PROTEINS: Structure, Function, and Genetics 29:334–347 (1997)

Glucoamylase Structural, Functional, and Evolutionary Relationships

Pedro M. Coutinho and Peter J. Reilly*
Department of Chemical Engineering, Iowa State University, Ames, Iowa

ABSTRACT **To correlate structural features with glucoamylase properties, a structure-based multisequence alignment was constructed using information from catalytic and starch-binding domain models. The catalytic domain is composed of three hydrophobic folding units, the most labile and least hydrophobic of them being missing in the most stable glucoamylase. The role of *O*-glycosylation in stabilizing the most hydrophobic folding unit, the only one where thermostabilizing mutations with unchanged activity have been made, is described. Differences in both length and composition of interhelical loops are correlated with stability and selectivity characteristics. Two new glucoamylase subfamilies are defined by using homology criteria. Protein parsimony analysis suggests an ancient bacterial origin for the glucoamylase gene. Increases in length of the belt surrounding the active site, degree of *O*-glycosylation, and length of the linker probably correspond to evolutionary steps that increase stability and secretion levels of *Aspergillus*-related glucoamylases. Proteins 29:334–347, 1997.** © 1997 Wiley-Liss, Inc.

Key words: evolution; glucoamylase; hydrophobic folding; protein parsimony analysis; structure/function

INTRODUCTION

Glucoamylase [α(1,4)-D-glucan glucohydrolase, EC 3.2.1.3, GA] converts raw and solubilized starch and a variety of other carbohydrate polymers and oligomers into glucose. Different microorganisms produce GAs,[1–3] but the industrial focus has been on GAs from *Aspergillus niger* (identical to *Aspergillus awamori* GA) and *Rhizopus oryzae* (also known as *Rhizopus delemar*). Both enzymes have a catalytic domain (CD) and a starch-binding domain (SBD) connected by an *O*-glycosylated linker. Partial or total proteolytic excision of the SBD leads to enzyme forms able to degrade solubilized starch but not its raw or granular form. Both GAs have been extensively characterized through the years.[4–8] However, more recent efforts to modify GA properties to better suit industrial uses require a better knowledge of GA structure/function relationships.

The three-dimensional model of the *A. awamori* var. *X100* GA CD has been obtained by protein crystallography in its native state and when complexed with different inhibitors,[9–14] and the structure of the *A. niger* GA SBD has been recently determined by NMR.[15,16] The CD is an $(\alpha/\alpha)_6$-barrel founded by an *O*-glycosylated belt, the active site lying in its center. In contrast to the high α-helical content of the CD, the SBD is comprised of a set of eight β strands in a fold characteristic of SBDs found in other glucohydrolases.[17,18] Belonging to Family 15 of glycosyl hydrolases,[19–21] GA has a CD whose topology is similar to those of glycosidases of Families 8 and 9, represented by *Clostridium thermocellum endo*-glucanases CelA[22] and CelD,[23] respectively. Unlike these enzymes, whose active sites are surface grooves where the substrate binds before undergoing *endo*-cleavage, the GA active site is a pocket where the nonreducing end of substrates binds and is hydrolyzed in a characteristic *exo*-acting mode to release β-glucose.[24] In *A. niger* GA, the CD and SBD are independent, separated by an *O*-glycosylated linker.[25,26] Differential scanning microcalorimetric studies of *A. niger* GA have recently indicated that the SBD is an independent folding domain and that the CD contains three independent folding units.[27]

The GA active site is composed of conserved loops whose conserved amino acid residues interact with substrates and inhibitors.[10,28–32] Maltooligosaccharide interaction in the first four subsites of *A. awamori* var. *X100* GA was defined in studies with maltotetraoase analogues,[12–14] the third and fourth subsites being multiple and pH-dependent. However, up to seven subsites have been found in the *A. niger*/*A. awamori* GA active site by subsite mapping.[33–35] Five subfamilies of related GAs, designated by the most common genera of the species included in them (*Aspergillus*, *Rhizopus*, *Saccharo-*

Contract grant sponsor: U.S. Department of Energy: Consortium for Plant Biotechnology Research; Contract grant sponsor: Genencor International, Inc.; Contract grant sponsor: Programa CIENCIA, JNICT (Lisboa, Portugal); Contract grant number: B.D./1057/90-IF.
*Correspondence to: Professor Peter J. Reilly, Department of Chemical Engineering, 2114 Sweeney Hall, Iowa State University, Ames, IA 50011-2230.
E-mail: reilly@iastate.edu
Received 3 March 1997; Accepted 20 May 1997

TABLE I. Hydrophobic Folding Units in *A. awamori* var. *X100* GA Models*

GA model	Conditions	Hydrophobic folding units† (hydrophobicity index)				
		Lh	LMh	Mh	MHh	Hh
3gly	Native, 2.2 Å pH 6.0	335–377 (0.73)		224–334 378–410 411–423 (0.79)	10–134 202–209 (0.79)	135–160 458–471 (0.74)
1glm	Native, 2.4 Å pH 4.0	335–377 (0.73)		224–334 378–411 (0.79)	1–223 451–471 (0.82)	
1agm	Complex, 2.4 Å pH 6.0	328–382 (0.75)		1–14 173–194 218–327 382–452 (0.82)		15–172 195–217 452–471 (0.84)
1dog	Complex, 2.4 Å pH 6.0	332–378 (0.74)		1–14 170–195 218–331 380–452 (0.83)		15–169 196–217 453–471 (0.84)
1gah	Complex, 2.0 Å pH 6.0		269–279 326–382 (0.79)	280–325 383–411 (0.76)		15–169 195–217 452–472 (0.84)
1gai	Complex, 1.7 Å pH 4.0	327–381 (0.75)		1–14 173–194 218–326 381–452 (0.81)		15–172 195–217 452–473 (0.83)
Average		330–379		1–14 172–194 218–329 380–452		15–171 195–217 453–470

*The hydrophobicity of each unit is indicated as well as an average. Segments follow *A. niger* GA numbering. Some segments were not assigned in some protein models.
†Classified by hydrophobicity units: Lh, lower; Mh, medium; Hh, higher. For mixed results, two other categories, not based on hydrophobicity, were used: LMh, lower/medium; MHh, medium/higher.

mycopsis, *Saccharomyces*, and *Clostridium*) have been identified based on hydrophobic cluster analysis of their GA sequences.[30–32] The different GA subfamilies share the same overall CD arrangement.

Only *Aspergillus*-related and *Rhizopus* GAs possess SBDs, the former at the C terminus and the latter at the N terminus. In *Aspergillus*-related GAs, most conserved SBD residues are either structurally important or are putatively involved in substrate binding,[31] substrate interactions being based in the interactions of the homologous SBD from *Bacillus circulans* strain *251* cyclodextrin glucosyltransferase (CGTase) with maltose[36] and α-cyclodextrin.[37]

In this study, a division of the CD into hydrophobic folding units, which are likely to have an important role in GA stability, is made. Different GA sequences are characterized based on newly released sequence information and on the new NMR model of the SBD, with the identification of two new GA subfamilies, *Arxula* and *Methanococcus*. The new structure-based alignment allows a more complete comparison of expected structural features in the different GAs and suggests that the omission of the least stable hydrophobic folding unit in bacterial and archaeal GAs stabilize these proteins. Protein parsimony analysis of the alignment gives a powerful insight into the evolution of the CD, linker, and SBD structures.

METHODS

Secondary and Tertiary Structures

All A. awamori var. *X100* CD models in the Brookhaven Protein Databank, structures **3gly**, **1glm**, **1dog**, **1agm**, **1gah**, and **1gai**,[9–14] were partitioned into hydrophobic folding domains by the World Wide Web service of the National Cancer Institute.[38] The results were averaged into three hydrophobic folding units, their limits yielding a typical uncertainty of ±3 amino acid residues (Table I).

Secondary strucure information was extracted from three-dimensional models of the CD from *A. awamori* var. *X100* GA[9,11] and that of the SBD of *A. niger* GA[16] (structures **3gly** and **1kum** of the Brookhaven Protein Databank, respectively) using STRIDE (Euro-

pean Molecular Biology Laboratory, Heidelberg, Germany).[39]

Primary Structures

GA gene sequences were collected from the filamentous fungi *A. awamori*,[40] *A. awamori* var. *kawachi*,[41] *A. awamori* var. *X100*,[42] *A. niger*,[43] *A. niger* var. *T21*,[44] *Aspergillus oryzae*,[45] *Aspergillus shirousami*,[46] *Aspergillus terreus*,[47] *Corticium rolfsii*,[48] *Hormoconis resinae GamP*,[49] *Humicola grisea* var. *thermoidea*,[50] *Neurospora crassa*,[51] and *R. oryzae*,[5] the yeasts *Arxula adeninivorans*,[52] *Saccharomyces cerevisiae*,[53,54] *Saccharomyces diastaticus*,[55] *Saccharomycopsis fibuligera*,[29] *Saccharomycopsis fibuligera KZ*,[56] *Schizosaccharomyces pombe*,[57] and *Schwanniomyces occidentalis*,[58] the eubacterium *Clostridium* sp. *G0005*,[59] and the archaeon *Methanococcus jannaschii*.[60]

Partial protein sequence information was also collected from peptide mapping of fungal *H. resinae* GA *S*[61] (R. Fagerström, personal communication) and *Trichoderma reesei* GA,[62] and from N-terminal sequencing of fungal *Aspergillus phoeniciens* (also known as *Aspergillus saitoi*) GA,[63] *Aspergillus* sp. *K-27* GA,[64] and three *Chalara paradoxa* GA forms.[65] Partial genomic sequencing in *N. crassa* also yielded segments of a putative GA sequence, designated as GA-2,[66] a GA structurally different from that reported earlier.[51]

Significant errors in some sequences were manually corrected. These can strongly affect data interpretation, as in obscuring the identification of conserved catalytic residues in *Saccharomyces* GA sequences[32] that were only definitely corrected following sequencing of *S. cerevisiae* chromosome IX.[54] The corrected GA sequences were:

1. *A. awamori* var. *X100*, where a shift in the position of a small set of residues between helices H11 and H12 was found
2. *A. niger* var. *T21*,where the catalytic acid was missing
3. *H. grisea* var. *thermoidea*, where the active site Arg54 (*A. niger* numbering) was a proline residue
4. *S. diastaticus*, to incorporate the correction of the frameshift error[32] as obtained for *S. cerevisiae* GA in chromosome IX.[54]

Also, only segments of the partial peptide mapping of *H. resinae* GA *S* and *T. reesei* GA[61,62] that could be matched to segments of other GAs were considered. An unidentified residue in one of the *H. resinae* GA *S* fragments was assumed to be *N*-glycosylated by homology to other sequences.

Multisequence Alignment

A structure-based alignment of both complete and partial GA protein sequence information was constructed and edited by using DCSE 3.4 (University of Antwerp, Belgium),[67] a sequence editor and analysis tool that allows simultaneous assignment of primary and secondary structure to protein sequences. Based on the principles of hydrophobic cluster alignment (HCA),[68,69] a multisequence alignment of all related GA sequences was constructed manually. Secondary structural elements were identified by homology to the existing three-dimensional models, maintaining hydrophobic residues in a regular arrangement in helices and β strands, and placing most insertions or deletions in nonstructured segments often associated with the positioning of turns. α-Helical, β-sheet, and turn assignments were based on the three-dimensional models used,[9,11,16] but 3_{10}-helical information, along with SBD turn assignments made by STRIDE, were also followed. The alignment served as a template for manual correction of some GA sequences and for alignment of peptide sequences derived from partial peptide mapping.

Binding Sites

The residues involved in the different binding subsites in the CD were evaluated using the GA/D-*gluco*-dihydroacarbose complex (structure **1gai**).[13,14] D-*gluco*-Dihydroacarbose is likely to represent the multiple binding conformations of maltotetraose in GA active site. Amino acid residues within 7 Å of any of the two inhibitor conformers were selected and were assigned to a specific subsite based on a distance criteria. Several residues can have a role in different subsites but only the closest will be shown.

For the putative binding sites in the SBD, those residues within 7 Å of maltose in the homologous SBD from *Bacillus circulans* strain *251* CGTase (structure **1cgt**)[36] were evaluated and the structurally equivalent residues in the *A. niger* SBD[16] were then selected.

Evolutionary Analysis

An evolutionary tree was constructed for all complete GA sequences following protein parsimony analysis. An unrooted distance tree was calculated with the Protpars and Drawtree programs, both components of the PHYLIP package.[70,71]

RESULTS AND DISCUSSION

Hydrophobic Folding Units

The different CD models were partitioned into hydrophobic folding units (Table I). Even though large differences on the limits of the folding units sometimes appeared, they tended to vary around the same values. Averaged results in the assignment of three folding units are shown in Figure 1. The least hydrophobic unit comprises helix H11, which is not part of the $(\alpha/\alpha)_6$ barrel structure, and portions of helices H10 and H12. The most hydrophobic unit contains parts of helices H1 and H7, all of helices H2 to H6, and second half of the *O*-glycosylated belt that follows helix H13. The unit of medium hydrophobic-

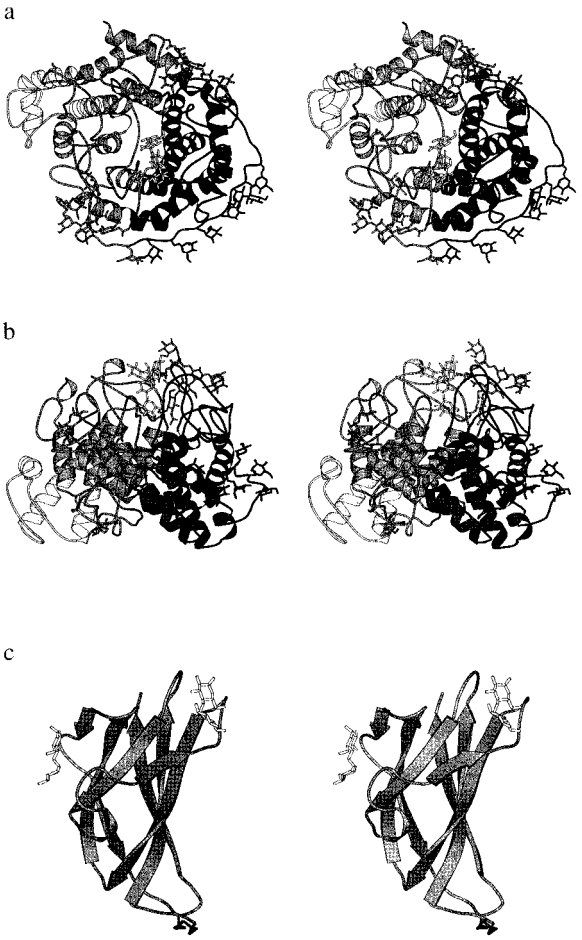

Fig. 1. CD and SBD of *Aspergillus* GAs visualized with MOLSCRIPT.[72] Stereoscopic plots of **a:** Hydrophobic folding units of CD of *A. awamori* var. *X100* GA with D-*gluco*-dihydroacarbose in active site[13,14] (front view). **b:** Same (view from below). **c:** Average minimized SBD of *A. niger* GA[16] with two maltose molecules interacting with the two surface binding sites as found in a CGTase.[36] Hydrophobic intensities of the CD are indicated by color: (clear) low hydrophobicity; (medium grey) medium hydrophobicity; and (dark gray) high hydrophobicity. *N*- and *O*-glycosylation is indicated with the colors of the respective folding units. The ligands found in the CD and SBD are colored in white.

ity contains the rest. Three folding units have been assigned to the *A. niger* GA CD in differential calorimetric studies,[27] with a fourth unit being assigned to the SBD. A recent combination of site-directed mutagenesis and differential calorimetry of the resulting mutated *A. awamori* GA shows that stabilization of helix H11 increased the melting temperature of only the first folding unit and did not affect overall enzyme thermostability (Y. Li et al., unpublished results). On the other hand, all mutated amino acid residues whose expressed proteins possess increased thermostability and at least wild-type activity[73] (Y. Li et al., M. J. Allen et al., unpublished results) are located in the most hydrophobic domain.

Sequence Alignment

A structure-based alignment of GA CDs and SBDs is shown in Figure 2. This alignment contains all existing GA sequences except those from *S. pombe*[57] and *S. occidentalis*[58] and the partial sequence from the putative *N. crassa* GA-2 GA,[66] which are not structurally related to the remaining GAs. They instead belong to glycosyl hydrolase Family 31,[19–21] a diverse family also containing α-glucosidases and sucrase-isomaltases, and will not be further discussed here. The SBDs of only the *Aspergillus*-related GAs were considered in this alignment, given that homology of the N-terminal SBD of *R. oryzae* GA with the corresponding N-terminal segment of *A. adeninivorans* GA invalidated some of the assumptions necessary to relate this domain to the remaining SBDs.[31] These SBDs might belong to a different family of evolutionarily related SBDs than those described earlier.[17]

Two new subfamilies of GA sequences, named *Arxula* and *Methanococcus* after their two members, show less than 50% amino acid identity to any other GA sequence, the same homology criterion used earlier.[30,31] The latter is the first archaeal GA subfamily described.

The alignment shows a clear relationship between the *Aspergillus*-related GAs, some homology between *Rhizopus* and *Arxula* GAs, and some between *Saccharomycopsis* and *Saccharomyces* GAs, where both sequence insertions and deletions tend to occur at the same sites. Both *Methanococcus* and *Clostridium* GAs are significantly different from the other GAs, the latter more than the former. However, they share some unique features with each other, namely, the location of important deletions in regions present in all other GAs that could have implications for both stability and function.

Alignment of partial sequences revealed that *H. resinae* GA *S* also contains an SBD, likely placed at the CD C terminus as in all *Aspergillus*-related GAs. *T. reesei* GA segments are closely related to the *Aspergillus* subfamily, as are the N termini of both *C. paradoxa* and *Aspergillus* sp. *K-27* GAs. The inclusion of the short *A. phoeniciens* GA segment in the same category is only tentative, but is likely because of the obviously close relationship of this species to other subfamily members.

Catalytic Domain Analysis

Most GA CDs share the same architecture, being comprised of thirteen helices, twelve paired two by two in an $(\alpha/\alpha)_6$ barrel. The extra helix H11 along with the terminal segments of helices H10 and H12 are missing from both *M. jannaschii* and *Clostridium* GAs. This suggests that the least hydrophobic folding unit, the one that unfolds first with an increase of temperature, is not present in these enzymes. Given that these two GAs are thermo-

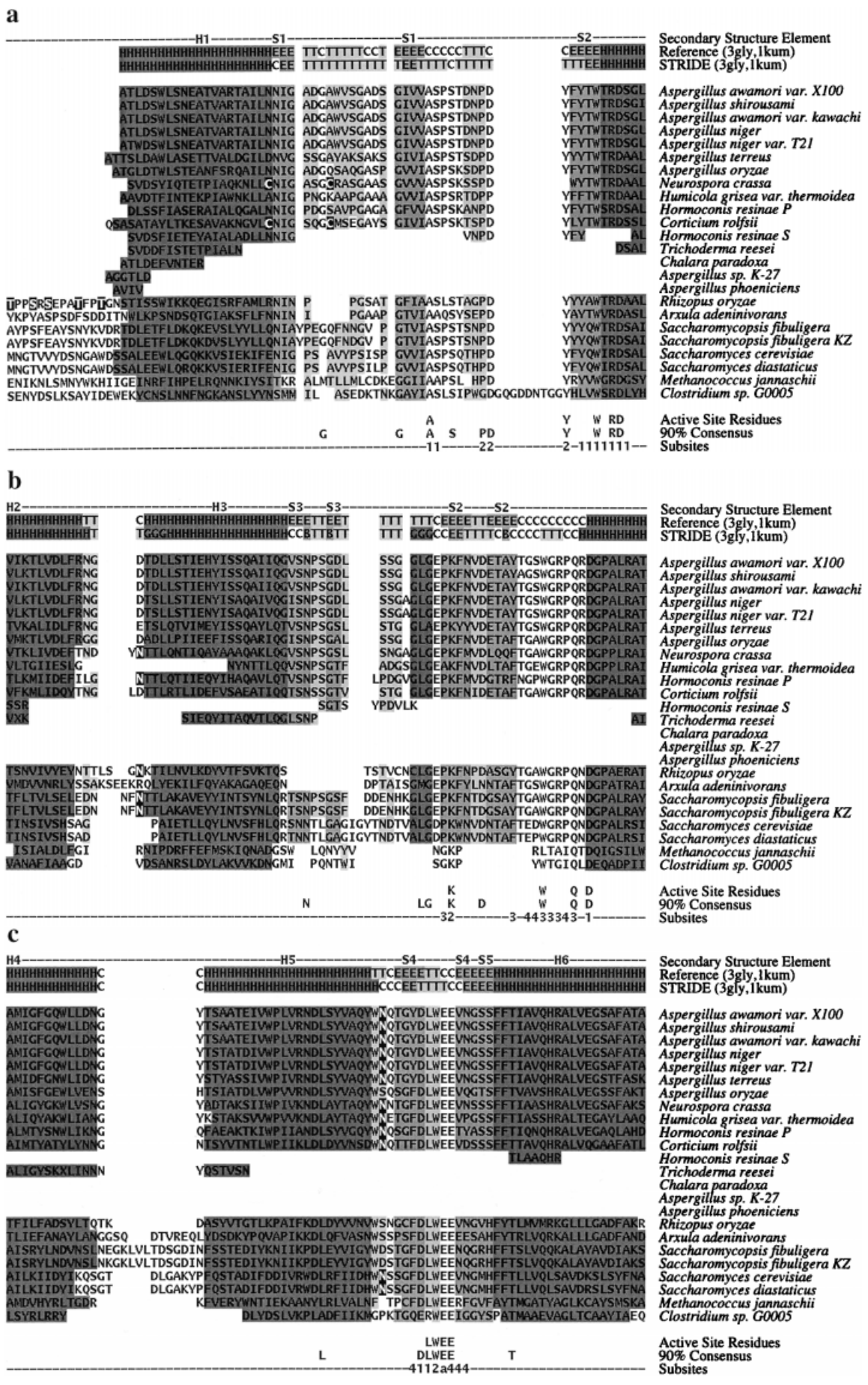
a
Secondary Structure Element
Reference (3gly,1kum)
STRIDE (3gly,1kum)
Aspergillus awamori var. X100
Aspergillus shirousami
Aspergillus awamori var. kawachi
Aspergillus niger
Aspergillus niger var. T21
Aspergillus terreus
Aspergillus oryzae
Neurospora crassa
Humicola grisea var. thermoidea
Hormoconis resinae P
Corticium rolfsii
Hormoconis resinae S
Trichoderma reesei
Chalara paradoxa
Aspergillus sp. K-27
Aspergillus phoeniciens
Rhizopus oryzae
Arxula adeninivorans
Saccharomycopsis fibuligera
Saccharomycopsis fibuligera KZ
Saccharomyces cerevisiae
Saccharomyces diastaticus
Methanococcus jannaschii
Clostridium sp. G0005
Active Site Residues
90% Consensus
Subsites
b
Secondary Structure Element
Reference (3gly,1kum)
STRIDE (3gly,1kum)
Aspergillus awamori var. X100
Aspergillus shirousami
Aspergillus awamori var. kawachi
Aspergillus niger
Aspergillus niger var. T21
Aspergillus terreus
Aspergillus oryzae
Neurospora crassa
Humicola grisea var. thermoidea
Hormoconis resinae P
Corticium rolfsii
Hormoconis resinae S
Trichoderma reesei
Chalara paradoxa
Aspergillus sp. K-27
Aspergillus phoeniciens
Rhizopus oryzae
Arxula adeninivorans
Saccharomycopsis fibuligera
Saccharomycopsis fibuligera KZ
Saccharomyces cerevisiae
Saccharomyces diastaticus
Methanococcus jannaschii
Clostridium sp. G0005
Active Site Residues
90% Consensus
Subsites
c
Secondary Structure Element
Reference (3gly,1kum)
STRIDE (3gly,1kum)
Aspergillus awamori var. X100
Aspergillus shirousami
Aspergillus awamori var. kawachi
Aspergillus niger
Aspergillus niger var. T21
Aspergillus terreus
Aspergillus oryzae
Neurospora crassa
Humicola grisea var. thermoidea
Hormoconis resinae P
Corticium rolfsii
Hormoconis resinae S
Trichoderma reesei
Chalara paradoxa
Aspergillus sp. K-27
Aspergillus phoeniciens
Rhizopus oryzae
Arxula adeninivorans
Saccharomycopsis fibuligera
Saccharomycopsis fibuligera KZ
Saccharomyces cerevisiae
Saccharomyces diastaticus
Methanococcus jannaschii
Clostridium sp. G0005
Active Site Residues
90% Consensus
Subsites

Fig. 2.

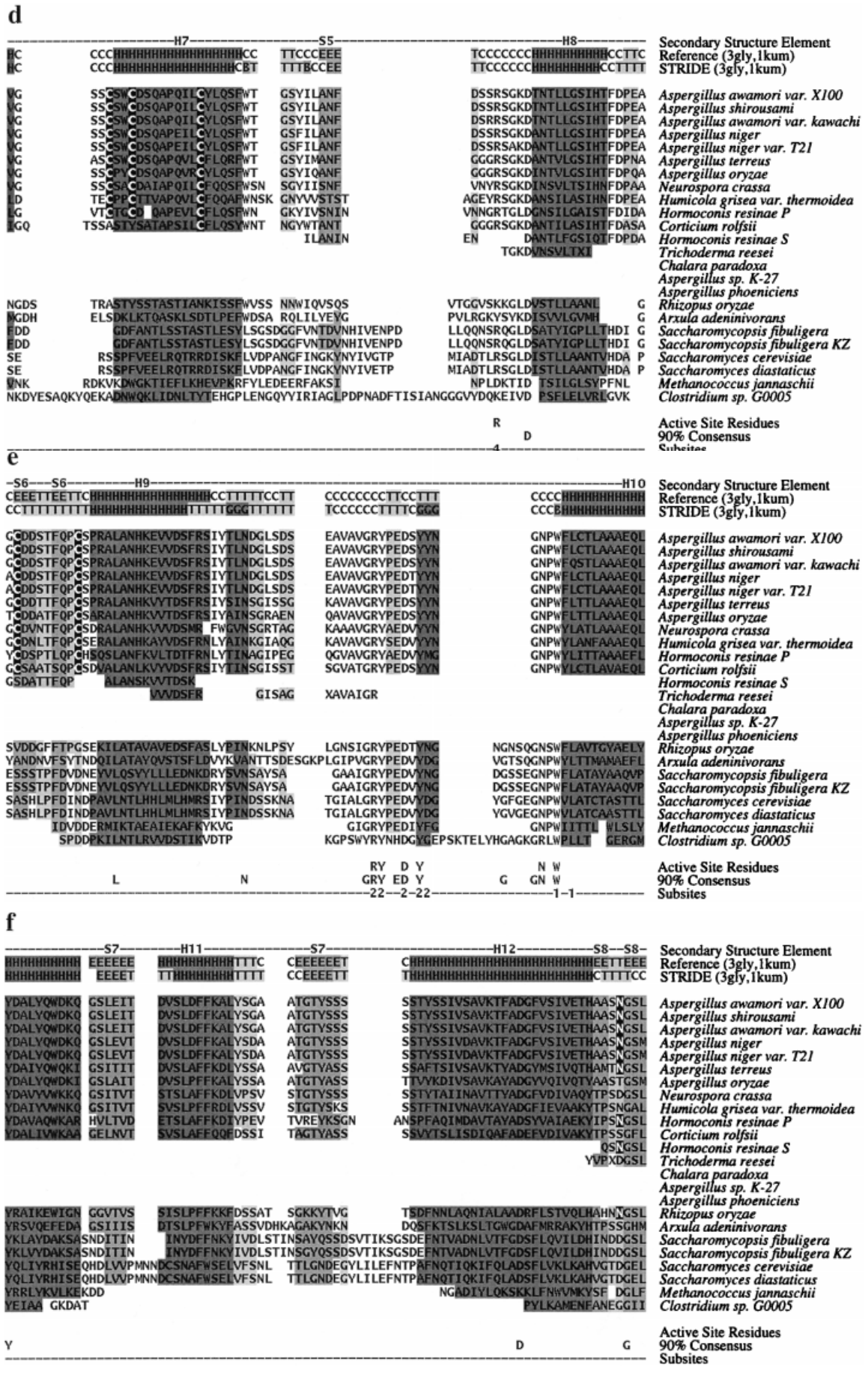

Figure 2. (Continued.)

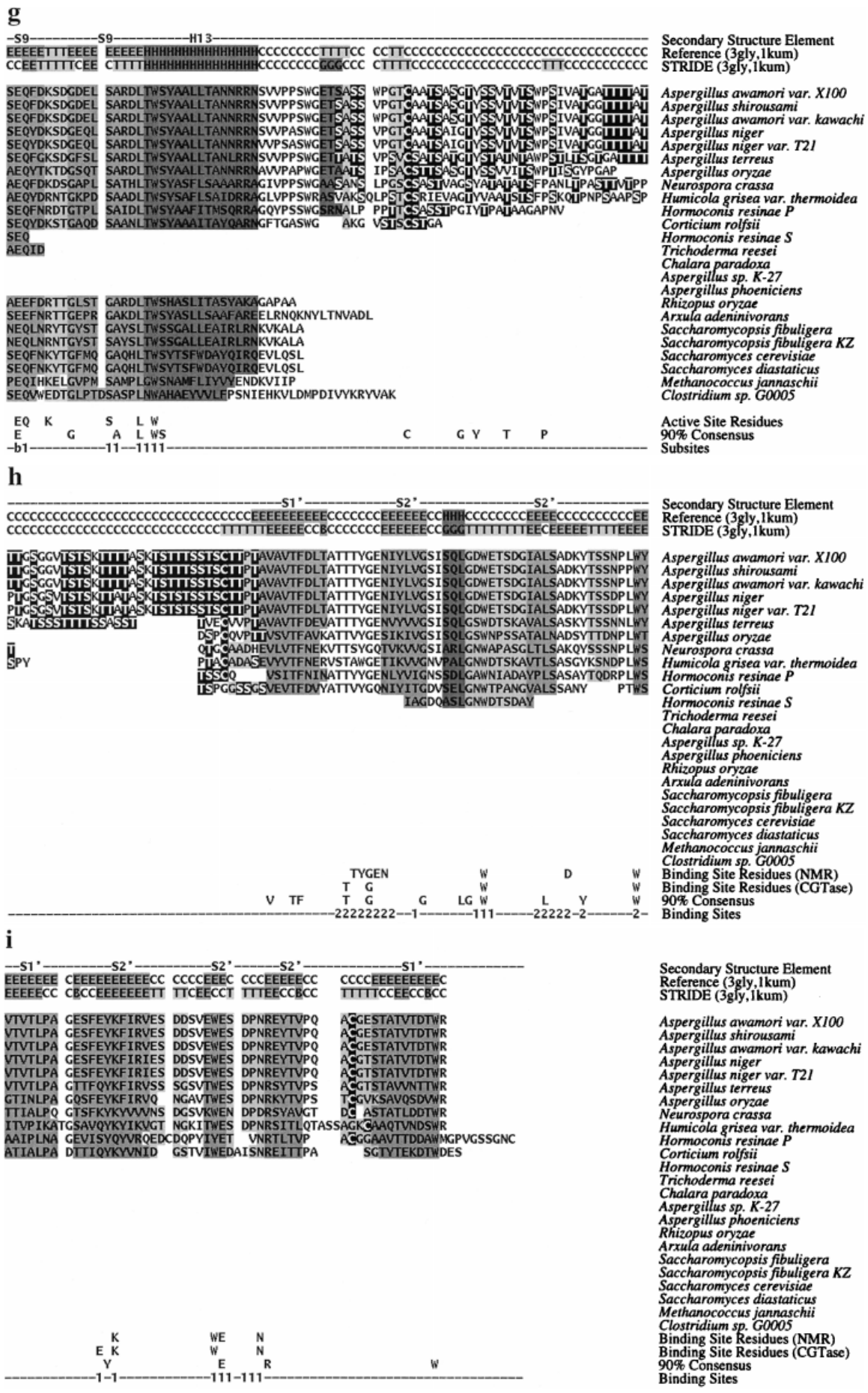

Figure 2. (Continued.)

stable, especially the one from *M. jannaschii*, a thermophilic organism that lives above 85°C,[60] the absence of this folding unit could be associated with stability. It could also simply indicate that these two more divergent enzymes are more primitive than the remaining GAs and never developed such a unit.

In *H. grisea* var. *thermoidea* GA significant segments of helices H2 and H3 are deleted, the remnants of the helices being connected by a much shorter loop than those found in other GAs. Small variations in the length of the loop between the two helices in different GA sequences also occur. This region exhibits a larger temperature factor in the *A. awamori* var. *X100* GA models than that found for the most hydrophobic folding domain, but still lower than those found in the last 10 residues on the *O*-glycosylated belt with which it interacts, suggesting that a shorter loop here might be associated with increased stability. Typically, short interhelical loops are found on the side of the CD opposite to the active site in *Aspergillus* GAs in the regions between helices H2 and H3, H4 and H5, and H6 and H7. The loop between H2 and H3 contains small insertions in *Arxula*, *Saccharomycopsis*, and *Rhizopus* GAs, while that between H4 and H5 is longer in GAs from all three yeast subfamilies. The loop between H6 and H7 appears to be significantly longer only in *Methanococcus* and *Clostridium* GAs, but their highly divergent sequences render interpretation difficult in this region. In *Aspergillus*-related GAs, a highly structured region exists between the short helix H8 and H9, containing both β sheet S6 and the Cys262–Cys270 (*A. niger* numbering) disulfide bridge (Fig. 2e). This region is shorter in the other GAs, an identical deletion being found in *Rhizopus*, *Arxula*, *Saccharomycopsis*, and *Saccharomyces* GAs. As mentioned before, the extra helix H11 is sequentially placed between the two long helices H10 and H12, the two interconnecting loops forming the antiparallel β sheet S7.[9] Such features are apparently shared by all sequences except those of *M. jannaschii* and *Clostridium* GAs (Fig. 2f), with insertions in both *Saccharomycopsis* and *Saccharomyces* GAs.

The interhelical loops surrounding the active site contain most of the residues conserved in GA. The majority of them are involved in the active site,[30–32] as can also be observed in the consensus residues and first four subsites (Fig. 2). The segment between helices H1 and H2 is a long structured region containing several active-site residues involved in subsites 1 and 2, most found between the second strand of β sheet S1, the first strand of S2, and the first loop of helix H2. As found earlier,[31] a major insertion is found in that region in the *Clostridium* GA. Also, significant deletions occur in the loop between both S1 β strands in both *Rhizopus* and *Arxula* GAs. Significantly, both *N. crassa* and *C. rolfsii* GAs form a disulfide bridge (Fig. 2a) between Cys20 and Cys27 (*A. niger* GA numbering). When both this bridge and the Ser30→ Pro mutation are incorporated into *A. awamori* GA, a more thermostable enzyme results (Y. Li et al., M. J. Allen et al., unpublished results). This segment might be associated with GA stability in all cases.

The long structured loop between helices H3 and H4 also contains several active-site residues, mostly residues involved in subsites 3 and 4 but also some closer to subsites 1 and 2. Most are found in the segment preceeding H4 but some occur in the two S2 β strands found there. Some small insertions and deletions are found after β sheet S3, but the most significant variation is the deletion of a large segment containing the two S2 β strands in both *Methanococcus* and *Clostridium* GAs. It is unclear if the single lysine residue found there in these two obviously related sequences takes the role of Lys108 (Fig. 2b), a residue essential for GA catalysis (H.-C. Liu et al., unpublished results), and if the assignment of a leucine residue to substitute for the critical Trp120[74] is correct for *M. jannaschii* GA.

The short but critical loop between helices H5 and H6 holding the catalytic acid Glu179[75] has the same structure in all cases, with a minor deletion next to H5 in *M. jannaschii* GA. This loop has active-site residues interacting in critical subsites 1 and 2, but with many involved in subsite 4. However, when three *H. resinae* GA *P* residues in the second segment of S4 and in S5, involved mostly in subsite 4, and four other residues in the loop between H9 and H10, likely affecting subsite 2, replaced the corresponding seven residues in *A. niger* GA, a doubling of GA isomaltose hydrolysis catalytic efficiency occurred.[76] Separate replacement of the two sets of residues decreased activity with no significant changes in selectivity.

Fig. 2. Structure-based alignment of GA sequences. Secondary structures as described in **3gly**[12] and **1kum**,[16] or as determined by STRIDE.[39] Structural elements: (B) β bridges; (E) β strands; (G) 3_{10}-helices; (H) α helices; (T) turns. Secondary structure in each sequence is indicated by background colors: (dark gray) α helices or 3_{10}-helices; (medium gray) β strands or β bridges; (light gray) turns. Other important features are: (black background) Cys residues involved in disulfide bridges or Asn, Thr, or Ser residues potentially involved in *N*- or *O*-glycosylation. CD residues involved in the active site[31] and in the first four subsites (this study) are indicated with: (a) catalytic acid; (b) catalytic base; (1, 2, 3, or 4) subsite. SBD residues involved in the binding sites as described earlier based on CGTase,[31] after obtaining the NMR structure,[16] or in this study are indicated with: (1 or 2) binding site. The residues conserved in more than 90% of the sequences are also shown. Subfamilies: **a: *Aspergillus*** (*A. awamori* var. *X100*; *A. shirousami*; *A. awamori* var. *kawachi*; *A. niger*/*A. awamori*; *A. niger T21*; *A. terreus*; *A. oryzae*; *C. paradoxa*; *C. rolfsii*; *H. grisea* var. *thermoidea*; *H. resinae P*; *H. resinae S*; *N. crassa*; *T. reesei*); **b: *Rhizopus*** (*R. oryzae*) **c: *Arxula*** (*A. adeninivorans*); **d: *Saccharomycopsis*** (*S. fibuligera*; *S. fibuligera KZ*); **e: *Saccharomyces*** (*S. cerevisiae*; *S. diastaticus*); **f: *Methanococcus*** (*M. jannaschii*); **g: *Clostridium*** (*Clostridium* sp. *G0005*). Minor errors were eliminated from different sequences manually.

A longer and less conserved loop is found between H7 and H8. Residues in this region are close to subsite 4 and support some active-site residues through both side-chain contacts[30] and participation in β sheet S5, but do not interact directly with the substrate in the first subsite. Large insertions appear to exist after the S5 β strand segment in *Clostridium* GA, and to a minor degree in *Saccharomycopsis* and *Saccharomyces* GAs.

No significant regular structural features occur in the active-site loop between helices H9 and H10, apart from a 3_{10}-helical segment derived from the STRIDE analysis in *A. niger* GA residues 311 to 313 (Fig. 2e). Several active-site residues are found around this segment,[30,31] interacting exclusively with subsites 1 and 2. All GAs except those of the *Aspergillus* subfamily and, most interestingly, *M. jannaschii* GA have an insertion of four residues following the segment that could affect both stability and selectivity, with *Clostridium* GA having a very long insert in addition. This region could have an important role in GA selectivity, mainly by affecting the second subsite, as the inclusion of four *H. resinae* GA *P* residues found here into *A. niger* GA mentioned earlier[76] has shown. In fact, interactions of some residues in the second subsite have been already described,[30,77] different substrates having slightly different interactions there. Furthermore, the large insert in *Clostridium* sp. *G0005* GA found here could be associated with its unusually high isomaltose hydrolytic activity.[59]

The last active-site loop is delimited by helices H12 and H13. A very structured loop, it contains several important active-site residues all especially important in the first subsite including the catalytic base Glu400.[10,78] Here a one-residue insertion around the second β strand of S9 is found only in *Clostridium* GA. Some selectivity properties might also be associated with the composition of this loop. The substitution of active-site residue Ser411, conserved in *Aspergillus*-related and other GAs, with a glycine residue, the only other alternative at the same position in the alignment, leads to a reduced ability to form isomaltose from glucose by *A. awamori* GA (T.-Y. Fang et al., unpublished results). A similar natural variation in this residue between the closely related *S. fibuligera* and *S. fibuligera KZ* GAs might explain some of their differences in kinetics and thermostability.[79]

Rhizopus, *Saccharomycopsis*, *Saccharomyces*, and *Methanococcus* GAs have their C-termini shortly after the end of helix H13, with slightly longer sequences in *Arxula* and *Clostridium* GAs. The *Aspergillus* subfamily, on the other hand, has a long partially *O*-glycosylated belt partially surrounding the CD that extends to Pro467, the last conserved residue in this domain in all but the *C. rolfsii* GA (Fig. 2g).[31] Deletions in the *O*-glycosylated region, comprising part of the belt and linker, from the SBD to the belt residue Ser460 lead to loss of secretion and stability in recombinant *A. awamori* GA.[80] The only GA in the *Aspergillus* subfamily produced by a Basidiomycetes, the recently reported *C. rolfsii* GA,[48] is missing part of the belt found in the remaining *Aspergillus*-related GAs, all belonging to Ascomycetes (a group that also includes yeasts), suggesting that this GA is more primitive than other members of the subfamily. Basidiomycetes and Ascomycetes are, however, closely related fungi.

The *C. rolfsii* GA CD possesses the Cys222–Cys449 (*A. niger* GA numbering) disulfide bridge (Fig. 2d,g) that connects the belt to the rest of the CD in *Aspergillus* subfamily GAs and appears to be an old feature helping to stabilize their most hydrophobic folding unit. It also contains the common Cys262–Cys270 bridge (Fig. 2e) and, along with *N. crassa* GA, an extra bridge between Cys20 and Cys27 (Fig. 2a). However, it is the only *Aspergillus*-related GA to lack the Cys210–Cys213 disulfide bond (Fig. 2d).

Starch-Binding Domain Analysis

The SBDs aligned in this study belong uniquely to *Aspergillus* subfamily GAs. HCA studies show their sequences to be mostly conserved,[31] suggesting that the corresponding structures all contain the eight β strands found in the NMR structures of the *A. niger* GA SBD.[15,16] Many conserved residues are found here, most near the binding sites assigned earlier,[16,31] while others appear to have a structural role. The major divergence is again that of *C. rolfsii* GA, which has two features that render it unique: 1) no disulfide bridges are found, unlike all other SBDs from *Aspergillus*-related GAs,[25,30,31] but like SBDs in other enzymes except an *A. niger* acid-stable α-amylase;[81] 2) a deletion of four residues in the second loop of binding site 2, using the numbering used with *B. circulans* strain *251* CGTase[36] and in NMR structural studies of the *A. niger* GA SBD.[16] This loop in the GAs is different from that in CGTase and in all other glycosidases possessing an SBD,[17,18] and these could explain differences in their abilities to degrade raw starch. In fact, while in CGTases binding site 2 is involved in guiding linear starch chains into the active site,[82] an action not likely to occur in GAs, where the SBD is either separated from the CD by a long linker or found opposite the active site, in GAs this site could be involved in the active solubilization of raw starch. Binding site 1, the site most conserved in GA SBDs,[31] is very similar in both CGTases and GAs and has an important role in raw starch binding in the former,[82] suggesting a similar role in the latter. In fact, a stronger role in raw starch binding by this binding site had already been indicated by kinetic analysis of *A. awamori* GAs with partially deleted SBDs.[83]

Beside allowing raw starch hydrolysis, SBDs also support GA adsorption to cell walls,[84,85] where a local increase of enzyme concentration may result in

enhanced glucose flow to the cell. At more favorable substrate concentrations, the linker can be cleaved proteolytically to release the CD. This has a parallel in the release of cell-bound *Clostridium thermosaccharolyticum* GA under rapid growth conditions.[86]

Role of *N*- and *O*-glycosylation

Only three potential *N*-glycosylation sites, all located in the CD, appear to be conserved in GAs from different subfamilies. They are not otherwise involved in secondary structural elements, given that they typically are located in backbone turns. Two sites, located at Asn171 and Asn395 following helices H5 and H12, respectively (Fig. 2c,g), have been identified following both peptide mapping[87] and x-ray crystallography[9] of *A. niger* and *A. awamori* var. *X100* GAs, respectively. The first site is shared by most *Aspergillus*-related and *Saccharomyces* GAs and the second by several *Aspergillus*-related and *Rhizopus* GAs. Elimination of the latter site in recombinant *A. awamori* GA greatly decreases enzyme secretion and thermostability.[88] A potential third site is shared by GAs from *N. crassa, H. resinae* (GA *P*), *Rhizopus*, and *Saccharomycopsis*, and also by *H. grisea* var. *thermoidea* GA using an alternative alignment[31] (Fig. 2b). Given that glycosylation is found in *Saccharomycopsis* GAs[79,89] that lack the two other conserved *N*-glycosylation sites and that this site is both at the surface and between secondary structural elements, it is probably *N*-glycosylated. The three-dimensional model suggests that glycosylation at this point in *A. awamori* var. *X100* and closely related GAs could make steric contacts with the linker and so negatively affect stability.

O-glycosylation is found in a 72-residue-long segment in *A. niger* GA comprising the belt circling part the CD, as well as in the linker and the N-terminus of the SBD.[31,87,90] Glycosylation density is lower in the belt than in the linker. This suggests that in the former it helps the rigidification of the CD or only the corresponding hydrophobic folding unit adjacent to it, with mostly hydrophobic residues anchoring the belt to the rest of the protein. Of the 10 *O*-glycosylation sites that occur in the CD of *A. awamori* var. *X100* GA, seven are found around the most hydrophobic unit. Extensive *O*-glycosylation further rigidifies the 40-residue-long linker in a solvated environment, contributing to the physical separation of the CD and SBD.[25–27] Such separation has been estimated to be of about 100 Å,[31] which agrees with the interdomain distance in the order of 95 Å obtained by scanning tunneling microscopy.[91] All this agrees with recent differential scanning microcalorimetry studies indicating that the carbohydrate moiety in *A. niger* GA and other glycoproteins has not only a stabilizing effect, but also prevents unfolded and partially folded protein from aggregating.[92]

Heavily glycosylated segments are predicted for all *Aspergillus* GAs except that from *A. oryzae*. GAs in the *Aspergillus* subfamily but not from the *Aspergillus* genus have shorter linkers with a predicted *O*-glycosylation pattern similar to that of the belt, indicating that their SBDs may be physically close to their CDs. Deletion studies in the linker region of *A. awamori* GA have shown its importance for high levels of GA expression,[93] so a long *O*-glycosylated segment might improve enzyme secretion. *O*-glycosylation of the N-terminal region of *S. diastaticus* GA also has a definite role in enzyme secretion.[94]

The microheterogeneity pattern observed in *O*-glycosylation[95] explains the diverse lengths of the carbohydrate moieties present.[90] Given that one of the roles of *O*-glycosylation is to protect the enzyme from proteolysis, those GA forms with lower degrees of glycosylation are more susceptible to protease action.[96] The preferred cleavage sites are found at the N-terminus of the SBD after Pro512 in *A. niger* GA,[6] or at the beginning of the linker region shortly after the conserved Pro467 in both *A. niger* GA[26,96] and *H. resinae* GA *P*.[97]

Glucoamylase Evolution

Parsimony analysis of the different complete GA protein sequences is shown in Figure 3. The sequence alignment leading to this evolutionary plot shows that *Clostridium* sp. *G0005* GA, the only GA of eubacterial origin, possesses uniquely long insertions and deletions compared to the remaining GAs, which are only partially present in the GA from the archaeon *M. jannaschii*. Given that the only other enzymes sharing the same architecture, the glycosyl hydrolases from Families 8 and 9,[20,21] are present only in bacteria, we suspect that the remote common origin of all these enzymes[22] is indeed bacterial. The assumption that all GAs are derived from a bacterial ancestral gene is further justified by their presence in both archaeal and fungal organisms.

Seven subfamilies of related GAs, *Clostridium* (bacterial), *Methanococcus* (archaeal), *Arxula*, *Saccharomyces*, and *Saccharomycopsis* (yeast), and *Rhizopus* and *Aspergillus* (filamentous fungal), have now been described. The evolution of fungal and yeast GAs has not followed phylogenetic lines, probably due to the multiplicity of GA forms existing in ancestral fungi. Only that can explain why *A. adeninivorans* and *R. oryzae* GAs share several features, including a putative newer-type SBD at the N terminus, even though Zygomycetes like *Rhizopus* diverged from Ascomyctes (which includes all other fungal and yeast species described here, with the exception the Basidomycetes *C. rolfsii*) before the appearance of yeast. The yeast GA subfamilies *Saccharomyces* and *Saccharomycopsis* are closely related to each other, sharing the apparent lack of SBDs.

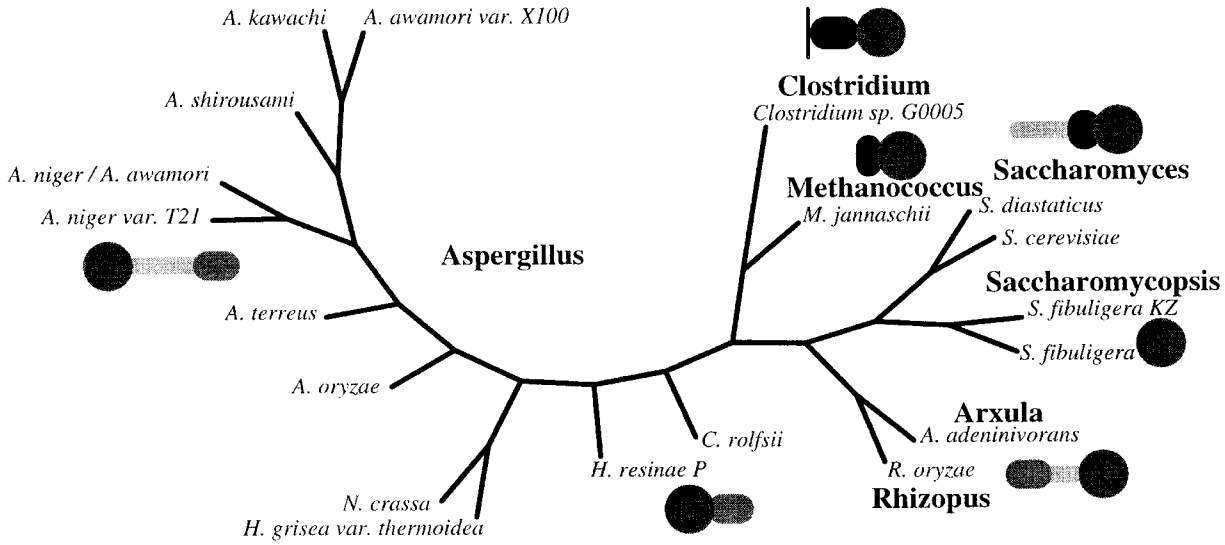

Fig. 3. The evolution of GA sequences from a common bacterial ancestral gene as obtained from protein parsimony analysis.[70,71] Microbial species are indicated in italics while GA subfamilies are indicated in bold. Approximate domain distribution is shown: (dark gray) CD; (medium gray) SBD; (clear) *O*-glycosylated region or linker; (black) other domains; (vertical bar) cell wall in the cell-bound *Clostridium* sp. *G0005* GA.

The *Aspergillus* subfamily definitely corresponds to a successful GA design. The evolutionary tree suggests that upon the attachment of a primitive SBD, exemplified by that of *C. rolfsii* GA, which lacks an internal disulfide bridge and has a gap in one of the binding sites, an evolution by stages occurred. The more evolved *H. resinae* GA *P* acquired both a full belt surrounding the CD and a disulfide bridge in the SBD (in fact a second disulfide bridge might also be present there). Longer linkers and more *O*-glycosylation in the belt surrounding the CD are found in *N. crassa* and *H. grisea* var. *thermoidea* GAs, with a variation in the position of one of the cysteines involved in the SBD disulfide bridge in the latter enzyme (Fig. 2i). A slightly shorter linker region is found in *A. oryzae* GA, which becomes a fully *O*-glycosylated linker in *A. terreus* GA. The latter also has two *N*-glycosylation sites not present in the former. A longer fully *O*-glycosylated linker is found in the remaining *Aspergillus* GAs, where now only minor variations in sequence occur.

Upon addition of the SBD, much of the later GA evolution appears to be oriented toward increased secretion and more effective GAs. The gradual increase in length and *O*-glycosylation of the belt surrounding the most hydrophobic unit of the CD suggests a gradual effort to increase its stability. Also, the increased length of the linker with its heavy glycosylation, beside affecting enzyme secretion,[93] can increase overall GA stability by physically separating two domains that unfold independently,[26,27] avoiding protein denaturation by contact of partially or totally unfolded CD and SBD domains in the same GA molecule.

CONCLUSIONS

In this study, the three CD hydrophobic folding units described by differential calorimetric studies of *A. niger* GA[27] were physically identified, which will help to guide efforts to genetically engineer more thermostable GAs. A structure-based alignment of known GA amino acid sequences was presented. Two new GA subfamilies were identified, from the archaeal *Methanococcus* and the yeast *Arxula*. Important structural elements in each sequence were identified, with a tentative correlation of these features with specific GA properties. The least hydrophobic folding unit identified in existing GA models is not found in the thermostable *M. jannaschii* and *Clostridium sp. G0005* GAs.

An evolutionary tree of GA sequences was generated by protein parsimony analysis, allowing identification of the traits that are likely to lead to increased stability and secretion in *Aspergillus*-related GAs. Better understanding of GA structure and stability derived from this study constitutes a base for protein engineering to improve the properties of industrially used GAs.

ACKNOWLEDGMENTS

We thank the U.S. Department of Energy through the Consortium for Plant Biotechnology Research, Genencor International, Inc., and Programa CIÊNCIA, JNICT, Lisboa, Portugal (P.M.C.) (B.D./1057/90-IF) for financial support.

REFERENCES

1. Manjunath, P., Shenoy, B.C., Raghavendra Rao, M.R. Fungal glucoamylases. J. Appl. Biochem. 5:235–260, 1983.

2. Vihinen, M., Mäntsälä, P. Microbial amylolytic enzymes. Crit. Rev. Biochem. Mol. Biol. 24:329–418, 1989.
3. Pandey, A. Glucoamylase research: An overview. Starch/Stärke 47:439–445, 1995.
4. Takahashi, T., Kato, K., Ikegami, Y., Irie, M. Different behavior towards raw starch of three forms of glucoamylase from a *Rhizopus* sp. J. Biochem. 98:663–671, 1985.
5. Ashikari, T., Nakamura, N., Tanaka, Y., Kiuchi, N., Shibano, Y., Tanaka, T., Amachi, T., Yoshizumi, H. *Rhizopus* raw-starch-degrading glucoamylase: Its cloning and expression in yeast. Agric. Biol. Chem. 50:957–964, 1986.
6. Svensson, B., Larsen, K., Gunnarsson, A. Characterization of a glucoamylase G2 from *Aspergillus niger*. Eur. J. Biochem. 154:497–502, 1986.
7. Neustroev, K.N., Firsov, L.M. Acid proteinase and the multiplicity of forms of glucoamylase from *Asp. awamori*. Biokhimiya 55:776–785, 1990.
8. Dalmia, B.K., Nikolov, Z.L. Characterization of glucoamylase adsorption to raw starch. Enzyme Microb. Technol. 13:982–990, 1991.
9. Aleshin, A., Golubev, A., Firsov, L.M., Honzatko, R.B. Crystal structure of glucoamylase from *Aspergillus awamori* var. *X100* to 2.2-Å resolution. J. Biol. Chem. 267:19291–19298, 1992.
10. Harris, E.M.S., Aleshin, A.E., Firsov, L.M., Honzatko, R.B. Refined structure for the complex of 1-deoxynojirimycin with glucoamylase from *Aspergillus awamori* var. *X100* to 2.4-Å resolution. Biochemistry 32:1618–1626, 1993.
11. Aleshin, A.E., Hoffman, C., Firsov, L.M., Honzatko, R.B. Refined crystal structures of glucoamylase from *Aspergillus awamori* var. *X100*. J. Mol. Biol. 238:575–591, 1994.
12. Aleshin, A.E., Firsov, L.M., Honzatko, R.B. Refined structure for the complex of acarbose with glucoamylase from *Aspergillus awamori* var. *X100* to 2.4-Å resolution. J. Biol. Chem. 269:15631–15639, 1994.
13. Stoffer, B., Aleshin, A.E., Firsov, L.M., Svensson, B., Honzatko, R.B. Refined structure of the complex of D-*gluco*-dihydroacarbose with glucoamylase from *Aspergillus awamori* var. *X100* to 2.2 Å resolution: Dual conformations for extended inhibitors bound to the active site of glucoamylase. FEBS Lett. 358:57–61, 1995.
14. Aleshin, A.E., Stoffer, B., Firsov, L.M., Svensson, B., Honzatko, R.B. Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: Relationship of stereochemical distortions at the non-reducing end to the catalytic mechanism. Biochemistry 35:8319–8328, 1996.
15. Jacks, A.J., Sorimachi, K., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D.B., Williamson, M.P. ^{1}H and ^{15}N assignments and secondary structure of the starch-binding domain of glucoamylase from *Aspergillus niger*. Eur. J. Biochem. 233:568–578, 1995.
16. Sorimachi, K., Jacks, A.J., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D.B., Williamson, M.P. J. Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. Mol. Biol. 259:970–987, 1996.
17. Svensson, B., Jespersen, H., Sierks, M.R., MacGregor, E.A. Sequence homology between putative starch binding domains from different starch degrading enzymes. Biochem. J. 264:309–311, 1989.
18. Jespersen, H.M., MacGregor, E.A., Sierks, M.R., Svensson, B. Comparison of the domain-level organization of starch hydrolases and related enzymes. Biochem. J. 280:51–55, 1991.
19. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequnce similarities. Biochem. J. 280:309–316, 1991.
20. Henrissat, B., Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequnce similarities. Biochem. J. 293:781–788, 1993.
21. Henrissat, B., Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316:695–696, 1996.
22. Alzari, P.M., Souchon, H., Dominguez, R. The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. Structure 4:265–275, 1996.
23. Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.L., Claeyssens, M., Béguin, P., Aubert, J.-P. Three-dimensional structure of a thermostable bacterial cellulase. Nature 357:89–91, 1992.
24. Hehre, E.J., Okada, G., Genghof, D.S. Configurational specificity: Unappreciated key to understanding enzymic reversions and de novo glycosidic bond synthesis. I. Reversal of hydrolysis by alpha-, beta- and glucoamylases with donors of correct anomeric form. Arch. Biochem. Biophys. 135:74–89, 1969.
25. Williamson, G., Belshaw, N.J., Williamson, M.P. *O*-glycosylation in *Aspergillus* glucoamylase. Conformation and role in binding. Biochem. J. 282:423–428, 1992.
26. Williamson, G., Belshaw, N.J., Noel, T.R., Ring, S.G., Williamson, M.P. *O*-glycosylation and stability. Unfolding of glucoamylase induced by heat and guanidine hydrochloride. Eur. J. Biochem. 207:661–670, 1992.
27. Tanaka, A., Fukuda, H., Takahashi, K. Differential scanning calorimetric studies on the domain structure of *Aspergillus* glucoamylase. J. Biochem. 117:1024–1028, 1995.
28. Tanaka, Y., Ashikari, T., Nakamura, N., Kiuchi, N., Shibano, Y., Amachi, N., Yoshizumi, H. Comparison of amino acid sequences of three glucoamylases and their structure function relationships. Agric. Biol. Chem. 50:965–969, 1986.
29. Itoh, T., Ohtsuki, I., Yamashita, I., Fukui, S. Nucleotide sequence of the glucoamylase gene GLU1 in the yeast *Saccharomycopsis fibuligera*. J. Bacteriol. 169:4171–4176, 1987.
30. Coutinho, P.M., Reilly, P.J. Structure-function relationships in the catalytic and starch-binding domains of glucoamylase. Protein Eng. 7:393–400, 1994.
31. Coutinho, P.M., Reilly, P.J. Structural similarities in glucoamylases by hydrophobic cluster analysis. Protein Eng. 7:749–760, 1994.
32. Henrissat, B., Coutinho, P.M., Reilly, P.J. Reading-frame shift in *Saccharomyces* glucoamylases restores catalytic base, extends sequence and improves alignment with other glucoamylases. Protein Eng. 7:1281–1282, 1994.
33. Savel'ev, A.N., Sergeev, V.R., Firsov, L.M. Study of the active site of glucoamylase from *Aspergillus awamori*. Biokhimiya 47:390–397, 1982.
34. Meagher, M.M., Nikolov, Z.L., Reilly, P.J. Subsite mapping of *Aspergillus niger* glucoamylases I and II with malto- and isomaltooligosaccharides. Biotechnol. Bioeng. 34:681–688, 1989.
35. Ermer, J., Rose, K., Hübner, G., Schellenberger, A. Subsite affinities of *Aspergillus niger* glucoamylase II determined with p-nitrophenylmaltooligosaccharides. Biol. Chem. Hoppe-Seyler 374:123–128, 1993.
36. Lawson, C.L., van Monfort, R., Strokopytov, B., Rozeboom, H.J., Kalk, K.H., de Vries, G.E., Penninga, D., Dijkhuizen, L., Dijkstra, B.W. Nuceolotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form.J. Mol. Biol. 236:590–600, 1994.
37. Knegtel, R.M.A., Strokopytóv, B., Penninga, D., Faber, O.G., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L., Dijkstra, B.W. Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 with natural products and substrates. J. Biol. Chem. 270:29256–29264.
38. Tsai, C.J., Nussinov, R. Hydrophobic folding units derived from dissimilar monomer structures and their interactions. Prot. Sci. 6:24–42, 1997.
39. Frishman, D., Argos, P. Knowledge-based protein secondary structure assignment. Proteins 23:566–579, 1995.
40. Nunberg, J.H., Meade, J.H., Cole, G., Lawyer, F.C., McCabe, P., Schweickart, V., Tal, R., Wittman, V.P., Flatgaard, J.E., Innis, M.A. Molecular cloning and characterization of the glucoamylase gene of *Aspergillus awamori*. Mol. Cell. Biol. 4:2306–2315, 1984.
41. Hayashida, S., Kuroda, K., Ohta, K., Kuhuara, S., Fukuda, K., Sakaki, Y. Molecular cloning of the glucoamylase I gene of *Aspergillus awamori* var. *kawachi* for localization of

raw-starch-affinity site.Agric. Biol. Chem. 54:923–929, 1989.
42. Diagne, I., Pechexonov, V.T., Bulat, S.A., Firsov, L.M. Genbank Entry # U59303, 1996.
43. Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K.E., Fiil, N.P. Glucoamylases G1 and G2 from *Aspergillus niger* are synthesized from two different but closely related mRNAs. EMBO J. 3:1097–1102, 1984.
44. Zhong, L., Tang, G., Yang, K. Isolation and sequencing of glucoamylase gene from a glucoamylase overproducing strain. Wei Sheng Wu Hsueh Pao 34:184–190, 1994.
45. Hata, Y., Tsuchiya, K., Kitamoto, K., Gomi, K., Kumagai, C., Tamura, G., Hara, S. Nucleotide sequence and expression of the glucoamylase-encoding gene (glaA) from *Aspergillus oryzae*. Gene 108:145–150, 1991.
46. Shibuya, I., Gomi, K., Iimura, Y., Takahashi, K., Tamura, G., Hara, S. Molecular cloning of the glucoamylase gene of *Aspergillus shirousami* and its expression in *Aspergillus oryzae*. Agric. Biol. Chem. 54:1905–1914, 1990.
47. Ventura, L., Gonzalez, C.L., Perez, G.J., Ramon, D. Molecular cloning and transcriptional analysis of the *Aspergillus terreus* gla1 gene encoding a glucoamylase. Appl. Environ. Microbiol. 61:399–402, 1995.
48. Nagasaka, Y., Muraki, N., Kimura, A., Suto, M., Yokota, A., Tomita, F. Cloning of *Corticium rolfsii* glucoamylase cDNA and its expression in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 44:451–458, 1995.
49. Joutsjoki, V.V., Torkkeli, T.K. Glucoamylase P gene of *Hormoconis resinae*: Molecular cloning, sequencing and introduction into *Trichoderma reesei*. FEMS Microbiol. Lett. 99:237–244, 1992.
50. Berka, R.M., Rey, M.W., Thompson, S.A., Gray, G.L., Carmona, C.L., Power, S.D. GenBank, Entry # HUIGLA1A, 1993.
51. Stone, P.J., Makoff, A.J., Parish, J.H., Radford, A. Cloning and sequence analysis of the glucoamylase gene of *Neurospora crassa*. Curr. Genet. 24:205–211, 1993.
52. Bui, D.M., Kunze, I., Forsters, S., Watarmann, T., Horstmann, C., Manteuffel, R., Kunze, G. Cloning and expression of an *Arxula adeninivorans* glucoamylase gene in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 44: 610–619, 1996.
53. Yamashita, I., Nakamura, M., Fukui, S. Gene fusion is a possible mechanism underlying the evolution of STA1. J. Bacteriol. 169:2142–2149, 1987.
54. Barrell, B.G., Badcock, K., Bankier, A.T., Bowman, S., Brown, D., Churcher, C.M., Connor, R., Copsey, T., Dear, S., Devlin, K., Fraser, A., Gentles, S., Hamlyn, N., Horsnell, T.S., Hunt, S., Jagels, K., Jones, M., Louis, E., Lye, G., Moule, S., Moule, T., Odell, C., Pearson, D., Rajandream, M.A., Riles, L., Rowley, N., Skelton, J., Smith, V., Walsh, S.V., Whitehead, S. EMBL Entry # Z38125, 1994.
55. Yamashita, I., Suzuki, K., Fukui, S. Nucleotide sequence of the extracellular glucoamylase gene STA1 in the yeast *Saccharomyces diastaticus*. J. Bacteriol. 161:567–573, 1985.
56. Hostinová, E., Balanová, J., Gašperík, J. The nucleotide sequence of the glucoamylase gene GLA1 from *Saccharomycopsis fibuligera* KZ. FEMS Microb. Lett. 83:103–108, 1991.
57. Pearson, D., Churcher, C.M. EMBL Entry # Z67961, 1995.
58. Dohmen, R.J., Strasser, A.W.M., Dahlmens, U.M., Hollenberg, C.P. Cloning of the *Schwanniomyces occidentalis* glucoamylase gene (GAM1) and its expression in *Saccharomyces cerevisiae*. Gene 95:111–121, 1990.
59. Ohnishi, H., Kitamura, H., Minowa, T., Sakai, H., Ohta, T. Molecular cloning of a glucoamylase gene from a thermophilic *Clostridium* and kinetics of the cloned enzyme. Eur. J. Biochem. 207:413–418, 1992.
60. Bult, C.J. White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, H.-P., Fraser, C.M., Smith, H.O., Woese, C.R., Venter, J.C. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. Science 273:1058–1073, 1996.
61. Fagerström, R. Subsite mapping of *Hormoconis resinae* glucoamylases and their inhibition by gluconolactone. J. Gen. Microbiol. 137:1001–1008, 1991.
62. Fagerström, R., Kalkkinen, N. Characterization, subsite mapping and partial amino acid sequence of glucoamylase from the filamentous fungus *Trichoderma reesei*. Biotechnol. Appl. Biochem. 21:223–231, 1995.
63. Inokuchi, N., Takahashi, T., Irie, M. Purification and characterization of a minor glucoamylase from *Aspergillus saitoi*. J. Biochem. 90:1055–1067, 1981.
64. Abe, J., Nakajima, K., Hizukuri, S. Alteration of the properties of *Aspergillus* sp. *K-27* glucoamylase on limited proteolysis with subtilisin. Carbohydr. Res. 203:129–138, 1990.
65. Monma, M., Kainuma, K. Heterogeneity of the glucoamylase components of the raw-starch-digesting amylase from *Chalara paradoxa*. Carbohydr. Res. 227:385–388, 1992.
66. Ebbole, D. J., Madi, L. Identification of a gene from *Neurospora crassa* with similarity to a glucoamylase gene from *Schwanniomyces occidentalis*. Fungal Genetics Newsletter 43:23–24, 1996.
67. De Rijk, P., De Wachter, R. DCSE, an interactive tool for sequence and secondary structure research. Comput. Appl. Biosci. 9:735–740, 1993.
68. Gaboriaud, C., Bissery, V., Benchtrit, T., Mornon, J.-P. Hydrophobic cluster analysis: An efficient new way to compare and analyse amino acid sequences. FEBS Lett. 224:149–155, 1987.
69. Lemesle-Verloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A., Mornon, J.-P. Hydrophobic cluster analysis: Procedures to derive structural and functional information from 2-D-representation of protein sequences. Biochimie 72:555–574, 1990.
70. Felsenstein, J. Numerical methods for inferring evolutionary trees. Quart. Rev. Biol. 57:379–404, 1982.
71. Felsenstein, J. Phylogenies from molecular sequences: Inference and reliability. Annu. Rev. Genet. 22:521–565, 1988.
72. Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:946–950, 1991.
73. Chen, H.-M., Li, Y., Panda, T., Buehler, F.U., Ford, C., Reilly, P.J. Effect of replacing helical glycine residues with alanines on reversible and irreversible stability and production of *Aspergillus awamori* glucoamylase. Protein Eng. 9:499–505, 1996.
74. Sierks, M.R., Ford, C., Reilly, P.J., Svensson, B. Site-directed mutagenesis at the active site Trp120 of *Aspergillus awamori* glucoamylase. Protein Eng. 2:621–625, 1989.
75. Sierks, M.R., Ford, C., Reilly, P.J., Svensson, B. Catalytic mechanism of fungal glucoamylase as defined by mutagenesis of Asp176, Glu179, and Glu180 in the enzyme from *Aspergillus awamori*. Protein Eng. 3:193–198, 1990.
76. Fierobe, H.-P., Stoffer, B.B., Frandsen, T.P., Svensson, B. 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 35:8696–8704, 1996.
77. Lemieux, R.U., Spohr, U., Bach, M., Cameron, D.R., Frandsen, T.P., Svensson, B., Palcic, M.M. Chemical mapping of the active site of the glucoamylase of *Aspergillus niger*. Can. J. Chem. 74:319–335, 1996.
78. Frandsen, T.P., Dupont, C., Lehmbeck, J., Stoffer, B., Sierks, M.R., Honzatko, R.B., Svensson, B. Replacement of the catalytic base Glu 400 and two neighboring residues, Tyr 48 and Gln 401, hydrogen bonded to the γ-carboxylate group of Glu 400 in glucoamylase from *Aspergillus niger*. Biochemistry 33:13808–13816, 1994.
79. Gašperík, J., Hostinová, E. Glucoamylases encoded by variant *Saccharomycopsis fibuligera* genes: Structure and properties. Curr. Microbiol. 27:11–14, 1993.

80. Evans, R., Ford, C., Sierks, M., Nikolov, Z., Svensson, B. Deletion analysis of the *O*-glycosylated region of *Aspergillus* glucoamylase: Effect on activity and thermal stability. Gene 91:131–134, 1990.
81. Kaneko, A., Sudo, S., Takayasusakamoto, Y., Tamura, G., Ishikawa, T., Oba, T. J. Molecular-cloning and determination of nucleotide-sequence of a gene encoding an acid-stable α-amylase from *Aspergillus kawachii*. Ferment. Bioeng. 81:292–298, 1996.
82. Penninga, D., van der Veen, B.A., Knegtel, R.M.A., van Hijum, S.A.F.T., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W., Dijkhuizen, L. The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. J. Biol. Chem. 271:32777–32784, 1996.
83. Chen, L., Coutinho, P.M., Nikolov, Z., Ford, C. Deletion analysis of the starch binding domain of *Aspergillus* glucoamylase. Protein Eng. 8:1049–1055, 1995.
84. Wösten, H.A., Moukha, S.M., Sietsma, J.H., Wessels, J.G. Localization of growth and secretion of proteins in *Aspergillus niger*. J. Gen. Microbiol. 137:2017–2023, 1991.
85. Neustroev, K.N., Valter, S.N., Timchenko, N.V., Firsov, L.M., Golubev, A.M., Khokhlov, S.E. Adsorption of glucoamylase from *Aspergillus awamori* X-100/D27 on cell walls. Biochem. Mol. Biol. Internat. 30:115-120, 1993.
86. Specka, U., Mayer, F. Purification and properties of a thermoactive glucoamylase from *Clostridium thermosaccharolyticum*. Arch. Microbiol. 160:284–287, 1993.
87. Svensson, B., Larsen, K., Svendsen, I., Boel, E. The complete amino acid sequence of the glycoprotein, glucoamylase G1, from *Aspergillus niger*. Carlsberg Res. Commun. 48:529–544, 1983.
88. Chen, H.-M., Ford, C., Reilly, P.J. Substitution of Asn residues in *Aspergillus awamori* glucoamylase by site-directed mutagenesis to eliminate *N*-glycosylation and inactivation by deamidation. Biochem. J. 301:275–281, 1994.
89. Solovicová, A., Gašperík, J., Hostinová, E. High-yield production of *Saccharomycopsis fibuligera* glucoamylase in *Escherichia coli*, refolding, and comparison of the nonglycosylated and glycosylated enzyme forms. Biochem. Biophys. Res. Commun. 224:790–795, 1996.
90. Gunnarsson, A., Svensson, B., Nilsson, B., Svensson, S. Structural studies on the *O*-glycosidically linked carbohydrate chains of glucoamylase G1 from *Aspergillus niger*. Eur. J. Biochem. 145:463–468, 1984.
91. Kramer, G.F.H., Gunning, A.P., Morris, V.J., Belshaw, N.J., Williamson, G. Scanning tunnelling microscopy of *Aspergillus niger* glucoamylase. J. Chem. Soc. Faraday Trans. 89:2595–2602, 1993.
92. Wang, C., Eufimi, M., Turano, C., Giartosino, A. Influence of the carbohydrate moiety on the stability of glycoproteins. Biochemistry 35:7299–7307, 1996.
93. Libby, C.B., Cornett, C.A.G., Reilly, P.J., Ford, C. Effect of amino acid deletions in the *O*-glycosylated region of *Aspergillus awamori* glucoamylase. Protein Eng. 7:1109–1114, 1994.
94. Yamashita, I. The threonine- and serine-rich tract of the secretory glucoamylase can direct β-galactosidase to the cell envelope. Agric. Biol. Chem. 53:483–489, 1989.
95. Neustroev, K.N., Golubev, A.M., Ibatullin, F.M., Mosichuk, A.V. Microheterogeneity in O-type sugar chains of carbohydrases secreted by *Aspergillus awamori*. Biochem. Mol. Biol. Internat. 30:107–113.
96. Le Gal-Coëffet, M.-F., Jacks, A. J., Sorimachi, K., Williamson, M. P., Williamson, G., Archer, D. B. Expression in *Aspergillus niger* of the starch-binding domain of glucoamylase. Comparison with the proteolytically produced starch-binding domain. Eur. J. Biochem. 233:561–567, 1995.
97. Fagerström, R. Evidence for a polysaccharide-binding domain in *Hormoconis resinae* glucoamylase P: Effects of its proteolytic removal on substrate specificity and inhibition by β-cyclodextrin. J. Gen. Microbiol. 140:2399–2407, 1994.