

Individual recombinant repeats of MUC16 display variable binding to CA125 antibodies

Chien-Wei Wang^{1,2}, Eliza K. Hanson^{1,2}, Lisa Minkoff² and Rebecca J. Whelan^{1,2*}

Affiliations:

1 Department of Chemistry, University of Kansas, Lawrence, KS, United States of America

2 Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN, United States of America

*Corresponding author: Rebecca J. Whelan, University of Kansas, Multidisciplinary Research Building 220E, University of Kansas, Lawrence, KS, United States of America. Tel.: + 1-785-864-4670; E-mail: rwhelan1@ku.edu

Abstract.

BACKGROUND: Despite its importance in the clinical management of ovarian cancer, the CA125 biomarker—located on the mucin protein MUC16—is still not completely understood. Questions remain about MUC16’s function and structure, specifically the identity and location of the CA125 epitopes.

OBJECTIVE: The goal of this study was to characterize the interaction of individual recombinant repeats from the tandem repeat domain of MUC16 with antibodies used in the clinical CA125 II test.

METHODS: Using *E. coli* expression, we isolated nine repeats from the putative antigenic domain of CA125. Amino acid composition of recombinant repeats was confirmed by high-resolution mass spectrometry. We characterized the binding of four antibodies—OC125, M11, “OC125-like,” and “M11-like”—to nine recombinant repeats using Western blotting, indirect enzyme-linked immunosorbent assay (ELISA), and localized surface plasmon resonance (SPR)

spectroscopy.

RESULTS: Each recombinant repeat was recognized by a different combination of CA125 antibodies. OC125 and “OC125-like” antibodies did not bind the same set of recombinant repeats, nor did M11 and “M11-like” antibodies.

CONCLUSIONS: Characterization of the interactions between MUC16 recombinant repeats and CA125 antibodies will contribute to ongoing efforts to identify the CA125 epitopes and improve our understanding of this important biomarker.

Keywords: CA125; MUC16; ovarian cancer; tandem repeats; antibodies; affinity

1 Introduction

Considerable effort has been directed toward expanding the suite of biomarkers available for diagnosing and monitoring high-grade serous ovarian cancer (HGSOC) [1 – 8]. Although new biomarkers—most significantly human epididymis protein 4 (HE4) [9, 10]—have been identified and are proving to be transformative, enabling new assays and algorithms [11 – 16], no biomarker has supplanted cancer antigen 125 (CA125), which remains the clinical gold standard for monitoring response to treatment and detecting cancer recurrence [17 – 20]. The FDA-approved assay for CA125 is widely used, despite the fact that the CA125 epitopes have not been defined and controversy persists regarding the minimal functional unit necessary for antibody detection [21 – 23]. In other words, the test that underlies vital decisions in ovarian cancer care employs a mechanism that is not understood.

The CA125 epitopes are carried on MUC16, a large mucin [24, 25]. Some structural features of MUC16 have been determined, including a highly glycosylated N-terminal domain, an immunologically active tandem repeat region containing many similar but non-identical subdomains, and a C-terminal domain including a transmembrane region and cytoplasmic tail [26, 27]. The circulating form of MUC16 detected with the CA125 II ELISA does not contain the transmembrane or cytoplasmic regions. Based on competition studies, CA125 antibodies have been sorted into three groups: OC125-like (group A), M11-like (group B), and OV197-like (group C) [28, 29]. The CA125 II ELISA test is a double determinant immunoassay using two different antibodies (M11 and OC125) as capture and tracer [30]. Earlier iterations of the assay used OC125 as both capture and tracer [31]. The location and identity of the CA125 epitopes remain undefined, and diverse experimental approaches to determine the epitopes have been

reported [32 – 36]. Experiments by Bressan and co-workers using Western blot analysis of six recombinantly expressed repeat domains (R2, R7, R9, R11, R25, and R51) revealed that the antibodies used in the CA125 II test (OC125 [31] and M11 [37]) do not recognize all repeat domains uniformly [35]. We hypothesized that variation in antibody recognition may be observed in other recombinant repeats and that the nature of the molecular recognition assay may influence whether binding is observed, particularly if the CA125 epitope is conformational [38].

Here we report the expression of nine recombinant repeats from the putative antigenic domain of CA125 (R2, R5, R6, R7, R9, R11, R25, R34, and R58, in the O'Brien numbering system, sequence alignment shown in SI Figure 1) [26]. Using Western blotting, indirect ELISA, and localized surface plasmon resonance (SPR) spectroscopy, we characterized the interactions of expressed and purified recombinant repeats with four CA125-binding monoclonal antibodies (OC125, M11, “OC125-like,” and “M11-like”). Consistent with our hypothesis and previous reports, the epitopes were found to be distributed nonuniformly over the recombinant repeats, and variation across assay method was observed. Without knowledge of the CA125 epitopes, it is impossible to characterize individual variation in MUC16 proteoforms or to determine whether MUC16 expression changes during cancer development, in response to treatment, or during recurrence. To improve the long-term survival of ovarian cancer patients, there is a vital need to improve the diagnostic value of CA125. This study represents the largest set of MUC16 recombinant repeats, the largest number of antibodies, and the most molecular interaction assay methods reported to date and contributes to ongoing efforts to understand the molecular nature and immunological activity of CA125. The rationale of this study is that its results will ultimately enable us to reinvent the CA125 test by developing new affinity reagents to

supplement or replace the antibodies in current use, since a biomarker is only as good as the tools available to detect it.

2 Materials and methods

2.1. Recombinant repeat expression and purification

The MUC16 coding sequence as described by O'Brien and co-workers [26] was obtained from NCBI (GenBank: AF414442.2). The nucleotide sequences of nine tandem repeats were synthesized and cloned into pET14b vector (GenScript, Piscataway, NJ), which was used to express protein with N' 6xHistidine-tagging (6xHis) by XhoI and BamHI sites. Each plasmid was transformed into the SHuffle T7 Express *E. coli* strain (New England Biolabs, Beverly, MA). Bacteria clones were grown in Luria–Bertani (LB) broth (Thermo, Waltham, MA) containing 100 µg/mL ampicillin at 30°C to exponential phase, and protein expression was induced with 400 µM IPTG for 4 hours. Cells were harvested and lysed by freeze-thaw cycles in phosphate buffered saline (PBS) containing cOmpletetTM protease inhibitor (Thermo). The 6xHis-tagged recombinant repeats were then purified on Ni-nitrilotriacetic acid (NTA) beads (Thermo) and verified by Western blot with anti-6xHis antibody and mass spectrometry as described below.

2.2. Mass spectrometry

The bottom-up proteomics workflow, including sample preparation and liquid chromatography-tandem mass spectrometry (LC–MS/MS), closely followed our previously published reports [39, 40]. Briefly, 10 µg of recombinant repeat protein was precipitated in 70% acetone at –20 °C overnight and resuspended in 100 mM triethylammonium bicarbonate (TEAB) buffer with 0.2%

deoxycholic acid (DCA), 6% sodium dodecyl sulfate (SDS), and 10 mM tris(2-carboxyethyl)phosphine (TCEP). Protein was denatured and reduced by incubation at 95°C for 10 min. Reduced protein was alkylated with 10 mM iodoacetamide (IAA) for 30 min at room temperature (RT) in the dark. The alkylation reaction was quenched by 1.2% phosphoric acid. The protein solution was spun onto an S-Trap device (Protifi, Farmingdale, NY) and digested by 750 ng trypsin (Promega, Madison, WI) in 100 mM TEAB buffer at 37°C overnight. Digested peptides were eluted by three buffers: 100 mM TEAB, 0.1% formic acid (FA) in water, and 50% acetonitrile (ACN) with 0.1% FA. Eluted peptides were combined and desalted using C18 ZipTips. Desalted peptides were reconstituted in water with 4% ACN containing 0.5% FA and analyzed by gradient elution–tandem mass spectrometry on a Waters NanoAcquity LC system coupled to a Q-Exactive mass spectrometer (Thermo).

2.3. *Western blot*

Recombinant repeats were resolved by 16% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to a polyvinylidene difluoride membrane. The membrane was blocked with 5% non-fat milk in TBS-T buffer (Tris buffered saline with 0.05% Tween 20) at RT for 1 hour. The membrane was hybridized at 4°C overnight with one of five primary antibodies: anti-6xHis (Thermo, HIS.H8, 1:2000), anti-CA125 epitope group A (Sigma, OC125, 1:200), anti-CA125 epitope group A (Fitzgerald, OC125-like, M61704, 1:2000), anti-CA125 epitope group B (Agilent, M11, 1:100), or anti-CA125 epitope group B (Fitzgerald, M11-like, M61703, 1:2000). The membrane was washed and hybridized with goat anti-mouse secondary antibody conjugated with horseradish peroxidase (HRP) (Thermo, 62-6520, 1:5000) at RT for 1 hour. Signals were

developed using Pierce ECL Western Blotting Substrate (Thermo), and membranes were imaged on a ChemiDoc system (Bio-Rad, Hercules, CA).

2.4. *ELISA*

Pierce Nickel-Coated Plates (Thermo) were used to immobilize recombinant repeats through their 6xHis tag for indirect ELISA. Five hundred ng of recombinant repeat in 100 μ L PBS buffer were used for immobilization. HE4 protein was used in place of recombinant repeats as a negative control, and wells with no immobilized protein were used as additional negative controls. After three washes with 200 μ L PBS containing 0.01% Tween 20, 100 μ L primary antibody—anti-6xHis (Thermo, HIS.H8, 1:2000), anti-CA125 epitope group A (Sigma, OC125, 1:200 or Fitzgerald, M61704, 1:2000) or anti-CA125 epitope group B (Agilent, M11, 1:100 or Fitzgerald, M61703, 1:2000)—was added to the well and incubated at RT for 1 hour. After three washes with 200 μ L PBS-T, 100 μ L anti-mouse HRP (Thermo, 62-6520, 1:2000) secondary antibody was added to the well and incubated at RT for 1 hour. Signals were developed using SuperSignal ELISA Femto Substrate (Thermo). Relative luminescence units at 425 nm were detected on a SpectraMax M5 plate reader (Molecular Devices, San Jose, CA).

2.5. *Surface plasmon resonance*

All SPR experiments were performed on a Nicoya Lifesciences OpenSPR-XT 2-channel instrument (Kitchener, ON) using Nicoya NTA-functionalized standard sensor chips. PBS with 0.05% Tween 20 (PBS-T; PBS from GrowCells, Irvine, CA; Tween 20 from Fisher Scientific, Waltham, MA) was used as both the immobilization and running buffer. Autosampler tray and sensor temperatures were set to 20°C. Recombinant repeat concentration was estimated on a

ThermoFisher NanoDrop 2000 (Waltham, MA) using the Protein A₂₈₀ function with the 1 Abs = 1 mg/mL setting. Recombinant repeats were diluted to 125 nM in PBS-T except for repeat 6, which required 500 nM to achieve acceptable immobilization levels. Regeneration was performed at the beginning of a run and after every ligand-analyte pair was injected, using two injections of 10 mM Glycine-HCl (VWR LifeSciences, Radnor, PA) and one injection of 350 mM EDTA (TCI America, Portland, OR).

After surface regeneration with EDTA, the surface was activated with 40 mM NiCl₂ (BeanTown Chemical, Hudson, NH) and recombinant repeat was injected to achieve ligand immobilization. Recombinant His-tagged streptavidin (Fitzgerald Industries, Acton, MA) at a concentration of 0.75 μM was used as a negative control molecule, and then antibody was injected as the analyte. M11 was injected at stock concentration, OC125 was diluted 1:200, and both M11-like and OC125-like were diluted 1:2000. The surface was then regenerated, and the repeat was re-immobilized for the next antibody. Three replicates were collected per recombinant repeat (R5 had an additional replicate). All injections were at a flow rate of 20 μL/min, except for Glycine-HCl (150 μL/min) and EDTA (100 μL/min). All injections had a dissociation time of 270 s. Corrected signals (reference channel subtracted from ligand-immobilized channel) were used to calculate average signals for antibody binding over the interval 500–525 seconds at the end of the dissociation phase using TraceDrawer Analysis 1.9.2 (Ridgeview Instruments, Uppsala, Sweden).

3 Results

3.1. Recombinant tandem repeat protein expression and confirmation

Recombinant repeats were expressed in *E. coli* and purified on Ni-NTA beads. Yield and purity of material covered from Ni-NTA beads were confirmed by visualization on polyacrylamide gel with Coomassie blue staining (SI Figure 3). All recombinant repeats were first analyzed by Western blot with anti-6xHis antibody to confirm expression (Figure 1). A single band at the expected molecular weight (~ 19.4 kDa) was observed for R2, R5, R6, R7 and R25. A minor additional band was observed for R9, R11, R34 and R58. The minor band could result from incomplete denaturation during SDS-PAGE sample preparation, protein degradation, or post-translational modification. The first two possibilities were ruled out by using stronger denaturing conditions for sample preparation and increasing the concentration of protease inhibitor in the lysis buffer. The recognition of both major and minor bands by anti-CA125 antibodies (Figure 1) suggests they are the same protein with different post-translational modifications that do not affect antibody binding. The amino acid sequences of recombinant repeats were verified by mass spectrometry using a proteomic workflow our lab previously reported for MUC16 detection. All recombinant repeat proteins have at least 95% coverage by mass spectrometry (Table 1 and SI Figure 2) except R6, for which 67% coverage was observed.

3.2. *Western blot*

Recombinant repeats were probed via Western blot with antibodies OC125, OC125-like, M11 and M11-like, and anti-6xHis antibody as a loading control (Figure 1). Epitope group A antibodies (OC125 and OC125-like), recognized repeats R5, R6, R9 and R25, with clear bands observed at the expected MW (~19.4 kDa). Epitope group B antibody M11 recognized all repeats with different affinities, while M11-like antibody strongly bound to R2 and had faint bands for R5, R6 and R25. The uneven affinity among different recombinant repeats and

antibodies is consistent with observations reported by Bressan and co-workers that epitopes were not uniformly distributed over the recombinant repeats they studied. In contrast to their observations, however, we do not detect binding between R11 and OC125 antibody, and we only observe weak binding of M11 antibody to R7 and R9. We observe that antibodies within the same epitope group have different recognition patterns at the level of individual expressed tandem repeats, and we suggest that this difference may also be observed on intact MUC16 *in vitro*. Further investigation is needed to confirm that these differences in binding are observed on intact MUC16.

3.3. ELISA

Recombinant repeats were immobilized onto Ni-NTA-coated 96-well plates for indirect ELISA. Luminescence values were normalized to the maximum signal for each antibody (Figure 2, shown as dot plots in SI Figure 4). Out of the four antibodies tested, OC125 showed evidence of binding to the fewest recombinant repeats (Figure 2A). OC125 showed maximum binding to R11 and strong binding to R5, R9, and R58, whereas R2, R6, and R7 had lower signals than the no-antigen control. OC125-like antibody (Figure 2B) displayed strong binding to R5, R7, R9, and R11, and little binding to R6 and R58. M11 (Figure 2C) had its maximum binding to R58, strong binding to R2, R5, R7, R9, and R11, and no apparent binding to R6. M11-like (Figure 2D) demonstrated maximum binding to R2, with strong binding to R5, R7, R9, R11, and R58. There were no repeats that showed no binding to any of the four antibodies, but R6 had smaller luminescence signals than the other recombinant repeats. This is not conclusive evidence that binding is not occurring, as the immobilization of protein within the wells was not quantified. The slightly higher luminescence signal in no antigen control wells compared to the HE4

controls indicates that nonspecific binding of antibodies to the wells may contribute to the luminescence signal. While the dilutions of antibodies varied, and the ranges of luminescence values varied from antibody to antibody, comparing the relative affinities of each clone to the repeats with one another can give some insight into the similarity of epitopes. Each antibody showed maximum binding signal to a different tandem repeat, indicating the binding locations may vary from antibody to antibody. OC125 and the OC125-like clone do not have similar patterns of binding from repeat to repeat (Figure 2A, 2B). M11 and the M11-like clone had similar binding patterns (Figure 2C, 2D).

3.4. *Surface plasmon resonance (SPR)*

Representative sensorgrams are found in SI Figure 6. Average corrected binding signal for each antibody interacting with recombinant repeat was normalized to the maximum signal for each antibody (Figure 3, shown as dot plots in SI Figure 5). A negative normalized corrected value occurs when signal resulting from non-specific binding to the sensor chip is greater than signal from recombinant repeat–antibody binding. Due to variable antibody dilution, direct comparisons based on corrected signals cannot be made. However, the pattern of normalized binding for each recombinant repeat–antibody pair demonstrates variability across combinations of recombinant repeat and antibodies. OC125 (Figure 3A) and M11-like (Figure 3D) show the strongest response to R58, while M11 (Figure 3C) shows the strongest binding to R2, and OC125-like (Figure 3B) shows the strongest binding to R5. As with the Western blot and indirect ELISA, the differences in binding patterns between the clinical antibodies and their “-like” counterparts suggest that the location of their epitopes differs. OC125 and the OC125-like antibody show variability in binding patterns: OC125 shows the strongest binding response to

R58, but OC125-like shows no binding to this repeat. Differences are also seen for M11 and the M11-like antibody: M11 shows stronger binding to R2 than to R5, while M11-like has similar affinity for both repeats. R34 and R58 also show different patterns of binding for M11 and M11-like: while the two repeats show comparable binding for M11, there is stronger binding to R58 for M11-like by a factor of 7.

4 Discussion

4.1. Comparison of analytical methods

Figure 4 displays the combined data for the three assays and enables visualization of differences in binding across recombinant repeat/antibody combinations (the raw data used to generate the heat map are found in SI Table 1). We observe that the extent of binding differs across the three analysis methods. One potential explanation of these differences is the nature of the assays: prior to Western blotting, recombinant repeats were denatured and treated to reduce disulfide bonds, while in ELISA and SPR assays the recombinant repeats remained as expressed with disulfide bonds intact. While it is not certain that the folded state of the expressed and purified recombinant repeats is structurally identical to the native state of a MUC16 tandem repeat *in vivo*, the discrepancy in antibody binding ability when recombinant protein is denatured supports the hypothesis that the epitopes for these antibodies is conformational [29, 36]. R6 and R25 are notable exceptions to the pattern. For these two recombinant repeats, Western blot results indicate antibody-repeat binding and the other binding analysis methods showed little evidence of interaction. In SPR experiments, R6 and R25 consistently had significantly lower immobilization signal response than other repeats, and yields during NTA affinity chromatography-based purification were consistently lower for these two recombinant proteins

compared to the other seven. This observation implies that the lower binding signals in SPR and ELISA for R6 and R25 may result at least in part from inefficient immobilization of these recombinant repeats. The patterns of which repeat–antibody combinations show strong, weak, or minimal binding evidence look similar between ELISA and SPR, two methods in which the recombinant repeats were not denatured and were immobilized in the same orientation. ELISA generally showed more consistent binding signals across recombinant repeats than did SPR. Corrections for nonspecific binding could account for these differences. ELISA results may have been affected by variable immobilization or nonspecific binding, which were not tracked through data collection. In SPR assays nonspecific binding signal was subtracted from binding signal.

4.2. *Working hypothesis regarding the nature of the CA125 epitopes*

Structural models for each of the nine tandem repeat proteins were obtained using the iTasser server [41 – 43]. In all cases, the top threading template was 7sa9, the human MUC16 SEA5 domain reported by White et al. [23]. This structure was also the top identified structural analog in the Protein Data Bank for all nine tandem repeat proteins, with good topological similarity, initiated by TM-scores of 0.749, 0.760, 0.760, 0.761, 0.763, 0.761, 0.758, 0.765, and 0.758 for repeats R2, R5, R6, R7, R9, R11, R25, R34, and R58, respectively, where a TM score of 1 indicates a perfect match. The sequence identity of the tandem repeats with 7sa9 were 0.861, 0.967, 0.943, 0.885, 0.795, 0.811, 0.992, 0.811, and 0.836, indicating that this previously reported model is an excellent template for other MUC16 tandem repeat domains. The predicted models overlay the MUC16 SEA5 domains well (SI Figure 7), with the notable exception of the proline-serine-threonine (PST) rich region which is largely unstructured in the iTasser predictions of the tandem repeats. The SEA5 domain structure does not contain the PST region.

The two major accessible faces noted by White et al. in their analysis of the crystal structure of SEA5 are seen in the tandem repeat domain proteins modeled here. It is not likely that the differences in antibody binding that we observe results from structural differences at these faces. Structural divergence between the SEA5 model (which contains a Serine at position 66) and the subset of tandem repeat proteins (R2, R7, R9, R11, R34, and R58) that contain Proline at that position is notable in the unstructured loop between beta strands that comprise the “C-loop” [32]. The center of the C-loop contains highly charged and polar amino acids and is found, in the model predicted by i-Tasser, to be adjacent to another loop that could be close enough to comprise a conformational epitope. Based on these studies, we hypothesize that the CA125 epitope involves the C-loop and charged and polar amino acids in the neighboring coil (residues 82-89). Notably, the tandem repeat protein that was found to be least immunogenic by Western blotting (R7) has aspartic acid replaced by asparagine within the loop that may be involved in conformational epitope presentation. Determining the exact location of the epitopes is the focus of our ongoing work.

5 Conclusions

This study contributes to ongoing efforts to identify the peptide epitopes of CA125. We anticipate that CA125 can be a more informative clinical biomarker if the nature of its recognition by antibodies is elucidated. Such knowledge may enable development of affinity reagents that recognize all repeats of MUC16. The current clinical test may under-report concentrations of CA125 in blood samples if the antibodies used for detection do not reliably recognize the subdomains in the immunogenic region, as data reported here suggest. Detection of smaller amounts of CA125 may be possible, enabling earlier detection of CA125 resurgence,

which correlates strongly with cancer recurrence. Future work following from this study includes structural characterization of free and antibody-bound recombinant repeats and the use of recombinant repeats as targets for the generation of novel affinity reagents such as nucleic acid aptamers, immunoaffinity reagents, and vaccines.

Acknowledgments

The authors thank Bill Boggess and the Notre Dame Mass Spectrometry and Proteomics Facility for technical assistance and Professor Brian Blagg for use of a ChemiDoc imager. The authors thank Professor Matthew Champion for helpful discussions and access to equipment for bacteria culture. Naviya Schuster-Little provided the structural alignment in SI Figure 1, and we thank her for many helpful discussions about protein structure prediction and comparisons. Daryl Good and Marko Jovic at Nicoya were pivotal in developing and optimizing SPR assays. This work was supported by award R21CA267532 from the National Cancer Institute and a Medical Research Program award from Tell Every Amazing Lady® About Ovarian Cancer Louisa M. McGregor Ovarian Cancer Foundation (T.E.A.L.®). Additional support came from a University of Notre Dame Advancing Our Vision Award in Analytical Science and Engineering (to RJW) and from the Department of Chemistry and the Office of Research at the University of Kansas.

Author contributions

Conception: CWW, EKH, LM, RJW

Interpretation or analysis of data: CWW, EKH, LM, RJW

Preparation of the manuscript: CWW, EKH, LM, RJW

Revision for important intellectual content: CWW, EKH, LM, RJW

Supervision: RJW

References

- [1] C. Kuk, V. Kulasingam, C. Gunawardana, C. Smith, I. Batruch, E. Diamandis, Mining the ovarian cancer ascites proteome for potential ovarian cancer biomarkers, *Molecular & Cellular Proteomics*, **8** (2009) 661-669.
- [2] A. Sinha, A. Hussain, V. Ignatchenko, A. Ignatchenko, K. Tang, V. Ho, B. Neel, B. Clarke, M. Bernardini, L. Ailles, T. Kislinger, N-Glycoproteomics of patient-derived xenografts: a strategy to discover tumor-associated proteins in high-grade serous ovarian cancer, *Cell Systems*, **8** (2019) 345-351.
- [3] M. Galey, A. Young, V. Petukhova, M. Wang, J. Wang, A. Salvi, A. Russo, J. Burdette, L. Sanchez, Detection of ovarian cancer using samples sourced from the vaginal microenvironment, *Journal of Proteome Research*, **19** (2020) 503-510.
- [4] P. Giampaolino, V. Foreste, L. Della Corte, C. Di Filippo, G. Iorio, G. Bifulco, Role of biomarkers for early detection of ovarian cancer recurrence, *Gland Surgery*, **9** (2020) 1102-1111.
- [5] H. Kim, M. Lim, J. Kim, S. Shin, Y. Cho, C. Cho, Circulating tumor cells enumerated by a centrifugal microfluidic device as a predictive marker for monitoring ovarian cancer treatment: a pilot study, *Diagnostics*, **10** (2020) 249.
- [6] C. Trinidad, A. Tetlow, L. Bantis, A. Godwin, Reducing ovarian cancer mortality through early detection: approaches using circulating biomarkers, *Cancer Prevention Research*, **13** (2020) 241-252.
- [7] H. Whitwell, J. Worthington, O. Blyuss, A. Gentry-Maharaj, A. Ryan, R. Gunu, J. Kalsi, U. Menon, I. Jacobs, A. Zaikin, J. Timms, Improved early detection of ovarian cancer using longitudinal multimarker models, *British Journal of Cancer*, **122** (2020) 847-856.

- [8] P. Giamougiannis, P. Martin-Hirsch, F. Martin, The evolving role of MUC16 (CA125) in the transformation of ovarian cells and the progression of neoplasia, *Carcinogenesis*, **42** (2021) 327-343.
- [9] I. Hellstrom, J. Raycraft, M. Hayden-Ledbetter, J. Ledbetter, M. Schummer, M. McIntosh, C. Drescher, N. Urban, K. Hellstrom, The HE4 (WFDC2) protein is a biomarker for ovarian carcinoma, *Cancer Research*, **63** (2003) 3695-3700.
- [10] R. Drapkin, H. von Horsten, Y. Lin, S. Mok, C. Crum, W. Welch, J. Hecht, Human epididymis protein 4 (HE4) is a secreted glycoprotein that is overexpressed by serous and endometrioid ovarian carcinomas, *Cancer Research*, **65** (2005) 2162-2169.
- [11] R. Moore, A. Brown, M. Miller, S. Skates, W. Allard, T. Verch, M. Steinhoff, G. Messerlian, P. DiSilvestro, C. Granai, R. Bast, The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass, *Gynecologic Oncology*, **108** (2008) 402-408.
- [12] R. Moore, D. McMeekin, A. Brown, P. DiSilvestro, M. Miller, W. Allard, W. Gajewski, R. Kurman, R. Bast, S. Skates, A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass, *Gynecologic Oncology*, **112** (2009) 40-46.
- [13] I. Hellstrom, P. Heagerty, E. Swisher, P. Liu, J. Jaffar, K. Agnew, K. Hellstrom, Detection of the HE4 protein in urine as a biomarker for ovarian neoplasms, *Cancer Letters*, **296** (2010) 43-48.
- [14] R. Moore, M. Jabre-Raughley, A. Brown, K. Robison, M. Miller, W. Allard, R. Kurman, R. Bast, S. Skates, Comparison of a novel multiple marker assay vs the Risk of Malignancy Index

for the prediction of epithelial ovarian cancer in patients with a pelvic mass, *American Journal of Obstetrics and Gynecology*, **203** (2010) 228.e1-e6.

[15] E. Potenza, G. Parpinel, M. Laudani, C. Macchi, L. Fuso, P. Zola, Prognostic and predictive value of combined HE-4 and CA-125 biomarkers during chemotherapy in patients with epithelial ovarian cancer, *International Journal of Biological Markers*, **35** (2020) 20-27.

[16] L. Salminen, K. Gidwani, S. Grenman, O. Carpen, S. Hietanen, K. Pettersson, K. Huhtinen, J. Hynninen, HE4 in the evaluation of tumor load and prognostic stratification of high grade serous ovarian carcinoma, *Acta Oncologica*, **59** (2020) 1461-1468.

[17] N. Fleming, L. Cass, C. Walsh, B. Karlan, A. Li, CA125 surveillance increases optimal resectability at secondary cytoreductive surgery for recurrent epithelial ovarian cancer, *Gynecologic Oncology*, **121** (2011) 249-252.

[18] G. Soletormos, M. Duffy, S. Abu Hassan, R. Verheijen, B. Tholander, R. Bast, K. Gaarenstroom, C. Sturgeon, J. Bonfrer, P. Petersen, H. Troonen, G. CarloTorre, J. Kulpa, M. Tuxen, R. Molina, Clinical use of cancer biomarkers in epithelial ovarian cancer updated guidelines from the European group on tumor markers, *International Journal of Gynecological Cancer*, **26** (2016) 43-51.

[19] K. Terry, H. Schock, R. Fortner, A. Husing, R. Fichorova, H. Yamamoto, A. Vitonis, T. Johnson, K. Overvad, A. Tjonneland, M. Boutron-Ruault, S. Mesrine, G. Severi, L. Dossus, S. Rinaldi, H. Boeing, V. Benetou, P. Lagiou, A. Trichopoulou, V. Krogh, E. Kuhn, S. Panico, H. Bueno-de-Mesquita, N. Onland-Moret, P. Peeters, I. Gram, E. Weiderpass, E. Duell, M. Sanchez, E. Ardanaz, N. Etxezarreta, C. Navarro, A. Idahl, E. Lundin, K. Jirstrom, J. Manjer, N. Wareham, K. Khaw, K. Byrne, R. Travis, M. Gunter, M. Merritt, E. Riboli, D. Cramer, R.

Kaaks, A prospective evaluation of early detection biomarkers for ovarian cancer in the European EPIC cohort, *Clinical Cancer Research*, **22** (2016) 4664-4675.

[20] O. Blyuss, M. Burnell, A. Ryan, A. Gentry-Maharaj, I. Marino, J. Kalsi, R. Manchanda, J. Timms, M. Parmar, S. Skates, I. Jacobs, A. Zaikin, U. Menon, Comparison of longitudinal CA125 algorithms as a first-line screen for ovarian cancer in the general population, *Clinical Cancer Research*, **24** (2018) 4726-4733.

[21] D. Warren, K. Nustad, J. Beard, T. O'Brien, Expression and epitope characterization of a recombinant CA 125 repeat: fourth report from the ISOBM TD-1 workshop, *Tumor Biology*, **30** (2009) 51-60.

[22] M. Felder, A. Kapur, J. Gonzalez-Bosquet, S. Horibata, J. Heintz, R. Albrecht, L. Fass, J. Kaur, K. Hu, H. Shojaei, R. Whelan, M. Patankar, MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress, *Molecular Cancer*, **13** (2014) 129.

[23] B. White, M. Patterson, S. Karnwal, C. Brooks, Crystal structure of a human MUC16 SEA domain reveals insight into the nature of the CA125 tumor marker, *Proteins-Structure Function and Bioinformatics*, **90** (2022) 1210-1218.

[24] B. Yin, K. Lloyd, Molecular cloning of the CA125 ovarian cancer antigen - Identification as a new mucin, MUC16, *Journal of Biological Chemistry*, **276** (2001) 27371-27375.

[25] B. Yin, A. Dnistrian, K. Lloyd, Ovarian cancer antigen CA125 is encoded by the MUC16 mucin gene, *International Journal of Cancer*, **98** (2002) 737-740.

[26] T. O'Brien, J. Beard, L. Underwood, R. Dennis, A. Santin, L. York, The CA 125 gene: An extracellular superstructure dominated by repeat sequences, *Tumor Biology*, **22** (2001) 348-366.

- [27] T. O'Brien, J. Beard, L. Underwood, K. Shigemasa, The CA 125 gene: A newly discovered extension of the glycosylated N-terminal domain doubles the size of this extracellular superstructure, *Tumor Biology*, **23** (2002) 154-169.
- [28] M. Nap, A. Vitali, K. Nustad, R. Bast, T. O'Brien, O. Nilsson, P. Seguin, M. Suresh, O. Borner, T. Saga, H. deBruijn, S. Nozawa, F. Kreutz, D. Jette, H. Sakahara, M. Gadnell, K. Endo, E. Barlow, D. Warren, E. Paus, S. Hammarstrom, P. Kenemans, J. Hilgers, Immunohistochemical characterization of 22 monoclonal antibodies against the CA125 antigen: 2nd report from the ISOBM TD-1 workshop, *Tumor Biology*, **17** (1996) 325-331.
- [29] K. Nustad, R. Bast, T. O'Brien, O. Nilsson, P. Seguin, M. Suresh, T. Saga, S. Nozawa, O. Borner, H. deBruijn, M. Nap, A. Vitali, M. Gadnell, J. Clark, K. Shigemasa, B. Karlsson, F. Kreutz, D. Jette, H. Sakahara, K. Endo, E. Paus, D. Warren, S. Hammarstrom, P. Kenemans, J. Hilgers, Specificity and affinity of 26 monoclonal antibodies against the CA 125 antigen: First report from the ISOBM TD-1 workshop, *Tumor Biology*, **17** (1996) 196-219.
- [30] P. Kenemans, G. van Kamp, P. Oehr, R. Verstraeten, Heterologous double-determinant immunoradiometric assay CA125 II: reliable second-generation immunoassay for determining CA125 in serum, *Clinical Chemistry*, **39** (1993) 2509-2513.
- [31] T. Klug, R. Bast, J. Niloff, R. Knapp, V. Zurawski, Monoclonal antibody immunoradiometric assay for an antigenic determinant (CA 125) associated with human epithelial ovarian carcinomas, *Cancer Research*, **44** (1984) 1048-1053.
- [32] Z. Berman, L. Moore, K. Knudson, R. Whelan, Synthesis and structural characterization of the peptide epitope of the ovarian cancer biomarker CA125 (MUC16), *Tumor Biology*, **31** (2010) 495-502.

[33] F. Weiland, K. Fritz, M. Oehler, P. Hoffmann, Methods for Identification of CA125 from Ovarian Cancer Ascites by High Resolution Mass Spectrometry, *International Journal of Molecular Sciences*, **13** (2012) 9942-9958.

[34] F. Weiland, K. Martin, M. Oehler, P. Hoffmann, Deciphering the Molecular Nature of Ovarian Cancer Biomarker CA125, *International Journal of Molecular Sciences*, **13** (2012) 10568-10582.

[35] A. Bressan, F. Bozzo, C. Maggi, M. Binaschi, OC125, M11 and OV197 epitopes are not uniformly distributed in the tandem-repeat region of CA125 and require the entire SEA domain, *Disease Markers*, **34** (2013) 257-267.

[36] L. Marcos-Silva, Y. Narimatsu, A. Halim, D. Campos, Z. Yang, M. Tarp, P. Pereira, U. Mandel, E. Bennett, S. Valchrushev, S. Levery, L. David, H. Clausen, Characterization of Binding Epitopes of CA125 Monoclonal Antibodies, *Journal of Proteome Research*, **13** (2014) 3349-3359.

[37] T. O'Brien, L. Raymond, G. Bannon, D. Ford, H. Hardardottir, F. Miller, J. Quirk, New monoclonal antibodies identify the glycoprotein carrying the CA 125 epitope, *American Journal of Obstetrics and Gynecology*, **165** (1991) 1857-1864.

[38] H. Davis, V. Zurawski, R. Bast, T. Klug, Characterization of the CA 125 antigen associated with human epithelial ovarian carcinomas, *Cancer Research*, **46** (1986) 6143-6148.

[39] N. Schuster-Little, S. Madera, R. Whelan, Developing a mass spectrometry-based assay for the ovarian cancer biomarker CA125 (MUC16) using suspension trapping (STrap), *Analytical and Bioanalytical Chemistry*, **412** (2020) 6361-6370.

- [40] N. Schuster-Little, R. Fritz-Klaus, M. Etzel, N. Patankar, S. Javeri, M.S. Patankar, R.J. Whelan, Affinity-free enrichment and mass spectrometry analysis of the ovarian cancer biomarker CA125 (MUC16) from patient-derived ascites, *Analyst*, **146** (2021) 85-94.
- [41] W. Zheng, C. Zhang, Y. Li, R. Pearce, E.W. Bell, Y. Zhang, Folding non-homology proteins by coupling deep-learning concept maps with I-TASSER assembly simulations. *Cell Reports Methods*, **1** (2021) 100014.
- [42] C. Zhang, P.L. Freddolino, Y. Zhang, COFACTOR: improved protein function prediction by combining structure, sequence, and protein-protein interaction information. *Nucleic Acids Research*, **45** (2017) W291-299.
- [43] J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*, **43** (2015) W174-W181.

Repeat	Sequence	Number	Coverage
<u>R2</u>	AAGPLLMPFTLNFTITNLQYEEDMRRTGSRKFNTMESVLQGLLKPLFKNTSVGPLYSGCRLTLLRPEKDGAAATGVD AIC THRLDPKSPGLNREQLYWELSKLTNDIEELGPYTLDRNSLYVNGFTHQSSVSTTSTPGTSTVDLRTSGTPSSLSSPTIM	12294- 12451	100%
R5	AAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLTLLRSEKDGAAATGVD AIC THRLDPKSPGVDRREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPST	12763- 12918	100%
R6	SAGPLLVPFTLNFTITNLQYEEDMRHPGSRKFNTTERVLQGLLKPLFKNTSVGPLYSGCRLTLLRSEKDGAAATGVD AIC THRLDPKSPGVDRREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPST	12919- 13074	67%
<u>R7</u>	SAGPLLVPFTLNFTITNLQYEEDMHHPGSRKFNTTERVLQGLLGPMFKNTSVGLLYSGCRLTLLRPEKDGAAATGMD AIC SHRLDPKSPGLNREQLYWELSQLTHGIKELGPYTLDRNSLYVNGFTHRSSVAPTSTPGTSTVDLGTSGTPSSLPST	13075- 13230	100%
<u>R9</u>	TAGPLLVPFTLNFTITNLQYEEDMRHPGSRKFNATERVLQGLLSPIFKNSSVGPLYSGCRLTSLRPEKDGAAATGMD AIC LYHPNPKRPGLDREQLYWELSQLTHNITEGLPYSLDRDLSLYVNGFTHQNSVPTTSTPGTSTVYVATTGTPSSFPGHT	13387- 13542	96%
<u>R11</u>	EPGPLLIPFTLNFTITNLHYEENMQHPGSRKFNTTERVLQGLLKPLFKNTSVGPLYSGCRLTLLRPEKHEAATGVD TIC THRVDPIGGLDRERLYWELSQLTNSITELGPYTLDRDLSLYVNGFNPRSSVPTTSTPGTSTVHLATSGTPSSLPST	13699- 13854	100%
<u>R25</u>	TAGPLLVPFTLNFTITNLQYEEDMHCPSGRKFNTTERVLQGLLGPMFKNTSVGPLYSGCRLTLLRSEKDGAAATGVD AIC THRLDPKSPGVDRREQLYWEKSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPST	15884- 16039	97%
R34	APVPLLIPFTLNFTITDLHYEENMQHPGSRKFNTTERVLQGLLKPLFKNTSVGPLYSGCRLTLLRPEKHGAATGVD AIC TLRLDPTGPGGLDRERLYWELSQLTNSITELGPYTLDRDLSLYVNGFNWSSVPTTSTPGTSTVHLATSGTPSSLPST	17287- 17442	100%
R58	ATGPVLLPFTLNFTITNLQYEEDMRHPGSRKFNTTERVLQGLLMPFKNTSVSSLYSGCRLTLLRPEKDGAAATRVDAIC THRPDPKSPGLDRERLYWELSQLTHGITELGPHYTLDRHLSLYVNGFTHQSSMTTTRTPDSTMHLATSRTPASLSGPT	21020- 21185	95%

Table 1. Recombinant repeats expressed and characterized in this study. Repeat numbering, amino acid sequences and corresponding amino acid number in reference sequence AF414442.2, as reported by O'Brien and co-workers [26], are shown in left three columns. The SEA domain extends from repeat 1 to 126 and is followed by a region rich in proline, serine, and threonine. Repeats with underlined numbers were previously examined by Bressan et al. [35]. Tandem repeat numbering increases from N- to C-terminus. Peptide coverage determined from bottom-up proteomics experiments to confirm expression is shown in the right column.

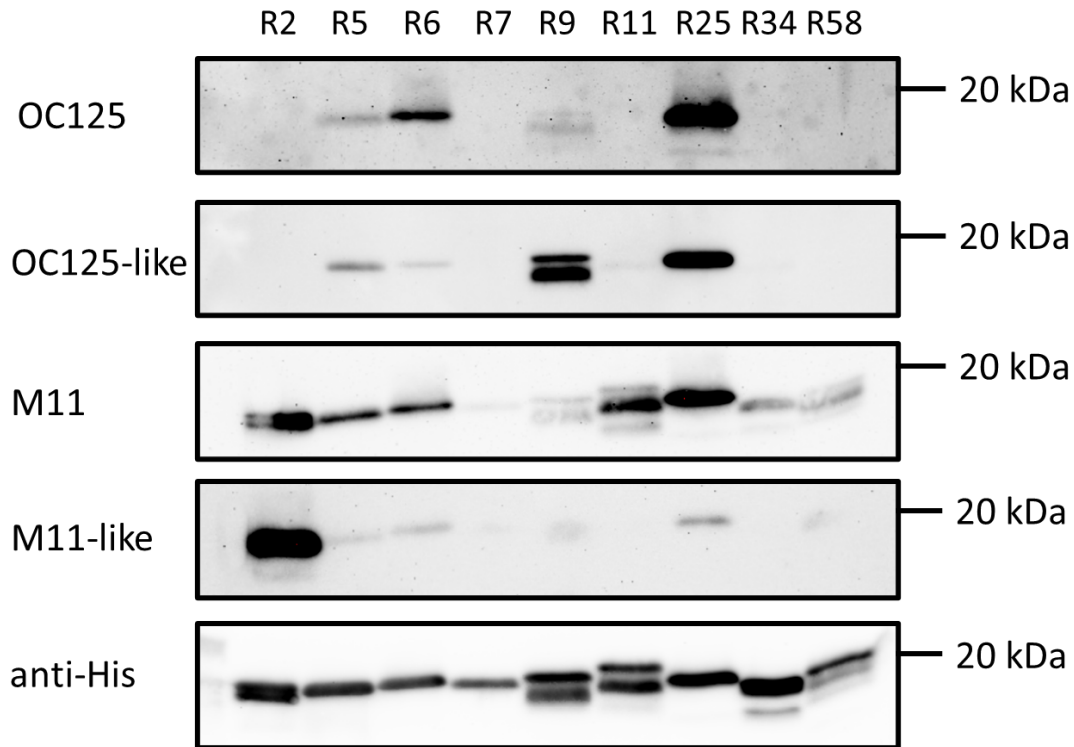


Figure 1. Western blot images of recombinant repeats from MUC16 probed with antibody clones OC125 (1:200), M61704 (OC125-like, 1:2000), M11 (1:100), and M61703 (M11-like, 1:2000), and an anti-His (His.H8, 1:2000) loading control.

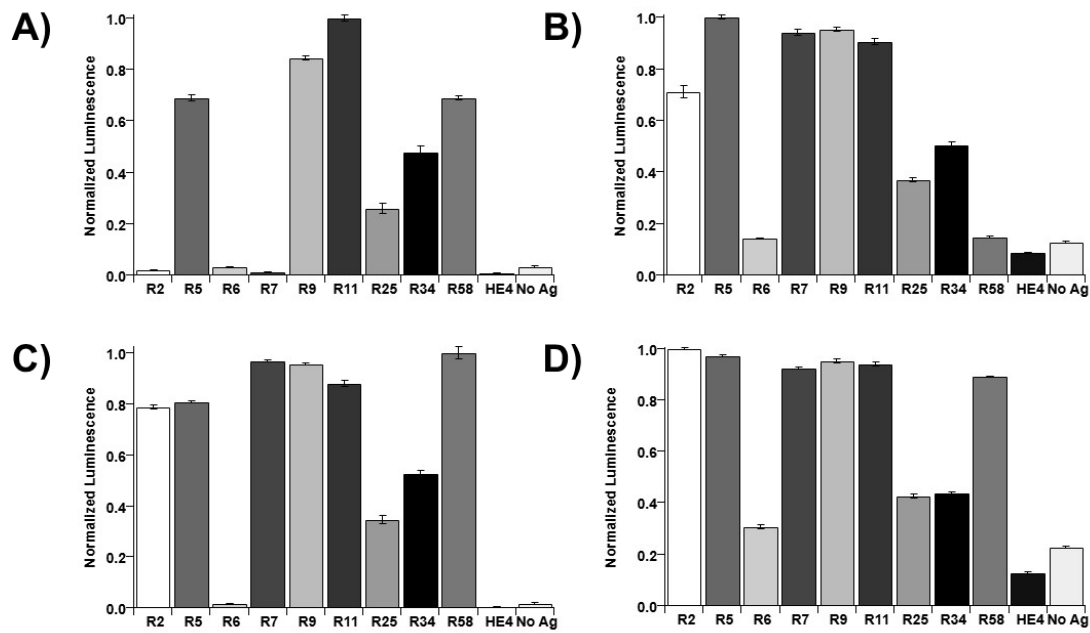


Figure 2. Results of indirect ELISA. Luminescent optical density values were normalized (to max = 1.0) within each data set. Recombinant repeats from MUC16 were probed with A) OC125 (1:200), B) M61704 (OC125-like, 1:2000), C) M11 (1:100), and D) M61703 (M11-like, 1:2000). Error bars are standard error of the mean (n = 3). HE4 and no antigen were used as negative controls.

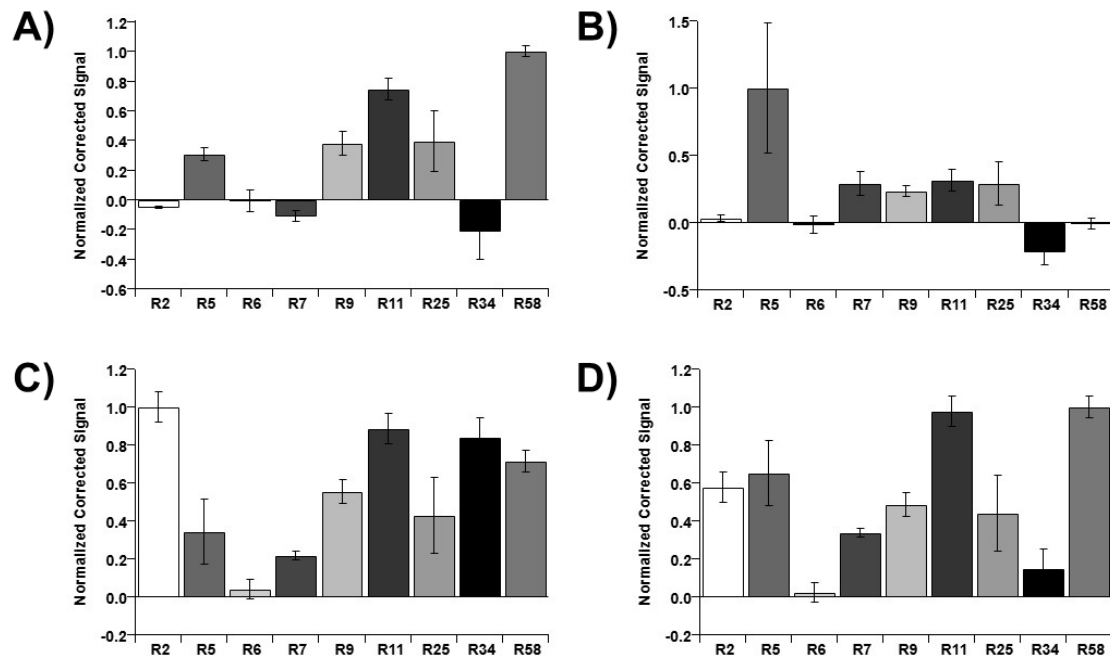


Figure 3. SPR data on binding of recombinant repeats from MUC16 to CA125 antibodies. Normalized (to max = 1.0) binding signals of A) OC125 (1:50), B) M61704 (OC125-like, 1:2000), C) M11 (stock concentration), and D) M61703 (M11-like, 1:2000) are shown. Error bars are standard error of the mean ($n = 3$, with the exception of R5, for which $n = 4$).

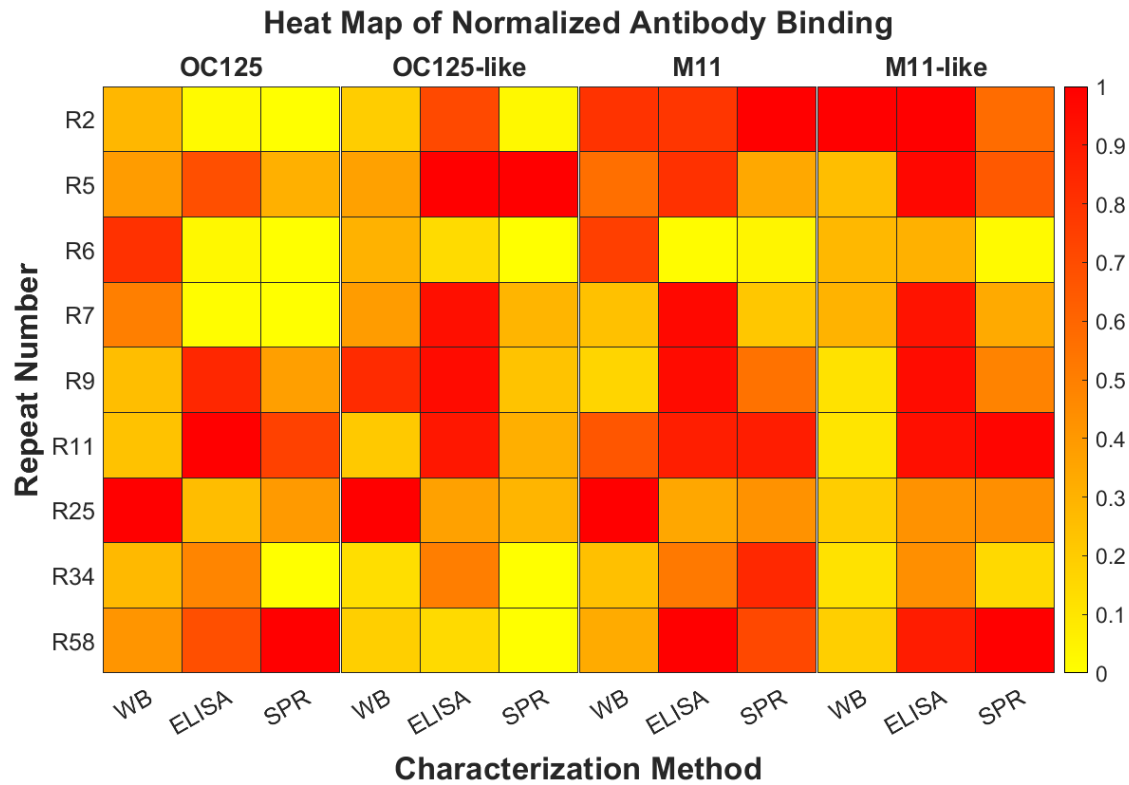
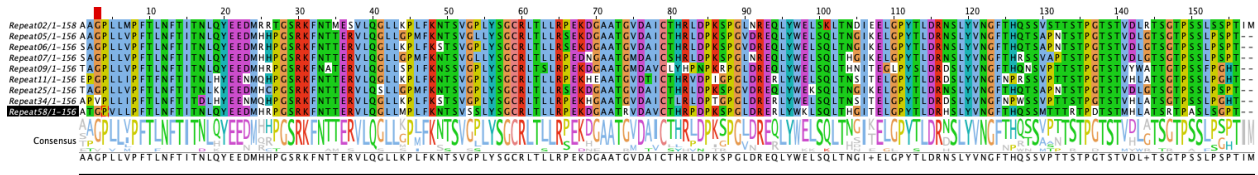


Figure 4: Heat map generated using normalized binding signals of antibodies to recombinant repeats from MUC16. Data used to generate the heat map are in SI Table 1. Any negative normalized binding signals were set to zero for the generation of the heat map.



SI Figure 1: Sequence alignment of the MUC16 tandem repeat proteins. The alignment was created using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and visualized using Jalview.

R2

AAGPLLPFTLNFTITNLQYEEDMRRTGSRKFNTMESVLOGLLKPLFKNTSVGPLYSGCRLLTLRPEKDGAAATGVDAICTHRLDPKSPGLNREQLYWELSKLTNDIEELGPYTLDRNSLYVNGFTHQSSVSTTSTPGTSTVDLRTSGTPSSLSPTIM

R5

AAGPLLPFTLNFTITNLQYEEDMHHHPGSRKFNTERVLOGLLGPMFKNTSVGLLYSGCRLLTLRSEKDGAAATGVDAICTHRLDPKSPGVDRREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPPT

R6

SAGPLLPFTLNFTITNLQYEEDMRHPGSRKFNTERVLOGLLKPLFKNTSVGPLYSGCRLLTLRSEKDGAAATGVDAICTHRLDPKSPGVDRREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPPT

R7

SAGPLLPFTLNFTITNLQYEEDMHHHPGSRKFNTERVLOGLLGPMFKNTSVGLLYSGCRLLTLRPEKGAATGMDAICSHRLDPKSPGLNREQLYWELSQLTHGIKELGPYTLDRNSLYVNGFTHRSSVAPTSTPGTSTVDLGTSGTPSSLPSPPT

R9

TAGPLLPFTLNFTITNLQYEEDMRHPGSRKFNTERVLOGLLSPIFKNSSVGPLYSGCRLLTLRPEKDGAAATGMDAVCLYHPNPKRPGLDREQLYWELSQLTHNITELGPYSIDRDSLIVNGFTHQNSVPTTSTPGTSTVYVWATTGTPSSFPGHT

R11

EPGPLLPTLNFTITNLHYEENMQHPGSRKFNTERVLOGLLKPLFKNTSVGPLYSGCRLLTLRPEKHAATGVDTICTHRVDPGGLDRERLYWELSQLNSITELGPYTLDRDSLIVNGFNPRSSVPTTSTPGTSTVHLATSGTPSSLPGHT

R25

TAGPLLPFTLNFTITNLQYEEDMHCPSGRKFNTERVLOGLLGPMFKNTSVGPLYSGCRLLTLRSEKDGAAATGVDAICTHRLDPKSPGVDRREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLGTSGTPSSLPSPPT

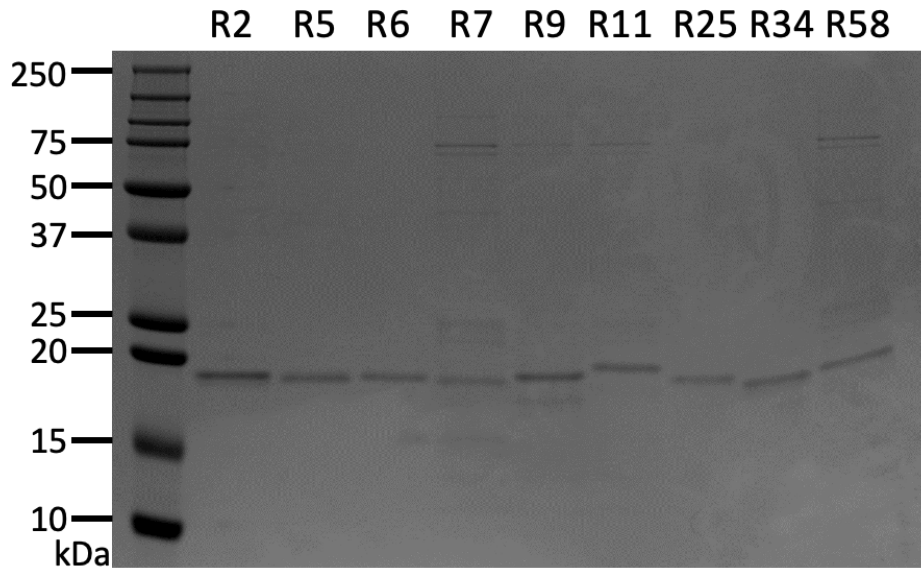
R34

APVPLLPTLNFTITDLHYEENMQHPGSRKFNTERVLOGLLKPLFKNTSVGPLYSGCRLLTLRPEKGAATGVDAICTRLDPTGGLDRERLYWELSQLNSITELGPYTLDRDSLIVNGFNPWSSVPTTSTPGTSTVHLATSGTPSSLPGHT

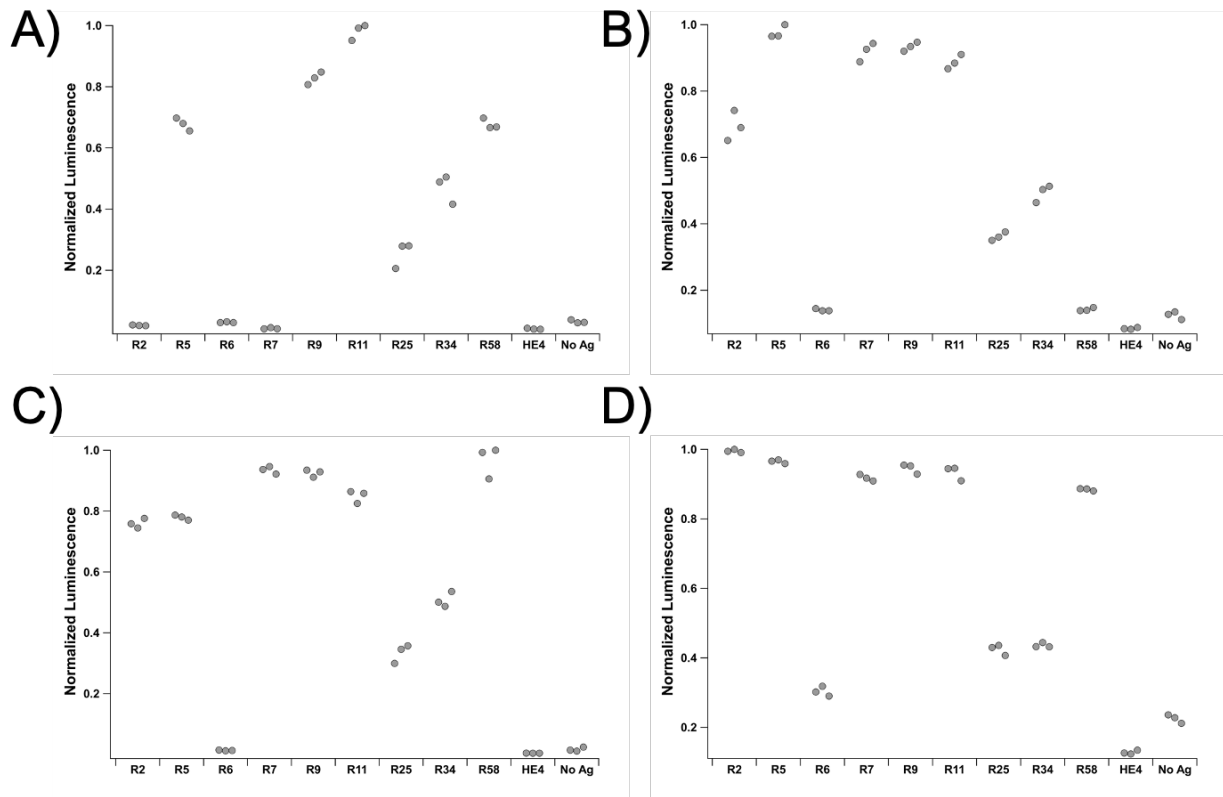
R58

ATGPVLLPFTLNFTITNLQYEEDMRHPGSRKFNTERVLOGLLMPLFKNTSVSSLYSGCRLLTLRPEKDGAAATRVDAVCTHRPDPKSPGLDRERLYWKLSQLTHGITEGYPYTLDRHSLIVNGFTHQSSMITTRTPDTSTMHLATSRTPASLSGPT

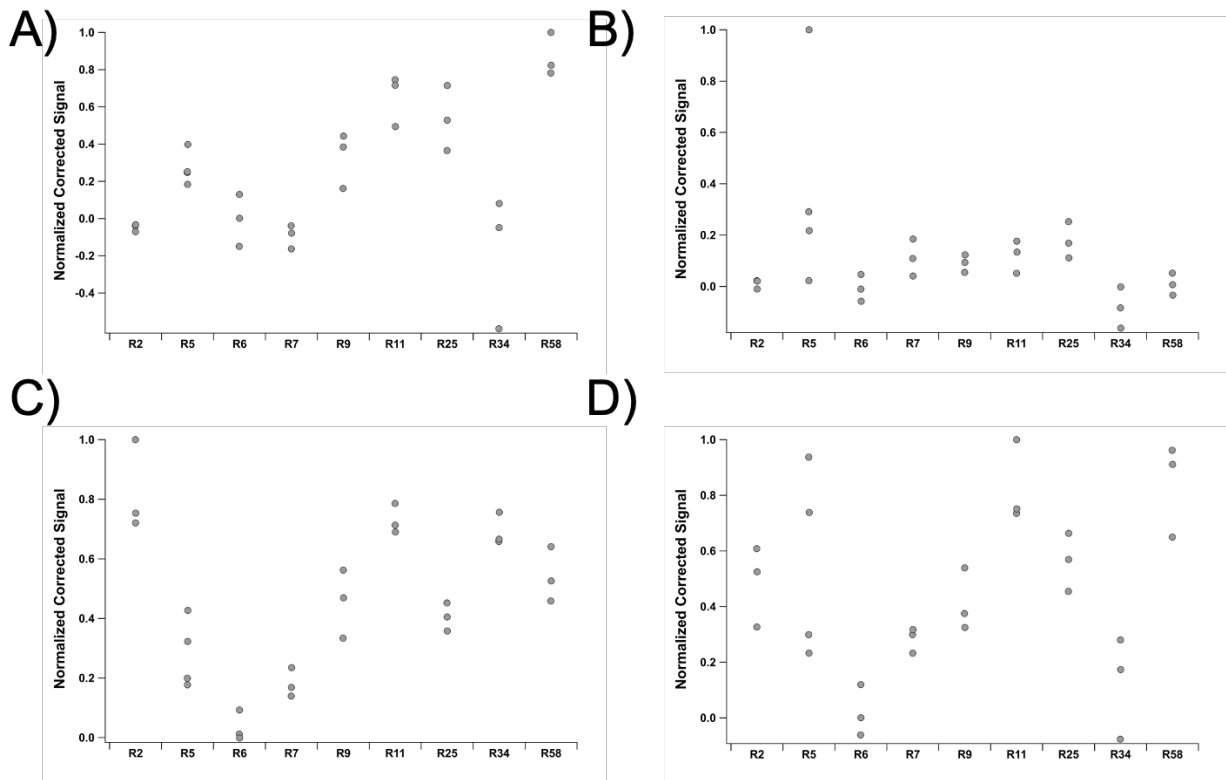
SI Figure 2: Mass spectrometry coverage map. The nine recombinant repeats studied here are shown in one-letter amino acid notation. Underlined amino acids were detected in tryptic peptides via LC-MS/MS.



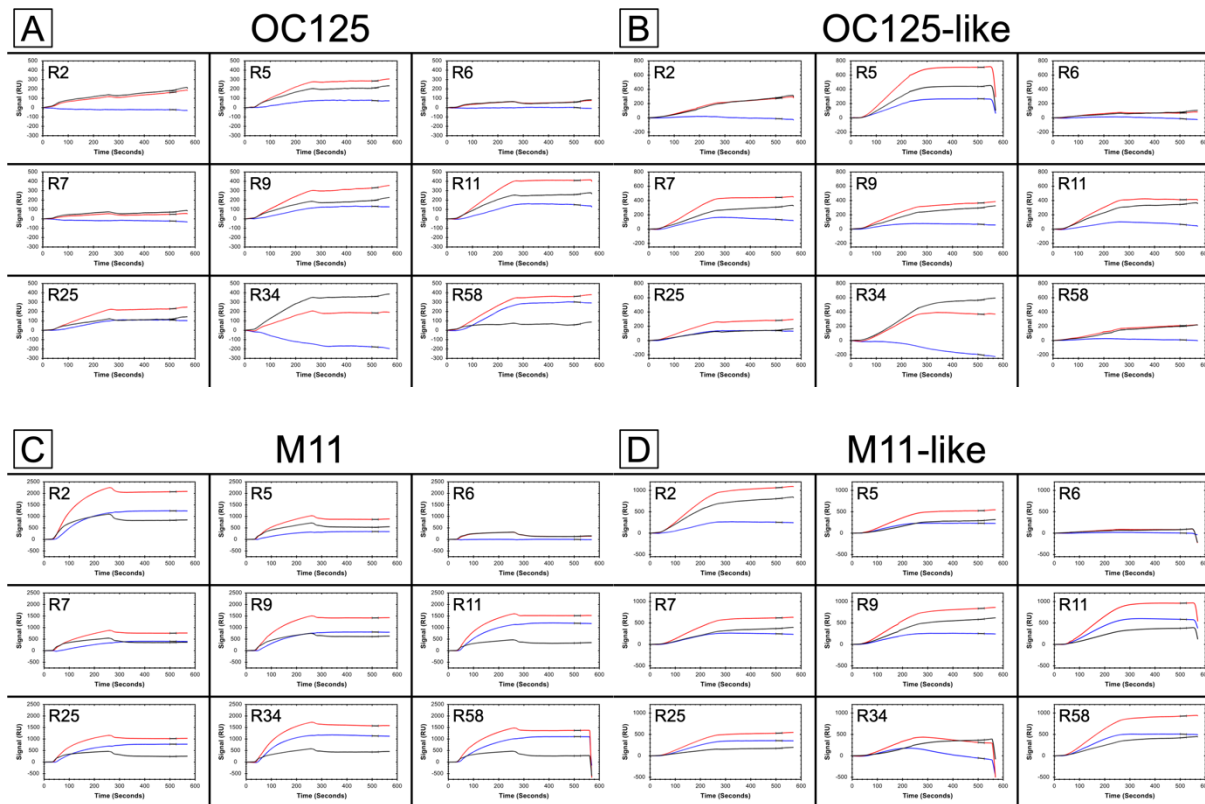
SI Figure 3. Coomassie blue stained gel image of recombinant repeats from MUC16 purified on Ni-NTA beads. Proteins eluted from Ni-NTA beads were separated by 16% SDS-PAGE at 100V for 2 hours. The gel was stained by Coomassie brilliant blue R-250 (Bio-Rad) and imaged on an Azure 400 imaging system (Azure).



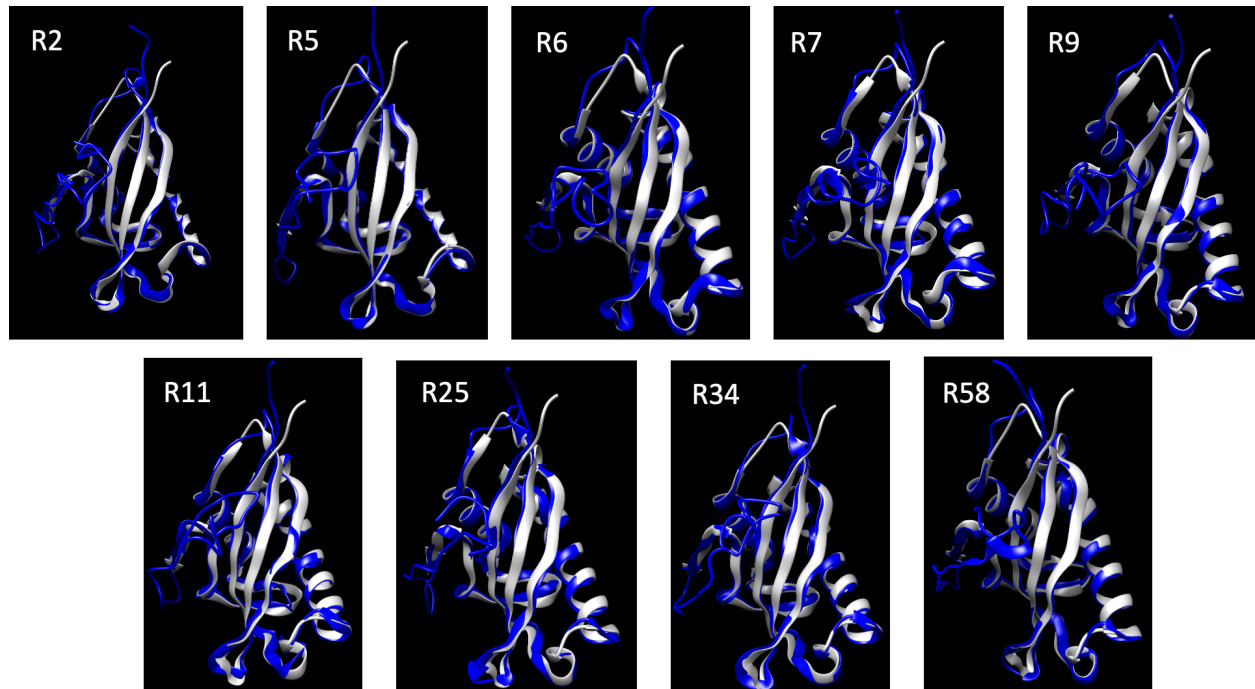
SI Figure 4. Results of indirect ELISA, same as Figure 2, shown as a dot plot. Luminescent optical density values were normalized (to max = 1.0) within each data set. Recombinant repeats from MUC16 were probed with A) OC125 (1:200), B) M61704 (OC125-like, 1:2000), C) M11 (1:100), and D) M61703 (M11-like, 1:2000).



SI Figure 5. SPR data on binding of recombinant repeats from MUC16 to CA125 antibodies, same as Figure 3, shown as a dot plot. Normalized (to max = 1.0) binding signals of A) OC125 (1:50), B) M61704 (OC125-like, 1:2000), C) M11 (stock concentration), and D) M61703 (M11-like, 1:2000) are shown.



SI Figure 6: Representative sensorgrams displaying real-time binding of antibodies to recombinant repeats. A) OC125, B) OC125-like, C) M11, and D) M11-like. The black trace corresponds to Channel 1 signal, the red trace corresponds to Channel 2, and the blue trace shows Corrected Signal (Channel 2 – Channel 1). The average value of the Corrected Signal over the interval between 500 and 525 seconds was used for quantitative assessment of binding.



SI Figure 7: Ribbon models of the MUC16 tandem repeat proteins predicted by iTasser (blue) and overlaid with the human SEA domain 5 (7sa9) as determined by White et al. (2022) (light grey). The cysteine loop (C-loop) comprises the middle of the right two antiparallel beta strands and the intervening loop (oriented toward the bottom of the image). Overlaid plots prepared in Chimera.

	OC125			OC125-like			M11			M11-like		
	WB	ELISA	SPR	WB	ELISA	SPR	WB	ELISA	SPR	WB	ELISA	SPR
R2	0.283	0.020	-0.054	0.197	0.711	0.029	0.797	0.786	1	1	1	0.575
R5	0.388	0.691	0.31	0.369	1	1	0.565	0.807	0.342	0.255	0.970	0.652
R6	0.805	0.030	-0.007	0.302	0.144	-0.018	0.752	0.014	0.042	0.279	0.305	0.023
R7	0.502	0.010	-0.107	0.389	0.941	0.291	0.246	0.968	0.219	0.297	0.923	0.334
R9	0.257	0.844	0.378	0.830	0.956	0.237	0.165	0.957	0.552	0.114	0.950	0.488
R11	0.239	1	0.746	0.211	0.908	0.316	0.661	0.879	0.885	0.103	0.938	0.978
R25	1	0.26	0.395	1	0.371	0.291	1	0.346	0.426	0.198	0.426	0.440
R34	0.274	0.479	-0.213	0.130	0.505	-0.216	0.248	0.526	0.841	0.117	0.438	0.149
R58	0.417	0.691	1	0.188	0.145	-0.010	0.329	1	0.715	0.188	0.889	1

SI Table 1: Data used to generate heat map in Figure 4.