

## Secondary Structure and Glycosylation of Mucus Glycoproteins by Raman Spectroscopies

Heather S. Davies,<sup>†,⊗</sup> Prabha Singh,<sup>‡</sup> Tanja Deckert-Gaudig,<sup>§</sup> Volker Deckert,<sup>‡,§</sup> Karine Rousseau,<sup>†</sup> Caroline E. Ridley,<sup>†</sup> Sarah E. Dowd,<sup>||</sup> Andrew J. Doig,<sup>||</sup> Paul D. A. Pudney,<sup>⊥</sup> David J. Thornton,<sup>†</sup> and Ewan W. Blanch<sup>\*,†,‡,#</sup>

<sup>†</sup>Wellcome Trust Centre for Cell-Matrix Research, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PL, United Kingdom

<sup>‡</sup>Institute of Physical Chemistry and Abbe Center of Photonics, Friedrich-Schiller University Jena, Helmholtzweg 4, 07743 Jena, Germany

<sup>§</sup>Leibniz Institute of Photonic Technology, Albert-Einstein-Strasse 9, 07745 Jena, Germany

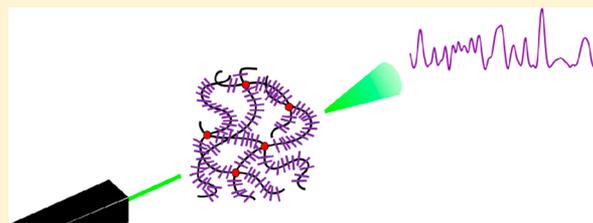
<sup>||</sup>School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN, United Kingdom

<sup>⊥</sup>Unilever Discover, Sharnbrook, Bedfordshire MK44 1LQ, United Kingdom

<sup>#</sup>School of Science, RMIT University, Melbourne, Victoria 3001, Australia

### Supporting Information

**ABSTRACT:** The major structural components of protective mucus hydrogels on mucosal surfaces are the secreted polymeric gel-forming mucins. The very high molecular weight and extensive O-glycosylation of gel-forming mucins, which are key to their viscoelastic properties, create problems when studying mucins using conventional biochemical/structural techniques. Thus, key structural information, such as the secondary structure of the various mucin subdomains, and glycosylation patterns along individual molecules, remains to be elucidated. Here, we utilized Raman spectroscopy, Raman optical activity (ROA), circular dichroism (CD), and tip-enhanced Raman spectroscopy (TERS) to study the structure of the secreted polymeric gel-forming mucin MUC5B. ROA indicated that the protein backbone of MUC5B is dominated by unordered conformation, which was found to originate from the heavily glycosylated central mucin domain by isolation of MUC5B O-glycan-rich regions. In sharp contrast, recombinant proteins of the N-terminal region of MUC5B (D1-D2-D'-D3 domains, NT5B), C-terminal region of MUC5B (D4-B-C-CK domains, CT5B) and the Cys-domain (within the central mucin domain of MUC5B) were found to be dominated by the  $\beta$ -sheet. Using these findings, we employed TERS, which combines the chemical specificity of Raman spectroscopy with the spatial resolution of atomic force microscopy to study the secondary structure along 90 nm of an individual MUC5B molecule. Interestingly, the molecule was found to contain a large amount of  $\alpha$ -helix/unordered structures and many signatures of glycosylation, pointing to a highly O-glycosylated region on the mucin.



Saliva and other mucous secretions are viscoelastic hydrogels that are important protective barriers at epithelial surfaces. The major structural components of these hydrogels are the polymeric gel-forming mucins that are extremely large, densely glycosylated proteins. The gel-forming mucin network protects the body's surfaces from dehydration, injury, and colonization by pathogens<sup>1</sup> and in the gastrointestinal (GI) tract provides a niche for commensal microbes. Furthermore, gel-forming mucins in the GI tract also interact with dietary molecules, as has recently been shown for the salivary mucin MUC5B and green tea polyphenols, which may affect the absorption of nutrients and alter mucin network organization.<sup>2</sup>

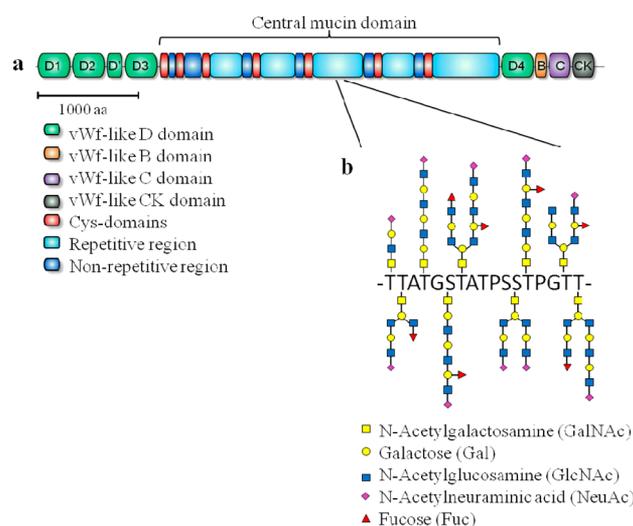
The five gel-forming mucin family members (MUC2, MUC5B, MUC5AC, MUC6, and MUC19) have different tissue localization but similar domain organization (e.g., Figure 1a depicts the domain organization of a MUC5B monomer, the

major mucin in saliva and respiratory mucus). The central mucin domain is repetitive and contains STP-rich regions that are substantially glycosylated with structurally diverse O-linked glycan chains (Figure 1b). This central highly O-glycosylated portion of gel-forming mucins is interrupted with, and flanked by, cysteine-rich, substantially less glycosylated protein domains. At the N- and C-terminal regions, these cysteine-rich domains are essential for mucin polymerization via disulfide bridges into very high weight (MW), polydisperse (2–40 MDa) macromolecules that entangle and cross-link (mainly by noncovalent interactions) to form an organized

Received: August 9, 2016

Accepted: October 28, 2016

Published: October 28, 2016



**Figure 1.** Domain organization of MUC5B. (a) Schematic representation of a MUC5B monomer. The N-terminal region is composed of vWF-like domains (D1, D2, D', D3), the central region is highly repetitive, heavily glycosylated and is interrupted by Cys-rich subdomains (Cys-domains), and the C-terminal region is composed of vWF-like domains (D4, B, C, CK). (b) Mucin glycan diversity. Oligosaccharides that decorate the STP-rich central mucin domain (repetitive and nonrepetitive sequences) are composed of 5 major monosaccharides but differ in composition, branching, length, and modification by sulfation and acetylation, contributing to the heterogeneity of mucins.

network.<sup>3–7</sup> The small cysteine-rich domains (Cys-domains) that interrupt the central O-glycan rich region of mucins may also support mucin cross-linking.<sup>8</sup>

The length (0.5–10  $\mu\text{m}$ ), polydispersity, and extended conformation of polymeric gel-forming mucins has been demonstrated by electron microscopy (EM), atomic force microscopy (AFM), light scattering, small-angle X-ray scattering (SAXS), and neutron scattering (SANS) studies.<sup>7,9–11</sup> These techniques have also demonstrated the extended topography of isolated mucin O-glycan rich regions and globular nature of mucin N- and C-terminal regions.<sup>8,12–15</sup>

Despite these breakthroughs, there are important questions that remain regarding the architecture of gel-forming mucins. For instance, what is the secondary structure of the various mucin subdomains; what are the locations of the various mucin subdomains along a single mucin chain; and do glycosylation patterns exist along individual molecules? Answers to these questions have been hindered by the size, polydispersity, and dense glycosylation of mucins, which make them unsuitable for investigation by conventional structural techniques. Here, we aim to shed light on these finer details of mucin architecture by utilizing a novel toolbox of complementary vibrational spectroscopies.

Raman spectroscopy analyzes characteristic vibrational modes within a molecule that depend on the nature of the functional groups. A derivative of Raman spectroscopy is Raman optical activity (ROA), which is sensitive to molecular chirality and particularly sensitive to the secondary and tertiary structures of biomolecules. Circular dichroism (CD) in the UV region is similarly sensitive to chirality but monitors electronic transitions rather than vibrational ones and is also sensitive to protein secondary structure. Therefore, analysis by Raman, ROA, and CD offers a powerful approach for deciphering the

secondary structure of complex biological molecules, such as full-length mucins and mucin subdomains.<sup>16</sup>

In addition, tip-enhanced Raman spectroscopy (TERS) is a novel technique that combines the chemical specificity of Raman spectroscopy with the spatial resolution of AFM and offers a unique opportunity to examine mucin structure along the length of an individual molecule. In contrast to ROA and CD, TERS exclusively probes the surface of a sample enabling a differentiation of core and surface composition of proteins, as has been demonstrated for amyloid fibrils.<sup>17,18</sup>

Here, we isolate the high MW, polymeric gel-forming mucin MUC5B from human whole saliva and generate MUC5B subdomains of the various “sugar-rich” regions (central O-glycan-rich regions) and “protein-rich” regions (N- and C-terminal regions and Cys-domains, which are substantially less glycosylated). We investigate their structures by Raman, ROA, and CD to reveal previously unknown information about the secondary structure of the various mucin subdomains. Furthermore, we measure the first TER spectra with spatial resolution along the length of a single MUC5B molecule and identify signatures of mucin secondary structure and glycosylation. Such spatially resolved structural information has not previously been reported for even simple glycoproteins, let alone the highly complex mucins. Our application of Raman, ROA, CD, and TERS spectroscopies to polymeric mucins has provided a new level of structural understanding of these complex glycoproteins.

## MATERIALS AND METHODS

**Materials.** Agarose, cesium chloride (CsCl), and dithiothreitol (DTT) were purchased from Melford Laboratories (Ipswich, U.K.). Monosaccharides and amino acids were purchased from Sigma-Aldrich (Poole, U.K.).

**Mucin Purification.** For MUC5B isolation, healthy volunteers with no overt sign of oral pathologies and who had provided their written consent were asked to donate saliva into a sterile container at least 1 h after consumption of food or drink. Ethical approval for this research was obtained from the University of Manchester. Saliva samples were immediately pooled and solubilized in CsCl/0.10 M NaCl at a starting density of 1.4 g/mL, in the presence of a protease inhibitor cocktail (Supporting Information) at 4 °C overnight. Under these conditions, MUC5B was purified by isopycnic density gradient centrifugation, followed by density gradient centrifugation in CsCl/0.10 M NaCl at a starting density of 1.5 g/mL, in a Beckman L-90 ultracentrifuge (Beckman Ti45 rotor, 72 h, 40 000 rpm, 15 °C, Figure S1).<sup>19</sup>

**Mucin Subdomain Preparation.** The O-glycan-rich regions of MUC5B were generated from purified, polymeric MUC5B by reduction and alkylation and subsequent trypsin digestion, followed by removal of tryptic peptides, as described previously.<sup>2</sup> The average MW of glycan-rich regions was found to be 546 kDa by size-exclusion chromatography coupled with multi-angle laser light scattering.<sup>20</sup> An N-terminal construct of MUC5B, consisting of D1-D2-D'-D3 domains (NT5B, residues 26–1304) and a C-terminal construct of MUC5B, consisting of D4-B-C-CK domains (CT5B, residues 4958–5765), were created, expressed, and purified.<sup>2</sup> A construct of the seventh Cys-domain of MUC5B (residues 4128–4235) was created, expressed, and purified in the same way as NT5B and CT5B, with the following exception: the size-exclusion chromatography step was performed on a Superdex 75 column 10/300GL (GE Healthcare).

**Sample Preparation for Raman and ROA.** Mono-saccharides were prepared to 40 mg/mL in 0.10 M NaCl, pH 6.0. MUC5B and MUC5B subdomains were dialyzed into 0.10 M NaCl, pH 6.0 and concentrated to 5–15 mg/mL in Sartorius Vivaspin 5–10 kDa molecular weight cut-off (MWCO) columns.

**Raman and ROA Data Acquisition.** Raman and ROA spectra were measured using a BioTools ChiralRaman spectrometer. The instrument was set up in a backscattering geometry and operated with a Nd:VO<sub>4</sub> laser with an excitation wavelength of 532 nm and spectral resolution of 7 cm<sup>-1</sup>. Samples were measured in quartz microfluorescence cells. Mucin samples were measured with a laser power of 600 mW at the sample, laser illumination period of 1.24 s and data were gathered over a 36–48 h time period, whereas monosaccharide and amino acid spectra were acquired for 4–12 h. Raman spectral baselines were corrected based on an approach reported elsewhere,<sup>21</sup> averaged and smoothed with a 15 point Savitzky-Golay filter. Sample and reference Raman spectra were then normalized to an invariant band (270 cm<sup>-1</sup>), and then the reference Raman spectrum of 0.10 M NaCl was subtracted from the sample Raman spectrum. ROA spectra were averaged and smoothed with a 15 point Savitzky-Golay filter.

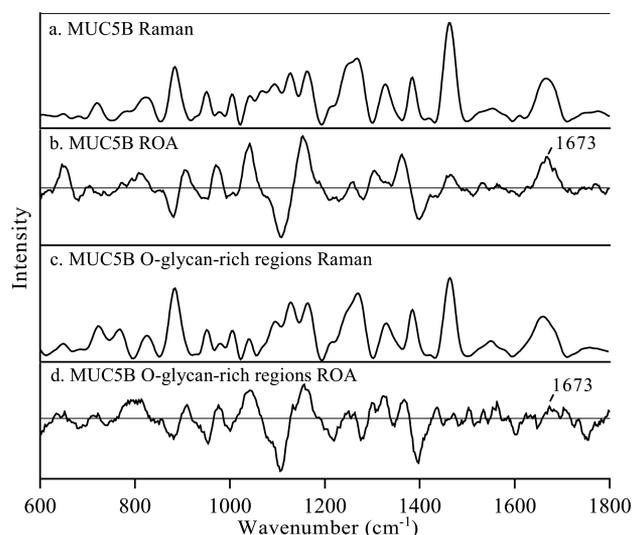
**CD Data Acquisition.** MUC5B subdomains (NT5B, CT5B, and Cys-domain) were prepared to 1 mg/mL for CD. CD measurements were made using an Applied Photophysics Chirascan qCD spectropolarimeter, step size 0.5 nm, 5 s per time point, and 4 repeats. Measurements were made in a 0.5 mm quartz cell at 12 °C in 10 mM Tris-HCl, 50 mM NaCl, pH 7.4.

**TERS Sample Preparation.** Glass coverslips were washed in 100% ethanol overnight and air-dried for 2 h before addition of the sample. Purified MUC5B at 10 μg/mL was extensively dialyzed against PBS, pH 7.4, and was prepared on a glass substrate by drop-deposition of 100 μL onto pre-cleaned glass for 30 s, followed by washing with 6 × 300 μL of ddH<sub>2</sub>O and air-dried for at least 24 h before being imaged. Mono-saccharides for TERS analysis were prepared to 1 mM in ddH<sub>2</sub>O, drop-deposited onto pre-cleaned glass substrates, and dried under argon, based on established protocols.<sup>22</sup>

**TERS Data Acquisition.** TER spectra and corresponding AFM images were collected with a Nanowizard II atomic force microscope (JPK Instrument AG, Germany) mounted on an inverted microscope (Olympus IX70, Japan) with a confocal Raman spectrometer (Acton Advanced SP2750 A, SI GmbH, Germany) and a 400 pixel charge-coupled device (CCD) camera. Tap190Al-G (Budget Sensors) AFM-tips were evaporated with 25 nm silver, as described previously.<sup>23</sup> An oil immersion objective (40×, 1.35 NA, Olympus) was used and a 532 nm laser as the excitation source. Each TER spectrum was acquired for 5 s. TER spectra were truncated before 560 cm<sup>-1</sup> to remove the silicon peak from the AFM-tip. Spectral baselines were corrected and cosmic ray strikes were removed using an in-house Raman toolbox in Matlab software, developed by Dr. Ben Gardner.<sup>24</sup> Data sets of mucin TERS spectra were normalized to the silicon band at ~940 cm<sup>-1</sup> and then normalized between 0 and 1. Color-coded heatmaps were generated using the “msheatmap” function in Matlab.

## RESULTS AND DISCUSSION

The Raman and ROA spectra of polymeric, high MW MUC5B (purified from human whole saliva; Figure S1) are shown in Figure 2a,b. The vast majority of peaks in the Raman and ROA



**Figure 2.** Raman and ROA spectra of MUC5B and MUC5B subdomains. Polymeric MUC5B (a,b) and O-glycan-rich regions (c,d) were purified and concentrated to ~15 mg/mL in 0.10 M NaCl, pH 6.0 and their Raman (a,c) and ROA (b,d) spectra measured.

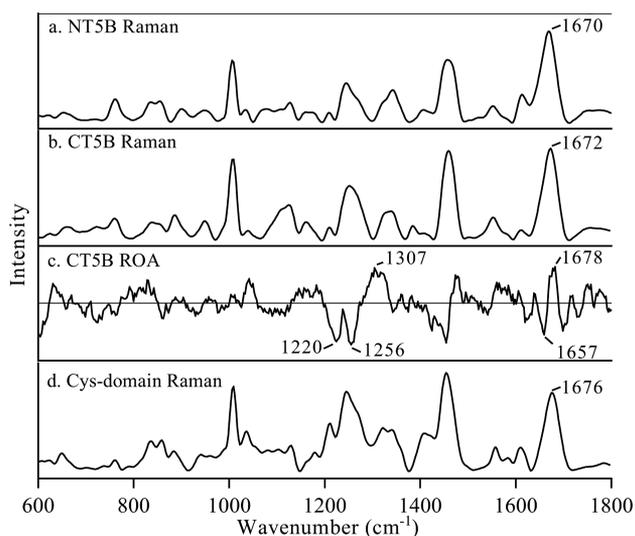
spectra of polymeric MUC5B could be assigned to the major monosaccharide (Fuc, Gal, GalNAc, GlcNAc, NeuAc) and amino acid (Pro, Ser, Thr, Cys) building blocks of the central mucin domain (Figures S2 and S3, Tables S1 and S2), demonstrating that the spectra are dominated by structures within the “sugar-rich” central region of the mucin. This is further demonstrated by the Raman and ROA spectra of the MUC5B O-glycan-rich regions (generated from purified MUC5B by reduction/alkylation and trypsin digestion; Figure 2c,d), which strongly resemble the Raman and ROA spectra of polymeric MUC5B.

The Raman and ROA spectra of polymeric MUC5B measured here share similarities with those of bovine submaxillary mucin<sup>25</sup> and porcine gastric mucin<sup>26</sup> previously reported, reflecting the common building blocks of the various mucin types. Despite this, differences could also be observed. For example, it is noteworthy that the negative peak at ~1109 cm<sup>-1</sup> is intense in the ROA spectra of polymeric MUC5B and bovine submaxillary mucin (putative bovine Muc5b) and is absent in the ROA spectrum of the porcine gastric mucin (putative Muc5ac and Muc6). This could reflect differences in the level of NeuAc, GlcNAc, and Fuc (found to contribute to the band at ~1109 cm<sup>-1</sup>, Table S3), which indeed is in agreement with compositional analysis of similar mucin preparations.<sup>27–29</sup> On the other hand, the negative peak at ~1109 cm<sup>-1</sup> is reported to be a signature of disaccharides,<sup>30,31</sup> and thus may indicate an increased length of attached O-glycans in MUC5B/Muc5b compared to commercial porcine gastric mucin. Perhaps most likely is that such differences arise due to interplay between different monosaccharide composition and length of the O-glycans on the various mucin types.

In order to study mucin secondary structure, the amide I region (1640–1700 cm<sup>-1</sup>) of Raman and ROA spectra was examined. This region contains a major contribution from the C=O stretching mode from the polypeptide backbone, with peak position informative about secondary structure. The amide I band in the Raman spectrum of polymeric MUC5B is broad, suggesting flexible conformation. Furthermore, within the amide I region of the ROA spectrum of polymeric MUC5B

there is a weak positive feature at  $\sim 1673\text{ cm}^{-1}$ , which has been observed in the ROA spectra of proteins with unordered conformation.<sup>31–38</sup> This feature can also be observed in the previously reported ROA spectrum of bovine submaxillary mucin,<sup>25</sup> demonstrating shared secondary structure of mucins between species. The unordered conformation of mucins is in agreement with mucins historically being considered as having a random coil conformation attributable to the central mucin domain that is heavily glycosylated with bulky, negatively charged sugars, which prevent any order of the protein backbone. Indeed, the ROA spectrum of MUC5B O-glycan-rich regions measured here also displays the weak positive feature at  $\sim 1673\text{ cm}^{-1}$  and has no features assigned to  $\alpha$  helix or  $\beta$ -structure, demonstrating a largely unordered protein backbone (Figure 2d).

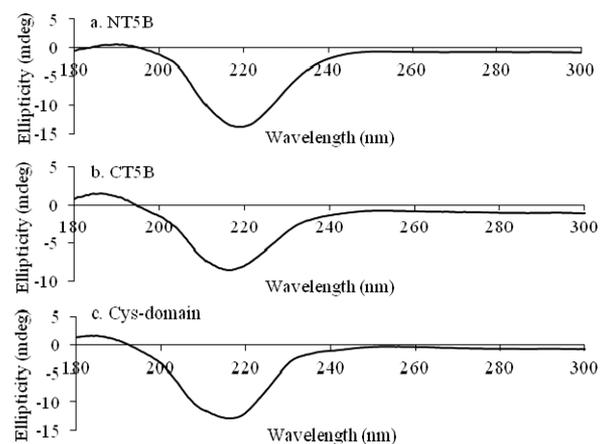
While the heavily glycosylated mucin domain dominates mucin molecular weight and length, the substantially less glycosylated domains that flank the mucin domain, which are crucial for mucin polymerization, contribute far less to mucin molecular weight, making it difficult to detect their structural signatures by examination of full-length mucins alone. Therefore, recombinant proteins of the N-terminal region (NT5B, domains D1-D2-D'-D3), C-terminal region (CT5B, domains D4-B-C-CK), and a Cys-domain of MUC5B were generated and their Raman spectra analyzed (Figure 3). The sharp peak in



**Figure 3.** Raman and ROA spectra of MUC5B subdomains. Recombinant proteins of (a) the N- and (b) the C-terminal regions of MUC5B (NT5B and CT5B, respectively), and (d) a Cys-domain of MUC5B were expressed and purified, and their Raman spectra collected at  $\sim 5\text{ mg/mL}$ . CT5B was purified and concentrated to  $\sim 15\text{ mg/mL}$  and its ROA spectrum measured (c).

the amide I region of the Raman spectra of NT5B, CT5B, and Cys-domain is centered at  $\sim 1670$ ,  $1672$ , and  $1676\text{ cm}^{-1}$ , respectively. Such peaks have been reported in the Raman spectra of proteins that are rich in  $\beta$ -structure,<sup>39–44</sup> suggesting that these MUC5B subdomains contain  $\beta$ -structure. Measurement of the ROA spectrum of CT5B provided further evidence of  $\beta$ -structure (Figure 3c), including the negative band at  $\sim 1220\text{ cm}^{-1}$ ,<sup>31</sup> the couplet with a negative band at  $\sim 1256\text{ cm}^{-1}$  and a positive band at  $\sim 1304\text{ cm}^{-1}$ ,<sup>33,45</sup> and the couplet with a negative band at  $\sim 1657\text{ cm}^{-1}$  and a positive band at  $\sim 1678\text{ cm}^{-1}$ .<sup>33</sup> Furthermore, the CD spectra of NT5B, CT5B, and Cys-domain display minima at  $\sim 216\text{ nm}$ , which is characteristic

of the  $\beta$ -sheet (Figure 4). Taken together, the Raman, ROA, and CD data presented here indicate that the  $\beta$ -sheet is a major feature in the “protein-rich” subdomains of MUC5B. This well-ordered structure of the N- and C-terminal regions of MUC5B is in agreement with EM and SAXS analysis of these regions, which have identified their globular shape.<sup>8,12–15</sup> The data presented here may indicate that the  $\beta$ -sheet is important for the functions of these globular regions of mucins, such as in forming intramolecular and intermolecular interactions that are paramount for mucin cross-linking and polymerization and thus the integrity of saliva and mucus gels.

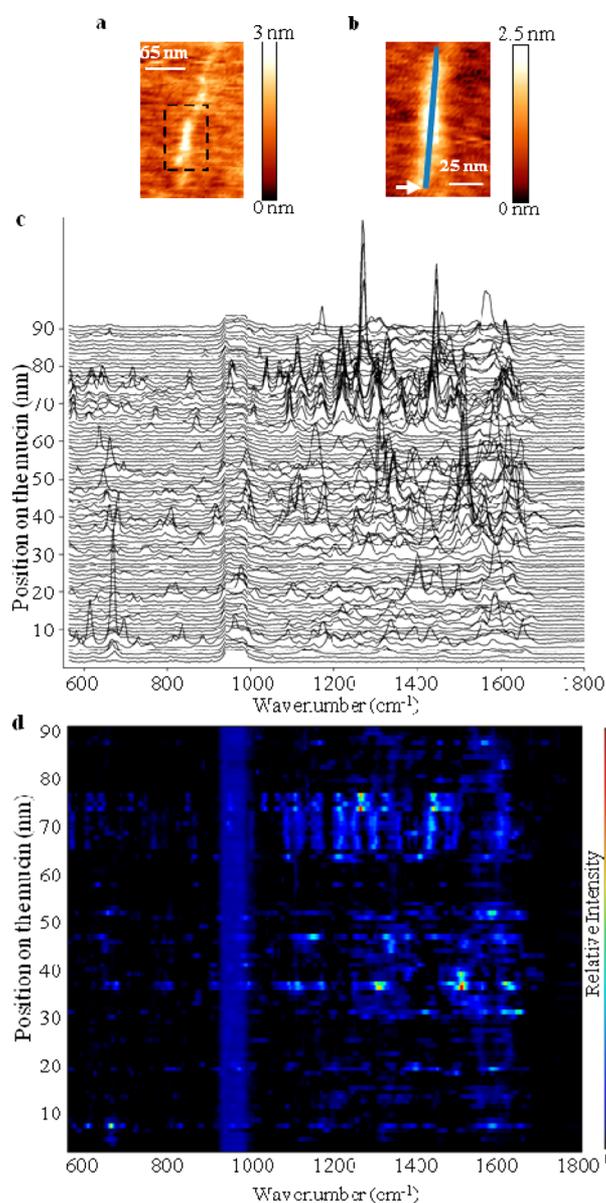


**Figure 4.** CD spectra of MUC5B subdomains. CD measurements of (a) NT5B, (b) CT5B, and (c) Cys-domain at  $12\text{ }^\circ\text{C}$  in  $10\text{ mM}$  Tris-HCl,  $50\text{ mM}$  NaCl,  $\text{pH } 7.4$ .

Following the Raman/ROA/CD analyses of the various mucin subdomains, polymeric MUC5B was analyzed by a novel spectroscopic technique, TERS. TERS combines the spatial resolution of AFM with the chemical specificity of Raman spectroscopy, enabling structure features to be examined along the length of individual molecules. Therefore, we aimed to generate the first TERS spectra along the length of an individual mucin molecule and identify signatures of glycosylation and secondary structure.

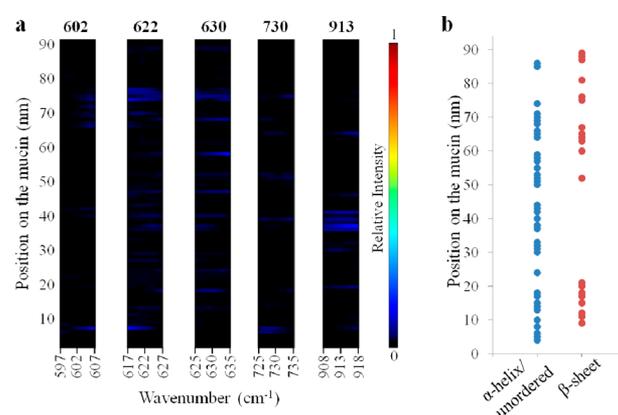
The AFM topography of an individual MUC5B molecule was visualized (Figure 5a,b), and TER spectra were measured every  $1\text{ nm}$  along a  $90\text{ nm}$  segment of the molecule (Figure 5c, and color-map Figure 5d). The spectra have excellent signal-to-noise characteristics and the complexity is highlighted by the vast array of peak positions and intensities. This likely reflects both the sensitivity and specificity of TERS enabling the differentiation of amino acids within a protein at nanometer resolution<sup>17,18</sup> and the structural complexity/diversity of the mucin molecule.

The presence of glycan chains on the mucin renders the interpretation of the data set not a trivial task, since the study of even simple carbohydrates by TERS is in its infancy. Indeed, this is the first report of TER spectra of such a complex glycoprotein with high spatial resolution. By collection of the TER spectra of the major monosaccharide building blocks of mucins (Figure S4), we generated a library of monosaccharide marker bands (Tables S3 and S4) in order to examine signatures of glycosylation along the length of the molecule. All sugar marker bands were identified within the MUC5B TERS data set and show distinct distribution patterns along the length of the molecule (Figure 6a). This may reflect the glycosylation



**Figure 5.** TER spectra along the length of a MUC5B molecule. (a and b) AFM topography of a single MUC5B molecule; the region identified by the dotted line is shown in greater detail in part b. (b) AFM topography of a MUC5B molecule, with the area mapped by TERS shown by the blue line, starting at the arrow. (c) MUC5B TER spectra, measured every 1 nm along 90 nm of the single MUC5B molecule. Each spectrum was truncated below  $560\text{ cm}^{-1}$ , and the background was electronically subtracted and normalized to between 0 and 1. The broad region common to all spectra between  $950$  and  $1000\text{ cm}^{-1}$  is due to the overtone band of the silicon of the TERS-tip. (d) Color-map of MUC5B TER spectra, with peak intensities color-coded (as shown by the scale bar).

of the molecule (i.e., the presence of particular monosaccharides, Table S4), or different orientation of sugars on the glass substrate, which affects their detection by TERS. At this stage it is not possible to unequivocally distinguish between these possibilities; this would be a fruitful area for longer term study. Despite this, it is striking that monosaccharide marker bands are displayed throughout the MUC5B TERS data set, suggesting that the area imaged includes at least part of a highly glycosylated mucin domain.



**Figure 6.** Glycosylation and secondary structure signatures along a MUC5B molecule. (a) Color-maps of monosaccharide marker bands in MUC5B TERS spectra. The appearance and intensity of the marker bands of monosaccharides at  $602$ ,  $622$ ,  $630$ ,  $730$ , and  $913\text{ cm}^{-1}$  in the MUC5B TERS data set are shown. A band position allowance of  $\pm 5\text{ cm}^{-1}$  was used. (b) Plot of detected bands assigned to different secondary structures along the length of the MUC5B TERS data set. Amide I peaks in each spectrum were assigned as  $\alpha$ -helix/unordered (blue,  $1640$ – $1664\text{ cm}^{-1}$ ) or  $\beta$ -sheet (red,  $1665$ – $1678\text{ cm}^{-1}$ ). For the plot band intensities were not considered.

In recent years TERS has proven to be an effective tool for investigating the secondary structure distribution along the lengths of protein structures. Therefore, we examined the amide I band positions along the TERS data set of MUC5B. The position of the amide I band was assigned as  $\alpha$ -helix/unordered ( $1640$ – $1664\text{ cm}^{-1}$ ) or  $\beta$ -sheet ( $1665$ – $1678\text{ cm}^{-1}$ ), based on previous TERS analysis of proteins,<sup>17</sup> providing a map of secondary structure signatures along the MUC5B molecule (Figure 6b). The plot in Figure 6b shows that  $\alpha$ -helix/unordered conformation was identified in the majority of spectra all along the area imaged. Since our Raman, ROA, and CD analyses showed that unordered structure is present in MUC5B O-glycan-rich regions and there is no evidence of  $\alpha$ -helical structure in any of the MUC5B subdomains, this suggests that these TERS signals originate from unordered conformation rather than  $\alpha$ -helix. This would indicate that the area imaged by TERS includes part of an extended heavily O-glycosylated region of the mucin, in strong agreement with the observation of sugar bands along the majority of the molecule (Figure 6a). Indeed, such regions are reported to be  $\sim 100$ – $150\text{ nm}$  in length.<sup>9,46,47</sup> Additionally, among the  $\alpha$ -helix/unordered secondary structure are discrete regions ( $10$ – $20\text{ nm}$  and  $60$ – $90\text{ nm}$ ) that contain  $\beta$ -sheet structure. This may represent a movement of the TERS tip over a protein-rich subdomain (NT5B, CT5B, or Cys-domain) of MUC5B, which were found by Raman, ROA, and CD here to possess largely a  $\beta$ -sheet conformation.

Our discovery of the secondary structures of the various mucin subdomains by Raman/ROA/CD has provided a powerful opportunity to investigate mucin secondary structure along the length of the individual molecule analyzed here by TERS. However, it must be considered that there are differences between these techniques and TERS; far field Raman/ROA/CD detects the core structures of biomolecules, while TERS is extremely sensitive to structures present on the surface of molecules. Therefore, although it is likely that the secondary structure signatures identified in the TERS data set represent the locations of the mucin subdomains (by

correlation with Raman/ROA/CD analysis), it cannot be ruled out that they may have another local origin, thus an absolute statement regarding the conformation on the surface from the presented experiments alone cannot be made at this time.

Here, we present the first demonstration of TER spectra acquisition with high spatial resolution along a gel-forming mucin where different secondary structures and glycosylation signatures can be discerned. Undoubtedly, more in-depth analysis of these data and additional data sets may offer more information regarding mucin structure, such as glycosylation patterns and differentiation of amino acids. Presently, for the purpose of this manuscript, we have focused on the identification of signatures of secondary structure and glycosylation.

Going forward, studies that build upon this work will likely help to forward the capabilities of Raman spectroscopies. For instance, elaborate multivariate data analysis, that are beyond the scope of this manuscript, have recently been demonstrated for the distinction of a glycosylated protein from the nonglycosylated form.<sup>48</sup> Raman spectroscopy and ROA of carbohydrates are also being advanced by *ab initio* approaches to model disaccharides<sup>49–51</sup> and potentially will lead to the ability to model the Raman and ROA spectra of more complex glycoconjugates. Such computational modeling can also certainly be transmitted to the analysis of the TER spectra of glycoproteins.

## CONCLUSIONS

Here, we report the first measurements of the Raman, ROA, and TER spectra of a purified, polymeric, gel-forming mucin, MUC5B. The combined use of Raman, ROA, and CD revealed that polymeric MUC5B has a largely unordered polypeptide backbone arising from its heavily O-glycosylated regions, while the substantially less glycosylated subdomains of MUC5B at the N- and C-terminal regions and central Cys-domains are dominated by  $\beta$ -sheet. This was extended to the single molecule level, and we reported the first TER spectra along the length of a single MUC5B molecule and provided a method to tentatively identify the regions of secondary structure and areas that are glycosylated. This work highlights the potential of combined Raman spectroscopies as a strategy for mapping the secondary structure of complex glycoproteins.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b03095.

Protease inhibitor cocktail methodology; TERS library of monosaccharide marker bands methodology; purification of MUC5B; Raman and ROA spectra of mucin monosaccharides and amino acids; band assignments of the MUC5B Raman and ROA spectrum; TER spectra of monosaccharides; and TERS monosaccharide marker bands (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: ewan.blanch@rmit.edu.au.

### Present Address

⊗H.S.D.: University Grenoble Alpes, Laboratory of Interdisciplinary Physics, Grenoble, France.

## Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank BBSRC and Unilever for funding this work and a travel award to HSD from the Helen Muir Fund, University of Manchester. We acknowledge the Wellcome Trust Centre for Cell-Matrix Research, University of Manchester, which is supported by core funding from the Wellcome Trust [Grant Number 088785/Z/09/Z]. P.S. acknowledges support from the Deutsche Forschungs Gemeinschaft (Grant FR 1348/19-1). Thanks also go to Dr. Ben Gardner for helpful discussions and guidance with data analysis and Nixon's Farm shop abattoir (Manchester, U.K.) for their donation of porcine tissue.

## REFERENCES

- (1) Veerman, E. C. I.; Ligtenberg, A. J. M.; Schenkels, L. C. P. M.; Walgreen-Weterings, E.; Nieuw Amerongen, A. V. *J. Dent. Res.* **1995**, *74*, 351–357.
- (2) Davies, H. S.; Pudney, P. D.; Georgiades, P.; Waigh, T. A.; Hodson, N. W.; Ridley, C. E.; Blanch, E. W.; Thornton, D. J. *PLoS One* **2014**, *9*, e108372.
- (3) Bansil, R.; Turner, B. S. *Curr. Opin. Colloid Interface Sci.* **2006**, *11*, 164–170.
- (4) Raynal, B. D. E.; Hardingham, T. E.; Sheehan, J. K.; Thornton, D. J. *J. Biol. Chem.* **2003**, *278*, 28703–28710.
- (5) Sellers, L. A.; Allen, A.; Morris, E. R.; Ross-Murphy, S. B. *Carbohydr. Res.* **1988**, *178*, 93–110.
- (6) Perez-Vilar, J.; Hill, R. L. *J. Biol. Chem.* **1999**, *274*, 31751–31754.
- (7) Thornton, D. J.; Sheehan, J. K. *Proc. Am. Thorac. Soc.* **2004**, *1*, 54–61.
- (8) Ambort, D.; van der Post, S.; Johansson, M. E. V.; MacKenzie, J.; Thomsson, E.; Kregel, U.; Hansson, G. C. *Biochem. J.* **2011**, *436*, 61–70.
- (9) Thornton, D. J.; Davies, J. R.; Kraayenbrink, M.; Richardson, P. S.; Sheehan, J. K.; Carlstedt, I. *Biochem. J.* **1990**, *265*, 179–186.
- (10) Round, A.; Berry, M.; McMaster, T.; Stoll, S.; Gowers, D.; Corfield, A.; Miles, M. *Biophys. J.* **2002**, *83*, 1661–1670.
- (11) Georgiades, P.; di Cola, E.; Heenan, R. K.; Pudney, P. D. A.; Thornton, D. J.; Waigh, T. A. *Biopolymers* **2014**, *101*, 1154–1164.
- (12) Ridley, C.; Kouvatso, N.; Raynal, B. D.; Howard, M.; Collins, R. F.; Desseyn, J.-L.; Jowitt, T. A.; Baldock, C.; Davis, C. W.; Hardingham, T. E.; Thornton, D. J. *J. Biol. Chem.* **2014**, *289*, 16409–16420.
- (13) Godl, K.; Johansson, M. E. V.; Lidell, M. E.; Mörgelin, M.; Karlsson, H.; Olson, F. J.; Gum, J. R.; Kim, Y. S.; Hansson, G. C. *J. Biol. Chem.* **2002**, *277*, 47248–47256.
- (14) Nilsson, H. E.; Ambort, D.; Backstrom, M.; Thomsson, E.; Koeck, P. J.; Hansson, G. C.; Hebert, H. *J. Mol. Biol.* **2014**, *426*, 2567–2579.
- (15) Backstrom, M.; Ambort, D.; Thomsson, E.; Johansson, M. E.; Hansson, G. C. *Mol. Biotechnol.* **2013**, *54*, 250–256.
- (16) Tu, A. T. *Raman Spectroscopy in Biology: Principles and Applications*; John Wiley & Sons Inc.: New York, 1982; p 464.
- (17) Kurouski, D.; Deckert-Gaudig, T.; Deckert, V.; Lednev, I. K. *Biophys. J.* **2014**, *106*, 263–271.
- (18) vandenAkker, C. C.; Deckert-Gaudig, T.; Schleegeer, M.; Velikov, K. P.; Deckert, V.; Bonn, M.; Koenderink, G. H. *Small* **2015**, *11*, 4131–4139.
- (19) Carlstedt, I.; Lindgren, H.; Sheehan, J. K.; Ulmsten, U.; Wingerup, L. *Biochem. J.* **1983**, *211*, 13–22.
- (20) Thornton, D. J.; Carlstedt, I.; Howard, M.; Devine, P. L.; Price, M. R.; Sheehan, J. K. *Biochem. J.* **1996**, *316*, 967–975.

- (21) Eilers, P. H. C.; Boelens, H. F. M. *Baseline Correction with Asymmetric Least Squares Smoothing*; Leiden University Medical Centre Report, 2005.
- (22) Kurouski, D.; Deckert-Gaudig, T.; Deckert, V.; Lednev, I. K. *J. Am. Chem. Soc.* **2012**, *134*, 13323–13329.
- (23) Rasmussen, A.; Deckert, V. *J. Raman Spectrosc.* **2006**, *37*, 311–317.
- (24) Gardner, B. *Untangling Fibrillogenesis: Investigations of the Mechanisms of Amyloid Formation*. Ph.D. Thesis, The University of Manchester, Manchester, U.K., 2014.
- (25) Johannessen, C.; Pendrill, R.; Widmalm, G.; Hecht, L.; Barron, L. D. *Angew. Chem., Int. Ed.* **2011**, *50*, 5349–5351.
- (26) Ashton, L.; Pudney, P. D.; Blanch, E. W.; Yakubov, G. E. *Adv. Colloid Interface Sci.* **2013**, 199–200, 66–77.
- (27) Loomis, R. E.; Prakobphol, A.; Levine, M. J.; Reddy, M. S.; Jones, P. C. *Arch. Biochem. Biophys.* **1987**, *258*, 452–464.
- (28) Starkey, B. J.; Snary, D.; Allen, A. *Biochem. J.* **1974**, *141*, 633–639.
- (29) Gunning, A. P.; Kirby, A. R.; Fuell, C.; Pin, C.; Tailford, L. E.; Juge, N. *FASEB J.* **2013**, *27*, 2342–2354.
- (30) Bell, A. F.; Hecht, L.; Barron, L. D. *J. Am. Chem. Soc.* **1994**, *116*, 5155–5161.
- (31) Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F. *Prog. Biophys. Mol. Biol.* **2000**, *73*, 1–49.
- (32) Barron, L. D.; Blanch, E. W.; Hecht, L. In *Advances in Protein Chemistry*; Academic Press: London, 2002; pp 51–90.
- (33) McColl, I. H.; Blanch, E. W.; Gill, A. C.; Rhie, A. G.; Ritchie, M. A.; Hecht, L.; Nielsen, K.; Barron, L. D. *J. Am. Chem. Soc.* **2003**, *125*, 10019–10026.
- (34) McColl, I. H.; Blanch, E. W.; Hecht, L.; Kallenbach, N. R.; Barron, L. D. *J. Am. Chem. Soc.* **2004**, *126*, 5076–5077.
- (35) Smyth, E.; Syme, C. D.; Blanch, E. W.; Hecht, L.; Vašák, M.; Barron, L. D. *Biopolymers* **2001**, *58*, 138–151.
- (36) Syme, C. D.; Blanch, E. W.; Holt, C.; Jakes, R.; Goedert, M.; Hecht, L.; Barron, L. D. *Eur. J. Biochem.* **2002**, *269*, 148–156.
- (37) Johannessen, C.; Kapitan, J.; Collet, H.; Commeyras, A.; Hecht, L.; Barron, L. D. *Biomacromolecules* **2009**, *10*, 1662–1664.
- (38) Zhu, F.; Kapitan, J.; Tranter, G. E.; Pudney, P. D.; Isaacs, N. W.; Hecht, L.; Barron, L. D. *Proteins: Struct., Funct., Genet.* **2008**, *70*, 823–833.
- (39) Carey, P. In *Biochemical Applications of Raman and Resonance Raman Spectroscopies*; Academic Press: London, 1982.
- (40) Miura, T.; Thomas, G., Jr. In *Proteins: Structure, Function, and Engineering*; Biswas, B. B.; Roy, S., Eds.; Springer: New York, 1995; pp 55–99.
- (41) Chen, M. C.; Lord, R. C. *J. Am. Chem. Soc.* **1974**, *96*, 4750–4752.
- (42) Shao, Z.; Vollrath, F.; Sirichaisit, J.; Young, R. J. *Polymer* **1999**, *40*, 2493–2500.
- (43) Yu, T. J.; Lippert, J. L.; Peticolas, W. L. *Biopolymers* **1973**, *12*, 2161–2175.
- (44) Dong, J.; Wan, Z. L.; Chu, Y. C.; Nakagawa, S. N.; Katsoyannis, P. G.; Weiss, M. A.; Carey, P. R. *J. Am. Chem. Soc.* **2001**, *123*, 7919–7920.
- (45) Blanch, E. W.; Hecht, L.; Syme, C. D.; Volpetti, V.; Lomonosoff, G. P.; Nielsen, K.; Barron, L. D. *J. Gen. Virol.* **2002**, *83*, 2593–2600.
- (46) Thornton, D. J.; Howard, M.; Khan, N.; Sheehan, J. K. *J. Biol. Chem.* **1997**, *272*, 9561–9566.
- (47) Carlstedt, I.; Lindgren, H.; Sheehan, J. K. *Biochem. J.* **1983**, *213*, 427–435.
- (48) Cowcher, D. P.; Deckert-Gaudig, T.; Brewster, V. L.; Ashton, L.; Deckert, V.; Goodacre, R. *Anal. Chem.* **2016**, *88*, 2105–2112.
- (49) Cheeseman, J. R.; Shaik, M. S.; Popelier, P. L. A.; Blanch, E. W. *J. Am. Chem. Soc.* **2011**, *133*, 4991–4997.
- (50) Mutter, S. T.; Zielinski, F.; Cheeseman, J. R.; Johannessen, C.; Popelier, P. L.; Blanch, E. W. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6016–6027.
- (51) Zielinski, F.; Mutter, S. T.; Johannessen, C.; Blanch, E. W.; Popelier, P. L. *Phys. Chem. Chem. Phys.* **2015**, *17*, 21799–21809.