10-1-2010

Synthesis and structural characterization of the peptide epitope of the ovarian cancer biomarker CA125 (MUC16)

Rebecca J. Whelan
Oberlin College

Zach T. Berman

Lee J. Moore
Oberlin College

Kathleen E. Knudson

Title: Synthesis and Structural Characterization of the Peptide Epitope of the Ovarian Cancer Biomarker CA125 (MUC16)

Authors: Zach T. Berman, Lee J. Moore, Katherine E. Knudson, and Rebecca J. Whelan

Institution: Department of Chemistry and Biochemistry, Oberlin College

Short title: CA125 Peptide

Full address of corresponding author:
Rebecca J. Whelan
Department of Chemistry and Biochemistry
119 Woodland Street
Oberlin College
Oberlin OH 44074

Telephone: 440-775-8941

Fax: 440-775-6682

E-mail: rwhelan@oberlin.edu

Key words: CA125, MUC16, ovarian cancer, biomarker, synthetic peptide, repeat domain, epitope, infrared spectroscopy, FTIR

Abstract:

A highly conserved region of 21 amino acids flanked by cysteine residues, contained within a larger repeated domain, has been proposed to be the antibody-binding site in the ovarian cancer biomarker CA125 (MUC16). In this study solid-phase peptide synthesis with Fmoc protection chemistry was used to assemble a 21-mer peptide corresponding to the most frequently occurring antibody binding sequence in CA125. Potentially significant sequence variants were also synthesized. Peptide secondary structure was investigated using Fourier transform infrared spectroscopy, revealing the consensus sequence peptide to be largely unstructured at physiological pH whether the cysteine residues were reduced or were oxidized to form an intramolecular disulfide bond. Substitution of serine for proline at position 8 (P8S) results in β-sheet formation in peptides involved in intramolecular disulfide bonds. This β-sheet structure does not persist in peptides incapable of intramolecular disulfide bonding because of sequence, nor in peptides treated with the reducing agent dithiothreitol. In CA125 P8S is predicted to occur in ~25% of repeat domains, suggesting that this structural motif is a non-negligible contributor to overall structure and function. These findings suggest that future structural characterization efforts of CA125 should be especially mindful of the amino acid sequence and oxidation state of the protein.

Introduction

Ovarian cancer antigen CA125 (MUC16) is the best-validated and most widely assayed biomarker of the disease, which is responsible for approximately 14,600 deaths in the United States annually [1]. Originally identified by Bast et al. as the antigenic determinant of the OC125 antibody [2], CA125 was shown in 1983 to be elevated in a substantial percentage of patients with surgically demonstrated ovarian carcinoma and to correlate well with progression or regression during treatment [3]. Although isolated measurements of serum CA125 are insufficiently sensitive and specific to be used for early screening, serial measurements over time may be highly specific [4]. Extensive epidemiological studies have investigated the power of CA125 assays to predict ovarian cancer survival; this literature was recently reviewed [5]. It is expected that as new ovarian cancer biomarkers continue to be discovered and validated, simultaneous assays for CA125 in combination with other biomarkers will emerge that are more informative than assays of CA125 alone [6, 7]. However, CA125 remains the "gold standard" against which other ovarian cancer biomarkers are compared.

Relatively little is known about the structure of CA125, although meaningful progress has been made in the last decade, most significantly with the cloning of the CA125 gene, reported independently by two groups in 2001 [8, 9], with follow-up reports in 2002 expanding and refining the original description [10, 11]. Insights on the CA125 gene enabled the description of the CA125 protein as containing the following structural elements: (1) a massive N-terminal domain rich in serine and threonine residues that serve as sites of O-linked glycosylation, (2) 60+ tandem repeats that contain the antibody binding sites, probably within a highly conserved cysteine-bound region 21 amino acids long, (3) a transmembrane domain, and (4) a cytoplasmic

C-terminus containing sites for phosphorylation. Of particular interest to investigators wishing to understand the CA125/antibody interaction is the site (or sites) at which antibody binding occurs. In their study, O'Brien and co-workers determined the extracellular repeat domain as comprising 156 amino acid segments, each containing a loop formed by intramolecular disulfide bonding of two conserved cysteine residues. This "C-loop" is believed to contain the binding site for OC125 and M11-type antibodies [8]. A significant, possibly immunological, role for the amino acids flanked by these two cysteines is suggested on the basis of the cysteines' total conservation, the role of the C-loop in pushing the amino acids it contains away from the protein core, the hydrophilicity of the amino acids at the center of the loop, and the relative absence of nearby glycosylation. These observations are consistent with greater availability of the amino acids in the C-loop for interaction with antibodies or other binding partners. The independent work of Yin and Lloyd identified an incomplete set of partially conserved tandem repeats (156 amino acids each) each containing two conserved cysteine residues [9], consistent with the observations of O'Brien and co-workers. Despite the valuable insights afforded by gene cloning efforts, work remains to be done: there is currently no solid- or solution-phase structural model of human CA125. Detailed knowledge of the structure of CA125 will undoubtedly aid the development of novel detection and therapy schemes.

One potentially valuable approach to the study of CA125's structure is the characterization of individual repeat domains, or of peptides comprising sub-sections of those domains. This approach is particularly suitable to the study of mucins, in which repeat domain structure is ubiquitous [12, 13]. The investigation of a single repeat domain can provide insights on the structure and function of physiologically and clinically important mucins in a smaller, more tractable form. For example, synthetic peptides representing the epitope region of the breast cancer-associated mucin MUC1 have been particularly fruitful objects of study. Peptides ranging from 5 to 120 amino acids in length have been tested for their interaction with International Society of Oncology and BioMarkers (ISOBM) antibodies [14], and overlapping 20-mers have been used to map the amino acid sequence of the antibody-binding site [15]. The structure of MUC1 peptides and their interactions with known antibodies have been studied via 2D NMR [16]. Multiple repeats of MUC1 synthetic peptides have been shown to achieve native conformation [17], giving credibility to their use in characterizing the structure and function of native MUC1. DNA aptamers (high affinity nucleic acid ligands) have been selected against MUC1 synthetic peptides of differing lengths; these aptamer molecules exhibit affinity for native MUC1 and for MUC1-expressing breast cancer cells and may be useful in the development of novel diagnostic assays [18].

In contrast with the well-established use of small peptides in the investigation of MUC1, this strategy has not yet been widely applied in studies of CA125/MUC16. Recently reported work represents a significant step in this direction: in an extension of earlier efforts towards the structural characterization of CA125, Warren et al. reported the expression and epitope characterization of the 156-amino acid repeat domain of CA125 [19]. They found that although a number of antibodies were observed to interact with single 156-amino acid repeats, the prototypical CA125-specific antibody (OC125) showed low binding activity. Binding of OC125 was enhanced by prior incubation with antibodies of group B (M11-like), suggesting an "activation" of epitopes within the expressed 156-amino acid peptide. The authors predicted that the availability of a recombinant CA125 would be of great use in clarifying the biological functions of intact CA125.

In this report, we describe a complementary strategy for obtaining structure/function information about CA125. Rather than expression of a protein or peptide, we used chemical synthesis to produce the putative functional region of CA125: the 21-amino acid domain to which antibody binding is proposed to occur. Solid-phase synthesis enables precise control over the primary sequence of the peptide and gives a product that is easily subjected to structural characterization [20]. With it, we have made the predominant sequence and several important variants. We then used Fourier transform infrared spectroscopy (FTIR) to study the peptides' secondary structure. The infrared (IR) spectra of proteins and peptides contain structural information largely as a result of the sensitivity of the amide I band (1600 to 1700 cm^{-1}), which arises primarily from carbonyl stretching in the peptide backbone, to its local environment. The frequency of the amide I band may vary by more than 50 cm^{-1} in different secondary structures [21] and the spectra associated with structural features such as α-helices and β-sheets have been rationalized by quantum mechanical calculations and characterized experimentally in a variety of proteins. In addition, the amino-acid side-chain IR bands have been well characterized [22]. In this study, the IR spectra of a series of model peptides were measured and the position of the amide I bands were interpreted to gain insight on the structure of a small, but likely important, section of the CA125 repeat domain.

Materials and Methods

Peptide Synthesis

Peptides were synthesized using a PS3 automated peptide synthesizer from Protein Technologies (Tucson, AZ) and 9-fluorenylmethoxycarbonyl (Fmoc)-protected amino acids from Novabiochem/EMD Chemicals (Gibbstown, NJ). Insoluble resin beads functionalized with the Fmoc-protected C-terminal amino acid were placed in a reaction flask; protected amino acids and activator (HBTU) were loaded together into individual vials. The resin was swelled with *N*,*N*-dimethylformamide (DMF) followed by deprotection of the N-terminus with 20% piperidine in DMF. Amino acid coupling proceeded in the presence of 0.4 M *N*-methylmorpholine in DMF. The time allotted for coupling can be changed depending on the reaction efficiency but was typically 20 min. Inclusion of two pseudoproline dipeptides—Fmoc-Leu-Thr($\Psi^{Me,Me}$pro) in positions 3,4 and Fmoc-Ala-Thr($\Psi^{Me,Me}$pro) in positions 14,15—was found to be essential for obtaining high yields.

Completed peptide on-resin was washed twice with dichloromethane:methanol (3:2) and once with neat dichloromethane before being dried in a vacuum desiccator. A cleavage mixture containing 92.5% trifluoroacetic acid (TFA), 2.5% ethanedithiol (EDT), 2.5% water, and 2.5% triisopropyl silane was mixed with the resin and stirred for 3.5 hr to cleave the peptide from the resin and remove side-chain protecting groups. The mixture was filtered through glass wool to remove the resin, and the volume of the peptide-containing solution was reduced to one-quarter under a stream of nitrogen. Addition of excess ice-cold diethyl ether precipitated the peptide, which was then pelleted by centrifugation at 10,000 rpm for 10 min, followed by two washings with ice-cold ether. The peptide pellet was dissolved in 10% aqueous acetic acid, frozen in liquid nitrogen, and lyophilized. Product identity and purity were confirmed by electrospray ionization mass spectrometry (MS) (in-house) and amino acid analysis (Commonwealth Biotechnologies, Richmond, VA and Biosynthesis, Lewisville, TX).

Most peptides were prepared as peptide acids with the C- and N-termini uncapped. To prepare capped peptides, synthesis was performed using Rink amide resin. Prior to cleavage the

peptide was treated with acetic anhydride to acetylate the N-terminus, and cleavage using TFA produced the C-terminal amide.

Peptide Purification

Crude peptide was purified by semi-preparative high performance liquid chromatography using a Varian Microsorb column (5 µm particle packing, C18 stationary phase, 21.4 mm diameter, 25 cm length). Chromatography used gradient elution with 0.1% (v/v) TFA in water as the aqueous phase and 0.1% (v/v) TFA in acetonitrile as the organic phase. The separation was followed by monitoring peptide backbone absorbance at 214 nm. Fractions containing the desired peptide product, as confirmed by MS analysis, were collected and lyophilized. Purified lyophilized peptide was dissolved in 50 mM HCl and lyophilized again to remove residual TFA. Acid-treated peptide was then oxidized to form the intramolecular disulfide bond by preparing a < 1 mg/mL aqueous solution at pH 7 or pH 11. Peptide oxidation was followed by colorimetric assay using Ellman's reagent or by monitoring the m/z values of both the oxidized and reduced forms of the peptide by MS (described below.) All peptides were subjected to the same oxidation treatment whether or not they were capable of intramolecular disulfide bonding. Peptide was frozen and lyophilized again, then stored in a desiccator at 4˚C until needed.

Determination of Oxidation State

Peptide oxidation state was determined either by colorimetric assay using Ellman's reagent or by MS. Ellman's reagent (5,5'-dithio-*bis*-(2-nitrobenzoic acid); Pierce, Rockford, IL) reacts quantitatively with free thiols under neutral and alkaline conditions to give the yellow 2-nitro-5-thiobenzoate dianion (λ_{max} = 412 nm; ε = 14,150 $M^{-1}cm^{-1}$). Absorbance at 412 nm was monitored by spectrophotometry, and plots of A_{412} versus time were fit with single exponentials to determine oxidation half-life ($t_{1/2}$). Selected ion monitoring (SIM) mode of the mass spectrometer was used to find abundances of the reduced and oxidized forms of the peptide. Plots of the m/z values corresponding to oxidized peptide were plotted versus time and fit with single exponentials to find $t_{1/2}$. Fits to the reduced peptide signal (m/z) versus time gave the same $t_{1/2}$ values, but fits to the oxidation data were found to be more robust, consistent with the fact that reduced peptide is consumed completely in the oxidation reaction.

FTIR Spectroscopy

Purified peptide was dissolved in 100 mM phosphate buffer (PB) in D_2O at pH 7.4 to a final concentration of ~20 mg/mL. Samples were placed between CaF_2 plates separated by a 100-µm PTFE spacer. The temperature of the sample holder was varied between 5˚C and 75˚C by means of a circulating thermostatted water bath. FTIR spectra were collected in a dry nitrogen atmosphere using a Bruker Vector 22 spectrometer (Bruker Optics, Billerica, MA). For each spectrum, 512 scans were collected over a range of 4000 to 400 cm^{-1} with 4 cm^{-1} resolution. The spectrum of the blank buffer (recorded at the same temperature) and of water vapor (collected each day) were digitally subtracted from each peptide spectrum using Grams spectroscopy software (Thermo Scientific, Waltham, MA).

Results

Identification of the Target Sequence

Using the CA125 sequence data reported by O'Brien et al., we examined the 63 unique instances of the antigenic repeat domain [8]. Table 1 shows the relative abundances of amino acids at each position in the putative peptide epitope. Using this information, we identified the peptide sequence comprised of the most commonly occurring residues (position 1 to 21, reading from amino to carboxyl): CRLTLLRPEKDGAATGVDAIC. This peptide corresponds to positions 59 to 79 in the larger 156-amino acid repeat domain. Although this particular sequence only occurs 8 times out of the 63 repeats, it is the most common sequence, and we hypothesized that it is the peptide most likely to be recognized by CA125 antibodies. The same target peptide is selected using the shorter deduced amino acid sequence reported by Yin and Lloyd [9]. We also identified variants that we expected to possess significant structural features, such as the substitution of Ala for Cys at either terminus (which represents a peptide unable to form an intramolecular disulfide bond or C-loop) and the substitution of Ser for Pro at position 8 (proline is often important in determining structure but is only present in 60% of repeat domains). Table 2 lists the peptides synthesized and characterized in this study.

Oxidation studies

Peptide 1 was found to oxidize spontaneously in air under mild conditions (10 mM phosphate buffered saline, PBS, pH 7.4) with $t_{1/2} = 38.5 \pm 5.5$ min. Raising the pH to 11 by drop-wise addition of aqueous sodium hydroxide increased the rate of air oxidation, with $t_{1/2} = 5.5 \pm 0.2$ min. Peptide 4, in which proline at position 8 is replaced with serine, exhibited identical oxidation kinetics to peptide 1, with $t_{1/2} = 5.7 \pm 0.3$ min at pH 11. Peptides with uncharged, capped ends exhibited slow air oxidation kinetics ($t_{1/2} > 10$ hr) in PBS at physiological pH, but oxidized readily at pH 11 ($t_{1/2} = 30 \pm 1$ min).

FTIR spectroscopy

IR spectra were acquired for peptides dissolved in PB/D_2O, which provides a neutral solution of physiological ionic strength while avoiding spectroscopic interferences from H_2O. To avoid interference from the absorbance of water, FTIR spectra of peptides are typically collected in D_2O solution, in which case the relevant band is referred to as amide I'. The primary feature in the amide I' region for peptide 1 was a broad band centered around 1645 cm^{-1}, characteristic of random coil configuration. Figure 1 shows a representative spectrum of peptide 1 at 25˚C. In addition to the dominant 1645 cm^{-1} band, spectra of peptide 1 contained an unidentified shoulder at 1606 cm^{-1} and a minor feature at 1585 cm^{-1}, which we assign to a superposition of protonated Arg (literature value: 1581 to 1586 cm^{-1}) and deprotonated Asp (literature value: 1584 cm^{-1}) [22]. Between 5˚C and 75˚C, slight shifting of the 1645 cm^{-1} band to higher frequencies was observed, but no new features emerged, and no hysteresis was observed during thermal cycling. Peptides 2 and 3 differ from peptide 1 in the substitution of a terminal Cys with Ala (C1A in peptide 2 and C21A in peptide 3) and are therefore unable to form intramolecular disulfide bonds. Their IR spectra are essentially identical to that of peptide 1, as shown in figure 1. Peptide 4 differs from peptide 1 in the substitution of Pro at position 8 with Ser (P8S). Spectra of peptide 4 show a smaller 1645 cm^{-1} feature relative to peptide 1, along with two new features consistent with β-sheet structure: a dominant narrow band at 1620 cm^{-1} and a minor feature at 1684 cm^{-1}. Samples of peptide 4 were opaque and viscous, consistent with aggregation of β-sheets. Between 5˚C and 75˚C, the 1620 cm^{-1} and 1684 cm^{-1} bands increase at the expense of the 1645 cm^{-1} feature as shown in figure 2a. Figure 2b shows that this increase in β-sheet character persists upon returning to 25˚C. Addition of reducing agent dithiothreitol at basic pH clears the solution;

spectra of peptide 4 collected under these reducing conditions show that β-sheet character is eliminated (Figure 3). Peptide 5 is also P8S, but is unable to form intramolecular disulfide bonds because of a Cys substitution with Ala (C21A). As shown in figure 1, this peptide exhibits only random coil secondary structure. Peptide 6 is identical to peptide 1 except that its ends have been capped to eliminate free charges. This peptide was examined after intramolecular disulfide bond formation and found to be identical to peptide 1 (figure 1). Finally, peptide 7 is C21A with capped ends. Its FTIR spectrum is also dominated by the 1645 cm^{-1} feature associated with random coil, as seen in figure 1.

Discussion

To identify our synthetic targets we examined the 63 instances of the highly conserved cysteine-flanked peptide epitope reported by O'Brien et al. [8]. Table 1 shows the percent occurrence of amino acids at each position. Except for a single instance where serine appears at position 1, the cysteines at positions 1 and 21 are completely conserved, suggesting that under the oxidizing conditions of serum, the 21-mer will form a closed loop, as the thiol groups of the cysteine side chains undergo disulfide bonding. This C-loop structure may push the amino acids comprising the loop away from the protein core, making them more accessible to interaction with antibodies or other binding partners. Among the 63 repeats, some residues are completely or very highly conserved (>96%) including arginine at positions 2 and 7; leucine at positions 3 and 6; lysine at position 10; alanine at position 14; and aspartic acid at position 18. Others are more variable, such as position 8, where proline occurs in 60% of the repeats, and position 11, where aspartic acid occurs in 54%. Our primary synthetic target was the consensus sequence (peptide 1) comprised of the most frequently occurring amino acids. We also synthesized several sequence variants that we predicted might have differing structure (peptides 2 through 7). We were particularly interested in characterizing the effects of disulfide bonding, on the assumption that the C-loop was significant to secondary structure. We were also interested in the incomplete conservation of proline, given the potential significance of that amino acid to secondary structure. Proline is unique in having its side chain bonded to the backbone nitrogen, and the resulting lack of amide hydrogen impairs hydrogen bonding at that site. In addition, the 5-membered ring significantly constrains rotation around the peptide backbone, which limits the range of secondary structures that may be achieved by proline-containing peptides.

FTIR spectroscopy, which is sensitive to protein and peptide secondary structure, reveals that 6 of the 7 peptides we synthesized are largely unstructured, as indicated by the broad 1645 cm^{-1} band characteristic of random coil configuration. It is not surprising that a 21-mer peptide is unable to achieve a well-defined secondary structure. Upon substitution of serine for proline at position 8 (peptide 4), however, the random coil signature is replaced during thermal cycling by two narrow peaks (at 1620 cm^{-1} and 1684 cm^{-1}) associated with β-sheet structure. Concomitant with this alteration in the secondary structure, we also observed increased viscosity in solutions of peptide 4, consistent with aggregation. The β-sheet structure increases upon heating and exhibits persistence after thermal cycling. This observation suggests that peptide 4 as synthesized initially exists in a variety of conformations, some of which occupy local energetic minima and are kinetically (though not thermodynamically) stable. Heating enables the semi-stable conformations to overcome the energetic barrier to attain the global energetic minimum of the thermodynamically stable state. We therefore conclude that for oxidized peptide 4, β-sheet

structure is thermodynamically favored. Addition of a reducing agent to solutions of peptide 4 caused them to clarify, and a return to random coil configuration is indicated by the FTIR spectra. Peptide 5, which is also P8S but unable to form intramolecular disulfide bonds, showed no evidence of β-sheet formation. Thus two conditions that favor β-sheet formation and aggregation are the presence of a disulfide bond (in our case an intramolecular bond) and the P8S sequence substitution.

Aggregation of mucins is vital to their role in forming a hydrated, gel-like protective coating on the epithelial cells that secrete them. For CA125/MUC16, our observation of sequence- and oxidation-state dependent aggregation suggests the intriguing possibility of a role for some, but not all, repeated peptide epitopes in intermolecular aggregation. Differences in structure may also result in differing affinities for CA125 antibodies; assays to investigate this question are ongoing.

A structural model has been proposed for an SEA (sea urchin sperm protein, enterokinase, and agrin) domain from a murine homogue of CA125 [23]. In their NMR solution structure, Maeda et al. located two cysteine residues (Cys-57 and Cys-78 in their numbering scheme) near each other on two β strands with an intervening unstructured loop. They found no evidence for disulfide bonding between Cys-57 and Cys-78 although those amino acids are seen in the NMR structure to be close enough for such bonding to occur. For several reasons, we suggest that this structural model is of limited utility as a model for human CA125. Importantly, the primary sequence between Cys-57 and Cys-78 for the murine protein they characterized differs significantly from the human consensus sequence that we identify from the sequence information reported by O'Brien et al. [8]. The 22-amino acid sequence CQVLAFRSVSNNNNHTGVDSLC found in the mouse protein characterized by Maeda differs significantly from our 21-amino acid-long consensus sequence: CRLTLLRPEKDGAATGVDAIC. At position 8 in our numbering scheme (64 in Maeda's notation), the murine protein contains a Ser, which is found in approximately 25% of repeated sequences in human CA125 and, most significantly, was uniquely responsible for the alteration of secondary structure from largely random coil to aggregated β-sheet structure in disulfide-bonded peptides. In addition, NMR spectroscopy was performed on peptides in a solution containing DTT, a reducing agent that we show alters the structure of our model peptides, destroying β-sheet structure. The oxidizing environment of serum is not well modeled by a DTT-rich solution. Further structural characterization of true human CA125 and its repeat domains is needed, and these efforts must pay particular attention to the identity of the residue at position 8 and to the peptide oxidation state.

Gubbels and co-workers demonstrated N-glycan dependent, high-affinity binding between mesothelin (a cell-surface protein normally present in the peritoneal cavity) and CA125 expressed on the ovarian tumor cell line OVCAR-3 [24]. Such interactions likely serve as a mechanism for peritoneal metastasis of ovarian tumors. The binding site on mesothelin for CA125 has been identified, and incubation with a single-chain monoclonal antibody has been found to block the mesothelin-CA125 interaction [25]. The synthetic, glycan-free peptide epitopes that we report here may find use in the continuing characterization of metastasis by decoupling the role of peptide primary and secondary sequence from the role of glycans present on cell-expressed or ascites-derived CA125. These peptides may also serve as therapeutic agents by interfering with cell-cell adhesion or be used in the development of a targeted drug-delivery strategy. Targeting antibodies armed with cytotoxic drugs to repeat domains of CA125 has been

demonstrated to be effective in an animal model [26] and may represent a useful mode of therapy in humans. The investigation of such applications is a focus of our current work.

Acknowledgements

The authors thank Manish Mehta, Sean Decatur and Kalani Seu for material and technical assistance. Funding was received from the Camille and Henry Dreyfus Foundation, the Oberlin-Michigan-Kalamazoo (OKUM) faculty exchange grant and Oberlin College.

References

1 American Cancer Society: All about ovarian cancer. 2009; http://www.cancer.org.

2 Bast RC, Feeney M, Lazarus H, Nadler LM, Colvin RB, Knapp RC: Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest 1981; 68: 1331–1337.

3 Bast RC, Klug TL, St. John E, Jenison E, Niloff JM, Lazarus H, Berkowitz RS, Leavitt T, Griffiths CT, Parker L, Zurawski VR, Knapp RC: A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. N Engl J Med 1983; 309: 883–887.

4 Bast RC: Status of tumor markers in ovarian cancer screening. J Clin Oncol 2003; 21: 200s–205s.

5 Gupta D, Lis CG: Role of CA125 in predicting ovarian cancer survival–a review of the epidemiological literature. J Ovarian Res 2009; 2: 13.

6 Bast RC, Badgwell D, Lu Z, Marquez R, Rosen D, Liu J, Baggerly KA, Atkinson EN, Skates S, Zhang Z, Lokshin A, Menon U, Jacobs I, Lu K. New tumor markers: CA125 and beyond. Int J Gynecol Cancer 2005; 15 (suppl 3): 274–281.

7 Zhang Z, Yu Y, Xu F, Berchuck A, van Haaften-Day C, Havrilesky LJ, de Bruijn HWA, van der Zee AGJ, Woolas RP, Jacobs IJ, Skates S, Chan DW, Bast RC: Combining multiple serum tumor markers improves detection of stage I epithelial ovarian cancer. Gynecol Oncol 2007; 107: 526–531.

8 O'Brien TJ, Beard JB, Underwood LJ, Dennis RA, Santin AD, York L: The CA 125 gene: An extracellular superstructure dominated by repeat sequences. Tumor Biol 2001; 22: 348–366.

9 Yin BWT and Lloyd KO: Molecular cloning of the CA125 ovarian cancer antigen: Identification as a new mucin, MUC16. J Biol Chem 2001; 276: 27371–27375.

10 O'Brien TJ, Beard JB, Underwood LJ, Shigemasa K: The CA 125 gene: A newly discovered extension of the glycosylated N-terminal domain doubles the size of this extracellular superstructure. Tumor Biol 2002; 23: 154–169.

11 Yin BWT, Dnistrian A, Lloyd KO: Ovarian cancer antigen CA125 is encoded by the *MUC16* mucin gene. Int J Cancer 2002; 98: 737–740.

12 Hattrup CL, Gendler SJ: Structure and function of the cell surface (tethered) mucins. Annu Rev Physiol 2008; 70: 431–457.

13 Singh AP, Senapati S, Ponnusamy MP, Jain M, Lele SM, Davis JS, Remmenga S, Batra SK: Clinical potential of mucins in diagnosis, prognosis, and therapy of ovarian cancer. Lancet Oncol 2008; 9: 1076–1085.

14 Imai S, Haga S, Kiyozuka Y: Epitope characterization of MUC1 antibodies. Tumor Biol 1998; 19 (suppl 1): 30–34.

15 Schol DJ, Meulenbroek MFA, Snijdewint FGM, von Mensdorff-Pouilly S, Verstraeten RA, Murakami F, Kenemans P, Hilgers J: "Epitope fingerprinting" using overlapping 20-mer peptides of the MUC1 tandem repeat sequence. Tumor Biol 1998; 19 (suppl 1): 35–45.

16 Grinstead JS, Koganty RR, Krantz MJ, Longenecker BM, Campbell AP: Effect of glycosylation on MUC1 humoral immune recognition: NMR studies of MUC1 glycopeptide–antibody interactions. Biochem 2002; 41: 9946–9961.

17 Fontenot JD, Tjandra N, Bu D, Ho C, Montelaro RC, Finn OJ: Biophysical characterization of one-, two-, and three-tandem repeats of human mucin (muc-1) protein core. Cancer Res 1993; 53: 5386–5394.

18 Ferreira CSM, Matthews CS, Missailidis S: DNA aptamers that bind to MUC1 tumour marker: design and characterization of MUC1-binding single-stranded DNA aptamers. Tumor Biol 2006; 27: 289–301.

19 Warren DJ, Nustad K, Beard JB, O'Brien TJ: Expression and epitope characterization of a recombinant CA 125 repeat: Fourth report from the ISOBM TD-1 workshop. Tumor Biol 2009; 30: 51–60.

20 Chan WC and White PD, eds: *Fmoc solid phase peptide synthesis: a practical approach*; Oxford University Press: New York, 2000.

21 Diem, M: *Introduction to Modern Vibrational Spectroscopy*; John Wiley and Sons: New York, 1993

22 Barth A, Zscherp C: What vibrations tell us about proteins. Q Rev Biophys 2002; 35: 369–430.

23 Maeda T, Inoue M, Koshiba S, Yabuki T, Aoki M, Nunokawa E, Seki E, Matsuda T, Motoda Y, Kobayashi A, Hiroyasu F, Shirouzu M, Terada T, Hayami N, Ishizuka Y, Shinya N, Tatsuguchi A, Yoshida M, Hirota H, Matsuo Y, Tani K, Arakawa T, Carninci P, Kawai J, Hayashiaki Y, Kigawa T, Yokoyama S: Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16). J Biol Chem 2004; 279: 13174–13182.

24 Gubbels JAA, Belisle J, Onda M, Rancourt C, Migneault M, Ho M, Bera TK, Connor J, Sathyanarayana BK, Lee B, Pastan I, Patankar MS: Mesothelin-MUC16 binding is a

high-affinity, N-glycan dependent interaction that facilitates peritoneal metastasis of ovarian tumors. Mol Cancer 2006; 5: 50.

25 Kaneko O, Gong L, Zhang J, Hansen JK, Hassan R, Lee, B, Ho M: A binding domain on mesothelein for CA125/MUC16. J Biol Chem 2009; 284: 3739–3749.

26 Chen Y, Clark S, Wong T, Chen Y, Chen Y, Dennis MS, Luis E, Zhong F, Bheddah S, Koeppen H, Gogineni A, Ross S, Polakis P, Mallet W: Armed antibodies targeting the mucin repeats of the ovarian cancer antigen, MUC16, are highly efficacious in animal tumor models. Cancer Res 2007; 67: 4924–4932.

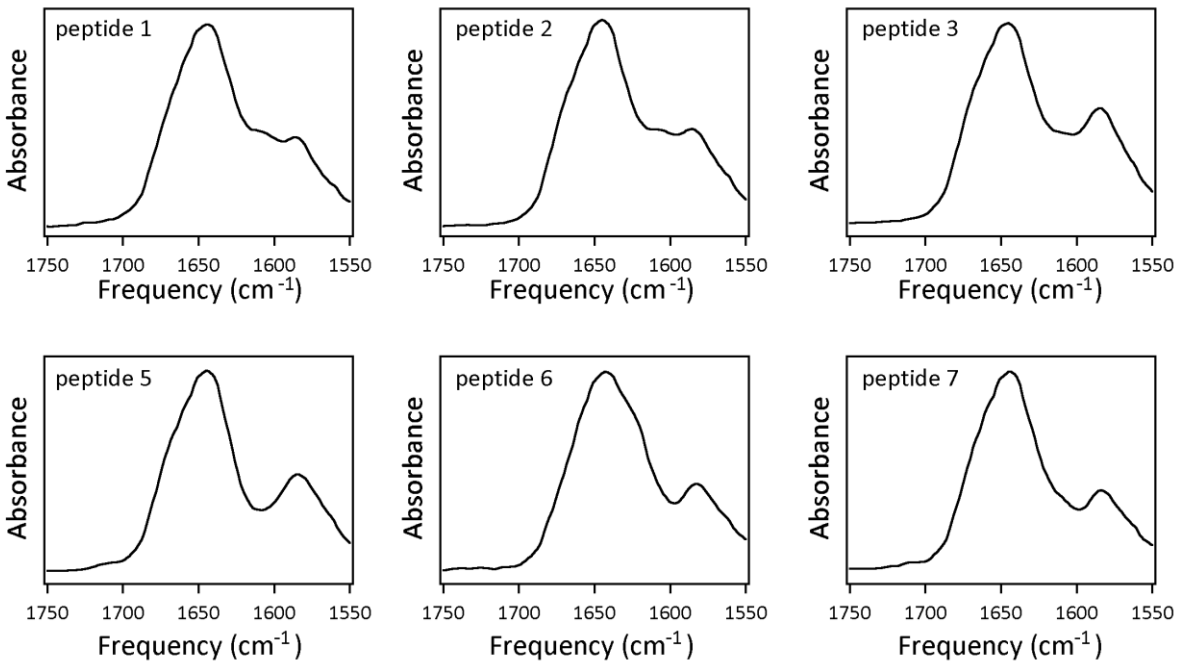

Figure 1. FTIR spectra of 21-residue peptides (labeled according to Table 2) measured at 25˚C. The 1645 cm^{-1} band characteristic of random coil configuration is clearly pronounced.

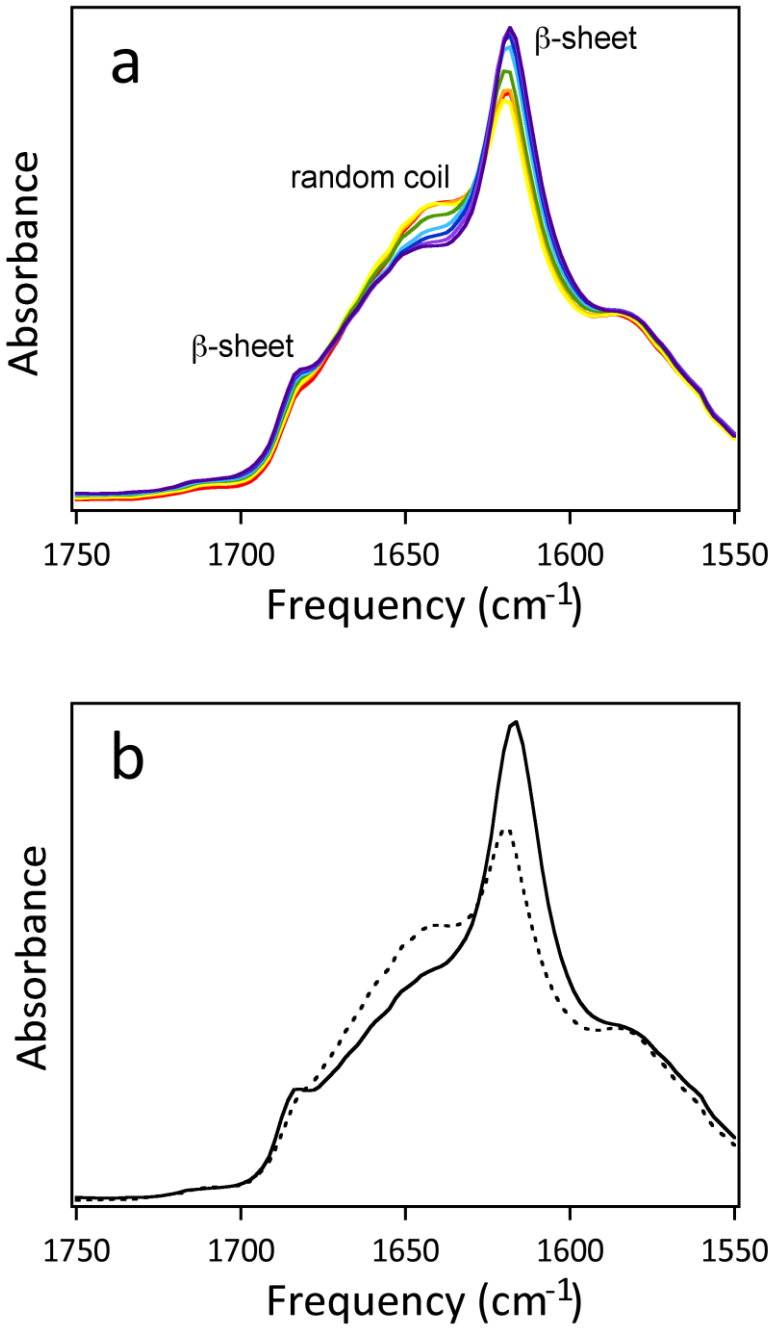

Figure 2. (a) FTIR spectra of peptide 4 recorded at temperatures ranging from 5˚C (red) to 75˚C (violet). The features associated with β-sheet structure increase with temperature at the expense of the random coil peak. (b) FTIR spectra for peptide 4 recorded at 25˚C before (dashed line) and after (solid line) heating of the peptide to 75˚C, showing the persistence of the β-sheet structure.

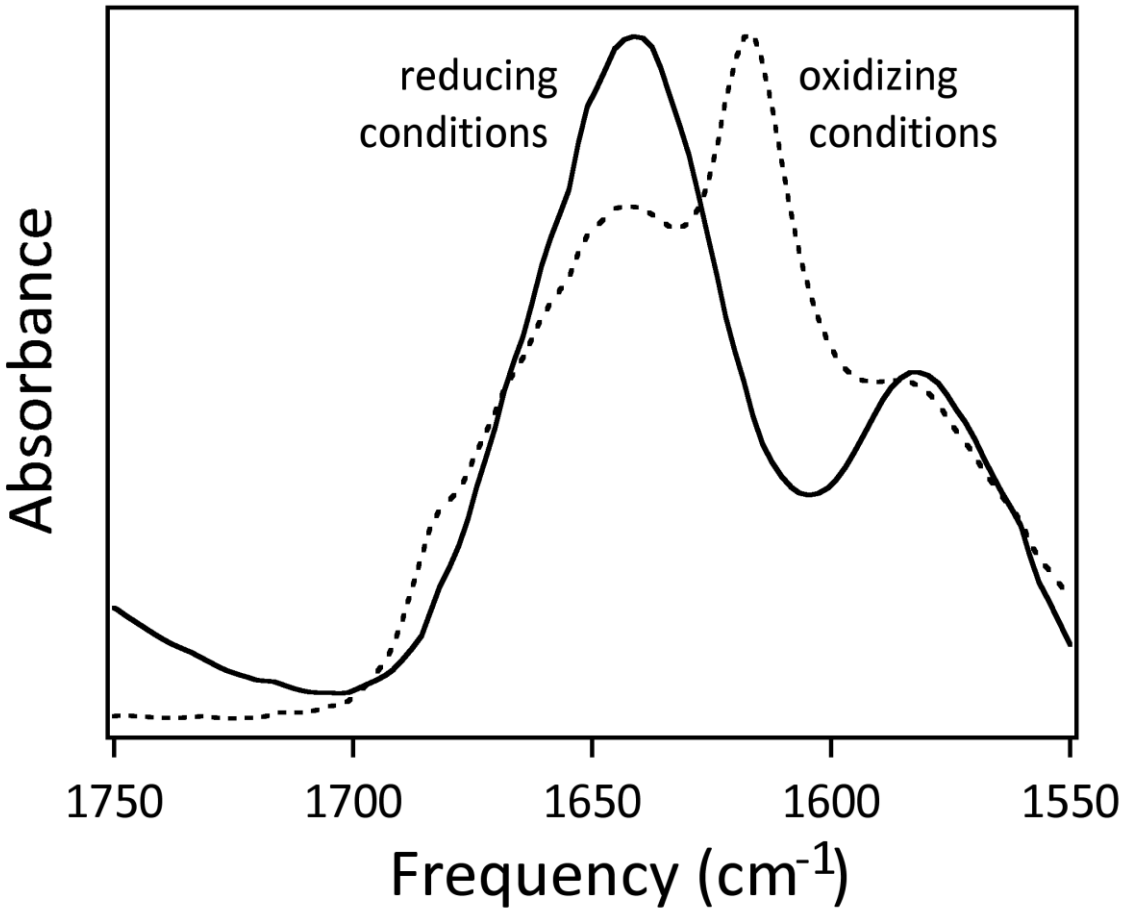

Figure 3. FTIR spectra of peptide 4 recorded at 25˚C in the absence (dashed line) and presence (solid line) of reducing agent DTT, showing the sensitivity of β-sheet structure to oxidation state.

position	amino acid	% occ.	position	amino acid	% occ.	position	amino acid	% occ.
1	C	98.4	11	D	54.0	16	A	4.8
	S	1.6		N	9.5		K	4.8
2	R	98.4		H	9.5		X	15.9
	Q	1.6		Q	6.3	17	V	87.3
3	L	98.4		R	4.78		M	12.7
	V	1.6		X	15.9	18	D	96.8
4	T	88.9	12	G	68.2		A	1.6
	A	4.8		E	9.5		T	1.6
	I	6.3		S	4.8	19	A	63.5
5	L	76.2		K	1.6		T	15.9
	S	22.2		X	15.9		I	1.6
	A	1.6	13	A	90.5		L	1.59
6	L	100		S	4.7		V	1.6
7	R	100		V	1.6		X	15.9
8	P	60.3		T	1.6	20	I	63.5
	S	23.8		E	1.6		V	11.1
	X	15.9	14	A	98.4		T	1.6
9	E	92.0		E	1.6		A	3.8
	K	6.4	15	T	95.2		C	3.8
	V	1.6		M	4.8		L	1.6
10	K	98.4	16	G	61.9		X	15.9
	D	1.6		R	12.7	21	C	100

Table 1. Occurrence frequencies of amino acids at each position in the CA125 antigenic determinant, based on reported data (O'Brien et al., 2001). "X" refers to an unspecified amino acid.

name	sequence
peptide 1	CRLTLLRPEKDGAATGVDAIC
peptide 2	**A**RLTLLRPEKDGAATGVDAIC
peptide 3	CRLTLLRPEKDGAATGVDAI**A**
peptide 4	CRLTLLR**S**EKDGAATGVDAIC
peptide 5	CRLTLLR**S**EKDGAATGVDAI**A**
peptide 6	Ac-CRLTLLRPEKDGAATGVDAIC-NH_2
peptide 7	Ac-CRLTLLRPEKDGAATGVDAI**A**-NH_2

Table 2. One-letter abbreviations of the peptide sequences synthesized in this study. Sequence variants are designated by a bold letter showing the amino acid at which the peptide differs from peptide 1, the consensus sequence. Peptides 1-5 have charged free ends. Peptides 6 and 7 were capped to remove end charges; the C-terminus was amidated and the N-terminus was acetylated.