Carlsberg Res. Commun. Vol 48, p. 517–527, 1983

AMINO ACID SEQUENCE OF TRYPTIC FRAGMENTS OF GLUCOAMYLASE G1 FROM ASPERGILLUS NIGER

by

BIRTE SVENSSON, KJELD LARSEN[1)]

and

IB SVENDSEN

Department of Chemistry, Carlsberg Laboratory,
Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby

[1)] Present address: Harvard Medical School,
Center for Biochemical and Biophysical Sciences and Medicine,
Seeley G. Mudd Bldg., 250 Longwood Ave., Boston, Mass. 02115

Keywords: Glycopeptides, deglycosylation, sites of glycosylation, high pressure liquid chromatography.

The glycoprotein glucoamylase (EC 3.2.1.3) G1 from Aspergillus niger was digested with trypsin after 2-pyridylethylation and the resulting peptide fragments were separated by gel filtration, followed by reverse phase HPLC. A different set of peptide fragments was obtained from the citraconylated, 2-pyridylethylated enzyme. These were separated by gel filtrations, affinity chromatography on Con A-Sepharose, and reverse phase HPLC. The amino acid sequence of the isolated peptide fragments was determined by automated Edman degradation and digestion with carboxypeptidases Y and B.

The majority of the carbohydrate of glucoamylase G1 was located in a fragment which carried approximately 35 units of neutral sugar linked O-glycosidically to threonine and serine residues, while a minor fraction was located in a different tryptic fragment which contained a single N-glycosylated asparagine residue.

Abbreviations: ca = citraconyl; CHO = carbohydrate; Con A = concanavalin A; DFP = diisopropylfluorophosphate; DPCC = diphenylcarbamyl chloride; EDTA = ethylenediaminetetraacetic acid, disodium salt; SDS = sodium dodecyl sulfate; HPLC = high pressure liquid chromatography; PTH- = phenylthiohydantoin-; 2-pe- = 2-pyridylethyl- ; d signifies deglycosylated. G1 designates the larger of the two forms of glucoamylase from A. niger (31).

Springer-Verlag 0105-1938/83/0048/0517/$02.20

1. INTRODUCTION

Fungal glucoamylases (1,4-α-D-glucan glucohydrolase, EC 3.2.1.3) generally exist in multiple forms of which only the larger, designated G1, has the capacity to adsorb to and digest raw starch (36). Aspergillus niger, thus produces two forms of glucoamylase, G1 and G2, which structurally are very similar, since degradation of these forms with cyanogen bromide resulted in fragments with identical N-terminal sequences. However, the G1 form had a C-terminal fragment which was lacking in G2 (31). In order to investigate the relationship between structure and function of glucoamylases, a more detailed characterization of the primary structure of Aspergillus niger G1 was initiated and the present study describes the amino acid sequence of tryptic fragments.

Sites of glycosylation have been determined for two of the peptide fragments: one which carried a single N-glycosidic oligosaccharide, containing amino sugar, and another which contained the majority of the neutral carbohydrate of G1 in the form of short sugar units linked to a high number of serine and threonine residues.

This work has been presented in part at the XIth International Carbohydrate Symposium, Vancouver, August 1982.

2. MATERIALS AND METHODS

2.1. Materials

Glucoamylase G1 was purified as earlier described (31) from AMG 200L (Batch ASN 5192), a commercial Aspergillus niger glucoamylase preparation obtained as a gift from Novo Industries, Bagsværd, Denmark. The enzyme was homogeneous both by SDS polyacrylamide gel electrophoresis and by N-terminal sequence analysis (31).

Bio-Gels P-100, P-60, P-10 and P-6 were from the Bio-Rad Laboratories, Richmond, Calif. Con A-Sepharose was a product of Pharmacia Fine Chemicals, Uppsala, Sweden. Dithiothreitol, α-methyl-D-mannoside, DFP-treated carboxypeptidase B and DPCC-treated trypsin were from Sigma, St.Louis, Mo., 2-vinylpyridine was from Aldrich-Europe, Beerse, Belgium, citraconic anhydride and 2-propanol z. Anal. were from Merck, Darmstadt, F.R.G., 1-propanol z. Anal. was from Ferak, Berlin, acetonitrile HPLC Grade SS and trifluoroacetic acid Sequence Grade were from Rathburn, Walkerburn, Scotland. Crystalline thermolysin (thermophilic bacterial protease) was a product of Daiwa Kasei K.K., Osaka, Japan. Carboxypeptidase Y was prepared in this laboratory. Hydrogen fluoride was from Gerlingholz, Hamburg, F.R.G. Reagents and solvents used in the sequencing were from Rathburn except ethylacetate which was obtained from Merck.

2.2. Methods

2.2.1. Chemical modifications

The disulfide bridges of glucoamylase G1 were reduced and 2-pyridylethylated by a modification of the procedure described by FRIEDMAN et al. (9). G1 (500 mg) was dissolved, and reduced for 2 hours by dithiothreitol (100 mg) in 7 M-guanidinium chloride, 0.2 M-Tris and 5 mM-EDTA, pH 7.6 (25 ml) which had been flushed with nitrogen. Three portions of 2-vinylpyridine (3 · 250 µl) were added to this solution over a period of 30 minutes. The pH of the reaction mixture was brought to 4.0 by addition of glacial acetic acid and excess of reagents were removed by dialysis against 2% acetic acid, 5% in ethanol. This stock solution of reduced and 2-pyridylethylated G1 was stored at -18 °C.

The primary amino groups of 2-pyridylethylated G1 were citraconylated essentially as described by DIXON and PERHAM (2). Typically 120 mg of the protein was dissolved in 50 ml of 6 M-guanidinium chloride and 100 mg of citraconic anhydride was added over 15 minutes. The pH was maintained at 8.2 by titration with 1 M-sodium hydroxide in a pH-stat. When base uptake had ceased the pH was adjusted to 7.0 and the solution was made 0.1 M in hydroxylamine. Following three hours at room temperature the reaction mixture was finally dialyzed against 0.1 M-ammonium bicarbonate pH 7.8.

Complete deglycosylation of O-glycosylated peptides was performed a.m. MORT and LAMPORT (22). The glycopeptide (1-10 mg) was treated for 3 hours at 0 °C with 1 ml anhydrous hydrogen fluoride containing 100 µl of anisole. The HF was removed in the cold by aspiration with a water pump. To the residue was added water, followed by neutralization with ammonia, and desalting by dialysis against 0.05 M-ammonium

bicarbonate or by gel filtration on Bio-Gel P-6 in 30% acetic acid.

2.2.2. Enzymatic cleavages

2-Pyridylethylated G1, or G1 which had both been citraconylated and 2-pyridylethylated, (1.6 mg · ml^{-1}) was typically digested with trypsin (1:50, wt/wt) in 0.1 M-ammonium bicarbonate, 0.1 mM-calcium chloride pH 7.8 for 2 hours at 37 °C. A resulting highly glycosylated tryptic fragment (Tca2-1) was after deglycosylation and decitraconylation in HF, further digested with trypsin at the same conditions as above. Tca2-1 was also cleaved by thermolysin (1:30, wt/wt) in 0.1 M-ammonium bicarbonate pH 7.8 for 6 hours at 37 °C before, as well as after, deglycosylation.

2.2.3. Purification of peptides

The tryptic digests of G1 were initially fractionated on Bio-Gel P-100 in 0.2 M-ammonium bicarbonate pH 7.8. Thermolytic peptides of the fragment Tca2-1 were separated on Bio-Gel P-60 in 0.1 M-ammonium bicarbonate.

Glycopeptides in pool 2 from the Bio-Gel P-100 chromatography (10-20 mg) were adsorbed to a column of Con A-Sepharose (1.5×5 cm) equilibrated in 0.1 M-sodium acetate, 0.2 M-sodium chloride pH 6.0, containing 1 mM of Mn^{2+}, Mg^{2+}, and Ca^{2+}, respectively. Elution was performed with the same buffer containing 0.2 M-α-methyl-D-mannoside. The glycopeptides subsequently were desalted on Bio-Gel P-10 in 0.05 M-ammonium bicarbonate followed by lyophilization. The highly glycosylated fragment, Tca2-1, was finally obtained after gel filtration on Bio-Gel P-100 (1.5×90 cm) in 0.2 M-ammonium bicarbonate.

Peptides obtained after initial gel filtration on Bio-Gel P-100, or after tryptic digests of HF-treated Tca2-1 (see *2.2.2.*), were fractionated by HPLC using a Waters liquid chromatograph consisting of a Model U6 K injector, a Model 660 solvent programmer, two Model 6000A pumps, a Model 450 variable wavelength detector, and a Hewlett-Packard 3390 A integrator. Three support and solvent systems (20) were used: 1) a Waters Radial Pak C_{18} (10 μ) reverse phase column was eluted with a linear gradient over 40 minutes from 0.1% trifluoroacetic acid to 0.1% trifluoroacetic acid in 80% acetonitrile; 2) a Wide Pore C_{18} column, i.d. 4.6 mm and length 25 cm, from J.T. Baker Research Products, was eluted over 80 minutes by a linear gradient as above using 80% 2-propanol as organic phase; 3) the Wide Pore C_{18} column was eluted as above with 80% 1-propanol as organic phase. The flow rate was 0.5 ml · min^{-1} and the effluent was monitored at 220 or 230 nm. Fractions were collected and lyophilized prior to rechromatography or amino acid and sequence analyses. For rechromatography, narrow gradients (corresponding to 12% increase in the concentration of organic solvent) were used covering the predicted elution conditions.

2.2.4. Amino acid analysis

Amino acid compositions were determined as described previously (31) using a Durrum D-500 amino acid analyzer after acid hydrolysis for 24 hours at 110 °C. For certain peptides, hydrolysis was allowed to proceed for 48 and 72 hours. The colour yield of S-β-(2-pyridylethyl)-L-cysteine was determined to 67% of the value of lysine. An estimate of the number of O-glycosylated threonines and serines was made after reductive β-elimination a.m. DOWNS and PIGMAN (3).

2.2.5. Carbohydrate analyses

Total carbohydrate analysis was performed by the phenolsulphuric acid procedure of DUBOIS et al. (4) using mannose as standard. Glucosamine was determined on the amino acid analyzer after hydrolysis in 4 M-HCl at 110 °C for 6 hours in vacuo.

2.2.6. Amino acid sequence analysis

Peptides were sequenced in a Beckman Sequencer 890C as described by EDMAN and BEGG (6) using the conditions specified by JOHANSEN et al. (15). Polybrene was added when shorter peptides were sequenced (16). The PTH-amino acids were identified as described by SVENDSEN et al. (30). In addition, selected samples were subjected to back-hydrolysis with HCl/$SnCl_2$ (21) followed by amino acid analysis. Dansyl-Edman degradation was performed essentially as described by KLEMM (17) and GRAY (10).

The highly glycosylated fragment Tca2-1 (1.8 mg · ml^{-1}) was digested with carboxypeptidase Y

Table I.
Amino acid sequence of tryptic peptides from 2-pyridylethylated glucoamylase G1.

Pool number (Figure 1)	Fragment	Sequence of peptide	Elution conditions (%, vol/vol)
2	T2-1 (Tca2-1)	Asn - Ser - Val - Val - Pro - Ala - Ser - Trp - Gly - Glu - Thr - Ser* - Ala - Ser* - Ser* - (5, 10, 15) Val - Pro - Gly - Thr - Cys - Ala - Ala - Thr* - Ser* - Ala - Ile - Gly - Thr - Tyr - Ser* - (20, 25, 30) Ser* - Val - Thr* - Val - Thr* - Ser* - Trp - Pro - Ser* - Ile - Val - Ala - Thr* - Gly - Gly - (35, 40, 45) Thr* - Thr* - Thr* - Thr* - Ala - Thr - Pro - Thr* - Gly - Ser* - Gly - Ser* - Val - Thr* - Ser* - (50, 55, 60) Thr* Ser* Lys - Thr - Thr* - Ala - Thr - Ala - Ser* - Lys, Thr - Ser - Thr - Ser - Thr - (65, 70, 75) Ser - Ser - Thr - Ser - Cys - Thr - Thr (Thr, Pro, Ala) Val - Ala - Val - Thr - Phe (80, 85, 90)	40 CH_3CN
5	T5-1	Thr - Phe - Ala - Asp - Gly - Phe - Val - Ser - Ile - Val - Glu - Thr - His - Ala - Ala - (5, 10, 15) CHO Ser - Asn - Gly - Ser - Met - Ser - Glu - Gln - Tyr - Asp - Lys (20, 25)	48 CH_3CN 39 2-prOH
	T5-2	Asn - Gly - Asp - Thr - Ser - Leu - Leu - Ser - Thr - Ile - Glu - Asn - Tyr - Ile - Ser - (5, 10, 15) Ala - Gln - Ala - Ile - Val - Gln - Gly - Ile - Ser - Asn - Pro (Ser) Gly - Asp - Leu - (20, 25, 30) Ser - Ser - Gly - Ala - Gly - Leu - Gly - Glu - Pro - Lys (35, 40)	58 CH_3CN 45 2-prOH
6	T6-1	Ile - Glu - Ser - Asp - Asp - Ser - Val - Glu - Trp - Glu - Ser - Asp - Pro - Asn - Arg (5, 10, 15)	32 CH_3CN 24 2-prOH 13 1-prOH
	T6-2	Ser - Ile - Tyr - Thr - Leu - Asn - Asp - Gly - Leu (Asx, Ser, Ser) Glu, Ala, Val, (5, 10, 15) Ala - Val - Gly - Arg	42 CH_3CN
	T6-3	Asp - Ala - Asn - Thr - Leu - Leu - Gly - Ser - Ile - His - Thr - Phe - Asp - Pro (Asx, (5, 10, 15) Asx, Thr, Ser, Glx, Glx, Pro, Ala, Ala, Phe, Cys) Cys - Ser - Pro - Arg (20, 25)	36 2-prOH
	T6-4	Thr - Ala - Ile - Leu - Asn - Asn - Ile - Gly - Ala - Asp - Gly - Ala - Trp - Val - Ser - (5, 10, 15) Gly - Ala - Asp - Ser - Gly - Ile - Val - Val - Ala - Ser - Pro - Ser - Thr - Asp - Asn - (20, 25, 30) Pro (Thr) Tyr - Phe - Tyr - Thr - Arg (35)	54 CH_3CN 38 2-prOH
7	T7-1	Ser - Asp - Gly - Glu - Gln, Leu, Ser, Ala, Arg (5)	24 CH_3CN 6 1-prOH
	T7-2	Asp - Ser - Gly - Leu - Val - Leu - Lys (5)	34 CH_3CN

(cont.)

Table I. *(contd.)*
Amino acid sequence of tryptic peptides from 2-pyridylethylated glucoamylase G1.

Pool number (Figure 1)	Fragment	Sequence of peptide	Elution conditions (%, vol/vol)
	T7-3	Glu - Val - Val, Asp, Ser, Phe, Arg	34 CH_3CN
	T7-4	Ala - Leu - Tyr - Ser - Asp - Ala - Ala - Thr - Gly - Thr - Tyr - Ser - Ser - Ser - Ser - Ser (Asx, Thr, Ser, Ser, Ala, Val, Val, Ile, Tyr) Lys	44 CH_3CN
	T7-5	Ala - Thr - Leu - Asp - Ser - Trp - Leu - Ser - Asn - Glu - Ala - Thr - Val - Ala - Arg	46 CH_3CN
	T7-6	Tyr - Thr - Ser - Ser - Asp - Pro - Leu - Trp - Tyr - Val - Thr - Val - Thr - Leu - Pro Ala, Gly, Glu, Ser, Phe, Glu, Tyr - Lys	51 CH_3CN
	T7-7	Glu - Tyr - Thr - Val - Pro - Gln - Ala - Cys - Gly - Thr - Ser - Thr - Ala - Thr - Val - Thr - Asp - Thr - Trp - Arg	54 CH_3CN 15 1-prOH
	T7-8	Thr, Leu, Val, Asp, Leu, Phe, Arg	55 CH_3CN
	T7-9	Ala - Thr - Ala - Met - Ile - Gly - Phe - Gly - Gln - Trp - Leu - Leu - Asp - Asn - Gly - Tyr - Thr - Ser - Thr - Ala - Thr - Asp - Ile - Val - Trp - Pro - Leu - Val - Arg	62 CH_3CN
8	T8-1	Asp - Gly (Pro, Ala, Leu) Arg	30 CH_3CN
	T8-2	Asp - Ser - Trp - Leu - Ser - Asn - Glu - Ala - Thr - Val - Ala - Arg	47 CH_3CN
	T8-3	Ala, Leu, Ala, Asp, His, Lys	10 1-prOH
	T8-4	Asn - Val - Asp - Glu - Thr - Ala - Tyr - Thr - Gly - Ser, Trp, Gly, Arg	26 1-prOH
	T8-5	Phe - Asn - Val - Asp - Glu - Thr - Ala - Tyr - Thr - Gly - Ser, Trp, Gly, Arg	32 1-prOH
	T8-6	Asp - Leu - Thr - Trp - Ser - Tyr - Ala - Ala - Leu, Leu, Thr, Ala, Asn, Asn, Arg	58 CH_3CN
9	T9-1	Phe, Ile, Arg	43 CH_3CN

The tryptic fragments were eluted in HPLC-system 1,2, or 3 (see 2.2.3.) at the indicated concentration of organic solvent. In sequences confirmed by the aid of different peptide fragments (to be published) the amino acid residues are separated by commas. For incompletely sequenced fragments, the amino acid composition of the unidentified parts are in brackets. T2-1 is described in detail together with Tca2-1 in section 3.2. The asterisks indicate glycosylated residues. The majority of the residues in the sequence Thr(71)-Thr(82) are presumably glycosylated, but the fragment Thr(71)-Phe(90) could only be obtained in deglycosylated form. T6-4 starts at Thr(16) in the N-terminal part of G1 (31); position (32) appeared vacant and was suggested to be a threonine residue from the amino acid composition. Positions (27) in T5-2 and (62) in T2-1 were in the same way determined to be a serine residue. T7-5 and T8-2 are N-terminal peptides of the two forms of G1 (31). T7-7 is the C-terminal peptide of G1. T5-1 contained 2 glucosamine and 8 residues of neutral sugar assumed to be mannose. T5-1 and T7-9 are overlapping fragments to the cyanogen bromide peptides (31).

(0.035 mg · ml^{-1}) in 0.05 M-N-ethylmorpholine acetate buffer pH 6.0. Norleucine was included as internal standard and the peptide concentration was determined from an aliquot by amino acid analysis after hydrolysis. The release of free amino acids was followed by removing aliquots at appropriate time intervals, adjusting to pH 0.9 with HCl, lyophilization and redissolution in 0.07 M-citrate pH 2.2 prior to amino acid analysis. Other tryptic fragments were digested with carboxypeptidase B in 0.05 M-N-ethylmorpholine acetate pH 8.5 prior to carboxypeptidase Y at pH 5.0, or by using carboxypeptidase Y first at pH 7.0 and then at pH 4.5.

3. RESULTS

3.1. Separation and sequence analysis of tryptic peptides

The elution pattern of the tryptic fragments from Bio-Gel P-100 suggested that a few long and several shorter peptides had been generated (Figure 1). The fragments of pools 2 and 3 contained an average of 35 and 31 wt.pct. of carbohydrate, respectively, whereas the remaining pools contained from zero to 6 wt.pct. of carbohydrate.

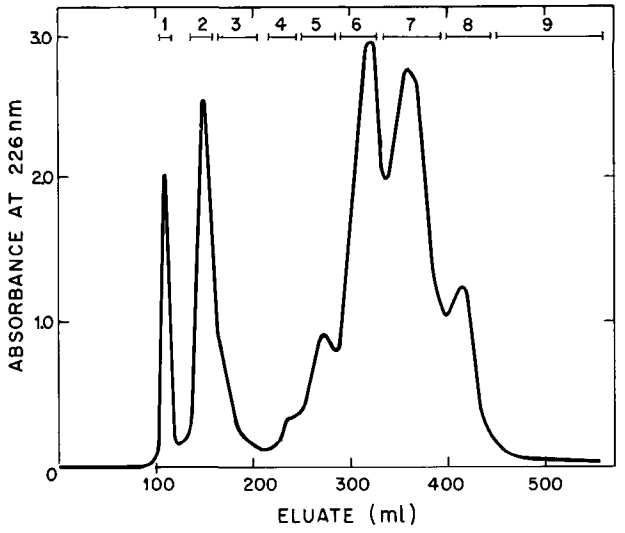

Figure 1. Separation of tryptic fragments of 2-pyridylethylated glucoamylase G1 (35 mg) on Bio-Gel P-100 (2.5×90 cm) in 0.2 M-ammonium bicarbonate at a flow rate of 6 ml · h^{-1}. The absorbance of the eluate was monitored at 226 nm. Pooled fractions are indicated by bars.

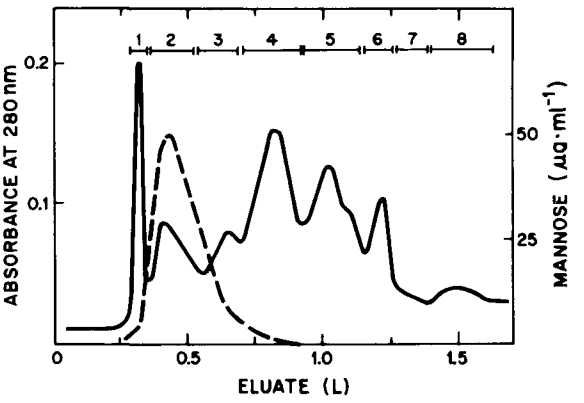

Figure 2. Separation of tryptic peptides (60 mg) from citraconylated and 2-pyridylethylated glucoamylase G1 on Bio-Gel P-100 (5×90 cm) in 0.2 M-ammonium bicarbonate at a flow rate of 30 ml · h^{-1}. The absorbance of the eluate was monitored at 280 nm. The concentration of carbohydrate in selected fractions was expressed as μg · ml^{-1} of mannose (- - -). Pooled fractions are indicated by bars.

Homogeneous peptides were isolated by HPLC from pools 2 and 5-9 in yields sufficient for sequencing and the results are shown in Table I. The largest fragment (T2-1) was 90 amino acid residues long and highly glycosylated; it is further described in section 3.2. The largest of the other fragments (T5-2) contained 40 amino acid residues (Table I). Peptide T7-5 was the N-terminal sequence of G1 and T8-2 was the N-terminal sequence of the form of G1 previously reported to lack the three N-terminal residues (31). The C-terminus of G1 has been reinvestigated and found to be Arg(20) of fragment T7-7. Increasing effectiveness for elution of peptides from the Wide Pore C_{18} column was seen with 1-propanol compared to 2-propanol, and both of these solvents were more effective than elution with acetonitrile from the Radial Pak C_{18} column (Table I).

Citraconylation of 2-pyridylethylated G1 restricted the tryptic cleavage to take place at arginine residues. The elution profile of the fragments resembled that of fragments prepared from non-citraconylated G1 both in UV-absorbance and carbohydrate content (Figure 2). In Table II are listed sequences of fragments obtained by further purifications from selected

Table II.
Amino acid sequence of tryptic fragments of citraconylated and 2-pyridylethylated glucoamylase G1.

Pool number (Figure 2)	Fragment	Sequence of peptide
1	Tca1-1	Ala - Leu - Val - Glu - Gly - Ser - Ala - Phe - Ala - Thr - Ala - Val - Gly - Ser - Ser - (5, 10, 15) Cys - Ser - Trp - Cys - Asp - Ser - Gln - Ala - Pro - Glu - Ile - Leu - Cys - Tyr - Leu - (20, 25, 30) Gln - Ser - Phe - Trp - Thr, Glu, Ser, Phe, Ile, Leu, Ala, Asn, Phe, Asp, Ser, (35, 40, 45) Arg
2	Tca2-1	Sequence identical with T2-1 (Table I)
3	Tca3-1	CHO, ca Ala - Ser - Asn - Gly - Ser - Met - Ser - Glu - Gln - Tyr - Asp - Lys - Ser - Asp - Gly - (5, 10, 15) Glu - Gln, Leu, Ser, Ala, Arg (20)
	Tca3-2	ca Ser - Gly - Lys - Asp - Ala - Asn - Thr - Leu - Leu - Gly - Ser - Ile - His - Thr - Phe - (5, 10, 15) Asp - Pro (Asx, Asx, Thr, Ser, Glx, Glx, Pro, Ala, Ala, Phe, Cys) Cys - Ser - (20, 25, 30) Pro - Arg

Tca1-1 was rechromatographed on Bio-Gel P-100. Tca2-1 was purified by affinity chromatography on Con A-Sepharose, followed by gel filtration on Bio-Gel P-100. Tca3-1 was obtained by gel filtration on Bio-Gel P-60 or eluted at 34% acetonitrile using HPLC system 1. Tca3-2 was eluted in HPLC system 2 at 38% 2-propanol. ca signifies citraconyl.

pools of Figure 2. These fragments include sequences known from studies of the non-citraconylated G1. The fragment Tca1-1 was of low solubility and eluted near the front from Bio-Gel P-100, presumably due to aggregation. Although no lysines were found in Tca1-1, this fragment was obtained in highest yield from tryptic digests of citraconylated G1. Similarly, Tca2-1 was obtained in a yield of 32%, utilizing adsorption to Con A-Sepharose for the purification, whereas the non-citraconylated T2-1 by the same procedure was isolated in 22% yield. Direct purification of T2-1 from pool 2 (Figure 1) by HPLC resulted in a lower yield.

The fragment Tca3-1 (Table II) resulted from an unexpected cleavage between Ala(14) and Ala(15), found in the sequence of T5-1 (Table I). Dansyl-Edman analysis of Tca3-1 indicated position three from its N-terminus to be either aspartic acid or asparagine, while this position was vacant in the automated sequencing. Furthermore, the amino acid composition of fragments T5-1 and Tca3-1, agreed with asparagine found at the position in question. Tca3-1 contained 8 neutral sugar residues assumed to be mannose and 2 glucosamine residues, which were released upon acid hydrolysis.

3.2. Chemical characterization and subfragmentation of the glycopeptides T2-1 and Tca2-1

It was difficult to establish the purities of the highly glycosylated fragments T2-1 and Tca2-1 (Tables I and II) since these fragments did not fix and stain after polyacrylamide gel electrophoresis. However, the shape of the elution profile from Bio-Gel P-100, the amino acid composition, the results from 6 cycles of N-terminal sequencing, and the absence of residual staining in polyacrylamide gel electrophoresis indicated the fragments to be pure.

Tca2-1 was sequenced through 54 cycles and

Table III.
Carbohydrate content and amino acid composition of Tca2-1 and subfragments

Amino acid	Residues/fragment			
	Tca2-1	Th1-Tca2-1	T2-dTca2-1	T3-dTca2-1
Aspartic acid	1.2 (1)			
Threonine[a]	(26)	(19)	7.5 (8)	2.9 (3)
Serine[a]	(20)	(10)	4.8 (5)	1.0 (1)
Glutamic acid	0.90(1)			
Proline	4.8 (5)	1.7 (2)	1.0 (1)	
Glycine	6.6 (7)	4.2 (4)		
Alanine	10.9 (11)	4.7 (5)	2.1 (2)	2.1 (2)
Valine	8.5[b](9)	1.7 (2)	1.9 (2)	
Isoleucine	1.9[b](2)	0.8 (1)		
Tyrosine	1.1 (1)			
Phenylalanine	1.0 (1)		1.0 (1)	
2-pe-Cysteine[c]	1.4 (2)[g]	1.0 (1)	0.4 (1)[g]	
Lysine	1.9 (2)	1.8 (2)		1.0 (1)
Tryptophan[d]	2 (2)	n.d.		
Total number of residues	90	46	20	7
O-Glycosylated threonine[h]	19	14		
O-Glycosylated serine[h]	16	8		
Carbohydrate (wt.pct.)[e]	50	67		
N-terminal yield (%)[f]	50	65		

Th1-Tca2-1 was purified on Bio-Gel P-60. T2-dTca2-1 and T3-dTca2-1 eluted at 20% and 11% of 1-propanol in HPLC system 3, respectively.

a) Values extrapolated to zero time hydrolysis or the value after 24 hours multiplied by 1.06 and 1.15 in the case of threonine and serine, respectively.
b) Value from 72 hours of hydrolysis.
c) 2-pyridylethyl-cysteine (9).
d) Estimated from the UV-absorbance of Tca2-1 (5). n.d. means not detected.
e) Determined by the phenol-sulfuric acid procedure (4) using mannose as the standard.
f) The yield of PTH-amino acid in the first step of the automated sequencing based on a peptide content determined from amino acid analysis and the indicated length of fragment.
g) Sequence studies of Tca2-1, T2-dTca2-1 and amino acid composition of Th1-Tca2-1 suggested two 2-pe-cysteine residues in Tca2-1.
h) Determined by amino acid analysis after reductive β-elimination (3).

many positions appeared vacant because the O-glycosylated anilinothiazolinone-amino acids were not extracted from the reaction film of the spinning cup. After deglycosylation sequence analysis showed as expected serine and threonine at the previously vacant positions. O-Glycosylated amino acid residues identified in this manner are indicated by asterisks in T2-1 (Table I).

Thermolysin, in contrast to trypsin, efficiently catalyzed the hydrolysis of peptide bonds adjacent to glycosylated amino acid side chains. One major (Th1-Tca2-1) and several smaller glycopeptides resulted from digestion of Tca2-1 with thermolysin. The contents of amino acids, carbohydrate and glycosylated amino acids in both Th1-Tca2-1 and Tca2-1 are shown in Table III. The results from sequencing Tca2-1 and Th1-Tca2-1, starting at Ile(40) (see T2-1, Table I), and the presence in Tca2-1 of a single residue of asparagine, glutamic acid, tyrosine, and phenylalanine, respectively, suggested this fragment

to contain 90 amino acid residues. The large thermolytic subfragment was sequenced through 31 cycles and was estimated to contain 46 amino acid residues. In both fragments the majority of the serine and threonine residues were O-glycosylated. Tca2-1 became only upon deglycosylation and decitraconylation susceptible to tryptic digestion with cleavage after both Lys(63) and Lys(70). Digestion with carboxypeptidases B and Y of the fragment Thr(64)-Lys(70) (T3-dTca2-1, Table III) suggested the positions (67) and (69) (unidentified by automated sequencing of the deglycosylated version of Th1-Tca2-1) to be glycosylated threonine and glycosylated serine , respectively. From the total number of glycosylated positions and the number of identified glycosylated positions in Tca2-1 (Tables I and III) the sequence Thr(71)-Phe(90) (T2-dTca2-1, Table III) was seen to be extensively glycosylated. Position (62) was suggested to be glycosylated serine from the amino acid compositions and sequence information of Tca2-1 and its subfragments.

Digestion of Tca2-1 with carboxypeptidase Y released phenylalanine, threonine, and valine in that order (Figure 3) and only trace amounts of alanine appeared after valine. N-Terminal sequencing through 12 cycles and digestion with carboxypeptidase Y of the deglycosylated tryptic subfragment Thr(71)-Phe(90) identified the C-terminal sequence of Tca2-1 as reported in Table I. The amino acid composition of Th1-Tca2-1 (Table III) indicated its C-terminus to be at residue (85) (see T2-1, Table I), which, however, could not be released with carboxypeptidase Y.

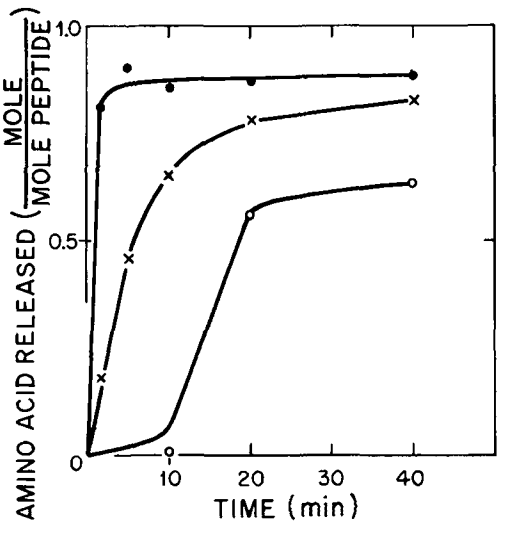

Figure 3. Release of free amino acids from Tca2-1 by digestion with carboxypeptidase Y. Phenylalanine (o), threonine (×), and valine (●).

4. DISCUSSION

The reported tryptic fragments represent 491 different amino acid residues corresponding to more than 80% of the G1 molecule. This is based on a minimum size of G1 of about 614 residues as calculated from the amino acid composition (31) assuming a content of 18 arginines. In the literature the molecular weights of G1 from A. niger ranged from 52,000 to 110,000 (8, 19, 26, 31). Consequently an accurate size determination of the enzyme and a discussion of the conflicting findings await the completion of the amino acid sequence. The gene of glucoamylase from A. awamori has recently been cloned (24) and subjected to sequencing (13), thus the complete amino acid sequence might be determined with the aid of results from nucleotide sequencing.

Comparison of the portions of the G1 sequence now available with the primary structures of α-amylases from Bacillus amyloliquefaciens (1, 29, 32), Aspergillus oryzae (Taka-amylase A) (33), mouse pancreas and salivary gland (12), and porcine pancreas (18, 25) and the partial amino acid sequence of Bacillus subtilis var. amylosacchariticus α-amylase (23) did not indicate any homology. Among the glucoamylases no extensive sequence information has been published at present. However, as in G1 from A. niger, the N-terminal and the C-terminal residues of A. awamori glucoamylase G1 were found to be alanine and arginine, respectively (35). In contrast, the four N-terminal residues reported for glucoamylase from A. saitoi were different from those found in G1 and G2 from A. niger (14, 31).

The carbohydrate composition of the N-glycoside unit in Tca3-1 appeared to be the same as reported for the single carbohydrate group of Taka-amylase A (34), no sequence homology was found between this enzyme and

glucoamylase. Glucoamylase G1 from A. niger, in addition, contained a highly glycosylated region consisting of 70 amino acid residues which possessed very few potential cleavage points leading to overlapping fragments. However, treatment with anhydrous HF removed essentially all the carbohydrate from the O-glycosylated amino acid side chains without damaging the polypeptide moiety. Most of the glycosylated positions could then be identified by automated sequencing of the rather long fragments. The carbohydrate groups in this region are calculated to contain an average of two sugar residues per unit, which is similar to G1 itself (27, 31). Two other fungal proteins, cellobiohydrolase from Trichoderma viride (11) and mycodextranase from Penicillium melinii (28), were also reported to carry a high number of short carbohydrate chains. Moreover in cellobiohydrolase, it has likewise been suggested that a region of about 20 amino acid residues is very rich in carbohydrate (7). As both glucoamylase G1 and this enzyme have the capacity to adsorb to and digest an insoluble substrate viz. starch and cellulose, it is conceivable that the highly glycosylated regions are involved in the formation of the corresponding enzyme-substrate complexes. However, no homology has been detected between the primary structures of glucoamylase and the cellobiohydrolase (7).

ACKNOWLEDGEMENTS

The authors are indebted to Professor MARTIN OTTESEN for his continuous encouragement and advice during the work and for his critical revision of the manuscript. The help of Drs. KLAUS BOCK and CHRISTIAN PEDERSEN (Department of Organic Chemistry, The Technical University of Denmark, Lyngby) with deglycosylation and of Dr. VIBEKE BARKHOLT (Institute of Biochemical Genetics, University of Copenhagen) with manual peptide sequencing is gratefully acknowledged. Dr. ESPER BOEL (Laboratory of Genetics, NOVO Research Institute, Bagsværd, Denmark) is thanked for sequence information obtained from cDNA corresponding to the mRNA of glucoamylase from A. niger. Ms. EDITH FLØJSTRUP is thanked for expert technical assistance at all stages of the work. Mss. LONE SØRENSEN, BODIL CORNELIUSSEN, SIDSEL EHLERS, and HANNE CHRISTIANSEN are acknowledged for excellent technical assistance with amino acid analysis, sequencing, and peptide separation. Dr. ANTHONY CLARKE is thanked for linguistic correction of the manuscript.

REFERENCES

1. CHUNG, H. & F. FRIEDBERG: Sequence of the N-terminal half of Bacillus amyloliquefaciens α-amylase. Biochem J. 185, 387-395 (1980)
2. DIXON, H.B.F. & R.N. PERHAM: Reversible blocking of amino groups with citraconic anhydride. Biochem. J. 109, 312-314 (1968)
3. DOWNS, F. & W. PIGMAN: Determination of O-glycosidic linkages to L-serine and L-threonine residues of glycoproteins. Meth. Carbohyd. Chem. 7, 200-204 (1972)
4. DUBOIS, M., K.A. GILLES, J.K. HAMILTON, P.A. REBERS & F. SMITH: Colorimetric method for determination of sugars and related substances. Anal. Chem. 28, 350-356 (1956)
5. EDELHOCH, H.: Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 6, 1948-1954 (1967)
6. EDMAN, P. & A. BEGG: A protein sequenator. Eur. J. Biochem. 1, 80-91 (1967)
7. FÄGERSTAM, L.G.: Cellulases from Trichoderma reesei QM 9414. Enzymatic and structural properties. Dissertation, Acta Universitatis Upsaliensis no. 610, Uppsala 1981
8. FREEDBERG, I.M., Y. LEVIN, C.M. KAY, W.D. MCCUBBIN & E. KATCHALSKI-KATZIR: Purification and charaterization of Aspergillus niger exo-1,4-glucosidase. Biochim. Biophys. Acta 391, 361-381 (1975)
9. FRIEDMAN, M., J.C. ZAHNLEY & J.R. WAGNER: Estimation of the disulfide content of trypsin inhibitors as S-β-(2-pyridylethyl)-L-cysteine. Anal. Biochem. 106, 27-34 (1980)
10. GRAY, W.R.: End-group analysis using dansyl chloride. Meth. Enzymol. 25, 121-138 (1972)
11. GUM, E.K. & R.D. BROWN: Structural characterization of a glycoprotein cellulase, 1,4-β-D-glucan cellobiohydrolase C from Trichoderma viride. Biochim. Biophys. Acta 446, 371-386 (1976)
12. HAGENBÜCHLE, O., R. BOVEY & R.A. YOUNG: Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell 21, 179-187 (1980)
13. INNIS, M.: Personal communication
14. INOKUCHI, N., T. TAKAHASHI & M. IRIE: Purification and characterization of a minor

glucoamylase from Aspergillus saitoi. J. Biochem. 90, 1055- 1067 (1981)
15. JOHANSEN, J.T., C. OVERBALLE-PETERSEN, B. MARTIN, V. HASEMANN & I. SVENDSEN: The complete amino acid sequence of copper, zinc superoxide dismutase from Saccharomyces cerevisiae. Carlsberg Res. Commun. 44, 201-217 (1979)
16. KLAPPER, D.G., C.E. WILDE III & J.D. CAPRA: Automated amino acid sequence of small peptides utilizing Polybrene. Anal. Biochem. 85, 126-131 (1978)
17. KLEMM, P.: The complete amino-acid sequence of the K88 Antigen, a fimbrial protein from Escherichia coli. Eur. J. Biochem. 117, 617-627 (1981)
18. KLUH, I.: Amino acid sequence of hog pancreatic α-amylase isoenzyme I. FEBS Letters 136, 231-234 (1981)
19. LINEBACK, D.R. & L.A. AIRA: Structural characterization of two forms of glucoamylase from Aspergillus niger. Cereal. Chem. 49, 283-296 (1972)
20. MAHONEY, W.C. & M.A. HERMODSON: Separation of large denatured peptides by reverse phase high performance liquid chromatography. Trifluoroacetic acid as a peptide solvent. J. Biol. Chem. 255, 11199-11203 (1980)
21. MENDEZ, E. & LAI, C.Y.: Regeneration of amino acids from thiazolinones formed in the Edman degradation. Anal. Biochem. 68, 47-53 (1975)
22. MORT, A.J. & D.T.A. LAMPORT: Anhydrous hydrogen fluoride deglycosylates glycoproteins. Anal. Biochem. 82, 289-309 (1977)
23. NAGATA, Y., S. SUGA, O. KADO & B. MARUO: N-terminal amino acid sequence of α-amylase from Bacillus subtilis var. amylosacchariticus: Comparison with that of "liquefying" type α-amylase. Agri. Biol. Chem. 44, 215-216 (1980)
24. NUNBERG, J., V. SCHWEICKART, F. LAWYER, M. INNIS, J. FLATGAARD, V. WITTMAN, & J. MENDE: Cloning of the Aspergillus awamori glucoamylase gene. Abstract from "Gene manipulation in the exploitation and study of fungi". Amer. Soc. Microbiol., May 1983, South Bent, Indiana.
25. PASERO, L., B. ABADIE, Y. CHICHEPORTICHE, Y. MAZZEI, D. MOINIER, J.P. BIZZOZERO, M. FOUGEREAU & G. MARCHIS-MOUREN: Porcine pancreatic α-amylase I: Sequence between the 35th and 410th amino acid. Biochimie 63, 71-76 (1981)
26. PAZUR, J.H., H.R. KNULL & A. CEPURE: Glucoenzymes: Structure and properties of the two forms of glucoamylase from Aspergillus niger. Carbohyd. Res. 20, 83-96 (1971)
27. PAZUR, J.H., Y. TOMINAGA, L.S. FORSBERG & D.L. SIMPSON: Glycoenzymes: an unusual type of glycoprotein structure for a glucoamylase. Carbohyd. Res. 84, 103-114 (1980)
28. ROSENTHAL, A.L. & J.H. NORDIN: Enzymes that hydrolyze fungal cell wall polysaccharides. The carbohydrate constitution of mycodextranase, an endo-α(1→4)-D-glucanase from Penicillium melinii. J. Biol. Chem. 256, 5295-5303 (1975)
29. SACHDEV, O. & F. FRIEDBERG: Sequence of cyanogen bromide fragments D and E of B. amyloliquefaciens alpha amylase. Int. J. Peptide Protein Res. 18, 228-236 (1981)
30. SVENDSEN, I., B. MARTIN & I. JONASSEN: Characteristics of Hiproly barley III. Amino acid sequences of two lysine-rich proteins. Carlsberg Res. Commun. 45, 79-85 (1980)
31. SVENSSON, B., T.G. PEDERSEN, I. SVENDSEN, T. SAKAI & M. OTTESEN: Characterization of two forms of glucoamylase from Aspergillus niger. Carlsberg Res. Commun. 47, 55-69 (1982)
32. TÄKKINEN, K., R.F. PETTERSSON, N. KALKKINEN, I. PALVA, H. SÖDERLUND & L. KÄÄRIÄNEN: Amino acid sequence of α-Amylase from Bacillus amyloliquefaciens deduced from the nucleotide sequence of the cloned gene. J. Biol. Chem. 258, 1007-1013 (1983)
33. TODA, H., K. KONDO & K. NARITA: The complete amino acid sequence of Taka-amylase A. Proc. Japan Acad. 58, Ser. B, 208-212 (1982)
34. YAMAGUCHI, H., T. IKENAKA & X. MATSUSHIMA: The complete sequence of a glycopeptide obtained from Taka-amylase A. J. Biochem. 70, 587-594 (1971)
35. YOSHINO, E. & S. HAYASHIDA: Enzymatic modification of glucoamylase of Aspergillus awamori var. kawachi. J.Ferment. Technol. 56, 289-295 (1978)
36. UEDA, S.: Fungal glucoamylases and raw starch digestion. Trends Biochem. Sci. 6, 89-90 (1981)