HHS Public Access

Author manuscript

Proteins. Author manuscript; available in PMC 2022 November 01.

Published in final edited form as:

Proteins. 2022 May ; 90(5): 1210–1218. doi:10.1002/prot.26303.

Crystal Structure of a Human MUC16 SEA Domain Reveals Insight Into the Nature of the CA125 Tumour Marker

Brandy White[1], **Michelle Patterson**[1], **Saloni Karnwal**[1], **Cory L. Brooks**[1]

[1]Department of Chemistry and Biochemistry, California State University Fresno, 2555 E. San Ramon Ave, Fresno, CA, 93740.

Abstract

MUC16 is a membrane bound glycoprotein involved in the progression and metastasis of pancreatic and ovarian cancer. The protein is shed into the serum and the resulting cancer antigen (CA125) can be detected by immunoassays. The CA125 epitope is used for monitoring ovarian cancer treatment progression, and has emerged as a potential target for antibody mediated immunotherapy. The extracellular tandem repeat domain of the protein is composed of repeating segments of heavily glycosylated sequence intermixed with homologous SEA domains. Here we report the purification and the first X-ray structure of a human MUC16 SEA domain. The structure was solved by molecular replacement using a Rosetta generated structure as a search model. The SEA domain reacted with three different MUC16 therapeutic antibodies, confirming that the CA125 epitope is localized to the SEA domain. The structure revealed a canonical ferredoxin-like fold, and contained a conserved disulphide bond. Analysis of the relative solvent accessibility of side chains within the SEA domain clarified the assignment of N-linked and O-linked glycosylation sites within the domain. A model of the glycosylated SEA domain revealed two major accessible faces, which likely represent the binding sites of CA125 specific antibodies. The results presented here will serve to accelerate future work to understand the functional role of MUC16 SEA domains and antibody recognition of the CA125 epitope.

Keywords

Mucin; CA125; ovarian cancer; pancreatic cancer; unknown function

1. Introduction

The mucin family glycoprotein MUC16 (cancer antigen 125, CA125) is the second largest protein (~3 – 5 million Da) in the human proteome. The protein plays an important role in several malignancies, and detection of the protein in serum assays (CA125 biomarker) can be used to monitor treatment progression in ovarian cancer patients[1]. Structurally, MUC16 is a type I transmembrane protein, consisting of three regions: a C-terminal region with a short cytoplasmic tail and a single transmembrane helix; the N-terminal region consist of a tandem repeat (TR) domain and a large unstructured N-terminal domain[1]. Like all mucins,

Correspondence: cbrooks@csufresno.edu.

MUC16 is heavily glycosylated. While the protein contains both N-linked and O-linked type glycans, up to 77% of the protein's weight is derived from O-linked glycans[2].

The tandem repeat domain contains ~ 60 repeats of 156 amino acids. Unlike other mucins, the tandem repeat is interspersed with 16 homologous SEA (Sea urchin sperm protein, Enterokinase, Argin) domains ~120 residues) flanked by sequences rich in proline, serine and threonine. This PST rich sequence consists of ~ 30 residues on either side of the SEA domain[3]. The SEA domains are sequentially numbered from the N-terminus (SEA1-SEA16), with the C-terminal domains being closer to the membrane interface (SEA16 being nearest to the membrane). Phylogenetics suggested that the multitude of SEA domains found in MUC16 are the result of repeated duplication events, with the most ancient domains being located at the C-terminal end of the protein[3]. The 12 SEA domains closest to the N-terminus (SEA1 – SEA12) share the highest degree of sequence homology (>70% sequence identify), while the 4 SEA domains proximal to the membrane (SEA13–16) are rather distinct (~ 20–64 % sequence identify, supporting information: Table S1, Figure S1).

The SEA domain is glycosylated and contains three N-linked glycosylation sites and 5 O-linked glycosylation sites[4]. The PST rich sequences flanking the SEA domain are heavily O-glycosylated and are likely unstructured[4,5]. The specific function of the tandem repeat domain and associated SEA domains is unknown however, it has been hypothesized to play a role in protein-protein interactions and/or interaction with other glycans/glycoproteins[6]. The SEA domain plays an important role in the metastasis of ovarian cancer. N-linked glycans localized to the SEA domain are essential for interaction of MUC16 with mesothelin[7]. In dysregulated, tumour associated MUC16, the tandem repeat domain appears to be highly immunogenic, as immunization of mice using MUC16 isolated from cancer patient ascites results in antibodies specific for the tandem repeat domain[8].

While MUC16 was initially discovered as a tumour marker overexpressed in ovarian cancer[9], the protein has emerged as an important contributing factor in the development and progression of several other diseases including pancreatic and breast cancer[1]. Due to MUC16's central role in the metastasis and progression of pancreatic and ovarian cancer, several groups and companies have developed monoclonal antibodies which target the tandem repeat domain for immunotherapy, antibody-drug conjugates and for image-guided surgery[10–13]. However, despite the importance of human MUC16 SEA domains in the immunogenic properties of MUC16, the only structure reported to date is an NMR structure of a murine SEA domain homologue[5]. Here we report the purification and X-ray structure of a human SEA domain. We demonstrate reactivity of the purified domain with several MUC16 specific therapeutic antibodies and construct a glycosylated model of the protein. The structure presented here will accelerate future work to delineate the role of SEA domains in MUC16 function.

2. Materials and Methods

2.1 Expression and purification of MUC16 SEA5 and recombinant antibodies

The human MUC16 SEA domain was produced as a N-terminal thioredoxin (Trx) fusion protein. The gene for a 6x-His tag Trx (Uniprot code: P0AA26), a Tobacco etch virus

(TEV) protease cut site, and the human MUC16 SEA5 domain including PST rich flanking sequences (Uniprot code Q8WXI7, residues 12665–12857) was codon optimized for *E. coli* expression and produced as double stranded DNA fragments (GenParts, Genscript Inc, Piscataway, NJ). The GenParts were cloned into the plasmid vector pD451 using the ELECTRA cloning system (ATUM Bio Inc, Newark, CA), and transformed in *E. coli* BL21 (DE3). For protein expression and purification, cells were grown to an OD_{600} of 0.5 (37°C, 250 RPM) and induced overnight with isopropyl β-d1-thiogalactopyranoside [IPTG] (0.4 mM, 37°C, 250 RPM). Cells were harvested by centrifugation (5,000 × g, 10 min, 4°C) and suspended in TBS buffer (50 mM Tris pH 8.0, 0.15 M NaCl, 1 mM phenylmethylsulfonyl fluoride [PMSF]). Cells were lysed using a sonic dismembrator (Fisherbrand Model 505) and the lysate clarified by centrifugation (12,000 × g, 30 min, 4°C). The supernatant was bound to Ni-NTA resin (Thermo Fisher Scientific, Weltham, MA) in batch mode. The resin bound protein was placed into a chromatography column and the protein eluted with a step gradient of imidazole (50 mM Tris pH 8.0, 0.3 M NaCl, 0.1 – 1 M imidazole). Protein purity was confirmed by SDS-PAGE.

For purification of the SEA domain, the fusion protein was digested using TEV protease[14] and purified using reverse Ni-NTA chromatography. Briefly, digested fusion protein was dialyzed (50 mM Tris, 0.3 M NaCl) and batch bound to Ni-NTA resin as described above. The flow-through and wash fractions containing the SEA domain were pooled and purity confirmed by SDS-PAGE.

Anti-MUC16 monoclonal antibodies AR9.6[15], 3A5[13] and H1H8794P2[12] were produced by transient transfection in expiCHO cells as described previsouly[16]. Briefly, genes coding for IgG heavy and light chains were produced by gene synthesis and cloned into pcDNA3.1 and transfected into expiCHO cells according to the manufacturer's instruction (Thermo Fisher Scientific, Weltham, MA). Culture supernatant was harvested when cell viability decreased below 80% and IgG was purified by Protein A affinity chromatography.

2.2 Enzyme linked immunosorbent assay

MUC16 fragments (Trx-1.2TR, SEA5) and Trx (negative control), (150 ng/well) were immobilized on ELISA plates (MaxiSorp, Immulon 4 HBX, Thermo Scientific, Weltham, MA) and blocked overnight (4°C) with bovine serum albumin (1% in PBS). Recombinant primary human anti-MUC16 antibodies (AR9.6, 3A5, H1H8794P2) and a human IgG isotype control (Genscript Inc, Piscataway, NJ) were 5-fold serially diluted (starting concentration of 50 μg/ml, 333 nM) and applied to the plate. Primary antibodies were detected by the addition of HRP-conjugated goat anti-human kappa secondary antibody (1:40,000 dilution, Novus Biologicals, Littleton, CO). For signal development, TMB substrate (TMB Ultra, Thermo Fisher, Weltham, MA) was added and the reaction was stopped with the addition of 0.18 M H_2SO_4. Absorbances at 450 and 540nm were measured using a 96 well plate reader (BioTek Synergy HT, Gen5 2.04); blanks were subtracted. Binding was plotted against the log concentration of antibody and the data was fit using a 4-parameter logistic curve and IC_{50} values determined from the fit (GraphPad Prism 9). All trials were performed in triplicate.

2.3 Crystallization, X-ray data collection, structure determination and refinement

The purified MUC16 SEA domain was concentrated to 10 mg/ml and crystal screening carried out in 96 well sitting drop plates using commercially available sparse matrix screens (Index HT and PEGRx HT, Hampton Research, Aliso Viejo, CA). Initial crystals appeared in several conditions in both screens. Crystals were optimized by hanging drop vapor diffusion using 24 well plates. Large, diffraction quality crystals appeared in 0.2 M ammonium acetate, 0.1 M Tris (pH 6.0 −9.0) and PEG 10,000 (12 – 18%). Crystals were flash frozen in liquid nitrogen and X-ray data was collected at the 08ID-1 Canadian Light Source (CLS) beamline[17]. X-ray data was processed using Xia2[18]. The structure was solved by molecular replacement with Phaser as implemented in Phenix[19], using a search model generated by Rosetta[20]. Model building and refinement were performed with Phenix and Coot[19,21]. Final model and X-ray data statistics are given in Table I.

2.4 Computing solvent accessibility of putative glycosylated residues

The absolute Accessible Surface Area (ASA) of each putative glycosylation site residue was computed using the DSSP program[22]. The Relative Solvent Accessibility (rASA) values were obtained dividing absolute ASA values in $Å^2$ by residue-specific maximal accessibility values, as extracted using the scale described by Miller[23]. Residues were classified as buried if the rASA was below 20%, and exposed (E) if above the 20% threshold.

3. Results and Discussion

3.1 Purification of human SEA5 domain

Here, the SEA5 domain of MUC16 was chosen as a model to study MUC16 SEA domains as it had been previously characterized in terms of glycosylation as well as reactivity to several MUC16/CA125 antibodies[4]. Furthermore, the SEA5 domain is a suitable representative of the MUC16 SEA domains generally, as it is >80% homologous to SEA domains further from the membrane (SEA1–12, supporting information, Table S1).

Previously reported efforts to produce recombinant human MUC16 SEA domains in *E. coli* resulted in the protein localized to inclusion bodies, which required solubilization in chaotropic agents[4,10,24]. Similarly, our initial attempts to produce the protein using a pET system vector resulted in accumulation of the protein in inclusion bodies. Efforts to refold the protein from urea solubilized inclusion bodies were unsuccessful. The sequence of the human MUC16 SEA5 domain (Uniprot code Q8WXI7, residues 12697–12818) contains 3 cysteine residues, two of which are conserved[25]. The presence of free cysteines and/or disulphide bonds in the SEA domain may be a significant contributing factor to the difficulty in producing properly folded protein in the reducing environment of the *E. coli* cytoplasm. To overcome this challenge, a thioredoxin (TrxA) fusion protein strategy was employed. This approach has been found to reduce protein accrual in inclusion bodies[26]. To this end, 1.2 copies of a MUC16 tandem repeat (1.2 TR), which included the SEA5 domain and flanking regions rich in PST residues (Uniprot code, Q8WXI7, residues 12665–12857) were produced as an N-terminal TrxA fusion, separated by a TEV cut site. The fusion protein (called Trx-1.2TR) resulted in soluble expression in the *E. coli* cytoplasm with a yield of 30 mg per L of culture. Following TEV protease digestion, the MUC16 1.2TR fragment

was purified from the Trx tag using reverse Ni-NTA affinity chromatography (Figure 1A). Interestingly, once the 1.2TR protein was separated from Trx, the fusion protein was susceptible to slow proteolytic degradation. A time course proteolytic stability experiment revealed that following TEV protease digestion, the molecular weight of the protein shifted from ~22 kDa (corresponding to the 1.2TR) to ~ 14 kDa (corresponding to the SEA domain) sometime between 24- and 48-hours (Figure 1A). Although autoproteolytic activity has been reported for some mucin SEA domains including MUC1[27], the MUC16 SEA5 domain lacks the GSVV proteolytic cleavage motif and thus is an unlikely explanation for the observed proteolysis. Based on the resultant size of the fragment, the PST rich sequences flanking the SEA5 domain were possibly cleaved by residual protease activity of contaminating proteases, resulting in a stable SEA domain. The resulting x-ray structure (section 3 below) confirmed this hypothesis.

3.2 Binding of human SEA5 domain to MUC16 therapeutic antibodies.

MUC16 has emerged as a potential target for antibody mediated immunotherapy for the treatment of several malignancies including ovarian and pancreatic cancer[1]. The majority of the potential therapeutic antibodies developed to date bind the tandem repeat domain, targeting the CA125 epitope[1]. While the specific nature of the CA125 epitope is not entirely clear, it may contain at least a portion of an intact SEA domain[4,24]. To determine if a single fully folded human SEA domain is sufficient to bind distinct MUC16 antibodies, an indirect ELISA was carried out to measure the binding of three MUC16 specific antibodies (AR9.6, 3A5, and H1H8794P2) to Trx1.2TR and SEA5 (Figure 1B). Each of these three MUC16 antibodies has previously been shown to bind MUC16[10,12,13]. AR9.6 is under investigation as both an immunotherapy and tumour imaging agent[11,15,16]. Antibody 3A5 has undergone a phase I clinical trial as an antibody-drug conjugate for the treatment of ovarian cancer[28], while H1H8794P2 is derived from a study developing a T-cell engaging bispecific antibody for the treatment of ovarian cancer[12,29]. All three mAbs bound both the Trx-1.2TR and SEA5. In all three cases, the antibodies displayed ~2-fold higher affinity for the SEA5 domain compared to Trx-1.2TR, possibly due to steric interference from the Trx tag. The specific epitopes recognized by these antibodies is not reported in the literature, however, all three antibodies were generated by immunizations using different sources of MUC16. Antibody AR9.6 was generated by immunization with MUC16 purified from the ascites of an ovarian cancer patient. 3A5 was generated by immunization using a recombinant MUC16 fragment containing domains SEA5-SEA9. Similarly, H1H8794P2 was generated using a recombinant MUC16 fragment contain SEA domains 12–16. Despite differences in the antigen used for immunization, all three antibodies cross-reacted with this specific SEA5 domain, albeit with differing apparent affinities (Figure 1B). These results would seem to support the hypothesis that the CA125 is conformational in nature[4] and that this conformational epitope is at least partially conserved in different SEA domains.

3.3 Structure determination of MUC16 SEA5 domain

Initially, molecular replacement (MR) was attempted using the structure of the homologous murine MUC16 SEA domain, which had been determined using NMR spectroscopy (PDB code:1IVZ)[5]. This approach yielded no clear MR solution, with more than 30 partial solutions with top TFZ scores of only ~5. The murine SEA domain is 37% identical to

that of the human SEA5 domain, which approaches the limitation of ~30–35% sequence homology for a suitable search model for MR. An additional complication in using the murine MUC16 SEA domain as a MR search model is that NMR models themselves have traditionally been difficult to use for MR[30]. Similarly, the recently published structure of the MUC1 SEA domain could not be determined by molecular replacement using an NMR model, even with complete sequence identity[27]. Although NMR models historically have been difficult to use as MR models, refinement with Rosetta has been found to greatly improve the phasing power of NMR models[31]. Indeed, Rosetta has emerged as a powerful tool to facilitate MR of difficult targets[32].

To solve the structure of the human SEA5 domain, models were generated using the Robetta server implementing the trRosetta modelling protocol[20]. The resulting top model had a confidence score of 0.82. Using the default Rosetta model resulted in a failure of the top MR solutions to pack. To solve this problem, the Rosetta generated model was modified by deleting residues with error estimates greater than 2 Å (Figure 2). This approach yielded a clear MR solution with a TFZ of 9.2 and LLG 79. Subsequent auto-building and refinement resulted in the final structure of the protein (Table I) and highlights the utility of Rosetta in determining X-ray structures by molecular replacement.

3.4 Structure of MUC16 SEA Domain

The structure of the human MUC16 SEA5 domain crystallized in the monoclinic space group C2 and diffracted to a resolution of 1.8 Å (Table I). The structure contained two molecules in the asymmetric unit, arranged in a head-to-tail fashion (Figure 3A). As expected from the observed proteolysis during purification of the protein, the structure lacked any residues from the flanking PST rich loop regions (residues 12665–12694; 12822–12857) and consisted solely of the core SEA domain (Uniprot residues 12695–12821, structure residue numbering 35–160). The structure contained a short-disordered region consisting of the loop connecting β3-α2 (residues 114–117). The overall structure conforms to that of a canonical SEA domain, consisting of a Ferredoxin-like fold of a α/β sandwich (Figure 3B). The structure contained two anti-parallel β-sheets. The first sheet is composed of four strands, three long β-strands (residues 39–47, 87–99, 104–114) and a short strand of only four residues (residues 152–154). There are two classic anti-parallel β-bulges (residues 89–90, and 94–95) in the first β-sheet, which introduces a major twist in the 3rd strand of the sheet. The second β-sheet is composed of only two very short strands, each consisting of three residues (residues 139–141, 144–146). Each of the β-sheets contain β-hairpins. The first hairpin is class 4:4 (residue 100–103) and connects the second and third strands of β-sheet A. The second hairpin is class 2:2 IIP (residues 142–143) and connects the two short strands in β-sheet B. The structure also contains two longer α-helices (20 and 13 residues long, residues 61–81, 123–135) that pack against each other. The structure also contains three shorter α-helical insertions (residues 53–58, 80–84, and 100–103) (Figure 3B).

The SEA5 domain contains three cysteine residues. One of the cysteine residues (Cys58) is located in a solvent exposed loop, and is unique to SEA5 among the 16 homologous SEA domains found in MUC16. The remaining two cysteine residues (Cys91, Cys11) are

conserved in all 16 human MUC16 SEA domains,[25] and form a disulphide bond, which connect strands β2 and β3 (Figure 3B).

3.5 Comparison of the human and murine MUC16 SEA domains

To date the only other reported MUC16 SEA structure is an NMR structure of a murine SEA homologue[5]. At the primary sequence level, the murine homologue is 37% identical to the human SEA5 domain reported here. Unsurprisingly, the overall architecture of the two proteins is very similar, with an alignment of Cα atoms yielding an r.m.s.d. of 1.36 Å. However, despite the global similarity of the two proteins, there were small structural differences. The human SEA domain contained two short α-helical insertions between strands β2 and β3 (residues 100–10) and helix-α2 and strand β3 (residues 80–84). These regions are loop structures in the murine SEA domain (Figure 4A). The central β-sheet in the murine SEA domain is also different, as it contains 5 strands *vs.* 4 strands in the human domain. The length of the strands within the sheet are generally shorter in the murine domain compared to the human domain (Figure 4A).

A distinguishing feature identified in the murine SEA domain was the “TY-turn”, a tight type I β-turn which connected the α1 - α2 helices. In the murine domain, the turn was composed of 4 residues (Gln31, Pro32, Ser33, Thr34). The conserved Thr residue (Thr33) hydrogen bonded with a conserved Tyr (Tyr37) residue on α2, capping the helix (hence the designation TY-turn). The second residue in the turn (Pro32) stacked with the side chain of Tyr37 and was thought to stabilize the turn (Figure 4B)[5]. A similar feature exists in the human SEA domain structure (Cys58, Pro59, Gly60, Ser61), albeit with small differences. The structure of the turn was that of a type II β-turn rather than the type I found in the murine domain. The distinguishing Thr and Tyr residues found in the murine domain were conservatively substituted by Ser and Phe (Ser61, Phe64) respectively. The side chain of Ser61 serves to cap the helix by interacting with the amide nitrogen of Phe64. Analogous to the murine structure, the turn is stabilized by a stacking interaction between Pro59 and Phe64 (Figure 4B). Overall, differences between the murine NMR structure and the X-ray structure presented here are rather small. They may represent subtle variation in structure, or in part may be due to crystal packing vs. the in-solution NMR structure. Regardless, the structures point to the general conservation of the fold despite a low degree of sequence homology.

3.6 Glycosylation of SEA5

As a mucin protein, the glycosylation of MUC16 is critical for the structure and function of the protein. The tandem repeat domain contains both N-linked and O-linked glycosylation sites within the SEA domain and flanking PST regions. Glycosylation of the human 1.2TR (SEA5 domain, and flanking PST rich residues) was recently investigated in a proteomic study, where the protein was produced in a CHO cell line with a *cosmc* knock-out, to produce truncated O-glycans[4]. This work confirmed that the SEA domain contained 3 N-linked glycosites, and 5 O-linked glycosites. The three N-linked glycans were unambiguously identified (Asn residues 12704, 12725, 12741). Of the 5 O-linked glycosites, three were unambiguously identified (Thr12702, Ser12721, Ser12758), another was located at one of two residues (Thr12726 or Thr12727). The 5th O-glycosite could only be assigned

to a peptide consisting of residues 12760–12774, a stretch of sequence which contains several possible O-linked glycosylation sites[4]. As glycosylation occurs post-translationally, the side-chains of modified residues must be accessible to solvent; at least in the case of N-linked glycosylation, this has been shown to be the case[33,34].

In an attempt to clarify the ambiguous assignment of the glycosylation sites in the SEA5 domain, the Relative Solvent Accessibility (rASA) was calculated for the ambiguous and unambiguously assigned glycosites from both molecules in the asymmetric unit of the crystal structure (Table II). As expected, the N-linked glycosylated residues (Uniprot residues 12704, N12725, N12741; structure residues 44, 65, 81) exhibited a rARA greater than 20%, and were likely solvent exposed. Interestingly, while two of unambiguously assigned O-linked glycosylation sites (Uniprot residues 12702, 12758, structure residues 42, 98) displayed a rARA greater than 20%, one of the sites (Uniprot residue 12721, structure residue 61) displayed borderline solvent accessibility (Chain A), or was likely buried (Chain B) (Table II). This residue is located in the conserved TY-turn region (Figure 4B). It is difficult to reconcile the clearly unambiguous assignment of the glycosite in the mass spectrometry data, and the solvent accessibility of the residue. One possible explanation is that the glycosylated protein may exhibit structural changes which further exposes the side chain of Ser61 for glycosylation.

Based on the rASA values, we could infer the location of the remaining two O-linked glycosylation sites which could not be unambiguously identified in the mass spectrometry study. Between the pair of residues Thr12726/Thr12727, only residue Thr12726 was solvent exposed, and is thus likely the glycosylation site. In the region of 12760–12774 there are two Thr residues, only one of which (T12772) was solvent exposed. Taken together, the likely N-linked glycosylation sites are N12704, N12725, N12741 and the five likely O-linked glycosylation sites are T12702, S12721, T12726, S12758 and T12772 (Table II).

Like all glycoproteins, the glycosylation of MUC16 is complex and highly heterogenous. The N-linked glycans of MUC16 isolated from OVCAR-3 cells contained ~80% complex type N-linked glycans and 20% high mannose type[35]. The most abundant of the complex N-linked glycans were mono-fucosylated, bisected bi-antennary[35]. Similarly, a survey of the glycosylation state of MUC16 isolated from serum found an increase in fucosylated bi-antennary monosialylated N-linked glycans in cancer sera[36]. The O-linked glycans were mono/di-sialylated core type I and core type II[35] (Figure 5A). Circulating MUC16 in ovarian cancer sera carries STn type O-glycans[37] (Figure 5A). Based on this we constructed a simple model of glycosylated SEA5 using the GLYCAM force field[38]. The structural model included one of the most abundantly found complex N-linked glycans[35] and the STn antigen (Figure 5B). The resulting glycosylated model provides some useful insights into the nature of the CA125 epitope.

3.7 The CA125 epitope and SEA5

Despite the discovery of the CA125 epitope more than 40 years ago, little is known regarding the specific epitope recognized by CA125 mAbs. The binding of CA125 antibodies had been broadly categorized into three epitope bins based on competitive ELISA studies. These epitope bins were named after prototypic antibodies and include: OC125-like;

M11-like; and OV197-like[39,40]. It has recently been shown that the epitope recognized by mAbs OC125 and M11 is complex and conformational in nature, also importantly that antibody binding was unaffected by both N-linked and O-linked glycosylation[4].

In our model of the glycosylated SEA domain, the carbohydrates form a “glycan belt” across the narrow edge of the SEA domain (Figure 5C). This glycan belt may limit the interaction surface available for antibody binding to the SEA protein. The glycan distribution results in 3 accessible faces on the SEA domain which may correspond to antibody epitope regions. One of the faces would form the interface with an adjacent SEA domain in the native MUC16 structure. The linker sequence connecting the two SEA domains is heavily O-glycosylated, and thus this face would likely be inaccessible in an intact MUC16 protein, and seems to be an implausible antibody binding site. The other two faces represent possible antibody epitope locations (Figure 5D). Face A contains residues from the 2nd β-hairpin, the first long α-helix, and loops connecting β strand 1 and the long α-helix. Face B is composed of residues in the TY-turn as well as the β-sheet. Based on this analysis we hypothesise that these two faces likely contain the three CA125 epitope types, Face A contains the larger surface area, and may house two epitope locations, while Face B may contain the third epitope. Future efforts will focus on obtaining antibody X-ray structures in complex with the SEA domain to further clarify and delineate the CA125 epitope.

4.0 Conclusion

This work provided the first high-resolution X-ray crystal structure of an SEA domain derived from the cancer antigen MUC16. The structure provides new insights into the nature of the CA125 cancer antigen. Interestingly, three different MUC16 antibodies, all isolated using different MUC16 antigens bound recombinant SEA5. This suggests the CA125 epitope may be conserved across homologous SEA domains found in the MUC16 tandem repeat. Initial attempts to solve the structure by molecular replacement using a homologous murine NMR structure were unsuccessful. However, a homology model generated with Rosetta resulted in a clear solution, highlighting the utility of machine-learning derived structural models for molecular replacement. Using a combination of previously reported glycosylation data from a mass-spectroscopy study[4] and the solvent accessibility of residues calculated from the structure, a model of the fully glycosylated SEA domain was constructed. This model revealed two accessible faces which may represent sites of antibody interaction. Future work will focus on obtaining antibody structures in complex with the SEA domain to further clarify and delineate the specific nature of the CA125 epitope.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding information:

Research reported in this publication/was supported by National Cancer Institute of the National Institutes of Health under award number **R15CA242349**

5. References

1. Aithal A, Rauth S, Kshirsagar P, et al. MUC16 as a novel target for cancer therapy. Expert opinion on therapeutic targets. 2018;22(8):675–686. [PubMed: 29999426]
2. Lloyd KO, Yin BW, Kudryashov V. Isolation and characterization of ovarian cancer antigen CA 125 using a new monoclonal antibody (VK-8): identification as a mucin-type molecule. International journal of cancer. 1997;71(5):842–850. [PubMed: 9180155]
3. Duraisamy S, Ramasamy S, Kharbanda S, Kufe D. Distinct evolution of the human carcinoma-associated transmembrane mucins, MUC1, MUC4 AND MUC16. Gene. 2006;373:28–34. [PubMed: 16500040]
4. Marcos-Silva L, Narimatsu Y, Halim A, et al. Characterization of binding epitopes of CA125 monoclonal antibodies. J Proteome Res. 2014;13(7):3349–3359. [PubMed: 24850311]
5. Maeda T, Inoue M, Koshiba S, et al. Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16). J Biol Chem. 2004;279(13):13174–13182. [PubMed: 14764598]
6. Haridas D, Ponnusamy MP, Chugh S, Lakshmanan I, Seshacharyulu P, Batra SK. MUC16: molecular analysis and its functional implications in benign and malignant conditions. Faseb j. 2014;28(10):4183–4199. [PubMed: 25002120]
7. Gubbels JA, Belisle J, Onda M, et al. Mesothelin-MUC16 binding is a high affinity, N-glycan dependent interaction that facilitates peritoneal metastasis of ovarian tumors. Mol Cancer. 2006;5(1):50. [PubMed: 17067392]
8. Nustad K, Bast RC Jr., Brien TJ, et al. Specificity and affinity of 26 monoclonal antibodies against the CA 125 antigen: first report from the ISOBM TD-1 workshop. International Society for Oncodevelopmental Biology and Medicine. Tumour Biol. 1996;17(4):196–219. [PubMed: 8685601]
9. Bast RC Jr., Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC. Reactivity of a monoclonal antibody with human ovarian carcinoma. The Journal of clinical investigation. 1981;68(5):1331–1337. [PubMed: 7028788]
10. Thomas D, Sagar S, Liu X, et al. Isoforms of MUC16 activate oncogenic signaling through EGF receptors to enhance the progression of pancreatic cancer. Mol Ther. 2020.
11. Olson MT, Wojtynek NE, Talmon GA, et al. Development of a MUC16-Targeted Near-Infrared Fluorescent Antibody Conjugate for Intraoperative Imaging of Pancreatic Cancer. Mol Cancer Ther. 2020;19(8):1670–1681. [PubMed: 32404409]
12. Crawford A, Haber L, Kelly MP, et al. A Mucin 16 bispecific T cell-engaging antibody for the treatment of ovarian cancer. Sci Transl Med. 2019;11(497).
13. Chen Y, Clark S, Wong T, et al. Armed antibodies targeting the mucin repeats of the ovarian cancer antigen, MUC16, are highly efficacious in animal tumor models. Cancer Res. 2007;67(10):4924–4932. [PubMed: 17510422]
14. Raran-Kurussi S, Cherry S, Zhang D, Waugh DS. Removal of Affinity Tags with TEV Protease. Methods Mol Biol. 2017;1586:221–230. [PubMed: 28470608]
15. Thomas D, Sagar S, Liu X, et al. Isoforms of MUC16 activate oncogenic signaling through EGF receptors to enhance the progression of pancreatic cancer. Mol Ther. 2021;29(4):1557–1571. [PubMed: 33359791]
16. Sharma SK, Mack KN, Piersigilli A, et al. ImmunoPET of Ovarian and Pancreatic Cancer with AR9.6, a Novel MUC16-Targeted Therapeutic Antibody. Clin Cancer Res. 2021.
17. Grochulski P, Fodje MN, Gorin J, Labiuk SL, Berg R. Beamline 08ID-1, the prime beamline of the Canadian Macromolecular Crystallography Facility. J Synchrotron Radiat. 2011;18(Pt 4):681–684. [PubMed: 21685687]
18. Winter G, Lobley CM, Prince SM. Decision making in xia2. Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 7):1260–1273. [PubMed: 23793152]
19. Liebschner D, Afonine PV, Baker ML, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol. 2019;75(Pt 10):861–877. [PubMed: 31588918]

20. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. Proc Natl Acad Sci U S A. 2020;117(3):1496–1503. [PubMed: 31896580]
21. Casañal A, Lohkamp B, Emsley P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. Protein Sci. 2020;29(4):1069–1078. [PubMed: 31730249]
22. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22(12):2577–2637. [PubMed: 6667333]
23. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol. 1987;196(3):641–656. [PubMed: 3681970]
24. Bressan A, Bozzo F, Maggi CA, Binaschi M. OC125, M11 and OV197 epitopes are not uniformly distributed in the tandem-repeat region of CA125 and require the entire SEA domain. Disease markers. 2013;34(4):257–267. [PubMed: 23396293]
25. O'Brien TJ, Beard JB, Underwood LJ, Dennis RA, Santin AD, York L. The CA 125 gene: an extracellular superstructure dominated by repeat sequences. Tumour Biol. 2001;22(6):348–366. [PubMed: 11786729]
26. LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. Biotechnology (N Y). 1993;11(2):187–193. [PubMed: 7763371]
27. Noguera ME, Jakoncic J, Ermácora MR. High-resolution structure of intramolecularly proteolyzed human mucin-1 SEA domain. Biochim Biophys Acta Proteins Proteom. 2020;1868(3):140361. [PubMed: 31923589]
28. Liu JF, Moore KN, Birrer MJ, et al. Phase I study of safety and pharmacokinetics of the anti-MUC16 antibody-drug conjugate DMUC5754A in patients with platinum-resistant ovarian cancer or unresectable pancreatic cancer. Ann Oncol. 2016;27(11):2124–2130. [PubMed: 27793850]
29. Haber L, Smith E, Kelly M, Kirshner JR, Coetzee S, Crawford A, Nittoli T, Liu Y, Inventor; Regeneron Pharmaceuticals Inc, assignee. Anti-MUC16 antibodies, antibody-drug conjugates and sispeficic antigen binding molecules that bind MUC16 and CD3 and uses therof. US patent US20200317810A12020.
30. Chen YW. Solution solution: using NMR models for molecular replacement. Acta Crystallogr D Biol Crystallogr. 2001;57(Pt 10):1457–1461. [PubMed: 11567160]
31. Mao B, Tejero R, Baker D, Montelione GT. Protein NMR structures refined with Rosetta have higher accuracy relative to corresponding X-ray crystal structures. J Am Chem Soc. 2014;136(5):1893–1906. [PubMed: 24392845]
32. DiMaio F Rosetta Structure Prediction as a Tool for Solving Difficult Molecular Replacement Problems. Methods Mol Biol. 2017;1607:455–466. [PubMed: 28573585]
33. Lee LY, Lin CH, Fanayan S, Packer NH, Thaysen-Andersen M. Differential site accessibility mechanistically explains subcellular-specific N-glycosylation determinants. Front Immunol. 2014;5:404. [PubMed: 25202310]
34. Suga A, Nagae M, Yamaguchi Y. Analysis of protein landscapes around N-glycosylation sites from the PDB repository for understanding the structural basis of N-glycoprotein processing and maturation. Glycobiology. 2018;28(10):774–785. [PubMed: 29931153]
35. Kui Wong N, Easton RL, Panico M, et al. Characterization of the oligosaccharides associated with the human ovarian tumor marker CA125. J Biol Chem. 2003;278(31):28619–28634. [PubMed: 12734200]
36. Saldova R, Struwe WB, Wynne K, Elia G, Duffy MJ, Rudd PM. Exploring the glycosylation of serum CA125. Int J Mol Sci. 2013;14(8):15636–15654. [PubMed: 23896595]
37. Chen K, Gentry-Maharaj A, Burnell M, et al. Microarray Glycoprofiling of CA125 improves differential diagnosis of ovarian cancer. J Proteome Res. 2013;12(3):1408–1418. [PubMed: 23360124]
38. Kirschner KN, Yongye AB, Tschampel SM, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem. 2008;29(4):622–655. [PubMed: 17849372]

39. Nap M, Vitali A, Nustad K, et al. Immunohistochemical characterization of 22 monoclonal antibodies against the CA125 antigen: 2nd report from the ISOBM TD-1 Workshop. Tumour Biol. 1996;17(6):325–331. [PubMed: 8938947]

40. Nustad K, Bast RC Jr., Brien TJ, et al. Specificity and affinity of 26 monoclonal antibodies against the CA 125 antigen: first report from the ISOBM TD-1 workshop. International Society for Oncodevelopmental Biology and Medicine. Tumour Biol. 1996;17(4):196–219. [PubMed: 8685601]

41. Mehta AY, Cummings RD. GlycoGlyph: a glycan visualizing, drawing and naming application. Bioinformatics. 2020;36(11):3613–3614. [PubMed: 32170934]

42. Varki A, Cummings RD, Aebi M, et al. Symbol Nomenclature for Graphical Representations of Glycans. Glycobiology. 2015;25(12):1323–1324. [PubMed: 26543186]

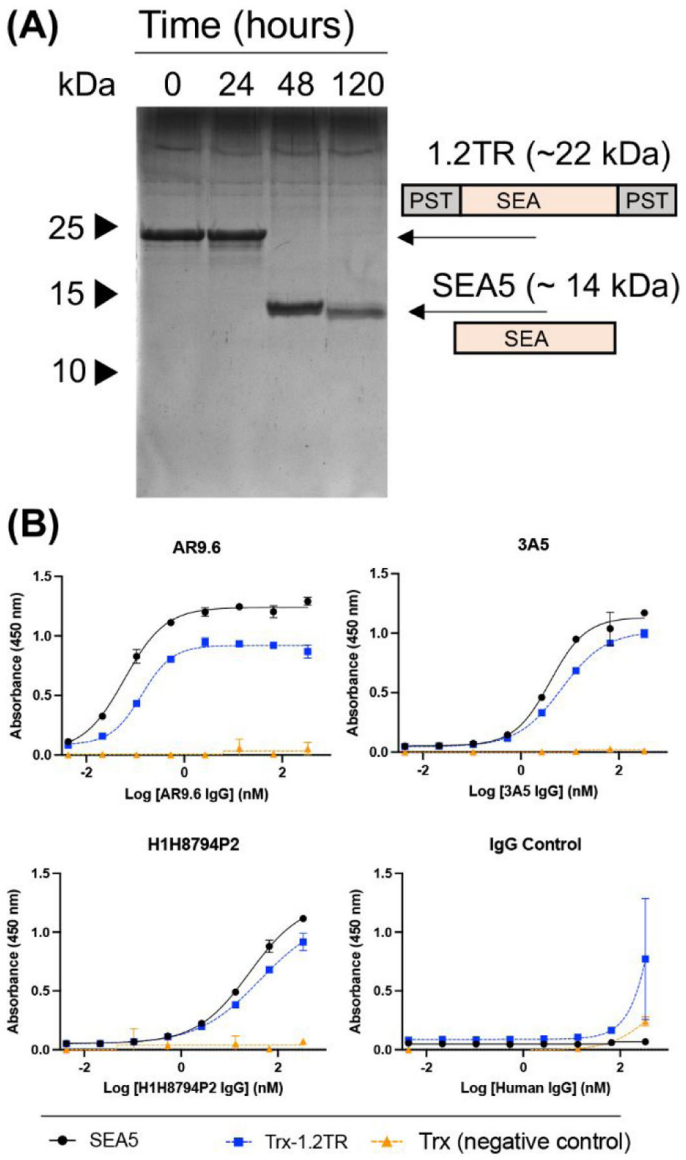

Figure 1 –. Stability and Reactivity of Purified Human MUC16 SEA5 Domain.

(A) Time course stability of purified MUC16 1.2TR (SEA5 domain with flanking PST rich sequences) using purified MUC16 SEA domain run on a 14% acrylamide gel stained with Coomassie blue. The 1.2TR protein (~ 22 kDa) undergoes slow proteolytic degradation producing a stable SEA domain (~ 14 kDa) within 48 hours. (**B**) Indirect ELISA confirming reactivity of three MUC16/CA125 antibodies (3A5, AR9.6, H1H8794) with both the 1.2TR and SEA5 proteins. Purified thioredoxin (Trx) was used as an irrelevant coating protein control, and a human IgG isotype control was used as a negative antibody reactivity control.

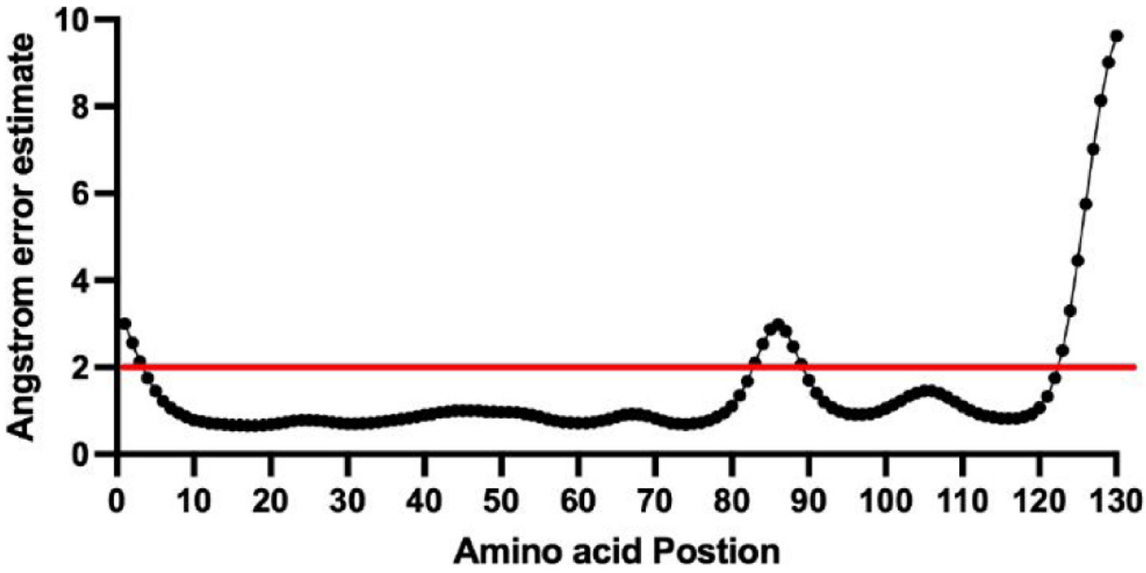

Figure 2 –. Error Estimate of Rosetta Produced MUC16 SEA5 Model for Molecular Replacement.

Truncation of residues in the model with an error estimate greater than 2 Å resulted in successful molecular search model.

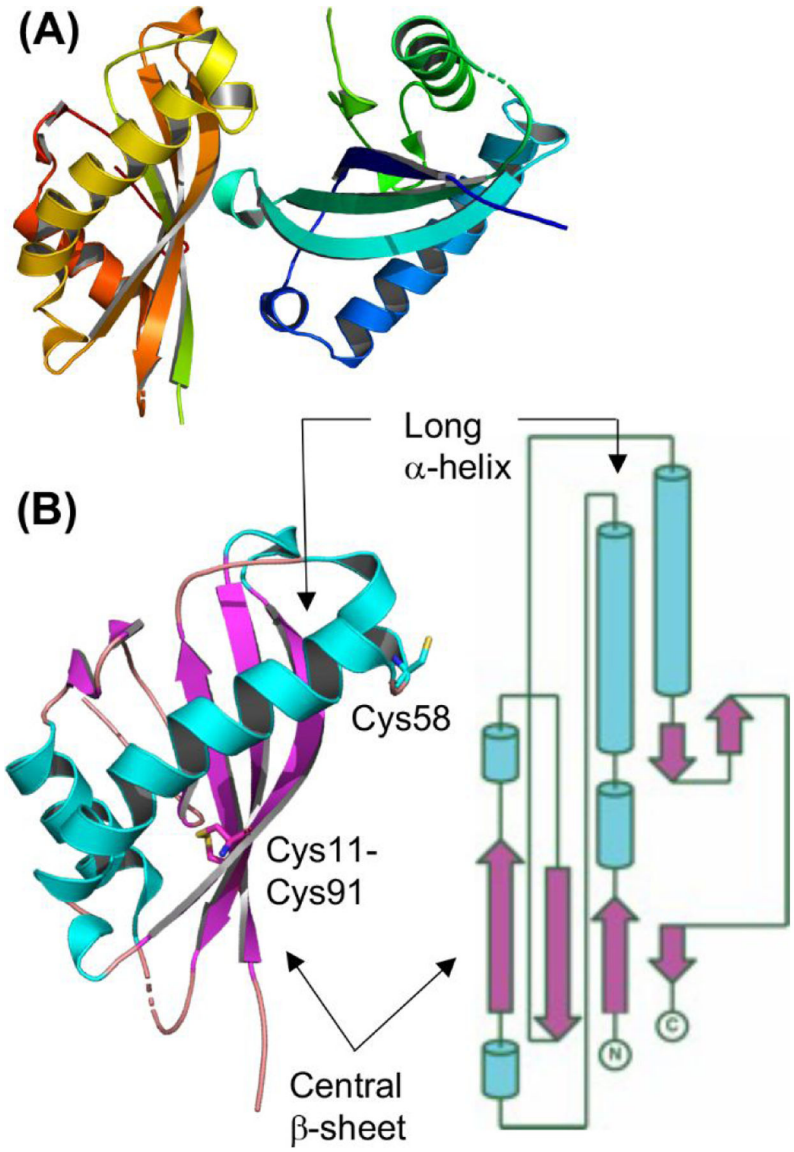

Figure 3 –. Structure of the Human MUC16 SEA5 Domain.

(A) The structure contained two molecules in the asymmetric unit arranged in a twisted head-to-tail fashion (**B**) Topology of the human SEA5 domain.

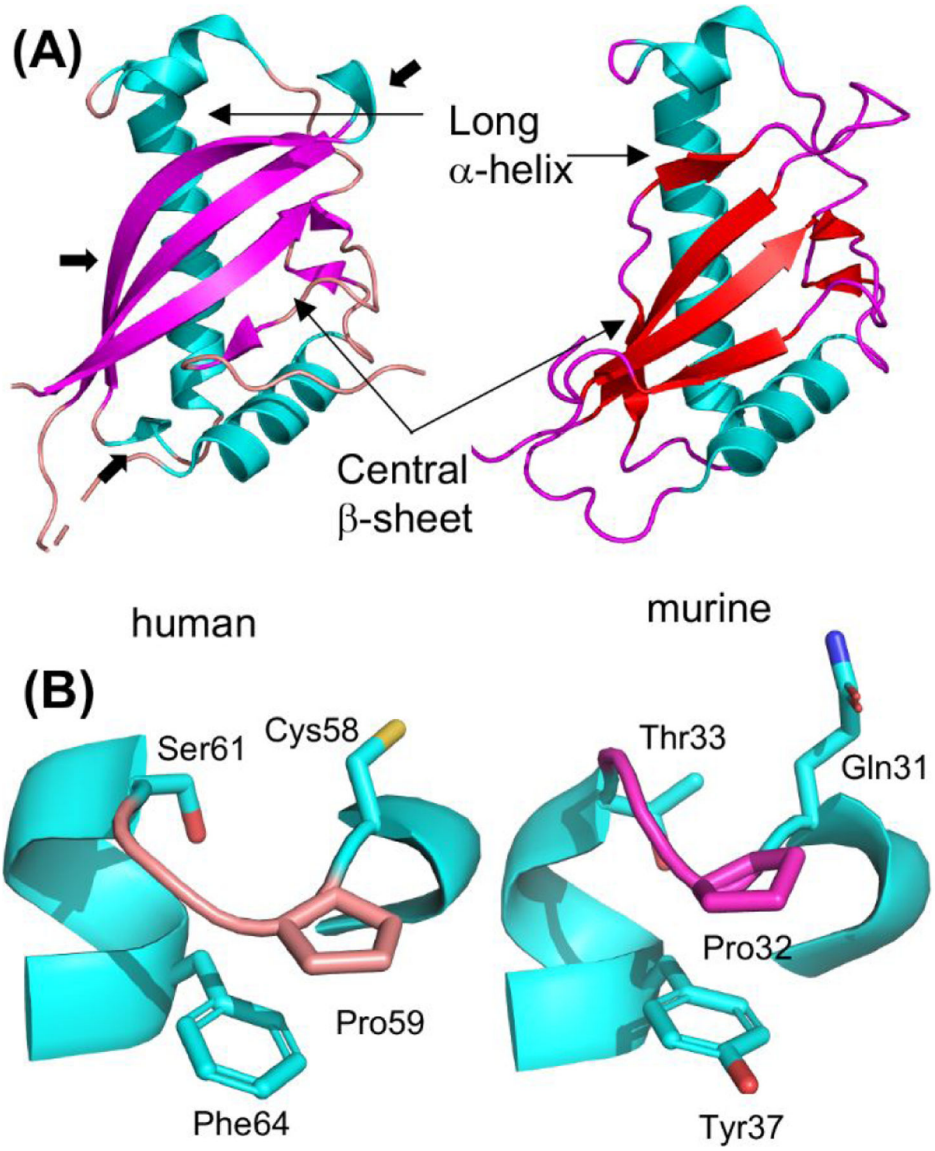

Figure 4 –. Comparison of Human and Murine MUC16 SEA Domains.

(**A**) The human SEA5 domain contains a 5 stranded β-sheet with generally longer strands, and two small α-helical insertions (arrows). (**B**) The TY-turn is generally conserved between the human and murine SEA domains.

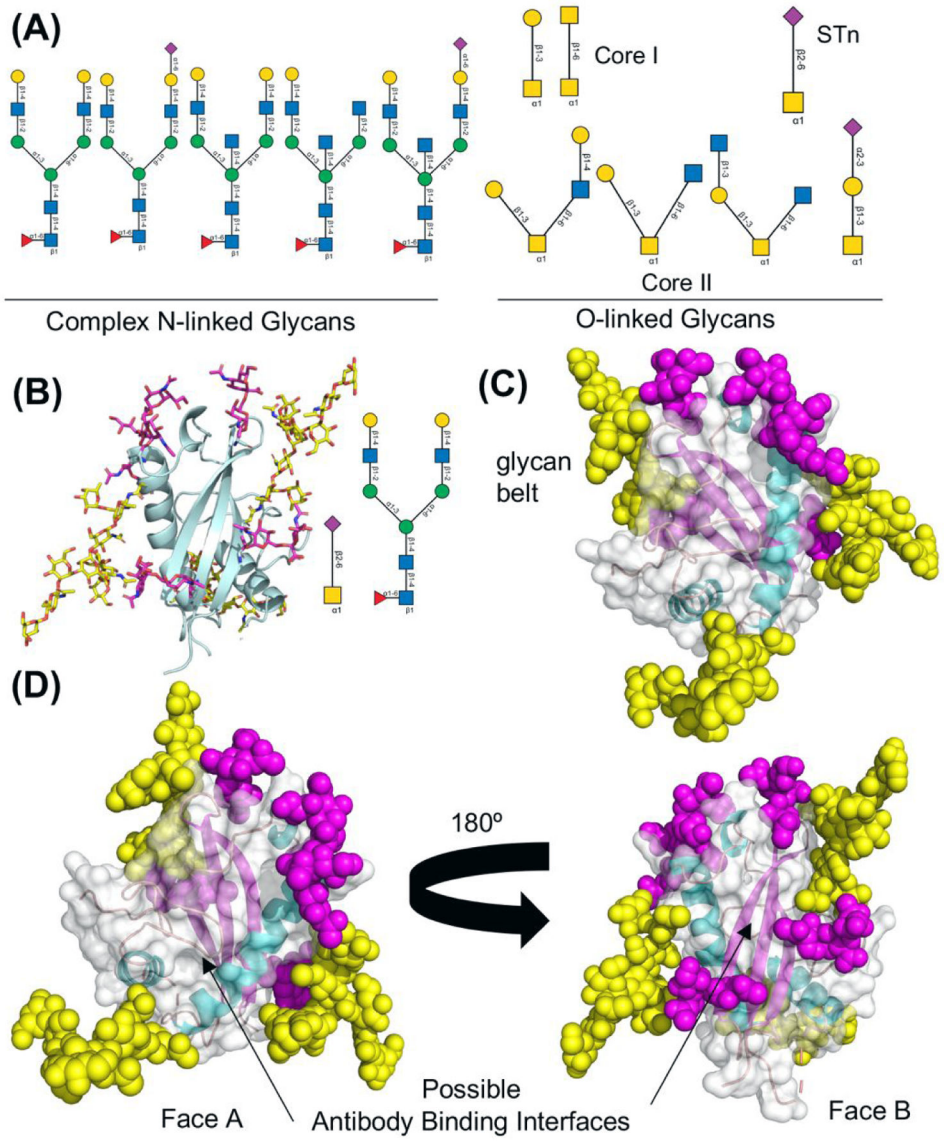

Figure 5 –. Glycosylation of MUC16.

(A) Major N-linked and O-linked glycan structures identified in cancer associated MUC16. Carbohydrates were drawn with GlycoGlyph[41], symbols for carbohydrates follows the standard symbol nomenclature[42]. **(B)** Model of glycosylated MUC16 SEA5 domain containing a complex N-linked glycan (yellow) and the STn O-linked glycan (magenta). (**C**) The glycosylated model contains a “glycan belt” (space filling model) which coats the apex of the protein. (**D**) The glycosylated MUC16 SEA5 domain model contains two major accessible faces which likely represent the sites of CA125 antibody interaction.

Table I

Data Collection and Refinement Statistics

Parameter	MUC16 SEA5 Domain
PDB Code	7SA9
Resolution	58.36 – 1.8 (1.864 – 1.8)
Space group	C 1 21 1
Unit cell	78.54 90.15 42.82 90
	102.88 90
Total reflections	87950 (8643)
Unique reflections	26651 (2632)
Multiplicity	3.3 (3.3)
Completeness	98.88 (97.70)
Mean I/σI	11.42 (2.08)
Wilson B-factor	24.81
R_{merge}	0.055 (0.547)
P_{pim}	0.036 (0.299)
CC1/2	0.996 (0.684)
R_{work} (%)	19.48 (33.50)
R_{free} (%)	21.37 (38.38)
Number of non-hydrogen atoms	2135
Macromolecule	1931
Solvent	204
R.M.S. bonds (Å)	0.004
R.M.S. angles (°)	0.67
Ramachandran favoured (%)	98.29
Ramachandran outliers (%)	0.43
Average B-factor	33.27
macromolecules	32.81
solvent	37.63
Number of TLS groups	15

Table II

Glycosylation Sites in MUC16 SEA5 Domain

Residue number (Uniprot)	Residue number (Structure)	Residue	Glycosylation assignment*	rASA (Chain A)	rASA (Chain B)	Solvent Accessibility
			N-linked			
12704	44	Asn	unambiguous	0.35	0.31	Exposed
12725	65	Asn	unambiguous	0.49	0.55	Exposed
12741	81	Asn	unambiguous	1.05	1.09	Exposed
			O-linked			
12702	42	Thr	unambiguous	0.34	0.38	Exposed
12721	61	Ser	unambiguous	0.20	0.13	Exposed/Buried
12726	66	Thr	ambiguous	0.40	0.32	Exposed
12727	67	Thr	ambiguous	0.05	0.06	Buried
12758	98	Ser	unambiguous	0.48	0.43	Exposed
12765	105	Thr	ambiguous	0	0	Buried
12772	112	Thr	ambiguous	0.27	0.29	Exposed

*4