Glycosylation by *Pichia pastoris* decreases the affinity of a family 2a carbohydrate-binding module from *Cellulomonas fimi*: a functional and mutational analysis

Alisdair B. BORASTON*†‡, R. Antony J. WARREN*† and Douglas G. KILBURN*†‡[1]

*The Protein Engineering Network of Centres of Excellence, PENCE Inc., National Business Centre, 750 Heritage Medical Research Centre, Edmonton, Alberta, Canada T6G 2S2, †Department of Microbiology and Immunology, University of British Columbia, 300-6174 University Boulevard, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3, and ‡The Biotechnology Laboratory, 237-6174 University Boulevard, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3

When produced by *Pichia pastoris*, three of the five Asn-Xaa-Ser/Thr sequences (corresponding to Asn-24, Asn-73 and Asn-87) in the carbohydrate-binding module CBM2a of xylanase 10A from *Cellulomonas fimi* are glycosylated. The glycans are of the high-mannose type, ranging in size from $GlcNAc_2Man_8$ to $GlcNAc_2Man_{14}$. The N-linked glycans block the binding of CBM2a to cellulose. Analysis of mutants of CBM2a shows that glycans on Asn-24 decrease the association constant (K_a) for the binding of CBM2a to bacterial microcrystalline cellulose approx. 10-fold, whereas glycans on Asn-87 destroy binding. The K_a of a mutant of CBM2a lacking all three N-linked glycosylation sites is the same when the polypeptide is produced by either *Escherichia coli* or *P. pastoris* and is approx. half that of wild-type CBM2a produced by *E. coli*.

Key words: cellulose-binding module, mapping, mass spectrometry, N-linked, post-translational modification.

INTRODUCTION

Carbohydrate-binding modules (CBMs) are present in many glycoside hydrolases [1]. For example, the family 2a CBM, CBM2a (previously called CBD_{Cex}), of xylanase 10A from *Cellulomonas fimi*, binds tightly to crystalline and amorphous cellulose and to chitin, the affinity constants (K_a) being micromolar [2,3]. CBM2a is a nine-stranded all-β-sheet polypeptide with a β-barrel topology [4]. Three tryptophan residues that form a ridge on one surface of the polypeptide are critical to ligand binding [5].

Examples of CBM hybrid proteins incorporating CBM2a have been successfully produced in *Pichia pastoris* [6], which has become an attractive organism for the high-level production of recombinant proteins [7–9]. However, as with many recombinant proteins produced in eukaryotic hosts, glycosylation of these proteins, particularly the CBM, was an important issue. Preliminary results indicated that CBM fusion proteins produced in *P. pastoris* could bind cellulose more weakly than their counterparts produced in *Escherichia coli*.

As a result of the importance of *P. pastoris* as a host for the production of recombinant proteins, the post-translational modification performed by this organism has become a matter of practical interest. *P. pastoris* can attach N-linked glycans to secreted heterologous proteins [10–13]. These glycans are of the high-mannose type typical of fungal systems. Here we show that CBM2a produced by *P. pastoris* is N-glycosylated; we identify the sites of glycosylation and present a preliminary characterization of the glycans. We show also that the N-linked glycans decrease the affinity of CBM2a for cellulose and we discuss the implications of this for cellulose recognition.

MATERIALS AND METHODS

Strains and vectors

All subcloning steps were performed with pZeRO 1.1 (Invitrogen, San Diego, CA, U.S.A.). The *P. pastoris* expression/shuttle vector was pPICZαA (Invitrogen). All DNA manipulations were performed in *E. coli* TOPP 10F. Expression clones were obtained by electrotransformation of *P. pastoris* strain GS115 in accordance with the supplier's recommendations (Invitrogen). pTUGKH, a derivative of pTUGA encoding kanamycin resistance, was used for bacterial expression [14].

DNA manipulations

Agarose-gel electrophoresis, small-scale plasmid isolation and *E. coli* transformations were performed as described previously [15]. Large-scale plasmid isolation for sequencing was done with Qiagen Tip-100 columns (Qiagen, Chatsworth, CA, U.S.A.). Restriction enzymes were used as recommended by the manufacturer. All PCR products and DNA fragments resulting from restriction digests were purified from agarose gels after electrophoresis by using Qiaex II (Qiagen) in accordance with the manufacturer's protocol. DNA was sequenced by the Nucleic

Abbreviations used: CBM, carbohydrate-binding module; ConA, concanavalin A from *Canavalia ensiformis*; EndoF1, endoglycosidase F1 from *Chryseobacterium meningosepticum*; K_a, association constant; MALDI–TOF-MS, matrix-assisted laser desorption ionization–time-of-flight MS; TFA, trifluoroacetic acid.

[1] To whom correspondence should be addressed at Biotechnology Laboratory, University of British Columbia (e-mail kilburn@interchange.ubc.ca).

Table 1 Oligonucleotide primers used in the PCR generation of CBM2a variants

Underlining indicates the cloning restriction endonuclease site sequences. Bold indicates the screening/silent restriction endonuclease site sequences. Italics denote substituted bases to introduce the amino acid substitution and/or silent restriction endonuclease sites.

Primer	Amino acid substitution introduced	Cloning restriction enzyme site	Screening restriction enzyme site	Sequence (5′ → 3′)
1	–	*Not*I	*Nar*I	AACGCGGCCGCTTATTAACCAACGGTGCAAGGGGTACCGTTCAGAGAGAAAGCGGTT**GGCGC**CGCGTTGGTACCG
2	Asn-24 → Gln	*Hpa*I	*Bst*EII	GGTGTTAACCAGTGGAACACCGGTTTCACCGCT*CA*G**GTTACC**GTTAAA
3	Ser-31 → Gly	*Hpa*I or *Psp*1406I	*Bpu*1102	GGTGTTAACCAGTGGAACACCGGTTTCACCGCTAACGTTACCGTTAAAAACACG*G***GCTC***A***GC**TCCGGTTGAC
4	Ser-31 → Gly with Asn-24 → Gln	*Bst*EII	*Bpu*1102	GGTTTCACCGCTCAGGTTACCGTTAAAAACACG*G***GCTC***A***GC**TCCGGTTGAC
5	Ser-75 → Asn	*Pst*I	*Ssp*I	TCCACCTGCAGG***AATATT***ACCGTTCCACGG
6	Asn-87 → Gln	*Nar*I	*Sty*I	GGTTGGCGCCGCGTTGGTACCGGTGTGAGAA**CC***T***T***G***G**AAACCGAACTG
7	Thr-105 → Ala	*Not*I	*Apa*I	AACGCGGCCGCTTATTAACCAACGGTGCAAG**GGG***CC***C**CGTTCAGAGAGAA
6	–	*Aat*II or *Hpa*I		GACTGGTTCCAATTGACAAG

Table 2 CBM2a mutants

Amino acid substitutions are given in single-letter amino acid codes. The column headed 'N-glycosylation site' shows remaining N-glycosylation sites.

Amino acid substitutions	N-glycosylation site	Polypeptide abbreviation
N24Q S31G N87Q T105A	Asn-73	CBM2a.4
N24Q S31G S75N N87Q T105A	None	CBM2a.5
N24Q S75N N87Q T105A	Asn-29	CBM2a.6
N24Q S31G S75N N87Q	Asn-103	CBM2a.7
S31G S75N N87Q T105A	Asn-24	CBM2a.8
N24Q S31G S75N T105A	Asn-87	CBM2a.9
None	All	CBM2a

Acids and Protein Services Unit (Biotechnology Laboratory, University of British Columbia, Canada) with the AmpliTaq dye termination cycle sequencing protocol and an Applied Biosystems Model 377 sequencer.

Standard PCR procedures were used to introduce base-pair substitutions encoding specific amino acid substitutions into the CBM2a gene fragment by using oligonucleotide primers (Table 1). Standard cloning procedures were used to generate combinations of mutations (Table 2) in the cloning vector pZeRO 1.1 (Invitrogen). An N-terminal hexahistidine tag, separated from the CBM2a gene by a Factor Xa cleavage site, was incorporated. Finally, the mutated CBM2a gene fragments were inserted into the *P. pastoris* shuttle vector pPICZαA to generate fusions encoding the *Saccharomyces cerevisiae* α-factor leader peptide, hexahistidine tag, Factor Xa cleavage site and the altered CBM2a gene fragment. Bacterial expression plasmids were prepared by inserting the mutated CBM2a gene fragments into pTUGKH by using 5′ *Aat*II and 3′ *Hin*dIII restriction sites to produce fusion with the Xyn10A secretion leader and a hexahistidine tag as described previously [5].

Production and purification of CBM2a and CBM2a mutants

Small-scale expression in *P. pastoris* was done at 30 °C in 50 ml culture tubes containing 5 ml of buffered glycerol complex medium [BMGY: 1 % (w/v) yeast extract/2 % (w/v) peptone/100 mM potassium phosphate (pH 6.0)/1.34 % (w/v) yeast nitrogen base/4×10^{-5} % biotin/1 % (v/v) glycerol]. Cultures were grown to a D_{600} of 2.0 and induced by removing the supernatant and replacing the medium with buffered minimal medium [BMM: 100 mM potassium phosphate (pH 6.0)/1.34 % (w/v) yeast nitrogen base/4×10^{-5} % biotin] containing 0.5 % (v/v) methanol and supplemented with 0.04 % L-histidine. Cultures were incubated for 48 h at 30 °C. Supernatants were obtained by centrifugation at 4 °C for 10 min at 27000 ***g***. Large-scale cultures were started by inoculation of 350 ml of BMGY in 2-litre baffled flasks with 1 ml of a 5 ml culture grown overnight in BMGY at 30 °C. Cultures were grown to a D_{600} of approx. 2.0 and induced by replacing the supernatant with 350 ml of BMM. Cultures were incubated for 48 h at 30 °C with additions of methanol to 0.5 % (v/v) every 12 h. The supernatant was obtained by removal of the cells by centrifugation at 4 °C for 20 min at 11000 ***g***. The pH of the supernatant was adjusted to 8.0 with NaOH; the supernatant was then filtered under vacuum through a 0.7 μm cut-off glass-fibre filter to remove particulate material.

CBM2a and its mutants produced in *P. pastoris* were purified by batch immobilized metal-affinity chromatography. Ni^{2+}-charged His-Bind resin (Novagen, Milwaukee, WI, U.S.A.) (20 ml) was stirred with the clarified supernatant at 4 °C for 2 h. The resin was recovered by vacuum filtration with a 0.7 μm cut-off glass-fibre filter. The resin was washed on the filter with 500 ml of binding buffer. Adsorbed proteins were eluted by washing the resin on the filter with 50–100 ml of binding buffer containing 500 mM imidazole. The eluted protein was assessed for purity by SDS/PAGE and then exchanged into 50 mM potassium phosphate, pH 7.0, or 25 mM Tris/HCl, pH 7.4, and concentrated in a stirred ultrafiltration unit (Amicon, Beverly, MA, U.S.A.) with a 1 kDa cut-off filter (Filtron, Northborough, MA, U.S.A.). Yields were typically 0.5–5 mg/l of culture supernatant. Samples used for quantitative binding studies were further purified by affinity chromatography on concanavalin A (ConA)–Sepharose (Pharmacia, Uppsala, Sweden) in accordance with the manufacturer's recommendations. The fractions were assayed for protein by SDS/PAGE and exchanged into 50 mM potassium phosphate, pH 7.0, as above.

CBM2a and its mutants were produced in *E. coli* and purified by immobilized metal-affinity chromatography as described previously [5]. Yields were typically 30–60 mg/l of culture supernatant. The polypeptides were concentrated and exchanged into 50 mM potassium phosphate, pH 7.0, as above.

Adsorption studies

For qualitative adsorption experiments, 1 ml samples of culture supernatant were mixed (4 °C) end-over-end with 5 mg of cellu-

lose (Avicel™; Little Island, County Cork, Ireland). After 1 h the Avicel was collected by centrifugation (13000 ***g***, 4 °C) and washed three times with 1 ml of potassium phosphate buffer (50 mM, pH 7.0). After the last wash, the cellulose was collected by centrifugation and boiled for 5 min after the addition of 40 μl of SDS loading buffer. Fractions of 20 μl were then analysed for protein by Western blotting. Quantitative binding to bacterial microcrystalline cellulose was determined as described previously [3,5].

Western immunoblotting

Proteins separated by electrophoresis through 16 % (w/v) polyacrylamide gels were electroblotted to PVDF membranes (Immobilon™; Millipore, Bedford, MA, U.S.A.). CBMs were detected with rabbit polyclonal anti-CBM2a antibodies at a dilution of 1:10000. Goat anti-rabbit IgG antibodies conjugated with horseradish peroxidase were used as secondary antibodies at a dilution of 1:10000. Glycoproteins were detected with ConA conjugated to horseradish peroxidase (ConA–HRP; Seikagaku, Tokyo, Japan) at a dilution of 1:2000. Antibodies or ConA–HRP were diluted in 10 ml of PBS containing 0.5 % BSA and 0.05 % (v/v) Tween 20. These solutions were incubated with the PVDF blot for 1 h at room temperature. Blots were washed three times with 75 ml of PBS containing 0.05 % (v/v) Tween 20 after probing. Blots were developed with a chemiluminescent horseradish peroxidase detection kit from Amersham.

Fluorophore-assisted carbohydrate electrophoresis

Monosaccharide composition analysis and glycan size profiling were done with kits purchased from Glyko (Novato, CA, U.S.A.) with 50–200 μg of protein. All reactions and electrophoretic separations were performed in accordance with the protocols supplied by the manufacturer.

Protease digests

Purified protein (200 μl; 100–200 μM) was desalted and exchanged into distilled water by overnight drop dialysis with VS 0.025 μm filters (Millipore). The concentration of the exchanged protein was measured by A_{280}; 3 nmol of this protein was dried in a vacuum concentrator followed by dissolution in 5 μl of 8 M urea containing 10 mM dithiothreitol. This was incubated for 30 min at 50 °C. HCO_3 (50 mM, 145 μl) was added slowly to the sample followed by the addition of 1 μl of chymotrypsin (1 mg/ml) or 1 μl of chymotrypsin and 1 μl of trypsin (both at 1 mg/ml). The protease digests were incubated at 37 °C for 4 h. Samples of the reactions were diluted 1:4 in 70 % (v/v) acetonitrile/0.1 % (v/v) trifluoroacetic acid (TFA) before preparation for MS.

Enzymic deglycosylation and preparation of glycans for MS

All deglycosylation reactions were performed with recombinant endoglycosidase F1 from *Chryseobacterium meningosepticum* (EndoF1) purchased from Boehringer Mannheim (Laval, Quebec, Canada). Deglycosylation reactions for SDS/PAGE were performed in 20 μl volumes containing 100 mM potassium phosphate buffer, pH 6.0, approx. 10 μg of protein and 0.05 unit of EndoF1 [one unit is defined as the enzyme activity required to hydrolyse 1 mg of dansyl-Asn$(GlcNac)_2(Man)_5$ within 60 min at 37 °C at pH 5.0]. Reaction mixtures were incubated for 2 h at 37 °C. Reactions for the preparation of glycans for MS were performed in 20–100 μl volumes containing 25 mM ammonium acetate buffer, pH 6.0, approx. 3 nmoles protein and 0.5–1 unit of EndoF1. Reaction mixtures were incubated for 4 h at 37 °C. Ethanol [2 vol., 95 % (v/v)] chilled to −20 °C was added and the reactions were held at −20 °C for 30 min. Precipitated material was pelleted by centrifugation at 27000 ***g*** and 4 °C for 10 min. The supernatant fraction containing the released glycans was removed and evaporated to dryness in a vacuum concentrator. The dried material was dissolved in 5 μl of distilled water and samples were diluted 1:1, 1:4 and 1:9 in 70 % (v/v) acetonitrile/0.1 % (v/v) TFA before preparation for MS.

Matrix-assisted laser desorption ionization–time-of-flight (MALDI–TOF) MS

All MALDI–TOF-MS spectra were collected with a SELDI-MassPhoresis mass spectrometer (Ciphergen, Palo Alto, CA, U.S.A.). Samples of intact CBM were prepared for MS by diluting 2 μl of a 10 pmol/μl solution of CBM with an equal volume of 70 % (v/v) acetonitrile/0.1 % (v/v) TFA saturated with sinipinic acid (Sigma). Bovine superoxide dismutase (12230.9 Da) prepared in the same matrix was used to calibrate the mass spectrometer. Samples of protease digests were prepared by diluting 2 μl of the prediluted reaction with an equal volume of 70 % (v/v) acetonitrile/0.1 % (v/v) TFA saturated with cinnamic acid (Sigma). Glycan samples were prepared by diluting 2 μl of the pre-diluted glycans with an equal volume of 10 mg/ml dihydroxybenzoic acid (Sigma) in 70 % (v/v) acetonitrile/0.1 % (v/v) TFA. Angiotensin (1296.5 Da) and fibrinopeptide B (1570.6 Da) were used as external calibrants for the protease-digested samples and the glycan samples. The same matrix was used for the calibrants as for the experimental samples. Sample (1 μl) was spotted on the MALDI target and left to dry at room temperature under atmospheric pressure. Spectra were acquired and analysed with the SELDI-MassPhoresis software supplied with the mass spectrometer.

Protein concentration determination

The concentration of purified CBM2a was determined from A_{280} by using a calculated molar absorption coefficient [16] of 27625 $M^{-1} \cdot cm^{-1}$.

RESULTS

Production of CBM2a in *P. pastoris*

Production of the PpCBM2a (see Table 1 for polypeptide designations) construct with the *S. cerevisiae* α-factor leader sequence led to the secretion of three polypeptides into the culture supernatant, all of them larger than CBM2a produced by *E. coli* (Figure 1, lane 2). None of the polypeptides bound strongly to cellulose (Figure 1, lanes 3 and 4). When tunicamycin, an inhibitor of N-linked glycosylation [17], was added at the time of induction, only a single polypeptide of the same size as that produced by *E. coli* was produced by *P. pastoris*; this bound tightly to cellulose (Figure 1, lanes 5–7). Thus N-linked glycosylation of CBM2a by *P. pastoris* decreased its affinity for cellulose. The glycosylation seemed to be non-uniform, as indicated by the heterogeneous mixture of products.

N-linked glycan composition and profiling

Mannose and *N*-acetylglucosamine residues were present in the fluorophore-assisted-carbohydrate-electrophoresis ('FACE') monosaccharide analysis of glycosylated PpCBM2a (results not shown). The masses of the neutral glycans released from PpCBM2a by EndoF1, which cleaves N-linked glycans at the β1-

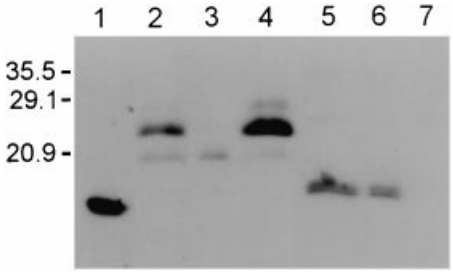

Figure 1 N-glycosylation of CBM2a and its effect on cellulose binding

Duplicate cultures of *P. pastoris* containing the CBM2a gene were grown for 8 h at 30 °C, then induced with 0.5% (v/v) methanol for 24 h. At induction one of the duplicate cultures was treated with 30 μg/μl tunicamycin (lanes 5–7). Samples were prepared as described in the Materials and methods section, separated by SDS/PAGE and blotted. The blot was probed with anti-CBM2a antibodies. Lane 1, *E. coli*-produced CBM2a; lanes 2 and 5, *P. pastoris* culture supernatant; lanes 3 and 6, Avicel-bound polypeptides desorbed by boiling with SDS/PAGE loading buffer; lanes 4 and 7, *P. pastoris* culture supernatant after preincubation with Avicel to remove cellulose-binding proteins. The concentrations of all samples were normalized by volume to the material loaded in the unbound samples to allow semiquantitative comparisons. The positions of molecular mass markers are indicated (in kDa) at the left.

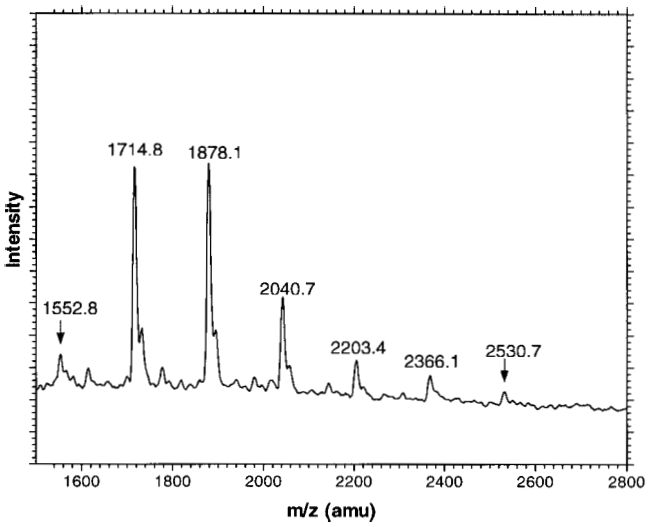

Figure 2 MALDI–TOF-MS spectra of PpCBM2a glycans released by EndoF1

The glycans were prepared for MS as described in the Materials and methods section. The masses of neutral glycans obtained are given as K^+ adducts and ranged from $GlcNAc_1Man_8$ to $GlcNAc_1Man_{14}$. The calculated relative molecular masses of the oligosaccharides were as follows: $GlcNAc_1Man_8$, 1518.3 (1557.4); $GlcNAc_1Man_9$, 1680.5 (1719.6); $GlcNAc_1Man_{10}$, 1842.6 (1881.7); $GlcNAc_1Man_{11}$, 2004.7 (2043.8); $GlcNAc_1Man_{12}$, 2166.9 (2206); $GlcNAc_1Man_{13}$, 2329.0 (2368.1); $GlcNAc_1Man_{14}$, 2491.2 (2530.3); $GlcNAc_1Man_{15}$, 2653.3 (2692.4). The masses in parentheses are those of theoretical K^+ adducts. Abbreviation: amu, atomic mass units.

4 linkage of the chitobiose core, leaving a single GlcNAc attached to the protein [18,19], corresponded well to the masses of K^+ adducts of glycans ranging in size from $GlcNAc_1Man_8$ to $GlcNAc_1Man_{14}$ (Figure 2). Fluorophore-assisted-carbohydrate-electrophoresis analysis of the size distribution of the N-glycans released by peptide-N-glycosidase F was consistent with the MS results (results not shown).

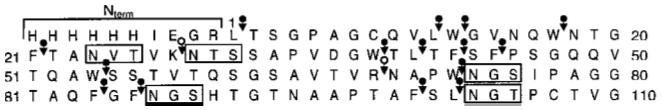

Figure 3 Potential N-glycosylation sites of CBM2a

Boxed sequences indicate potential N-glycosylation sites. Arrowheads attached to filled circles indicate chymotrypsin cleavage sites; arrowheads attached to open circles denote trypsin cleavage sites. The numbering is in accordance with the co-ordinate file of the structure of CBM2a. The sequence marked N_{term} includes the His_6 affinity tag and the IEGR FXa recognition site.

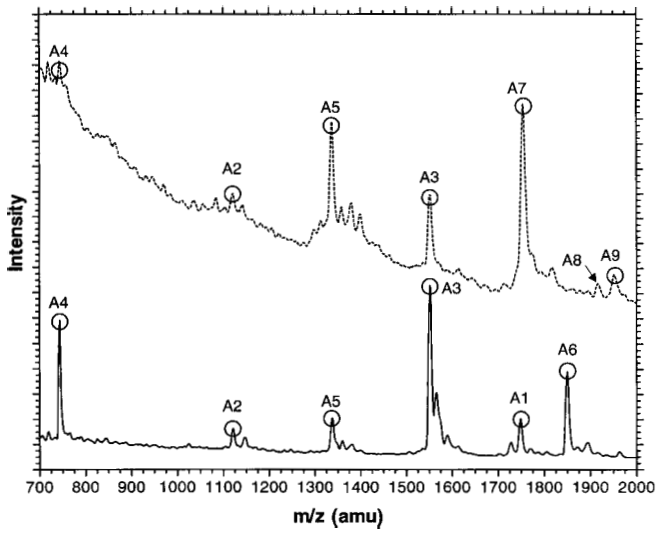

Figure 4 MALDI–TOF-MS spectra of peptides generated by chymotrypsin cleavage of CBM2a produced by *P. pastoris*

Samples of CBM2a purified from *E. coli* cultures (continuous line) or CBM2a purified from *P. pastoris* cultures (dotted line) were treated with EndoF1 followed by denaturation with urea and treatment with protease. Peak labels are described in Table 3. Abbreviation: amu, atomic mass units.

Locations of N-linked glycans

The GlcNAc residue left on the protein by EndoF1 can be used as a marker of N-linked glycosylation. There are five Asn-Xaa-Ser/Thr sequences in CBM2a, corresponding to Asn-24, Asn-29, Asn-73, Asn-87 and Asn-103, separated by chymotrypsin or trypsin cleavage sites (Figure 3). Peptides from EndoF1-treated PpCBM2a and EcCBM2a were obtained by treatment of denatured polypeptide with either chymotrypsin or a mixture of chymotrypsin and trypsin. Analysis of these mixtures by MALDI–TOF-MS (Figures 4 and 5) showed that peptides containing the Asn-24 (peak A9), Asn-73 (Peak B10) and Asn-87 residues (peaks A7, A8, B8 and B9) from PpCBM2a had masses consistent with the presence of a GlcNAc residue (Tables 3 and 4). Furthermore, chymotrypsin cleaved completely at Trp-72 of EcCBM2a, resulting in peak A6, but not at Trp-72 of PpCBM2a, as indicated by the absence of peak A6 in the digest of PpCBM2a. Presumably the presence of GlcNAc at Asn-73 inhibited cleavage at this site. A small peak corresponding to the expected mass of peptide Gly-85–Phe-100 with attached GlcNAc and a single hexose sugar was also present (peaks A8 and B9). The identification of three N-glycosylation sites is consistent with the presence of three polypeptide species identified by SDS/

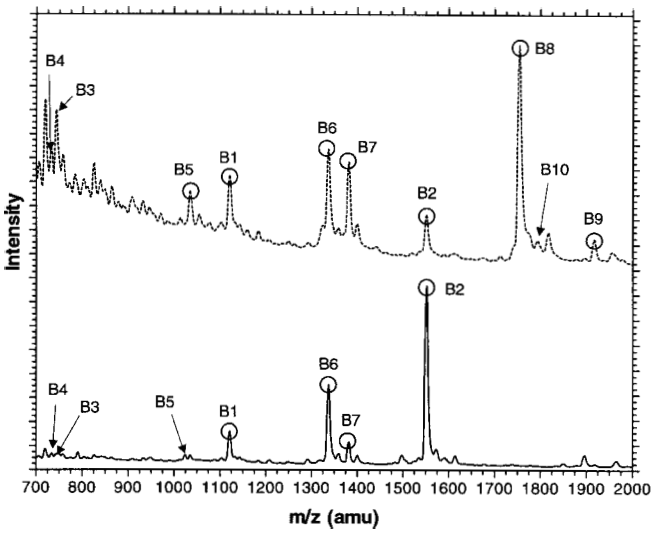

Figure 5 MALDI–TOF-MS spectra of peptides generated by cleavage of CBM2a with chymotrypsin and with trypsin

Samples of CBM2a purified from *E. coli* cultures (continuous line) or CBM2a purified from *P. pastoris* cultures (dotted line) were treated with EndoF1 followed by denaturation with urea and treatment with protease. Peak labels are described in Table 4. Abbreviation; amu, atomic mass units.

PAGE. The presence of the hexose sugar suggested that a small amount of O-glycosylation had occurred in addition to N-linked glycosylation.

The locations of the N-linked glycosylation acceptor sites were verified by site-directed mutation of key amino acids of the Asn-Xaa-Ser/Thr sequences (Table 2). All of the CBM2a mutants were secreted by *P. pastoris*. Both Asn-24 and Asn-87 were glycosylated, as indicated by the increased molecular masses of PpCBM2a.8 and PpCBM2a.9 (Figure 6). Treatment of purified PpCBM2a, PpCBM2a.8 and PpCBM2a.9 with EndoF1 decreased their masses to that of EcCBM2a as judged by SDS/PAGE (data not shown). The PpCBM2a.4 mutant, which had only the Asn-73 glycosylation site, was not glycosylated. In contrast, MALDI–TOF mapping (above) indicated that the Asn-73 site in the wild-type PpCBM2a was glycosylated.

Binding to cellulose

PpCBM2a.9 did not bind to Avicel (Figure 6). Only a small amount of PpCBM2a.9 with a molecular mass similar to that of unglycosylated CBM was bound. This indicates that glycosylation at Asn-87 abolished the ability of PpCBM2a.9 to bind cellulose. Binding of glycosylated wild-type PpCBM2a sample to Avicel seemed limited to the lower-molecular-mass bands. N-glycosylated PpCBM2a.8 bound relatively well to Avicel, as did the remaining CBM mutants that were not N-glycosylated when expressed in *P. pastoris*.

Table 3 MALDI–TOF-MS data for CBM2a after deglycosylation and digestion with chymotrypsin

Peak labels are those in Figure 4. Amino acids and numbering refer to those used in Figure 3. Abbreviations: Hex, protein-linked hexose sugar adding a mass of 162.12 Da; GlcNAc, protein-linked *N*-acetylglucosamine sugar adding a mass of 203.2 Da.

		Molecular mass (Da)		
Peak label	Peptide	Predicted	Found (*E. coli*)	Found (*P. pastoris*)
A1	Thr-22–Trp-38	1746.9	1746.6	–
A2	Asn-73–Phe-84	1119.2	1119.2	1119.2
A3	Gly-85–Phe-100	1549.6	1549.6	1549.6
A4	Asn-103–Gly-110	747.8	741.8	743.2
A5	Ser-43–Trp-54	1335.4	1335.0	1335.0
A6	Ser-55–Trp-72	1848.8	1847.9	–
A7	Gly-85–Phe-100 + GlcNAc	1752.7	–	1752.3
A8	Gly-85–Phe-100 + GlcNAc + Hex	1914.7	–	1913.5
A9	Thr-22–Trp-38 + GlcNAc	1950.0	–	1949.9

Table 4 MALDI–TOF-MS data for CBM2a after deglycosylation and digestion with chymotrypsin and with trypsin

Peak labels are those in Figure 5. Amino acids and numbering refer to those used in Figure 3. Abbreviations: Hex, protein-linked hexose sugar adding a mass of 162.12 Da; GlcNAc, protein-linked *N*-acetylglucosamine sugar adding a mass of 203.2 Da.

		Molecular mass (Da)		
Peak label	Peptide	Predicted	Found (*E. coli*)	Found (*P. pastoris*)
B1	Asn-73–Phe-84	1119.2	1119.2	1119.2
B2	Gly-85–Phe-100	1549.6	1549.6	1549.6
B3	Asn-103–Gly-110	747.8	742.8	741.8
B4	Thr-22–Lys-28	731.8	732.4	731.7
B5	Asn-29–Trp-38	1033.1	1033.6	1033.1
B6	Ser-43–Trp-54	1335.4	1335.2	1335.3
B7	Ser-55–Arg-68	1379.5	1379.6	1379.6
B8	Gly-85–Phe-100 + GlcNAc	1752.7	–	1752.4
B9	Gly-85–Phe-100 + GlcNAc + Hex	1914.7	–	1914.2
B10	Asn-69–Phe-84 + GlcNAc	1790.9	–	1790.7

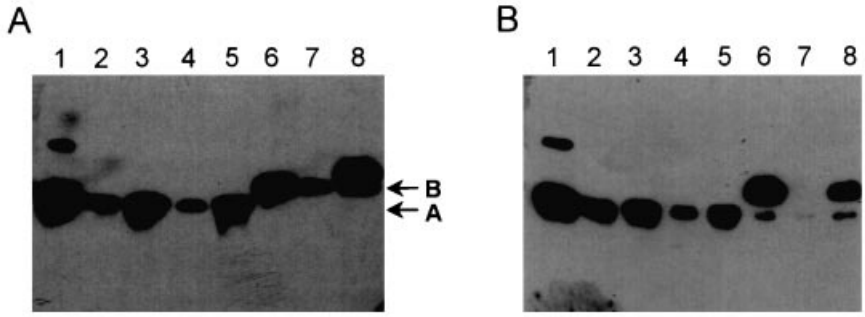

Figure 6 SDS/PAGE mobility of CBM2a mutants produced by *P. pastoris* and detected by Western blotting with anti-CBM2a antibodies

Cultures of *P. pastoris* containing the mutated CBM2a genes were grown for 8 h at 30 °C, then induced with 0.5% (v/v) methanol for 24 h. (**A**) Samples of cleared supernatant. (**B**) Samples of cleared supernatant bound to cellulose and desorbed by boiling. Equivalent amounts of sample were loaded in each panel. Lane 1, EcCBM2a; lane 2, PpCBM2a.4; lane 3, PpCBM2a.5; lane 4, PpCBM2a.6; lane 5, PpCBM2a.7; lane 6, PpCBM2a.8; lane 7, PpCBM2a.9; lane 8, PpCBM2a (see Table 2 for abbreviations). Arrow A indicates masses corresponding to CBM2a produced in *E. coli*; arrow B indicates glycoform(s).

Table 5 Affinity of CBM2a variants for bacterial microcrystalline cellulose (BMCC)

CBM abbreviations correspond to those in Table 1. Abbreviations: n.d., not determined; n.q., not quantifiable. Binding capacity is defined as the number of CBM-binding sites per unit mass of cellulose. Results are means ± S.E.M. from a non-linear regression.

		Binding parameters for CBM from *E. coli*		Binding parameters for CBM from *P. pastoris*	
CBM mutant	Glycosylation site	K_a (μM^{-1})	Binding capacity (μmol/g of BMCC)	K_a (μM^{-1})	Binding capacity (μmol/g of BMCC)
CBM2a.5	None	1.8 ± 0.6	10.7 ± 1.1	1.5 ± 0.1	11.8 ± 0.3
CBM2a.8	Asn-24	2.8 ± 0.4	10.2 ± 0.3	0.3 ± 0.0	10.0 ± 0.7
CBM2a.9	Asn-87	3.5 ± 1.1	12.4 ± 1.1	n.q.	n.q.
CBM2a	All	2.9 ± 0.7	12.5 ± 0.9	n.d.	n.d.

The mutations had no affect on the K_a of the CBMs for cellulose or on the binding capacities (Table 5). Glycosylation at Asn-24 (PpCBM2a.8) decreased the association constant 10-fold relative to the same mutants produced in *E. coli*. Glycosylation of Asn-87 (PpCBM2a.9) resulted in binding too low to be quantifiable, which was consistent with the qualitative binding experiments. The K_a of PpCBM2a.5 was similar to that for EcCBM2a.5.

DISCUSSION

N-linked glycosylation

The size distribution and sugar content of the N-glycans on PpCBM2a were very similar to those found previously on other N-glycosylated proteins produced by *P. pastoris* [10,12,20]. The relative uniformity of the N-glycosylation in *P. pastoris* distinguishes it from *S. cerevisiae*, where in general the glycans are quite heterogeneous in size [21,22]. Furthermore, *S. cerevisiae* has a tendency to hyperglycosylate, forming extremely large glycans containing more than 50 mannose residues. This has been observed only infrequently in *P. pastoris* [10] and was not found for PpCBM2a. The relative homogeneity of the N-linked glycans attached to proteins secreted by *P. pastoris* is one of the biotechnological advantages of this expression system [10,12].

Of the five sites that could potentially act as N-linked glycan acceptors, only three were glycosylated. Several factors have been proposed to determine the suitability of a potential N-glycosylation site. Positive determinants are proximity to the N-terminus of the polypeptide and placement in a turn or loop; a negative determinant is the presence of side chains that are prone to being buried during protein folding [23]. The structure of CBM2a is known [4]; of the five potential acceptors, four are on the surface: Asn-24, Asn-73, Asn-87 and Asn-103. The fifth, Asn-29, is located in a turn but is not an appropriate acceptor because it is buried. Asn-103, although exposed, is near the C-terminus of the polypeptide. It is thought that sites at the C-terminus of a polypeptide have less exposure time to the glycosylation mechanism during co-translational modification than sites at the N-terminus, so they are glycosylated less frequently [24]. The size profile of glycans on the PpCBM2a.8 mutant was identical with that of the wild-type PpCBM2a (results not shown). This suggests that the size distributions of the N-glycans at each of the sites on CBM2a do not differ significantly.

It is interesting that the Asn-73 site functions as an acceptor of N-glycosylation in the polypeptide containing the other N-glycosylation sites but does not when all the other sites are absent. The reason for this phenomenon is unknown. The occupancy of an adjacent N-glycosylation site can affect the occupancy of another potential site [24]. However, this usually occurs at sites separated by fewer than four amino acids in the primary sequence and is a negative effect: glycosylation at one site hinders glycosylation at the other.

Glycosylation and cellulose binding

The N-glycosylation of CBM2a greatly decreased its ability to bind cellulose. In particular, PpCBM2a bound poorly to cellulose and PpCBM2a.9 bound not at all, whereas glycosylation at Asn-24 decreased the binding of PpCBM2a.8 only 10-fold. The most

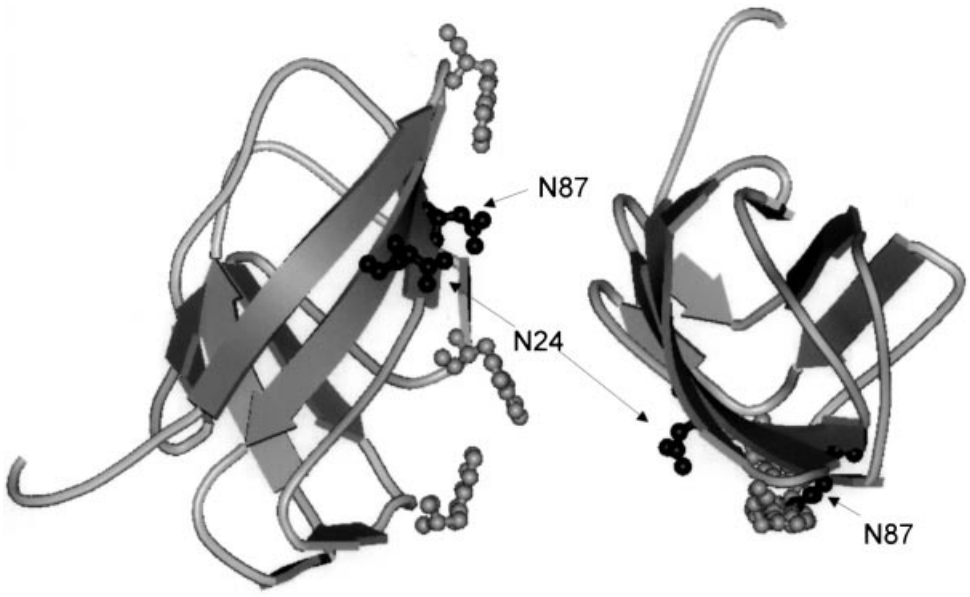

Figure 7 Placement of the Asn-24 (N24) and Asn-87 (N87) glycosylation sites (black and labelled) in side view (left panel) and end-on view (right panel) of CBM2a

The tryptophan residues involved in binding to crystalline cellulose are shown in light grey.

plausible explanation is steric effects resulting from the large sugar moiety hindering the interaction of the binding face of the CBM with the cellulose surface. Asn-87 is directly on the binding face between two tryptophan residues that are essential to binding (Figure 7) [5]. The presence of a large glycan at this position in PpCBM2a and PpCBM2a.9 would surely impose steric hindrances, blocking binding. This effect is not complete in PpCBM2a.8, as indicated by the retention of considerable affinity for cellulose. The interaction of CBM2a with cellulose is a dynamic process. It has been proposed that only two of the three surface tryptophan residues interact with the cellulose surface at any one time, with the centre tryptophan residue, Trp-54, acting as the pivotal residue [5]. Replacement of the peripheral tryptophans (Trp-17 and Trp-72) with alanine results in an approx. 10-fold decrease in affinity [5], similar to the effect of glycosylation at Asn-24. It is possible that a bulky glycan on Asn-24 sterically inhibits the interaction of one of the tryptophans, probably Trp-17 on the basis of its proximity to Asn-24, with cellulose.

The thermal and chemical stability of the CBM2a mutants, both glycosylated and unglycosylated, was not significantly altered (results not shown), indicating a lack of gross structural changes resulting from glycosylation at any of the sites. Thus the effects of glycosylation on binding are probably a result of the proposed steric effects rather than structural effects.

Biotechnological implications

The most significant practical finding was the construction of a CBM2a mutant that lacked N-glycosylation sites and retained its affinity for cellulose when produced in *P. pastoris*. The poor binding of N-glycosylated CBM presents a barrier to the use of these modules as fusion partners for the immobilization of proteins if the hybrid proteins are produced in hosts that perform N-linked glycosylation. The use of the CBM2a.5 variant will permit the construction of CBM gene fusions for expression in eukaryotic hosts that will bind cellulose with a high affinity. Such hosts are frequently the only option for producing complex eukaryotic proteins.

We thank Emily Amandoron-Akow and Emily Kwan for technical assistance. This work was supported by the National Sciences and Engineering Research Council, the Protein Engineering Network Centres of Excellence and CBD Technologies.

REFERENCES

1 Tomme, P., Warren, R. A., Miller, Jr, R. C., Kilburn, D. G. and Gilkes, N. R. (1995) Cellulose-binding domains: classification and properties. In Enzymatic Degradation of Insoluble Polysaccharides (Saddler, J. N. and Penner, M., eds.), pp. 142–163, American Chemical Society, Washington, DC

2 Ong, E., Gilkes, N. R., Miller, Jr, R. C., Warren, R. A. and Kilburn, D. G. (1993) The cellulose-binding (CBD_{Cex}) domain of an exoglucanase from *Cellulomonas fimi*: production in *Escherichia coli* and characterization of the polypeptide. Biotechnol. Bioeng. **42**, 401–409

3 Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. and Haynes, C. A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. Proc. Natl. Acad. Sci. U.S.A. **93**, 12229–12234

4 Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E. and Harvey, T. S. (1995) Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. Biochemistry **34**, 6993–7009

5 McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A. and Kilburn, D. G. (2000) Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. Protein Eng. **13**, 801–809

6 Guarna, M. M., Cote, H. C., Amandoron, E. A., MacGillivray, R. T., Warren, R. A. and Kilburn, D. G. (1996) Engineering factor X fusions for expression in *Pichia pastoris*. Ann. N.Y. Acad. Sci. **799**, 397–400

7 Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K. and Romanos, M. A. (1991) High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. Biotechnology (N.Y.) **9**, 455–460

8 Cregg, J. M., Vedvick, T. S. and Raschke, W. C. (1993) Recent advances in the expression of foreign genes in *Pichia pastoris*. Biotechnology (N.Y.) **11**, 905–910

9 Higgins, D. R. and Cregg, J. M. (1998) Introduction to *Pichia pastoris*. Methods Mol. Biol. **103**, 1–15

10 Grinna, L. S. and Tschopp, J. F. (1989) Size distribution and general structural features of N-linked oligosaccharides from the methylotrophic yeast, *Pichia pastoris*. Yeast **5**, 107–115

11 Duman, J. G., Miele, R. G., Liang, H., Grella, D. K., Sim, K. L., Castellino, F. J. and Bretthauer, R. K. (1998) O-Mannosylation of *Pichia pastoris* cellular and recombinant proteins. Biotechnol. Appl. Biochem. **28**, 39–45
12 Miele, R. G., Nilsen, S. L., Brito, T., Bretthauer, R. K. and Castellino, F. J. (1997) Glycosylation properties of the *Pichia pastoris*-expressed recombinant kringle 2 domain of tissue-type plasminogen activator. Biotechnol. Appl. Biochem. **25**, 151–157
13 Bretthauer, R. K. and Castellino, F. J. (1999) Glycosylation of *Pichia pastoris*-derived proteins. Biotechnol. Appl. Biochem. **30**, 193–200
14 Graham, R. W., Greenwood, J. M., Warren, R. A., Kilburn, D. G. and Trimbur, D. E. (1995) The pTugA and pTugAS vectors for high-level expression of cloned genes in *Escherichia coli*. Gene **158**, 51–54
15 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
16 Mach, H., Middaugh, C. R. and Lewis, R. V. (1992) Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. Anal. Biochem. **200**, 74–80
17 Elbein, A. D. (1984) Inhibitors of the biosynthesis and processing of N-linked oligosaccharides. CRC Crit. Rev. Biochem. **16**, 21–49
18 Plummer, T. H. J. and Tarentino, A. L. (1991) Purification of the oligosaccharide-cleaving enzymes of *Flavobacterium meningosepticum*. Glycobiology **1**, 257–263
19 Tarentino, A. L., Quinones, G., Schrader, W. P., Changchien, L. M. and Plummer, T. H. J. (1992) Multiple endoglycosidase (Endo) F activities expressed by *Flavobacterium meningosepticum*. Endo F1: molecular cloning, primary sequence, and structural relationship to Endo H. J. Biol. Chem. **267**, 3868–3872
20 Miele, R. G., Castellino, F. J. and Bretthauer, R. K. (1997) Characterization of the acidic oligosaccharides assembled on the *Pichia pastoris*-expressed recombinant kringle 2 domain of human tissue-type plasminogen activator. Biotechnol. Appl. Biochem. **26**, 79–83
21 Tanner, W. and Lehle, L. (1987) Protein glycosylation in yeast. Biochim. Biophys. Acta **906**, 81–99
22 Lehle, L. (1992) Protein glycosylation in yeast. Antonie Van Leeuwenhoek **61**, 133–134
23 Dwek, R. A. (1995) Glycobiology: 'towards understanding the function of sugars'. Biochem. Soc. Trans. **23**, 1–25
24 Gavel, Y. and von Heijne, G. (1990) Sequence differences between glycosylated and non-glycosylated Asn-X- Thr/Ser acceptor sites: implications for protein engineering. Protein Eng. **3**, 433–442

Received 26 February 2001/27 April 2001; accepted 18 June 2001