Characterization of recombinant β subunit of human MUC4 mucin (rMUC4β)

Prakash G. Kshirsagar1, Mansi Gulati1,6, Wade M. Junker1,2,6, Abhijit Aithal1,6, Gaelle Spagnol1, Srustidhar Das1, Kavita Mallya1, Shailendra K. Gautam1, Sushil Kumar1, Paul Sorgen1, Krishan K. Pandey1, Surinder K. Batra1,2,4,5 & Maneesh Jain1,4*

MUC4 is a transmembrane mucin expressed on various epithelial surfaces, including respiratory and gastrointestinal tracts, and helps in their lubrication and protection. MUC4 is also aberrantly overexpressed in various epithelial malignancies and functionally contributes to cancer development and progression. MUC4 is putatively cleaved at the GDPH site into a mucin-like α-subunit and a membrane-tethered growth factor-like β-subunit. Due to the presence of several functional domains, the characterization of MUC4β is critical for understanding MUC4 biology. We developed a method to produce and purify multi-milligram amounts of recombinant MUC4β (rMUC4β). Purified rMUC4β was characterized by Far-UV CD and I-TASSER-based protein structure prediction analyses, and its ability to interact with cellular proteins was determined by the affinity pull-down assay. Two of the three EGF-like domains exhibited typical β-fold, while the third EGF-like domain and vWD domain were predominantly random coils. We observed that rMUC4β physically interacts with Ezrin and EGFR family members. Overall, this study describes an efficient and simple strategy for the purification of biologically-active rMUC4β that can serve as a valuable reagent for a variety of biochemical and functional studies to elucidate MUC4 function and generating domain-specific antibodies and vaccines for cancer immunotherapy.

Abbreviations
Ab  Antibody
EGFR  Epidermal growth factor receptor
FPLC  Fast protein liquid chromatography
MW  Molecular weight
MUC4  Mucin 4
MWCO  Molecular weight cut-off
PCR  Polymerase chain reaction
TB  Terrific broth
LB  Luria broth
BME  β-Mercaptoethanol
vWD  Von Willebrand factor type D domain
PIC  Protease inhibitor cocktails
FPLC  Fast pressure liquid chromatography

Mucins comprise a family of 21 high molecular weight secretory or membrane-bound glycoproteins that provide the structural framework of a ‘mucus-like’ layer covering the epithelial surfaces within our body. Their main roles are to protect epithelial surfaces from various insults and promote their regeneration and repair1–4. One of the mucins, MUC4, is a large, multi-domain, transmembrane glycoprotein involved in multiple physiological and
MUC4 structure

MUC4α-subunit

Histidine-tagged, Thrombin site-encoded Human MUC4α sequence

MUC4β-subunit

Bacterial expression vector pET-28a (+)

Bacterial expression vector pET-28a-MUC4β

rMUC4β expression in C41(DE3)

rMUC4β expression in Rosetta 2(DE3)

Comparative rMUC4β expression

ÄKTA-FPLC purification of rMUC4β

16°C C  25°C C  37°C C

Crude lysate  Inclusion bodies  Supernatant  Pre-ÄKTA pellet  Pre-ÄKTA Sup.  Pure rMUC4β

(kDa)  16  25  37

U  I  U  I  U  I

(kDa)  72  55  72

IPTG [mM]  0  0.25  1.0

IPTG [mM]  0  0.25  1.0

IPTG [mM]  0  0.25  1.0
pathological processes. Full-length MUC4 is synthesized as a single polypeptide that is putatively cleaved in an auto-catalytic manner at the Gly-Asp-Pro-His (GDPH) site located within the vWD domain resulting in two subunits, MUC4α and MUC4β. The large extracellular N-terminal subunit MUC4α contains a characteristic tandem repeat (TR) domain along with NIDO and AMOP domains. The smaller, membrane-tethered, C-terminal subunit MUC4β includes a truncated vWD domain, three (EGF)-like domains, and a cytoplasmic tail. Due to its aberrant overexpression in multiple carcinomas (pancreas, lung, cervix, ovary, colon, skin), MUC4 has emerged as a promising biomarker and therapeutic target. Furthermore, MUC4 is also a potential target for cancer immunotherapy, owing to its aberrant glycosylation and the existence of several splice variants.

We have previously cloned MUC4 "minigene" containing all the MUC4 domains but only 10% of the TR domain and used it to discern the functional role of MUC4 in cancer cell lines using overexpression studies. However, direct structure–function studies of MUC4 have been hampered by its large molecular size, lack of domain-specific reagents, and complexities associated with the existence of diverse glycoforms and splice variants in cancer. Purification of the mega-Dalton-sized, 'full-length' MUC4 by presently available molecular biology techniques is extremely challenging. Due to its potential retention on the cell membrane post-cleavage, and the presence of several functional domains, we directed our efforts to clone, produce, and purify a recombinant MUC4β (rMUC4β) subunit using a bacterial expression system. Herein, we report the optimization of the parameters for the production and purification of rMUC4β and its preliminary biophysical characterization. The purified rMUC4β was recently evaluated for vaccine development and the generation of domain-specific monoclonal antibodies (mAbs). We also tested the utility of purified rMUC4β for identifying MUC4-interacting partners. The optimized production and purification method yielded multi-milligram amounts of pure and soluble rMUC4β. In affinity pull-down assays, rMUC4β recognized known (EGFR, HER2, and HER3) and novel (Ezrin) MUC4-interacting partners. Our studies suggest that rMUC4β produced and purified using the optimized method described herein, retains the characteristics of native MUC4, and can serve as a valuable tool to identify MUC4-interacting partners and undertake structure–function studies.

Results

Cloning and expression of rMUC4β. The schematic structure of MUC4 and the overall scheme of cloning strategy for the construction of the pET28a-MUC4β expression plasmid is described in Fig. 1A. The final construct was sequenced to confirm in-frame insertion of MUC4β with N-terminal hexahistidine (His6) tag (Fig. S1). The overall scheme for the expression and purification of rMUC4β is described in Fig. S2A. Initially, C41(DE3) cells transformed with pET28a-MUC4β were grown until the mid-log phase (Fig. S2B) and induced by the addition of IPTG. The SDS-PAGE and immunoblot analyses of culture supernatants indicated a distinct ~72 kDa rMUC4β band following IPTG induction (Fig. 1B, lane 7 and 8), whereas negative control lanes did not show any expression (Fig. 1B, lanes 1–6). The initial purification yield of rMUC4β from a 4 L culture of C41(DE3) cells was found to be low (~2–3 mg/L). Hence, we optimized various growth and purification conditions (Fig. S2C-D) to improve the yield and purity of rMUC4β protein.

Optimization of culture parameters for rMUC4β expression. To improve the yield of purified rMUC4β, we optimized various parameters and examined their effect on protein expression. Competent cells of five different E. coli (DE3) strains, namely C41, BL21, C41pLysS, C43pLysS, and Rosetta 2, were transformed and tested for rMUC4β expression. Of all the strains tested, Rosetta 2(DE3) exhibited the highest rMUC4β expression, followed by C41(DE3), while the expression was very low in other strains (Fig. S3A-B). Comparative assessment of rMUC4β expression at variable IPTG concentrations showed higher rMUC4β expression in Rosetta 2(DE3) as compared to C41(DE3) (Fig. 1C). We also examined the effect of different culture media (i.e., Terrific Broth (TB) and LB) on the expression of rMUC4β in Rosetta 2(DE3) at different IPTG concentrations. We observed (~1.3-fold) higher expression of rMUC4β in the TB compared to LB (data not shown). However, the cost/benefit provided by the TB media was not substantial enough to justify its use, and thus LB was employed for all further studies. A distinctly higher expression of rMUC4β was observed at 25 °C and 37 °C following IPTG induction, while a relatively low expression was seen at 16 °C (Fig. 1D). To minimize nucleic
differences (Fig. 2A, S3F). Analysis of the secondary structure recorded at 7 °C is presented in Fig. 2B. The CD structure prediction. The Far-UV spectrum collected at 7 °C and higher temperatures did not show significant 

ging, and structural integrity of the renatured rMUC4β protein, we performed CD analysis and i-TASSER-based 

amino acid sequence was submitted for 3-D structure prediction to an online server i-TASSER31–33. The server-

protein from cell lysate includes both soluble and insoluble protein. 

(violet) is depicted as a mostly random coil structure (Fig. 2E, right panel). Meanwhile, the vWD domain (Blue) 

structure was rotated 180 degrees. The model revealed that the EGF-like domain 1 (Cyan) and EGF-like domain 

were arranged according to their decreasing confidence scores (Fig. S4B–E). The tertiary structure of MUC4β 

sen for further analysis as it has the highest confidence score (Fig. 2E). The remaining predicted structures 

of α-helical structures in the 5 predicted models ranged between 1.7 and 6%, while extended β-strands ranged 

generated five different models were analyzed for the composition of the secondary structure. The percentage 

rMUC4β (Table 1). The purified rMUC4β was observed as a single 72 kDa band in SDS-PAGE (Fig. 1E. Lane 

we accomplished a ~ ninefold increase (i.e., ~ 18 mg/L, + 50% recovery) in the total yield of purified-refolded 

detergent mass in the final lyophilized protein. Using the optimized expression and purification conditions, 

significant precipitation when urea was removed following dialysis (Fig. S2D). To enhance the protein solubil-

and subjected to purification as outlined in Fig. S2A. The eluted protein was soluble in the 6 M urea but exhibited 

relatively pure (Fig. 1E, Lane 2). Thus, inclusion bodies were isolated by centrifugation, solubilized in 6 M urea, 

and subjected to purification as outlined in Fig. S2A. The eluted protein was soluble in the 6 M urea but exhibited 

significance precipitation when urea was removed following dialysis (Fig. S2D). To enhance the protein solubility, 

a zwitterionic detergent CHAPS was used instead of Triton X-100 in both wash and elution buffers during 

puriﬁcation. This led to increased solubility of rMUC4β during the refolding steps and minimized the residual 
detergent mass in the final lyophilized protein. Using the optimized expression and purification conditions, 

we accomplished a ~ ninefold increase (i.e., ~ 18 mg/L, + 50% recovery) in the total yield of purified-refolded 

rMUC4β (Table 1). The purified rMUC4β was observed as a single 72 kDa band in SDS-PAGE (Fig. 1E. Lane 

6). Immunoblot analysis of purified protein was also performed for various fractions and conditions to validate 

using anti-His tag antibody (Fig. 1B–D). Interestingly, immunoblot analysis of the purified protein following 

BN-PAGE indicated the tendency of rMUC4β to form discrete covalent dimers (~ 150 kDa under non-reducing 

conditions, which was resolved into a monomeric form in the presence of BME (Fig. S3E).

Table 1. Purification of rMUC4β from E. coli Rosetta 2(DE3). aFractions collected during ÄKTA-FPLC. bTotal 

protein quantified using NanoDrop One-C (Thermo Scientific). c rMUC4β concentration (mg) determined 

by the relative quantification of Coomassie Blue stained SDS-PAGE gels compared to a BSA standard curve 

using Biorad Gel Doc XR + Imaging System (BioRad). dPurity of refolded rMUC4β in each fraction obtained 

by dividing the rMUC4β band area (intensity) by total area (intensity) of all stained bands in the same lane. 

ePercent recovery or yield of rMUC4β was calculated on a “per fraction” basis by dividing the amount of 

MUC4β (mg) in that fraction by the amount present in the previous fraction and multiplying by 100. fTotal 

protein from cell lysate includes both soluble and insoluble protein.

Prediction of rMUC4β secondary structure by CD spectroscopy. To evaluate the stability, folding, 

and structural integrity of the renatured rMUC4β protein, we performed CD analysis and i-TASSER-based 

structure prediction. The Far-UV spectrum collected at 7 °C and higher temperatures did not show significant 

differences (Fig. 2A, S3F). Analysis of the secondary structure recorded at 7 °C is presented in Fig. 2B. The CD spectrum (λ = 190–260 nm) of the refolded rMUC4β (Fig. 2A) is characterized by one broad minima around 

215 nm typical of β-strand structure and a lower than expected maxima at about 195 nm that is most likely 

caused by a significant amount of random coil content (typical shows as a pronounced minima at 197 nm)28. 

Further, the assessment of the CD spectrum (Fig. 2B–D) was performed separately using the CDSSTR and 

Contin/LL (Provencher & Glockner Method) programs provided by the online analysis tool DichroWeb29,30. 

Both analyses revealed an average secondary structure composition of ~10% for α-helical structure, ~ 48% for 

β-strand, and ~ 41% for random coil (Fig. 2C,D). Comparison of the secondary structure prediction data from 

the amino acid sequence (GOR IV), I-TASSER, and CD (average composition from the CDSSTR or CONTIN/ 

LL) analyses are shown in Fig. 2B.

3-D structure prediction of MUC4β by I-TASSER. To predict the tertiary structure of MUC4β, the 

amino acid sequence was submitted for 3-D structure prediction to an online server i-TASSER31–33. The server-
gen generated five different models were analyzed for the composition of the secondary structure. The percentage 
of α-helical structures in the 5 predicted models ranged between 1.7 and 6%, while extended β-strands ranged 

from 10.1 to 29.4%, and random coil structure ranged from 68 to 84% (PDB code: 5G56). Model 1 was chosen 

for further analysis as it has the highest confidence score (Fig. 2E). The remaining predicted structures 

were arranged according to their decreasing confidence scores (Fig. 2B–E). The tertiary structure of MUC4β 

(Model 1) is displayed in a rainbow-colored spectrum (Fig. 2E, left panel), with the N-terminus and C-terminus 
denoted in blue and red color, respectively. To better visualize the EGf-like domains contained in MUC4β, the 
structure was rotated 180 degrees. The model revealed that the EGf-like domain 1 (Cyan) and EGf-like domain 
2 (Magenta) are composed of the expected typical β folds (anti-parallel β-sheet), while the EGf-like domain 3 
(red) is depicted as a mostly random coil structure (Fig. 2E, right panel). Meanwhile, the vWD domain (Blue)
appears as a highly random coiled structure; three small α-helix and three β-strands forming a parallel β-sheet) are observed.

**Affinity pull-down and mass spectrometry to detect MUC4 interacting partners.** To determine if the rMUC4β exhibits a similar conformation to native MUC4, we tested its ability to interact with known and potentially novel MUC4-interacting partners using a His-affinity pull-down assay as outlined in Fig. 3A. We
Figure 3. Affinity pull-down and tandem mass spectrometry to determine the interacting partners of MUC4. (A) Schematic outline of affinity pull-down assays (I) Pre-clearing of total lysate [input] of MUC4 expressing (CD18/HPAF) and non-expressing (PANC-1) pancreatic cancer cell lines. (II) Affinity pull-down reactions were performed in the presence [rMUC4β (+)] or absence [rMUC4β (−), i.e., RIPA buffer] of rMUC4β (bait) added with washed Ni–NTA beads. Respective supernatants [i.e., Flow-through (FT) and last wash (W5)] were collected before the final elution. (III) An aliquot of rMUC4β-bound Ni-resin was eluted separately to serve as a positive control. All saved fractions (as shown in pink color, italic font) from CD18/HPAF (B) and PANC1 (C) were resolved on SDS-PAGE gel and stained with Coomassie Blue. The ~ 80 kDa band (arrowed) was identified, excised, and assessed by the tandem MS/MS and proteomic analyses.
observed that rMUC4β pulled down a ~ 80 kDa band from the lysates of both MUC4-expressing (CD18/HPAF) and non-expressing (PANC-1) cell lines (Fig. 3B,C, lane 13), which was not detected in the pull-down eluates when beads alone were used (Fig. 3B,C, lane 6). This band was excised, in-gel trypsin digested, and the resulting peptides were analyzed by tandem MS/MS. The most significant top hits were matched to Ezrin in CD18/HPAF and PANC-1 cells (Fig. S5A-B). Analysis of the pulled-down complexes by immunoblotting with the antibodies of known MUC4-interacting proteins (EGFR family receptors) indicated that rMUC4β interacts with EGFR, HER2, and HER3 in both MUC4-expressing and non-expressing cells (Fig. 4A,B). Interaction with Ezrin was confirmed by a pull-down assay using an anti-His tag Ab-coupled to magnetic beads to immobilized rMUC4β (Fig. 4C). Finally, the ability of endogenous MUC4β to interact with Ezrin was further confirmed by CO-IP (i.e., immunoprecipitation with anti-MUC4β mAb 6E8 (Fig. 4D). Overall, these results suggest that the rMUC4β recapitulates the functional conformation of native MUC4β and can recognize its natural interacting partners.

Discussion

Due to their differential expression in various malignancies, mucins have been extensively investigated as determinants of disease initiation and progression and as diagnostic and therapeutic targets. MUC4 is a high molecular weight, multi-domain, and heavily glycosylated protein that is differentially overexpressed in multiple cancers, including pancreatic, breast, lung, and head and neck cancers, where it functionally contributes to disease initiation, progression, metastasis, and chemoresistance. It is also a potential biomarker for the diagnosis and prognosis of various cancers and a novel target for cancer therapeutics.

Most information regarding MUC4 function has been discerned using cell lines engineered for MUC4 overexpression or knockdown, while its pathobiological significance has been determined from correlative studies involving the analysis of biological samples for MUC4 transcripts and protein. However, direct structure-function studies for human MUC4 have not been extensively performed due to its large genomic and molecular size, and extensive glycosylation. Due to the large size of the MUC4 gene and transcript, cloning and expressing a full-length MUC4 protein is practically impossible. We have previously cloned the MUC4 “minigene” in a eukaryotic expression system, containing only 10% of the tandem repeat sequence but with all other domains intact. While this construct has been useful for studying the impact of MUC4 overexpression in non-expressing cells to
discern MUC4 function, it is unsuitable for large-scale purification. An alternative approach is to clone, express, and purify smaller fragments containing functional domains. We have previously cloned and expressed various domains of the MUC4α subunit and used them developing domain-specific anti-MUC4 monoclonal antibodies. Due to the presence of three EGF-like domains and a vWD domain, and its proximity to the cell membrane, several functions of MUC4, particularly the ones involving interactions with the EGF receptor family, have been attributed to MUC4β subunit.

Further, our recent analysis suggests that the immunogenic peptides in MUC4 are enriched in the β-subunit. Due to these attributes, the MUC4β subunit is a potential target for immunotherapy and targeted therapies. Thus, we directed our efforts to clone, express, and purify the MUC4β subunit with a goal to utilize the purified recombinant protein to develop domain-specific antibodies and pancreatic cancer immunotherapy. Recently, Liberele et al. generated fusion constructs containing GFP fused to either intact human MUC4β or truncated mutants containing only EGF-like domains. These constructs were expressed in CHO cells and used to characterize MUC4-ErbB2 interactions. However, the studies were performed using CHO cell lysates and recombinant ErbB2, and no efforts were made to purify MUC4β-GFP.

Initially, the MUC4β sequence was cloned into the pET-28a (+) expression vector. Initial scale-up attempts using 1L culture of pET28a-MUC4β plasmid transformed in C41(DE3) cells resulted in a low yield of purified MUC4β (i.e., ~ 1.8 mg/L of LB media). Analysis of MUC4β sequence, with a rare codon calculator (RaCC), indicated the presence of 32 rare or minor codons including AGA, AGG, CGA, CGG, CUA, CCC, and AUA. Rare or minor codons are underutilized codons by the bacterial translational machinery, resulting in low protein yield. We screened five different genetically engineered E. coli strains for MUC4β expression, of which Rosetta 2(DE3) provided the maximum improvement (> tenfold than the C41 strain) in protein yield (Table 1). This specific strain expresses the less utilized tRNAs to compensate for rare codons that are generally not used in E. coli. Prior to scaling up the production, we optimized various parameters, including post-induction incubation temperature, the concentration of IPTG, and culture media composition, to improve the rMUC4β yield. Under conditions optimized for high expression (IPTG: 0.5 mM; growth medium-LB broth), rMUC4β accumulated in inclusion bodies. We developed a comprehensive purification strategy that included cell lysis, removal of nucleic acids, isolation and solubilization of inclusion bodies, immobilized metal affinity chromatography, and dialysis to facilitate urea removal and protein refolding. Due to the high amount of nucleic acids, the high viscosity of the bacterial lysates can potentially interfere with downstream processing and impact protein yield. Reduction in cell lystate viscosity can be achieved by supplementing bacterial lysate with Benzonase nuclease. The use of CHAPS instead of non-ionic Triton X-100 in the solubilization and elution steps prevented protein aggregation during urea removal and contributed to the high solubility of rMUC4β in water that can be attributed to the high CMC value and zwitterionic nature of CHAPS. Immuno blotting analysis following BN-PAGE under reducing and non-reducing conditions suggested the tendency of purified rMUC4β to form covalent dimers via disulfide bonds. The possibility of higher-order soluble aggregates can be ruled out as we did not observe higher molecular weight bands even by immuno blotting.

To gain insight into protein folding, we employed a GOR IV server to predict the secondary structure by amino acid sequences. Further, we performed structural characterization of the purified rMUC4β protein by conducting CD spectroscopy. The CDSSTR and Contin/LL methods were employed to analyze the Far-UV CD spectrum. CD data demonstrated the presence of α-helix, β-strand, and random coils. While understanding the secondary structure, we observed structural disparity in the β-strands derived from CD and other secondary structure prediction methods. Supportively, some reports have shown some discrepancies in CD analysis and sequence-based structure predictions regarding the quantification for β-structure because of their morphological and structural diversity. To gain further insights into the tertiary structure and understand the folding of the rMUC4β, we performed structure prediction analysis using an online-based I-TASSER server. The server-generated five models. The model with the highest confidence score was selected to understand the folding pattern and its interacting partners. The predicted tertiary structure showed typical β-folds for two EGF-like domains, while the third EGF-like domain showed a random coiled structure. The vWD domain is predominantly disordered, but we did observe a parallel β-sheet comprising of three β-strands (3–4 residues each). Three small α-helical structures (4 residues each) were also observed in the vWD domain. The presence of a high percentage of random coils (disordered structures) in MUC4 can be attributed to the presence of a high proportion of charged or polar residues. The decreased hydrophobic residues make the protein less susceptible to present a hydrophobic ordered core. In a recent study analyzing the intrinsically disordered structures of mucins, we reported that MUC4 is 77% disordered; the study also highlighted the vWD domain as a highly disordered domain. However, these prediction studies are based on amonucin backbones, and the impact of glycosylation on the intrinsically disordered content of mucins (including MUC4) remains unexplored and remains to be discerned. While the recombinant MUC4β protein is non-glycosylated, our affinity pull-down studies demonstrated the ability of MUC4β to interact with EGFR and HER2 in a manner similar to that reported previously for native protein. Thus, some of the interactions of MUC4 with its binding partners can be glycosylation independent. We observed some differences in the predicted structures by GOR IV and i-TASSER; these differences can be attributed to the different algorithms employed by these two servers. GOR IV is a second-
MUC4 plays multifaceted roles in cell adhesion, migration, proliferation to promote tumorigenesis and metastasis in pancreatic cancer and other epithelial malignancies\textsuperscript{2,6,37,38}. One of the ways by which it mediates these roles is by interacting with cell surface molecules such as HER2, HER3, and integrins and soluble mediators in the extracellular microenvironment such as galectins\textsuperscript{16,18,20,55}. Interestingly, we could also detect EGFR in the eluted fraction. This can either be due to direct interaction with the protein or indirectly because of its association with HER2. Given the folding pattern observed in our structure prediction analysis, it is possible that the three EGF-like domains of MUC4 bind to the EGF binding pocket of EGFR. Further studies are needed to confirm or rule out these observations. It will be interesting to determine if exogenously added rMUC4β can potentially compete with the endogenous MUC4 for binding partners and serve as a potential therapeutic agent.

In summary, we successfully developed an expression and purification system to produce multi-milligram amounts of rMUC4β. The recombinant protein fragment retained the binding characteristics of the native MUC4β and can be a valuable tool to undertake MUC4 structure–function analysis. Further, purified rMUC4β has been successfully formulated and evaluated as a nanovaccine\textsuperscript{22,23} and used as an immunogen for the development of domain-specific anti-MUC4 antibodies\textsuperscript{24,25}.

Materials and methods

Materials. C41(DE3) and BL21(DE3) E. coli strains were purchased from MilliporeSigma. C41pLysS, C43pLysS, and Rosetta 2(DE3) E. coli strains were procured from Thermo scientific. Bacto tryptone, Bacto yeast, and Bacto Agar were purchased from BD Biosciences. Anti-His tag Ab (Clone 27E9, Cat. No. 2366) and anti-EGFR mAb (Clone D38B1, Cat. No. 4267) were purchased from Cell Signaling Technology, while anti-Ezrin Ab (Clone H-276, Cat No. SC-20773) was bought from Santa Cruz Biotechnology. HisTrap HP affinity Ni-NTA column was purchased from GE healthcare. Anti-His tag Ab Magnetic bead conjugate (Cat. No. 8811) and Protein A/G agarose beads (Cat. No. 9863) were purchased from Cell Signaling Technology. Anti-MUC4β mAb (Clone 6E8) was generated using recombinant pET-28a-MUC4β as an immunogen. More details on the construction of recombinant pET-28a-MUC4β are mentioned in supporting information.

Optimization of culture parameters to enhance rMUC4β yield. Selection of E. coli host for optimal production. Five different chemically competent E. coli (DE3) strains (C41, BL21, C41pLysS, C43pLysS, and Rosetta 2) were tested for optimal expression of rMUC4β protein. Briefly, selected strains transformed with pET28a-MUC4β plasmid were grown in LB/antibiotics media overnight at 37 °C. A freshly-made pre-inoculum culture (2%) from each cell type was inoculated into 20 mL LB media and incubated at 37 °C until the mid-exponential growth phase. When OD\textsubscript{600} nm reached ~0.6, cultures were induced with 0.5 mM IPTG for 4 h at 37 °C (a set of negative control was incubated under similar conditions without IPTG). Next, both induced and uninduced cultures were harvested and disrupted in an appropriate volume of lysis buffer (500 mM Tris HCl, 2 mM MgCl\textsubscript{2}, 10 mM NaCl, pH 8.0) to adjust the OD\textsubscript{600} nm to ~1.2/mL. Equivalent amounts of crude lysate were loaded, ran on SDS-PAGE, and stained with Coomassie brilliant blue.

Optimization of rMUC4β expression. C41(DE3) and Rosetta 2(DE3) competent cells transformed with pET28a-MUC4β plasmid were grown in LB culture at 37 °C. After reaching the mid-exponential phase, all bacterial cultures were separately induced at variable IPTG concentrations (0, 0.25, 0.5, and 1.0 mM) for 4 h at 37 °C. Uninduced controls were prepared without adding IPTG. The effect of three different post-induction temperatures on the enrichment of rMUC4β expression was evaluated in pET-28a-MUC4β transformed Rosetta 2(DE3) cells\textsuperscript{26}. Bacterial culture was induced at ~0.6 OD by adding 0.5 mM of IPTG. After induction, each tube was incubated separately at 16 °C, 25 °C, and 37 °C, respectively, for 4 h. To investigate the effects of different culture media on the expression efficiency of rMUC4β in E. coli, freshly-made pre-inoculum culture (2%) of Rosetta 2(DE3) cells transformed with pET-28a-MUC4β plasmid was added to 20 mL of culture media (LB and TB) and grown at 37 °C until medium exponential phase. One set of cultures was induced with 0.5 mM IPTG and incubated for 4 h at 37 °C. The second control set was grown without IPTG addition. A sample from each culture condition was pelletted and lysed in an appropriate volume of lysis buffer to maintain equal ODs (~1.2/
Scale-up production of rMUC4β protein in Rosetta 2(DE3). Large-scale production of rMUC4β protein was undertaken using optimized culture conditions. The transformed cells were grown on LB agar plates containing kanamycin (50 µg/mL) and chloramphenicol (40 µg/mL). After overnight incubation at 37 °C, a single colony was picked and grown overnight into 10 mL of LB/antibiotics culture in a shaking incubator at 37 °C. Following day, 50 mL of LB/antibiotics culture was inoculated by 500 µL of overnight-grown culture (1%) and incubated overnight at 37 °C, 250 rpm. The next day, overnight grown pre-inoculum (15.0 mL) was inoculated into 1 L of LB/antibiotics media. Subsequently, the culture was incubated under continuous agitation at 37 °C until OD600 nm reached ~0.5–0.6. MUC4β expression was induced by adding IPTG [0.5 mM] and growing for another 4 h at 37 °C. Cells were harvested by centrifugation (4300 × g for 30 min, at 4 °C), washed once with ice-cold 1 × PBS, and stored overnight at -80 °C. The bacterial pellet was thawed and resuspended in ice-cold lysis buffer [50 mM Tris–HCl (pH 8.0), 2 mM MgCl₂, and 150 mM NaCl, EDTA-free protease inhibitor cocktail (250 µL/10 g of the bacterial pellet) and benzonase nuclease (~10 U/mL pre-lysis culture)]. Following 15 min incubation on ice, cells were lysed with three passages through an EmulsiFlex-C3 (Avestin) at ~15,000 psi. Cell debris was removed by centrifugation (4300 × g for 30 min at 4 °C). The supernatant was treated again with benzonase nuclease (10 U/mL cell lysate) and incubated for 2 h at 4°C. The rMUC4β protein present in the insoluble inclusion bodies was collected by high-speed centrifugation (39,000 × g for 60 min at 4 °C). The recovered pellet was resuspended in 40 mL of buffer A (1 × PBS, 6 M urea, 20 mM imidazole, 300 mM NaCl, 0.5% CHAPS, and 2 mM β-mercaptoethanol (BME), pH 8.0), incubated overnight at 4 °C and again treated with Benzonase nuclease. Insoluble debris was removed by centrifugation (39,000 × g, 60 min, at 4 °C), and the pre-AKTA supernatant was passed through a 0.22 µm Sterilip vacuum filtration unit (EMD Millipore). More details on the purification and refolding of rMUC4β protein can be found in supporting information.

Purification and protein refolding. The solubilized inclusion bodies (pre-AKTA supernatant) recovered after multi-step centrifugation was filtered and applied to a HisTrap Ni-NTA column (GE Healthcare) connected to an ÄKTA-FPLC (GE Healthcare). The column was washed with buffer A. Protein elution was performed using a step gradient consisting of 20%, 30%, 50%, 60%, 75% and 100% Buffer B (1 × PBS, 6 M urea, 500 mM imidazole, 300 mM NaCl, 0.5% CHAPS, and 2 mM BME, pH 8.0). Eluted fractions were run on SDS-PAGE gels and were visualized by Coomassie staining. Fractions containing His₆₆-tagged MUC4β were pooled and concentrated using Amicon Ultra-4 centrifugal filters (50 kDa MWCO) to exclude lower molecular weight bacterial protein impurities. For protein refolding, a Float-A-lizer G2 dialysis tube (50 kDa MWCO, Spectrum Labs) was used, and step-wise dialysis was performed against 1 × PBS (pH 8.0, volume 2 L) containing decreasing concentrations of urea (6 M, 5 M, 4 M, 3 M, 2 M, 1 M, and 1 × PBS) to gradually minimize the denaturant concentration and allow for a smooth refolding of rMUC4β. Each dialysis step was performed at 4 °C for 3 h against 2 L volumes of dialysis buffer. The final dialysis was done against Hyclone (GE) endotoxin-free water (for 15 h with three changes every 5 h). Purified rMUC4β was also resolved using Blue Native PAGE (BN-PAGE), following thermal denaturation under both reducing and non-reducing conditions, subjected to western blotting and probed with anti-His tag mAb to determine the presence of covalent and/or non-covalent dimers or aggregates.

Circular dichroism (CD) spectroscopy. Circular dichroism (CD) analysis of rMUC4β was performed on a Jasco J-815 spectrometer in the Far-UV spectral region (λ = 190 to 260 nm) using a 0.1 mm path length quartz cuvette. CD spectra of rMUC4β at a concentration of 39 µmol/L in 1 × PBS were recorded at 7 °C, 25 °C and 37 °C using a scan rate of 50 nm/min, an integration time of 1 s, and a bandwidth of 1.0 nm. The final spectrum was obtained by averaging five scans and was corrected by subtracting the solvent spectrum acquired under identical conditions. CD data were processed using the spectra analysis function of Jasco Spectra Manager II (http://www.jasco.com/). Analysis of the CD spectrum was performed separately using two methods, CDSSTR and Contin/LL.

Iterative threading ASSEMBly refinement (I-TASSER). I-TASSER is a bioinformatics program that predicts the three-dimensional (3-D) model structure of the desired protein using amino acid (AA) sequences. We used this intelligent program to model the 3-D structure of MUC4β protein by submitting the corresponding AA sequences to the I-TASSER online server. After obtaining five different predicted models of MUC4β, the structure with the highest confidence score was selected for comparative analysis with CD data.

Affinity pull-down assays. High-affinity Ni-NTA resin (GenScript, Cat. No. L00250) was washed and equilibrated with ice-cold wash buffer (1 × PBS, containing 1% CHAPS, 20 mM imidazole, and 2 mM BME, 300 mM NaCl, pH 8.0). The equilibrated Ni-NTA slurry was incubated with either CD18/HPAF or PANC-1 cell lysates for 4 h at 4 °C to pre-clear the cell lysates. Ni-NTA resin, after pre-clearing for each sample, was used as a "non-specific binding control." A portion of each pre-cleared lysate supernatant was saved and used as 'Input' control. The lysates were incubated with 60 µg of rMUC4β protein [rMUC4β (+)] or without rMUC4β protein [rMUC4β (−)], in total 500 µL reaction volume overnight at 4 °C with end-over-end mixing. The samples were centrifuged (2000 rpm, for 2 min at 4 °C), and the supernatant containing unbound proteins was designated as "flow-through" (FT) and saved for further analysis. The Ni-NTA resin was washed five times (2000 rpm, for 2 min at 4 °C) with ice-cold wash buffer (1 × PBS, containing 20 mM imidazole, 1% CHAPS, 2 mM BME, 300 mM NaCl, pH 8.0) and the last wash (Wash 5) was saved for analysis. Finally, the 'bait-prey' complexes
were eluted with elution buffer (1 × PBS supplemented with 500 mM imidazole, 1% CHAPS, and 2 mM BME, 300 mM NaCl, pH 8.0) ice for 15 min. The eluted proteins (Eluate) were recovered in the supernatant following centrifugation. An aliquot of rMUC4β-bound Ni-resin was also eluted with elution buffer to serve as a positive control (i.e., Eluted Recombinant Protein. "ERP (s)" control). All collected samples were solubilized in 6 × Laemmli sample buffer, heated at 95 °C for 5 min, and subjected to SDS-PAGE and immunoblot analyses briefly described in supporting information.

**Mass spectrometry (MS/MS) characterization of excised protein band.** The protein profiles of pull-down eluted complexes in two different cell lysates (CD18/HPAF and PANC-1) were visualized by Coomassie staining and silver staining. The distinct protein band present in the eluted fraction of the rMUC4β (+) panel in both CD18/HPAF and PANC-1 lysates was destained and stored in 50% methanol overnight at 4 °C. The selected band (MW ~ 80 kDa) was excised and digested with trypsin by the UNMC Mass Spectrometry and Proteomics Core. Following digestion, samples were dried by SpeedVac, resuspended in 20 µl of 0.5% trifluoroacetic acid, and were cleaned up using ZipTip µC18 columns (EMD Millipore). Further, the samples were resuspended in 0.1% formic acid and injected onto an Eksigent SpeedVac, resuspended in 20 µl of 0.5% trifluoroacetic acid, and were cleaned up using ZipTip µC18 columns (EMD Millipore). The selected band (MW ~ 80 kDa) was excised and digested with trypsin by the UNMC Mass Spectrometry and Proteomics Core. Following digestion, samples were dried by SpeedVac, resuspended in 20 µl of 0.5% trifluoroacetic acid, and were cleaned up using ZipTip µC18 columns (EMD Millipore). Further, the samples were resuspended in 0.1% formic acid and injected onto an Eksigent chipLC column (75 µm × 15 cm ChromXP C18-CL 3 µm 120 Å), and 6600 TripleTOF instruments (gradient (EMD Millipore). Further, the samples were resuspended in 0.1% formic acid and injected onto an Eksigent chipLC column (75 µm × 15 cm ChromXP C18-CL 3 µm 120 Å), and 6600 TripleTOF instruments (gradient 2−60% ACN in 60 min).

**References**

1. Kaur, S., Kumar, S., Momi, N., Sasson, A. R. & Batra, S. K. Mucins in pancreatic cancer and its microenvironment. *Nat. Rev. Gastroenterol. Hepatol.*, 10, 837–848 (2013).
2. Kufe, D. W. Mucins in cancer: Function, prognosis and therapy. *Nat. Rev. Cancer*, 4, 65–60 (2004).
3. Aithal, A. et al. MUC16 as a novel target for cancer therapy. *Expert Opin. Ther. Targets*, 22, 675–686 (2018).
4. Chaturvedi, P. & Batra, S. K. MUC4 mucin: Protection and control of the cell surface. *Nat. Rev. Cancer*, 4, 65–60 (2004).
5. Moniaux, N. et al. Alternative splicing generates a family of putative secreted and membrane-associated MUC4 mucins. *Eur. J. Biochem.*, 267, 4536–4544 (2000).
6. Moniaux, N. et al. Complete sequence of the human mucin MUC4: A putative cell membrane-associated mucin. *J. Biochem.*, 338, 325–333 (1999).
7. Batina, S. et al. MUC4, a multifunctional transmembrane glycoprotein, induces oncogenic transformation of NIH3T3 mouse fibroblast cells. *Can. Res.*, 68, 9231–9238 (2008).
8. Carraway, K. L., Theodoropoulos, G., Kooloski, G. A. & Carothers Carraway, C. A. Muc4/MUC4 functions and regulation in cancer. *Carcinogenesis*, 29, 809–910 (2008).
9. Gautam, S. K. et al. MUC4 mucin—a therapeutic target for pancreatic ductal adenocarcinoma. *Expert Opin. Ther. Targets*, 21, 657–669 (2017).
10. Gautam, S. K. et al. MUCIN-4 (MUC4) is a novel tumor antigen for pancreatic cancer immunotherapy. *Semin. Immunol.*, 47, 101391; https://doi.org/10.1016/j.smim.2020.101391 (2020).
11. Singh, A. P. & Batra, S. K. Emerging roles of MUC4 in cancer: A novel target for diagnosis and therapy. *Can. Res.*, 67, 433–436 (2007).
12. Kumar, S. et al. Genetic variants of mucins: Unexplored conundrum. *Carcinogenesis*, 38, 671–679 (2017).
13. Albrecht, H. & Carraway, K. L. MUC1 and MUC4: Switching the emphasis from large to small. *Cancer Biol. Ther.*, 26, 261–271 (2011).
14. Chaturvedi, P. et al. MUC4 mucin interacts with and stabilizes the HER2 oncoprotein in human pancreatic cancer cells. *Can. Res.*, 68, 2065–2070 (2008).
15. Jonckheere, N. et al. The mucin MUC4 and its membrane partner ErbB2 regulate biological properties of human CAPAN-2 pancreatic cancer cells via different signalling pathways. *PLoS ONE*, 7, e32232 (2012).
16. Lakshmanan, I. et al. Novel HER3/MUC4 oncogenic signaling aggravates the tumorigenic phenotypes of pancreatic cancer cells. *Oncotarget*, 6, 21085–21099 (2015).
17. Liberele, M. et al. MUC4-ErbB2 oncogenic complex: Binding studies using microscale thermophoresis. *Sci. Rep.*, 9, 16678 (2019).
18. Ponnusamy, M. P. et al. MUC4 activates HER2 signalling and enhances the motility of human ovarian cancer cells. *Br. J. Cancer*, 99, 520–526 (2008).
19. Moniaux, N. et al. Human MUC4 mucin induces ultra-structural changes and tumorigenicity in pancreatic cancer cells. *Br. J. Cancer*, 97, 345–357 (2007).
20. Banerjee, K. et al. Amphilphilic polyanhydride-based recombinant MUC4β-nanovaccine activates dendritic cells. *Genes Cancer*, 10, 52–62 (2019).
21. Liu, L. et al. Polyanhydride nanoparticles stabilize pancreatic cancer antigen MUC4β. *J. Biomed. Mater. Res. Part A*, 109(6), 893–902 (2021).
22. Aithal, A. et al. Targeting MUC4 in pancreatic cancer using non-shed cell surface bound antigenic epitopes. *Pancreas*, 48(10), 1401–1402 (2019).
23. Orzechowski, C. et al. Generation and characterization of novel antibodies against the β subunit of MUC4A. *Cancer Res.*, 79(13), 3829–3839 (2018).
24. Boynton, Z. L. et al. Reduction of cell lysate viscosity during processing of poly(3-hydroxyalkanoates) by chromosomal integration of the staphylococcal nuclease gene in pseudomonas putida. *Appl. Environ. Microbiol.*, 65, 1524–1529 (1999).
25. Rodríguez Gamero, J. E. et al. Nuclease expression in efficient polyhydroxyalkanoates-producing bacteria could yield cost reduction during downstream processing. *Biore. Technol.*, 261, 176–181 (2018).
26. Manavalan, P. & Johnson, W. C. Sensitivity of circular dichroism to protein tertiary structure class. *Bioinformatics*, 27, 1630–1636 (2011).
Acknowledgements
This work and the authors of this manuscript are supported, in parts, by grants from the National Institutes of Health (U01 CA213862, R01 CA195586, R01 CA247471, P01CA217798 and R44 DK117472). The authors acknowledge the invaluable technical support from Nicole Japp.

Author contributions
P.G.K.: Conceptualization, Methodology, Validation, Formal analysis, Prepared figures, Writing-Original draft, Writing- Review & Editing. M.G.: Methodology, Validation, Formal analysis, Prepared figures, Writing- Review & Editing. W.M.J.: Methodology, Validation, Formal analysis, Writing- Review & Editing. A.A.: Methodology, Validation, Formal analysis, Writing- Review & Editing. G.S.: Formal analysis, Writing- Review & Editing. S.K.G.: Formal analysis, Writing- Review & Editing. K.M.: Formal analysis, Writing- Review & Editing. S.K.: Formal analysis, Writing- Review & Editing. P.S.: Formal analysis, Resources, Writing- Review & Editing. K.K.P.: Formal analysis, Writing- Review & Editing. S.K.B.: Conceptualization, Resources, Supervision, Writing- Review & Editing. M.J.: Conceptualization, Resources, Supervision, Formal analysis, Writing- Review & Editing.
Competing interests
SKB is one of the co-founders of Sanguine Diagnostics and Therapeutics, Inc. The other authors disclose no potential conflicts of interest.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-02860-5.

Correspondence and requests for materials should be addressed to S.K.B. or M.J.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021