# β-Glucosylarginine: a new glucose–protein bond in a self-glucosylating protein from sweet corn

David G. Singh[a], Joseph Lomako[a], Wieslawa M. Lomako[a], William J. Whelan[a,*], Helmut E. Meyer[b], Maria Serwe[b], Jörg W. Metzger[c]

[a]Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, (M823), PO Box 016129, Miami, FL 33101, USA
[b]Institut für Physiologische Chemie, Ruhr-Universität Bochum, Universitätsstrasse 150, D-44780 Bochum, Germany
[c]Institut für Organische Chemie, Eberhard-Karls-Universität Tübingen, Auf der Morgenstelle 18, D-72076 Tübingen, Germany

Received 5 October 1995; revised version received 18 October 1995

**Abstract In the search for a protein primer for starch synthesis, an autocatalytic self-glucosylating protein has been isolated from sweet corn. Several tryptic peptides were obtained from the [$^{14}$C]glucosylated protein and were sequenced, corresponding to over 40% of the estimated total sequence (molecular mass 42 kDa). There is no homology with the amino acid sequence of the autocatalytic glycogen primer, glycogenin, nor in respect of the nature of the union between the autocatalytically added glucose and the protein, which, in the case of the corn protein, now named amylogenin, is a novel glucose–protein bond, a single β-glucose residue joined to an arginine residue.**

*Key words:* Amylogenin; β-Glucosylarginine; Self-glucosylating protein; Autoglucosylation

## 1. Introduction

Muscle glycogen synthesis is primed by a self-glucosylating protein – glycogenin – of molecular mass 38 kDa. An average of eight α-glucose residues are added autocatalytically from UDP-glucose, the first to Tyr-194 and the subsequent ones in 1,4-linkage to the preceding glucose residue. At completion of autocatalysis, the maltosaccharide chains, which then prime the synthesis of glycogen by glycogen synthase and branching enzyme, are not of uniform length, but contain from 7 to 11 glucose residues. The now extensive literature on glycogenin has recently been reviewed [1].

We have been interested to learn whether the initiation of the synthesis of storage polysaccharides by an autocatalytic protein is a general phenomenon. Accordingly, we tested crude extracts of mammalian and plant tissues for protein glucosylation. After isolating glycogenin from muscle glycogen and obtaining an antibody, we were able to detect in muscle extracts what seemed to be a glycogen-free version of the same protein, both by Western blotting and by incubating the extract with 5 μM UDP-[$^{14}$C]glucose (this concentration is near the $K_m$) in presence of 5 mM $Mn^{2+}$ ($Mn^{2+}$ (or $Mg^{2+}$) is an obligatory activator). SDS-PAGE and radioautography of the digest revealed a $^{14}$C-labelled band with the $M_r$ of glycogenin. When similar experiments were performed on extracts from a wide variety of life forms, we frequently found ourselves able to radiolabel proteins of similar $M_r$ and, not infrequently, also observed a positive immunoblot [2]. We decided to study such a protein in sweet corn which underwent [$^{14}$C]glucosylation. The present report is concerned with identifying the nature of the glucose–protein bond in this protein, which, like the earlier discovery of the α-glucose–tyrosine bond in glycogenin [1], has proven to be yet another new type of union between glucose and protein. Although, like glycogenin, the corn protein is self-glucosylating, using UDP-glucose in a reaction activated by $Mn^{2+}$, and has a similar $M_r$, there is no sequence homology. Accordingly we have named the new corn protein, amylogenin.

## 2. Materials and methods

### 2.1. Materials

Commercial sweet corn was purchased from local supermarkets. α-Amylase, β-amylase, glucoamylase, α- and β-glucosidases were purchased from Sigma. Trypsin was from Pierce and Promega. Chromatography media came from Pharmacia, except Biogel P4 (BioRad). UDP-[$^{14}$C]glucose was from ICN Biochemicals. HPLC-grade water and acetonitrile were from Aldrich and trifluoroacetic acid from Pierce. All reagents for SDS-PAGE and protein determination were from BioRad. All other routine chemical reagents were from Sigma or Mallinckrodt. Other suppliers of apparatus and reagents are noted at the point of mention.

### 2.2. Labelling and quantitation of amylogenin

The [$^{14}$C]glucosylation of amylogenin was carried out by incubating a sample of the enzyme for 30 min at room temperature in a 100 μl reaction mixture containing 2 μM UDP-[$^{14}$C]glucose, 50 mM Tris-HCl buffer, pH 7.4 and 5 mM $MnCl_2$. (Larger scale digests for preparative labelling of amylogenin were similarly constituted.) The reaction was stopped by adding 1 ml of 10% trichloroacetic acid. This assay measures the total amount of glucose that can be incorporated, not the rate of glucosylation. The specific activity of the UDP-glucose was arranged so that about 120,000 counts/min were in the digest. The amount of enzyme was arranged so that no more than half the $^{14}$C was incorporated into protein. The $^{14}$C-labelled protein was passed through a nitrocellulose filter, type HA 0.45μm (Millipore), washed with 3 ml of 10% trichloroacetic acid and 10 ml of water and dried under an infrared lamp. The incorporated $^{14}$C was then measured in a scintillation counter. The precipitated protein, similarly prepared, was also subjected to SDS-PAGE [3] in 25 mM Tris/192 mM glycine/0.001% SDS, pH 8.3 at 25 mA for 1 h. Protein bands were stained and the gels then soaked in Amplify (Amersham) for 15 min [4], dried onto Whatman 3MM paper in a slab dryer (BioRad) at 60°C and then exposed for fluorography to Kodak X-OMAT XAR-5 film at −70°C.

### 2.3. Isolation of [$^{14}$C]glucosylated amylogenin

Full details of the purification of amylogenin to homogeneity from sweet-corn cobs [5] will be reported elsewhere. The following procedure followed the same method, from extraction of the cobs with a buffer containing peptidase inhibitors [6], acid precipitation at pH 5.0 and fractionation on a DEAE-Sepharose column, where the emerging fractions were tested for protein glucosylating activity and the most active combined. The product from seven cobs was concentrated to 25 ml in a Centriprep-10 concentrator (Amicon), was further divided 1:3 and

*Corresponding author. Fax: (1) (305) 324-5665.
E-mail: wwhelan@mednet.med.miami.edu


SDI 0014-5793(95)01247-8

the smaller portion allowed to undergo $^{14}C$-labelling under standard conditions; the larger portion was similarly glucosylated with unlabelled UDP-glucose. The two portions were combined and passed through Sepharose CL-6B (3 × 85 cm) in 50 mM Tris-HCl, pH 7.4 containing 0.02% Triton. Elution was performed with the same buffer and fractions (5.5 ml each) were collected and tested for $^{14}C$ content. The $^{14}C$ emerged in a 66% yield between fractions 43–60, well after a major 280 nm-absorbing peak (26–37). Aliquots of fractions 43–60 were examined by SDS-PAGE and silver staining. Fractions 43–51 (2.3 mg protein) were superior in degree of purity and were combined separately from 52–60. They were again concentrated by centrifugation using a Centriprep-10 concentrator and subjected to reverse-phase HPLC with water-acetonitrile in a Perkin Elmer system using a Vydac $C_4$ column (8 × 40 mm). Several portions (450 μl each, containing 0.26 mg protein and representing 75% of the protein in fractions 43–51) were successively applied and fractions combined on the basis of $^{14}C$ content. The $^{14}C$ emerged between 42% and 46% acetonitrile and represented 58% of the $^{14}C$ applied. A second fractionation was necessary to obtain a product which emerged as a single peak (149 μg) absorbing at 215 nm. SDS-PAGE and Coomassie blue staining showed it to be a single band of mass 42 kDa. The sample was reduced and carboxymethylated to lower its resistance to trypsinolysis. This was accomplished by mixing the sample (149 μg, 3.3 ml) with 5 M guanidinium chloride, 0.5 M Tris-HCl, pH 8.6 (400 μl) and concentrating to 0.4 ml on a Speed-Vac. 1 M dithiothreitol (5 μl) was added and the tube flushed with argon. Reaction proceeded at 50°C for 2 h when 1 M iodoacetic acid (15 μl) was added and argon again used to flush the tube which was stored for 30 min at room temperature in the dark. The two procedures were repeated in the same way using 7.5 μl of dithiothreitol and 15 μl of iodoacetic acid. Finally, 17.5 μl of dithiothreitol was added and the solution dialyzed against 50 mM Tris-HCl, pH 7.6, 1 mM $CaCl_2$ in a Pierce Microdialysis System 500. The buffer (2 litres) was continuously delivered for 2.5h. The volume was then 550 μl and the pH 7.8.

*2.4. Preparation and analysis of tryptic peptides*

Trypsin modified by reductive alkylation (Promega) was chosen for its resistance to autolytic digestion [7]. The enzyme (15 μg in 15 μl of 50 mM acetic acid) was added to 530 μl of the purified, labelled protein for a trypsin : protein ratio of 1:25. Digestion was carried out for 15 h at 37°C.

The tryptic peptides were separated by reverse-phase HPLC on a Perkin Elmer System using a Vydac $C_{18}$ column (100 × 4 mm). The entire tryptic digest was applied to the column and eluted with a linear gradient between 0.1% aqueous trifluoroacetic acid and 0.09% trifluoroacetic acid in 84% acetonitrile during 100 min at a flow rate of 0.4 ml/min. The absorbance of the effluent was monitored simultaneously at 215 nm and 295 nm and fractions collected manually based on absorbance. Individual peak fractions were further purified on a smaller $C_{18}$ column. In order to remove acetonitrile, each peak was concentrated to half its volume on a Speed-Vac and restored to the original volume with water. The sample was applied to the column (50 × 2 mm) and eluted with an acetonitrile gradient generated from 10 mM ammonium acetate, pH 6.0, and the same buffer containing 84% acetonitrile. The gradient was applied over a period of 60 min at 0.15 ml/min. Peptides were detected as before and peaks collected manually for sequencing. A total of eight homogeneous peptides was obtained.

The amino acid compositions of the peptides and the original, labelled protein were determined by the PTC derivatization procedure [8]. Sequencing was performed by the Edman procedure on an Applied Biosystems 470A gas-phase sequencer equipped with an on-line model 120 phenylthiohydantoin analyzer using the 03R phenylthiohydantoin standard program. No sequence data was obtained for the intact protein, suggesting that the N-terminus was blocked. The sequences of the eight peptides are shown in Table 1 and further data on peptide T2 in Table 2.

*2.5. Mass spectrometry of $^{14}C$-labelled tryptic peptide T2*

Peptide T2 (Tables 1 and 2) was subjected to ion-spray mass spectrometry [9–11]. The instrument was a Sciex (Toronto) API 111 triple-quadrupole mass spectrometer with 2400 Da mass range equipped with an ion-spray source. It was operated under unit-mass resolution conditions. Ion-spray voltage was 5 kV. The spectrum (Fig. 1) revealed only one prominent signal with a mass/charge ratio (m/z) of 888.9, corresponding to m/z 1776.8 for a single protonated ion. The calculated m/z for peptide T2, assuming that position 7 is occupied by a monoglucosylated arginine, is 1776.97. No other of the 19 amino acids found in proteins gives a corresponding calculated m/z ratio when substituted for arginine. The nearest is tyrosine (1783.96). The mass/charge ratios of peptides T1, T3, T4, T7 and T8 were also measured in the same way. The results agreed with the values calculated from the amino acid sequences (Table 1).

Tandem ion-spray mass spectrometry of peptide T2 was also carried out on a 150 pmol sample. A daughter ion spectrum obtained by collisional activation in which the doubly protonated parent ion was bombarded with argon, showed two signals. (The spectrum has been published in [12], Fig. 5.) The more prominent signal had a mass/charge ratio of 889.0 and the minor a ratio of 807.5. The major signal corresponds to the doubly protonated parent ion, monoglucosylated T2. The ratio of the minor signal corresponds to a doubly charged daughter ion that has lost a 163 Da glucosyl fragment. The presence of only one daughter ion with the loss of a 163 Da fragment is confirmation of the presence of a single glucose residue and that the in vitro glucosylation of amylogenin involves the direct glucosylation of arginine.

*2.6. Anomeric configuration of the glucose-amylogenin linkage*

A mixture of tryptic peptides from [$^{14}C$]glucosylated amylogenin (200 μl, containing 25,400 counts/min), prepared essentially as above, but using TPCK-treated enzyme (Pierce), was passed through a Biogel P4 column (80 × 0.8 cm) in 50 mM ammonium bicarbonate. Fractions (500 μl) were collected and tested for radioactivity using 50 μl.

The radioactivity emerged between fractions 36–44 and was collected and freeze-dried. Two portions, equivalent to 5350 counts/min, were digested in one case in 100 mM sodium phosphate (pH 7.0) with 5 U of α-glucosidase, the other in 100 mM sodium acetate (pH 7.0) with 5 U of β-glucosidase, both at room temperature for 72 h. The digests were applied to the same Biogel column as above, and eluted. The $^{14}C$ in the

Table 1
Amino acid sequences and mass/charge ratios of tryptic peptides from [$^{14}C$]glucosylated amylogenin

| | Measured m/z | Calc. m/z |
|---|---|---|
| T1 - Y V D A V M T I P K | 1135 | 1135.4 |
| T2 - E G A D F V X G Y P F S L R* | 1777 | 1770.0 |
| T3 - Y D D M W A G W C V K | 1432 | 1431.6 |
| T4 - E G A H T A V S H G L W L N I P D Y D A P T Q L V K P K | 3056 | 3058.4 |
| T5 - L G D A M V T W I E A W D E L N P S T P (A A A D G K) | | |
| T6 - N L L S P S T P F F F N T L Y D P Y R E G A D F V X G Y P F S L R* | | |
| T7 - G I F W Q E D I I P F F Q N V T I P K | 2292 | 2292.7 |
| T8 - N L D F L E M W R P F F Q P Y H L I I V Q D G D P T K | 3319 | 3320.8 |

*Peptides T2 and T6 were radioactive. During their sequencing, no phenylthiohydantoin amino acid was detected at positions 7 and 26, respectively (shown as X), and most of the applied radioactivity remained bound to the glass fiber disc. Comparison of the amino acid composition of T2 with the sequence revealed that the difference was constituted by there being one fewer arginine residue in the sequence (Table 2). Mass/charge ratios were measured by ion-spray mass spectrometry (section 2.5) with an error range of ±O.7 Da. The measured values for m/z are obtained from m/z for the doubly protonated species and expressed as m/z for the mono-protonated form. m/z ratios were not measured for T5 and T6. The assignment of the last six amino acids in T5, in parentheses, is tentative.

α-glucosidase digest emerged in a single peak between fractions 31–41. In the case of β-glucosidase treatment, no $^{14}C$ was detected before fraction 66. All the $^{14}C$ emerged between fractions 66–70, which represented the total volume of the column and was the position at which free glucose appeared.

## 3. Results and discussion

We have purified the active corn protein more than 2000-fold to homogeneity in 3% yield by conventional methodology, most of the purification taking place in the last step by molecular sieving on Sepharose CL-6B. This procedure [5] will be reported elsewhere, together with a description of the properties of the homogeneous protein which, like glycogenin, proved to be autocatalytic, also using UDP-glucose as donor, but not ADP-glucose. What we are reporting here is a procedure designed to purify the protein after [$^{14}C$]glucosylation so that sequence information could be obtained from tryptic peptides in order to clone the cDNA for the protein and so that the $^{14}C$ label could be used to learn the nature of the glucose to protein bond.

The homogeneous protein was unstable, so for the purposes of determining the site of [$^{14}C$]glucosylation we decided to carry out the glucosylation after partial purification. This overcame the problem posed by the unstable activity and permitted the further purification to include denaturing conditions. The partly purified, $^{14}C$-labelled protein was fractionated on Sepharose CL-6B. Inspection of the individual fractions for $^{14}C$ content and protein purity, using SDS-PAGE, enabled the purest labelled fractions to be pooled. The protein was further fractionated by HPLC and then reduced, carboxymethylated and treated with trypsin. Eight peptides were obtained and sequenced (Table 1). Two were radioactive, with overlapping sequences, and in each case no signal was seen at one cleavage step during sequence analysis, indicating that this was the position of glucosylation. There is no homology between the sequences of these peptides and the amino acid sequence of glycogenin [13].

Comparison of the sequences of the labelled peptide T2 with the amino acid composition suggested that the glucosylated amino acid was arginine (Table 2). This was confirmed by mass spectrometry. Ion-spray mass spectrometry of peptide T2 yielded only one prominent signal with a mass/charge ratio of 883.9 corresponding to a mono-protonated ion having m/z 1776.8 (Fig. 1). The calculated mass for a monoglucosyl T2

Table 2
Amino acid composition of amylogenin tryptic peptide T2 as revealed by compositional analysis and compared with the composition seen in its sequencing (Table 1)

| Amino acid | Amino acid analysis | Sequence analysis |
|---|---|---|
| D | 1.0 | 1 |
| E | 1.0 | 1 |
| S | 1.2 | 1 |
| G | 2.8 | 2 |
| R | 2.0 | 1 |
| A | 1.2 | 1 |
| P | nd | 1 |
| Y | 1.1 | 1 |
| V | 1.0 | 1 |
| L | 1.5 | 1 |

nd = not determined.

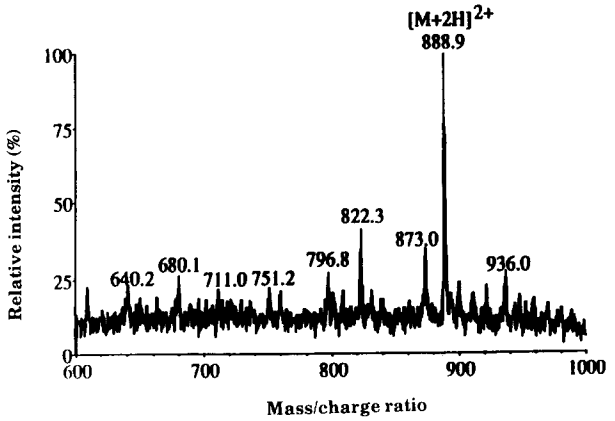


Fig. 1. Ion-spray mass spectrometry of the $^{14}C$-labelled tryptic peptide T2 (Tables 1 and 2) from sweet-corn amylogenin. The total number of ions obtained in the spectrum was 287,000. $[M+2H]^{2+}$ is the doubly protonated peptide T2.

peptide in which position 7 in peptide T2 (Tables 1 and 2) is arginine is 1776.97. Tandem ion-spray mass spectrometry yielded two signals corresponding to parent and daughter ions related by the loss of a glucosyl residue (163 Da). This is further confirmatory evidence that the [$^{14}C$]glucosylation of amylogenin involves the attachment of a single glucose residue to arginine. To our knowledge, glucose has not previously been found in linkage to arginine, although the glycosylation of arginine is well known in the cholera toxin-catalysed ADP-ribosylation of G proteins. The glucose residue is in the β-configuration. This was shown by its complete removal by β-glucosidase from tryptic peptides. (β-Glucosidase action on the parent protein was incomplete.) None of four enzymes that split α-glucosidic bonds could remove the $^{14}C$-label. That the glucose–arginine linkage has the β-configuration is another contrast with glycogenin, which autocatalytically adds α-glucose to Tyr-194 followed by an average of seven α-linked glucose residues [14].

The evidence that peptide T2 contained a glucose residue linked to arginine is therefore: (i) the amylogenin was self-[$^{14}C$]glucosylated in vitro; (ii) peptide T2 was radioactive and therefore was substituted by [$^{14}C$]glucose; (iii) comparison of the composition of T2 by amino acid analysis and sequencing revealed a non-detectable, modified arginine (Table 1); (iv) the $^{14}C$-label was removed by β-glucosidase; and (v) mass spectrometry of T2 (Fig. 1) confirmed it to have the mass calculated for a monoglucosylated derivative (Table 1).

We embarked on a search for a protein primer for starch synthesis in corn. Rothschild and Tandecarz [15] recently reported having detected a protein in developing maize endosperm (mass 38 kDa) that is glucosylated under conditions similar to amylogenin and then primes chain elongation by phosphorylase and starch synthase. We do not know whether amylogenin also fills this role and are now exploring the question. We have now cloned the cDNA for amylogenin (unpublished). It encompasses all the tryptic peptides in Table 1.

A search of GenBank for homologous sequence data showed significant matches of some of the peptides with partial cDNA and transcribed sequences from rice (RIC1486A, RICS5091A, RICC2546A) and *Arabidopsis thaliana* (ATT50381, T23020, T46745, ATTS3942, T22507, T22943, T44934), but none is

associated specifically with a 42 kDa, autoglucosylating protein.

*Acknowledgements:* We thank Dr. Peter L. Keeling for advice and encouragement. The work was supported by grants to W.J.W. from the Florida High Technology and Industry Council and ICI Seeds, from the Deutscheforschungs Gemeinschaft (Me 765/3-2 to H.E.M. and SFB 323,C3 to J.W.M.) and by the Fond der Chemischen Industrie (to H.E.M.).

## References

[1] Alonso, M.D., Lomako, J., Lomako, W.M. and Whelan, W.J. (1995) FASEB J. 9, 1126–1137.
[2] Singh, D., Yadav, S., Marmostein, D., Lomako, J., Lomako, W.M. and Whelan, W.J. (1988) FASEB J. 2, A557.
[3] Laemmli, J.K. (1970) Nature 227, 680–685.
[4] Chamberlain, J.P. (1970) Anal. Biochem. 98, 132–135.
[5] Singh, D.G. (1994) Ph.D. Thesis, University of Miami, Miami, FL, USA.
[6] Lomako, J., Lomako, W.M. and Whelan, W.J. (1988) FASEB J. 2, 3097–3103.
[7] Rice, R.H., Means, G.E. and Brown, W.D. (1977) Biochim. Biophys. Acta 49, 316–321.
[8] Bidlingmeyer, B.A., Cohen, S.A. and Tarvin, T.L. (1986) J. Chromatogr. 336, 93–104.
[9] Bruins, A.P., Covey, T.R. and Henion, J.D. (1987) Anal. Chem. 59, 2642–2646.
[10] Thomson, B.A., Iribarne, J.V. and Dziedzic, P.J. (1982) Anal. Chem. 54, 2219–2224.
[11] Heilmeyer, L.M.G., Jr., Serwe, M., Weber, C., Metzger, J., Hoffmann-Posorske, E. and Meyer, H.E. (1992) Proc. Natl. Acad. Sci. USA 89, 9554–9558.
[12] Meyer, H.E. (1994) in: Analyzing Post-Translational Protein Modifications (Kellner, R., Lottspeich, F. and Meyer, H.E. eds.) Microcharacterization of Proteins, pp. 131–146, VCH Verlagsgesellschaft, Weinheim.
[13] Viskupic, E., Cao, Y., Zhang, W., Cheng, C., DePaoli-Roach, A.A. and Roach, P.J. (1992) J. Biol. Chem. 267, 25759–25763.
[14] Alonso, M.D., Lomako, J., Lomako, W.M., Whelan, W.J. and Preiss, J. (1994) FEBS Lett. 353, 222–226.
[15] Rothschild, A. and Tandecarz, J.S. (1994) Plant Sci. 97, 119–127.