



HAL
open science

Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation

Stine K Knudsen, Allan Stensballe, Magnus Franzmann, Uffe B Westergaard,
Daniel E Otzen

► **To cite this version:**

Stine K Knudsen, Allan Stensballe, Magnus Franzmann, Uffe B Westergaard, Daniel E Otzen. Effect of glycosylation on the extracellular domain of the Ag43 bacterial autotransporter: enhanced stability and reduced cellular aggregation. *Biochemical Journal*, 2008, 412 (3), pp.563-577. <10.1042/BJ20071497>. <hal-00478910>

HAL Id: hal-00478910

<https://hal.science/hal-00478910v1>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Effect of glycosylation on the extracellular domain of the Ag43
bacterial autotransporter:
Enhanced stability and reduced cellular aggregation

Stine K. Knudsen^{1§}, Allan Stensballe^{1§}, Magnus Franzmann¹, Uffe B. Westergaard¹ and
Daniel E. Otzen^{1,2*}

¹Department of Life Sciences, Aalborg University, Sohngaardsholmsvej 49, DK – 9000
Aalborg, DENMARK

²Interdisciplinary Nanoscience Center (iNANO), Department of Molecular Biology, Aarhus
University, Gustav Wieds Vej 10C, DK - 8000 Aarhus C, DENMARK.

Running Title: Glycosylation effects on the Ag43 passenger domain

*To whom correspondence should be addressed at Department of Molecular Biology, Aarhus
University: Telephone: +45 89 42 50 46; Fax: +45 86 12 31 78; e-mail: dao@inano.dk

§ These authors contributed equally.

Stage 2(a) POST-PRINT

SUMMARY

Autotransporters constitute the biggest group of secreted proteins in Gram-negative bacteria and contain a membrane-bound β -domain and an α -domain secreted to the extracellular environment via an unusually long N-terminal sequence. Several α -domains are known to be glycosylated by cytosolic glycosyl transferases, promoting bacterial attachment to mammalian cells. Here we describe the effect of glycosylation on the extracellular α -domain of the *E. coli* autotransporter Ag43 α , which induces frizzy colony morphology and cell settling. We identify 16 glycosylation sites and suggest two possible glycosylation motifs for Ser and Thr. Glycosylation stabilizes against thermal and chemical denaturation and increases refolding kinetics. Unexpectedly, glycosylation also reduces the stabilizing effect of Ca²⁺ ions, removes the ability of Ca²⁺ to promote cell adhesion, reduces the ability of Ag43 α -containing cells to form bacterial amyloid and increases the susceptibility of the resulting amyloid to proteolysis. In addition, our data indicate that Ag43 α folds without a stable intermediate, unlike pertactin, indicating that autotransporters may arrive at the native state by a variety of different mechanisms despite a common overall structure. A small but significant fraction of Ag43 α can survive intact in the periplasm if expressed without the β -domain, suggesting that it is able to adopt a protease-resistant structure prior to translocation over the membrane. Our work demonstrates that glycosylation may play significant roles in structural and functional properties of bacterial autotransporters at many different levels.

Keywords: Autotransporters, glycosylation motifs, stability, folding, amyloid.

INTRODUCTION

Autotransporters are found in Gram-negative bacteria where they constitute the biggest group of secreted proteins, with over 500 members [1, 2]. Mature autotransporters are composed of two domains: the β -domain, located in the outer membrane, and the passenger- or α -domain, which is secreted to the extracellular environment with the assistance of the C-terminal β -domain, which is proposed to act as a pore [1]. The β -domain forms a β -barrel [3], while >97% of all passenger domains are predicted to form a right-handed parallel β -helix structure [2], as seen for 3 of the 4 available crystal structures (PDB codes 1dab, 1wxr, 1p9h and 1s7m). Autotransporters are unusual in that all information for translocation, insertion of the β -domain and translocation of the passenger domain are contained within the protein sequence [2, 3].

Once exposed to the extracellular environment, the passenger domain is thought to fold cooperatively in a manner that will drive translocation. A recent study on the thermal and chemical unfolding of the pertactin passenger domain from *Bacillus pertussis* revealed the existence of a stable intermediate with a protease-protected C-terminal region [2]. Given that the C-terminal part of the passenger domain is translocated first, it is possible that it acts as a template for further translocation-coupled folding. *In vitro* membrane-insertion and folding of isolated autotransporter β -domains [4, 5] and the pertactin passenger domain [2] is extremely slow, lasting many hours. This is incompatible with biological time scales and suggests that various conditions in the cell, including the vectorial folding process, may speed up the process significantly.

The present study focuses on the folding, stability and post-translational modification of the 499-residue passenger domain of antigen 43 (Ag43 α), a product of the *flu* gene in *Escherichia coli*. Expression of Ag43 α leads to a frizzy colony morphology [6] and bacterial cell settling in standing liquid cultures, which has been suggested to be derived from a self-recognizing dimerization property located in the N-terminal domain [7]. Ag43 α is cleaved from the 488-residue β -domain but stays non-covalently attached and is most efficiently removed by brief heating to 60°C [8]. While no structure

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

of Ag43 α is available, it is predicted to adopt a parallel β -helix conformation [2]. Klemm and co-workers have classified Ag43 as a self-associating autotransporter together with the two other *E. coli* autotransporters TibA and AIDA [9]. All 3 proteins promote bacterial aggregation. Unusual for prokaryotic proteins [10], they are all glycosylated by heptose units [11-14], which may promote adherence to mammalian cells. Glycosylation is carried out by heptosyl transferases which can cross-glycosylate each other's substrates. As the heptosyl transferases do not contain signal sequences designating them for export, glycosylation is likely to take place in the cytosol, which means that Ag43 α is already modified by the time it has to translocate and fold. While glycosylation of Ag43 does not affect its flocculation properties [13], it may affect its folding and stability. Protein glycosylation can reduce irreversible unfolding and aggregation by reducing dynamic fluctuations and chemical modifications [15] and increasing solubility [16]. In addition, they may increase thermal stability due to specific hydrogen bonding or steric interactions with the native state [17]. However, there have been very few reports about the structural and dynamic consequences of glycosylation of bacterial proteins.

Here we identify 18 glycosylated Ser and Thr residues in Ag43 α and use this to suggest two possible glycosylation motifs. Glycosylation has a number of effects on Ag43 α properties: As might be expected for a protective group, glycosyl groups make Ag43 α less prone to native-state proteolysis of the C-terminal region, which occurs around residue 367, stabilizes the protein against thermal and chemical denaturation and increases its refolding rate. Unexpectedly, glycosylation also reduces the stabilizing effect of Ca²⁺ ions, removes the ability of Ca²⁺ to promote cell adhesion and inhibits the ability of Ag43 α to promote formation of bacterial amyloid. In addition, our data indicate that Ag43 α folds and unfolds without a stable intermediate, unlike pertactin, indicating that autotransporters may fold by a variety of different mechanisms despite a common structural motif. Finally, a small but significant fraction of Ag43 α can survive intact in the periplasm if expressed without the β -domain, suggesting that it is able to adopt a protease-resistant structure prior to membrane translocation.

MATERIALS AND METHODS

Strains: The two *E. coli* strains expressing Ag43 were generously provided by Professor Per Klemm, Technical University of Denmark. The strains are based on the *E. coli* K-12 MG1655 strain where the *flu* gene is deleted [13]. OS64 contains two plasmids, namely chloramphenicol-resistant pKKJ128 (containing the *flu* gene from *E. coli* MG1655 in pACYC184 [18]) and ampicillin-resistant pOS28 (containing the *E. coli* heptosyl-transferase *aaH* gene in pBADmyc-HisA [13]). Upon induction with arabinose, this leads to glycosylated Ag43. PKL1061 contains the ampicillin-resistant pKK143 plasmid with the *flu* gene from *E. coli* MG1655 in pBADmyc-HisA [19]. Induction with arabinose leads to expression of non-glycosylated Ag43. The control strain OS60 contains the empty vectors pACYC and pBAD.

Expression and purification of Ag43 α : Cells were grown at 37°C in LB-media with appropriate antibiotics, induced with 0.2% L-arabinose at OD₆₀₀ \approx 0.5 and harvested at OD₆₀₀ \approx 1.7. Harvested cells were resuspended in 25 ml PBS per litre cell culture, with the addition of 50-fold diluted Protease Inhibitor Cocktail containing pepstatin A, bestatin, EDTA, E-64 and 4-(2-aminoethyl)benzenesulfonyl fluoride (Sigma-Aldrich, St. Louis, MO). To avoid proteolytic degradation of Ag43 α , the following steps were performed as quickly as possible. The passenger domain was released from the cell surface by heating to 60°C for 3 min, and subsequently cooled on ice for 5 min. The cell suspension was then centrifuged (4,000 g, 20 min, 4°C) and the supernatant containing the released Ag43 α was collected and immediately precipitated with ice-cold acetone (final concentration 75% v/v) and stored at -20°C prior to subsequent purification steps. Ag43 α was purified in two purification steps. The pellet was resuspended in 20 mM Tris pH 8.3, sonicated and filtered (0.45 μ m) before loading on a quarternary ammonium Source 15-Q column at a flow rate of 1.0 ml/min. Protein was eluted with a salt gradient of 0-500 mM NaCl over 300 mL at a flow rate of 3.0 mL/min. Fractions containing Ag43 α were identified by SDS-PAGE, pooled and loaded on a gel

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

filtration (HiLoad™ 26/60 Superdex™ 75 prep grade) column in 50 mM Tris pH 8.3 and 150 mM NaCl. Protein was loaded at 1.0 mL/min and eluted at 2.5 mL/min. Fractions containing Ag43 α were identified by SDS-PAGE and concentrated using an Amicon Centriprep Centrifugal Filter device with a 30 kDa cut-off filter (Millipore, Bedford, MA). Ag43 α concentration was determined by absorbance at 280 nm using an extinction coefficient of 27960 M⁻¹ cm⁻¹, frozen in liquid nitrogen and stored at -80°C for later use.

MS analysis of glycosylated and non-glycosylated Ag43 α . For analysis of glycosylation and limited proteolysis experiments, Ag43 α was isolated by SDS PAGE and stained with Coomassie Blue. Prior to analysis by MALDI MS and LC-MSMS analysis, the protein bands containing Ag43 α were digested *in situ* with trypsin essentially according to [20]. The excised gelbands of Ag43 α were washed with acetonitrile (50% v/v) in 0.1M ammonium bicarbonate. The gel pieces were dehydrated with acetonitrile and rehydrated in a minimal volume of digest buffer containing 50 mM ammonium bicarbonate and 12.5 ng/ μ L trypsin (Promega, Madison, WI) followed by overnight incubation at 37 °C. Formic acid (5% v/v) and absolute acetonitrile were added to gel pieces to extract peptides. Extraction was repeated two times. Peptide samples were dried by a centrifugal concentrator under vacuum to near dryness and rehydrated in 5% formic acid (v/v) prior to use. Beta-elimination/Michael addition of tryptic peptides of Ag43 with dithiothreitol was performed according to [21]. An aliquot of tryptic Ag43 peptides (1 μ L; 5% of gel band) was suspended in 100 μ L β -elimination/Michael addition solution (pH 12.5–13) containing 1.5% triethylamine (TEA), 0.15% NaOH, and 20 mM of DTT and incubated for 1.5 h at 52°C and acidified to below pH 5 with addition of TFA to 2%.

For MALDI TOF MS analysis, an aliquot of tryptic digest was mixed with MALDI matrix (2,5-dihydroxy-benzoic acid in 50% acetonitrile, 1% phosphoric acid) and analysed on MALDI-TOF MS (Bruker Daltronics Reflex III). External calibration of the mass spectrometer was performed using calibration standard immediately before data acquisition. Automated LC-ESI MS/MS was performed using a hybrid QTOF mass spectrometer and an iontrap mass spectrometer (Bruker Daltronics,

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Bremen, DE) and an Ultimate nano-HPLC system (LC Packings, The Netherlands) mounted with a vented-column setup [22]. Reversed phase columns (pre-column 2 cm, 75 μm id; separation column 12 cm, 50 μm internal diameter) were packed in-house with ReproSil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH, Ammer-Buch-Entringen, Germany) using a high-pressure vessel. Aliquots of the tryptic peptides were injected onto the pre-column with a flow rate of 3 $\mu\text{L}/\text{min}$ and subsequently eluted at 175 nL/min using a 35 min gradient of 5-40% acetonitrile in 0.6% acetic acid. The QTOF mass spectrometer was operated in data dependent mode to automatically switch between MS and MS/MS acquisition by collision induced dissociation (CID) using argon, selecting the three most abundant precursor ions. The tandem MS data was deconvoluted and de-isotoped and exported in a Mascot generic format. The ion-trap mass spectrometer was operated according to [23]. The iontrap mass spectrometer was set to acquire complementary tandem mass spectra using electron transfer dissociation (ETD) and CID.

MALDI MS data analysis was manually analyzed by the freeware m/z program (www.proteometrics.com). Tandem MS data was initially analyzed using an in house Mascot database search engine using a FASTA sequence database containing Ag43 α . Tandem MS search parameters were 0.05 Da accuracy; methionine oxidation and heptosylation allowed as variable modifications and semitryptic enzyme specificity [24]. Manual verification of heptosylation sites was aided using GPMAW v. 6.21 (Lighthouse Data, Odense, Denmark). Sequence logos were prepared using Weblogo 2.8.2 (<http://weblogo.berkeley.edu/>).

Spectroscopy: Far-UV CD spectra were recorded in a 1mm quartz cuvette on a Jasco J-810 spectropolarimeter (Jasco Spectroscopic Co. Ltd., Hachioji City, Japan) at 25°C. Wavelength scans were recorded using 5 μM Ag43 α in the wavelength range of 195-250nm with a band width of 2nm and a scanning speed of 50nm/min. Three accumulations were averaged to yield the final spectrum. Background contributions from the buffer were subtracted.

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Fluorescence spectra were recorded on a Cary eclipse fluorescence spectrophotometer (Varian Inc., Palo Alto, CA) heated by a Peltier heating element. Samples containing 1 μM glycosylated Ag43 α or 2 μM non-glycosylated Ag43 α in 20 mM Tris, pH 8.3 were heated from 20°C to 90°C and the fluorescence intensity was measured by excitation at 295 nm and emission at 358 nm, both with slit sizes of 10 nm. Data were collected with a data interval of 0.5 sec and a scanning rate of 2°C/min. A quartz cuvette with a 10 mm path was used (Hellma GmbH, Müllheim, Germany). These settings were also used to follow refolding kinetics and unfolding kinetics at different temperatures, in which the cuvette containing buffer was pre-equilibrated at the temperature in question for at least 15 minutes before Ag43 ($\leq 10\%$ of total volume) was added. For refolding data, the sample was only exposed to excitation light at the actual measurement, and was kept shielded from light between measurements (which extended over many hours) to avoid photobleaching. Thermal scan data were analyzed according to a two-state unfolding model [25] to obtain the midpoint for thermal denaturation t_m (in °C).

t_m values obtained from thermal denaturation in the presence of calcium were plotted against $[\text{Ca}^{2+}]$ on a logarithmic scale and fitted to a simple binding curve, where free Ag43 α and calcium are in equilibrium with the complex Ag43 α :Ca $^{2+}$ with an apparent equilibrium constant K_d . Making the simplifying assumption that free Ag43 α unfolds at a melting temperature t_m^o and the complex Ag43 α :Ca $^{2+}$ at t_m^{Ca} , we can fit the data to the following equation:

$$t_m = \frac{K_d * t_m^o + [\text{Ca}^{2+}] * t_m^{\text{Ca}}}{K_d + [\text{Ca}^{2+}]} \quad (1)$$

We emphasize that this is a very simplistic approach (since increasing amounts of Ca $^{2+}$ should stabilize the Ag43 α :Ca $^{2+}$ complex and thus increase t_m^{Ca}) and only use the equation for comparative purposes.

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Urea denaturation: 1-2 μM Ag43 α was incubated in 20 mM Tris pH 8.3 with 0-9.5 M urea, incubated 24 hrs at room temperature and fluorescence recorded on an LS-55 luminescence fluorimeter (Perkin Elmer, Wellesley, MA) with excitation at 295 nm and emission at 310-400 nm. Slit sizes of 10 nm, scanning rates of 300 nm/min with a 10 mm quartz cuvette were used. Each spectrum was the average of three measurements. Spectra containing urea and buffer were subtracted as blanks. Data were analyzed as described [25] to obtain [urea^{50%}], the urea concentration where 50% of the protein is denatured. As fluorescence signal we use the ratio of the emission at 356 and 346 nm, which gave the clearest signal change upon denaturation. Non-linear least-squares regression analysis was carried out with the program Kaleidagraph, version 3.5 (Synergy Software, Reading, PA).

Refolding kinetics were followed by first unfolding Ag43 α (10 μM glycosylated or 20 μM non-glycosylated) for 24 hrs in 5M urea, followed by 10x dilution to yield 0.5 M urea. Fluorescence spectra were recorded ca. 20 times over 24 hrs. To avoid photobleaching, the sample was only excited during the actual recording. Refolding kinetics could be fitted to a single exponential decay.

Trypsin experiments: To determine whether purified Ag43 α was correctly folded, 10 μM (500 $\mu\text{g/ml}$) Ag43 α was mixed with 100 $\mu\text{g/ml}$ trypsin in 20 mM Tris pH 8.3, incubated for 30 minutes at room temperature and then loaded onto an SDS-PAGE gel. To follow refolding kinetics, 10 μM Ag43 α pre-incubated at different temperatures (45-65°C) for 15 minutes was transferred to 25°C. At appropriate times, samples were digested with 100 $\mu\text{g/ml}$ trypsin for 30 minutes before SDS-PAGE. The reaction was stopped by addition of 2x SDS-PAGE loading buffer. Gels were scanned using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Pharmacia, Piscataway, NJ) and band intensities of refolded Ag43 α were quantified by ImageQuant 5.1 software.

Proteolysis and inhibition studies: To follow trimming of Ag43 α over longer time periods, 20 μM Ag43 α was incubated in 20 mM Tris pH 8.3 and 0.1% sodium azide in sterile eppendorf tubes for 0-168 hours at room temperature, after which they were mixed with SDS-PAGE loading buffer and run on a gel. Inhibition of Ag43 α trimming was studied by incubating Ag43 α as above for 168 hrs in the

presence of different protease inhibitors from the Protease Inhibitor Panel at concentrations specified by the manufacturer (Sigma-Aldrich).

N-terminal sequencing by Edman degradation: The stacking gel was allowed to polymerize one day prior to electrophoresis and samples were heated in SDS-PAGE loading buffer for 3 min at 80°C only. After electrophoresis, trimmed and non-treated Ag43 α bands were transferred to a polyvinylidene difluoride membrane (Immobilon-P, Millipore, Bedford, MA) in 10 mM CAPS, 10% (v/v) methanol, pH 11 as described [26]. Samples were analyzed by automated Edman degradation using a Procise amino acid sequencer (Applied Biosystems, Foster City, CA) with on-line phenylthiohydantoin analysis by HPLC.

Auto-aggregation assay: This was performed essentially as described [7]. Briefly, *E. coli* strains expressing glycosylated Ag43 (strain OS64) and non-glycosylated Ag43 (strain PKL1061) together with the negative control (strain OS60), were grown to OD₆₀₀~1.5. Subsequently, to each culture was added either 10 mM CaCl₂, 10 mM EDTA or none of the two and the concentration was standardized to OD₆₀₀=1.5 by dilution with LB media. The solutions were mixed well and left to settle in 500 ml conical flasks. At regular time intervals, a 1 mL sample was taken approximately 0.5 cm from top of the culture surface and OD₆₀₀ was measured. Each sample was measured in triplicate using 3 separate flasks.

Cloning and expression of Ag43 α without β -domain:

The α -domain of Ag43 was amplified from the pKKJ128 plasmid. The two 5'-primers (GTAGCTCATATGAAACGACATCTGAATACCTGCTACAGGCTGG and CAAGCTCATATGGCTGACATCGTTGTGCACCCGGGAGAAAC) together with the 3'-primer (AATACTCGAGGTCAATGGCACCGTTCAGCACAGTG) were used to amplify the coding sequence of the Ag43 α -domain (residues 53-551 of SwissProt entry P39180) with and without the native signal peptide (residues 1-52), respectively. The primers were designed to insert NdeI and XhoI restriction sites and a C-terminal hexa-histidine tag for purification and immunodetection

purposes. The PCR products were ligated into an NdeI/XhoI-predigested pET-30b(+) expression vector. DNA fragments were purified from gel prior to ligation and the final Ag43 α -containing plasmids were verified by sequencing. Plasmid-transformed cells, grown at 37°C, 150 rpm, were induced with 0.5 mM or 0.050 mM IPTG upon reaching OD₆₀₀ ~0.7-0.9. Cells were harvested 2-3 hrs later and the removed media was analyzed for the presence of Ag43 α in the extra-cellular environment. Harvested cells were resuspended in PBS and the presence of Ag43 α on the cell surface was investigated by heat removal of the α -domain by incubation at 60°C for 3 min followed by centrifugation (4.000 g, 20 min, 4°C). The supernatant was checked for the presence of Ag43 α by SDS-PAGE and western blotting, while an identical cell pellet, without prior heat treatment, was analyzed for the presence of Ag43 α in the periplasm or in the cytoplasm. The periplasmic space was isolated by a digestion of the outer cell membrane resulting in formation of spheroplasts by incubation with sucrose, lysozyme (final concentration of 32 μ g/mL) and EDTA as described [27]. Spheroplasts were isolated by centrifugation (4500 rpm, 10 min, 4°C) and the supernatant was analyzed for the presence of Ag43 α in the periplasm. Pellet containing spheroplasts was resuspended in PBS and spheroplasts were sonicated, thereby releasing the content of the cytoplasm. A last fractionation between proteins free in the cytoplasm and proteins captured in inclusion bodies was done by centrifugation (13000 rpm, 15 min, 4°C). Supernatant containing free cytoplasmic proteins together with resuspended pellet containing proteins as inclusion bodies were analyzed along with the other fractions by SDS-PAGE and western blotting.

Enzyme activity assays: To check the specificity of our fractionation procedure, the fractions containing medium, heat-treated supernatant, periplasm and cytosol was tested for the activity of both alkaline phosphatase and glucose-6-phosphate dehydrogenase. Alkaline phosphatase activity was determined by incubating 50 μ l of the enzyme fractions with 1200 μ l 50mM Tris-HCl pH 8.0, 1mM MgCl₂, 0.02% p-nitrophenyl-phosphate at 37°C. The absorbance at 410nm was measured at different time points up to 10 hours. The glucose-6-phosphate dehydrogenase activity assay was performed in

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

50mM Hepes pH 7.4, 2mM D-glucose-6-phosphate, 0.67mM β -nicotinamide adenine dinucleotide phosphate (β -NADP), 10mM $MgCl_2$ by adding 30 μ l enzyme fraction to 970 μ l and monitoring the increase in absorption at 340nm.

Western Blot Procedure: For Ag43 α with and without signal sequence, equivalent amounts of sample in terms of the number of cells from which the fractions were derived, were loaded in each SDS-PAGE lane to allow direct comparison. Proteins were transferred from the SDS-PAGE gel to a PVDF membrane by 0.8mA pr. 1 cm² membrane for 2 hrs. After transfer, the membrane was blocked in 2% Tween20, 20mM TrisHCl pH 7.4, 150mM NaCl at 4°C overnight. The membrane was washed 3x10 min in 0.05% Tween20, 20mM TrisHCl pH 7.4, 150mM NaCl followed by incubation with an anti-his tag antibody (1:2000 diluted Tetra-His IgG from, Qiagen, Hilden, Germany). After 2 hrs the washing procedure was repeated followed by incubation for 1 hrs with goat anti-mouse antibody conjugated with 1:1000 diluted Alexa Fluor 647 (Molecular Probes, Carlsberg, CA). After another washing procedure, bound antibody was visualized using the Typhoon 8600 phosphoimager.

Analysis of aggregative structures: To test for binding of Thioflavin T, the *E. coli* strains were grown overnight at 37°C, 150 rpm to OD₆₀₀=1.7 followed by mixing of 1 mL culture with Thioflavin T to yield a final Thioflavin T concentration of 1.5 μ M. Upon binding of Thioflavin T to amyloid structures, the excitation maximum of the dye changes from 385 nm to 450 nm and emission from 445 nm to 482 nm. The culture containing Thioflavin T was distributed on a gelatine coated cover slide and Thioflavin T coloring was detected by confocal laser scanning microscopy using a Zeiss LSM510 equipped with a meta-filter and a 630 X oil objective.

Binding of Thioflavin T and ANS to purified Ag43 α was performed and monitored as follows: 40 μ M Ag43 α was incubated with either 40 μ M Thioflavin T (excitation 450 nm, emission 470-490 nm) or 40 μ M ANS (excitation 350 nm, emission 420-550 nm) in 20 mM Tris pH 8.3.

Homology Modeling of AG43: Homology models of Ag43 were produced using the crystal structure of Heme Binding protein (HBP) from pathogenic *Escherichia coli* (PDB ID: 1WXR) as a template

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

using MODELLER 9v2 [28]. The models were subsequently optimized with respect to the Discrete Optimized Potential Energy (DOPE) score using the looprefine module of MODELLER. The sequence alignment of Ag43 and HBP used for modeling was produced using Multalin [29]. The two primary sequences have 23 % pairwise identity. The best Ag43 homology model obtained was energy minimized using the program Gromacs 3.3.1 [30] with the GROMOS96 force field [31] and done with the steepest descents method in steps of 1 femtoseconds with a cutoff for Vander Waals interactions of 1.4. The atomic accessible surface was calculated using the program NACCESS [32] and heptose molecules were added manually to the surface accessible Ser and Thr using the program PYMOL [33].

RESULTS

Purification of Ag43 α by heat treatment to unfold and release the protein

In order to perform biophysical analyses of Ag43 α , we need mg quantities of >95% pure and folded protein. We achieved this by over-expressing glycosylated and non-glycosylated Ag43 (henceforth named Ag43^{gly} and Ag43^{nongly}) from expression plasmids in an *E. coli* strain lacking endogenous Ag43. The endogenous heptosyl transferase for Ag43 remains unknown; instead we carry out glycosylation by co-expressing the *E. coli* heptosyl-transferase *aaH* gene that normally glycosylates AIDA but has also been shown to use Ag43 as substrate [13]. The relatively high expression level of the *aaH* gene is reported not to have a significant effect on the glycosylation [13].

After post-translational cleavage of the border between the extracellular α -domain and the membrane-bound β -domain, Ag43 α remains attached in a non-covalent manner to its cognate β -domain on the outer membrane surface [8]. We found that the most efficient way to release Ag43 α from the membrane was to heat the harvested and resuspended cells at 60°C for 3 minutes, followed by centrifugation to remove cells and retain Ag43 in the supernatant (Fig. 1A). Our subsequent experiments revealed that purified Ag43 α undergoes thermal denaturation around 50-55°C but can refold from the heat-denatured state over a period of several hours (see below). Thus, Ag43 α has to unfold in order to be released from the cell surface. It was important to include a protease inhibitor cocktail in order to avoid proteolytic degradation of Ag43 α , since the heat-denatured state is particularly vulnerable towards this. In addition, we carry out precipitation in acetone at an early stage in purification to inhibit proteases, and allow the protein to refold during the subsequent purification. Already after heat-release, the protein is relatively pure due to the small number of other proteins released by this treatment. Subsequent chromatographic steps using anion exchange and gel filtration yield a major dominating band by SDS-PAGE for Ag43^{nongly} (Fig. 1A) in addition to at least two smaller (ancillary) bands right below this band. The smaller bands are even more dominant for

Ag43^{gly} where they are clearly seen as two separate bands (data not shown). These bands represent Ag43 α proteins with heterogeneous C-termini (see below).

Verification of folded structure of heat-released Ag43 α by trypsin treatment

To investigate whether the purified Ag43 α ^{gly} and Ag43 α ^{nongly} were in a folded conformation after the heat-release step, we exposed them to limited proteolysis by trypsin. Due to its compactly folded conformation, the native state of a protein is generally much more resistant to proteases than the unfolded state [4]. When native Ag43 α ^{nongly} is exposed to trypsin, the major band around 54 kDa and the ancillary bands merge to a smaller band around 45kDa, corresponding to the lowest of the bands seen before trypsin treatment (Fig. 1A). Thus, trypsin appears to “trim” a rather heterogeneous collection of states to a more well-defined protein population by removing the non-structured C-terminal tail. In contrast to the limited proteolysis of the native state, we see that denaturation of Ag43 α prior to proteolysis by heating at 90°C leads to complete degradation. A similar trimming, although to a slightly smaller extent, is observed for Ag43 α ^{gly} (data not shown). Incubation of the two proteins over a week at room temperature under bacteriocidal conditions (0.1% sodium azide) increased the population of trimmed protein significantly. Proteolysis was inhibited by members of all four protease inhibitor classes (data not shown), including reagents targeting Cys residues which are absent from Ag43 α . Thus we conclude that trimming derives from small amounts of contaminating proteases rather than endogenous autoproteolytic activity seen for other autotransporter α -domains [34].

N-terminal sequencing of the trimmed band revealed the same N-terminal 6 residues for the trimmed and intact α -domain, namely ADIVVH (residues 51-56, just after the 50-residue signal sequence) in both cases. Thus, proteolysis must occur close to the C terminus. To identify the site of cleavage, we performed an exhaustive trypsin digestion coupled to mass spectrometric analysis. We isolated Ag43 α ^{gly} and Ag43 α ^{nongly} by SDS PAGE, resulting in a 55kDa band and a 45 kDa trimmed band, and

subsequently *in situ* exhaustively digested these protein bands with trypsin. The resulting peptide mixture of the two bands was analyzed by high precision LC-MS/MS to generate extensive peptide sequence information. The combined sequence coverage, confirmed by three individual experiments, is shown in Fig. 1B. The amino acid sequence coverage obtained by LC-MS/MS analysis of the tryptic digest of the 55kDa band (Fig. 1B, upper panel) is 91.4% (indicated in bold). We observed multiple overlapping peptide sequences of the C-terminal region, confirming their presence in the upper protein band. In contrast, LC-MS/MS analysis of the tryptic digest of the 45kDa band (Fig. 1B, lower panel) only covered 66.3%. The lack of peptides originating from 132 amino acid (13.1 kDa) residues of the C-terminal region confirms the previous assumption that a C-terminal trimming has occurred. By mining the mass spectrometric data for tryptic and non-tryptic peptides of the C-terminal region, our data identify Lys367 as the ultimate residue in the trimmed Ag43 α^{gly} and Ag43 α^{nongly} , highlighting the C-terminal 132 residues as essentially unstructured.

Identification of 16 sites of glycosylation on Ag43 α

Previous work by Klemm and co-workers, using a wildtype strain without plasmid-encoded heptosyl transferase, demonstrated that Ag43 α is glycosylated with heptose residues at several positions by an as yet unidentified glycosyl transferase [13]. Although several tryptic peptides were shown to be glycosylated, the residues involved in glycosylation were not identified in that work. In order to provide a more detailed view of this glycosylation pattern and possibly identify glycosylation motifs, we therefore undertook to identify the specific residues modified by glycosylation.

As an initial step, we determined the overall glycosylation of Ag43 α^{gly} by MALDI MS on intact Ag43 α^{gly} . In addition to non-glycosylated Ag43 α , we identified equally populated species with 2,3,4,5 and 6 heptose groups whereas only a small peak representing a possible singly glycosylated variant of the trimmed version of Ag43 α^{nongly} could be detected (Fig. 1C). The mobility in SDS-PAGE which we observe to be only slightly affected by the heptosylation is in good agreement with

the substoichiometric level of modifications. We estimate based on the current MS data that less than half of the total protein is glycosylated in our preparation with even distribution of up to 6 occupied sites.

Ag43 α^{gly} was excised from the gel after isolation by SDS-PAGE and digested *in situ* by trypsin followed by analysis by peptide mass mapping by MALDI TOF MS and nanoflow LC-tandem MS. Identification of glycosylated peptides were performed manually by analysis of peptide mass and automatically using the Mascot sequence database-searching software.

Peptide sequencing by tandem mass spectrometry of the heptosylated peptides obtained in this analysis identified several individual heptosylated residues (summarized in Table 1). This is confirmed (Fig. 2A) by identifying a doubly charged non-modified peptide (m/z 1072.0441) and the doubly charged heptosylated peptide (m/z 1168.0731) differing by 192.058 Da (theoretical mass 192.063 Da). Assignments of heptosylated sites were in all cases verified by manual inspection of the tandem mass spectra (Fig. 2B-C). However, the heptose moiety was observed to be highly labile under the conditions for collision-activated dissociation, whereby multiple tandem mass spectra of in particular multiheptosylated peptides lacked conclusive information to assign acceptor amino acid residues. Therefore we investigated two means of unambiguously confirming the sites of modification. Firstly, we performed peptide sequencing by electron capture dissociation which enables peptide sequencing without destruction of labile modifications [35]. The tandem mass spectrum of the heptosylated peptide 71-VNPGGSVSDTVISAGGGQSLQGR-93 (Fig. 2D) shows a near complete cleavage of the peptide backbone with no loss of the heptose structure, thereby confirming the modified S83.

Secondly, we investigated β -elimination/Michael addition using dithiothreitol to convert the labile modification to moieties which were stable under tandem mass spectrometric analysis. As an example, Fig. 3 shows the tandem MS peptide of the unmodified doubly charged peptide 61-TTVTSGGLQR-70 (Panel A) and the tandem MS spectrum of DTT-treated formerly heptosylated peptide (Panel B).

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

As described in the figure legend, the mass of the fifth residue increases from 87 Da (corresponding to unmodified S65) in the unmodified form to 223 Da in the modified form (corresponding to Ser with a DTT group minus an eliminated water molecule).

In this way the site of modification can be unambiguously assigned for most glycosylated peptides, although a few ambiguities remain (Table 1). In all cases, we only find one heptose attached to each glycosylated residue. Interestingly, one of the glycosylation sites is found at S370, which is close to the C-terminus of trimmed Ag43 α (K367). The presence of a glycosylation site in this area could explain why Ag43 α^{gly} is less easily trimmed down than Ag43 α^{nongly} (Fig. 1A). A comparison with the glycosylation regions identified by Klemm and co-workers [13] reveals that we have identified regions corresponding to 9 of the 11 glycosylated peptides reported in that study (we do not identify peptide 121-156 and 296-319) and have found 8 additional non-overlapping peptides. Based on all these peptides, we identify 16 glycosylation sites (T35, T61, S65, S83, T97, T98, T177, T180, T235/T236, S265, S344, S370, S280, T435, S439, S451). In addition, we have found 6 other sites that still need site specific confirmation.

In order to further investigate the sites of modification and potentially predict the location of ambiguous sites, we used sequence logos to find putative motifs surrounding the identified sites [36, 37]. All unambiguously assigned glycosylation sites were aligned in the -6 to +6 positions (with the modified Thr or Ser in position 0) and the frequency of residues within specific positions was used for a sequence logo graphical representation of the amino acid multiple sequence alignments using Weblogo 2.8.2. This representation (Fig. 4A) suggests that O-glycosylation at Thr shows a preference for Asn at position +2 and Val at position +5, whereas O-glycosylation at Ser has a preference for Val (followed by Ala) at position -2 and a weaker preference for Gly (followed by Thr and Ala) at position +2. These motifs are not seen in the non-glycosylated sequences containing Ser and Thr (Fig. 4B). The recent study of AIDA glycosylation by Mourez and co-workers [14] provide 13 additional Ser-glycosylated sequences (as well as 1 Thr-glycosylated sequence). Gratifyingly, these sequences

show the same overall trend as for Ag43 α in positions -2 and +2 (data not shown). When combined with our data, we see a reinforced preference for Val at -2, Gly at +2 and also Gly at +1 (Fig. 4C). AIDA is modified by its own heptosyl transferase aaH and thus presents an authentic glycosylation substrate, while Ag43 α is a foreign substrate. The close resemblance between the motifs provided by the Mourez AIDA peptides and our own Ag43 α peptides underline the motifs' generality. Ag43 α is glycosylated by an endogenous Ag43 α in the endogenous strain UTI536 [13], but the low expression level of Ag43 α in this strain made it very difficult to purify it in amounts suitable for a detailed MS analysis. However, from the SDS-PAGE bands of Ag43 α released upon heat treatment we identified 6 peptides identical to those seen for Ag43 α from OS64 (data not shown). When combined with the fact that over-expression of aaH has been shown not to affect glycosylation [13], this is a good indication that the level of glycosylation is comparable in Ag43 α from OS64 and UTI536.

Ca²⁺ stabilizes Ag43 and promotes autoaggregation

Having purified folded Ag43 α^{gly} and Ag43 α^{nongly} and identified regions of glycosylation, we embarked on a comparative analysis of the stabilities of the two autotransporters. First we compared their overall structure by far-UV CD spectroscopy (Fig. 5A). Although both Ag43 α^{gly} and Ag43 α^{nongly} showed the broad minimum around 216-218 nm characteristic of all- β proteins [38], there was a significant difference between the two proteins below 210 nm, where the signal of Ag43 α^{nongly} remains more negative than that of Ag43 α^{gly} . This difference could be caused by the presence of glycosyl groups which can make contributions in the far-UV CD spectrum [39].

Ag43 α contains 4 Trp residues, and there is a significant difference in the Trp fluorescence spectrum of folded and heat-denatured Ag43 α (Fig. 5B insert), allowing us to monitor thermal denaturation by fluorescence. For both proteins, this occurs as a cooperative transition that can be fitted to a simple two-state model, leading to denaturation temperatures (t_m) 57.0 \pm 0.3 and 51.4 \pm 0.1 °C for Ag43 α^{gly} and Ag43 α^{nongly} , respectively (Fig. 5B). Thus Ag43 α^{gly} is clearly more heat-stable than Ag43 α^{nongly} .

Ag43 shows 68.5% identity (72.8% similarity) to the *E. coli* autotransporter Cah, which has been shown to bind Ca^{2+} ions via the motif GGAG(N/D)D(R/T)LT that is suggested to form a tetralobate structure[40]. Although Ag43 does not contain the intact motif, we identified several sequences that were close to the consensus motif and contained 5 of the 9 conserved residues (data not shown). We therefore decided to investigate whether Ca^{2+} would play a role in stabilizing Ag43 α against thermal denaturation. There is indeed a significant increase in t_m as we increase the Ca^{2+} concentration (Fig. 6A). For both proteins, the data can be fitted to a simple equilibrium between free and Ca^{2+} -bound Ag43 α to yield an apparent binding constant K_d^{app} of 0.92 ± 0.34 mM for Ag43 α^{gly} and 0.76 ± 0.33 mM for Ag43 α^{nongly} (Fig. 6B). Although these K_d^{app} -values indicate that the two forms of Ag43 α thus bind Ca^{2+} with the same apparent stability, the effect of Ca^{2+} is more complex: Ag43 α^{gly} is more stable than Ag43 α^{nongly} in the absence of Ca^{2+} (t_m 57.0 ± 0.1 versus 51.4 ± 0.1 °C), but the roles reverse when Ca^{2+} is added (62.9 ± 0.3 versus 71.3 ± 1.5 °C at high $[\text{Ca}^{2+}]$). The same results were obtained using CaCl_2 and CaNO_3 , indicating that it is the Ca^{2+} ion rather than the anion that is the stabilizing agent (data not shown).

We decided to investigate the possible physiological consequences of this Ca^{2+} binding. The natural function of Ag43 α is to facilitate autoaggregation of *E. coli* in liquid culture [7]. Using a simple autoaggregation assay devised by Klemm and coworkers [7], we observe that Ca^{2+} significantly speeds up autoaggregation of Ag43 α^{nongly} -expressing cells compared to cells aggregating in the absence of exogenously added Ca^{2+} or in the presence of the Ca^{2+} -chelator EDTA. In contrast, Ca^{2+} and EDTA have no effect on cells expressing Ag43 α^{gly} (Fig. 6C). This correlates well with the increased stabilization of Ag43 α^{nongly} by Ca^{2+} compared to Ag43 α^{gly} .

Ag43 α^{gly} refolds faster than Ag43 α^{nongly} in one major step without a visible intermediate

After establishing by trypsin-digestion experiments that Ag43 α can be refolded from the heat-denatured state, we wanted to follow the kinetics of the process in order to identify any intermediates

accumulating during the folding process. Incubation at 55°C for 15 minutes was sufficient to unfold both forms of the protein to >99% according to Trp fluorescence and resistance to trypsin digestion (data not shown). We then followed the regain of native structure by both Trp fluorescence and the build-up of resistance to trypsin digestion by SDS-PAGE (Fig. 7A). The increase in Trp fluorescence fit well to a single exponential decay (data not shown). Similarly, by plotting band intensities versus time, we see a single exponential refolding phase (Fig. 7B). The rate constants summarized in Table 2 show a somewhat slower refolding rate according to Trp fluorescence, which may be attributed to the incubation period with trypsin during which Ag43 α can refold. However, in neither case is there evidence for the accumulation of an intermediate. Provided that an intermediate accumulates and decays slowly relatively to the measuring technique, such an accumulation would lead to double-exponential kinetics as seen for the autotransporter pertactin [2].

In order to make sure that the refolding kinetics we observe are not dependent on the type of denatured state we employ, we have also carried out refolding kinetics from the urea-denatured state, which is generally taken to represent the most completely unfolded state a protein can assume [41]. Initially we have to ascertain at what urea concentration Ag43 α^{gly} and Ag43 α^{nongly} unfold. Ag43 α^{gly} unfolds at a higher urea concentration than Ag43 α^{nongly} , namely 3.42 \pm 0.08 M as opposed to 2.31 \pm 0.05 M for Ag43 α^{nongly} (Fig. 8A). Both transitions can be fitted to a two-state unfolding mechanism without any stable intermediate, unlike pertactin [2]. We initiate refolding by diluting Ag43 α from 5 into 0.5M urea and follow refolding by Trp fluorescence (Fig. 8B). Despite some scatter, the data can be fitted to a single exponential decay which again differs from the double-exponential refolding kinetics reported for pertactin [2]. For Ag43 α^{nongly} , the refolding half life (7.42 \pm 0.97 hrs) agrees very well with spectroscopic data for the heat-denatured state (Table 2). Remarkably, refolding occurs around 3 times more rapidly (half life 2.43 \pm 0.54 hrs) for Ag43 α^{gly} . We cannot compare refolding for Ag43 α^{gly} from different heat-denatured states due to the low available amounts of this protein, but the

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

very good agreement between the different refolding methods for Ag43 α^{nongly} makes us confident that the data also are representative for Ag43 α^{gly} .

In vivo overexpression of the Ag43 α^{nongly} domain without the β -domain traps most but not all Ag43 α in inclusion bodies

The very slow refolding step of the isolated Ag43 α domain *in vitro* is not compatible with the fast doubling time of *E. coli* (less than an hour). *In vivo* conditions such as chaperones or the presence of the β -domain are likely to contribute to this process. To differentiate between the contributions of the β -domain and cellular chaperones, we prepared expression vectors overexpressing Ag43 α without the β -domain and either with (α^{sig}) or without (α°) the 52-residue signal peptide. In both cases, we observe that most (60-70%) of the protein accumulates as inclusion bodies, indicating that the β -domain is important for transport over both inner and outer membrane (Fig. 9). α^{sig} is found to a 3-4 fold larger extent than α° in the cytosol, in agreement with its lack of signal sequence. Intriguingly, small (ca. 10%) amounts of both proteins are found in the periplasm (Fig. 9), indicating that some protein can be exported to the periplasm and survive without degradation. Localization in the periplasm or the cytosol is not an artifact of the fractionation process. In our hands, 97-98% of all activity of the cytoplasmic enzyme glucose-6-phosphate is located in the cytosolic fraction and only 2-3% in the periplasm, which makes the presence of ca. 10% of α° and α^{sig} in the periplasm significant. Conversely, 85-87% of the activity of the periplasmic enzyme alkaline phosphatase is confined to the periplasm, making the large proportion of soluble α° and α^{sig} in the cytosol significant. We also observe that ca. 10% of α° and α^{sig} is released after the same heat treatment as wildtype Ag43 α . However, this is probably an artifact of the analysis, since 4-16% of our reporter enzymes are also released upon similar heat treatment.

Ag43 α^{nongly} induces bacterial amyloid

Under various conditions, many proteins have an ability to aggregate to regular arrays of β -sheet rich structures of indefinite length with a needle-like morphology [42]. Although this property has gained notoriety as a central event in the development of neurodegenerative diseases, it is also associated with beneficial functions such as the formation of bacterial biofilm that allows microbial organisms to adhere to cell surfaces [43].

In view of the central role of Ag43 α in promoting bacterial autoaggregation [7], we decided to test whether the protein was also involved in the formation of amyloid-like structures. We employed the dye Thioflavin T, which undergoes a band shift upon binding to fibrils, leading to bright green fluorescence at 485 nm when excited at 450 nm. Only PKL1061 and arabinose-free OS64 (predominantly expressing Ag43 α^{nongly}) give rise to the characteristic green colour, while arabinose- and OS60 (expressing no Ag43 α at all) remains colourless and induced OS64 (predominantly expressing Ag43 α^{gly}) shows very low levels of amyloid formation (Fig. 10).

Ag43 α is predicted to adopt a parallel β -helix or “ β -spiral” conformation [7] (and see our modeling below), and this regular β -sheet arrangement could be envisaged to bind Thioflavin T. However, purified and folded Ag43 α^{nongly} did not induce any spectral changes in Thioflavin T (data not shown), indicating that the amyloid structure reflects the particular environment of the surface of *E. coli* cells. Further albeit indirect support comes from an analysis of the Ag43 α sequence using the program TANGO [44], which only predicts a small number of regions which could be involved in β -sheet aggregation, but indicates a significant content of β -turns, in agreement with a β -helix conformation (data not shown). This indicates that Ag43 α should not have a high propensity to form higher order aggregates by itself. Naturally its mobility and associative properties will be strongly influenced by its anchoring to the bacterial surface.

We subsequently investigated in two ways how surface anchoring affects the conformational properties of Ag43 α . Firstly, direct trypsin-treatment of *E. coli* cells expressing Ag43 α^{gly} (OS64) and Ag43 α^{nongly} (PKL1061) and subsequent visualization on an SDS-PAGE gel indicated that neither of

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

the protein types were significantly susceptible to digestion (data not shown). This is in good agreement with deletion studies by Klemm and co-workers that implicate the N-terminal part is implicated in cell-cell contacts, leaving the C-terminal parts (which are trimmed when the purified protein is incubated with trypsin) close to the cell surface and thus protected from trypsin attack. Secondly, we find the sparse samples of amyloid formed by Ag43 α^{gly} in OS64 cells to be completely removed by trypsin treatment, in contrast to the Ag43 α^{nongly} in PKL1061 cells, which form larger amounts of amyloid that furthermore is resistant to trypsin digestion (data not shown). This suggests that Ag43 α^{nongly} form more sturdy or compact amyloid than Ag43 α^{gly} and emphasizes its superior ability to form organized aggregates.

DISCUSSION

Glycosylation plays a large role in the structural and functional properties of Ag43 α

Ag43 α is a member of a growing number of bacterial autotransporters which are known to be glycosylated by heptosyl transferases, with consequences for the ability to recognize mammalian cells [11]. Furthermore, bacterial autotransporters are targeted to the outer membrane and secreted by an unusually simple mechanism but fold extremely slowly *in vivo* according to the few reports available [2, 4, 5]. This makes it very relevant to analyze the effect of glycosylation on the folding properties of Ag43 α . Glycosylation of Ag43 α is very extensive, since we have been able to identify 18 different sites of glycosylation (in addition to 6 unassigned peptides), including three modifications in the N-terminal region (residues 1-70) that were not identified by Sherlock et al [13]. Based on these motifs, we have tentatively identified some sequence-preferences for O-glycosylation, which suggest different sequence requirements for glycosylation at Ser versus Thr. Together with a very recent study by Charbonneau et al. [14], these are the first examples of O-linked glycosylation motifs in bacteria. In contrast to N-linked glycosylation, which is coupled to the sequence motif Asn-X-Ser/Thr, O-linked glycosylation is generally not easy to identify from the sequence alone and appears to require a combination of tertiary structure and surface accessibility. The motifs are relatively weak and may not seem to provide obvious explanations for why only one protein out of several thousands is the target for the *aaH* heptosyl transferase. One reason for this weak signal could be over-glycosylation which will obscure the picture. Alternatively, specificity does not require a strong sequence motif but rather a special structure, namely the very uncommon repetitive β -helix proposed to occur in the Ag43 α domain. Thus it could be imagined that the transferase initially recognizes the β -helix structural motif and differentiates among the repetitive elements in the structure by virtue of these sequence motifs.

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Bacterial glycosylation is reported to have numerous functional roles [10], which include maintenance of protein conformation, resistance against proteolytic digestion, cell adhesion, enzymatic activity, surface recognition of host cells [11] as well as immune evasion. Several of these properties can be observed in this study and are to be expected from general properties of glycosylated proteins. These properties include increased stability against thermal denaturation, increased resistance to unfolding in urea as well as decreased rates of unfolding (unfolding data not shown). However, the differential effect on Ca^{2+} on the two forms of Ag43 α is unexpected and is discussed in the following section.

Differential effect of Ca^{2+} on Ag43 α^{gly} and Ag43 α^{nongly}

It seems paradoxical that Ca^{2+} within error has the same apparent binding affinity for Ag43 α^{gly} and Ag43 α^{nongly} in terms of the concentration range where it affects thermal stability, yet the overall increase in t_m is significantly larger for Ag43 α^{nongly} than for Ag43 α^{gly} . Metal chelators will generally increase stability by binding to the native state, e.g. ovotransferrin's melting temperature increases by 21°C in the presence of Fe^{3+} ions [45], and the degree of stabilization will be related to the affinity of Ca^{2+} for the native state. However, it should be borne in mind that the apparent binding affinity determined in Fig. 6B is not a true binding constant. The hydrated Ca^{2+} ion has a radius of around 9.6Å which is comparable to that of a glycan moiety (the radius of a glycan ring is around 3Å). Thus Ca^{2+} is sufficiently large to be blocked by glycan rings from binding to appropriate Asp or Glu side chains. It is possible that glycosylation simply reduces the number of binding sites of Ca^{2+} on Ag43 α^{gly} while retaining the same binding affinity at the accessible binding sites. The marked decrease in the cooperativity of unfolding (the $m_{\text{D-N}}$ value) upon glycosylation also suggests an extensive shielding of the denatured state; the same groups will also cover the native state and thus be able to block against Ca^{2+} binding. In order to visualize the extent of shielding, we have built a possible structure for Ag43 α based on the heme binding protein (see Materials and Methods) and

Knudsen *et al.*

Glycosylation effects on the Ag43 passenger domain

manually added heptoses to all the residues identified to be glycosylated by MS (Fig. 11). Gratifyingly, our surface area accessibility calculations indicate that the O γ atoms of all 18 modified residues except T61, T180 and S439 are accessible for modification. Bearing in mind that each Ag43 α can have up to 6 glycosyl groups and the groups cluster in the N-terminal part of the protein which is known to be involved in inter-protein and inter-cellular contacts, it does not seem inconceivable that Ca $^{2+}$ binding sites could be blocked to some extent by these groups. This is all the more pertinent given that Ca $^{2+}$ ions will have to bind in the N-terminal region to affect these contacts. In this light, Ca $^{2+}$'s ability to stimulate Ag43 α^{nongly} -mediated flocculation immediately suggests that Ca $^{2+}$ ions may bridge the self-recognition of different Ag43 α molecules if binding is not blocked by glycosyl groups. Self-recognition has been localized to the N-terminal third of the α -domain [7], and our studies indicate 7-8 heptosyl groups in this region (Table 1). Although there was no difference in the settling rate for Ag43 α^{gly} and Ag43 α^{nongly} in the absence of Ca $^{2+}$ or EDTA, Ag43 α glycosylation also plays a role in more general bacterial recognition, since cells expressing Ag43 α^{gly} formed much less bacterial amyloid than Ag43 α^{nongly} -expressing cells according to simple Thioflavin-T binding assays. Although we cannot rule out the trivial explanation that glycosylation blocks access of these dyes (which are significantly larger than Ca $^{2+}$ ions), it is also possible that the glycosylation prevents a regular lateral organization of Ag43 α mediated by intermolecular β -sheets, which would subsequently give rise to amyloid-like characteristics. Analogously, glycosylation often inhibits protein crystallization by blocking the regular contacts needed to build up the crystal lattice. Note that Sherlock *et al.* did not observe any effect of glycosylation on autoaggregation according to a simple cell settlement assay [13]. This suggests that there is not a direct link between amyloid formation and interbacterial interaction. Rather, the lack of glycosylation may affect the way in which the protein is organized on the surface of the cell.

It is also noteworthy that Ag43 α shows binding affinity for Ca $^{2+}$ at all, in view of the low conservation of the binding motifs observed in the Cah autotransporter, which additionally shows low

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

(millimolar) affinity for Ca^{2+} [40]. Interestingly, Cah is also associated with bacterial aggregation, characteristic colony morphology and is important in biofilm formation, although the role of Ca^{2+} in this context is unclear.

Glycosylation accelerates refolding, perhaps by blocking inappropriate contacts

In general, glycans do not affect productive refolding kinetics from the chemically denatured state, because they do not directly reduce the activation barrier for folding by stabilizing the transition state. The main effect is to reduce non-productive aggregation, towards which the deglycosylated state is more disposed [46]. In contrast, urea-denatured $\text{Ag43}\alpha^{\text{gly}}$ appears to fold significantly faster than $\text{Ag43}\alpha^{\text{nongly}}$. Folding and misfolding need however not be totally divergent phenomena. Folding has recently been suggested to involve “predetermined pathways with optional errors” [47]), where the protein can be temporarily trapped in numerous misfolded states *en route* to the native state. These misfolded states can be stabilized by the same type of suboptimal contacts as seen in aggregates. Most folding studies have been carried out with small rapidly folding model proteins where the native interactions are formed so quickly that the only competition they face is intermolecular contacts at high protein concentrations. However, large and slowly folding proteins will have time to probe a large number of inappropriate intramolecular contacts, and in this context judiciously placed glycans can play a large role by blocking these contacts and steering towards a folded structure where the glycans remain surface-exposed.

Refolding occurs in a single major step

$\text{Ag43}\alpha$ shows significantly different folding behaviour from pertactin, the only other passenger domain whose folding has been characterized [2]. Pertactin folds via a 21 kDa C-terminal stable core, which is suggested to adopt the canonical β -helix structure and thereby serve as a scaffold in further folding [2], whereas $\text{Ag43}\alpha$ folds in a single step according to both fluorescence and SDS-PAGE.

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Sequence comparison between Ag43 α and pertactin did not reveal any significantly conserved sequences in the C-terminal part which could intimate a conserved stable core in Ag43 α (data not shown), suggesting that there may be a number of different folding mechanisms for passenger domains. However, the pertactin and Ag43 α have the common property of folding on the hour-to-days scale *in vitro*. Such a slow folding mechanism might sustain the passenger domain in an unfolded conformation in the periplasm, which might be necessary prior to secretion across the outer membrane through the β -domain. We cannot conclude unequivocally whether the β -domain is a prerequisite for keeping passenger domain in this pre-folded conformation. The large majority of the Ag43 α molecules is retained in inclusion bodies in the cytosol, with or without the appropriate N-terminal signal sequence. This could be due to overexpression and consequent blockage of the export pathway or a genuine requirement for the β -domain. Nevertheless, our data show that Ag43 α can to some extent survive in the periplasm without the β -domain. A non-native state should be easy prey for proteases in the periplasm, but this could be counteracted by the large number of chaperones in the periplasm [48].

Another phenomenon which Ag43 α shares with pertactin is a disordered C-terminus. Pertactin's C-terminal 59 residues, which are disordered in the crystal structure, can be removed without affecting pertactin folding or unfolding [2]. Similarly, trypsin treatment (as well as activity by endogenous proteases) removes the C-terminal 132 residues of Ag43 α . SDS-PAGE analysis showed bands of several sizes between the full-length and the trimmed version, effectively ruling out the existence of a single well-defined C-terminal domain separated from the 367 preceding residues by a flexible linker region. Despite the flexibility of this region, it is only removed slowly during purification, and is retained both in the periplasm, cytosol and inclusion bodies upon overexpression. The role of this unstructured domain remains unknown; it is unlikely to be involved in the self-recognition motif since this has been localized in the N-terminal region of the protein [7]. Given that both Ag43 α^{gly} and Ag43 α^{nongly} are trypsin-resistant when bound to the cell surface, it is possible that the flexible C-

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

terminal region is stabilized upon contact with the membrane-bound β -domain of Ag43 α or other parts of the outer membrane.

Stage 2(a) POST-PRINT

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20071497

ACKNOWLEDGEMENTS

We thank Professor Per Klemm for introducing us to the Ag43 system and for generously providing us with the *E. coli* strains expressing Ag43. Dr. Steen Vang Pedersen and Professor Jan Enghild are gratefully acknowledged for performing N-terminal sequencing of trimmed Ag43 α and Professor Karen Welinder is thanked for helpful suggestions on proteolysis inhibition. We are grateful to Hanne Krone Nielsen for help with purification and to Poul Larsen for instructions on how to use the confocal scanning microscope. This work was funded by a pre-doctoral grant from the Danish Research Training Committee to the Graduate School in Molecular Biophysics. D.E.O. is funded by the Danish Research Foundation (inSPIN) and the Villum Kann Rasmussen Foundation (BioNET). U.B.W. is funded by a post-doc stipend from Villum Kann Rasmussen Foundation (BioNET). A.S. is funded by the Carlsberg Foundation and the Obel Foundation.

REFERENCES

- 1 Henderson, I. R., Navarro-Garcia, F. and Nataro, J. P. (1998) The great escape: structure and function of the autotransporter proteins. *Trends Microbiol.* **6**, 370-378
- 2 Junker, M., Schuster, C. C., McDonnell, A. V., Sorg, K. A., Finn, M. C., Berger, B. and Clark, P. L. (2006) Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4918-4923
- 3 Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J. and Gros, P. (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* **23**, 1257-1266
- 4 Mogensen, J. E., Tapadar, D., Schmidt, M. A. and Otzen, D. E. (2005) Barriers to folding of the Transmembrane Domain of the *Escherichia coli* Autotransporter Adhesin involved in diffuse adherence. *Biochemistry.* **44**, 4533-4545
- 5 Mogensen, J. E., Kleinschmidt, J. H., Schmidt, M. A. and Otzen, D. E. (2005) Misfolding of a Bacterial Autotransporter. *Prot. Sci.* **14**, 2814-2827
- 6 Diderichsen, B. (1980) *flu*, a metastable gene controlling surface properties of *Escherichia coli*. *J. Bact.* **141**, 858-867
- 7 Klemm, P., Hjerrild, L., Gjermansen, M. and Schembri, M. A. (2003) Structure-function analysis of the self-recognizing antigen 43 autotransporter protein from *Escherichia coli*. *Mol. Microbiol.* **51**, 283-296
- 8 Caffrey, P. and Owen, P. (1989) Purification and N-terminal sequence of the α subunit of antigen 43, a unique protein complex associated with the outer membrane of *Escherichia coli*. *J. Bact.* **171**, 3634-3640
- 9 Klemm, P., Vejborg, R. M. and Sherlock, O. (2006) Self-associating autotransporters, SAATs: Functional and structural similarities. *Int. J. Med. Microbiol.* **296**, 187-195
- 10 Schmidt, M. A., Riley, L. W. and Benz, I. (2003) Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol.* **11**, 554-561
- 11 Benz, I. and Schmidt, M. A. (2001) Glycosylation with heptose residues mediated by the *aah* gene product is essential for adherence of the AIDA-I adhesin. *Mol. Microbiol.* **40**, 1403-1413
- 12 Lindenthal, C. and Elsinghorst, E. A. (2001) Enterotoxigenic *Escherichia coli* TibA glycoprotein adheres to human intestine epithelial cells. *Infect. Immun.* **69**, 52-57
- 13 Sherlock, O., Dobrindt, U., Jensen, J. B., Vejborg, R. M. and Klemm, P. (2006) Glycosylation of the self-recognizing *Escherichia coli* Ag43 autotransporter protein. *J. Bact.* **188**, 1798-1807
- 14 Charbonneau, M.-E., Girard, V., Nikolakakis, A., Campos, M., Berthiaume, F., Dumas, F., Lépine, F. and Mourez, M. (2007) O-linked glycosylation ensures the normal conformation of the autotransporter Adhesin Involved in Diffuse Adherence. *J. Bact.* **189**, 8880-8889
- 15 Tams, J. W. and Welinder, K. G. (2001) Kinetic stability of designed glycosylation mutants of *Coprinus cinereus* mutants. *Biochem. Biophys. Res. Comm.* **286**, 701-706

- 16 Tams, J. W., Vind, J. and Welinder, K. G. (1999) Adapting protein solubility by glycosylation. N-Glycosylation mutants of Coprinus cinereus peroxidase in salt and organic solutions. *Biochim Biophys Acta*. **1432**
- 17 Gerken, T. A., Butenhof, K. J. and Shogren, R. (1989) Effects of glycosylation on the conformation and dynamics of α -linked glycoproteins: carbon-13 NMR studies of ovine submaxillary mucin. *Biochemistry*. **28**, 5536-5543
- 18 Kjærgaard, K., Schembri, M. A., Hasman, H. and Klemm, P. (2000) Antigen 43 from *Escherichia coli* induces inter- and intraspecies cell aggregation and changes in colony morphology of *Pseudomonas fluorescens*. *J. Bacteriol.* **182**, 4789-4796
- 19 Kjærgaard, K., Hasman, H., Schembri, M. A. and Klemm, P. (2002) Antigen 43-mediated autotransporter display, a versatile bacterial cell surface presentation system. *J. Bacteriol.* **184**, In press
- 20 Shevchenko, A. M., Wilm, M., Vorm, O. and Mann, M. (1996) Mass Spectrometric Sequencing of Proteins from Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **68**, 850-858
- 21 Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W. and Burlingame, A. L. (2005) Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. **5**
- 22 van der Heeft, E., ten Hove, G. J., Herberts, C. A., Meiring, H. D., van Els, C. A. and de Jong, A. P. (1998) A microcapillary column switching HPLC-electrospray ionization MS system for the direct identification of peptides presented by major histocompatibility complex class I molecules. *Anal. Chem.* **70**, 3742-3751
- 23 Kiehne, A., Brekenfeld, Baessmann, C. and Hartmer, R. (2006) Electron Transfer Dissociation for Sequence Characterization of larger Peptides and small Proteins. In *Int. Mass Spectr. Conf., Prague 2006*
- 24 Perkins, D. N., Pappin, D. J., Creasy, D. M. and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. **20**, 3551-3567
- 25 Mogensen, J. E., Ibsen, H., Lund, J. and Otzen, D. E. (2004) Elimination of an off-pathway folding intermediate by a single point mutation. *Biochemistry*. **43**, 3357-3367
- 26 Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**, 10035-10038
- 27 Tefsen, B., Geurtsen, J., Beckers, F., Tommassen, J. and de Cock, H. (2005) Lipopolysaccharide transport to the bacterial outer membrane in spheroplasts. *J. Biol. Chem.* **280**, 4504-4509
- 28 Šali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*. **234**, 779-815
- 29 Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881-10890
- 30 Lindahl, E., Hess, B. and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **7**, 306-317

- 31 van Gunsteren, W. F., Kruger, P., Billeter, S. R., Mark, A. E., Eising, A. A., Scott, W. R. P., Huneberger, P. H. and Tironi, I. G. (1996) Biomolecular simulation: the GROMOS96 manual and user Guide. Biomos Hochschulverlag AG, Groningen
- 32 Hubbard, S. J. and Thornton, J. M. (1993) NACCESS', computer program. Department of Biochemistry and Molecular Biology, University College London, London
- 33 DeLano, W. L. (2002) The PyMOL Molecular Graphics System. ed.)^eds.), DeLano Scientific, San Carlos, CA.
- 34 Pohlner, J., Halter, R., Beyreuther, K. and Meyer, T. F. (1987) Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. *Nature*. **325**, 458-462
- 35 Hogan, J. M., Pitteri, S. J., Chrisman, P. A. and McLuckey, S. A. (2005) Complementary Structural Information from a Tryptic N-Linked Glycopeptide via Electron Transfer Ion/Ion Reactions and Collision-Induced Dissociation. *J. Proteome Res.* **4**, 628-632
- 36 Schneider, T. D. and Stephens, R. M. (1990) Sequence logos: a new way to display consensus sequences. *Nucl. Acids Res.* **18**, 6097-6100
- 37 Crooks, G. E., Hon, G., Chandonia, J.-M. and Brenner, S. E. (2004) WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188-1190
- 38 Fasman, G. D., ed. (1996) Circular dichroism and the conformational analysis of biomolecules. Plenum Publ. Co.
- 39 Cronin, N. B., O'Reilly, A., Duclouhier, H. and Wallace, B. A. (2005) Effects of deglycosylation of sodium channels on their structure and function. *Biochemistry*. **44**, 441-449
- 40 Torres, A. G., Perna, N. T., Burland, V., Ruknudin, A., Blattner, F. R. and Kaper, J. B. (2002) Characterization of Cah, a calcium-binding and heat-extractable autotransporter protein of enterohaemorrhagic *Escherichia coli*. *Mol. Microbiol.* **45**, 951-966
- 41 Tanford, C. (1968) Protein Denaturation. Part A. Characterization of the denatured state. *Adv. Prot. Chem.* **23**, 121-217
- 42 Frokjaer, S. and Otzen, D. E. (2005) Protein drug stability - a formulation challenge. *Nat. Rev. Drug. Delivery.* **4**, 298-306
- 43 Otzen, D. E. and Nielsen, P. H. (2007) We find them here, we find them there: Functional bacterial amyloid. *Cell. Mol. Life Sci.* **Nov 24**; [Epub ahead of print]
- 44 Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and protein. *Nat. Biotechnol.* **22**, 1302-1306
- 45 Lin, L. N., Mason, A. B., Woodworth, R. C. and Brandts, J. F. (1994) Calorimetric studies of serum transferrin and ovotransferrin. Estimates of domain interactions, and study of the kinetic complexities of ferric ion binding. *Biochemistry*. **33**, 1881-1888
- 46 Graf, R., Lang, K., Vogl, H. and Schmid, F. X. (1987) The mechanism of folding of pancreatic ribonucleases is independent of the presence of covalently linked carbohydrate. *J. Biol. Chem.* **262**, 10624-10629
- 47 Krishna, M. M. G. and Englander, S. W. (2007) A unified mechanism for protein folding: Predetermined pathways with optional errors. *Prot. Sci.* **16**, 449-464
- 48 Mogensen, J. E. and Otzen, D. E. (2005) Interactions between periplasmic chaperones and bacterial outer membrane proteins. *Mol. Microbiol.* **57**, 326-346

Stage 2(a) POST-PRINT

TABLES

Table 1: Heptosylated peptides identified and sequenced by mass spectrometry.

Sequence ^a	Mr (Da)	Sequence	Sequence motif ^b	Modifications ^c
1 - 29*	3181.5163	ADIVVHPGETVNGGTLANHDNQIVFGTTN		(inconclusive)
32 - 60*	3171.4115	TISTGLEYGPDNEANTGGQWVQDGGTANK	NGMTISTGLEYGP	1: T35, +ETD
61 - 70*	1154.6400	TTVTSGGLQR	NKTTVTSGGLQRV	1: S65, +DTT, +ETD
61 - 70*	1402.6675	TTVTSGGLQR		1: S65; T61, +ETD
71 - 93	2334.1298	VNPGGSVSDTVISAGGGQSLQGR	VSDTVISAGGGQS	1: S83, +DTT, +ETD
71 - 93	2526.1932	VNPGGSVSDTVISAGGGQSLQGR		(inconclusive)
98 - 120*	2825.2912	TLNGGEQWMHEGAIATGTVINDK		2: T97 and T98
174 - 194*	2276.0192	AEGTANTTVVYAGGDQTVHGH	IVRAEGTANTTVV	1: (T177; inconclusive)
174 - 194*	2468.0826	AEGTANTTVVYAGGDQTVHGH	IVRAEGTANTTVV AEGTANTTVVYAG	2: T177 and T180
228 - 240*	1450.6899	NGGVAGNTTVNQK	GVAGNTTVNQKGR	1: T235 or T236

258 - 275	1907.9799	QGGALVTSTAATVTGINR	GGALVT <u>S</u> TAATVT	S265 +DTT, +ETD
258 - 275	2100.0433	QGGALVTSTAATVTGINR		(inconclusive)
265 - 275	1281.6412	STAATVTGINR	GGALVT <u>S</u> TAATVT	S265
276 - 285	1141.6126	LGAFSVVEGK	NRLGAF <u>S</u> VVEGKA	S280, +DTT
320 - 347	2728.2820	NGGTATTVSMGNGGVLLADSGAAVSGTR	DSGAAV <u>S</u> GTRSDG	S344
320 - 347	2920.3454	NGGTATTVSMGNGGVLLADSGAAVSGTR	-	(inconclusive)
320 - 347	3112.4088	NGGTATTVSMGNGGVLLADSGAAVSGTR		(inconclusive)
352 - 367*	1742.8990	AFSIGGGQADALMLEK	SDGKAF <u>S</u> SIGGGQA	S354, +DTT
368 - 392*	2665.2354	GSSFTLNAGDTATDTTVNGGLFTAR	MLEKG <u>S</u> SFTLNAG	S370, +ETD
424 - 439	1858.8167	EGDALLQGGSLTGNGS	LQGGSL <u>T</u> GNG <u>S</u> SLTGNG <u>S</u> VEKSGS	T435 and S439; semitryptic
443 - 458	1785.8843	SGSGTLTVSNNTTLTQK	SGTLTV <u>S</u> NNTTLTQ	1: S451; +DTT, +ETD

THIS IS NOT THE FINAL VERSION - SEE OUR JOURNAL WEBSITE FOR THE FINAL VERSION

Stage 2(a) POST-PRINT

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Notes:

^a * after peptide number indicates peptides not identified by Klemm and co-workers [13].

^b Modified residue is underlined.

^c +DTT indicates that modification site was identified by derivatizing with DTT. +ETD indicates that the assignment was confirmed electron transfer dissociation based peptide sequencing.

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

Table 2. Rate constants for refolding of Ag43 α from the heat- and urea-denatured state. All data are for Ag43 α^{nongly} unless otherwise indicated. Data are based on experiments in triplicates (except for SDS-PAGE where they were performed in duplicate).

Method	Denatured state	Refolding conditions	$t_{1/2}$ (hrs)
Trypsin-SDS PAGE	55°C	25°C	1.46±0.26
Trp fluorescence	55°C	25°C	5.6±0.8
Trp fluorescence	5M urea, 25°C	0.5M urea, 25°C	7.42±0.97 (Ag43 α^{nongly})
			2.43±0.54 (Ag43 α^{gly})

FIGURE LEGENDS

Figure 1

(A) Trypsin treatment of purified Ag43 α^{nongly} , indicating trimming of the protein. Arrow indicates trypsin. Numbers (kDa) refer to molecular weight markers. Similar patterns were observed for Ag43 α^{gly} .

(B) High mass precision LC-MS/MS analysis of Ag43 α^{nongly} peptide mixtures obtained after *in situ* digestion with trypsin. Numbering refers to α -domain only, and excludes the 52-residue signal sequence. Ag43 α amino acid sequence coverage (in bold) obtained by capillary LC-MS/MS analysis of peptide mixtures from intact Ag43 α^{nongly} (upper panel) and trimmed Ag43 α^{nongly} (lower panel). The underlined region in trimmed Ag43 α^{nongly} (lower panel) could not be identified by the MS data, which indicates that the C-terminal region from K367 onwards is absent in the lower band.

(C) m/z MS spectra of intact Ag43 α^{nongly} (from PKL1061) and Ag43 α^{gly} (from induced OS64). Arrows indicate positions of multiply glycosylated variants of Ag43 α^{gly} as well as a possible singly glycosylated variant of a trimmed version of Ag43 α^{nongly} . The m/z value of 49864 Da corresponds reasonably to the full length autotransporter domain (residues 1-499, theoretical molecular weight 49787 Da) while the m/z value of 44083 Da for the trimmed version of Ag43 α^{nongly} , in which part of the unstructured C-terminus has been removed by endogenous proteases during purification, corresponds well to the 45 kDa band in Fig. 1A and residues 1~422.

Figure 2

Automatic nanoflow LC-MS/MS analysis of Ag43 allows identification of sites of heptosylation.

(A) MS peptide data established the presence of heptosylation group on the doubly charged peptide 71-VNPGGSVSDTVISAGGGQSLQGR-93 by a mass difference of 96 Da corresponding to heptose group of 192 (= 2 x 96) Da.

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

(B) Tandem MS sequencing of heptosylated peptide $[H+2H]^{2+}$ m/z 1168.083 confirms S83 as acceptor site in panel (B) compared to the unmodified peptide in panel (C) based on near-complete b- and y-ion ion-series including b13 and y11, confirming the site of modification.

(D) Tandem MS spectrum of heptosylated peptide $[H+2H]^{2+}$ m/z 1168 using electron transfer dissociation enabling peptide sequencing with intact heptosylation indicated by a 274 Da mass difference between fragment ions z10 and z11, corresponding to the combined mass of serine (87 Da) and heptose (192 Da).

Figure 3

Identification of sites of heptosylation by chemical derivatization prior to peptide sequencing by tandem mass spectrometry.

(A) Tandem MS peptide of the unmodified doubly charged peptide 61-TTVTSGGLQR-70. The mass difference of 87 Da corresponds to Ser65.

(B) Tandem MS sequencing of BEMAD treated heptosylated peptide confirms S65 as acceptor site. The mass difference of 223 Da corresponds to the masses of Ser and DTT minus a water molecule.

Figure 4.

Identification of motifs for O-linked glycosylation in Ag43 α based on the identified glycosylated peptides.

(A) Sequence logo of Thr- and Ser- glycosylated peptides by multiple alignment of sequences -6 to +6 (right).

(B) Sequence logo of all non-glycosylated Thr- and Ser- residues.

(C) Sequence log of the combination of Ser-glycosylated peptides obtained in this study and the peptides provided by Mourez and co-workers [14].

Amino acids are colored according to their chemical properties: polar amino acids (G,S,T,Y,C,Q,N) are green/purple, basic (K,R,H) blue, acidic (D,E) red and hydrophobic (A,V,L,I,P,W,F,M) amino

Knudsen et al.

Glycosylation effects on the Ag43 passenger domain

acids are black. Maximum sequence conservation per site is $\log_2 20 \approx 4.32$ bits for proteins.
Generated by Weblogo 2.8.2.

Figure 5.

(A) Far UV CD spectra of native (N) and heat-denatured (D) Ag43 α^{nongly} and Ag43 α^{gly} .
(B) Thermal scans of Ag43 α^{nongly} and Ag43 α^{gly} monitored by fluorescence emission at 358 nm. Data fitted as described [25]. *Insert*: Trp fluorescence emission spectra of native (N) and heat-denatured (D) Ag43 α^{nongly} (excitation at 295 nm).

Figure 6. Effect of Ca²⁺ on Ag43 α stability and aggregative properties.

(A) Change in melting temperature of Ag43 α^{nongly} and Ag43 α^{gly} with Ca²⁺ concentration. The data have been fitted to eq. 1 to obtain apparent dissociation constants of 0.92 ± 0.34 mM for Ag43 α^{gly} and 0.76 ± 0.33 mM for Ag43 α^{nongly} .
(C) Settling profile of cells expressing Ag43 α^{nongly} (—) and Ag43 α^{gly} (-----) or transformed with an empty vector (-----) in the absence of added Ca²⁺ (◇), in the presence of 10 mM Ca²⁺ (■) or in the presence of 10 mM EDTA (○). Each settling profile is an average of a triple determination from three individual cell cultures, indicated by the error bars. Lines are drawn for visual guidance.

Figure 7. Refolding of Ag43 α^{nongly} and Ag43 α^{gly} from the heat-denatured state.

(A) SDS-PAGE of refolding monitored by trypsin digestion. Ag43 α^{nongly} was incubated at 55 or 65°C for 15 minutes, after which it was allowed to refold for 0-24 hrs at 25°C. At various time points, trypsin was added to digest unfolded Ag43 α^{nongly} .
(B) Band intensities of the gel in panel A (errors from duplicate measurements). Data are fitted to a single exponential decay to yield half lives of 1.46 ± 0.26 and 2.20 ± 0.03 hrs for 55 and 65°C, respectively.

Figure 8.

(A) Urea denaturation and renaturation of Ag43 α^{nongly} and Ag43 α^{gly} monitored by excitation at 295 nm and the emission ratio at 346/356 nm. Due to very low expression levels of Ag43 α^{gly} , it was not possible to carry out a complete refolding series for this protein. Data fitted as described in Materials and Methods.

(B) Kinetics of refolding of Ag43 α^{nongly} and Ag43 α^{gly} monitored as in panel A. Data are carried out in triplicate with associated error bars, fitted to a single exponential decay and summarized in Table 2. Data for Ag43 α^{gly} are displaced upwards by 0.20 to make it easier to distinguish the two proteins.

Figure 9.

Cellular localization of Ag43 α^{nongly} expressed without (α^0) and with (α^{sig}) signal sequence. Data represent a densitometric band analysis, based on a Western blot of fractionated samples. Data are normalized to the total band density. Error bars based on duplicates.

Figure 10.

Thioflavin T binding to cells expressing Ag43 α . Cells were visualized by (A) phase contrast images or (B) Thioflavin T fluorescence (excitation at 450 nm, emission at 482 nm).

Figure 11.

Two views of the modeled structure of Ag43 α with heptose groups added at all positions shown to be glycosylated. Note that our MS data only suggest up to 6 heptose groups per protein molecule. The figure's heptosylation level is therefore not representative but is shown for illustration. We have identified heptose groups at residues T35, T61, S65, S83, T97, T98, T177, T180, T235, T236, S265,

Knudsen et al.

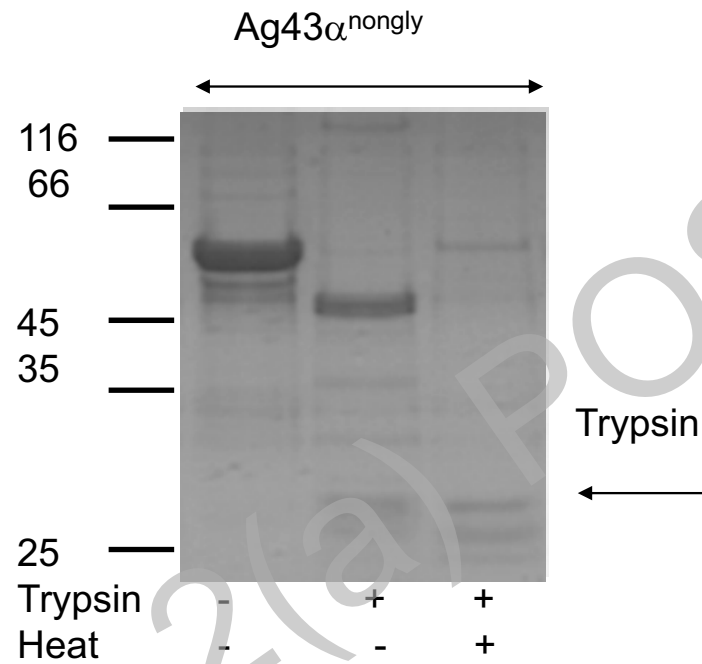
Glycosylation effects on the Ag43 passenger domain

S280, S344, S354, S370, T435, S439 and S451. Surface area accessibility calculations indicate that the O γ atoms of all these residues except T61, T180 and S439 are accessible for modification.

Stage 2(a) POST-PRINT

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20071497

Figure 1A Knudsen et al.



THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20071497

Stage 2(a) POST-PRINT

Figure 1B Knudsen et al.

1 ADIVVHPGET VNGGTLANHD NQIVFGTTNG MTISTGLEYG PDNEANTGGQ
 51 WVQDGGTANK TTVTSGGLQR VNPGGSVSDT VISAGGGQSL QGRAVNTTLN
 101 GGEQWMHEGA IATGTVINDK GWQVVKPGTV ATDTVNTGA EGGPDAENGD
 151 TGQFVRGDAV RTTINKNGRQ IVRAEGTANT TVVYAGGDQT VHGHALDTTL
 201 NGGYQYVHNG GTASDTVVNS DGWQIVKNGG VAGNTTVNQK GRLEVDAGGT
 251 ATNVTLKQGG ALVTSTAATV TGINRLGAFS VVEGKADNVV LENGGRLDVL
 301 TGHTATNTRV DDGGTLDVRN GGTATTVSMG NGGVLLADSG AAVSGTRSDG
 351 KAFSIGGGQA DALMLEKGSS FTLNAGDTAT DTTVNGGLFT ARGGTLAGTT
 401 TLNNGAILTL SGKTVNNDTL TIREGDALLQ GGSLTGNGSV EKSGSGTLTV
 451 SNTTLTQKAV NLNEGTTLN DSTVTTDVIA QRGTALKLTG STVLNGAID

1 ADIVVHPGET VNGGTLANHD NQIVFGTTNG MTISTGLEYG PDNEANTGGQ
 51 WVQDGGTANK TTVTSGGLQR VNPGGSVSDT VISAGGGQSL QGRAVNTTLN
 101 GGEQWMHEGA IATGTVINDK GWQVVKPGTV ATDTVNTGA EGGPDAENGD
 151 TGQFVRGDAV RTTINKNGRQ IVRAEGTANT TVVYAGGDQT VHGHALDTTL
 201 NGGYQYVHNG GTASDTVVNS DGWQIVKNGG VAGNTTVNQK GRLEVDAGGT
 251 ATNVTLKQGG ALVTSTAATV TGINRLGAFS VVEGKADNVV LENGGRLDVL
 301 TGHTATNTRV DDGGTLDVRN GGTATTVSMG NGGVLLADSG AAVSGTRSDG
 351 KAFSIGGGQA DALMLEKGSS FTLNAGDTAT DTTVNGGLFT ARGGTLAGTT
 401 TLNNGAILTL SGKTVNNDTL TIREGDALLQ GGSLTGNGSV EKSGSGTLTV
 451 SNTTLTQKAV NLNEGTTLN DSTVTTDVIA QRGTALKLTG STVLNGAID

DOI:10.1042/BJ20071497

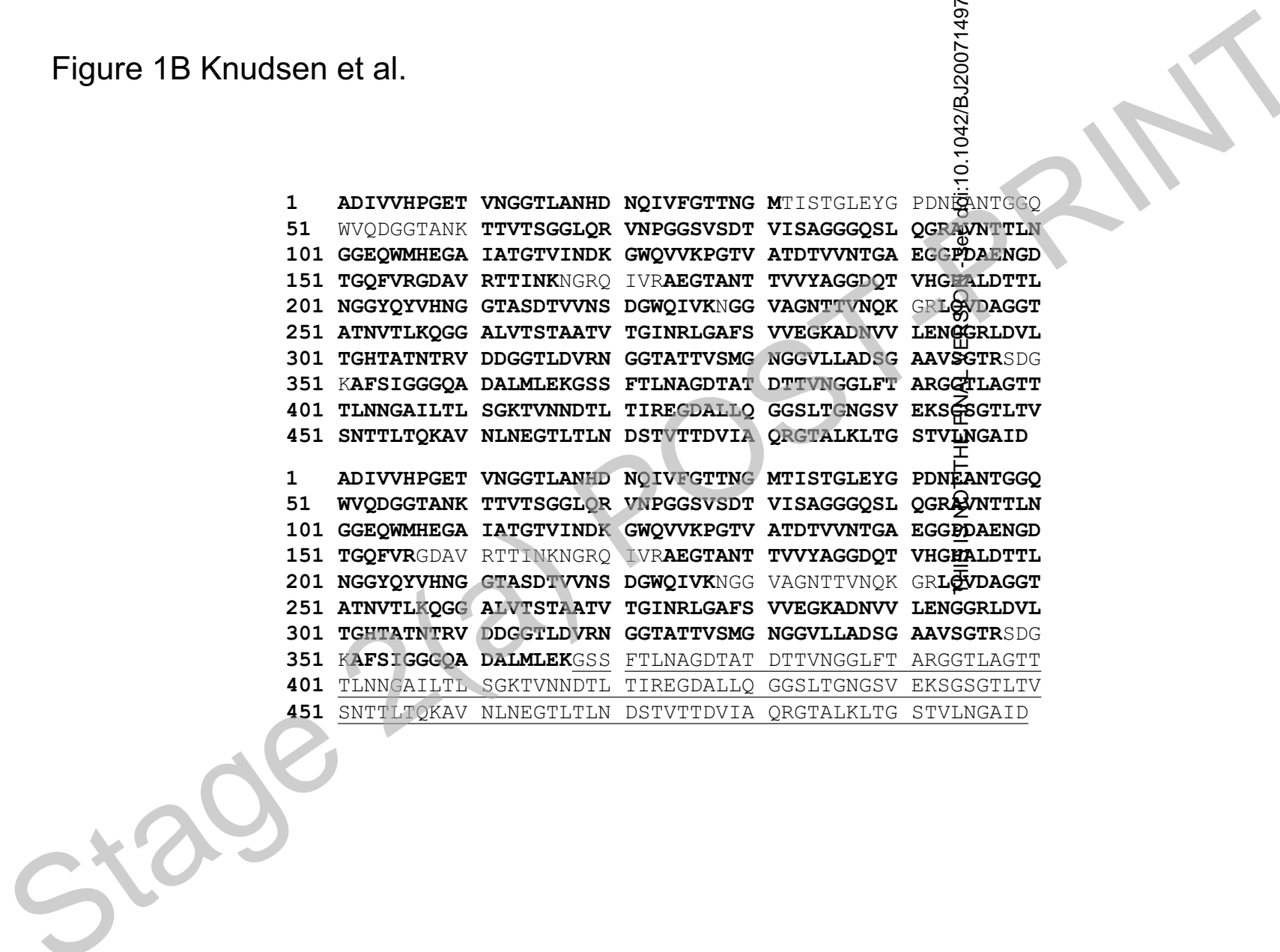


Figure 1C Knudsen et al.

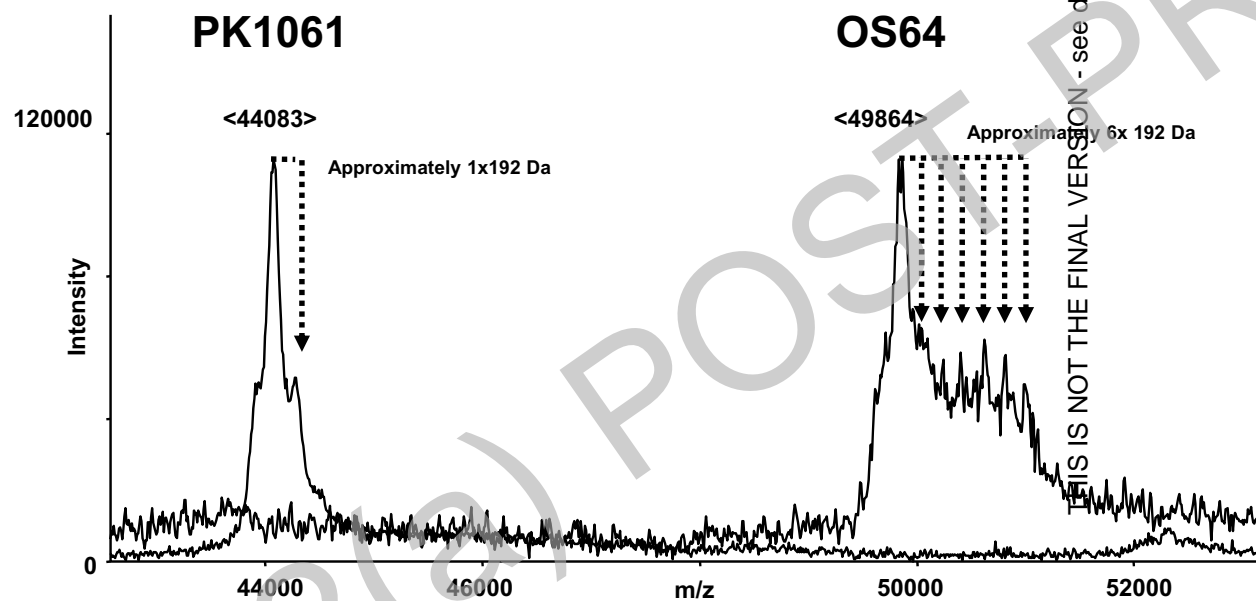
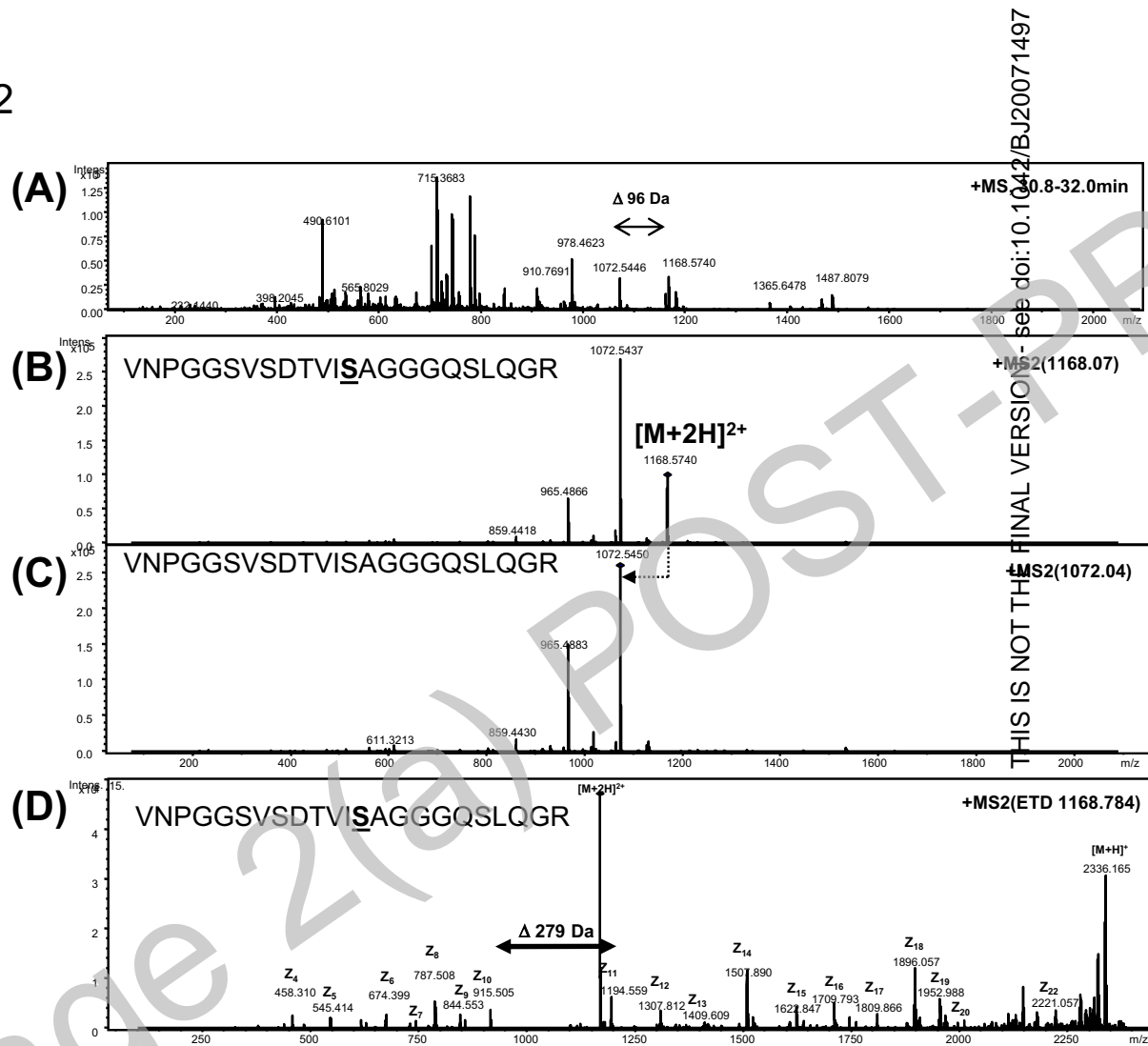


Figure 2



THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20071497

Figure 3 Knudsen et al.

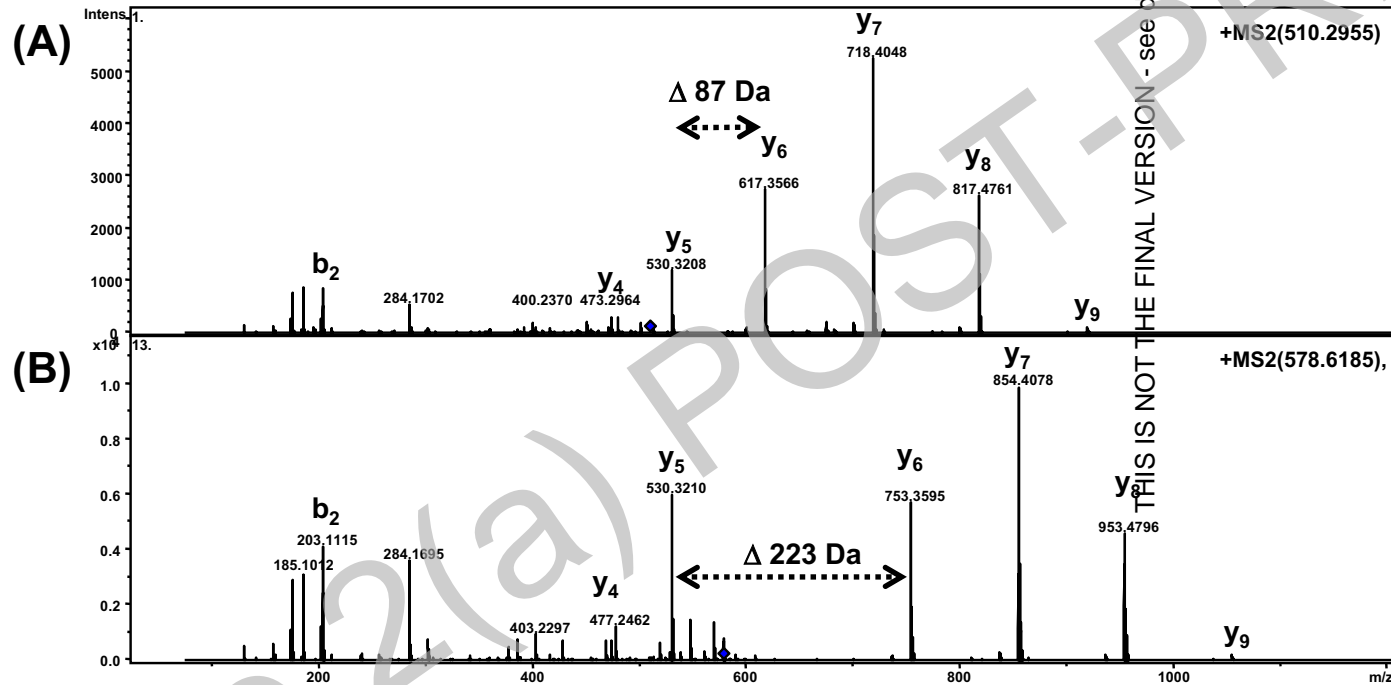


Figure 4 Nihsen et al.

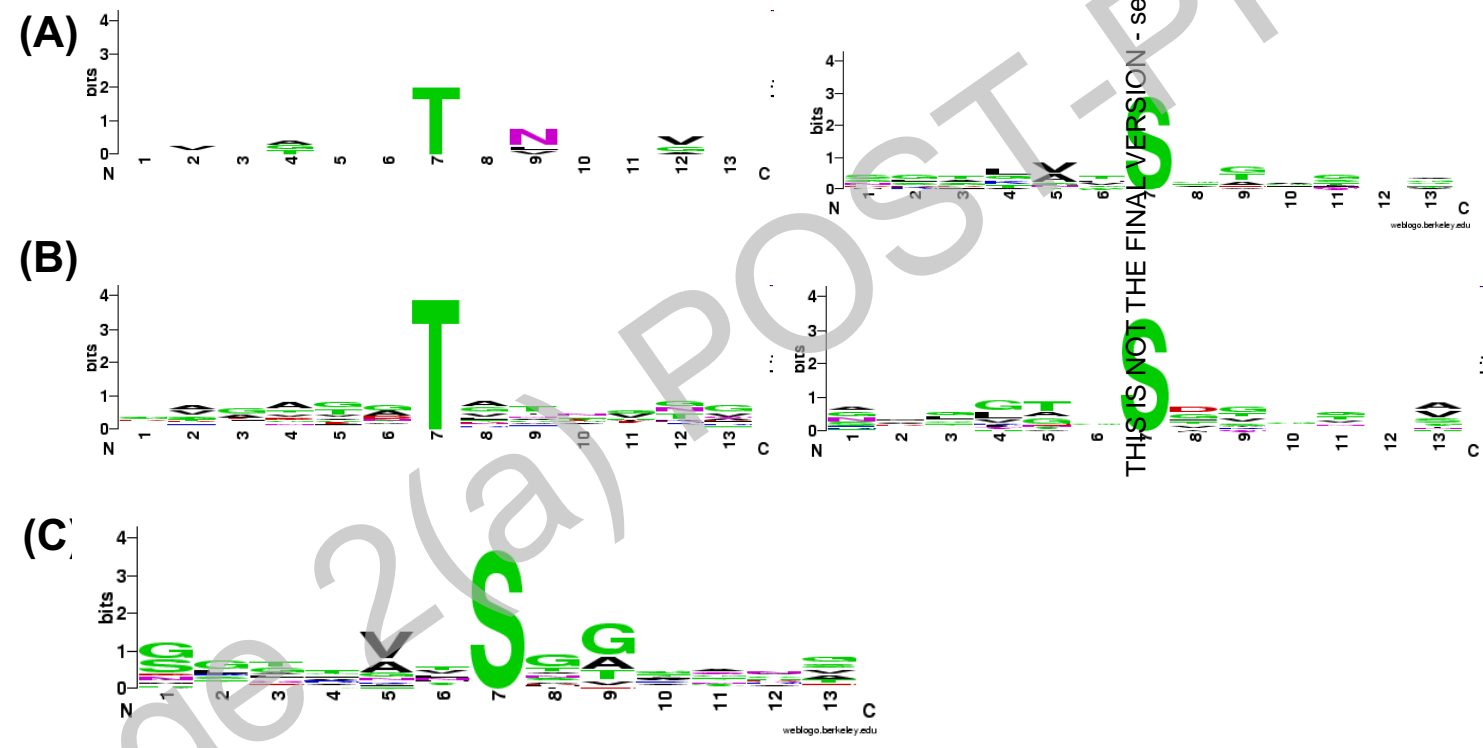


Figure 5 Knudsen et al.

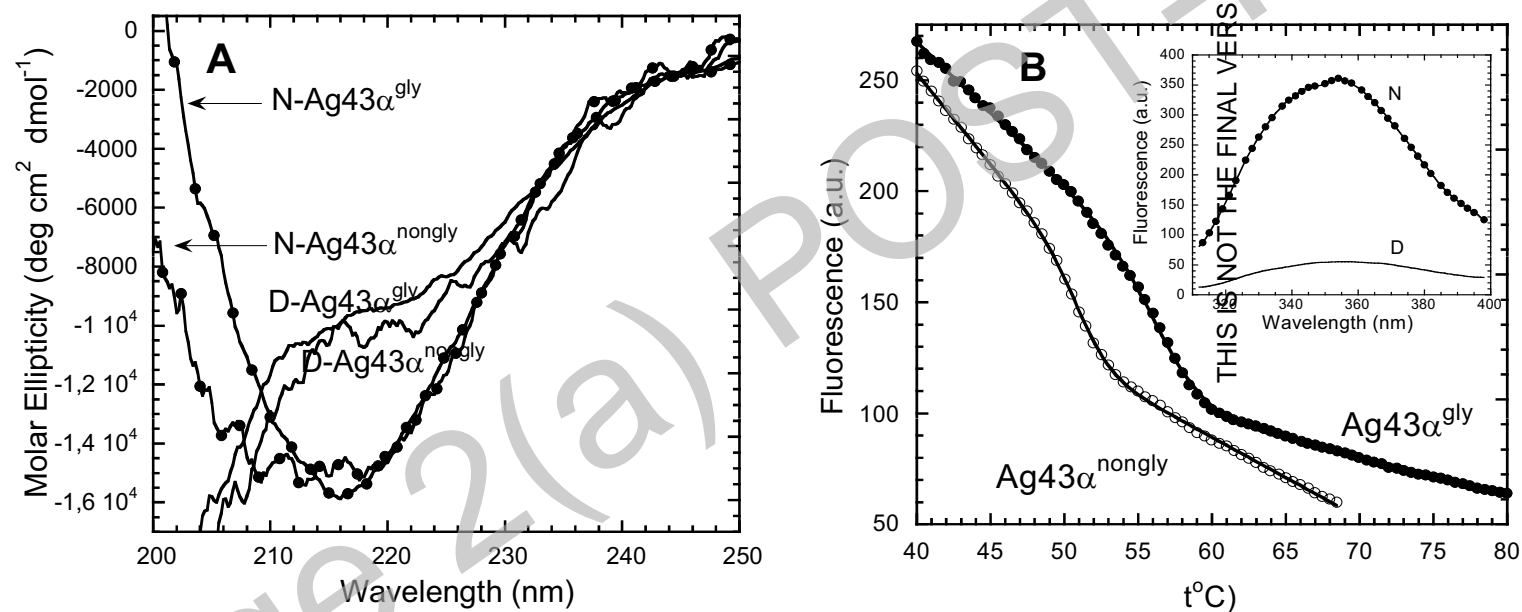
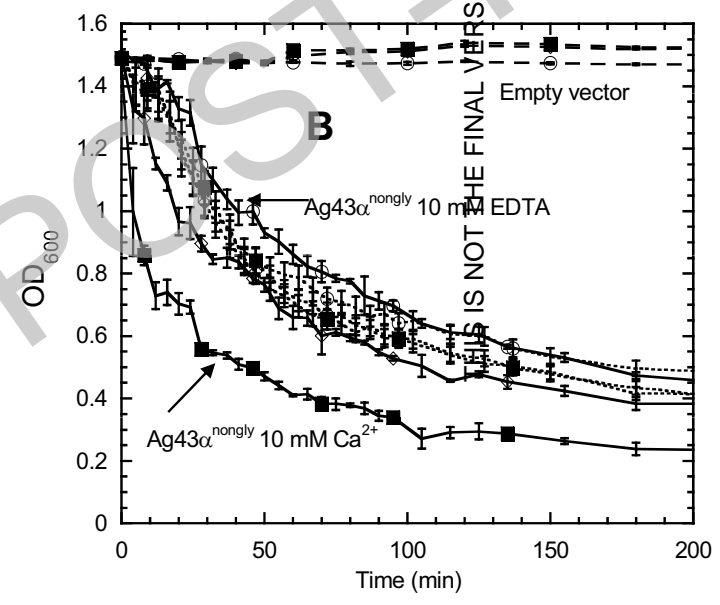
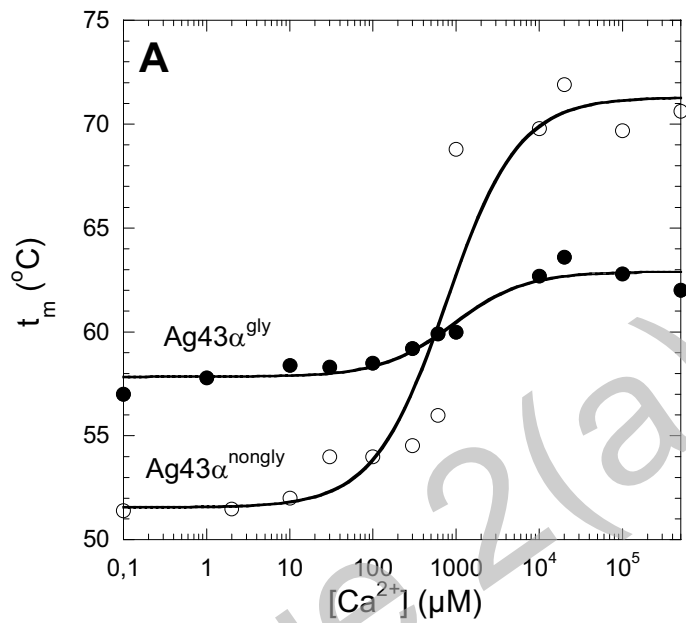


Figure 6 Knudsen et al.

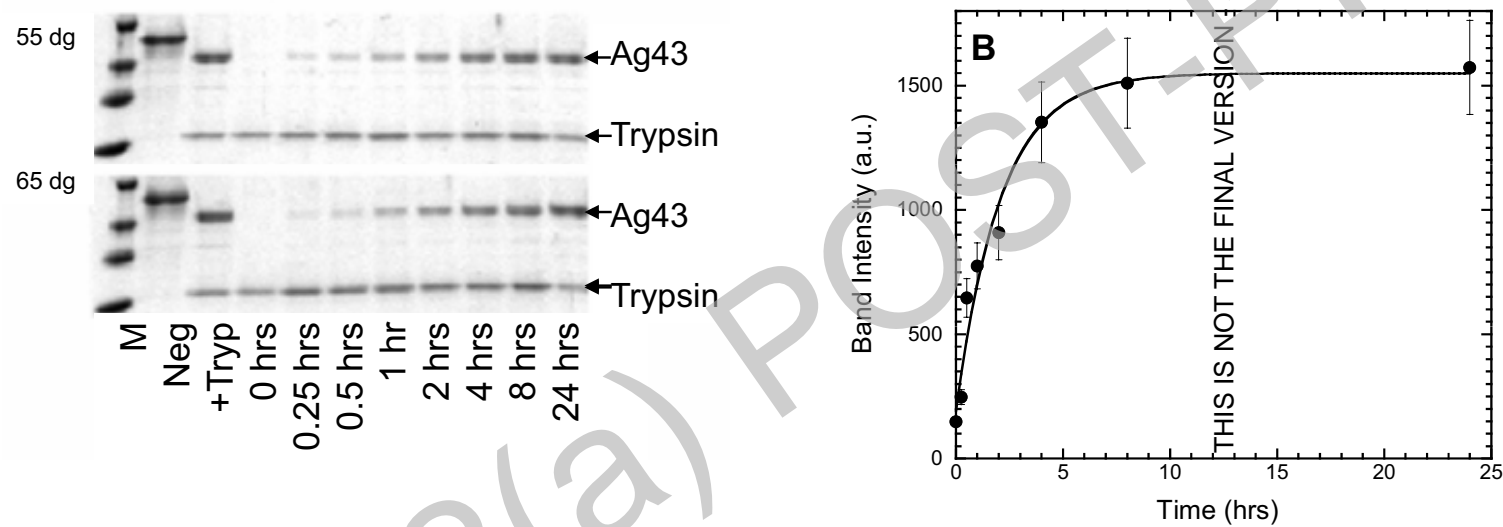


Stage 2(a)

POST-PRINT

THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20071497

Figure 7 Knudsen et al.



see doi:10.1042/BJ20071497

Time (hrs)

Figure 8 Knudsen et al.

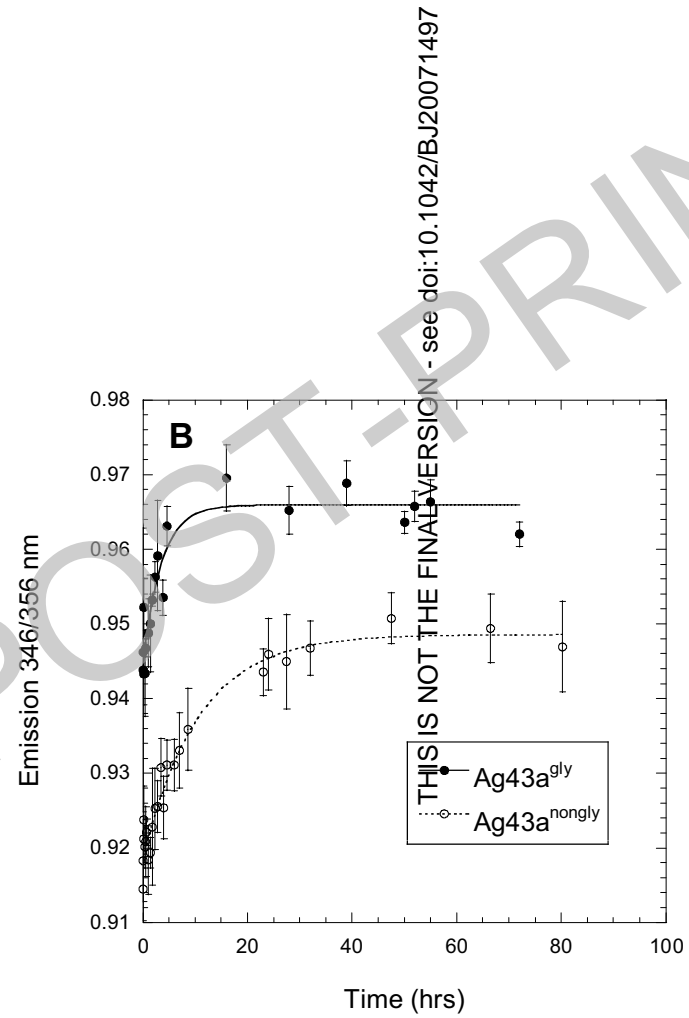
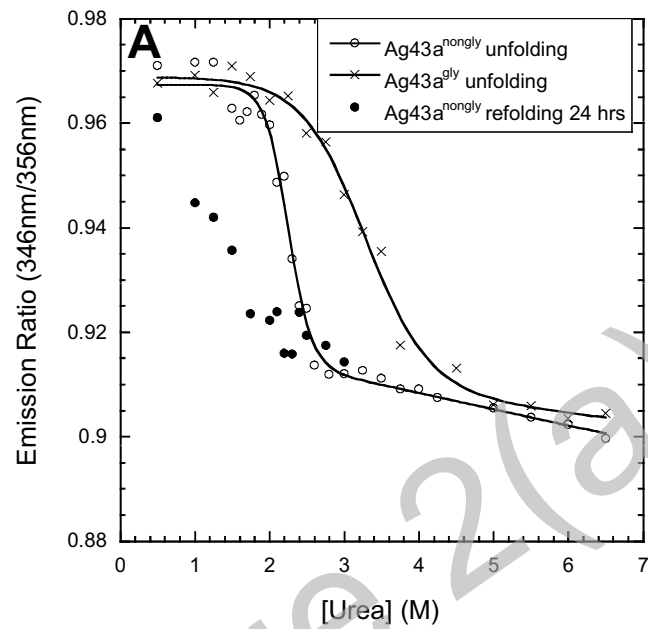
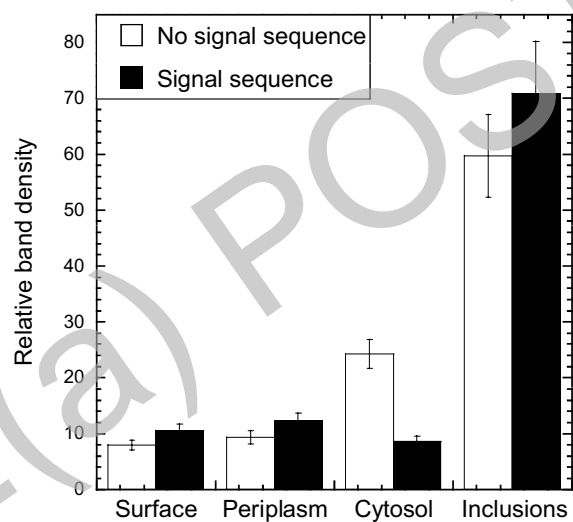


Figure 9 Knudsen et al.



THIS IS NOT THE FINAL VERSION - see doi:10.1042/BJ20071497

Figure 10 Knudsen et al.

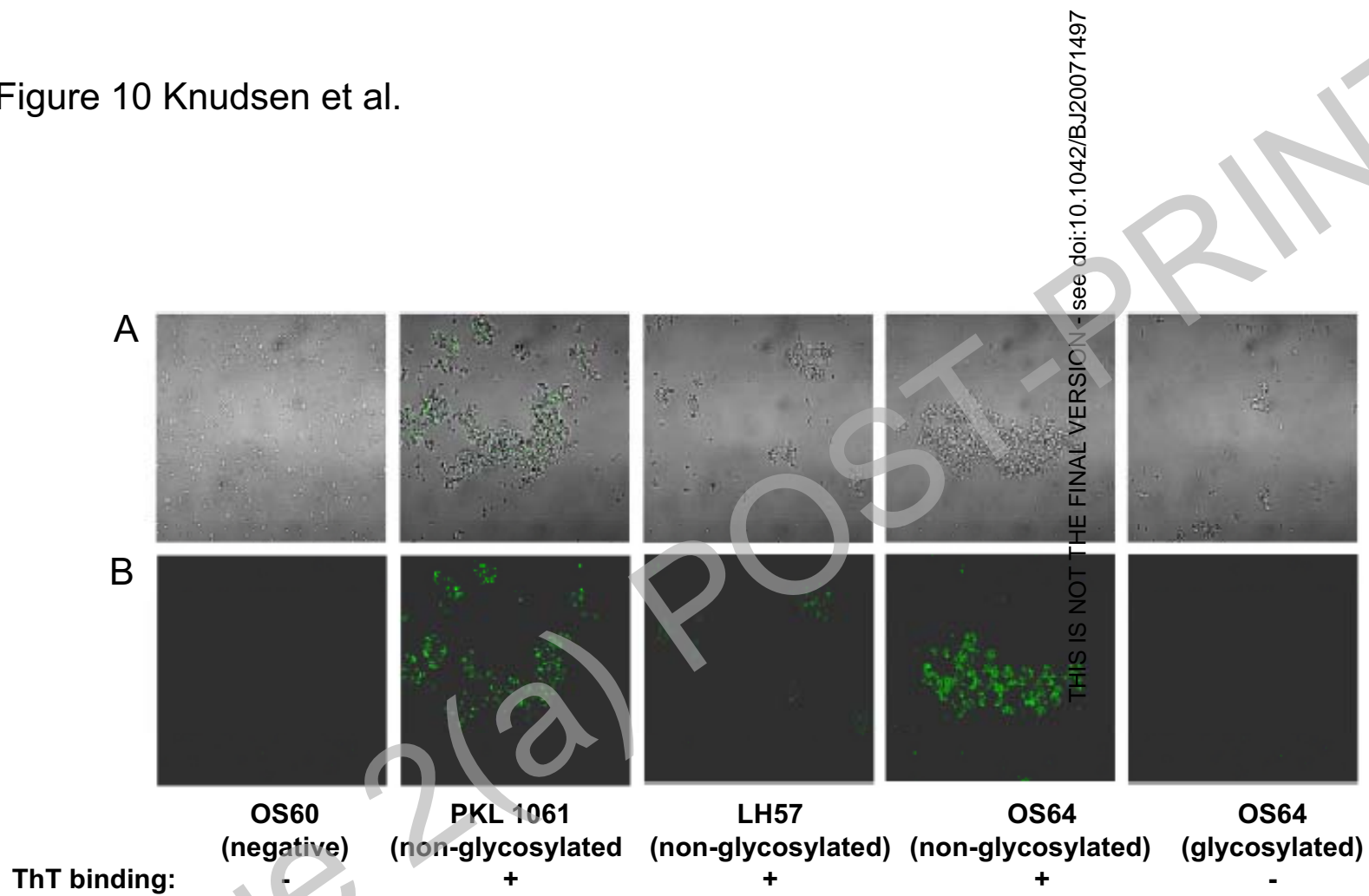


figure 11 Knudsen et al.

