J Mol Model (2011) 17:2953–2963
DOI 10.1007/s00894-010-0932-0

ORIGINAL PAPER

Modeling the human intestinal Mucin (MUC2) C-terminal cystine knot dimer

Vatsala D. Sadasivan · Sandeep R. Narpala · David E. Budil · Albert Sacco Jr · Rebecca L. Carrier

Received: 18 August 2010 / Accepted: 7 December 2010 / Published online: 12 February 2011

Abstract Intestinal mucus, a viscous secretion that lines the mucosa, is believed to be a barrier to absorption of many therapeutic compounds and carriers, and is known to play an important physiological role in controlling pathogen invasion. Nevertheless, there is as yet no clear understanding of the barrier properties of mucus, such as the nature of the molecular interactions between drug molecules and mucus components as well as those that govern gel formation. Secretory mucins, large and complex glycoprotein molecules, are the principal determinants of the viscoelastic properties of intestinal mucus. Despite the important role that mucins play in controlling transport and in diseases such as cystic fibrosis, their structures remain poorly characterized. The major intestinal secretory mucin gene, MUC2, has been identified and fully sequenced. The present study was undertaken to determine a detailed structure of the cysteine-rich region within the C-terminal end of human intestinal mucin (MUC2) via homology modeling, and explore possible configurations of a dimer of this cysteine-rich region, which may play an important role in governing mucus gel formation. Based on sequence–structure alignments and three-dimensional modeling, a cystine knot tertiary structure homologous to that of human chorionic gonadotropin (HCG) is predicted at the C-terminus of MUC2. Dimers of this C-terminal cystine knot (CTCK) were modeled using sequence alignment based on HCG and TGF-beta, followed by molecular dynamics and simulated annealing. Results support the formation of a cystine knot dimer with a structure analogous to that of HCG.

Keywords MUC2 · Mucin · Mucus · Cystine knot · Homology model · Dimer

Introduction

The mucosa, the inner layer of the intestinal wall, is the site of absorption of orally delivered compounds, including therapeutic drugs. The mucosa includes epithelial cells responsible for absorption and goblet cells that secrete a viscous mucus layer that lines the mucosa [1, 2]. Mucus provides an important selective barrier enabling absorption of nutrients and protecting against certain pathogens. The mucus layer has been identified as a significant barrier to absorption of some drugs [3–7]. However, the role of the mucus layer in drug transport has been largely neglected relative to that of the epithelial layer. In crossing the cell membrane of the epithelial layer, the lipophilic nature of the drug is generally of most importance. However, the drug must first diffuse through the hydrophilic mucus layer before encountering the epithelial layer. It is important for efficient drug delivery to understand mucus inter- and intra-

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0932-0) contains supplementary material, which is available to authorized users.

V. D. Sadasivan · A. Sacco Jr · R. L. Carrier (✉)
Department of Chemical Engineering, Northeastern University,
342 Snell Engineering Center, 360 Huntington Avenue,
Boston, MA 02115, USA
e-mail: r.carrier@neu.edu

S. R. Narpala
Department of Biology, Northeastern University,
134 Mugar Life Sciences Building, 360 Huntington Avenue,
Boston, MA 02115, USA

D. E. Budil
Department of Chemistry and Chemical Biology,
Northeastern University,
102 Hurtig Hall, 360 Huntington Avenue,
Boston, MA 02115, USA

molecular interactions and the role they could play in limiting molecular and particulate permeability.

Mucin glycoprotein structure

Secretory mucins are the main constituents of mucus [8–13]; they are responsible for the bulk of the viscoelastic and gel-forming characteristics as well as the protective functions of mucus. They are high molecular weight (10^6–10^7 g mol^{-1}) glycoprotein molecules composed approximately of 75% carbohydrates and 25% amino acids by weight [14]. These compounds have the ability to form intermolecular disulfide bridges, resulting in oligomeric mucin and gel formation. The oligosaccharides comprising a high percentage of mucin weight are O-linked to serine or threonine residues in the protein backbone. The middle region of the protein backbone contains numerous repeating sequences, including virtually all the O-linked oligosaccharide attachment sites.

In general, mucins are difficult to analyze by conventional biochemical and biophysical methods due to their large size and high degree of glycosylation. However, cDNA cloning and sequencing of mucin genes has provided considerable insight into the primary structure and, subsequently, the function of the molecules. The major intestinal secretory mucin gene, MUC2, has been identified and was the first human secretory mucin gene to be cloned and fully sequenced [15–17]. The protein core contains more than 5,100 amino acid (aa) residues in its most common allelic form [15, 16, 18]. The large middle tandem repeat domain consists of 23 aa (PTTTPITTTTTVTPTPTPTGTQT) repeating sequences, with the threonine (T) residues as potential sites of O-glycosylation and N-acetylgalactosamine at the reducing end of the carbohydrate chains [18]. In addition, cysteine rich domains similar to the D-domains of prepro-von-Willebrand platelet aggregating factor are situated towards the N and C termini of MUC2 and contain potential N-glycosylation sites [15, 16, 19, 20]. However, these potential glycosylation sites are fewer in number as compared to those of the central tandem repeat region, where the majority of the carbohydrates are attached [18].

Comparison of the sequence of MUC2 with that of purified intestinal mucin preparations demonstrates that it has the characteristic sequence and glycosylation patterns of the major secretory molecule. MUC2 mucin is therefore important to drug transport through mucus; it is responsible for volume exclusion and steric affects on drug diffusion associated with gel formation as well as potential molecular interaction with drugs, which can retard diffusion [3, 5, 6, 13, 21, 22]. Since details of the molecular structure of MUC2 mucin remain unknown, there is no clear understanding of its molecular interactions with drug molecules. A better understanding of the structure and expression of MUC2 may lead to better understanding of the pathology of several diseases, in addition to its influence on drug transport across the mucus layer.

Various techniques have been utilized to probe structural features of mucin molecules, including dynamic light scattering [23], atomic force microscopy [24], and small angle X-ray scattering [25], and there has been a recent focus on studies of structural features of mucin glycans [26]. However, tertiary model predictions of regions within MUC2 have not been reported. Although there have been some preliminary efforts carried out to investigate the 23-aa repeat unit of MUC2 [27, 28], little information is yet known regarding the structural motif of this region. In addition, detailed structural studies of the cystine knot motif present at the C-terminus have not been carried out on any mucin molecules, but X-ray crystallography of transforming growth factor-β2 (TGF- β2) and other growth factors has provided a precise picture of this unique structural feature [29, 30].

The cystine knot motif

The cystine knot three-dimensional (3D) structure is found in many extracellular molecules and is conserved among divergent species. To date, all known ten-membered cystine knot proteins have been found to be extracellular proteins, interacting with specific receptors and/or other extracellular proteins. However, identifying cystine knot structure-containing proteins using common approaches such as pairwise alignments is difficult because sequence homology is low among these proteins [31]. The cystine knot motif is most commonly found in transforming growth factor-β (TGF-β), nerve growth factors (NGF), glycoprotein hormones (GPH) and platelet-derived growth factor-β (PDGF-β) subfamily members [31–35]. The largest ten-membered cystine knot subfamily is the TGF-β family, which consists of growth factors such as transforming growth factor-β2 (TGF-β2), bone morphogenetic proteins (BMP) and glial cell line-derived neurotrophic factor (GDNF) [29, 36–38]. In addition to the well-known members of the cystine knot superfamily, several other proteins have been proposed to display a potential cystine knot structure. These include proteins of the mucin-like, slit-like and jagged-like subgroups [31]. The mucin like-subgroup includes von Willebrand factor (vWF) [39] and Norrie disease protein [40], as well as several bone morphogenetic antagonists [38, 41] and mucin-related proteins [31]. In addition to the well-known ten-membered cystine knot structure, the nerve growth factor family has been described to adapt a similar arrangement, with nine rather than three residues present between the second and third cysteine, leading to the formation of a 16-membered cystine knot [33].

The consensus sequence that is conserved among all the ten-membered cystine knot proteins is:

C_1-$(X)_n$-C_2-X-G-X-C_3-$(X)_n$-C_4-$(X)_n$-C_5-X-C_6

part of ring part of ring

A significant part of the cystine knot tertiary structure is composed of a consensus cysteine framework important for a homologous folding motif called the 'cysteine ring,' which was found to consist of four cysteine residues with a cysteine (C) spacing of C–$(X)_3$–C and C–X–C [42, 43]. These eight amino acids form a cysteine ring through two disulfide bridges (Fig. 1). Cysteines 2, 3, 5 and 6 form the cysteine ring through disulfide bonds between cysteines 2 and 5 as well as cysteines 3 and 6. This ring structure is conserved in many different groups of proteins, including the cystine knot superfamily [44]. Cystine knot superfamily members contain two additional cysteines forming a disulfide bond that penetrates the ring, thus forming a cystine knot with ten amino acids, of which six are cysteine residues [32] (Fig. 1). The third disulfide bond is formed between cysteines 1 and 4, where it penetrates through the ring forming the knot.

In addition to this framework of six cysteine residues, other features are also conserved among all the ten-membered cystine knot proteins. These include the conservation of a glycine residue between cysteines 2 and 3. The intrusion of the additional disulfide bond through the cysteine ring confines the amino residue between the second and third cysteine to a glycine with a positive φ backbone angle, because any other amino acid at this position would cause severe steric hindrance for the formation of the knot [33, 45]. In addition, in some cystine

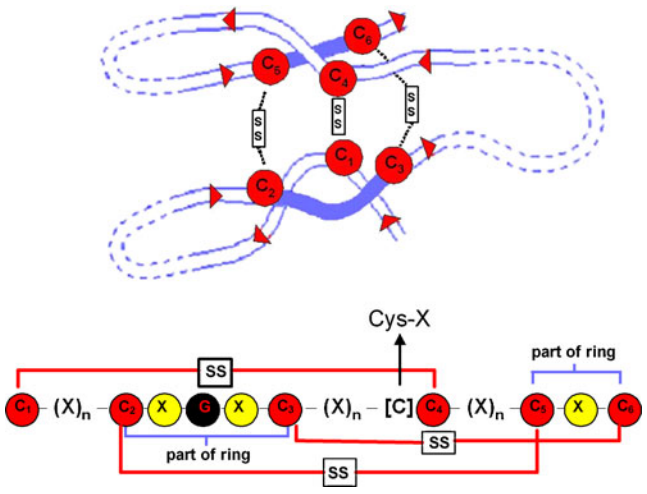

Fig. 1 Schematic of the cystine knot structure. *Arrows* indicate the direction (N to C terminal) of the amino acid side chains. *SS* Disulfide bonds. The six cysteine residues involved in the knot formations are labeled as C_1 to C_6 and their spacing is shown in the panel below

knot proteins, an additional cysteine is located between cysteines 3 and 4 (adjacent to cysteine 4) as shown in Fig. 1, which strengthens dimerization by forming a covalent disulfide bridge between the two subunits of the dimer [46, 47]. The peptide regions between the cysteine residues extend outwards from the central knot and twist into three variable external loops of approximately 13–16 residues each [48]. Two of the loops, between cysteines 1 and 2, as well as cysteines 4 and 5, consist of antiparallel β-strands that form finger-like projections, whereas the third domain between cysteines 3 and 4 usually contains an α-helical structure [38, 49]. Due to the arrangement of these loops, the 3D structure of these proteins is sometimes referred to as a "hand" containing two fingers and a middle heel [31].

In recent years, additional features have been identified in cystine knot proteins. Analysis of the residues between cysteines 2, 3, 5 and 6 revealed that none of the highly hydrophobic amino acids, Trp, Phe, Tyr, Ile, Leu, Val and Met, are present among the residues before or after cysteine 5 [31]. Although it is unclear how hydrophobic residues at these positions might interfere with cystine knot formation, this finding facilitates the analysis and search for cystine knot members. Overall, the amino acid residues surrounding the cystine knot have been found to have low sequence homology among existing cystine knot proteins [31, 40]. Among the existing cystine knot proteins with known 3D structures, it was found that the sequence identity does not exceed 10%, and similarity does not exceed 25%. Although these proteins have been found to have variable sizes (e.g., 55–5,400 with the majority at 400 amino acids), they have a similar size in their cystine knot signature. The length between cysteines 2 and 6 of the cystine knot was found to vary from 42 to 80 amino acids, and, in cystine knot proteins with known 3D structures, the knots were found to be located in the C-terminal region of the protein [C-terminal cystine knot (CTCK)] [31]. Because the size of the cystine knot signature in these proteins is not related to the total size of the protein, this motif likely represents a structural entity and not a random arrangement of cysteines. These findings also suggest that gene fusion events took place during evolution, and the same motif was used in diverse proteins with different functions.

As a first step towards understanding mucus molecular structure and interactions, the MUC2 mucin molecule was investigated. A detailed structural investigation of the cysteine-rich domain at the carbon terminus of MUC2 was carried out utilizing homology or comparative modeling. Efforts to analyze the structure of the C-terminus cysteine-rich domain of MUC2 revealed a cystine knot tertiary structure homologous to that of HCG based on sequence–structure alignments and 3D modeling. Potential

dimer structures based on alignment with HCG and TGF-β were explored.

Methodology

CTCK homology model development

The MUC2 sequence was obtained from the SWISS-PROT/TrEMBL database through the ExPASy (Expert Protein Analysis System) proteomics server of the Swiss Institute of Bioinformatics (SIB) [50]. The BLASTP program [51] available on the ExPASy server was used to compare the sequence of the C-terminal cysteine-rich domain of MUC2 against protein sequences in the Protein Data Bank (PDB) [52] to find structurally well-defined proteins (with experimentally solved 3D structure) whose sequences are homologous to the MUC2 protein sequence. MUC2 amino acid residues 5068–5159 were investigated. Having identified possible modeling templates, the T-COFFEE [53] and CLUSTAL W [54] programs were used to carry out pairwise sequence alignment and multiple sequence alignments, respectively. All experimentally derived protein structures were obtained from the Research Collaboratory for Structural Bioinformatics (RCSB) PDB [52]. The protein sequence displaying the highest homology with the carboxy-terminal end of MUC2—chain-B of human chorionic gonadotropin (HCG-B)—was used as the template for building the model. The structure of this molecule has been solved for amino acids 2–111 of the entire 145 amino acid sequence [34, 35]. An initial model was first developed with the SWISS MODEL server [55, 56] using the 'first approach mode.' The model was further improved by performing manual threading using the Deep View (Swiss PDB Viewer) program [56]. The manually threaded Deep View file was then resubmitted to the SWISS MODEL server for model building via the 'project/optimize mode.' The GROMOS96 force field [57] is used by SWISS MODEL server to perform energy minimization. The quality and structural features of the final optimized model were evaluated and compared with the template structure using various structural assessment methods including ANOLEA [58], PROCHECK [59], What Check [60] and Verify3D [61]. The Deep View (Swiss PDB Viewer) [56], BioMedCAChe Version 6.1.12,[1] and VMD [62] were used to view the final model.

[1] BioMedCAChe, Version 6.1.12 (2005); BioSciences Group, Fujitsu Computer Systems, Corp., 15244 NW Greenbrier Parkway, Beaverton, Oregon, 97006

Dimer model development

The final CTCK model was further used to design three dimer models. The 3D structure of HCG (PDB ID: 1HCN) was used as a template to construct the first and the second MUC2 CTCK dimer models, which differ only in that a disulfide bond was incorporated between the two chains of the dimer in the second model. These will be referred to as the HCG noCysX and HCG CysX models, respectively. The crystal structure of transforming growth factor-β2 (TGF-β2) was used as a template to construct the third model, which will be referred to as the TGFβ model. As in the HCG CysX model, a disulfide bond was incorporated between the two chains of this dimer at position 54 of the two chains. DaliLite [63], a tool for pair-wise comparison of protein structures, was used to translate the models with respect to the template.

All simulations were performed with NAMD (scalable molecular dynamics) [64] using the CHARMM22 force field. The systems were surrounded with a 10 Å layer of water molecules. No counter ions were introduced in the system, and the solvated proteins were minimized by 2,000 conjugate gradient steps followed by a 0.1 ns molecular dynamics (MD) simulation, utilized as an initial equilibration step. A simulated annealing (SA) simulation was started from the previous MD run, heating the system from 300 K to 1,050 K during 0.3 ns at temperature increments of 5 K every 1,000 steps, and cooled the system from 1,000 K to 300 K at temperature decrements of 5 K every 1,000 steps.

The resulting dimer models were analyzed with respect to general structural features and similarities, distance between residue 54 (Cys-X cysteine) sulfur atoms on each chain, and hydrophobic and total solvent accessible surface area (SASA). Distances and SASAs were measured in visual molecular dynamics (VMD) using a solvent radius of 1.4, and restricting the analysis to the hydrophobic residues to calculate hydrophobic surface area. The root mean squared deviation (RMSD) of the backbone atoms was calculated relative to the starting structures in the MD and simulated annealing trajectories. The local secondary structure was also determined for each residue throughout the simulation.

Results

Based on sequence–structure alignments and 3D modeling, a cystine knot tertiary structure at the C-terminus of MUC2 is predicted to be homologous to that of chain B HCG (HCG-B, PDB ID: 1HCN). The cysteine-rich domain at the C-terminus of MUC2 also shows sequence homology with other common existing cystine knot containing proteins. A

dimer model formed based on the structure of HCG appears to be stable compared to one based on TGF-β, another cystine-knot-containing dimer, with which the subunit containing the CTCK of MUC2 shares some sequence homology.

Sequence–structure alignment

Among all the experimental 3D structures currently available from the PDB, MUC2 exhibits the highest sequence identity and similarity with chain HCG-B, with values of 26% and 39%, respectively (Fig. 2). HCG is a glycoprotein hormone produced by trophoblastic cells of the placenta in both pregnancy and gestational trophoblastic diseases [35, 65]. Similar to other members of the glycoprotein family of hormones (luteinizing hormone, follicle-stimulating hormone and thyroid-stimulating hormone), chorionic gonadotropin is a non-covalent heterodimer composed of a 92-amino acid α-subunit with five disulfide bridges and two N-linked glycosylation sites, and a 145-amino acid β-subunit with six disulfide bridges and two N-linked and four O-linked glycosylation sites [66]. It is an unusual glycoprotein in that up to 35% of the molecular weight (MW) is from oligosaccharide side chains [65]. The 3D structure of HCG was determined based on experimental X-ray diffraction methods [35]. The 3D structure shows that each of its two different subunits (α/A and β/B) has a similar topology, with three disulfide bonds forming a cystine knot [35, 67].

Although the sequence homology between MUC2 and HCG-B is relatively low, they do share the cystine knot motif (Fig. 2a), which is conserved among all the ten-membered cystine knot proteins. The alignment of the MUC-2 primary sequence to HCG-B reveals several important conserved features. Most important is the conservation of the six cysteine residues (C_1–C_6, Fig. 2a), which appear with the correct spacing. Very few insertions and deletions have to be introduced to align MUC2 and HCG-B (four residues in three gaps). In particular, the knot forming cysteines representing the core of the protein fold can be aligned without gaps [cysteine ring framework consisting of four cysteine residues with a cysteine (C) spacing of C_2–$(X)_3$–C_3 and C_5–X–C_6 is conserved]. Consequently, the sizes of the cystine knots (C_1–C_6) in MUC2 and HCG-B are very similar, at 80 aa and 82 aa, respectively. In addition, a structurally important glycine between cysteine 2 and 3 with a positive φ angle—an unusual conformation—is conserved. There is an additional cysteine residue present adjacent to cysteine 4 in the MUC2 protein, which is also commonly observed among other cystine-knot-containing proteins but is missing in HCG-B. This cysteine is thought to play a role in the covalent

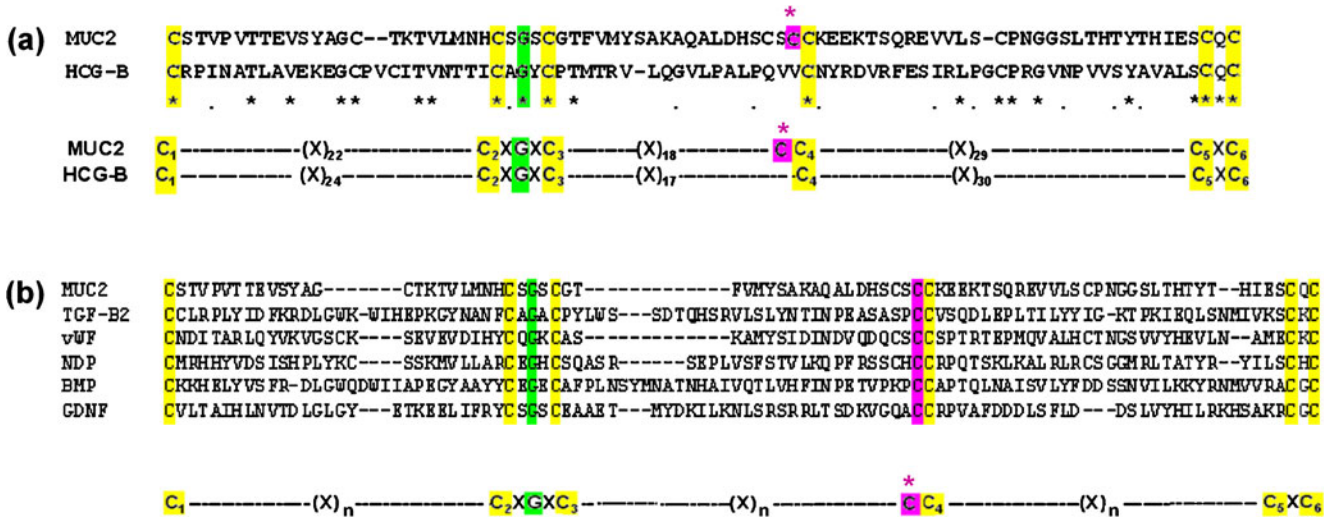

Fig. 2 Sequence alignments of the C-terminal cysteine-rich domain of major intestinal secretory mucin (MUC2) with the chain-B of human chorionic gonadotropin (HCG-B) and other cystine knot containing proteins. The length of the C-terminus cysteine-rich domain of MUC2 sequence, including C_1–C_6 only, is 80 aa (MUC2-human (UniProtKB/Swiss-Prot entry Q02817): aa 5075–5154 [50]). Alignment of the six cysteine residues (numbered C_1–C_6) and the glycine residue are highlighted in *yellow* and *green*, respectively. Cys-X is highlighted in *purple* and marked by an *asterisk* (*). **a** Optimal sequence alignment between MUC2 and HCG-B obtained using Deep View (Swiss PDB) [56] where manual threading was carried out. The number of amino acids (X_n) between conserved cysteine residues (C) is given. **b** Multiple sequence alignment of MUC2 and other cystine knot proteins using CLUSTAL W [54]. *TGF-β2* Transforming growth factor type β2 (PDB ID: 1TFG), *vWF* von Willebrand factor (experimentally derived structure consisting of the cystine knot region not available, sequence obtained from SWISS-PROT/TrEMBL (UniProtKB/Swiss-Prot entry Q00604): aa 2724–2804), *NDP* Norrie disease protein (experimentally derived structure not available, sequence obtained from SWISS-PROT/TrEMBL (UniProtKB/Swiss-Prot entry P04275): aa 39–128), *BMP* bone morphogenetic protein (PDB ID: 1BMP), *GDNF* glial cell line-derived neurotrophic factor (PDB ID: 1AGQ)

dimerization of rat MUC2 via formation of a disulfide bond. Following Bell et al. [51], we will refer to this residue as Cys-X.

Multiple sequence alignment of the C-terminal cysteine-rich domain of MUC2 and other cystine-knot-containing proteins further reveals the unique conservation of the ten-membered cystine knot motif as well as the additional cysteine residue adjacent to cysteine 4 in MUC2 (Fig. 2b). Differences can be observed in the length of the two twisted loops that extend from cysteine residues 1 and 2, from cysteine residues 4 and 5, and also in the third domain between cysteines 3 and 4.

3D model building

Based on the sequence–structure alignment of the C-terminal cysteine-rich domain of MUC2 with HCG-B, a full 3D model of the MUC2 domain was built. A comparison of these two structures is shown in Fig. 3a.

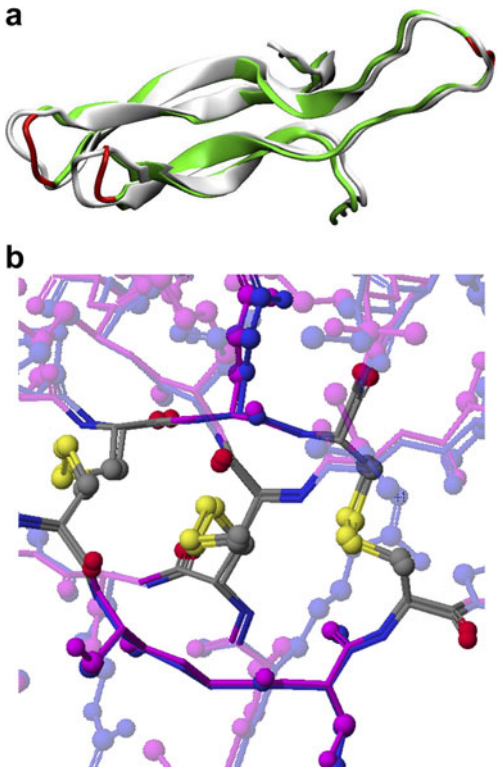

Fig. 3 Superposition of the 3D MUC2 model (aa 5071–5155 shown) and HCG-B (aa 6–91 shown). **a** Backbone ribbon representation: *green* 3D model of the C-terminus of MUC2, *white* HCG-B. *Red* areas in the 3D model indicate areas with poorer overlap in comparison to other regions in the entire structure. The root mean squared deviation (RMSD) (α-carbons) value is 2.38 Å. **b** Close up of the superposition of the ten-membered cystine knot structures emphasizing the alignment of the disulfide bonds (yellow); *pink* homology model of MUC2, *blue* HCG-B. The RMSD (α-carbons) value for the ten conserved cystine knot residues [C_1, C_2–$(X)_3$–C_3, C_4 and C_5–X–C_6] is 0.08 Å

The final optimized model displays the cystine knot formed by the six cysteine residues through three disulfide bonds as well as the highly twisted anti-parallel β-sheet structure consisting of three variable loops, a characteristic of cystine-knot-containing proteins [31, 48]. Close overlap between the backbones of MUC2 and HCG-B is observed, reflecting conservation of structural features between the two proteins. In both structures, the highly twisted anti-parallel β-sheet structure of the "finger" loops each contain an extended ladder of hydrogen bonds [34]. Hydrophobic residues are found to be located and spatially compact between the anti-parallel β-sheet "fingers," as well as at the tip of the third loop in both structures [35]. In addition, both MUC2 and HCG-B contain an additional disulfide bond linking the two anti-parallel β-sheet "fingers."

Finally, both HCG-B and the MUC2 model display a pair of β-sheet bulges in adjacent positions on each strand of the C-terminal β-sheet "fingers." All cystine knot growth factors possess these β-sheet bulges immediately adjacent to C_5 of the cystine knot, indicating they are an important feature of this motif [35]. Most importantly, the superposition indicates a precise alignment of the cystine knot region, where the ten conserved cystine knot residues [C_1, C_2–$(X)_3$–C_3, C_4 and C_5–X–C_6] in both molecules superimpose onto one another with precise alignment of the disulfide bonds (Fig. 3b). The α-carbon RMSD values were 2.38 Å for the entire structure (all residues) and 0.08 Å for the ten conserved cystine knot residues. This further indicates that the cystine knot signature in the C-terminus of MUC2 is highly conserved.

The 3D homology model was checked using various structural verification and assessment tools. The ANOLEA and PROCHECK methods were used; the results based on these methods are shown in Supplementary Figs. 1 and 2. The ANOLEA method of assessing mean force potential results in a plot in which the *y*-axis represents the energy for each amino acid of the protein chain (Supplementary Fig. 1). Negative energy values represent a favorable environment whereas positive values represent an unfavorable environment for each amino acid. The analysis indicated that the majority of amino acids in the model are in favorable energy environments, and that the energy profiles are very similar for the MUC2 model and HCG-B despite the lack of sequence homology. The Ramachandran plots generated by the PROCHECK [62] program (Supplementary Figure 2) display the psi (ψ) and phi (ϕ) dihedral angles of each residue in the protein structure to assess the "stereochemical quality" of a structure. The majority of the residues in the model correspond to regions on the plot where there are no steric clashes, and again the plots are similar for the MUC2 model and HCG-B. In general, these methods indicate

positive stereochemical quality and no significant structural errors in the 3D homology model when compared to HCG-B. Similar conclusions were made using What Check and Verify3D (data not shown). The close similarity of the quality and structural features of the homology model with the experimental HCG-B structure increases confidence in the reliability of the model.

Dimer models—general features

Figure 4 shows the models of the three possible dimers constructed from the homology model of the MUC2 Cys-rich domain. The models based on HCG have the characteristic "palm to palm" interaction of the cystine knots (using the "hand" analogy for the structure), as opposed to the intertwined nature of the interaction characteristic of TGF-β, which is also observed in the TGFβ model (Fig. 4). The characteristic antiparallel β-strands of the "fingers," a characteristic of cystine knots,

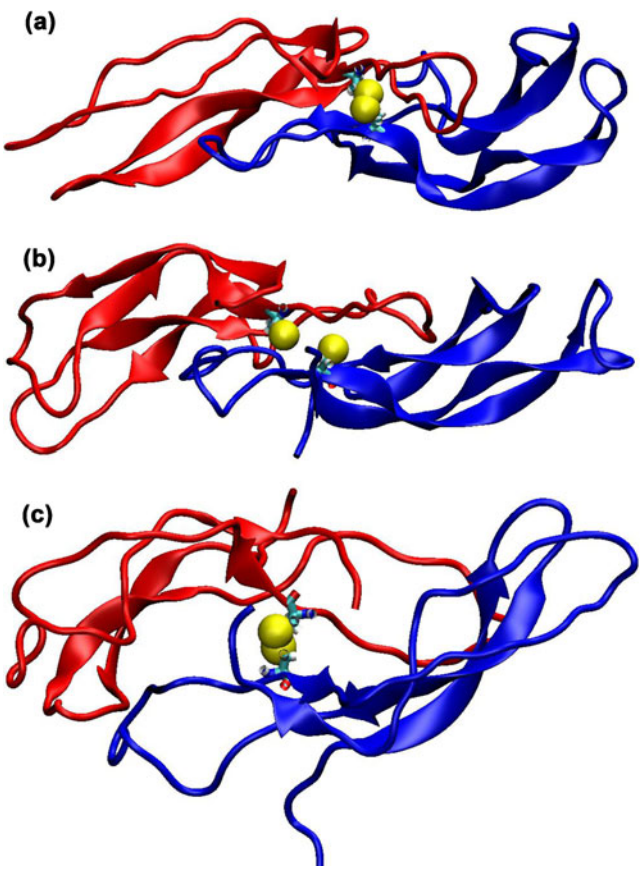

Fig. 4 Homodimer models of MUC2 in ribbon presentation with the chains colored red and blue. **a** HCG CysX model based on the 3D structure of HCG (1HCN) with an inter-chain disulfide bond between Cys54 of each chain. **b** HCG noCysX model; as in **a**, but without the interchain disulfide bond. **c** TGFβ model based on the 3D structure of TGF-β2

were conserved in the HCG CysX model, but not entirely in the other models. In general, the HCG CysX and HCG noCysX models exhibit some structural dissimilarities, even though they share gross structural features overall, as indicated by an overlay of the two models (Supplementary Figure 3).

Investigating the distance between the residue 54 (Cys-X) sulfur atoms on each strand of the dimer, results indicate that the Cys-X bond is not likely to be formed in the model based on HCG without explicit formation of this bond, as the distance between sulfur atoms in the relevant residues is much greater than a stable disulfide bond (2.05 Å) (Supplementary Table 1). In fact, this model contains a disulfide bond in each segment of the dimer between residue 54 (Cys-X) and residue 52. However, in the HCG CysX model, this bond appears to maintain a relatively stable length throughout both simulations. In the TGF-β model, the Cys-X bond length deviates somewhat from a stable disulfide bond length to 1.98–1.99 Å.

SASAs, particularly for hydrophobic residues, were examined (Supplementary Table 2). The formation of the Cys-X bond in the HCG-based model reduced the hydrophobic SASA to 3,151 $Å^2$, compared to 3,418 $Å^2$ and 3,870 $Å^2$ in the HCG noCysX and TGFβ models, respectively, indicating a more energetically favorable configuration in the aqueous environment of the mucus gel. Both HCG-based models had lower hydrophobic SASAs than the TGF-β based model. Visualization of hydrophobic residues on the three models (colored silver in Supplementary Figure 4) demonstrates that most of them are tucked inside the structure rather than exposed on the exterior of the molecule, especially in the HCG CysX model.

Average RMSD trajectories for the models during the MD (Fig. 5a) and SA (Fig. 5b) simulations indicate that all three models appeared to achieve a somewhat stable configuration during MD. The HCG CysX model leveled off most quickly and at the lowest value, around 1.5–1.75 Å. The HCG noCysX model leveled off at a slightly greater RMSD (~1.75–2.0 Å), while the TGFβ model demonstrated considerable fluctuations and a general rise in RMSD throughout most of the MD simulation, indicating a generally less stable starting structure. Parallel trends governed the SA portions of the simulations, with RMSD for the HCG CysX model leveling off first at approximately 4 Å, while the RMSD for the HCG noCysX model continued to rise to approximately 5 Å, and the TGFβ model's RMSD rose throughout most of the simulation to a value of approximately 6 Å. It is interesting to note that the relative stability of the HCG CysX model is due not only to the formation of the Cys-X bond, as the TGFβ model also contained this bond.

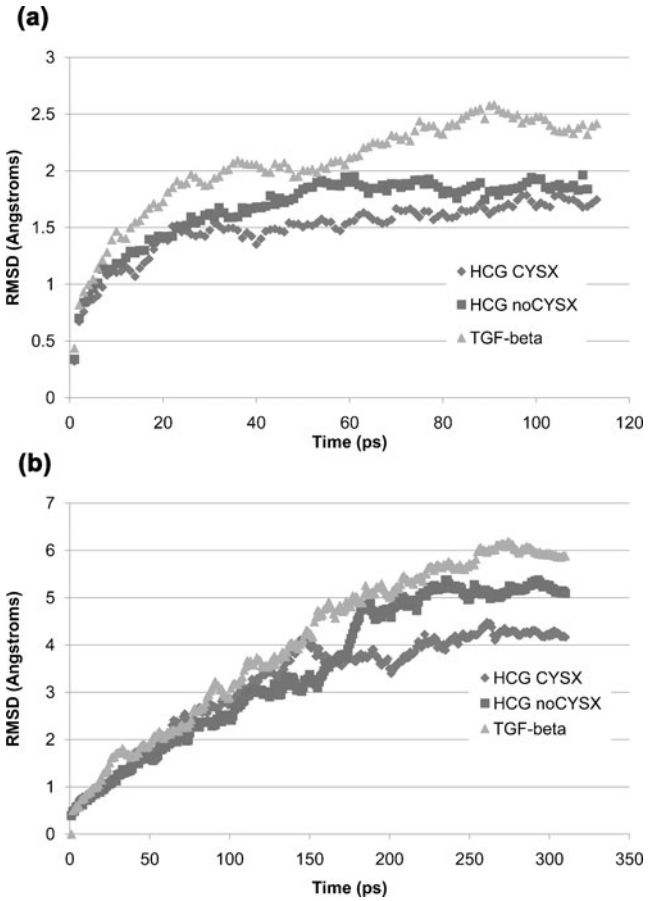

Fig. 5 Average RMSD values for each model during **a** molecular dynamics (MD) and **b** simulated annealing (SA) simulations

To test the hypothesis that the cystine-knot-containing portions of the molecules would be relatively stable in the dimer models, the average RMSD over each simulation was calculated for each residue, and residues were colored according to this value (Fig. 6). In general, the cystine knot portions of the models, corresponding to the central region of interaction of the two chains in each model, were more stable than the rest of the residues during the simulations, and the region of greatest stability (lowest RMSD) appeared to grow from the cystine knot region outward from the end of the MD to the end of the SA simulations.

Discussion

Based on sequence–structure alignments and 3D modeling, a cystine knot tertiary structure homologous to that of HCG is predicted at the C-terminus of MUC2. The 3D model of the C-terminus of MUC2 is, to the authors' knowledge, the first detailed structural study carried out on the cystine knot motif in MUC2, or on any secretory mucin, although gel-forming secretory mucins have been suggested to contain cystine knots through genomic analysis and mutagenesis [31, 48]. Although the sequence homology of MUC2 and HCG-B is low (26% sequence identity), it is still considered to be relatively high compared to the sequence homology among existing cystine knot proteins with known 3D structures, where the percentage figures do not exceed 10% for sequence identity and 25% for similarity. In addition, key similarities between the cystine knot portions of the two molecules support the homology model approach: the size of the cystine knot of MUC2 (80 aa) is very similar to that of HCG-B (82 aa), and only a few gaps had to be introduced to align the C-terminal end of MUC2 and HCG-B, resulting in the entire structures of the C-terminal ends of MUC2 and HCG-B being very similar.

The cystine knot forces the protein to adopt a 3D arrangement that exposes, in part, specific hydrophobic residues to the exterior of the molecule, likely facilitating protein–protein interactions. Thus, it is hypothesized that exterior hydrophobic residues, which have also been described for selected members of all cystine knot subfamilies, including HCG [33, 49, 68], contribute to the formation of homo- or hetero-dimers. Cys-X, which is not present in HCG-B, has also been hypothesized to play a key role in strengthening dimerization [48]. So far, all of the identified cystine knot proteins with known 3D structures are found to form dimers that bind to specific receptors [29–33, 35, 69, 70]. However, the mode of dimerization is different for each [32]. Taken together, these observations, together with the fact that several studies have indicated that the cysteine-rich domain at the C-terminus of MUC2 is involved in forming covalently linked dimers [48, 71], strongly supported exploration of dimer formation utilizing the developed model.

It can be concluded, based on RMSD analysis during the MD and SA simulations, length of the Cys-X bond, and observation of hydrophobic SASA, that a dimer model based on the HCG dimer, but with the Cys-X bond explicitly formed in the initial configuration, is more stable than an HCG-based model without the Cys-X formation or a TGF-β based model, which does have Cys-X formation. In addition, the six residues involved in the cystine knot formation of MUC2 appear to be important for dimerization, as indicated by the relative stability of this portion of the two strands of the dimer compared to the rest of the residues, probably by allowing correct positioning of Cys-X for stable intermolecular disulfide bond formation, as was suggested for rat muc2 [48].

Although the molecular function of the cystine knot in MUC2 is still unclear, these results support its significance in mucus gel formation. Linking at the C-terminus together

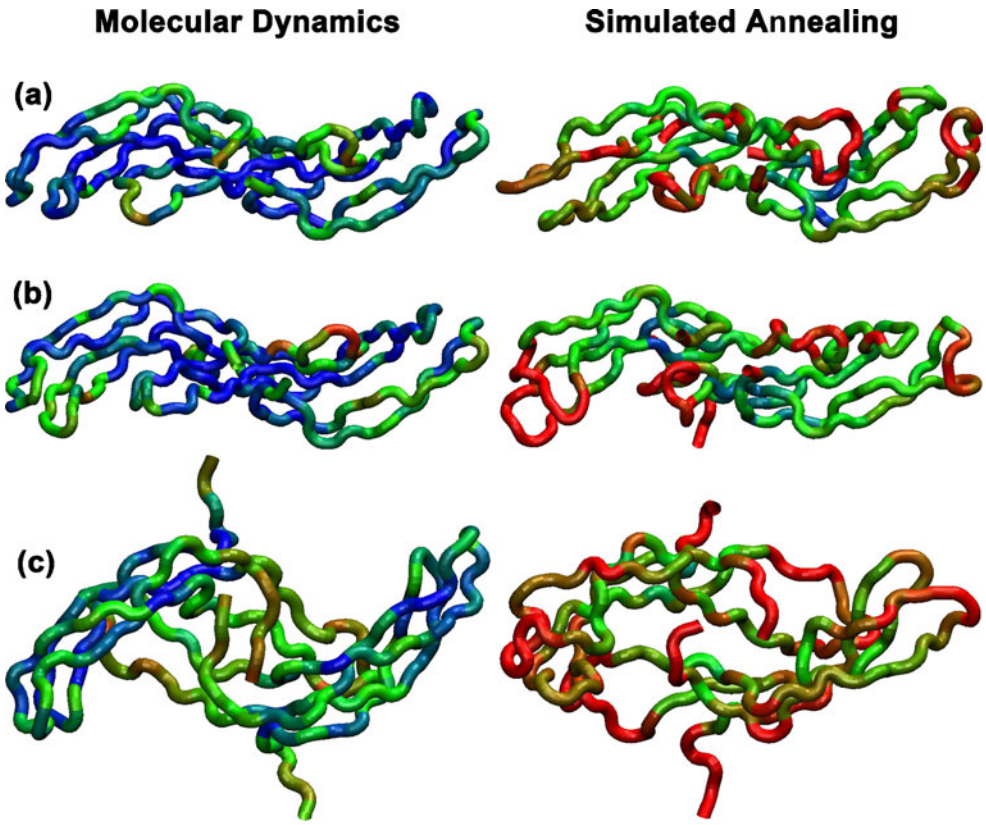

Fig. 6 Model dimers at the end of the MD (*left*) and SA (*right*) portions of the simulations, with residues color coded by RMSD for the HCG CysX (**a**), HCG noCysX (**b**) and TGFβ (**c**) models. RMSD was scaled between 0 and 5 Å for all molecules, with *blue* corresponding to the lowest RMSD and *red* to the highest

with cross-linking of the cysteine rich domains in the N-terminus (forming trimers) are known to result in formation of multimeric insoluble mucin gels [72, 73]. It could also be that the cystine knot in MUC2 functions to bind to a specific receptor as it does in most other cystine knot proteins, including HCG-B [29–33, 35, 69, 70]. It was demonstrated recently that the protozoan parasite *Entamoeba histolytica* secretes a cysteine protease that cleaves the cysteine-rich C-terminal domain of MUC2 mucin at residues 4322 and 4681 [74]. Cleavage at residue 4319 dissolves the mucin gel, releasing the portion of the molecule including the cystine-knot region discussed here. It is possible that this cleavage frees this portion of the molecule, facilitating binding to a receptor. Thus, further efforts to understand the cystine knot's biological function as well as its contribution to and impact on the overall structure of MUC2 would be meaningful.

Conclusions

The 3D model of the C-terminus of MUC2 predicts a highly conserved ten-membered cystine knot, thus providing strong evidence for the existence of the cystine knot in MUC2 and novel insight into the structure of MUC2. It is important to note that although the model is in agreement with a cystine knot structure, it must be considered as a first approximation and not a substitute for an experimentally derived tertiary structure. Nevertheless, it provides the basis for precise predictions of functional details that can be tested experimentally. Molecular details can be read from the model before the protein is expressed and its shape investigated experimentally. In addition, as demonstrated here, the model can be used to explore potential mucin dimer formation and its dependence on physicochemical parameters such as pH. The preliminary dimer modeling presented supports a dimer structure similar to that of HCG, but with a Cys-X disulfide bond formation. The manner in which the cystine knot affects the overall structure of MUC2, its impact on drug transport through the mucus layer, and its relation to any mucin-related diseases are open to investigation. For these areas of future research, the 3D model of the C-terminus of MUC2 could provide insight. Understanding of the impact of mucus structure and molecular interactions on drug transport will provide essential guidance for design of drugs and delivery agents, enabling more rapid drug development. In addition, understanding of the barrier properties of mucus will generally provide insight into the significance of this barrier in physiological function.

References

1. Sakata T, Engelhardt WV (1981) Luminal mucin in the large intestine of mice, rats and guinea pigs. Cell Tissue Res 219:629–635
2. Rozee KR, Cooper D, Lam K, Costertom JW (1982) Microbial flora of the mouse ileum mucous layer and epithelial surface. Appl Environ Microbiol 43:1451–1463
3. Thomson AB, Dietschy JM (1977) Derivation of the kinetics that describe the effects of unstirred water layers on the kinetic parameters of active transport processes in the intestine. J Theor Biol 64:277–294
4. Westergaard H, Dietschy JM (1974) Delineation of the dimensions and permeability characteristics of the two major diffusion barriers to passive mucosal uptake in the rabbit intestine. J Clin Invest 174:718–732
5. Smithson KW, Millar DB, Jacobs LR, Gray G (1981) Intestinal diffusional barrier: unstrirred water layer or membrane surface water coat? Science 214:1241–1244
6. Hughes DRL (1988) The influence of intestinal mucus on drug absorption. PhD thesis, Brighton Polytechnic, Brighton
7. Desai MA, Vadgama P (1991) Dependence of hydrochloric acid diffusion through gastric mucus: correlation with diffusion through a water layer using a membrane mounted glass pH electrode. Analyst 116:463–467
8. Neutra MR, Forstner JF (1987) Gastrointestinal mucus: synthesis, secretion and function. In: Johnson LR (ed) Physiology of the gastrointestinal tract. Raven Pss, New York
9. Allen A (1983) Mucus—a protective secretion of complexity. Trends Biochem Sci 8:169–173
10. Roussel P, Lamblin G, Lhermitte M, Houdret N, Lafitte JJ, Perini JM, Klein A, Scharfman A (1988) The complexity of mucins. Biochimie 70:1471–1482
11. Hansson GC, Sheehan JK, Carlstedt I (1988) Only trace amounts of fatty acids are found in pure mucus glycoproteins. Arch Biochem Biophys 266:197–200
12. Filipe MI (1979) Mucins in the human gastrointestinal epithelium: a review. Invest Cell Pathol 2:195–216
13. Khanvilkar K, Donovan MD, Flanagan DR (2001) Drug transfer through mucus. Adv Drug Deliv Rev 48:173–193
14. Bansil R, Stanley E, LaMont JT (1995) Mucin biophysics. Annu Rev Physiol 57:635–657
15. Gum JR, Hicks JW, Toribara NW (1992) The human muc2 intestinal mucin has cysteine-rich subdomains located both upstream and downstream of its central repetitive region. J Biol Chem 267:21375–21383
16. Gum JR Jr, Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) Molecular cloning of human intestinal mucin (muc2) cDNA. Identification of the amino terminus and overall sequence similarity to prepro-von Willebrand factor. J Biol Chem 269:2440–2446
17. Gendler SJ, Spicer AP (1995) Epithelial mucin genes. Annu Rev Physiol 57:607–634
18. Allen A, Hutton DA, Pearson JP (1998) The muc2 gene product: a human intestinal mucin. Int J Biochem Cell Biol 30:797–801
19. Toribara NW, Gum JR Jr, Culhane PJ, Lagace RE, Hicks JW, Petersen GM, Kim YS (1991) Muc-2 human small intestinal mucin gene structure. Repeated arrays and polymorphism. J Clin Invest 88:1005–1013
20. Gum JR, Byrd JC, Hicks JC, Toribara NW, Lamport DTA, Kim YS (1989) Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence for genetic polymorphism. J Biol Chem 264:6480–6487
21. MacAdam A (1993) The effect of gastro-intestinal mucus on drug absorption. Adv Drug Deliv Rev 11:201–220
22. Widdicombe JG (1997) Airway liquid: a barrier to drug diffusion? Eur Respir J 10:2194–2197
23. Yakubov GE, Papagiannopoulos A, Rat E, Easton RL, Waigh TA (2007) Molecular structure and rheological properties of short-side-chain heavily glycosylated porcine stomach mucin. Biomacromolecules 8:3467–3477. doi:10.1021/bm700607w
24. Brunelli R, Papi M, Arcovito G, Bompiani A, Castagnola M, Parasassi T, Sampaolese B, Vincenzoni F, De Spirito M (2007) Globular structure of human ovulatory cervical mucus. FASEB J 21:3872–3876. doi:101096/fj.07-8189com
25. Di Cola E, Yakubov GE, Waigh TA (2008) Double-globular structure of porcine stomach mucin: a small-angle X-ray scattering study. Biomacromolecules 9:3216–3222. doi:10.1021/bm800799u
26. Jensen PH, Kolarich D, Packer NH(2010) Mucin-type O-glycosylation—putting the pieces together. FEBS J 277:81–94. doi:10.1111/j.1742-4658.2009.07429.x
27. Uray K, Kajtár J, Vass E, Price MR, Hollósi M, Hudecz F (1999) Effect of solution conformation on antibody recognition of a protein core epitope from gastrointestinal mucin (muc2). Arch Biochem Biophys 361:65–74
28. Uray K, Price MR, Majer Z, Vass E, Hollosi M, Hudecz F (2003) Indentification and solution conformation of multiple epitopes recognized by a muc2 mucin-specific monoclonal antibody. Arch Biochem Biophys 410:254–260
29. Daopin S, Piez KA, Ogawa Y, Davies DR (1992) Crystal structure of transforming growth factor-beta 2: an unusual fold for the superfamily. Science 257:369–373
30. Schlunegger MP, Grutter MG (1992) An unusual feature revealed by the crystal structure at 2.2 Å resolution of human transforming growth factor-beta 2. Nature 358:430–434
31. Vitt US, Hsu SY, Hsueh JW (2001) Evolution and classification of cystine knot-containing hormones and related extracellular signaling molecules. Mol Endocrinol 15:681–694
32. McDonald NQ, Hendrickson WA (1993) A structural superfamily of growth factors containing a cystine knot motif. Cell 73:421–424
33. Sun PD, Davies DR (1995) The cystine-knot growth-factor superfamily. Annu Rev Biophys Biomol Struct 24:269–291
34. Wu H, Lustbader JW, Liu Y, Canfield RE, Hendrickson WA (1994) Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein. Structure 2:545–558
35. Lapthorn AJ, Harris DC, Littlejohn A, Lustbader JW, Canfield RE, Machin KJ, Morgan FJ, Isaacs NW (1994) Crystal structure of human chorionic gonadotropin. Nature 369:455–461
36. Lin LF, Doherty DH, Lile JD, Bektesh S, Collins F (1993) GDNF a glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. Science 260:1130–1132
37. Wozney JM, Rosen V (1998) Bone morphogenetic protein and bone morphogenetic protein gene family in bone formation and repair. Clin Orthop 346:26–37
38. Griffith DL, Keck PC, Sampath TK, Rueger DC, Carlson WD (1996) Three-dimensional structure of recombinant human osteogenic protein 1: structural paradigm for the transforming growth factor beta superfamily. Proc Natl Acad Sci USA 93:878–883
39. Katsumi A, Tuley EA, Bodo I, Sadler JE (2000) Localization of disulfide bonds in the cystine knot domain of human von Willebrand factor. J Biol Chem 275:25585–25594
40. Meitinger T, Meindl A, Bork P, Rost B, Sandler C, Haasemann M, Murken J (1993) Molecular modeling of the Norrie disease protein predicts a cystine knot growth factor tertiary structure. Nat Genet 5:376–380
41. Stanley E, Biben C, Kotecha S, Fabri L, Tajbakhsh S, Wang CC, Hatzistavrou T, Roberts B, Drinkwater C, Lah M, Buckingham M, Hilton D, Nash A, Mohun T, Harvey RP (1998) DAN is a secreted glycoprotein related to *Xenopus cerberus*. Mech Dev 77:173–184

42. Tamaoki H, Kobayashi Y, Nishimura S, Ohkubo T, Kyogoku Y, Nakajima K, Kumagaye S, Kimura T, Sakakibara S (1991) Solution conformation of endothelin determined by means of 1H-NMR spectroscopy and distance geometry calculations. Protein Eng 4:506–518
43. Kobayashi Y, Takashima H, Tamaoki H, Kyogoku Y, Lambert P, Kuroda H, Chino N, Watanabe TX, Kimura T, Sakakibara S (1991) The cysteine-stabilized alpha-helix: a common structural motif of ion-channel blocking neurotoxic peptides. Biopolymers 31:1213–1220
44. Tamaoki H, Miura R, Kusunoki M, Kyogoku Y, Kobayashi Y, Moroder L (1998) Folding motifs induced and stabilized by distinct cysteine frameworks. Protein Eng 11:649–659
45. Isaacs NW (1995) Cystine knots. Curr Opin Struct Biol 5:391–395
46. Prestrelski SJ, Arakawa T, Duker K, Kenney WC, Narhi LO (1994) The conformational stability of a non-covalent dimer of a platelet-derived growth factor-B mutant lacking the two cysteines involved in interchain disulfide bonds. Int J Pept Protein Res 44:357–363
47. Hui JO, Woo G, Chow DT, Katta V, Osslund T, Haniu M (1999) The intermolecular disulfide bridge of human glial cell line-derived neurotrophic factor: its selective reduction and biological activity of the modified protein. J Protein Chem 18:585–593
48. Bell SL, Xu G, Forstner JF (2001) Role of the cystine-knot motif at the c-terminus of rat mucin protein MUC2 in dimer formation and secretion. Biochem J 357:203–209
49. Scheufler C, Sebald W, Hulsmeyer M (1999) Crystal structure of human bone morphogenetic protein-2 at 2.7 Å resolution. J Mol Biol 287:103–115
50. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) Expasy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788
51. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
52. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
53. Notredame C, Higgins D, Heringa J (2000) T-coffee: a novel method for multiple sequence alignments. J Mol Biol 302:205–217
54. Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
55. Schwede T, Kopp J, Guex N, Peitsch MC (2003) Swiss-model: an automated protein homology-modeling server. Nucleic Acids Res 31:3381–3385
56. Guex N, Peitsch MC (1997) Swiss-model and the Swiss-pdbviewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723
57. Van Gunsteren WF, Billeter SR, Eising AA, Hunenberger PH, Krueger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulations: the GROMOS96 manual and user guide. Hochschulverlag, Zurich
58. Melo F, Feytmans E (1998) Assessing protein structures with a non-local atomic interaction energy. J Mol Biol 277:1141–1152
59. Laskowski RA, Chistyakov VV, Thornton JM (2005) Pdbsum more: New summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 33:D266–D268
60. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381:272
61. Eisenberg D, Luthy R, Bowie JU (1997) Verify3d: Assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404
62. Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. J Mol Graph 14:33–38
63. Holm L, Park J (2000) Dalilite workbench for protein structure comparison. Bioinformatics 16:566–567
64. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
65. Harris DC, Machin KJ, Evin GM, Morgan FJ, Isaacs NW (1989) Preliminary X-ray diffraction analysis of human chorionic gonadotropin. J Biol Chem 262:6705–6706
66. Pierce J, Parsons T (1981) Glycoprotein hormones: structure and function. Annu Rev Biochem 50:465–495
67. Murray-Rust J, McDonald NQ, Blundell TL, Hosang M, Oefner C, Winkler FK, Bradshaw RA (1993) Topological similarities in the TGF-beta 2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. Structure 1:153–159
68. Butler SA, Laidler P, Porter JR, Kicman AT, Chard T, Cowan DA, Iles RK (1999) The beta-subunit of human chorionic gonadotrophin exists as a homodimer. J Mol Endocrinol 22:185–192
69. McDonald NQ, Lapatto R, Rust JM, Gunning J, Wlodawer A, Blundell TL (1991) New protein fold revealed by a 2.3-Å resolution crystal structure of nerve growth factor. Nature 354:411–414
70. Oefner C, D'Arcy A, Winkler FK, Eggimann B, Hosang M (1992) Crystal structure of human platelet-derived growth factor BB. EMBO J 11:3921–3926
71. Kim YS, Gum JR Jr (1995) Diversity of mucin genes, structure, function and expression. Gastroenterology 109:999–1001
72. Axelsson MA, Asker N, Hansson GC (1998) O-Glycosylated MUC2 monomer and dimer from LS 174T cells are water-soluble, whereas larger MUC2 species formed early during biosynthesis are insoluble and contain non-reducible intermolecular bonds. J Biol Chem 273:18864–18870
73. Perez-Villar J, Eckhardt AE, DeLuca A, Hill RL (1998) Porcine submaxillary mucins forms disulfide-linked multimers through its amino acid-terminal D-domains. J Biol Chem 273:14442–14449
74. Lidell ME, Moncada DM, Chadee K, Hansson GC (2006) *Entamoeba histolytica* cysteine proteases cleave the MUC2 mucin in its C-terminal domain and dissolve the protective mucus gel. Proc Natl Acad Sci USA 103:9298–9303