Three-Dimensional Structure of an Intact Glycoside Hydrolase Family 15 Glucoamylase from *Hypocrea jecorina*

Richard Bott,*,‡ Mae Saldajeno,‡ William Cuevas,‡ Donald Ward,‡ Martijn Scheffers,§ Wolfgang Aehle,§ Saeid Karkehabadi,|| Mats Sandgren,|| and Henrik Hansson||

Genencor-A Danisco Division, 925 Page Mill Road, Palo Alto, California 94304, Genencor-A Danisco Division, Archimedesweg 30, 2333 CN Leiden, The Netherlands, and Department of Molecular Biology, Swedish University of Agricultural Sciences, Biomedical Center, P.O. Box 590, SE-751 24 Uppsala, Sweden

Received December 10, 2007; Revised Manuscript Received April 1, 2008

ABSTRACT: The three-dimensional structure of a complete *Hypocrea jecorina* glucoamylase has been determined at 1.8 Å resolution. The presented structure model includes the catalytic and starch binding domains and traces the course of the 37-residue linker segment. While the structures of other fungal and yeast glucoamylase catalytic and starch binding domains have been determined separately, this is the first intact structure that allows visualization of the juxtaposition of the starch binding domain relative to the catalytic domain. The detailed interactions we see between the catalytic and starch binding domains are confirmed in a second independent structure determination of the enzyme in a second crystal form. This second structure model exhibits an identical conformation compared to the first structure model, which suggests that the *H. jecorina* glucoamylase structure we report is independent of crystal lattice contact restraints and represents the three-dimensional structure found in solution. The proposed starch binding regions for the starch binding domain are aligned with the catalytic domain in the three-dimensional structure in a manner that supports the hypothesis that the starch binding domain serves to target the glucoamylase at sites where the starch granular matrix is disrupted and where the enzyme might most effectively function.

Glucoamylases [GAs,[1] α-(1,4)-D-glucan glucohydrolase, EC 3.2.1.3] are multidomain exoglucohydrolases that catalyze the removal of glucose units from the nonreducing end of maltodextrin chains, e.g., starch and related oligosaccharides. The GAs act by hydrolyzing α-1,4 linkages between the glucose units in the maltodextrin polymer with an inversion of the anomeric configuration of the reactive sugar unit (*1, 2*), resulting in the release of mainly *β*-D-glucose monomers. The GAs are also capable of hydrolyzing α-1,6 linkages in maltodextrin polymers, but at a much lower rate compared to α-1,4 linkages (*3, 4*). The specific activity toward α-1,6 linkages in the substrate is only 0.2% compared to the activity toward α-1,4 linkages (*5*−*7*).

GAs are extensively used in industrial processes for the conversion of starch into glucose (saccharification), which is used for food and ethanol production (*8*). It has been suggested that by suppressing the activity of GAs on α-1,6 linkages in starchy materials, we can increase the current conversion rates (>95%) and produce higher yields in the industrial saccharification of starch (*9*).

The GAs have been classified into glycoside hydrolase (GH) family 15 of the carbohydrate active enzyme (CAZy) database (http://cazy.org; *10*, *11*). The GAs can be found among a wide range of different organisms such as fungi, yeasts, eubacteria, and archea (*12, 13*).

The GA from *Aspergillus awamori* var. *X100* (AaGA) has been one of the most extensively studied glucoamylases. The AaGA enzyme consists of an N-terminal catalytic domain (CD) followed by a linker domain and a C-terminal starch binding domain (SBD), which has been classified as a member of carbohydrate binding module (CBM) family 21 in the CAZy database. The AaGA CD, as well as all other GAs with known three-dimensional structures, folds into an $(\alpha/\alpha)_6$ barrel. AaGA has been studied biochemically in two forms: form I, which is the intact enzyme with its SBD attached, and form II that lacks the SBD. Both forms show comparable activity on soluble starch, while form I has up to 80-fold more activity on insoluble starch (*8, 14, 15*). The structure of the *Aspergillus niger* (AnGA) SBD has been determined by NMR, both in free form (*16*) and in complex with *β*-cyclodextrin (*17*). The SBD from AnGA has two binding sites oriented perpendicular to each other (*17, 18*), and it has been proposed that the primary function of the SBD is to localize the CD near nonparallel and thus more open starch to enhance the amylolytic rate (*19*). The three-dimensional structures of the AaGA CD in free form (*20*)

* To whom correspondence should be addressed. E-mail: rick.bott@danisco.com. Telephone: (650) 846-5832. Fax: (650) 845-6508.

‡ Genencor-A Danisco Division, 925 Page Mill Road, Palo Alto, California.

§ Genencor-A Danisco Division, Archimedesweg 30, 2333 CN Leiden, The Netherlands.

|| Swedish University of Agricultural Sciences.

[1] Abbreviations: Bis-tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; CD, catalytic domain; GA, glucoamylase; GH, glycoside hydrolase; Glc, glucose; mme, monomethyl ether; MR, molecular replacement; NAG, *N*-acetylglucosamine; PEG, polyethylene glycol; rmsd, root-mean-square deviation; SBD, starch binding domain; Tris; 2-amino-2-(hydroxymethyl)-1,3-propanediol; WT, wild type.

10.1021/bi702413k CCC: $40.75
Published on Web 05/06/2008

and in complex with the inhibitors acarbose and D-glucodihydroacarbose (*21*) have been determined by X-ray crystallography.

The structure of an intact GA, including a CD, a linker, and a SBD, has until now not been available. The role of the different domains of the GAs has been studied by comparing the activity of the catalytic core alone with that of the intact enzyme. It has been shown that the SBD increases enzyme activity on insoluble starch but does not affect activity on soluble substrates (*15, 22*).

The GA from *Hypocrea jecorina* (HjGA) is also a multidomain enzyme, consisting of an N-terminal CD (453 amino acid residues), followed by a linker (37 amino acid residues), and finally a CBM family 21 SBD (109 amino acid residues), being 55, 24, and 56% identical in sequence with the corresponding *A. awamori* glucoamylase domains, respectively. Its activity on various starch substrates is comparable to superior relative to that of the *A. awamori* enzyme (*23*). In this study, we present the three-dimensional apo structure of intact HjGA. The structure was determined by X-ray crystallography, in two different space groups, and refined at 1.85 and 1.90 Å resolution.

RESULTS AND DISCUSSION

Crystallization, Structure Solution, and Quality of the Final Models. HjGA was crystallized in an intact form containing 599 residues and all post-translational modifications such as glycosylations that would normally occur during its natural expression in *H. jecorina*. HjGA crystallizes in two different forms under similar crystallization conditions, depending on the presence or absence of Ca^{2+} during crystallization. One belongs to orthorhombic space group $P2_12_12_1$, and the other belongs to monoclinic space group $C2$. A low-resolution X-ray data set for the $P2_12_12_1$ crystal form was initially collected on a home X-ray source at room temperature, giving the following approximate unit cell parameters: a = 52.2 Å, b = 99.1 Å, and c = 121.3 Å. This corresponded to a calculated V_m of 2.3 Å^3/Da (*24*) and an estimated solvent content of 47%, consistent with there being one protein molecule in the asymmetric unit. This initial data allowed the HjGA structure to be determined by molecular replacement using the structure coordinates of AaGA [PDB entry 1GLM (*20*)] as a search model. While adjusting the molecular replacement model to include the HjGA sequence in the molecular replacement model consisting of residues 1−471, we observed additional contiguous electron density that did not correspond to the expected density for the HjGA CD. Initially, this density consisted of some extending segments of density between five and 11 amino acids and was interpreted to correspond to the linker and SBD of HjGA. As these segments were built, the model, including the disconnected chains, was refined. After refinement, it was possible to identify density that would extend and connect the segments. In this way, the entire HjGA structure model, including the linker as well as the SBD, was fitted to the electron density calculated from the low-resolution data. Subsequently, a high-resolution data set was collected at a synchrotron source to 1.8 Å on the orthorhombic crystal form, giving the same cell parameters as the low-resolution data. The structure model was refined to 1.85 Å and yielded final R_{work} and R_{free} values of 0.15 and 0.18, respectively. In the final calculated electron density map, density was visible for all residues and the final structure model contained 599 amino acid residues, 326 water molecules, 11 sugar residues, and two Bis-tris buffer molecules (discussed later). Table 1 shows data collection and refinement statistics for both the orthorhombic and the C-face-centered monoclinic crystal form, for which a high-resolution data set was collected to 1.9 Å at a synchrotron source on a cryo-cooled crystal. The cell parameters for the $C2$ crystal form were as follows: a = 216.8 Å, b = 48.5 Å, c = 50.6 Å, and β = 90.5°. The final refined R_{work} and R_{free} were 0.20 and 0.25, respectively, for the refined structure model, which contained 599 residues as well as 349 water molecules, six sugar residues, and one Bis-tris molecule. Both models contained the Bis-tris buffer molecule and also four cis peptides, three in the catalytic core and one in the SBD.

Table 1: Data Collection, Processing, and Refinement Statistics

(A) Data Collection and Processing		
beamline	i911-5, MAXlab[a]	i911-5, MAXlab[a]
wavelength (Å)	0.90718	0.90718
space group	$P2_12_12_1$	$C2$
unit cell parameters (Å)	a = 52.19, b = 99.23, c = 121.24	a = 216.76, b = 48.49, c = 50.56
unit cell angles (deg)	α = 90.00, β = 90.00, γ = 90.00	α = 90.00, β = 90.49, γ = 90.00
resolution range (Å)[b]	76.70−1.85 (1.95−1.85)	16.98−1.90 (2.00−1.90)
no. of unique reflections	53591 (7661)	41660 (6044)
average multiplicity	8.5 (7.4)	5.4 (4.9)
completeness (%)	98.3 (97.7)	99.8 (99.5)
mean $I/\sigma(I)$	20.2 (5.3)	19.7 (3.0)
R_{merge} (%)[c]	8.8 (45.0)	8.0 (39.8)
(B) Refinement		
PDB entry	2vn4	2vn7
resolution used in refinement (Å)	76.70−1.85	16.7−1.90
no. of reflections		
working set	50812	39555
test set	2728	2102
R_{work} (%)	15.5	20.1
R_{free} (%)	18.2	25.3
no. of residues in protein	599	599
no. of residues with alternate conformations	20	−
no. of protein atoms	4572	4537
no. of water molecules	326	349
no. of other atoms	157	87
average atomic B-factor (Å^2)		
overall	22.8	24.0
protein	21.3	23.8
waters	33.1	24.0
other atoms	41.8	31.5
rmsd for bond lengths (Å)[d]	0.010	0.010
rmsd for bond angles (deg)[d]	1.20	1.22
Ramachandran outliers (%)[e]	1.5	1.3

[a] Swedish National Electron Accelerator Laboratory for Nuclear Physics and Synchrotron Radiation Research (Max-Laboratory). [b] Numbers in parentheses are for the highest-resolution bins. [c] $R_{merge} = \sum_{hkl}\sum_i|I - \langle I\rangle|/\sum_{hkl}\sum_i|I|$. [d] From ref *37*. [e] According to the stringent boundary definition of ref *38*. Statistics for structure models were calculated with Refmac 5.0 (*32*), O (*34*), MOLEMAN (*39*), and LSQMAN (*40*).

Overall Structure. The final structure model corresponds to the entire mature HjGA molecule consisting of 599

residues. While the HjGA catalytic core and SBD shared approximately 50% sequence identity with the previously determined glucoamylase from *A. awamori*, the linker, 37 residues with nine prolines, was substantially less homologous with only 24% sequence identity. In contrast to the HjGA linker, the AaGA linker has only one proline and is comprised of 69 residues, 32 residues longer than HjGA. We anticipated that the HjGA linker might be sufficiently ordered to allow visualization of an intact glucoamylase with the linker and SBD. The electron density map of the refined HjGA allows fitting of all 599 amino acids present in the mature enzyme. Residues 1−453 comprise the catalytic core of the molecule, and residues 491−599 comprise the SBD, which are joined by a strand formed by residues 454−490, the linker region. In addition, we find density for several glycosylation sites and for two molecules of Bis-tris buffer.

Catalytic Core. The HjGA catalytic core segment follows the same $(\alpha/\alpha)_6$ barrel topology described in ref *20* for the AaGA, consisting of a double barrel of α helices with the C-terminus of the outer helix leading into the N-terminus of an inner helix. It was possible to identify key differences in the electron density such as the disulfide bridge between residues 19 and 26 and an insertion (residues 257−260) relative to AaGA. The segment comprising residues 80−100 also underwent extensive model rebuilding. One major glycosylation site was identified at Asn 171, which has up to four glycoside moieties attached. A similar glycosylation site was identified in AaGA. Additionally, the catalytic core contains three cis peptides between residues 22 and 23, 44 and 45, and 122 and 123, and these are conserved between HjGA and AaGA. Overall, there is a rmsd of 0.54 Å between 409 of 453 Cα atoms when comparing the coordinates of the catalytic cores of HjGA and AaGA [PDB entry 1GAI (*25*)].

Active Site. Density closely associated with Asp 54 (residue not labeled) was found at the active site in the HjGA structure. This density was interpreted to correspond to a Bis-tris molecule (Figure 1), which was used as a crystallization agent for the enzyme. The Bis-tris molecule is bound near the A subsite defined in ref *21* on the basis of their analysis of acarbose binding. In particular, the Bis-tris molecule forms hydrogen bonds with Asp 54, Arg 309, and the catalytic Glu 179, all of which are reported to hydrogen bond to the acarbose molecule in AaGA. It is interesting to note that a Tris molecule was found bound to the reactive center of the glucoamylase from *Saccharomycopsis fibuligera* (*26*). It may be that Tris and Bis-tris molecules are competitive inhibitors for glucoamylase enzymes and may not be suitable as buffers for kinetic studies of these enzymes. A second Bis-tris molecule was found bound at a lattice contact to the SBD in the orthorhombic form.

The active site of AaGA and the detailed interactions that the inhibitor acarbose makes with the AaGA enzyme have been described previously (*21, 25*). The residues forming the AaGA and HjGA substrate binding sites are compared in Figure 1 with HjGa shown with cyan carbon coloring and AaGA shown with yellow carbon coloring. In this figure, the catalytic glutamic acids 179 and 404 along with the stacking tryptophan and tyrosine residues (47, 51, 120, and 315) are highlighted. All residues that were identified in ref *25* to be involved in the acarbose and substrate binding in AaGA are conserved in the HjGA structure.

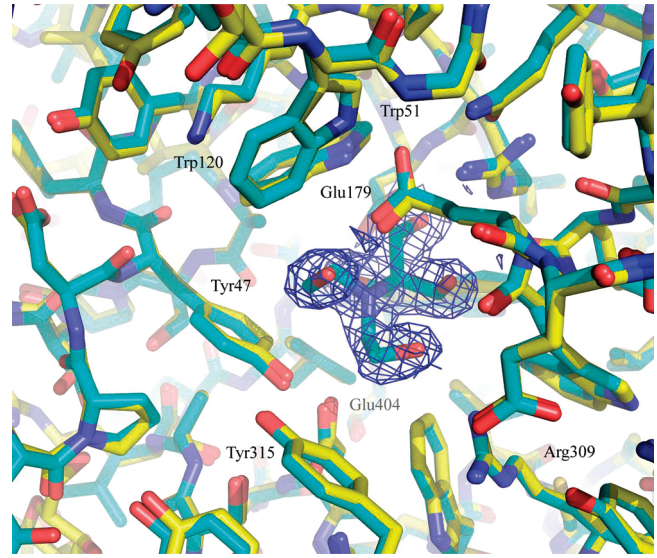

FIGURE 1: Electron density for a molecule of Bis-tris buffer bound at the active site of HjGA. The active sites of HjGA (cyan carbon coloring) and AaGA (yellow carbon coloring) are also compared. In AaGA, the residues comparable to Tyr 47, Trp 51, Trp 120, Glu 179, Arg 309, Tyr 315, and Glu 404 of the HjGA are reported to form key interactions with substrate analogues bound to the AaGA molecule. There are essentially no differences between side chain conformations of these two GAs at the catalytic site or in the substrate binding region as delineated by the acarbose binding to AaGA (see ref *25*).

Linker Region. The HjGA linker, residues 454−490, is defined as the segment spanning the region between two disulfide bridges: the last intradomain disulfide bridge of the CD, residues 222 and 453, and the first intradisulfide of the SBD, residues 491 and 587. As mentioned earlier, nine of the 35 residues (~25%) in the HjGA linker are prolines. The linker in HjGA extends from the back of the molecule in a wide arc followed by an abrupt turn after the lysine 477 residue in the Pro-Lys-Pro-Gly-Val-Pro sequence on the CD surface opposite the substrate binding site. The linker also contains four glycosylated residues, Thr 466, Thr 468, Thr 475, and Ser 482, each with one mannose bound. The linker extends as a random coil that is anchored by interactions of the side chains of Tyr 462, Pro 465, Phe 470, Gln 474, Pro 476, Lys 477, Val 480, and Tyr 486 with regions on the surface of the CD. A portion of the linker segment comprising residues 461−486 is shown in Figure 2, and this segment includes all eight anchoring residues, the four glycosylation sites, and seven of the nine proline residues. In Figure 2, the linker segment is represented as a stick model, extending over the CD represented by the gray surface to the SBD represented as the teal surface. Two of the glycosylated residues, Thr 475 and Ser 482, bracket the abrupt turn of residues 477−481. The electron density for this segment shown as an inset of Figure 2 can be seen to be continuous, and comparable density for the mannose moieties bonded to Thr 475 and Ser 482 can also be observed.

Along the linker there are eight residues with side chains that interact with residues of the CD. The side chains serve as anchors often filling cavities found along the surface of the CD. The OH group of the Tyr 462 side chain forms a hydrogen bond with Gln 168 OE1, while the ring atoms make van der Waals contacts with the side chains of Leu 98 and Gly 101. Pro 465 projects onto a hydrophobic surface patch

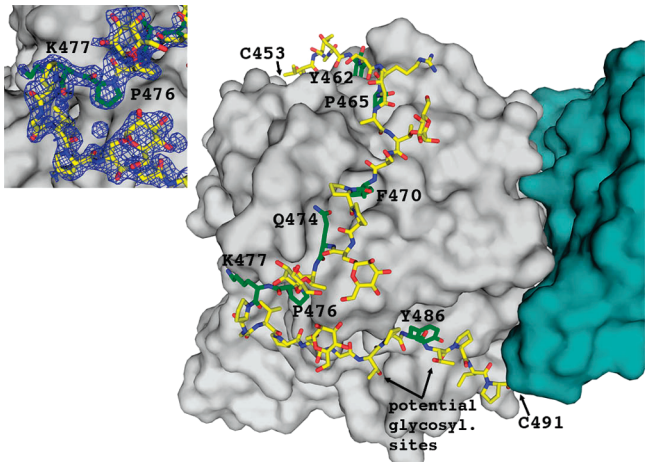

FIGURE 2: Interactions between the catalytic domain and linker region of *H. jecorina* GA. In this figure, the linker is shown as a stick figure interacting with the surface of the catalytic domain colored gray. The surface of the SBD is colored teal to differentiate it from the CD surface. The side chains of residues involved in the anchoring interactions (as described in the text) are color-coded in green. The electron density of residues 477−482 forming an abrupt turn in the linker is shown as an insert highlighting the quality of density for the linker and for the glycosylation at residues Thr 477 and Ser 582.

formed by the side chains of Val 86, Pro 157, and Ile 158. Between the anchoring side chains of Pro 465 and Phe 470, we find Thr 468. Thr 468 is fully glycosylated, projecting into the solvent and fixing the orientation to facilitate the interaction of the anchor residues with the surface of the CD. The side chain of Phe 470 projects into a hydrophobic patch created by the side chains of Ile 82, Leu 142, Thr 150, Val 154, and Ile 155. Gln 474 forms a hydrogen bond with the side chain of Thr 150. The OH group of Tyr 147 forms a hydrogen bond with the carbonyl O of Thr 475. The side chains of residues Pro 476 and Val 480 form hydrophobic interactions with the side chains of Tyr 70 and Trp 141 of the CD. These residues are found on a sharp bend extending from residue Thr 475, including Pro 476, Lys 477 (which hydrogen bonds with the carbonyl of Asn 144), Pro 478, and Gly 479, followed by Val 480 and ending with Ser 481. Both Thr 475 and Ser 481 are glycosylated, forming a hydrophilic canopy above the hydrophobic side chains of Pro 476 and Val 480. In addition to the glycosylation sites, proline residues comprise many of the surface-directed side chains in the segment between residues 461 and 486. The side chain of Tyr 486 stacks over a hydrophobic patch created by the side chains of Leu 14 and Leu 18. In Figure 2, the anchor side chains are colored green.

It is worth noting that the segment preceding the linker, residues 433−453, is also a random coil and appears in several ways to be quite similar to the linker domain. While it is tethered by the 222−453 disulfide bridge, it also shows a pattern of anchoring side chain interactions found with the linker. One could consider this region as a prelinker segment.

Starch Binding Domain. The SBD is composed of a β-sandwich of two twisted β sheets, tethered at one end by a disulfide bridge between Cys 491 and Cys 587 and at the other end having a series of loops that comprise a binding site for starch connected by long loops. The structure of HjGA SBD is quite similar to the averaged structure of the AnGA SBD determined by NMR (*17*) and the SBD of β-amylase from *Bacillus cereus* (*27*). Overall, the averaged

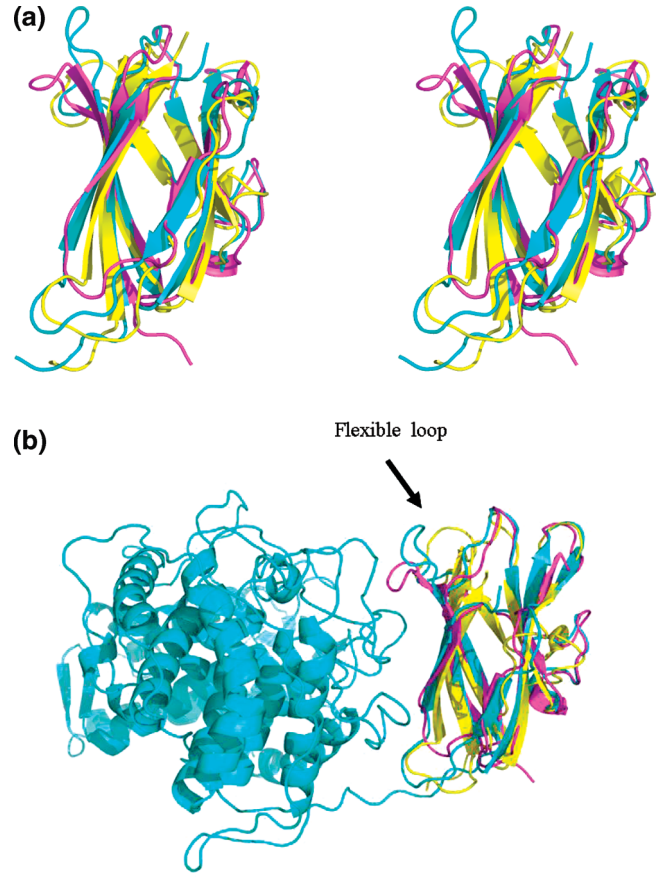

FIGURE 3: Comparison of SBD structures. (a) The three-dimensional folds of three SBDs are compared in a stereo cartoon representation. Colored blue is the *H. jecorina* SBD determined in this study, and overlaid in yellow and magenta are the starch binding domains of AnGA and that from the β-amylase from *B. cereus*, respectively. The overall agreement in the secondary structure and juxtaposition of strands from the three structures are very good. However, there is some variability in the loop regions. (b) The location of the variable loops is shown relative to the juxtaposition with the CD. The flexible loops are seen to be those nearest the catalytic domain.

SBD coordinates of *B. cereus* GA (PDB entry 1b90) can be aligned with the SBD of HjGA with a rmsd of 2.1 Å for 99 of a total of 109 Cα atoms. The SBD of *B. cereus* β-amylase (PDB entry 1AC0) is also aligned with a rmsd of 1.7 Å for 99 aligned Cα atoms. The secondary structure is highly conserved as shown in Figure 3a, where the SBD of HjGA in cyan is compared with an averaged NMR structure of the isolated SBD fragment from *A. awamori* (yellow) and the SBD of β-amylase from *B. cereus* (magenta). One loop stands out as being highly variable and is highlighted in Figure 3b, which shows the intact model, including the CD and the linker which are included to show the juxtaposition of the structural variable loop in the context of the entire structure. In HjGA, this loop corresponds to residues 537−543, whereas in *A. niger*, this loop corresponds to residues 554−560; in the β-amylase from *B. cereus* (colored magenta), this loop is shorter and corresponds to residues 462−465. In the NMR structure of β-cyclodextrin complexed to the isolated SBD of *A. niger* GA (*17*), the loop including residues Tyr 556 and Asp 560 is noted to shift substantially upon binding of β-cyclodextrin. The Asp 560 was noted to shift by more than 13 Å from its position in the uncomplexed SBD structure. Given that this loop stands out as being more

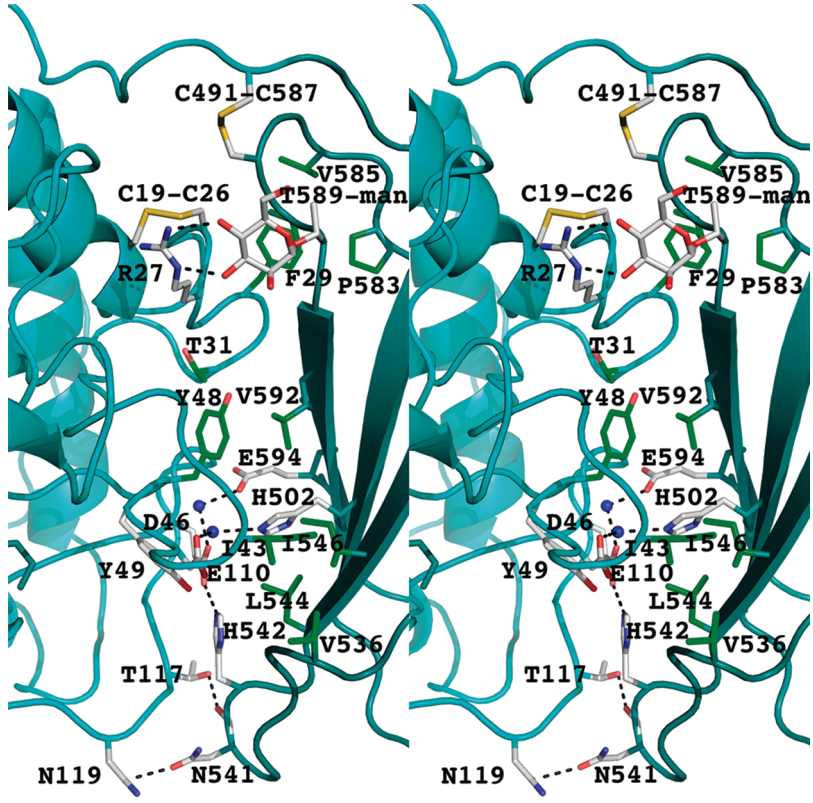

FIGURE 4: Interactions at the interface between the SBD and CD. Interacting residues are shown in stick representation; residues forming hydrophobic interactions are colored green, and residues involved in electrostatic or hydrogen bonding are shown with gray carbon coding. Potential hydrogen bonds are shown with dashed lines.

variable than others and that it undergoes a dramatic shift upon cylcodextrin binding, we designate this loop as the flexible loop (Figure 3b). This flexible loop forms part of "binding site 2" noted in ref *17*. These authors also identified a primary site "binding site 1" which shares similarities with other carbohydrate binding proteins. Overall, the degree of conservation of residues and even side conformations in the binding site 1 of these SBDs is very high. The side chains of Trp 525, Thr 527, and Lys 560 superpose perfectly (the alignment is based on 99 Cα atoms) on one another. The side chains of Glu 558, Val 570, Trp 572, and Asn 577 also superpose very well and would be positioned to make similar interactions with the starch chain. Five threonine residues (492, 524, 579, 581, and 589) on the HjGA SBD are glycosylated, and while most are found extending into the solvent, the O-linked mannose of Thr 589 participates in the interface contacts between the SBD and the CD as shown in Figure 4. The electron density for the mannose bound to Thr 589 is well-defined (not shown) and is positioned to form hydrogen bonds with the side chain of Arg 27 from the CD. This represents one of the numerous interactions that occur at the interface between the SBD and the CD as seen in Figure 4.

Figure 4 also illustrates that the CD comes in contact with one side of the SBD β-sandwich, which consists of three mixed strands. Most of the interactions between the CD and the SBD involve clusters of residues in two segments, residues 536−546 and residues 583−594. These segments include parts of the two outer strands on the side of the SBD β-sandwich. Only one residue from the middle strand, His 502, is involved in the contact with the CD. Overall, the catalytic domain−starch binding domain interactions are quite extensive and can be seen to involve clusters of hydrophilic or hydrophobic residues occurring as discrete areas of the interaction between the catalytic and starch binding domains.

The largest area involves hydrophilic residues, including His 502, which along with His 542 and Glu 594 forms the core of this large hydrophilic area or patch on the β-sandwich of the SBD. These residues interact with Asp 46, Tyr 49, and Glu 110 from the CD. Noteworthy among these is a potential salt bridge between residues Glu 110 on the CD and His 542 on the SBD. A solvent molecule mediates another potential salt link between Asp 46 on the CD and His 502 on the SBD. There is an adjunct section of hydrophilic interaction between the side chain of Asn 541 and the side chain of Asn 119 along with a hydrogen bond between the carbonyl O atom of Asn 541 and the side chain of Thr 117. In addition, we also find interactions between residues Pro 583, Val 585, and Val 588 from the SBD and Phe 29 from the CD near the N-terminus of the starch binding domain. Further, two more hydrophobic patches are found. One includes Val 536, Leu 544, and Ile 546 from the SBD, which interact with Ile 43, and a second, involving the side chain of Val 592 from the SBD, which serves as an anchor to the side chains of Thr 31 and Tyr 48 of the CD.

Taken together, there appears to be a common pattern for the interactions between the linker and SBD with the CD. The interaction is in the form of an anchoring side chain that interacts with the surface area of the neighboring domain. In general, the anchor residue is found on the linker segment. In the case of the interactions between the CD and SBD, the anchor residues can be contributed from either

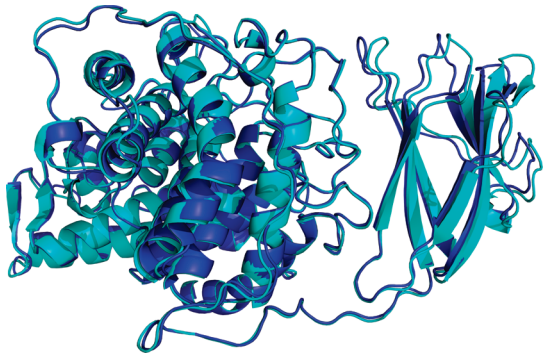

FIGURE 5: Comparison of HjGA in two crystal forms. The structures for HjGA determined in $P2_12_12_1$ (cyan) and $C2$ (dark blue) show the tight conservation of the catalytic and linker region. Between the two structure models only two variable loops are seen for the SBD. One of these is the variable loop seen between SBD from different enzymes and organisms. On the basis of this, we conclude that the juxtaposition between the catalytic domain and the starch binding domain that we see in both crystal forms is likely to be the same as in solution.

domain as in the case of residues Ile 43 and Phe 29, which come from the CD, or residue Val 592, which comes from the SBD.

Conservation of the SBD−CD Interactions. The juxtaposition found between the SBD and the CD in the $P2_12_12_1$ crystal form shows a compact organization. This could bear on the function of the SBD in the intact molecule, assuming that the conformation of the molecule we observe represents the one in solution. The alternative would be that the juxtaposition we observe is as a consequence of crystal packing. We were fortunate to obtain crystals of the intact enzyme in a second crystal form having a different space group and crystal packing arrangement. As one can see in Figure 5, when the structures from the two different crystal forms are compared, we find that the interactions between the CD and SBD are conserved and the only variation shown in Figure 5 is for isolated loops, including two in the region designated the "second binding site" in ref *17*. The two HjGA structures can be superimposed with a rmsd of 0.41 Å (599 aligned Cα atoms). If the conformation is the same in two different crystal forms, then it is a good indication that the observed structure is not an artifact of a particular crystal packing. There are loops in the SBD that are seen to vary between the structures determined in the two different crystal forms. One of these is the flexible loop that was seen to vary between the SBDs from different organisms (Figure 3b). The other is the loop formed by residues 563−569. Thus, while there is evidence of some conformational flexibility in the loops near binding site 2 on the SBD, the overall topology and juxtaposition of the SBD relative to the CD are taken to represent the solution structure of the intact enzyme.

The sequences of the residues forming the interface between the CD and SBD are variable for different glucoamylases. The residues involved in the interactions shown in Figure 5 are generally not conserved in the homologous positions. Complementary changes, however, are seen for closely related glucoamylases from *Humicola grisea* and *Talaromyces emersonii*. It is possible that for these enzymes, which also have comparable linker segment lengths, compared to HjGA, the interactions of the SBD with the respective CD would be similar. On the other hand, in the case of *A. awamori* GA, which has a longer linker segment, the SBD may be aligned differently. Reinforcing this view is the observation of an additional glycosylation site at Asn 395 on the CD of AaGA. This extensive glycosylation of AaGA would interfere with the SBD as it is seen to interact with the CD in the *H. jecorina* GA structures. In the *H. jecorina* and *T. emersonii* structures, the homologous residue is an Asp and thus glycosylation would not occur. In *H. grisea* GA, the situation is uncertain as there is an Asn at the homologous position, which could be glycosylated. Thus, the juxtaposition we find for the SBD and CDs further could be found in other glucoamylases.

Proposed Role of the SBD. Having determined the structure of an intact glucoamylase which represents the solution structure of *H. jecorina* and possibly other related glucoamylases, one can ask what can be inferred regarding the function of the SBD in glucoamylase. On the basis of the knowledge that the presence of the SBD has an impact on the hydrolysis of insoluble starch, it follows that we should expect an interaction of the SBD with larger starch molecules. Figure 6 summarizes what can be deduced from the structure of HjGA and earlier work done to determine the structure of the CD of AaGA and the SBD of *A. niger* GA. In Figure 6, we compare the structures of HjGA with the acarbose-bound CD of AaGA and the SBD from *A. niger* GA complexed with β-cyclodextrin, in the same orientation as the HjGA. It can be seen that the β-cyclodextrin bound at binding site 2 is close to the substrate location as indicated by the location of acarbose bound to *A. awamori* CD.

The coordinates of acarbose (green stick figure) from the structure model of the AaGA [PDB entry 1GAI (*20*)] can easily be accommodated into the HjGA active site when the two structures are superposed. We can similarly align the AnGA SBD structure model [PDB entry 1AC0 (*17*)] with the two β-cyclodextrin molecules (green stick figures) representing the two SBD binding sites with the SBD of the HjGA molecule. Taken together, we can examine how the acarbose model bound to HjGA derived by homology with the complex of acarbose with AaGA would juxtapose in our intact molecule using the model of β-cyclodextrin binding to the isolated SBD from AnGA, which is represented in Figure 6. The acarbose bound to AaGA was disordered with the C and D sugars adopting two conformations, both of which are included in Figure 6. One of the conformations appears to follow a direction toward the β-cyclodextrin bound at the second binding site identified for the SBD from AnGA determined by NMR. In that structure, it was noted that the loops involved in binding site 2 underwent conformational changes upon binding compared to the structure of the free SBD (*17*). The two β-cyclodextrin sites are oriented perpendicular, and the first site is on the side opposite the CD in HjGA. It is proposed that the SBD would localize the HjGA CD near disrupted starch and also prevent the enzyme from diffusing away from the substrate while releasing the product from the active site after the hydrolytic advent. The overall structure of the HjGA we have determined would be consistent with these proposed functions for the SBD. The SBD of HjGA would bind to starch along site 1, as we propose in Figure 6, and favor localization where a disrupted fragment (shown in the bottom figure) could bind to site 2 within a loose end that points into the catalytic site as we see for the acarbose.

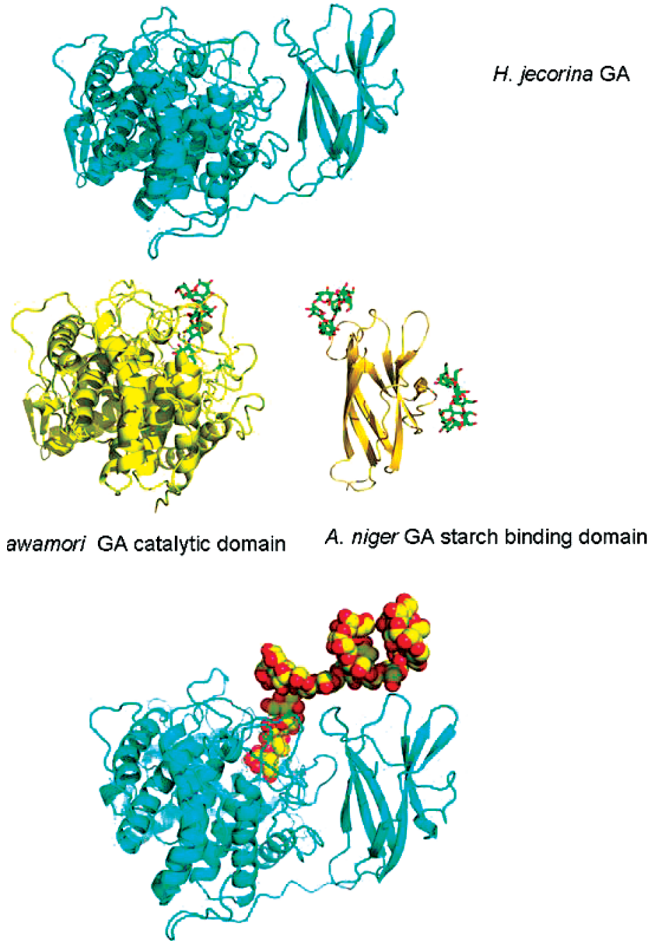

FIGURE 6: Role of the SBD in HjGA. In this composite, the top panel again shows the structure of HjGA, below the middle left is the crystal structure of the catalytic domain fragment of AaGA complexed with acarbose, and the middle right is the structure of the SBD from AnGA complexed with β-cyclodextrin where the β-cyclodextrin was seen to bind at two sites. At the bottom is again the structure of HjGA where a model of starch bound with the extended reducing segment has been modeled to overlap the sugar moieties of acarbose bound to the catalytic domain so that the intact helical starch structure overlaps one of the β-cyclodextrin sites. The variable loop seen for different SBD is involved in this β-cyclodextrin binding site. On the basis of this modeling, the hypothesis proposed in ref *17* that the SBD tends to localize in the GA near regions of disrupted starch is consistent with the intact structure observed in this study. Furthermore, the binding of starch to the SBD would provide a means of tethering the enzyme to starch during the product release step, which appears necessary for the release of glucose after catalysis.

In summary, the structure of HjGA presents a model of an intact glucoamylase, which has been determined in two crystal forms. The CD is essentially identical to other glucoamylases. In this model, we also visualize for the first time how the linker and SBD could contribute to the proposed function of the enzyme in facilitating the hydrolysis of insoluble starch. The amino acid side chains involved in the specific interaction among the CD, the linker, and the SBD are specific for *H. jecorina*; however, it appears that in other glucoamylases, complementary sequence changes would enable similar overall interactions and domain juxtaposition. The structure of HjGA should provide a basis for further experimentation and for engineering this and related enzymes for several important processes in the generation of fuels based on starch feed stocks.

MATERIALS AND METHODS

Protein Expression and Purification. The gene encoding *H. jecorina* GA was cloned and expressed according to the protocols described previously by Dunn-Coleman et al. (U.S. Patent 0094080 A1, 2006).

The HjGA protein material used for all crystallization experiments was initially purified in one step by anion exchange chromatography as follows. Concentrated culture supernatants of expressed HjGA, consisting of 180 mg/mL total protein, were prepared by diluting samples 1:10 in a 25 mM Tris-HCl buffer (pH 8.0). A HiPrep 16/10 Q Sepharose FF column (GE Healthcare) was employed for the anion exchange purification. The HiPrep column was equilibrated with 4 column volumes of starting buffer [25 mM Tris-HCl (pH 8.0)] followed by application of 10 mL of the diluted protein sample. An 8 column volume linear gradient of 0 to 140 mM NaCl in running buffer [25 mM Tris-HCl (pH 8.0)] was applied to elute bound protein. Bound HjGA eluted from the HiPrep Q Sepharose column at a salt concentration of approximately 80 mM NaCl. Fractions containing pure HjGA protein were pooled and concentrated to 50 mg/mL using a 20 mL Vivaspin centrifugal concentration tube (Viva Science) with a molecular mass cutoff of 10 kDa. Purified and concentrated HjGA material was buffer exchanged using a DG-10 desalting column (Bio-Rad) equilibrated with 50 mM sodium acetate buffer (pH 4.3). Protein concentrations were determined by measuring the absorbance at 280 nm using an extinction coefficient of 0.55 mg/mL/A_{280} unit. The initially purified and concentrated HjGA protein stock was thereafter stored at −20 °C.

Fractions collected during all purifications of HjGA were analyzed using sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE). Two additional purification steps, one additional anion exchange purification, and one size exclusion purification step were at a later stage introduced to enhance the crystal forming ability of the HjGA protein material. These two additional purification steps were performed as follows. In the first anion exchange purification step, a 10 mL MonoQ column (GE Healthcare) was employed. A sample of 1 mL of the initially purified and frozen HjGA material (50 mg of protein) was thawed, and the buffer was changed to 20 mM Tris-HCl (pH 8.0) by repeated dilution of the sample to 6 mL in the new buffer, followed by a concentration of the sample again to 0.5 mL using a 6 mL 5 kDa molecular mass cutoff concentration tube. The HjGA sample was after the last concentration step diluted in distilled water until a conductivity of the protein sample was reached that corresponded to the conductivity of the starting buffer of the anion purification, i.e., 25 mM Tris-HCl (pH 8.0). The MonoQ column was first equilibrated with 4 column volumes of starting buffer, followed by application of the diluted protein sample to the column. Bound protein was eluted from the MonoQ column by two different gradients. In the first, a 4 column volume linear pH gradient was applied where the pH of the starting buffer was decreased from 8.0 to 6.0. In the second gradient, an 8 column volume long salt gradient was applied in which the salt concentration was increased from 0 to 350 mM NaCl in the running buffer [25 mM Tris-HCl (pH 6.0)]. Bound HjGA was found to elute from the column during the second salt gradient at an approximate NaCl concentration of 150 mM.

Fractions containing HjGA were pooled and concentrated to 2 mL using a 6 mL 5 kDa molecular mass cutoff Vivaspin concentration tube. The concentrated HjGA sample was thereafter applied to a Superdex 200 16/60 size exclusion column (GE Healthcare) equilibrated with 4 column volumes of 20 mM Tris-HCl (pH 8.0) and 50 mM NaCl, which also was used as running buffer. Fractions from the main elution peak after the size exclusion purification were pooled and concentrated to an approximate protein concentration of 7.5 mg/mL using a 6 mL 5 kDa molecular mass cutoff Vivaspin concentration tube.

Protein Crystallization. The protein sample that was used to find the initial HjGA crystallization conditions was a sample of the HjGA material that was purified only once by anion exchange purification and thereafter stored at −20 °C. The HjGA protein sample was thawed and diluted with 50 mM sodium acetate buffer (pH 4.3) to approximately 12 mg/mL, prior to the initial crystallization experiments. This "less pure" HjGA protein sample was used to grow the first crystals from which the orthorhombic X-ray data set, used to determine the HjGA structure by molecular replacement (MR), and the high-resolution orthorhombic data set, used for the final orthorhombic space group HjGA structure model, were obtained. The orthorhombic HjGA crystals were found to grow in solution #94 of Qiagen's JCSG+ Suite [25% PEG 3350, 0.20 M ammonium acetate, and 0.10 M Bis-tris (pH 5.5)], using the vapor-diffusion method with hanging drops (*28*), at 20 °C. Crystallization drops were prepared by mixing equal amounts of protein solution (12 mg/mL) and crystallization agent to a final volume of 5 μL. The initial orthorhombic HjGA crystals appeared as aggregates containing multiple crystals. Multiple macro-seeding steps, at 4 °C using 5 μL of a 12 mg/mL protein sample and 5 μL of the crystallization reagent, were performed to obtain improved HjGA crystals suitable for X-ray data collection. The HjGA crystals, grown using the less purified protein material purified with only one anion exchange purification, were found to belong to orthorhombic space group $P2_12_12_1$ with the following approximate cell dimensions: $a = 52.2$ Å, $b = 99.2$ Å, and $c = 121.2$ Å. They were also found to have a calculated V_m of 2.3 Å^3/Da (*24*) with one molecule in the asymmetric unit.

It was found that it was much easier to grow uniform single HjGA crystals using a sample of HjGA protein, which was purified by the "improved" HjGA purification protocol. These new crystals were grown in a crystallization agent containing 25% PEG 3350, 0.1 M $CaCl_2$, and Bis-tris (pH 5.5), using the vapor-diffusion method with hanging drops, at 20 °C. These crystallization drops were prepared by mixing equal amounts of a protein solution (5 mg/mL) and crystallization agent to a final volume of 2 μL. These drops were also streak-seeded once to obtain single high-quality crystals suitable for X-ray data collection in a reasonable time. These new HjGA crystals were found to belong to monoclinic space group $C2$ with the following approximate cell dimensions: $a = 216.8$ Å, $b = 48.5$ Å, $c = 50.6$ Å, and $\beta = 90.5°$. They were also found to have a calculated V_m of 2.0 Å^3/Da (*24*) with one molecule in the asymmetric unit.

X-ray Data Collection. The two orthorhombic HjGA data sets were collected from single crystals mounted in sealed capillary tubes, at room temperature. The initial orthorhombic HjGA X-ray data set, used to determine the structure by molecular replacement methods (MR), was collected on a home X-ray source, an MSC/Rigaku (Molecular Structures Corp., The Woodlands, TX) Raxis IV++ image plate detector with focusing mirrors using Cu Kα radiation from a Rigaku RU200 rotating anode generator. This data set was processed, scaled, and averaged using the d*trek software provided by MSC/Rigaku. A second data set was also collected from the home source to a resolution of 2.0 Å. The C-centered monoclinic data set was collected from a single frozen HjGA crystal at 100 K, equilibrated in a cryoprotective agent comprised of 25% PEG 3350, 15% glycerol, 50 mM $CaCl_2$, and 0.1 M Bis-tris (pH 5.5) as a cryoprotectant, mounted in rayon-fiber loops, and plunge-frozen in liquid nitrogen prior to transportation to the synchrotron. The final high-resolution orthorhombic (1.85 Å) data set and the C-centric monoclinic data set (1.8 Å) were both collected at a synchrotron source, beamline 911:5 at MAX LABORATORY in Lund, Sweden. Both data sets that were collected at a synchrotron source were processed with MOSFLM (*29*) and scaled with SCALA included in the CCP4 program package (*30*). All subsequent data processing was performed using the CCP4 program package (*30*), unless otherwise stated. A set of 5% of the reflections from each data set was set aside and used for monitoring R_{free} (*31*). Data collection statistics for the different X-ray data sets are summarized in Table 1A.

Structure Solution. The HjGA structure was initially determined by MR with the automatic replacement program MOLREP, included in the CCP4 program package, using the initial orthorhombic data set, and using the coordinates of *A. awamori* GA variant X100 [PDB entry 1GLM (*20*)] as a search model. The *A. awamori* GA search model was edited to remove all sugar moieties attached to the protein molecule as N- and O-glycosylations and all solvent molecules before the MR experiments were carried out. All reflections between 36.8 and 2.8 Å resolution from the initial HjGA data set were used for the MR solution. The MR program found a single rotation function solution, with a maximum of 11.1σ above the background; the next highest maximum was 3.8σ above the background. The translation function solution gave an R_{work} of 48.7% and a contrast factor of 17.4. The MR solution was refined for 10 cycles of restrained least-squares refinement using Refmac 5.0 (*32*). This lowered the crystallographic R_{work} to 31.1%, while the R_{free} value dropped from 42.2 to 41.1%.

Model Fitting and Refinement. The refined MR solution model was used to calculate an initial density map from the orthorhombic HjGA data set. Electron density for a disulfide bridge between residues 19 and 26 of HjGA, a disulfide bridge not present in the *A. awamori* variant X100 structure model, could readily be identified in this electron density map. This was taken as an indication that the electron density map was of sufficient quality to be used to build a structure model of HjGA from its amino acid sequence. The initial HjGA structure model, based on the 2.0 Å resolution data set, was refined with alternating cycles of model building using Coot (*33*), and maximum likelihood refinement using Refmac 5.0 (*32*).

The resolution of the initial HjGA structure model was extended to a resolution of 1.85 Å by refining the initial HjGA structure model against the high-resolution synchrotron data set for 10 cycles of restrained refinement using Refmac

5.0. The monoclinic HjGA structure was determined by MR using the refined high-resolution orthorhombic HjGA structure as a search model. The two high-resolution HjGA structure models were refined with alternating cycles of model building using O (*34*), and maximum likelihood refinement using Refmac 5.0. The final two HjGA structure models contain all amino acid residues from position 1 to 599, and the final R_{work} and R_{free} values for the structure models are 15.5 and 18.2%, respectively, for the orthorhombic structure model and 20.3 and 25.3%, respectively, for the monoclinic structure model.

Most water molecules in the structure models were located automatically by using the water picking protocols in the refinement programs and then manually selected or discarded by inspection by eye. A summary of refinement statistics for the two structures is given in Table 1B.

All structural comparisons were made with either Coot (*33*) or O (*34*), and figures were prepared with PyMOL (*35*). Coordinates for the two HjGA structure models have been deposited in the Protein Data Bank (*36*) as entries 2vn4 and 2vn7, respectively.

ACKNOWLEDGMENT

We thank Roopa Ghirnikar for her editorial assistance in the preparation of the manuscript.

REFERENCES

1. Koshland, P. J. (1953) Stereochemistry and the mechanism of enzymatic reactions. *Biol. Rev. 28*, 416–436.
2. Sinnott, M. L. (1990) Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev. 90*, 1171–1202.
3. Weill, C. E., Burch, R. J., and van Dyk, J. W. (1954) *Cereal Chem. 31*, 150–158.
4. Pazur, J. H., and Ando, T. (1960) The hydrolysis of glucosyl oligosaccharides with α-D-(1−4) and α-D-(1−6) bonds by fungal amyloglucosidase. *J. Biol. Chem. 235*, 297–302.
5. Hiromi, K., et al. (1966) Kinetic studies on gluc-amylase. II. Competition between two types of substrate having alpha-1,4 and alpha-1,6 glucosidic linkage. *J. Biochem. 59* (4), 411–418.
6. Sierks, M. R., and Svensson, B. (1994) Protein engineering of the relative specificity of glucoamylase from *Aspergillus awamori* based on sequence similarities between starch-degrading enzymes. *Protein Eng. 7* (12), 1479–1484.
7. Fierobe, H. P., et al. (1996) Mutational modulation of substrate bond-type specificity and thermostability of glucoamylase from *Aspergillus awamori* by replacement with short homologue active site sequences and thiol/disulfide engineering. *Biochemistry 35* (26), 8696–8704.
8. Saha, B. C., and Zeikus, J. G. (1989) Microbial glucoamylases: Biochemical and biotechnological features. *Starch/Staerke 41*, 57–64.
9. Sauer, J., et al. (2000) Glucoamylase: Structure/function relationships, and protein engineering. *Biochim. Biophys. Acta 1543* (2), 275–293.
10. Henrissat, B., and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J. 293* (Part 3), 781–788.
11. Coutinho, P. M., and Henrissat, B. (1999) Carbohydrate-active enzymes: An integrated database approach, in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., et al., Eds.) pp 3−12, The Royal Society of Chemistry, Cambridge, U.K.
12. Coutinho, P. M., and Reilly, P. J. (1994) Structural similarities in glucoamylase by hydrophobic cluster analysis. *Protein Eng. 7* (6), 749–760.
13. Coutinho, P. M., and Reilly, P. J. (1997) Glucoamylase structural, functional, and evolutionary relationships. *Proteins 29* (3), 334–347.
14. Hayashida, S. (1975) *Agric. Biol. Chem. 39*, 2093–2099.
15. Svensson, B., et al. (1982) Characterization of two forms of glucoamylase from *Aspergillus niger*. *Carlsberg Res. Commun. 47*, 55–69.
16. Sorimachi, K., et al. (1996) Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. *J. Mol. Biol. 259* (5), 970–987.
17. Sorimachi, K., et al. (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. *Structure 5* (5), 647–661.
18. Williamson, M. P., et al. (1997) Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. *Biochemistry 36* (24), 7535–7539.
19. Southall, S. M., et al. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett. 447* (1), 58–60.
20. Aleshin, A. E., et al. 1994a Refined crystal structures of glucoamylase from *Aspergillus awamori* var. X100. *J. Mol. Biol. 238* (4), 575–591.
21. Aleshin, A. E., Firsov, L. M., and Honzatko, R. B. (1994) Refined structure for the complex of acarbose with glucoamylase from *Aspergillus awamori* var. X100 to 2.4-Å resolution. *J. Biol. Chem. 269* (22), 15631–15639.
22. Stoffer, B., et al. (1993) Production, purification and characterization of the catalytic domain of glucoamylase from *Aspergillus niger*. *Biochem. J. 292* (Part 1), 197–202.
23. Fagerström, R., and Kalkkinen, N. (1995) Characterization, subsite mapping and partial amino acid sequence of glucoamylase from the filamentous fungus *Trichoderma reesei*. *Biotechnol. Appl. Biochem. 21* (Part 2), 223–231.
24. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol. 33*, 491–497.
25. Aleshin, A. E., et al. (1996) Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: Relationship of stereochemical distortions at the nonreducing end to the catalytic mechanism. *Biochemistry 35* (25), 8319–8328.
26. Sevcik, J., et al. (1998) Structure of glucoamylase from *Saccharomycopsis fibuligera* at 1.7 Å resolution. *Acta Crystallogr. D54* (Part 5), 854–866.
27. Mikami, B., et al. (1999) Structure of raw starch-digesting *Bacillus cereus* β-amylase complexed with maltose. *Biochemistry 38* (22), 7050–7061.
28. McPherson, A. J. (1982) *Preparation and Analysis of Protein Crystals*, John Wiley and Sons, New York.
29. Leslie, A. G. W. (1992)*Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
30. Collaborative Computational Project Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* 760–763.
31. Brünger, A. T. (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature 355*, 472–475.
32. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
33. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60* (Part 12), 2126–2132.
34. Jones, T. A., et al. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.
35. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA.
36. Bernstein, F. C., et al. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*, 535–542.
37. Engh, R. A., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. A47*, 392–400.
38. Kleywegt, G. J., and Jones, T. A. (1996) Phi/Psi-chology: Ramachandran revisited. *Structure 4*, 1395–1400.
39. Kleywegt, G. J., and Jones, T. A. (1996) xdlMAPMAN and xdlDATAMAN: Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr. D52*, 826–828.
40. Kleywegt, G. J., and Jones, T. A. (1997) Detecting folding motifs and similarities in protein structures. *Methods Enzymol. 277*, 525–545.

BI702413K