.09	41					
7	−0.1715	0.6902	0.4026	0.18/0.21	0.00014	0.00023	−0.00086	0.00059	−0.00048	0.00097
Uranyl-2										
1	−0.1897	0.1366	0.0798	0.33/0.38	0.00129	−0.00006	0.00013	0.00234	0.00005	0.00568
2	−0.4887	0.1700	0.1099	0.73/0.82	0.00051	0.00000	0.00068	0.00087	−0.00029	0.00348
3	−0.0344	0.3682	0.3645	0.40/0.46	0.00025	−0.00025	0.00037	0.00002	0.00094	0.00028
4	−0.3971	0.9301	0.1958	0.09/0.09	54					
5	−0.3715	0.9361	0.2697	0.09/0.08	51					
Pt										
1	−0.1832	0.1408	0.0606	0.73/-	0.00172	0.00106	0.00012	0.00171	0.00197	0.00489
2	−0.4773	0.2124	0.1542	0.39/-	0.00044	0.00043	0.00059	0.00076	0.00084	0.00442
3	−0.3459	0.8898	0.4108	0.25/-	0.00099	−0.00058	−0.00031	0.00100	−0.00038	0.00039
4	−0.1244	0.7533	0.3367	0.19/-	113					

^a Chemical names for the derivatives appear in footnotes to Table I.^b x, y, and z are in fractions of the unit cell lengths a, b, and c, respectively.^c Occupancies are in fractions of the real and anomalous form factors for uranium (91, 13) and fractions of the real form factor for platinum (78).^d β_1 to β_6 are parameters of an anisotropic temperature factor $T = \exp(-\beta_1 h^2 - \beta_2 hk - \beta_3 kl - \beta_4 k^2 - \beta_5 kl - \beta_6 l^2)$. β_1 alone is a parameter for an isotropic temperature factor $T = \exp(-\beta_1 \sin^2 \Theta / \lambda^2)$.TABLE III
Phasing statistics from multiple isomorphous replacement

	11.7	8.3	6.4	5.2	4.4	3.8	3.4	3.0	Total
Resolution (Å)	190	415	723	1,101	1,539	2,090	2,737	3,455	12,250
Number of observations	0.89	0.90	0.89	0.89	0.88	0.84	0.78	0.68	0.80
Mean figure-of-merit ^a									
Uranyl-1 ^b									
Root mean square F_H^c	138.1	135.1	120.6	109.0	96.5	83.6	72.1	61.4	85.9
Root mean square residual ^d	30.5	25.6	26.1	24.0	26.9	28.7	29.0	28.7	28.0
Anomalous root mean square delta ^e	28.3	28.0	26.9	25.2	22.2	19.6	16.9	14.5	19.5
Anomalous root mean square residual ^f	10.7	13.0	12.1	10.7	11.3	12.5	15.7	18.7	13.5
Uranyl-2 ^b									
Root mean square F_H^c	155.8	148.7	137.4	126.2	112.4	99.0	88.8	78.5	102.1
Root mean square residual ^d	41.6	36.8	33.2	33.6	33.5	34.3	36.2	37.2	35.5
Anomalous root mean square delta ^e	35.4	33.2	31.3	28.9	25.4	23.6	20.9	18.7	23.4
Anomalous root mean square residual ^f	13.8	15.2	16.0	16.5	13.5	17.2	22.7	27.6	18.3
Pt ^b									
Root mean square F_H^c	130.2	116.8	106.3	87.9	72.7	61.7	50.4	39.2	66.4
Root mean square residual ^d	78.4	57.5	37.3	31.1	33.6	34.1	31.7	31.9	35.0

^a Calculated according to the definition of Blow and Crick (41).^b Chemical names of the derivatives appear in footnotes to Table I.^c Root mean square modulus of the structure factor for heavy atoms, based on the parameters of Table II.^d Root mean square difference in the moduli of observed and calculated structure factors for the derivative. The calculated structure factor for the derivative is $|F_P| \exp(i\alpha_{MIR}) + |F_H| \exp(i\alpha_H)$, where $|F_P|$ and $|F_H|$ are the moduli of the structure factors for the native crystal and for the heavy atoms of a particular derivative, α_{MIR} is the multiple isomorphous replacement phase angle for the native structure factor and α_H is the calculated phase angle for the structure factor of the heavy atoms.^e Root mean square difference in the moduli of Friedel pairs, based on the parameters of Table II.^f Root mean square difference in the calculated anomalous difference between Friedel pairs and the observed difference.

mannosyl residue of Ser⁴⁴³. Sites 4 and 5 of the uranyl acetate derivative are near Asp¹⁴⁴ and Asp²⁹⁷, respectively, whereas site 6 is within binding distance to both Glu²⁸⁰ and Asp³⁴⁹. Site 7 is close to Asp⁴⁷ and Tyr⁵⁰. Site 4 of the uranyl fluoride derivative is near Glu³⁸⁹ and Thr², whereas Site 5 is in prox-

imity to Glu³⁸⁹ alone. Sites 2 and 3 of the platinum derivative are close to Lys²⁷⁹ and His³⁹¹, respectively. However, site 4 of the platinum derivative is more than 4 Å from the nearest protein atom (side-chain of Thr⁴³). The thermal factor (113 Å²) for site 4 is also high relative to other heavy atom sites.

The platinum reagent may bind nonspecifically at site 4, but a more probable explanation is that the density here reflects a difference in the length of the glycosyl chain attached to Asn³⁹⁵ in the crystal of the platinum derivative relative to other crystals used in this study. Site 4 is only 5.48 Å from the 2-OH of residue D of the glycosyl chain at Asn³⁹⁵. Furthermore, the crystal employed in the platinum derivative came from a different preparation of the enzyme.

The model (Figs. 1 and 2) is an accurate representation of the conformation of the polypeptide chain. Electron density for the backbone of the polypeptide is continuous from the N to the C termini at a level of 3 σ . The density is weakest in the main chain for Ala¹ and Thr². Glucoamylase from the closely related strain of *A. niger* has 50% of the molecules missing the first 3 residues from the N terminus (10). Low density is present for side chains of Ala¹, Thr², Lys³³⁷, Lys³⁵², and Lys³⁷⁸. Some of the side chains listed above may be in two conformations, but only one conformation is presented in the current model. The electron density map and the amino acid sequence of glucoamylase-I from *A. awamori* var. kawachi disagree in five positions. The DNA sequence for *A. awamori* var. kawachi (8) calls for Val¹⁴¹, Arg²⁴⁰, Gln³¹⁹, Ser³²⁰, and Trp⁴⁴⁵, whereas we observe in the electron density Trp¹⁴¹, Ser²⁴⁰, Leu³¹⁹, Cys³²⁰, and Val⁴⁴⁵. As mentioned above the electron densities for these five positions conform to the sequence of glucoamylase from *A. niger* (6, 11). The amino

acid sequence used in modeling glucoamylase-II(471) appears in Fig. 3.

As would be anticipated, regions of high thermal parameters involve loops that connect regular secondary structure, as well as the highly glycosylated C-terminal span of 30 residues. However, no thermal parameter for a protein atom exceeds 40 Å². The solvent structure is not represented in this preliminary model nor is the modeling of the glycosylation complete (see below). The overall R-factor to 2.2-Å resolution is 0.23. Key statistics regarding the agreement between observed and ideal stereochemistry are in Table IV.

The Ramachandran plot (21) provides another illustration of the overall quality of the model. Only 2 nonglycine residues fall significantly outside of the allowed region of the Ramachandran plot (Fig. 4). Asn³¹³ ($\phi = 85^\circ$, $\psi = -10^\circ$) has a close contact (2.82 Å) between its β -carbon and the carbonyl oxygen of Tyr³¹². The conformation of Asn³¹³ is enforced by hydrogen bonds between carbonyl 312 and ND2 of Asn³¹⁵ (donor-acceptor distance of 2.84 Å), carbonyl 313 and NZ of Lys⁴⁰⁴ (donor-acceptor distance of 2.75 Å), amide 313 and carbonyl 310 (donor-acceptor distance of 2.75 Å), and amide 314 and carbonyl 311 (donor-acceptor distance of 2.75 Å). Ser⁴¹¹ ($\phi = -75^\circ$, $\psi = -170^\circ$) has a close contact (2.96 Å) between its β -carbon and amide 412. The close contact is enforced by hydrogen bonds between the OG of Ser⁴¹¹ and its amide nitrogen (donor-acceptor distance of 2.74 Å), between OG of Ser⁴¹¹ and OE2 of Glu⁴⁰⁰ (donor-acceptor distance of 2.68 Å), between carbonyl 410 and OG of Ser³⁹⁹ (donor-acceptor distance of 2.95 Å) and between carbonyl 411 and the 4-OH of residue F of the glycosyl chain attached to Asn³⁹⁵ (donor-acceptor distance of 2.85 Å). Interestingly, glucoamylases from other species have glycine at position 411 instead of serine.

The predominant type of secondary structure in glucoamylase-II(471) is the α -helix. Approximately 51% of all residues have ϕ and ψ angles that fall within $\pm 30^\circ$ of the typical values ($\phi = -64^\circ$, $\psi = -40^\circ$) for the α -helix. The 13 helical spans of glucoamylase-I, which average approximately 18 residues apiece, involve residues 1–20, 53–68, 72–89, 126–144, 148–168, 186–205, 211–227, 246–254, 272–286, 318–338, 345–354, 368–391, and 416–429. In contrast to the abundance of α -helices, there is little β -sheet structure. Our definition of a β -sheet requires that there be at least two hydrogen bonds between any given strand and its neighboring strand. Most of

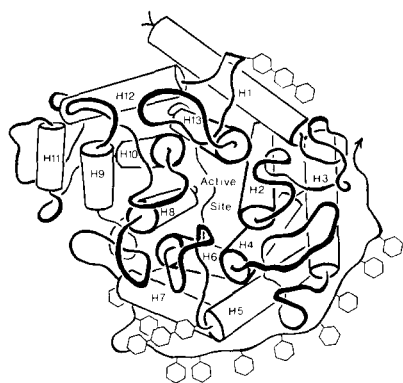


FIG. 1. Schematic of the glucoamylase-II(471). α -Helices are cylinders, sites of O-glycosylation are represented by single hexagons, and sites of N-glycosylation are symbolized by chains of three hexagons.

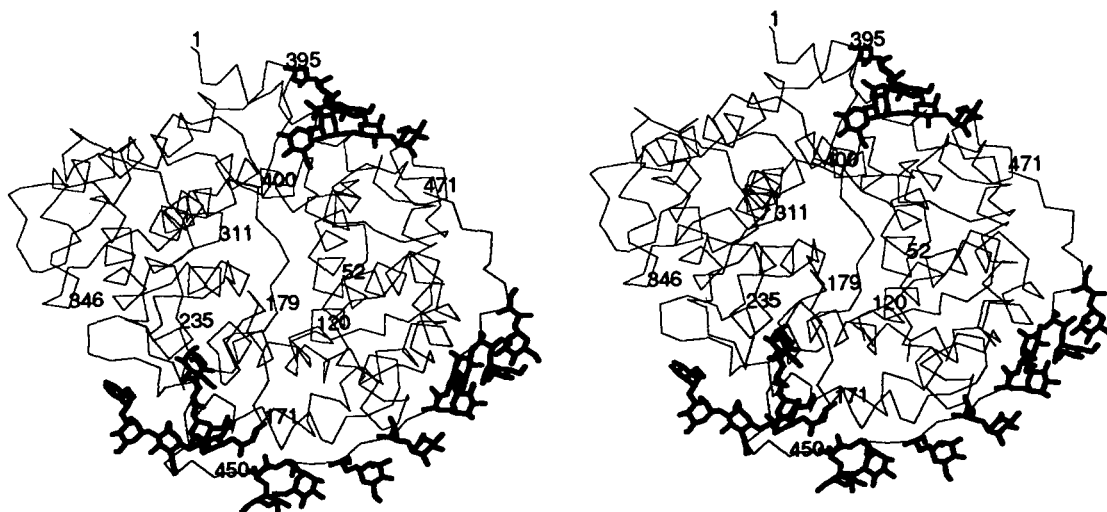


FIG. 2. Stereo view of an α -carbon plot of glucoamylase-II(471). Structures presented in bold lines are O- and N-linked glycosyl chains.

```

ATLDSWLSNEATVARTAILN 20
NIGADGAWVSGADSGIVVAS 40
PSTDNPDYFYFTWRDSSLVI 60
KTLVDLFRNGDITDLLSTIEH 80
YISSQAI IQGVSNPFGDLSS 100
G - GLGEPKFNVDETAYTGSW 120
GRPQRDGPALRATAMIGFGQ 140
WLLDNGYTSAAATEIVWPLVR 160
NDLSYVAQYWNQTGYDLWEE 180
VNGSSFFTIAVQHRALVEGS 200
AFATAVGSSSCSWCDSQAPQI 220
LCYLQSFWTGSYILANFDSS 240
RSGKDTNTLLGSIHTFDPEA 260
GCDSDSTFQPCSPRALANHKE 280
VVDSEFRSITYTLNDGLSDSEA 300
VAVGRYPEDSYINGNPWFLLC 320
TLAAAEQLYDALYQWDKQGS 340
LEITDVSLSLDFFKALYSGAAT 360
GTYSSTSSSTYSSTIVSAVKTF 380
ADGFVSIIVETHAASNGSLSE 400
QFDKSDGDELSARDLTWSYA 420
ALLTANNRRNSVPPPSWGET 440
SASSVPGTCAATSASGTYS 460
VTVTISWPSIVA 471

```

FIG. 3. Amino acid sequence used in the refinement of glucoamylase-II(471). The single letters correspond to amino acids as established by the IUPAC-IUB Commission on Biochemical Nomenclature, Biochemistry 9, 3471-3479.

TABLE IV

Agreement between target and observed stereochemistry for glucoamylase

Target root mean square deviations given in parentheses.

Root mean square deviation (in Å) for distances	
Two-center	0.014 (0.02)
Three-center	0.029 (0.03)
Four-center	0.039 (0.05)
Root mean square deviation (in Å) for planar groups	0.013 (0.02)
Root mean square deviation (in Å ³) for chiral volumes	0.138 (0.150)
Root mean square deviation (in Å) for non-bonded contacts	
Contacts determined by single torsion angles	0.222 (0.25)
Contacts determined by multiple torsion angles	0.157 (0.25)
Root mean square deviation (in degrees) for torsion angles	
Planar peptides ($\omega = 180$)	2.3 (3)
Staggered side-chains	13.9 (15)

the sheet structures in glucoamylase-II(471) are hairpin loops connecting adjacent α -helices. These loops pack together with only one or no hydrogen-bond interactions. Residues 49-52, 107-110, and 113-116 define three β -strands that come together to form an antiparallel sheet. The other sheet structures are two-stranded, antiparallel β -sheets: 1) residues 21-23 and 35-38, 2) residues 91-93 and 96-97, 3) residues 173-176 and 181-183, 4) residues 184-185 and 235-237, 5) residues 262-264 and 267-268, 6) residues 339-344 and 360-365, 7) residues 392-393 and 396-397, and 8) residues 398-403 and 407-415. Only 16% of all residues belong to the β -sheet structures, described above.

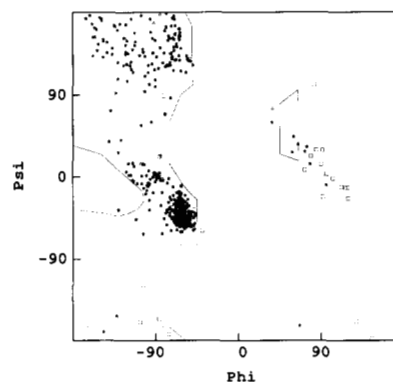


FIG. 4. Plot of ϕ and ψ angles of amino acid residues of glucoamylase-II(471). Open squares represent glycine residues, and open triangles represent the residues on the N- and C-terminal sides of *cis*-peptide linkages. Stars represent all other amino acids.

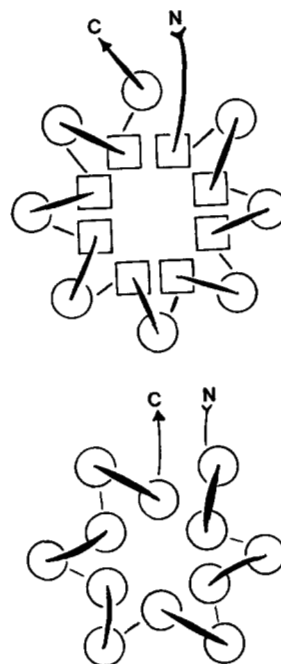


FIG. 5. Topology diagrams for the α/β -barrel (top) and the α/α -barrel (bottom). Helices are represented by circles, and β -strands are represented by squares.

All amylolytic enzymes with crystal structures reported in the literature (22-26) have the α/β -barrel-fold for their catalytic domains. Recent efforts to align the amino acid sequence of the catalytic domain of glucoamylase with those of other amylolytic enzymes have not revealed regions of similarity (27), not a surprising result in light of the difference between the polypeptide folds of glucoamylase and other amylolytic enzymes. Glucoamylase has a nontrivial variation of the α/β -barrel, which we call an α/α -barrel (Fig. 5). The α/α -barrel has a core of six, mutually parallel α -helices. The inner core of helices are connected to each other through a set of six peripheral α -helices. The set of peripheral helices are mutually parallel to each other but approximately antiparallel to the set of helices at the core.

At the level of topological folding, the α/β -barrel can be transformed into an α/α -barrel by omitting the first and last strands of the β -barrel, as well as its last two helices, and replacing the remaining six β -strands with a set of α -helices (Fig. 5). However, there is an additional difference between the α/β - and α/α -barrels which is not revealed by a topology



FIG. 6. Stereo view of an α -carbon plot of glucoamylase-II(471). Positions marked with solid circles represent sites of directed mutation and/or chemical modification that result in significant changes in catalytic activity. Bold lines represent the five regions of sequence similarity observed amongst glucoamylases from *Aspergillus*.

diagram. The active sites of enzymes possessing the α/β -barrel are comprised of loops bridging the C termini of the β -strands to the N termini of the peripheral helices. On the other hand, the packing void between the inner core of six helices of glucoamylase forms what we believe is the active site (see below). One end of the active site is blocked off by residues 430–440, whereas the opposite end is open to the solvent and substrates. The active site of glucoamylase is comprised of loops connecting the N termini of the inner set of helices to the C termini of the peripheral helices. Thus, the active sites of glucoamylase and other amylolytic enzymes lie on opposite ends of their respective barrels.

The *O*-glycosylated segment (residues 440–470) is in an extended conformation. The polypeptide chain curves over the “waist” of the α/α -barrel much as a belt. The significance of the *O*-glycosylated domain and the distribution of glycosyl chains over the surface of glucoamylase-II(471) is discussed below.

In addition to the broad topological features of the polypeptide fold, there are local structural features of some significance. The structure has three disulfide bonds between residues 210 and 213, residues 262 and 270, and residues 222 and 449. The third linkage, listed above, latches the *O*-glycosylated arm to the main body of the protein. In addition to the disulfide linkages, we observe three *cis*-peptide linkages, between Gly²³-Ala²⁴, Asn⁴⁵-Pro⁴⁶, and Arg¹²²-Pro¹²³. Finally, Pro¹⁵⁷ and Pro²¹⁸ occur in the middle of separate α -helices, creating a modest kink in each helix. The tendency for proline to introduce kinks in helices has been noted for other proteins (28, 29).

Directed mutations of Trp⁵²→Leu, Arg⁵⁴→Lys, Asp⁵⁵→Asn, Asp³⁰⁹→Asn,² Trp¹²⁰→Phe, Asp¹⁷⁶→Asn, Glu¹⁷⁹→Gln, Glu¹⁸⁰→Gln, and Tyr³⁰⁶→Phe² have resulted in mutant enzymes with altered catalytic properties of glucoamylases from *A. niger* (30, 31) or *Saccharomycopsis fibuligera* (32). In addition studies in chemical modification have placed Trp⁴¹⁷ in the vicinity of the active site (33–35) and studies of Itoh *et al.* (36) and Svensson *et al.* (27) have identified five regions of sequence similarity among glucoamylases. The residues listed above, as well as the regions of sequence similarity, cluster in or near the packing void of the core of six α -helices (Fig. 6). On the strength of these data, we have assigned the active site to the packing void. The “active pocket” is in the shape

of a funnel with a 15-Å diameter at the lip and a depth of 10 Å. In Fig. 7 we illustrate the spatial relationships between residues of the putative active site. Polar side chains of residues Tyr⁴⁸, Tyr⁵⁰, Lys¹⁰⁸, Gln¹²⁴, Arg²⁴¹, Asp²⁴⁵, His²⁵⁴, Arg³⁰⁵, Asn³¹⁵, Trp³¹⁷, Glu⁴⁰⁰, and Gln⁴⁰¹ are in the “active pocket” and conceivably could play a role in catalytic hydrolysis of the glycosidic bond or in the binding of substrate. These residues, thus, are suitable targets for additional studies in directed mutation or chemical modification.

Proposed mechanisms for the cleavage of the glycosidic bond by glucoamylase call for the participation of a catalytic acid and a catalytic base. Sierks *et al.* (31) suggest that (a) Glu¹⁷⁹ functions as the catalytic acid, transferring a proton to the oxygen of the glycosidic linkage; (b) interactions between the substrate and Trp¹²⁰ stabilize the transition state; and (c) Asp¹⁷⁶ is the catalytic base. Interestingly, the x-ray structure indicates a strong hydrogen bond between OE2 of Glu¹⁷⁹ and the hydrogens on NE1 of Trp¹²⁰ and Trp⁵² (donor acceptor distances of 2.82 and 2.91 Å, respectively). However, Asp¹⁷⁶ is buried under the loop of the polypeptide chain extending from Tyr¹⁶⁹ to Ser¹⁸⁵. The current crystallographic structure does not have Asp¹⁷⁶ and Glu¹⁷⁹ on opposite sides of a cleft or pocket as would be consistent with their putative catalytic roles. Either Asp¹⁷⁶ and the entire loop from residues 169 to 185 undergoes a major conformational displacement in response to enzyme-substrate interactions or the assignment of Asp¹⁷⁶ and/or Asp¹⁷⁹ as the catalytic base and acid, respectively, may be incorrect. Without structural information on inhibitor-enzyme complexes, there remains a great deal of uncertainty regarding specific roles for residues in the putative active site of glucoamylase.

Studies of glycosylation have focused primarily on glucoamylase from *A. niger*. Svensson *et al.* (11) discovered 11 sites of *O*-glycosylation in the region of 440–471, as opposed to 10 sites found in the crystal structure. The electron density map indicates single mannosyl residues at Ser⁴⁴³, Ser⁴⁴⁴, Thr⁴⁵², Ser⁴⁵³, Ser⁴⁵⁵, Thr⁴⁵⁷, Ser⁴⁵⁹, Ser⁴⁶⁰, Thr⁴⁶², and Thr⁴⁶⁴. For a specific sample of glucoamylase from *A. niger* *O*-glycosylation occurs, in addition, at Ser⁴⁴¹, Ser⁴⁶⁵, and Ser⁴⁶⁸, but not at position 455 and Thr⁴⁵⁷ (11). Position 455 in glucoamylase from *A. niger* is isoleucine, but reasons for the other differences in the sites of glycosylation between the two enzymes are not readily apparent. The electron density associated with *O*-linked sites of glycosylation is generally strong and well resolved. However, for sites Thr⁴⁶⁴ and Thr⁴⁵² the density is

² B. Svensson, Chemistry Dept., Carlsberg Laboratories, unpublished results.

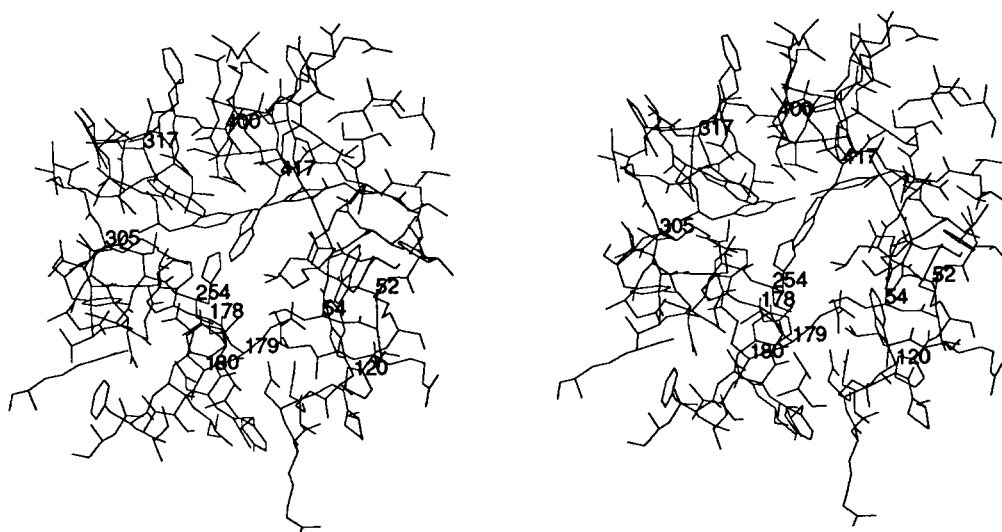


FIG. 7. Stereo view of the putative active site of glucoamylase-II(471). The orientation here is identical to that of Fig. 6.

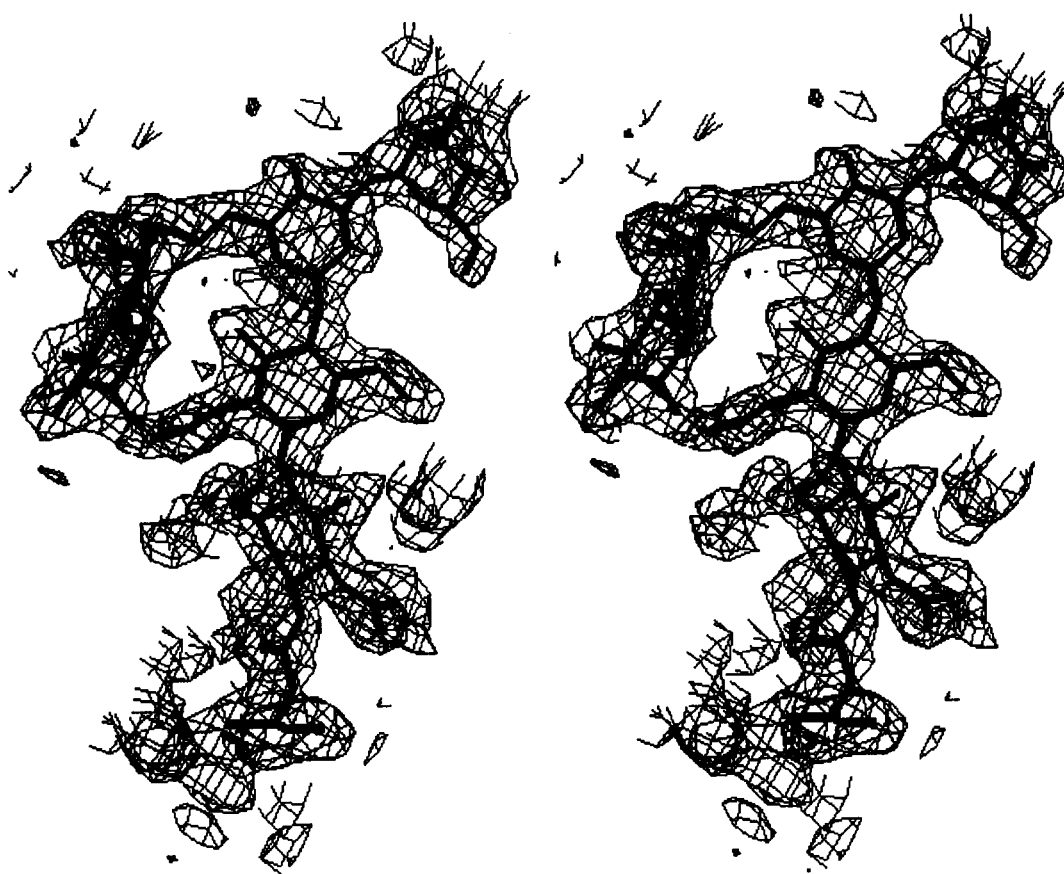
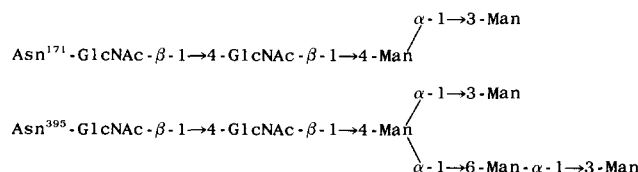


FIG. 8. Stereo view of the *N*-glycosyl chain at position 395 with associated electron density.

weak, indicating the possibility here of substantially less than 100% glycosylation. All sites of *O*-linked glycosylation are modeled as single mannopyranosyl units (C_1 chair conformation), with the OG of serine or threonine α -linked to the C1 position of the carbohydrate. We see no evidence at the present stage of the refinement for the addition of glycosyl residues beyond a single mannose unit, in contrast to the observed *O*-glycosylation of glucoamylase from *A. niger* (17), which indicates that approximately 75% of the *O*-linked glycosylation involves either 2 or 3 mannose residues. The dif-

ference in the glycosylation reported for glucoamylase from *A. niger* and the glycosylation observed in the crystal structure may arise from differences in the strain of *Aspergillus* and the growth conditions of the fungus, absent electron density due to disorder, and/or the segregation of the more complex glycosyl chains to the C-terminal half of the *O*-glycosylated domain. Several studies of glycosylation of glucoamylase from *A. niger*, for example, reveal significant variations in the composition of the *O*-glycosyl groups (17, 37, 38). Exhaustive refinement of the crystal structure, including a complete

Sites of *N*-glycosylation are two in number at Asn¹⁷¹ and Asn³⁹⁵, entirely consistent with the *N*-glycosylation observed for glucoamylase from *A. niger*. In each instance the ND2 of asparagine is β -linked to the C1 position of the first sugar residue (*N*-acetyl-D-glucosamine), followed by $\beta 1 \rightarrow 4$ linkages to the second (*N*-acetyl-D-glucosamine) and the third (D-mannose) residues. Electron density strongly suggests branching at the 3- and 6-hydroxyl positions of the mannose for site 395 (Fig. 8), but no branching at the mannose for the glycosyl chain linked to site 171. The structures of the two *N*-glycosyl chains as deduced from the electron density are as follows.



The C-terminal half of the *O*-glycosylated domain is absent in the crystal structure, presumably due to limited proteolysis. However, if the C-terminal half adopts an extended conformation similar to the N-terminal portion of the *O*-glycosylated domain, then the entire *O*-glycosylated domain would be a belt of glycosylation, connecting the two sites of *N*-glycosylation (Fig. 2). Thus, sites of glycosylation are localized, suggesting a specific functional role for these groups. Interestingly, when the *O*-glycosylated domain of a glucoamylase from yeast is fused with β -galactosidase, the resulting fusion protein appears in extracellular media (39). Thus, the *O*-glycosylated domain may be involved in transport, as well as serve as a structural linker between the catalytic domain and the starch binding domain. Finally, Hayashida *et al.* (40) have suggested that some of the starch binding properties of glucoamylase-I may reside in the *O*-glycosylated domain. Indeed, chemical modification studies (33–35) clearly indicate that the potent glucoamylase inhibitor, acarbose, and the substrate maltose will protect Trp²¹² and Trp⁴³⁷ from modification, which we know now to be near sites of *O*-glycosylation. The *O*-glycosylated domain, therefore, may well have some affinity for starch maltodextrin chains.

Addendum—Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeysens M., Béguin, P., and Aubert, J.-P. (1992) *Nature* **357**, 89–91, report recently that the endoglucanase CelD from *Clostridium thermocellum* also has an α/α -barrel similar to the one reported here for glucoamylase-II(471).

1. Saha, B. C., and Zeikus, J. G. (1989) *Starch/Stärke* **41**, 57-64
2. Weill, C. E., Burch, R. J., and Van Dyk, J. W. (1954) *Cereal Chem.* **31**, 150-158
3. Pazur, J. H., and Ando, T. (1960) *J. Biol. Chem.* **235**, 297-302
4. Robyt, J. H. (1984) in *Starch, Chemistry and Technology* (Whistler, R. L., BeMiller, J. N., and Paschall, E. F., eds) Ed. 2, pp. 87-123, Academic Press, New York
5. Pazur, J. H., and Kleppe, K. (1962) *J. Biol. Chem.* **237**, 1002-1006
6. Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K. E., and Fiil, N. P. (1984) *EMBO J.* **3**, 1097-1102
7. Nunberg, J. H., Meade, J. H., Cole, G., Lawyer, E. C., McCabe, P., Schweickart, V., Tal, R., Wittman, V. P., Flatgaard, J. E., and Innis, M. A. (1984) *Mol. Cell. Biol.* **4**, 2306-2315
8. Hayashida, S., Kuroda, K., Ohta, K., Kuhara, S., Fukuda, K., and Sakaki, Y. (1989) *Agric. Biol. Chem.* **53**, 923-929
9. Hayashida, S., Nomura, T., Yoshino, E., and Hongo, M. (1975) *Agric. Biol. Chem.* **40**, 141-146
10. Svensson, B., Pedersen, T. G., Svendsen, I., Sakai, T., and Ottesen, M. (1982) *Carlsberg Res. Commun.* **47**, 55-59
11. Svensson, B., Larsen, K., Svendsen, I., and Boel, E. (1983) *Carlsberg Res. Commun.* **48**, 529-544
12. Neustroyev, K. N., and Firsov, L. M. (1990) *Biokhimiia* **55**, 776-785
13. Wilson, A. J. C. (1942) *Nature* **150**, 151-152
14. Matthews, B. W., and Czerwinski, E. (1975) *Acta Cryst. A* **31**, 480-487
15. Cambillau, C., and Horjales, E. (1987) *J. Mol. Graphics* **5**, 174-177
16. Einsphar, H. M. (1989) *Trans. Am. Cryst. Assoc.* **25**, 1-22
17. Gunnarsson, A., Svensson, B., Nilsson, B., and Svensson, S. (1984) *Eur. J. Biochem.* **145**, 463-467
18. Chu, S. C., and Jeffrey, G. A. (1968) *Acta Cryst. B* **24**, 830-838
19. Longchambon, P. F., Avenel, D., and Neuman, A. (1976) *Acta Cryst. B* **32**, 1822-1826
20. Hendrickson, W. A., and Konnert, J. H. (1980) in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., and Venkutesan, K., eds) pp. 13.01-13.23, Indian Academy of Sciences, Bangalore
21. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) *J. Mol. Biol.* **7**, 95-99
22. Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L., and Woldike, H. F. (1990) *Biochemistry* **29**, 6244-6249
23. Buisson, G., Duée, E., Haser, R., and Payan, F. (1987) *EMBO J.* **6**, 3909-3916
24. Hofmann, B. E., Bender, H., and Schulz, G. E. (1989) *J. Mol. Biol.* **209**, 793-800
25. Klein, C., and Schulz, G. E. (1991) *J. Mol. Biol.* **217**, 737-750
26. Matsuura, Y., Kusunoki, M., Harada, W., and Kakudo, M. (1984) *J. Biochem.* **95**, 697-702
27. Svensson, B. (1991) *Depun Kagaku* **32**, 125-135
28. Richardson, J. S. (1981) *Adv. Protein Chem.* **34**, 167-339
29. Barlow, D. J., and Thornton, J. M. (1988) *J. Mol. Biol.* **201**, 601-619
30. Sierks, M. R., Ford, C., Reilly, P. J., and Svensson, B. (1989) *Protein Engineering* **2**, 621-625
31. Sierks, M. R., Ford, C., Reilly, P. J., and Svensson, B. (1990) *Protein Engineering* **3**, 193-198
32. Itoh, T., Sakata, Y., Akada, R., Nimi, O., and Yamashita, I. (1989) *Agric. Biol. Chem.* **53**, 3159-3167
33. Clarke, A. J., and Svensson, B. (1984) *Carlsberg Res. Commun.* **49**, 111-122
34. Clarke, A. J., and Svensson, B. (1984) *Carlsberg Res. Commun.* **49**, 559-566
35. Svensson, B., Clarke, A. J., and Svendsen, I. (1986) *Carlsberg Res. Commun.* **51**, 61-73
36. Itoh, T., Ohtsuki, I., Yamashita, I., and Fukui, S. (1987) *J. Bacteriol.* **169**, 4171-4176
37. Pazur, J. H., Tominaga, Y., Forsberg, L. S., and Simpson, D. L. (1980) *Carbohydr. Res.* **84**, 103-114
38. Lineback, D. R., and Aira, L. A. (1972) *Cereal Chem.* **49**, 283-296
39. Yamashita, I. (1989) *Agric. Biol. Chem.* **53**, 483-489
40. Hayashida, S., Nakahara, K., Kuroda, K., Miyata, T., and Iwanaga, S. (1989) *Agric. Biol. Chem.* **53**, 135-141
41. Blow, D. M., and Crick, F. H. C. (1959) *Acta Cryst.* **14**, 1195-1202