The Role of FCGBP in Mucus

Structure, Processing and Function

Erik Ehrencrona

Department of Medical Biochemistry and Cell Biology
Institute of Biomedicine

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2021

Cover illustration: FCGBP in human ileum by Erik Ehrencrona

The Role of FCGBP in Mucus: Structure, Processing and Function

erik.ehrencrona@medkem.gu.se

ISBN 978-91-8009-468-9 (PRINT)
ISBN 978-91-8009-469-6 (PDF)
http://hdl.handle.net/2077/69314

Printed in Borås, Sweden 2021
Printed by Stema Specialtryck AB

ABSTRACT

In this thesis, a bottom-up approach was used to study the IgG Fc-binding protein (FCGBP), which next to the MUC2 mucin is the second main component of secreted mucus in small intestine and colon. FCGBP is also found in airways during inflammatory conditions with static mucus. Although discovered as an IgG sequester, this function was not reproducible here using purified proteins. FCGBP includes many von Willebrand D domains (vWDs) with most having a GDPH (Gly-Asp-Pro-His) motif. Theoretically, hydrolysis of the DP peptide bonds should result in reactive Asp-anhydrides driving covalent crosslinking between FCGBP and MUC2. Using mass spectrometry (MS), recombinant proteins, electrophoresis and Western blot, we found that all motifs were cleaved but FCGBP remained intact as consecutive fragments were tethered by single disulphide bonds. Label-free MS quantification of proteins in murine mucus showed that Muc2 and Fcgbp are mostly not covalently bound, and *in silico* structural predictions further argued against such interactions, with these Asp-anhydrides being inaccessible for their suggested MUC2 substrates. Recombinant proteins were purified, analysed and used for generation of FCGBP antisera. The murine Fcgbp is smaller but highly similar to the human orthologue, making it ideal for functional and structural studies. Microscopy was used to study live and fixed tissue from mouse colon and airways to investigate its physiological role. Recombinant proteins formed C-terminal cysteine dimers with cryogenic electron microscopy (cryo-EM) showing a spring feather-like quaternary structure. Even larger linear structures were detected in cryo-EM micrographs, and electrophoresis showed large complexes in mucus. Immunohistochemistry (IHC) also revealed elongated ultrastructures in healthy intestine and airways of a murine chronic obstructive pulmonary disease (COPD) model. Results indicated less attached mucus in airways of *Fcgbp*$^{-/-}$ mice, and further a mucus expansion phenotype in colon. An N-terminal sequence linked to helical gliding was studied and alignments revealed that the murine sequence had partially been genetically lost. The remaining N-terminal sequence shared between human and mouse was found to be repeated prior to every vWD in the FCGBP sequence. In summary, these results indicate a role for FCGBP in mucus structure and attachment.

Keywords: Mucus, FCGBP, IgG, IBD, COPD, Mucins, Disulphide, GDPH

POPULÄRVETENSSKAPLIG SAMMANFATTNING

FCGBP (IgG Fc-binding protein) hittas i slemmet som hydrerar och skyddar tarmens innandöme. Avhandlingen syftade till att etablera grundläggande förståelse för detta protein genom kartläggning av struktur, interaktioner och funktion. Vidare, att utveckla metoder och verktyg för framtida studier. Proteinet namngavs tidigare efter data där bindning till IgG påvisats. Detta motbevisas här genom att undersöka interaktioner mellan upprenade proteiner. Projektet har genererat flera specifika antikroppar, rekombinanta proteiner och lämpliga djurmodeller för studiens genomförande. I syfte att evaluera musmodellens lämplighet har skillnader mellan FCGBP hos människa och mus kartlagts och verifierats med molekylärbiologiska metoder. Proteinet är uppbyggt av von Willebrand D (vWD) domäner som beskrivs som viktiga för proteininteraktioner. Aminosyrasekvensen GDPH (Glycin-Aspartat-Prolin-Histidin), som finns på 11 ställen i humant FCGBP och på 5 ställen i den kortare musvarianten, är fullständigt klyvd vilket teoretiskt genererar reaktiva grupper som kan binda starkt till MUC2, huvudkomponenten i slemmet. Modellering och masspektrometri visar dock att grupperna är gömda inuti proteinet och kan spjälkas av vatten. Detta överensstämmer med att inga proteiner starkt bundna till FCGBP kunde identifieras i mukus utöver en liten fraktion som binder till MUC2, sannolikt inte är fysiologiskt relevant. Data från olika metoder visar på kompakt proteinstruktur med disulfidmedierad C-terminal dimer där flera vWD domäner bildar ett spiralfjäderliknande komplex. Icke-kovalenta interaktioner påvisades som medförde ännu större komplex, där funktionen sannolikt är att bidra till slemmets organisation och mekaniska egenskaper. Genom histologi visades uttryck i alla tarmens bägarceller och att FCGBP utsöndras som täta strukturer vilka organiserar slemlagrets inre skikt och sannolikt resten av mukustäcket. Försök med levande vävnad visar att avsaknad av FCGBP inte gör slemmet mer genomsläpplig för partiklar av bakteriestorlek. Infärgning visade dock ett mindre strukturerat slemlager och även en förändrad slemtillväxt i tjocktarmen. FCGBP är normalt inte uttryckt i lunga men uppregleras exempelvis vid kronisk obstruktiv lungsjukdom (KOL) där slemmet blir segare och liknar det som hittas i tarm. Avsaknad av FCGBP i djurmodell för KOL resulterade i mindre fastsittande slem. Immunfärgningar visade täta nätverksliknande strukturer som kapslar in och håller kvar mukus längs epitelytan. Sammantaget avslöjas dess funktion som en strukturellt viktig komponent i slem för att vid behov bilda ett skyddande mukustäcke på kroppens slemhinnor.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. **Ehrencrona, E**. van der Post, S. Gallego, P. Recktenwald, C. Rodriguez-Pineiro, A.M. Garcia-Bonete, MJ. Trillo-Muyo, S. Bäckström, M. Hansson, GC. Johanson, MEV. **The IgG Fc-binding protein FCGBP is secreted with all GDPH sequences cleaved, but maintained by inter-fragment disulfide Bonds**
Journal of Biochemistry **2021;** 293(1):100871.

II. Fakih, D. **Ehrencrona, E**. Martinez-Abad, B. Arike, L., Ermund, A. Trillo-Muyo, S. Gallego, P. Johansson, M.E.V. and Hansson, G.C. **The FCGBP Protein Induced at Lung Disease Anchors the Mucus Layer to the Tracheobronchial Surface**

Manuscript.

III. **Ehrencrona, E***. Gallego, P*. Garcia-Bonete, M.J. Trillo-Muyo, S. van der Post, S.V.P. Recktenwald, C.V., Rodriguez-pineiro, A.M. Hansson, G.C. and Johansson, M.E.V. **The FCGBP Structure Reveals a Convoluted C-terminal Dimer Stabilised by Cysteine Bonds**

Manuscript. * Equal contribution

IV. **Ehrencrona, E.** Svensson, F. Gallego, P. Garcia-Bonete, M.J. Martinez Abad, B. Hansson, G.C. Johansson, M.E.V. **Functional Analyses of FCGBP and its Role in Organisation of the Colonic Mucus Barrier**

Manuscript.

CONTENT

ABBREVIATIONS

GI tract	Gastrointestinal tract
FCGBP	IgG Fc binding protein
IgG	Immunoglobulin G
MUC2	Mucin 2
MUC5AC	Mucin 5AC
MUC5B	Mucin 5B
Mr	Molecular mass
AMP	Anti-microbial protein or peptide
vWF	von Willeband factor
vWD	von Willeband factor D domain
vWA	von Willeband factor A domain
vWC	von Willeband factor C domain
CTCK	Cysteine knot
C8	Domain with 8 conserved cysteines
TIL	Trypsin inhibitor-like domain
CysD	Cysteine-rich domain
EGF	Epidermal growth factor
TSP1	Thrombospondin 1
EMI	Elastin microfibril interfacer domain
LDLRA	Low density lipoprotein receptor A domain
FA58C	Coagulation factor 58 C-terminal domain
PTPmu	Receptor protein tyrosine phosphatase Mu
MAM	Meprin A-5 protein and PTPmu
FCGBP_N	FCGBP N-terminal sequence
HCl	Hydrochloric acid
IML	Inner mucus layer
OML	Outer mucus layer
DC	Distal colon
PC	Proximal colon
Si8	distal 8^{th} portion of small intestine (Ileum)
ASL	Air surface liquid
TECTα	Tectorin alpha
SSPO	SCO spondin
SCO	Sub commissural organs
ZAN	Zonadhesin
OTOG	Otogelin
DMBT1	Deleted in malignant brain tumors 1
CLCA1	Chloride channel accessory 1
ADAMTS-13	A disintegrin and metalloproteinase with a thrombospondin type 1 motif, member 13
MEP1	Meprin 1
MMP	Matrix metalloproteinase
ZG16	Zymogen granule protein 16
AGR2	Anterior gradient protein 2

TFF3	Trefoil factor 3
NEO1	Neogenin-1
PIGR	Polymeric immunoglobulin receptor
IL8	Interleukin 8
TNFα	Tumour necrosis factor alpha
PTS domain	Proline-threonine-serine domain
GDPH motif	Glycine-aspartate-proline-histidine motif
vWF	von Willebrand factor
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
BN PAGE	Blue native polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SAXS	Small angle x-ray scattering
Cryo-EM	Cryogenic electron microscopy
MS	Mass spectrometry
GuHCl	Guanidium hydrochloride
COPD	Chronic obstructive pulmonary disease
CF	Cystic fibrosis
BALF	Bronchoalveolar lavage fluid
CRC	Colorectal cancer
TC	Thyroid carcinoma
HNSCC	Head and neck squamous cell cancer
MAMP	Microbe-associated molecular pattern
PAMP	Pathogen-associated molecular pattern
SNP	Single nucleotide polymorphism

1 INTRODUCTION

1.1 Mucus is the first line of defense

In order to understand the physiological function and composition of mucus, it may be beneficial to discuss the environmental challenges needed to be tackled on mucosal surfaces, and how similar issues are dealt with in other places and evolutionary contexts. In addition to being subjected to mechanical forces that can be injurious, humans are constantly in contact with microorganisms that can as well be harmful. As an adaptation, evolution has granted different strategies to protect the integrity of mucosal surfaces, with examples including physical barriers, an innate and adaptive immune system, along with symbiotic relationships with certain microorganisms. The mucosal surface of the intestine is a great example. It is lined only by a single layer of epithelial cells, yet the luminal content is a source of intense mechanical stress and a home to a lush microbiota. Without a more sophisticated shielding, humans would likely succumb due to intestinal hemorrhages or invasive infections caused by intestinal pathogens.

The lower GI tract is an incubator of life, containing a warm and nutritious microenvironment that constantly promotes bacterial growth. Unsurprisingly, the large intestine houses a highly concentrated and diverse microbial community not found anywhere else in the human body [1,2]. To protect and maintain homeostasis at the mucosal surfaces, a highly complex and compositionally dynamic film of glycosylated biopolymers is secreted by a specialised type of epithelial cell called goblet cells (GCs) [3,4]. This secreted material is what is known as mucus, and its primary constituents are the large glycoproteins called mucins. Mucins assemble into networks resembling a sieve where larger structures are trapped [5-8]. They have massive *O*-linked glycans stretching from the side chains of the amino acids threonine and serine in the PTS (Pro-Thr-Ser) domains. These glycans together with the many disulphide bonds that crosslink the protein backbone of the mucins, are thought to protect the mucinous networks against proteases and glycosidases introduced by the host or colonising bacteria [5,9]. Furthermore, *O*-glycans cause the mucins to resemble bottlebrushes. These glycans contain many acidic structures such as sulphate and sialic acid groups, causing them to become highly water absorbant, subsequently resulting in mucus having the properties of a hydrogel. The charges of these glycans likely keep the mucins stretched out as they are inherently repulsive. The secreted mucins are also called gel-forming mucins [5,10-12]. There is also a variety of membrane bound mucins found

throughout the GI tract [5]. These likely contribute to epithelial protection but do not extensively contribute to the secreted mucus barrier.

Mucus structure and properties differs in various locations such as the airways, urogenital and GI tract. However, the way the barrier function is maintained is essentially the same; the surface is kept hydrated, lubricated, and unwanted structures or particles are entrapped and cleared out. This could be bacteria in the colon, or dust in the lungs. The complex and not always well studied mechanism of mucus clearance is used to control the distance between these particles and the mucosa. In general, while bacteria may stick to the mucus, smaller molecules should diffuse through the barrier. These systems may falter, resulting in disease [5,13]. Upon binding the unwanted particle, it is also important to rid these by renewing and removing old mucus, otherwise the bacterial enzymes may succeed in degrading both protein and glycan components of the mucin framework, thereby reaching the epithelial cells.

Even though the epidermis of the skin deals with a resident microbiota, the barrier function is executed in a very different style. In a hypothetical scenario were the skin and intestinal barrier systems were swapped, the result would be a disaster. On one hand, in a non-aqueous or low humidity environment, a terrestrial animal normally having a mucus-coated surface would have problems conserving water and keeping the surface hydrated. The tissue would likely crack, leading to infection and possibly death. On the other hand, the presence of an epidermal-like barrier in the GI tract would lead to starvation. Nutrients would not be properly absorbed as the mammalian epidermis is a restrictive diffusion barrier. The epidermis has a thick layer of anucleated cells and a high content of lipids, creating a mechanically robust diffusion barrier, allowing for storage of water and controlling homeostasis in a dry environment [14,15]. However, it is likely that not only epidermal thickness is important for barrier function, but also renewal and shedding of the sheets of skin, so that damaged tissue may be removed and unwanted microorganisms will struggle to get a permanent footing on the surface.

In the small intestine, the movement of fluid, electrolytes and nutrients such as monosaccharides, amino acids and fatty acids, is essential for the intestinal physiology and function. The thin intestinal epithelium is therefore shielded with a loose hydrous mucus coat that enables diffusion [16]. Here, GCs differentiate from the stem cells in the crypts and move upwards renewing the tissue surface every few days [17]. Generally, GCs secrete mucus at a baseline rate but can also be stimulated to enhance mucus release [18]. The colonic mucus barrier is slightly different, having an outer layer housing bacteria, and an inner layer that is not normally accessible to colonisation [8]. As the airways are a

location for gas exchange, they must also have a barrier that does not impair diffusion. Illustrations showing the first barrier for the epidermis [14,15], oropharynx [19], airways [13,20] and colon [8,13] are shown in **figure 1** below.

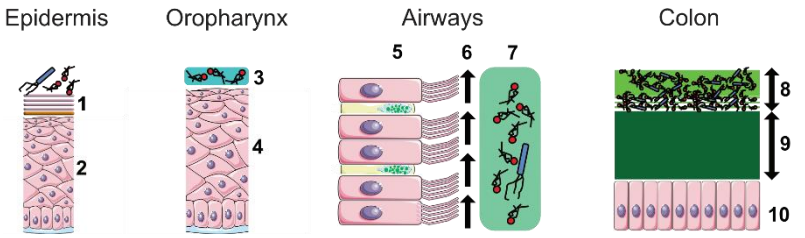

Figure 1. Histological illustrations of first barriers in different human anatomical locations. For the epidermis, the stratum corneum (1) is an important barrier for mechanical protection and water consolidation, in addition to the several layers of epithelial squamous cells (2). In the oropharynx, the multi-layered squamous cells (4), lacking a stratum corneum on top, are lubricated and partly protected by saliva (3). The trachea and bronchi are covered with a pseudostratified epithelium (5) where columnar cells have stereocilia (6) that transport away liquid and mucus (7). The colon has an inner (9) and and outer (8) mucus layer protecting a single layer of epith elial cells (10).

1.2 Physiology, histology and mucus composition in the upper and lower GI tract

There are 21 proteins listed within the human family of mucins. Here, the main gel-forming mucins are MUC2, MUC5AC, MUC5B and MUC6. There is also a host of transmembrane mucins, with examples including MUC1 and MUC17 [5]. While MUC5B is the predominant secreted mucin in the oral cavity and airways [21], MUC5AC is the prime component of the gastric mucus layer [22]. With some exceptions in muscle layer configuration, the traditional histological structure remains the same throughout the GI tract. However, due to the tissue-like properties of mucus, in some parts of the tract it may be considered an additional histological layer. As seen from a cross-sectional image of the tube, the GI tract is composed of a set of layers (**Fig. 2,** Colonic section as an example on the next page) [23].

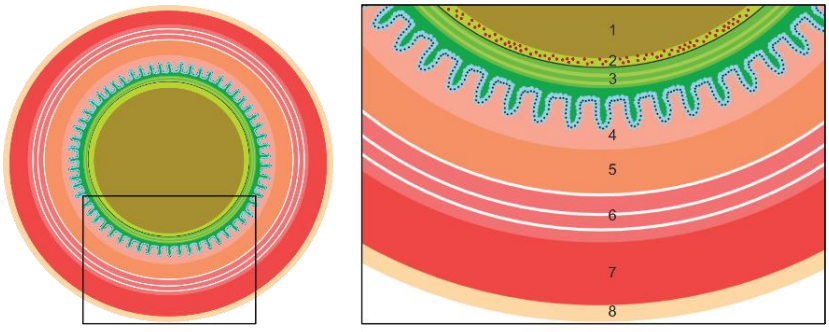

Figure 2. The histological layers of the distal colon. In the center of the lumen is the fecal pellet (1), covered also by the mucosa-lining mucus layers (2-3) that varies in composition throughout the GI tract. The mucosa (4), is covered by varying types of epithelial cells. Next, there is the submucosa (5), surrounded by the circular muscularis interna (6) and longitudinal muscularis externa (7). The outermost layers is the serosa (8), and layers 4-9 are present in all parts of the alimentary tract.

The GI tract is divided into two main sections; the upper and lower part. Whereas the upper GI tract stretches from the tongue to the end of the ileum [24], the latter reaches from the caecum to the anus [25]. At the beginning of the upper GI tract, the oropharynx and esophageal region have many layers of squamous cells protecting the surface. However, unlike the epidermis, these lack anucleated corneocytes [19]. Instead, saliva is complementary for the barrier function. In a review from 2008 [26], de Almeida *et Al.* presents an overview of the general composition and physiology of saliva. More than 90 % of this liquid is secreted from the sublingual, submandibular and parotid glands. Although it is less gel-like than mucus, saliva contains mucins and mucus components in addition to antimicrobial proteins (AMPs) and the digestive enzyme amylase. In addition to protection of the teeth, the water-retaining effect of mucins is believed to be important in maintaining lubrication. As in other locations, both the innate and adaptive immune system work in tandem to detect and react to pathogen-associated molecular patterns (PAMPs) and microbe-associated molecular patterns (MAMPs) in the mucus secretions. The oral cavity is the front door where all new microbes are introduced to the system. Therefore, in addition to immunoglobulins, a diverse set of AMPs are needed to prevent infection and modulate the microbiota through bactericidal and bacteriostatic effects. The protein deleted in malignant brain tumour 1 (DMBT1) is an agglutinin found in saliva [26] that has a series of scavenger receptors embedded within its sequence. It is highly versatile and has been shown to interact with a series of different structures from both viruses and bacteria. DMBT1 can also bind to secreted proteins in mucus, including MUC5B and IgA [27]. It is found also in the airways [28] and throughout the rest of the GI tract [28,29].

In the stomach, the parenchyma needs protection not only from mechanical forces, but also from a hostile acidic environment generated by the secretion of hydrochloric acid (HCl) from parietal cells deep within the gastric glands [30]. A review by Martinsen *et al.* from 2019 [31] details the composition and function of gastric juice. The stomach has an alternative approach to handling microbial content of the lumen, working as an effective sterilisation chamber [31] with a luminal pH of 2 [30], in addition to providing the first digestive enzymes targeting lipids and proteins. The low pH is further needed to denaturate the macromolecules in food [31]. The bolus is grinded to pieces by peristaltic movements as the ventricle is sealed from both ends by powerful sphincters [32]. Here, the wall muscle layers differs from the rest of the GI tract, having an additional oblique muscle layer that enables the grinding movement [33]. How the mucus layer is important in the protection against HCl in the stomach is summarised by Philipsson in a review from 2004 [30]. The GCs, here called foveolar cells, produce great amounts of bicarbonate and the main gel-forming mucin MUC5AC, in addition to the MUC6 mucin produced by cells inside the glands. Gastric mucus is considered to have two mucus layers, with the inner being firmly attached to the epithelium. The inner layer forms a pH gradient, being most acidic close the luminal content. The gradient is thought to be formed by bicarbonate secretion from underneath the mucus and acid from the lumen gradually mixing with the mucus as a diffusion barrier [30]. Pancreatic juice and bile acid is introduced to the chyme upon transition to the duodenum, introducing a myriad of enzymes and salts that are needed to digest the macromolecules of the chyme and emulsify the lipid content [34,35]. One of the most important functions is to neutralise the acidic content from the stomach, mainly through the secretion of bicarbonate [34].

In the rest of the GI tract, the main gel-forming mucin is the MUC2 [29]. Recombinant MUC2 forms N-terminal dimers through the von Willebrand Factor D3 domain (vWD3) [7] and C-terminal dimers between cysteine knots (CTCKs) [6]. Consequently, the interactions grant the formation of a polymeric network that constitutes the main framework within the mucus barrier of the small intestine and colon [13]. MUC5B and MUC5AC have a similar domain architecture (**Fig. 3**). The importance of MUC2 in epithelial protection is emphasised by studies of *Muc2*$^{-/-}$ mice which lack a normal mucus layer in the colon. These mice develop severe colitis and in some cases cancer [8,36,37]. MUC2 is densely packed in the GC vesicles, with a sudden drop in Ca^{2+}-levels upon secretion being thought to be a driver of mucus expansion [13,38]. Next to MUC2, IgG Fc- binding protein (FCGBP) is the second main secreted core protein in the mucus of the large and small intestine [29]. It was initially described to be sequestering IgG [39-42], but speculation on its function have shifted recently, rather suggesting a structural role [43,44]. The amino acid sequence of

FCGBP shows many similarities with the family of mucins. However, it lacks the central regions rich in proline, serine and threonine, and could thereby be considered a naked mucin. As heavy *O*-glycosylation is suggested to reduce the flexibility of mucins [5], FCGBP will likely be a much less rigid molecule. Neither does FCGBP have CysDs, CTCKs, or vWCs that are found but not limited to MUC2. Here, the main parallel is the D-assemblies (**Fig. 3**), also found in the von Willebrand factor (vWF) protein that binds together collagen and blood proteins during the coagulation cascade [45]. Both C8 and TIL are domains rich in cysteines and therefore shielded against proteolytic digestion.

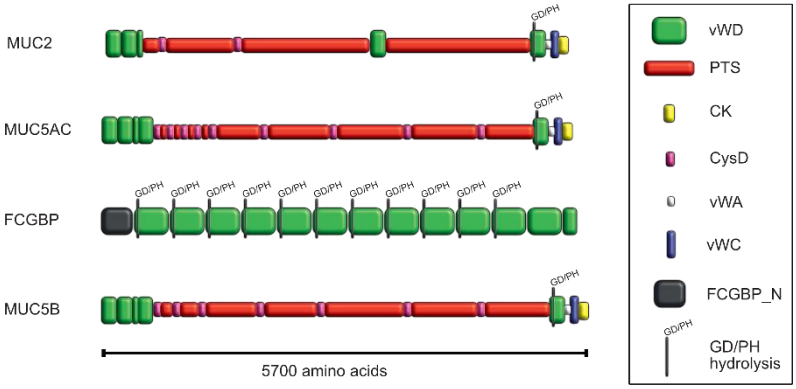

Figure 3. Illustrative Figures showing domain compositions of important mucus proteins. The conserved domains of the human mucins MUC2, MUC5AC and MUC5B are compared to those of FCGBP. The figure includes annotations for color-coded units representing each type of motif or domain.

Seen from an evolutionary perspective, the vWF naming scheme is considered slightly misleading as these domains originated from mucin-like proteins in early eukaryotic organisms that arose long before the vWF [5,46]. Currently, there are no *Fcgbp* $^{-/-}$ mice phenotypes described in literature.

Similar to gastric mucus, IHC experiments using Carnoy-fixed tissue has revealed an additional inner mucus layer (IML) that is firmly attached to the colonic epithelium [8]. It is proposed that structures above the IML, as seen in sections, is part locally generated and also consist of material gathered by the chyme during peristaltic transit through the colon, as mucus remnants still coat the surface of extracted fecal pellets [47-49]. Specific bacteria can thrive in the loose outer mucus layer (OML) [50]. In *ex vivo* experiments with fluorescent beads, only the interface area between the IML and OML is seen, as the rest has been removed during the preparation of the tissue [51] (**Fig. 4**).

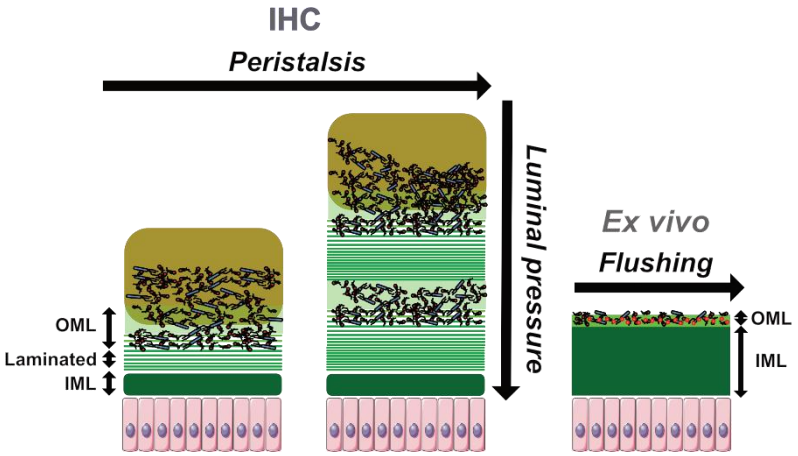

Figure 4. Illustrative figure showing mucus barrier organisation in fixed and live colonic tissue. During peristalsis, the colonic fecal pellet gathers layers of mucus as it is transported in a distal direction. This is visible in Carnoy fixed sections imaged *in vivo* using immunohistochemistry (IHC). In *ex vivo* imaging of colon using fluorescent beads and a confocal microscopy, preparation of the tissue removes most of the OML and other structures resting above it.

Having a thick compact mucus layer in the small intestine would likely negatively impact the diffusion and absorbance of nutrients due to the hydrostatic effect of the MUC2 *O*-glycans. Therefore, AMPs are used to a greater extent in order to protect the precious stem cells situated at the base of the crypt. In a review from 2011 [52], Bevins *et al.* did a detailed description of Paneth cells and AMPs found in the bottom of the small intestinal crypts, especially important for controlling the microbiota in the small intestine. Among AMPs produced by the host Paneth cells, defensins, lysozymes, phosholipases, lectins and ribonucleases are listed in the review. Collectively, these are able to shape how close to the epithelium microorganisms such as fungi, bacteria and viruses can reach [52]. In the the small intestine, there is only a loose mucus layer that can be easily detached *ex vivo*. Compared to that of colon, this mucus allows also for increased penetration by bacteria [13,50]. The *in vivo* model for mucus organisation and bacterial colonisation in the small intestine is shown in **figure 5**. The model is based on data from both living and fixed tissue visualised by microscopy [13].

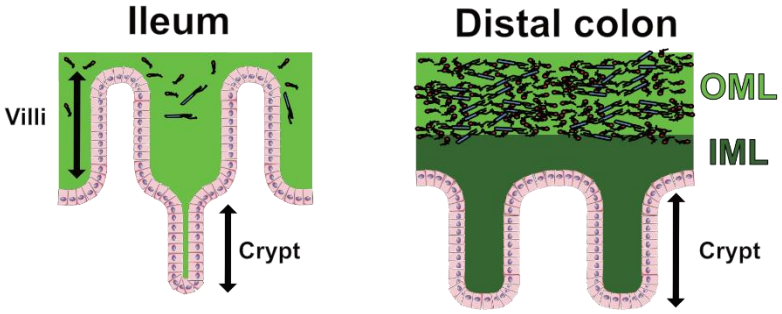

Figure 5. ***In vivo*** **models for mucus organisation in small intestine and colon.** The ileum has a loose mucus layer where bacteria can reach about a third of the length from the tip of the villi to bottom of the crypts. The distal colon has a distinct inner mucus layer (IML) firmly attached and not accessible to bacteria, and an outer mucus layer (OML), loosely attached and colonised by bacteria.

Clearance of mucus and chyme is also an important defense against microbes. In this context it has been shown that MEP1β is driving detachment of mucus in the small intestine by cleaving MUC2. Removal of calcium from MUC2 is also thought to be crucial for MEP1β-function [13,53]. In the colon, the CLCA1 metalloprotease is important as an additional modulator of mucus expansion, and thereby clearance [54]. Fluorescent lectins [54,55], bacteria-sized beads [51], and nucleic acid staining reveal the superstructural organisation of the mucus in an *ex vivo* setup. Strangely, the IML becomes penetrable to these beads in germ free (GF) mice, implicating the microbiota in shaping the mucus properties [13]. It has also been shown that a western style diet in mice, being low in fiber and high in fat, alters the microbial composition and increases the penetrability of the mucus, in addition to reducing the thickness of the IML. *Bifidobacterium* is an example of bacteria can directly affect mucus thickness and growth. The colonisation and expansion of these is promoted by dietary fibers [56].

It is important to stress that bacterial colonisation and digestion of mucin glycans is not a threat under normal conditions. The symbiotic relationship between the host and the gut microbiota is explored in depth in a review by Koropatkin *et al.* [9] The host benefits greatly from commensals such as the *Bacteroides thetaiotaomicron* digesting glycans in the OML. Short chain fatty acids (SCFAs) are metabolites of this process, which can in turn be absorbed back by the host, thereby recycling the mucus in a symbiotic fashion. Among the normal commensal bacteria, the Bacteroidetes, Firmicutes, Actinobacteria and Proteobacteria are the main phyla in adults and infants, although infants have certain adaptions correlating with the intake of maternal milk [9].

The protective mucus barrier can sometimes fail resulting in injury or infection by allowing bacteria to get closer to the epithelium. This triggers a massive reaction of mucus release from the upper crypt controlled by a specific sentinel

goblet cell [57]. In addition to mucins and FCGBP, the proteins anterior gradient protein 2 (AGR2), trefoil factor 3 (TFF3), zymogen granule protein 16 (ZG16) and transglutaminase 2 and 3 (TGM2/TGM3) are other important components of mucus in the small and large intestine [44]. ZG16 is a small 16 kDa protein that binds and aggregates Gram positive bacteria. *Zg16*$^{-/-}$ mice develop large deposits of visceral fat and 16S-sequencing shows bacterial infiltration of the spleen. It was therefore suggested that a dysfunctional barrier leads to chronic inflammation and weight gain, as possibly as a result of bacterial translocation [58]. AGR2 is thought to function as a protein disulphide isomerase with a specific role in biosynthesis of mucus proteins [59]. In addition, it has been shown to be secreted as a component of intestinal mucus [60]. TFF3 is a small protein believed to be important for tissue regeneration [61], protection against helminths [62], and suggested to be involved in stabilising mucus through covalent interactions with FCGBP and indirectly with mucins [43,63]. While *O*-glycans protect the protein core of MUC2, and disulphide bonds link the protein horizontally, the mucin framework is suggested to be stabilised vertically in the same way as the epidermal layer of the skin by isopeptide bonds catalysed by transglutamination [64,65].

1.3 Gastrointestinal mucus and the adaptive immune system

As part of the acquired immunity, mucosal plasma cells secrete dimeric immunoglobulin A (IgA) as the main antibody. The subject of IgA in mucosal immunity is well explained in a review by Brandtzaeg from 2011 [66]. It is translocated through epithelial cells to the intestinal lumen with the help of Polymeric IgG Receptor (PIGR). PIGR is a membrane protein which binds the IgA heavy chains. The complex undergoes cleavage along with bound IgA. This allows for the secretory component bound to IgA, forming secretory IgA, to be released into the mucus. The main function of secretory IgA is agglutination of bacteria to control the microbiota [66]. Pentameric IgM is not normally found in mucus of the small intestine and lower GI tract, or is at least it is secreted in very low amounts. However, in saliva it is secreted following the same pathway as IgA [66,67] In neonates with a premature adaptive immune system, IgA and smaller amounts of IgG is delivered to the GI tract through maternal milk [68]. This is especially important for IgA as it cannot pass through placenta to the fetus during development. With IgG being monomeric [69] , it would likely not be useful for agglutination of bacteria unless a mechanism of tethering IgG to mucus was available. FCGBP is indeed suggested to be such a sequesterer of IgG through binding to its many vWDs [41,42]. However, a link between FCGBP and IgA has never been detected. FCGBP has also been

suggested to be sequestering IgG in cervical mucus, trapping HIV particles [70]. The massive MUC16 mucin has also been suggested to act as a sequesterer of IgG in the female urogenital tract [71], but there is no other known potential Ig-binding site in intestinal mucus. However, IgG is likely secreted in very low concentrations in GI mucus, as it does not have a known dedicated translocation mechanism as seen with IgA and PIGR.

1.3.1 Mucus-related diseases in the GI tract

Although studies directly linking mucus to disease are scarce, it is widely regarded that an inadequate barrier function is one of the key drivers in the development of inflammatory bowel disease (IBD), including ulcerative colitis (UC) and Crohn's disease (CD). This inadequacy is suggested to arise from a more permeable mucus layer and an increased mucosal reaction and/or penetrability to bacteria (**Fig. 6**) [36,72-74], with increased levels of MAMPs such as LPS reaching the mucosa. Alternatively these factors acting in unison can result in parenchymal damage [74]. Long-term inflammation in UC can drive the development of colorectal cancer (CRC), mainly resulting in adenocarcinoma [75]. The link to mucus is clear as *Muc2*$^{-/-}$ mice with a disrupted mucus are more prone to develop cancer [37], and mucinous colorectal adenocarcinoma is known to show overexpression not only of MUC2, but also MUC5AC [76].

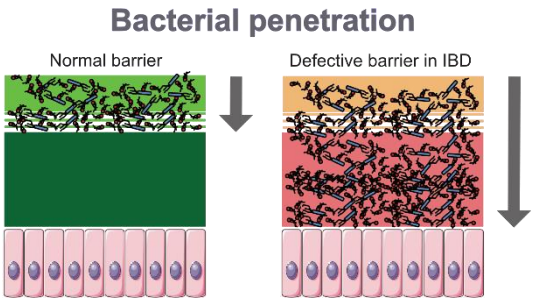

Figure 6. Relationship between colonic mucus and bacteria in health and IBD illustrated. Under normal conditions, bacterial colonization is limited to the OML. However, In IBD the barrier function of the IML is disrupted, allowing for bacteria to reach the surface of the epithelial cells.

General alterations of the proteomic profile [5,73] and mucin glycan composition [5,77,78] is associated with the disease. Just like *Muc2*$^{-/-}$ mice, the colonic IML of UC patients with active disease is penetrable to bacteria-sized fluorescent beads implicating alterations in mucus homeostasis [72]. In addition to being caused by an out of control immune responses due to penetrant antigens, UC is a complex disease with everything from ER stress to a dysbiosis and altered immune regulation being suggested to drive the condition [79]. Episodes of UC are limited to the colon, in contrast, CD can develop at virtually any place in

the GI tract. However, it is most commonly located to the terminal ileum. Unlike UC, the parenchymal damage in CD is not limited to the mucosa and submucosa. Here, all the histological layers can be compromised, and even fistulas can be formed between the intestine and nearby organs [80].

Cystic fibrosis (CF) is also related to GI mucus as it can lead to highly viscous attached plugs of mucus and fecal material, called meconium, forming in the intestine. This leads to blockage of fecal transit [81]. As *Cftr*$^{-/-}$ mice have firmly attached mucus in the small intestine, and this process cannot be reversed by recombinant MEP1β, but instead through the addition of bicarbonate, it is proposed that impaired calcium homeostasis is part of the pathology [53,82]. Calcium and low pH drive intracellular packing of MUC2 [7,38], and it was therefore suggested that mucin expansion by calcium chelation reveals the MEP1β cleavage sites [53,82]. The molecular details of the links between CF and the many mucus core proteins are otherwise scarcely studied. Mucus plugging can at least be associated with the *CLCA1* gene, as patients with the pAsn357Ser single nucleotide polymorphism (SNP) variant shows increased risk of plugging [83]. In concurrence with the SNP driven pathology, studies of a murine *Cftr*$^{-/-}$ model showed lower expression of the murine CLCA1 ortholog (Clca1, previously called Clca3) [84]. Transgenic *Cftr*$^{-/-}$ mice not being able to express Cftr, but overexpressing Clca1, in turn showed less damage to the villi and normalised mucus release. This effect on mucus retention was suggested to be partly mediated by its function as a metalloprotease [85], being able to aid in dissolving the accumulated mucus.

1.4 Lung

1.4.1 Airway mucus and clearance physiology in health and disease

Compared to the GI tract, the physiology pertaining to mucus is different in the lung under healthy conditions. The tracheal surface is lined by pseudostratified epithelium with embedded goblet cells which secrete mucus. Mucinous material is also secreted by submucosal glands. Under healthy circumstances, this glandular mucus mainly contains MUC5B. The glands are only found in the submucosa of the cervical part of murine trachea, and unlike human and pig, not at all present in the bronchi [5,13,20]. In human and pig, MUC5B forms bundles that extend horizontally in relation to the trachea [5,86,87], whereas the murine mucus has a mainly cloud-like appearance. Muc5ac is thought to be important for the formation of these clouds in mice [5,88]. Epithelial cells are apically covered by vertically projecting stereocilia continuously

beating overlaying material in a cephalic direction, moving it through the epiglottis into the esophagus where it swallowed together with the bolus. This mediates clearance of the mucus separated from epithelium by the ASL. The cilia help remove microbes stuck to the mucus by interacting directly with the mucin or components of the innate or adaptive immune system, and also eliminate particles such as dust inhaled from the environment [89]. It has been suggested that these MUC5B bundles or clouds are used to rake the surface, depending on the species (**Fig. 7**) [5].

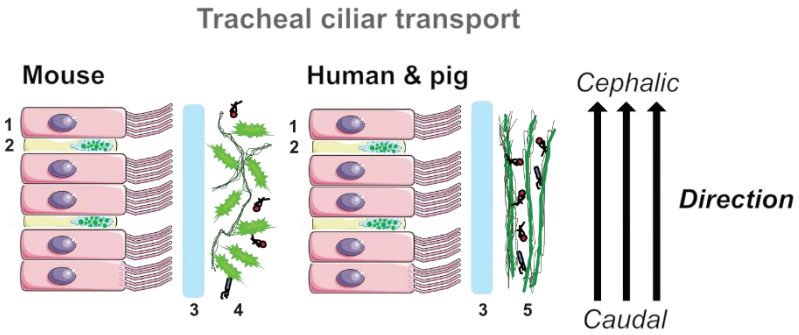

Figure 7. Illustrative Figures showing tracheal mucus transport in mouse, humans and pigs. The psuedostratified and ciliated epithelial cells (1) with embedded GCs (2) are shared histological features of human, pig and murine tracheas. As well the layer of air surface liquid (ASL) resting on top of the epithelium is shared (3). However, unique to the murine mucus organisation is cloud-like structures sweeping the surface (4) being transported in a cephalic direction, whereas human and pig airways have strands and bundles (5).

Chronic inflammatory conditions of the lung are most often associated with dysregulation or impairment of this mucus clearence system. After many years of chronic inflammation of the airways driven by tobacco smoke, a person may have developed COPD. Fundamental pathological traits of COPD are bronchitis, fibrosis and emphysema, making it difficult for the patient not only to inhale, but also disrupts the gas exchange in the alveoli [90]. In a review from 2017 [91], Barnes summarises the known molecular details underlying COPD. Toxic gasses trigger macrophages to release pro-inflammatory cytokines such as IL-8 and TNFα mainly through a reactive oxygen species (ROS)-induced pathway. Epithelial cells are as well affected by ROS, and further take part in secretion of cytokines. Altogether, this leads to a subsequent release of proteases such as elastase and matrix metalloproteinases (MMPs) by neutrophils and alveolar macrophages, causing mucus hypersecretion and airway damage, with the latter resulting in emphysema. As the epithelium is damaged, the ciliary function is also compromised. This results in impaired mucociliary clearance, with the patient having to cough in order to remove

mucus. Activation of latent TGF-β appears to be one of the more important pathways for induction of fibrosis, being linked mainly to epithelial damage [91]. Recent findings indicate that the mucus of COPD airways become more similar in organisation to that normally found in the colon (**Fig. 8**) [5,92].

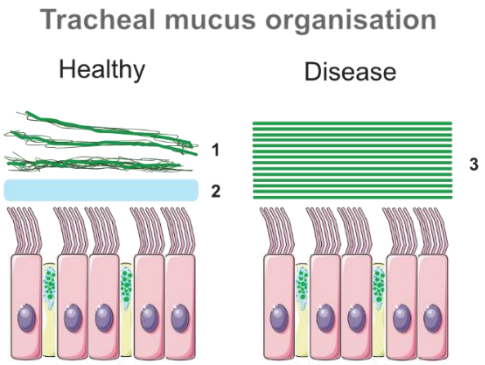

Figure 8. An illustraion of tracheal mucus organisation in health and under chronic inflammatory conditions. Under healthy conditions, the mucus bundles (1) are separated from the stereocilia by the ASL (2). In COPD, a stratified attached mucus layer (3) is formed, resembling that of the colon.

In COPD, levels of MUC5AC increase [92,93], and proteins normally associated with GI mucus appear at high concentrations, resulting in the formation of a firmly attached mucus layer. Examples of these proteins include CLCA1, FCGBP and TFF3 [92]. While MUC5B is produced in the submucosal glands, MUC5AC is produced by the goblet cells embedded within the pseudostratified epithelial surface [93]. It has therefore been suggested that similarly to the stomach, MUC5AC is used to mediate mucus attachment, here by providing an interface between the epithelial surface and MUC5B bundles that rake the epithelial surface [5].

Chronic inflammation, static mucus and parenchymal damage are traits particular also to CF [94]. Contemporary literature on CF is summarised in a review by Elborn CBE published in The Lancet in 2016 [95]. Mucus becomes highly static and the clearance system is impaired, generally believed to be mediated by increased mucus production and attachment to the epithelium. As well, further thickening of mucus by hyposecretion of water and thereby dehydration. Destruction of the lung is caused by chronic inflammation driven by microbes such as bacteria and fungi. The cystic fibrosis transmembrane receptor (CFTR) is a chloride channel important for the homeostasis of chloride and bicarbonate [95]. A disrupted secretion of bicarbonate has in turn been linked to a reduced pH [95,96], and the speculation that absence of $HCO3^-$ could actually be the main factor resulting in a firmly attached mucus layer [5,97].

Similarly, asthma shows hypersecretion and retention of mucus, however parenchymal damage is not a hallmark of the disease [98].

1.5 IgG Fc Gamma Binding Protein (FCGBP)

1.5.1 Structure, function and evolutionary relatives

The *FCGBP* gene is found on chromosome 19 in *Homo sapiens* and 7 in *Mus musculus*. First described by Kobayshi *et al.*, it was found in the GCs of the colon [39,40]. In addition to the larger airways, cystobiliar apparatus and cervix, IHC stainings revealed it to be expressed in submandibular glands and throughout the small intestine [40,42] . More recently, mass spectrometry (MS) analysis has showed that FCGBP can be found in the mucus of the small intestine and lower GI tract [29]. It has also been found in saliva [63], parts of the reproductive tract, including seminal plasma [99] and cervical mucus under normal conditions [100]. Peptides of FCGBP were also detected in the viscous cervical mucus plug generated during pregnancy [101]. In addition to being expressed with the mucins, CLCA1 and TFF3, it is usually expressed alongside DMBT1 [29,92]. MUC2, CLCA1 and FCGBP expression levels seem to be equally important in mucus as label free quantitative MS shows almost equimolar concentrations of the proteins expressed and secreted in mucus throughout the small intestine and lower GI tract [29]. Paralogues that are even more similar to FCGBP than mucins are the proteins Otogelin (OTOG) and α-Tectorin [46], both being components of the tectorial membrane of the inner ear. Thus, evolutionary related proteins are found in various extracellular matrices throughout the body. Mutations in the vWDs of α-Tectorin usually results in partial or complete hearing impairment, meaning that it is important for the structure and transduction of kinetic energy in the tectorial membrane [102]. In *Xenopus tropicalis*, the mucus coating of the tadpole epidermis contains Otogelin-like protein, Muc5E and possibly more than one version of FCGBP [103], depending on reliability of the available sequence data. The Otogelin-like protein of *Xenopus* is instead richly *O*-glycosylated and seems to function as the main mucin, with the suggested name MucXS [104]. The tadpoles have an adhesive organ called the cement gland that mainly secretes Muc5E, allowing the tadpoles to stick to surfaces. Due to similarities with the lung epithelium, as a result of the types of secreted proteins and function of the tissue as a gas exchange zone, *Xenopus tropicalis* is used as a model to study the lung mucus [103].

Another protein with domain composition similar to FCGBP is zonadhesin (ZAN), found in the acromion of spermatocytes, and believed to mediate binding to the zona pellucida during impregnation [105]. Furthermore, there is

also SCO spondin (SSPO), a fiber forming protein that is secreted from subcommissural organs (SCO) into the cerebrospinal fluid of the ventricles of the brain [106]. In *Danio rerio*, gene knockout of the SSPO orthologue leads to total absence of embryotic assembly of Reisner´s fiber in the central spinal canal, resulting in the body axis becoming bent and crooked during morphogenesis [107]. Of all these proteins, FCGBP contains the highest number of vWDs without interruption of any other type of domain. Excluding FCGBP orthologues, there are actually no other known proteins containing this many repeating vWDs [108]. Domain composition of FCGBP orthologues and paralogues is shown in **figure 3**. The X-ray crystal structure of von Willebrand factor D´D3 domain [109] and the MUC2 N-terminal cryo-EM structure [7] were recently published. These may prove to be useful tools for *in silico* predictions and aid other methods to solve the 3D-structure of FCGBP and related proteins. In general, other than protein-protein interactions, functions are not known for any of the FCGBP vWD domains. However, they have been shown to be involved in the biosynthesis and vesicle storage of the vWF [110,111] and assembly of MUC2 N-terminal polymers [7]. vWD-assemblies (D-assemblies), are with some exceptions, primarily constituted by a vWD domain, a cysteine rich domain (C8), trypsin inhibitor-like domain (TIL) and an E domain (E). Human FCGBP has 12 D-assemblies and a free von Willebrand D domain at the C-terminal end. Assemblies D3 to D11 are essentially 3 tandem repeats, or triplets, at the center of the human protein. Murine Fcgbp is truncated, having only one repeat, meaning a total 6 D-assemblies and a free C-terminal vWD domain (**Fig. 9**).

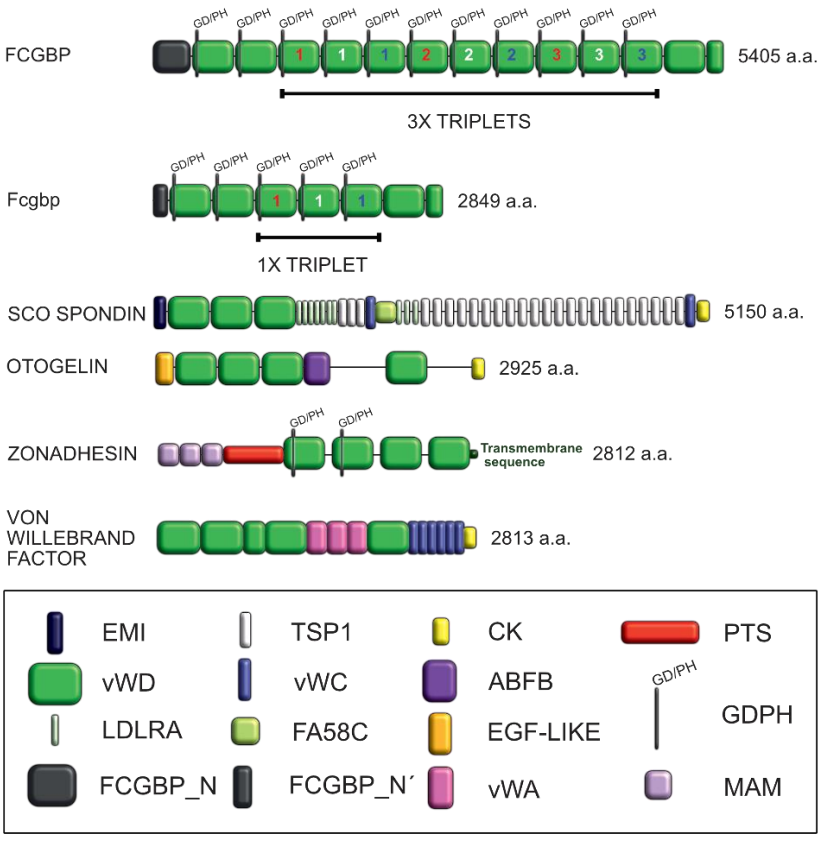

Figure 9. Domain compositions of FCGBP orthologues and paralogues. FCGBP domain structure in human and mouse compared to the paralogous proteins SCO spondin, otogelin and zonadhesin. The architecture of the von Willebrand factor is included for comparison. Color-coded motifs or domain structures are also included in the figure, along with annotation for each.

The domain composition of vWF and many of its interactions are summarised in a review by Haberichter in 2015 [112]. The first two vWDs together constitute a propeptide important for intracellular packing. D´D3 is site of covalent dimerisation and sequestering of factor VIII. This is followed by 3 vWAs, which together constitute a region that is susceptible to regulation by proteolysis and binding to endothelial collagen and platelets. The vWA#2 is cleaved by the ADAMTS-13 metalloprotease [113], but only when the protein is stretched by sheer stress, resulting in the target sequence being exposed, modulating the hemodynamics. The sheer-dependent A2-binding and cleavage is suggested to prevent clotting [114-116]. vWA#1 and #3 can bind collagen, and the former can also bind the platelet surface glycoprotein 1Iβ (GPIIβ) [117].

Interestingly, the link between vWA#1 and Platelet GPIβ is also shear dependent, where increased flow instead promotes clotting [113]. The vWAs are followed by the D4 assembly, later vWC#1-6, according the latest domain architecture consensus [117]. vWC#4, called vWC#1 in older references, mediates additional binding to the platelet surface glycoproteins GPIIβ and GPIIIα. Finally, at the C-terminus there is a cysteine knot [112] that results in dimerisation via 3 disulphide bonds [118]. As only interactions between vWD domains have been observed [7,109] it would be likely that FCGBP only interacts with itself in an intra- or intermolecular fashion or with other vWD containing molecules such as mucins. As calcium binding is important for mucin and vWF D-assembly binding and expansion [7,38,109], it is likely that a similar mechanism is used by FCGBP.

1.5.2 Current insight on GDPH motifs – Processing and function

In addition to being secreted together into mucus and having vWDs, a link between FCGBP and mucins, such as MUC2, MUC5AC and the membrane mucin MUC4, is the post-translational cleavage of vWDs [119-121]. In humans, MUC2 and MUC5AC have one self-cleaving GDPH (Gly-Asp-Pro-His) motif in the most C-terminal vWD. Cleavage of MUC2 is mediated by a low pH [120], which also drives but is not necessary for autocatalytic processing of MUC5AC [119]. MUC4 is as well GDPH cleaved, and as there are no cysteines crosslinking the domain, [121] the protein backbone is thought not to be covalently stable, allowing for the MUC4α peptide to dissociate from the β-peptide, resulting in its release from the cell membrane [122]. A similar mechanism has been detailed in MUC1, which is autocatalytically cleaved instead at a SEA domain, and not through a GDPH motif [123]. Here, as the cleaved fragments lack disulphide bond tethering, mechanical forces can detach MUC1 from the membrane. This is thought to function as a clearance mechanism used against penetrating bacteria, and as a means for sensing of mechanical stresses [124]. FCGBP carries a GDPH motif in all but the last 2 terminal vWDs, making a total of 11 motifs. This is a hallmark of FCGBP, with no other protein, other than orthologues, having more than 2 motifs. Given the large number of motifs, these are likely very important for structure and function of FCGBP.

Literature on GDPH motifs encompass many different areas of biology. For paralogues with vWDs more similar to those of FCGBP, ZAN has 2 GDPH motifs, whereas OTOG and SSPO have none. The motif is suggested to drive covalent crosslinking between FCGBP and other proteins through reactive Asp-anhydrides which are a result of the autocatalytic cleavage event. The described link between MUC2 and FCGBP was sensitive to neither a

chaotropic or reducing agents, suggesting a covalent isopeptide bond [44]. Thuveson *et al*., has shown that the Asp-anhydrides generated from GDPH cleavage could be used to form a stable link between the H3 heavy chain and the pre-alpha inhibitor and the chondroitin sulfate of Bikunin [125,126]. Similarly, for bacterial proteins, the FrpC toxin of *Neisseria meningitidis,* the ApxIVA of *Actinobacillus pleuropneumoniae* and the Nope1 protein of *Bradyrhizobium japonicum,* all have been shown to cleave through the GDPH motifs in the presence of calcium [127,128]. For Nope1, the cleavage took place under both acidic and alkaline conditions and was not affected by the addition of protease inhibitors [127]. In the case of FrpC and ApxIVA, it has been shown that the proteins assemble into covalently cross-linked multimers through isopeptide bonds between the Asp-anhydrides and lysine residues. It was also shown for FrpC that cleavage occurred in a wide pH range of 5.5 to 8.5 and was not affected by protease inhibitors [128]. High molecular mass (Mr) complexes that were not present upon mutation of the GDPH motif were also reported for Nope1, but these were not shown to be covalent in nature as they were only seen in Native-PAGE and not SDS-PAGE [127]. This could indicate the cleavage is important for correct conformation.

The repulsive guidance molecule (RGM) family of proteins induce a diverse set of effects upon regulatory signaling through interaction with Neogenin-1 (NEO1) at the cells surface [128,129], consequently, with mutations also leading to impairment of many different systems [129]. A published crystal structure of the NEO1-RGMB complex revealed that the GDPH motif constitutes a loop between beta strands in the vWD domain of RGMB, and that upon cleavage the aspartate was accessible on the surface, whereas the proline was hidden inside the vWD domain. Bell *et al.* predicted that half of the, at the time 14, documented mutations in RGMC that caused loss of function, leading to iron retention in Juvenile Hemochromatosis (JHH), were located in proximity to the GDPH motif, further stressing the functional importance of this motif [130]. Furthermore, the function of NEO1 is tightly linked to bone morphogenetic protein (BMP) signaling. NEO1 binds BMP, competing with BMP Endothelial cell precursor-derived regulators (BMP-ER) for binding the ligand. Therefore, NEO1 and BMP-ER together modulate BMP signalling. BMP-ER binds to the same BMP site as NEO1, and they together regulate iron metabolism [129]. Interestingly, BMP-ER has its own GDPH motif. Here, the GDPH motif is found in a C-terminal vWD domain, and upon cleavage, most of these two fragments are still held together by a single disulphide bridge. Lockhart-Cairns *et al.* stressed the difference between BMP-ER and RGM, where RGM was not secreted due to misfolding upon mutation of the GDPH motif, but this was not the case for BMP-ER [131].

A protein BLAST (Basic Local Alignment Search Tool) search of the GDPH motif in *Homo sapiens* showed that both intra and extracellular proteins have this motif (unpublished). A strange example in this search is the unconventional Myosin-15 (MYO15) where the GDPH is found in the MyTH4 1 tail region at position 2097-2100 (uniprot ID: Q9UKN7-1). Mutations of this protein have been shown to cause hearing loss. Mutations and deletions causing hearing loss without truncating the protein seem to be spread between amino acids 1253 and 3420. The described mutations closest to the motif can be found at position 2073, 2011, 2013 and 2014 [132]. However, it is unclear if the protein is even cleaved and what conditions could possibly drive cleavage. Current literature mentions no mutations associated with the motif. Furthermore, as the crystal structure of this protein has not been solved, not much can be said about the spatial localisation of the mutations in relation to the GDPH motif.

Another example of GDPH loss of function is the sushi domain containing 2 (SUSD2) protein which has been shown to be necessary for translocating Galectin-1 to the cell membrane [133], where it has a wide range of functions, involved in many different cellular functions [134]. Galectin-1 must be in a reduced state in order to exert its extracellular effect on T-cells [135]. Galectin-1 lacks a signal sequence and it could be speculated that it is not to going through the ER and Golgi because these compartments promote an oxidised state in contrast to the cytosol that is mainly a reductive [136]. The cleaved SUSD2 peptides were also tethered by a single disulphide bridge, with mutation of these cysteines inhibiting cleavage, and thereby surface transport. pH did not affect GDPH cleavage of SUSD2, instead, it was suggested to be cleaved by serine proteases [133].

All verified examples of cleaved GDPH motifs seem to be associated with proteins that reside in the extracellular environment. As Asp-anhydrides resulting from GDPH hydrolysis can react both with amino acids [128] and sugars [125], it was suggested that FCGBP was covalently bound to the MUC2 framework [44] through amide or ester bonds. The cleaved fragment of recombinant MUC5AC with a C-terminal aspartate showed reactivity with a primary amine through biotinylated ethylenediamine hydrobromide (B-EDA), indicating that the reactive group is actually preserved after cleavage [119]. However, a GAPH mutant negative control was not used. Hypothetically, this reactive group should be highly unstable and could theoretically react with water, thereby being hydrolysed. Therefore, binding to other molecules should happen early during biosynthesis and protein maturation as is the case for the pre-alpha inhibitor with H3[125] and FrpC toxin [128]. In addition, it has also been shown that FCGBP forms reducible high Mr complexes not only with MUC2

[44], but also TFF3 where heteromer formation is suggested to be mediated by disulphide bonds [43,63].

1.5.3 The FCGBP_N sequence and link to helical gliding bacteria

FCGBP in species ranging from vertebrates such as *Homo sapiens* to amphibians like *Xenopus tropicalis* contains a conserved N-terminal domain (FCGBP_N) described by phylogenetic and position-specific iterative basic local alignment search tool (PSI-BLAST) analysis by Lang *et al.* Most of this sequence cannot be found in *Mus musculus* or *Rattus norvegicus*, but it is found in closer relatives such as *Mesocricetus auratus* [108]. Currently, the PFAM database lists this region as IgG-binding, even though the initial publication on the subject shows that the D-assemblies mediate IgG sequestering [41]. Lang *et al.* showed that this sequence could be found early in evolution, long before the first antibodies emerged in cartilaginous fish [108,137]. The earliest known FCGBP_N-vWD-C8 composition emerged in the Ctenophores but TIL domains are not included until the emergence of the Chordata phylum. Therefore, mucins are considered older than FCGBP with complete D-assemblies already found in Ctenophora [108]. Furthermore, with the adaptive immune system being limited to vertebrates, it is further unlikely that FCGBP_N is involved in IgG sequestering as these sequences can also be found in several proteins produced in bacteria that utilise helical binding motility for movement, albeit without vWD, C8 or TIL domains. An example of this motility can be seen in *Myxococcus xhantus*, a widely studied biofilm-producing bacteria which possesses the FCGBP_N domain [108].

With the FCGBP and similar proteins are only found in eukaryotes [108], the link to prokaryotes through *Myxococcus xhantus* is interesting since researchers have compared them to multicellular organisms. The habitats and survival strategies are discussed by Muñoz-Dorado *et al.* in a review from 2016 [138]. The extracellular matrix (ECM), is considered a mesh-like network mainly composed of carbohydrates called exopolysaccharides (EPS), but there is also a protein component. In a biofilm, clusters of these bacteria form specialised regions that work symbiotically, much like cells in higher organisms [138]. Interactions with EPS and between cells is believed to be partly mediated by type 4 pili, and the ECM is also thought to be littered with protein-filled vesicles [138,139]. An organised system of movement, called social motility is coordinated by chemotaxis, with the kinetic pattern of the colony being compared to the rippling effect of water. However, when a single bacterium travels independently, called adventurous movement, a system of helical locomotion combined with focal adhesion, propels the bacteria forward

through rotational movement [138]. There are no known pathogenic or opportunistic strains of the myxobacteria mentioned in modern literature. However, they are found in the fecal material of some animal, mainly herbivores [140,141]. The FCGBP_N sequence is not linked to the biofilm formation itself. However, Lang *et al.* further noted that these domains were found together with conserved domain sequences associated with surface-bound proteins, including domain of unknown function 11 (DUF11) and laminin_G3 of *Algibacter lectus.* Interestingly, *Strongylocentrotus purpur* has an autoproteolytic motif between the FCGBP_N and a transmembrane domain suggesting that the FCGBP_N sequence could be released to the environment upon membrane incorporation. It was also found that the sequence exists in a protein containing a SprB domain. *Flavobacterium johnsoniae* uses a helical gliding motility system were the SprB forms filaments on the surface that mediate focal adhesion to surfaces [108,142]. The adoption the FCGBP_N sequence in higher organisms might be linked to shaping and controlling mucosal colonisation of myxococcales or similar types of bacteria encountered in soil-based food sources. This structure could be binding EPS found within the biofilm and tether it the protective mucus layers of the gut, resulting in the bacterial films being cleared out with the stool. Contrarily, it is also possible that this sequence emerged in higher organisms but was adopted by the lower organisms.

1.5.4 Links between FCGBP and disease

Currently function of FCGBP upregulation in disease is not well understood. As was previously presented, FCGBP is expressed at high levels during COPD where airway mucus function and morphology become more similar to colonic mucus, showing a firmly attached mucus layer [92]. In human bronchoalveolar lavage (BALF) from CFTR patients, there are increased levels of FCGBP but it is not contained within the 99 % confidence interval [143]. FCGBP is also one of the main constituents in the highly viscous mucus material found in the gallbladders of dogs with mucocele, which further highlights its importance for the rheological properties of the mucus [144]. In an article from 2014 [70], Schwartz suggested that copy-number variations is linked to FCGBP protection against human immunodeficiency virus 1 (HIV-1) in cervical mucosa of a specific group of seronegative HIV-exposed women, with the argument that additional IgG binding units in FCGBP enhances the IgG sequestering and immobilisation of HIV virus particles within the mucus network [70]. In relation to cancer, malignancies can be associated with both down and upregulation of FCGBP. Downregulation is observed in CRC [145] and thyroid cancer (TC) [146]. In contrast, ovarian cancer (OC) and head and neck squamous cell cancer (HNSCC) show upregulation of FCGBP [147,148]. The

upregulation in HNSCC is hypothesised to be caused by Human Papillomavirus (HPV) infection-associated gain of chromosome 19, the chromosome where the *FCGBP* gene is encoded [147]. Many of the published studies utilise transcriptomic data, however a small number of proteomic studies have been published. A study combining label-free proteomics with transcriptomic data from a public database reports that stage 2 CRC has a strong down-regulation of FCGBP on both protein and mRNA-level [145]. Another proteomic study suggests that FCGBP could be used as a marker to differentiate non-mucinous neoplastic cysts from other cysts in the pancreas [149].

2 AIM

2.1 General Aims

The aim of this thesis is to study the structural features of FCGBP, find possible interactors, and determine the importance of FCGBP for the integrity of the mucus barrier found in small and large intestine, and airways during disease. Furthermore, this thesis aims to generate new models for studying FCGBP, and map which cell types secrete FCGBP and the mechanisms of secretion.

2.2 Specific aims

Paper I – Map the extent of GDPH cleavage in FCGBP and determine its effect on the integrity of the protein. Study FCGBP expression *in vivo* and perform quantitative analysis of the covalent interaction with MUC2 to assess its physiological significance. Reproduce and reassess IgG Fc binding *in vitro* using purified proteins.

Paper II – Study FCGBP function in a murine COPD model, correlating molecular and ultrastructural data from both *in vitro* and *in vivo* experiments, to *ex vivo* and *in vivo* phenotypes.

Paper III – Utilise *in vitro* approaches to map the interactions driving homomeric oligomerisation of FCGBP and study its 3D-structure using the murine orthologue as a model.

Paper IV – Combine *in vitro*, *in vivo* and *ex vivo* approaches to study FCGBP ultrastructures and function in the GI tract focusing mainly on the colon. In addition, explore the function of the FCGBP N-terminal sequence in human, and uncover why parts of the sequence is missing in murine Fcgbp.

3 CONTRIBUTIONS

Paper I – I generated clones for recombinant proteins used for anti-serum generation. I was strongly involved in setting up, analysing the experiments and writing the paper. I was also highly involved in obtaining and setting up the *Fcgbp*$^{-/-}$ mouse strain, in addition to extracting and analysing the samples. I set up and performed purification of proteins and generated stable clones, performed electrophoresis, Western blot and analysed FCGBP content of bands through in gel digestion and MS, with data analysis. Furthermore, I detected peptides verifying the predicted signal sequences along with peptides from GDPH cleaved fragments. I did EndoH treatment of lysates from stable clones and performed sample preparation and label-free MS quantification of mucus proteins to assess covalent binding between FCGBP and MUC2. I also set up and performed binding assays with FCGBP and IgG. Finally, I performed the IHC experiments and analysis.

Paper II – I was involved in intranasal treatment of mice with elastase and extraction of tissue. I performed biochemical studies of *in vivo* and *in vitro* protein, including protein purification. I took part in setting up IHC staining of FCGBP and image acquisition.

Paper III – I cloned recombinant proteins and performed most protein purifications. I discovered the C-terminal cysteine dimer of FCGBP, performed the electrophoresis and size exclusion chromatography (SEC) experiments in addition to MS sample preparation and analysis. I did not perform the cryogenic electron microscopy (cryo-EM) and small angle x-ray scattering (SAXS) methods, but I got to observe and learn when my collegues performed these. My *in silico* sequence analyses and predictions helped in the workflow for elucidating the Fcgbp structure.

Paper IV – I designed and performed all experiments. I participated in measuring mucus thickness and designed the primer sequences. I also performed cDNA/DNA amplifications in addition to analysing the sequencing data.

4 METHODS AND MATERIAL

4.1 Molecular structure analysis and binding experiments

In this thesis, experiments were designed and analysed based on the uniprot FCGBP consensus sequences Q9Y6R7 and E9Q9C6, for human and mouse respectively. Structures, processing and Mr of *in vitro* and *in vivo* proteins were analysed by multiple approaches, including *in silico* methods. Both native and denaturing conditions were used for these studies. *In silico* sequence analysis was important for driving the project forward, as alignments and phylogenetic analyses enabled prediction of the vWD, C8 and TIL domain borders, along with determination of similarities between FCGBP orthologues and paralogues. The project pivoted towards *in vitro* analysis of FCGBP using the murine orthologue as a cornerstone, due to it being highly homogenous to the human variant (**papers I-III**), even though it is shorter and lacks tandem repeats at the center. N-terminals regions differed between species, and therefore a set of truncated N-terminal constructs based on both orthologues were used to find and study differences (**papers I, IV**).

Laemmli SDS-PAGE gels were used to anticipate Mr and study post-translational modifications (PTMs) both under reducing and non-reducing conditions [150] (**papers I-IV**). For analysis in absence of denaturing agents, Native-PAGE, developed by Schägger *et al.*[151], was used to study non-covalent interactions (**papers I-IV**), both homomeric (**papers II-IV**) and heteromeric interactions as seen in the case of FCGBP and IgG Fc (**paper I**). Electrophoresis and Western blot analysis was performed on both *in vivo* and *in vitro* material, comparing electrophoretic migration of recombinant protein (**papers I-IV**) and protein expressed in colonic mucus (**papers I, III**), as well as in lung BALF (**paper II**). This was valuable also for insight to whether or not *in vitro* protein was processed and folded in a comparable means to *in vivo* FCGBP.

Working with large proteins, heavily crosslinked by disulphides, and studying differences between oligomeric states under different oxidative states is challenging as the oxidised protein could be expected to migrate faster than reduced protein through polyacrylamide gels, due to lower friction arising from a more compact structure. As well, disulphide bonds can shuffle, forming new links within or between proteins, which can reduce the size of individual cleaved proteins, or create larger polymers by cross-linking. At a high pH, the disulphides will lose their protons, becoming reactive thiolates that can attack other disulphide bonds forming new links [152]. Mucins are thought to be stored

at a low pH inside the vesicles of GCs [38]. Such a pH would also be beneficial in preventing unwanted shuffling of cysteines.

The Native-PAGE method seeks to study protein structures in absence of denaturants. Here, the issue of protein charge which varies greatly between proteins depending on isoelectric points (pIs) and pH is mitigated by adding negatively charged Coomassie G-250 dye [151]. Using the NativeMark ladder from Invitrogen, complexes between 20 (soybean trypsin inhibitor) to 1236 (pentameric IgM) kDa can be resolved in a 4-16% gradient gel. Hypothetically, Coomassie G-250 could interfere with electrostatic interactions, thereby masking some binding events and conformations. In support, SEC was also used for size anticipation and interpolation for recombinant proteins. Using the low and high Mr gel filtration calibration kits from General Electric Healthcare, sizes within the range of 6.5 (aprotinin) to 669 (thyroglobulin) kDa could be interpolated by linear regression analysis (**papers II-III**). The method grants full control of the solvent used for the proteins and thereby simplifies studies of molecular dynamics. Here, charge is not normally a factor but just as is the case for Native-PAGE, the shape of the molecule will affect the results. Both Native-PAGE and SEC methods are based on molecular mass markers that are globular, meaning that more fibrillar proteins will likely show an overestimation of the Mr. To study IgG sequestering *in vitro*, all these methods were combined to screen for co-migration or co-elution of IgG and recombinant FCGBP (**paper I**).

In a collaborative study, SAXS (**paper III**) and cryo-EM (**papers II-III**) were used to study full-length recombinant murine Fcgbp. SAXS is useful as it provides information not only on the radius of gyration (Rg) and Mr, but as well the number of subunits in the complex. An envelope for the volume of the molecule can further give information on structural organisation and folding. It is also possible to perform the analysis using any buffer, allowing for studies of molecular dynamics [153]. Cryo-EM allowed for structural determination of less flexible regions of the molecule (**paper III**), as well as analysis of polymeric structures in micrographs (**papers II-III**). This technique is generally great for studying structure of large proteins. Proteins can be plunge frozen using any buffer, therefore also allowing for characterisation of structural dynamics depending on solvent. From imaged cryo-EM micrographs, a vast amount of molecules from different angles are found. Even though many regions of a molecule might be flexible, which would make structural determination by X-ray crystallography impossible, many particles in the Cryo-EM micrographs will share a specific conformation at a given time, allowing for structure to be determined by software by specifically picking these out and constructing an envelope [154]. Early studies of Fcgbp ultra

structure was performed with a low-resolution Talos L120C (**paper II**), whereas the further analysis was performed using a Titan Krios G2 (**paper III**), both from Thermo Fisher Scientific. Results from all methods were combined to draw conclusions on the structure of FCGBP.

4.2 Production of recombinant FCGBP and polyclonal antiserum

The main method used for cloning of N and C-terminal human FCGBP fusion proteins was through restriction enzyme linearisation of vectors and PCR amplification of inserts with up to 15 bp overhangs complementary to target vector (**papers I, III-IV**). Inserts and vectors were fused together by isothermal assembly, using the Gibson assembly master mix (New England Biolabs) [155]. For generating a D9-D11 vector, being the most similar to the one triplet found in the center of murine Fcgbp, two inserts were synthesised by GenScript and assembled into the psecTAG vector. The D9-D11 construct was not included in manuscripts as we did not manage to successfully express the protein in CHO cells.

Due to the sequences being highly unique and differing the most from paralogous sequences, the FCGBP_N and C-terminal FCGBPD13 regions were recombinantly expressed, purified by histidine affinity chromatography and used to immunise rabbits for generation of a highly specific polyclonal antiserum [**paper I**]. After comparing MUC2 peptides used for antisera, it was found that the MUC2C3 antiserum [8], would likely cross-react with GDPH cleavage products from FCGBP. This MUC2 antiserum was therefore also used to study FCGBP processing in **paper I**. There existed also two commercial polyclonal antibodies from Atlas Antibodies. These are the HPA003564 and HPA003517. The first of these targeted the human N-terminal sequence shared with helical gliding bacteria, whereas the latter was less specific, targeting the D5, 8 and 11 assemblies of human FCGBP. A full description of these and the rest of the antibodies used to study human and murine FCGBP is found in **paper I**.

A full-length murine Fcgbp expression vector was purchased from Origene as its sequence is in many parts similar to that of human FCGBP, and its expression easier to achieve in mammalian cell-lines due to a smaller size. However, in order to express the protein in adherent CHO cells, stable clones had to be generated. Transient expression in suspension-grown CHO-cells eventually resulted in good yields (**paper I**). Either affinity purification or ion exchange chromatography, sometimes followed by SEC were used to purify

the recombinant FCGBP proteins (**papers I-IV**). The low pIs of the FCGBP proteins facilitated anionic exchange chromatography, resulting in the proteins being found in the fractions eluted using high salt concentration. Bearing in mind that reactive internal anhydrides generated from putative GDPH cleavage should bind amine groups, HEPES and PBS were generally buffers used for storing the proteins, as Tris-HCl would potentially bind to the anhydrides. All FCGBP vectors used in this thesis are described in **table 1** below.

Table 1: Recombinant FCGBP expression vectors

Vector	Supplier	Species	Domains	Sequence	Tag	Antiserum	CHO Expression
pcDNA3.1	Invitrogen	Human	N	Q9Y6R7 (#1-471)	Myc, His	Rabbit	+
psec-TAG	Invitrogen	Human	N-D2	Q9Y6R7 (#1-1251)	Myc, His	No	+
psec-TAG	Invitrogen	Human	D9-D11	Q9Y6R7 (#3604-4865)	Myc, His	No	-
pcDNA3.1	Genescript	Human	D11-D12	Q9Y6R7 (#4466-5234)	Myc, His	No	+
psec-TAG	Invitrogen	Human	D12-D13	Q9Y6R7 (#4856-5405)	Myc, His	No	+
psec-TAG	Invitrogen	Human	D13	Q9Y6R7 (#5235-5405)	Myc, His	Rabbit	+
pCMV	Origene	Human	N-D13	Q9Y6R7 (#1-5405)	Myc, DDK	No	-
pcDNA3.1	Genescript	Mouse	N-D2	E9Q9C6 (#27-835)	Myc, His	No	+
pCMV	Origene	Mouse	N-D7	E9Q9C6 (#1-2583)	Myc, His	No	+

4.3 Animal experiments

A major issue encountered in beginning of my work was that the initial murine knockout model available, the TF0744 from Taconic, had a gene-trapping cassette in intron 18, meaning a stop sequence prior to exon 19. With the *Fcgbp* gene having 20 exons, amino acids #1-2490 out of 2583 could theoretically still be expressed, producing a 262 out of 272 kDa protein. This was a misstake as it had been generated before the *Fcgbp* gene had been properly sequenced. Thus, it was deemed unsuitable for studies as Fcgbp consists of repeats of many similar domains, therefore greatly stalling the project. Instead, an *Fcgbp*$^{-/-}$ variant with a LacZ insertion found after exon 3 was purchased from the EUCOMM library (EM05780) [**paper I**]. This model can be regarded as a full *Fcgbp*$^{-/-}$ as exon 2 corresponds to the 1st vWD domain (visualised in **Fig. 10**), theoretically only resulting in a 12 kDa peptide. The EM05780 is the model referred to as *Fcgbp*$^{-/-}$ in **papers I-IV** of this thesis. Data from the truncated *TF0744 variant is not included.*

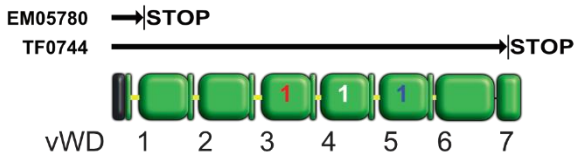

Figure 10. An overview of the translated product length in murine Fcgbp$^{-/-}$ variants. The original TF0744 ends translation in the 7th vWD, whereas the EM05780 from the EUCOMM library ends translation in the first vWD, resulting a substantially truncated product.

An *ex vivo* explant system for analysis of colonic mucus was used for studies (**papers I, III-IV**). Looking at barrier defects, unchallenged non-cohoused non-littermate WT and *Fcgbp*$^{-/-}$ mice were screened for phenotypes pertaining to mucus structure and organisation in the distal colon (DC). First introduced by Gustafsson *et al.* [51] micrometer-sized fluorescent carboxylate beads were added apically on the mucus to compare bead penetrability between *Fcgbp*$^{-/-}$ and WT controls. These beads are negatively charged to avoid binding to the acidic glycan structures of mucins. As well, fluorophore-labelled wheat germ agglutinin (WGA) and *Ulex Europeus* Agglutinin-1 (UEA1) lectins were used to stain and compare ultrastructures of mucus as described previously by Nyström *et al.* [55,156]. Mucus expansion was tested over-time through measurement of the thickness between mucus coated by 10 µm polystyrene beads and the epithelium, using a conventional stereo microscope and a 45° tilted glass capillary needle, performed at 37 °C with fresh Krebs's glucose buffer pumped basally (**paper IV**) [51].

Murine mucus for biochemical and MS analysis was extracted by directly scraping off the apical side of the tissue into PBS on a silicon-coated petri dish, whereas human mucus was extracted after 1 h growth in the mucus measurement chamber (**papers I, III**).

Studies pertaining to lung was done in collaboration. Intranasal administration of elastase was used to induce a COPD-like condition [92] in both *Fcgbp*$^{-/-}$ and *C57BL/6* mice, comparing them to naïve mice. Bronchoaleolar lavage fluid (BALF) was collected for MS and biochemical analyses. Airways were fixed and used for IHC experiments. Some mice were used for testing ciliary mucus transport velocity by a method previously described by Fakih *et al*. using an explant system [88] (**paper II**).

4.4 IHC and imaging

For *in vivo* IHC, paraffin-embedded sections from both co-housed littermates and non-cohoused non-littermates fixed in either Carnoy's solution (**papers I-IV**) or formaldehyde (**papers II, IV**) were studied. Images of lung (**paper II**) and GI tissue sections were obtained using a Nikon Eclipe E1000 (Nikon) epifluorescence microscope or a LSM-700 confocal system (Zeiss). The latter granted the ability to acquire z-stacks tracking FCGBP ultrastructures in mucus (**papers II, IV**).

4.5 MS analysis

In literature, proteomic analyses of secretions from different areas are often based on SDS-PAGE electrophoresis separation. However, this methodology is likely not always optimal for detecting mucins and other large proteins, depending also on oxidative state, with too large complexes not being able to enter gels with high polyacrylamide concentrations. Here, we relied on a mix of methods to study FCGBP and other mucus component (**papers I-III**). Three different instruments were used to analyse lysate and mucus from colon or lung BALF through in-gel digestions or total protein content analysis by filter-aided sample preparation (FASP) protocols and label-free quantification.

The instruments used here included LTQ Orbitrap XL, Q-Exactive HF and Q-Exactive HFX, all from Thermofisher Scientific. Mascot V. 2.6.0. (Matrix Science) (**papers I, II**I), PEAKS 2017 (Bioinformtics Solution Inc) (**paper I**) and MaxQuant V. 1.5.7.4 [157] (**papers I-II**) were programs used to analyse the data. In order to assess the portion total FCGBP peptides coming from the different GDPH cleavage products, thereby assessing which fragment the

peptides belonged to, manual processing was required (**papers I, III**). When studying covalent cross-linking between Fcgbp and Muc2 in presence of GuHCl at pH 8, as was done in **paper I,** N-Ethylmalemide (NEM) was added to avoid shuffling of disulphides by binding and neutrallising free thiolate groups [64].

4.6 Statistics

All statistical test and presented diagrams were generated using GraphPad Prism version 8 (Graphpad). Both parametric and non-parametric tests were performed for independent groups. The former included students T-test (**paper IV**) and the latter Mann Whitney-U test (**papers I-II**) or Kruskal-Wallis (**paper II**). A $p < 0.05$ was set as a threshold for statistical significance. Parametric tests were used for comparing mucus thickness as data follows normal distribution [54] (**Paper IV**). However, for the rest of the statistical analyses, we used non-parametric tests because of small sample sizes and lack of knowledge on variance and normal distribution (**Papers I, II**).

5 RESULTS AND DISCUSSION

5.1 Complete hydrolysis of FCGBP GDPH Motifs and no binding to IgG Fc or Muc2 (**Paper I**)

To set the direction for exploratory studies of the FCGBP protein, descriptive work examining and further scrutinising known and hypothetical aspects of function, expression and structure, was performed as summarised in this paper. A main focus of this paper was to study the many GDPH motifs which are the hallmark of the FCGBP protein. Although potential cleavages have been described previously [43,44], no extensive study has been made where the cleavage of the full protein has been examined. It was found that all GDPH motifs were cleaved *in vivo*, studying both human and murine FCGBP orthologues by electrophoresis and MS analyses of both cellular and secreted material. Mutating the GDPH sequence to GAPH in recombinant truncated human N-D2 also showed that the Asp was necessary for cleavage, but not for secretion of the protein. In contrast, absence of GDPH cleavage blocks extracellular translocation of RGM [129] and SSDO2 [133]. Working with recombinant murine Fcgbp revealed that intracellular GDPH cleaved fragments were sensitive to EndoH treatment, showing that the cleavage occurs in the endoplasmic reticulum. Furthermore, *in silico* modelling and MS analysis of recombinant proteins suggested that the reactive Asp-anhydrides resulting from GDPH cleavage can be hydrolysed. The fact that the GDPH motif is located within a non-exposed region of the domain structure, argues against the hypothesis that anhydrides could be exposed following cleavage and be involved in crosslinking FCGBP to other proteins such as MUC2 [44]. In support, quantitative *in vivo* approaches revealed murine Fcgbp to be mostly soluble in GuHCl, whereas Muc2 was highly insoluble. Hardly any Fcgbp was found upon MS-analysis of the insoluble Muc2 band separated in AgPAGE, arguing that a covalent link between MUC2 and FCGBP is not physiologically relevant.

For both the human and mouse FCGBP orthologue, GDPH cleaved fragments were still tethered by disulphide bonds. While MUC1 is cleaved in the SEA domain [123] and MUC4 by GDPH [121], none of these are covalently stable. In contrast, the disulphide bonds of FCGBP stabilised the protein backbone otherwise thought to be held together by two beta strands. However, these bonds could still be the weakest covalent links within the FCGBP structure, as

disulphides can be placed in different angles and are subjected to varying degrees of axial tension within different molecules. This affects how much energy is needed to break the bonds [158], which could be of relevance for FCGBP, resulting in a controlled dissociation and detachment of mucus in response to heavy sheer stress. *In silico* alignments and modelling showed that only a disulphide bridge stabilises each GDPH cleaved vWD of FCGBP, similar to what has been described for SUSD2 [133], and BMPER [131]. One of the cysteines in each disulphide holding together GDPH cleaved fragments is also part of a CXXC motif. CXXC motifs are mostly found in redox proteins such as thioredoxin, and they are thought to lower the redox potential, allowing for reducing agents to attack and reduce disulphide bonds [159]. It is possible that having this motif would allow for the disulphides to be more sensitive to a reducing environment. The human defensin HD1B is suggested to be activated by bacteria producing a reducing environment inside the intestinal lumen [160], and it is thought that the mucin networks also can be dissolved by sulphide-producing bacteria, not only by breaking the intermolecular links, but also the intramolecular bonds that protect against proteolytic degradation [161,162]. It could therefore be speculated that a reductive environment generated by a high bacterial load in the OML of the colon could also cause the FCGBP to break, thereby driving dissociation and detachment of mucus.

In vitro, the described binding between IgG and FCGBP [39-42] was tested under native conditions. IgG sequestering by FCGBP could not be detected in either Native-PAGE or SEC, arguing against this function, touted in literature as being of primary importance, by which the protein has also been named. However, with recent progress in studies on structure and function of similar proteins containing vWDs, a clear link to IgG sequestering has not really been indicated.

Additionally, MS analyses revealed the peptides for the signal sequences of murine and human orthologues, which would prove useful in future cloning endeavors.

5.2 Murine Fcgbp polymers drive mucus attachment in an elastase COPD model (**Paper II**)

With Fcgbp not normally being expressed in the murine lung, the use of an elastase induced model of inflammation helped partially elucidate its function in the organisation of the mucus layer covering the tracheal surface under COPD-like conditions [92]. It was found that for elastase treated mice with Fcgbp absent in BALF, there was less retention of mucus after washing and performing BALF extractions. Epithelial surfaces in the airways were less covered by attached mucus, and there was less airway obstruction in general. IHC stainings using the αFCGBP-D13 serum revealed dense trabecular Fcgbp structures in the mucus that did not mix with UEA1-stained material, thereby possibly devoid of Muc5b and Muc5ac mucins. However, Fcgbp trabeculi were still embedded within the mucinous materials, and they also formed a characteristic cap-like formation covering the stratified mucus. Such large structures could suggest that FCGBP is able to form linear polymers, and that there are also lateral interactions. *Ex vivo* assays revealed that elastase treated *Fcgbp*$^{-/-}$ mice had significantly higher mucus clearance transport velocity mediated by beating cilia. It has been speculated that MUC5AC is an important mediator of mucus attachment in the lungs, however *Muc5ac*$^{-/-}$ mice still displayed mucus attachment in the same *ex vivo* setup [88]. This suggests that rather than MUC5AC, FCGBP could be the key mediator of mucus attachment in the airways during disease, and thereby drive mucus plugging in COPD, asthma and CF. Elastase treated *Fcgbp*$^{-/-}$ mice also showed less parecnhymal inflammation, suggsting that abberant FCGBP production leading to mucus retention drives the inflammation in COPD, possibly indirectly by accumulation of bacteria.

The molecular details of the FCGBP superstructures were investigated. Size analyses using recombinant full-length murine Fcgbp showed a band close to 460 kDa in electrophoresis with SDS but without reducing agent. Proteins crosslinked by disulphides tend to migrate faster in SDS-PAGE in an oxidised states, and as the Fcgbp has a predicted size of 275 kDa, it could be argued that the band detected is covalent dimer. This band came from protein fractions separated by a Superose 6 increase column in a peak that had lower column-retention than thyroglobulin. The thyroglobulin molecular mass standard is 669 kDa, meaning that the size of the native protein might correspond to complexes even larger than a dimer, where non-covalent interactions drive

further polymerisation. Native-PAGE analysis as well revealed a size of at least 720 kDa, strengthening previous findings. Both the SDS-PAGE and Native-PAGE bands detected *in vitro*, were confirmed *in vivo* when analysing BALF from elastase treated mice using Western blot and the αFCGBP-D13 serum from **paper I**. This further suggested that the recombinant protein was folded and processed similarly to the *in vivo* protein expressed in the airways. Low-resolution cryo-EM images of the recombinant murine orthologue revealed linear polymers of varying length, likely mediated by transient interactions as such complexes were not indicated by neither Native-PAGE nor SEC.

In all, findings in this paper suggest that FCGBP drives retention of airway mucus through a combination of homo and heteromeric interactions, with the former having a covalent component. Pharmacological compounds targeting FCGBP synthesis or function could therefore decrease mucus retention and inflammation.

5.3 FCGBP forms a C-terminal disulphide dimer and a highly globular quarternary structure involving many D modules (**Paper III**)

In **paper III**, the homomeric polymerisation of FCGBP was further studied. The base model for various interactions mapped by recombinant proteins is summarised in **figure 11** on **page 36**. The same 720 kDa band observed using Native-PAGE *in vitro* and in BALF from elastase treated mice (**paper II)** was also found in mucus from murine distal colon. SAXS analysis of recombinant full-length Fcgbp also showed a Mr in the same range. However, further data analysis suggested that the complex was comprised of only two subunits. Analysing briefly reduced and alkylated murine protein in Native-PAGE showed a band that had migrated faster in the gel than non-reduced material. MS analyses of this band suggested is was actually a reduced monomer that had not yet been dissociated into individual GDPH fragments. The result suggested that the dimer was at least partly mediated by disulphide bonds. Expression of a truncated C-terminal construct, human FCGBPD11D12, corresponding to D5D6 of murine Fcgbp, did not form a dimer (**Fig. 11**). However, both the human D12D13 and D13 recombinant proteins showed reducible dimers using SDS-PAGE (**Fig. 11**), therefore mediated by the terminal D13 domain. *In silico* alignments showed that the cysteine configuration of the terminal vWD was conserved in both the human and murine orthologue. Further modelling and MS analysis of the recombinant D13

protein suggested that subunits were organised in an antiparallel manner (**Fig. 11**).

Through cryo-EM, using micrographs with full-length recombinant Fcgbp, it was possible to determine the structure of the sequence between D4 and D7. It formed a convoluted, or globular, antiparallel dimeric interface stabilised by inter-molecular bonds between D7´-D7 and D7´ to D4, and vice versa (**Fig. 11**, subunit Fcgbp´ shown in red, other subunit in green), with partially de-convoluted conformations also observed in micrographs. We chose to call the complete quaternary structure of the C-terminal interface the "core complex". The cysteine dimer was also found in the murine structure**,** and the resolved area was fully GDPH-cleaved. Due to flexibility, the structure of the sequence spanning from the N-terminus to the end of the D3-assembly was not resolved at his time. Superimposing the SAXS envelope with the cryo-EM structure suggested that N-terminals were protruding like flexible arms from the core-complex (**Fig. 11**). Native-PAGE migration and SEC Mr interpolation further suggested a Mr closer to a tetramer for the human D12D13 (**Fig. 11**).

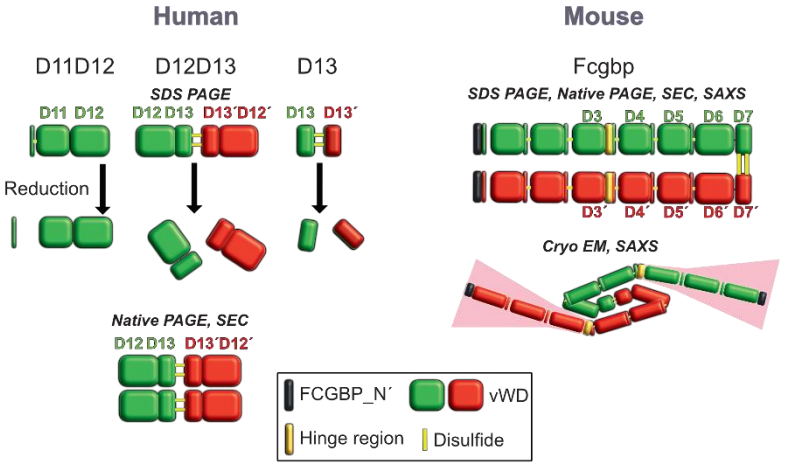

Figure 11. An illustrative figure visuallising interactions and structures found using recombinant FCGBP proteins. The hD11D12, hD12D13 and hD13 are human recombinant truncates of the the FCGBP C-terminus, whereas Fcgbp is the recombinant full-length murine protein. The figure includes description of methods used for analysis. The motifs, bonds and domains included in the image have a description and annotation. The different domains, bonds and and regions of interest are color-coded and annotated. When D-assemblies are forming a dimer or tetramer, both red and green colors are used to represent the domains of a specific subunit. The illustration of the murine Fcgbp 3D-structure down on the right has pink triangles marking the flexibility of the N´-D3 sequences.

It was speculated that partially de-convoluted FCGBP C-terminal interfaces could re-assemble into even bigger complexes. Using both conformations of murine Fcgbp detected in cryo-EM to search for filaments in high-resolution micrographs again revealed linear structures of varying length, as in **paper II**. With mucus attachment being a proposed function for FCGBP, it could further be speculated that the core complex can contribute to the elastic properties of mucus. The murine Fcgbp structure resembles a spring feather, likely being able to store and release kinetic energy, as there are no known covalent crosslinks that locks the structure internally. The transient nature of Fcgbp polymerisation likely allows for a much more dynamic and shapable material than what mucins can provide, as the linear polymers formed by MUC2 are stabilised by covalent interfaces at both termini [6,7]. However, in addition to not having *O*-glycans causing steric hindrances, great amount of D-assemblies in FCGBP, compared to the mucins, likely also contributes to far more possibilities in how FCGBP can assemble into larger complexes. It could be that shear stress modulates interactions with other structures in a similar manner as vWF, albeit without involvement of vWA domain, similarly resulting in binding or susceptibility to proteolytic cleavage [113], thereby affecting mucus structural properties and clearance.

5.4 FCGBP is important for colonic mucus structure and the N-terminal region genetically destroyed in mice forms non-covalent dimers (**Paper IV**)

With new knowledge on FCGBP structure (**papers I-III**), by *in silico* and *in vitro* work we tried to elucidate all the main differences in the amino acid sequences between the human and murine orthologues, as it was important to determine how much of the structural information was actually transferable between orthologues. Furthermore, we also wanted to closer study the protein function in its main sites of expression, the small intestine and lower GI tract [29].

The illusive FCGBP_N domain described by Lang *et al*.[108] was further investigated. It was found that the sequence unique to humans, here named FCGBP_N1 (N1), was limited to only one exon that was missing in mouse. cDNA amplification and sequencing did not reveal any new data compared to the mouse Fcgbp sequence already published. However, alignments of the murine genomic *Fcgbp* sequence with cDNA from species with complete N-terminal sequences showed that murine N1 had been genetically destroyed through several deletions and insertions. The FCGBP_N2 sequence (N2),

being the N-terminal sequence found the D1 assembly still shared between humans and mice, was originally called FCGBP_N´ (paper I). This new architecture nomenclature was deemed to be better as it is based on N-terminal positioning in human FCGBP. Alignments of all FCGBP D-assemblies with the N2 sequence showed that this sequence was actually repeated and found after the E domain in each D-assembly of both human and mouse FCGBP. In relation to the suggested flexible D3 hinge-region (**paper III**), the N2 sequence was located N-terminally. Both the longer human N1 and shorter N2 sequence shared with mouse are linked to helical gliding bacteria such as the myxococcales [108].

Another key inter-species difference identified was that vWD1 domain of human FCGBP had an unpaired cysteine close to the previously described CXXC motif, whereas murine Fcgbp had an extra cysteine in TIL1. The proximity to a CXXC motif could indicate involvement in disulphide shuffling. As found in **paper I**, one of the cysteines in the CXXC motif is believed to be part of the disulphide bridge tethering the GDPH fragments of the N-terminus and D1. Here, an unpaired cysteine seemed to cause a small proportion of the shorter recombinant human (FCGBPN-D2) and murine (FcgbpN-D2) N-terminal proteins to form cysteine dimers in SDS-PAGE. However, mostly monomers were observed under native conditions, suggesting that these SDS-PAGE findings might be an artefact. Interestingly, expressing the complete human N-terminal sequence in absence of D-assemblies, composed of N1 and N2 (FCGBP_N1-2), showed mostly non-covalent dimers, suggesting that the assemblies block an interaction under normal circumstances.

As was previously found (**paper I**), murine DC has an Fcgbp expression pattern similar to human sigmoid colon where FCGBP expression is seen in all GCs of the crypts. Similarly, it was observed that expression of FCGBP could be found along most of the ileal crypts of both species. However, expression varies when comparing murine proximal colon (PC) and human ascending colon. In human aschending colon, FCGBP can be seen expressed at all places of the crypts, but it is only found near the top of the crypt in mouse. Co-staining of Carnoy-fixed sections using UEA1 lectin, MUC2 and FCGBP antibodies suggested that FCGBP is expressed in all goblet cells, but that the proteins are segregated upon secretion. Fcgbp forms dense elongated structures in the colonic IML, similar to those found in lung (**paper II**). Stainings of PFA-fixed sections suggested that the proteins can be packed in the same vesicles. The sections also reveal stronger Muc2 signal in *Fcgbp*$^{-/-}$ compared to WT, indicating that Muc2 could be upregulated in these mice.

Using the MUC2C3 antibody, Johansson *et al.* first described a structure in the innermost part of the colonic IML, stained differently [8]. Our stainings further indicate cross-reactivity (**paper I**) with Fcgbp and that these differently stained structures could partly be composed of Fcgbp. Similar formations are not present in *Fcgbp*$^{-/-}$, and these structures can also be identified using the specific αFCGBP-D13 serum on WT sections. Z-stacks of UEA1 and FCGBP co-stained colons shows dense fiber-like and trabecular structures secreted from GCs, similar to those observed in tracheas of elastase treated mice (**paper II**). The Fcgbp and mucinous materials appear, in agreement, segregated and do not mix. Yet, Fcgbp structures are still embedded as a mesh within the UEA1 stained material. These dense structures are long and usually retain their link to the cells from which they were secreted suggesting that they provide a focal point for attachment.

The absence of these Fcgbp structures appears to negatively affect the organisation of mucus secreted by surface GCs *ex vivo*, stained by the WGA lectin. 1 h mucus growth-rate measurements showed a significantly higher expansion rate ($p < 0.05$) and more fragile mucus in *Fcgbp*$^{-/-}$, further suggesting that Fcgbp is complicit in mucus organisation and clearance. However, the sample size (n) for the mucus measurements is small, consisting of only 4 in each group. Using fluorescent micrometer-sized carboxylate beads on top of mucus *ex vivo* did not show a penetrability phenotype for these mice, arguing against a disrupted barrier.

Just like transglutaminases are thought to provide vertical stabilisation of the mucus layer by covalent bonds, making the mucus more insoluble [64] and likely promoting attachment, FCGBP could be providing another dimension to vertical stabilisation and attachment, by means of more dynamic non-covalent interactions.

6 CONCLUSION

In conclusion, the murine Fcgbp orthologue is an excellent model for studying FCGBP both *in vivo* and *in vitro* as the D-assemblies share a high sequence homology (**paper I, IV**) with the human protein. The murine orthologue is shorter, making it ideal for *in vitro* expression (**Paper I-III**). As it does not have any additional domains compared to human FCGBP, data on structure and function should be directly transferable to the human protein. However, human FCGBP has a longer N-terminal sequence, here dubbed FCGBP_N1, not found in the murine orthologue. This sequence is genetically lost in evolution and just like the FCGBP_N2 sequence shared with mouse, it is linked to helical gliding bacteria [108]. Therefore, the full human N-terminus must be studied by other means. *In vitro* analyses suggest that the human N-terminus can form non-covalent dimers, a function that is normally blocked or modulated by the D-assemblies (**paper IV**). Fcgbp forms long discrete trabecular or network-like super structures in mucus, embedded within mucin material but not necessarily mixing. FCGBP structures, most likely homopolymers, are sometimes seen directly connected to GCs from which they were secreted, thereby resembling mucus anchors. Both *in vivo* and *ex vivo* data further suggest that FCGBP is important for the structural organisation of the mucus layers. Interestingly, it was also found to be an important mediator of mucus attachment in lungs under diseased conditions (**papers II, IV**). *In vitro* studies support the hypothesis that FCGBP assembles into homopolymers. The base variant of Fcgbp structure is a C-terminal covalent dimer, with D4 to D7 of both subunits forming a non-covalent complex that could promote elastic properties. Linear complexes of varying length were also detected suggesting even larger dynamic complexes can be formed. These were mediated by non-covalent interactions (**papers II-III**), enabling much more dynamic structures in mucus compared to MUC2 polymers crosslinked by covalent bonds [6,7].

Finally, both human and murine FCGBP is fully cleaved through its many GDPH motifs. These cleavages were hypothesised to drive covalent crosslinking between FCGBP and MUC2, by generating a reactive Asp-anhydride. However, no physiologically significant covalent links between these proteins were found, suggesting that such interactions would, if existed, instead likely be mediated by transient non-covalent interactions. Each GDPH

cleaved fragment is tethered by a disulphide bridge (**Paper I**), and an altered redox environment driven by the microbiota could therefore disassemble FCGBP fragments, thereby increasing mucus clearance or mucus detachment.

7 FUTURE PERSPECTIVES

Mucus clearance is likely highly important in the relationship between host and bacteria as is exemplified in CF were increased mucus retention drives inflammation in both the lungs [94] and small intestine [163]. Clearence mechanisms pertaining to MUC2 have been described. Both MEP1β and CLCA1 can cleave the MUC2 protein backbone, thereby modulating mucus clearance [53,54]. As structural organisation and attachment of mucus are here the proposed functions of FCGBP (**paper II, IV**), the general co-expression of FCGBP and CLCA1 [29,92] is interesting. It is possible that these two proteins are important in modulating the attachment and expansion of intestinal mucus, and that a similar relationship might occur in the airways during COPD. The murine CLCA1 orthologue is however not upregulated in elastase treated *Fcgbp*$^{-/-}$ mice (**paper II**). Still, it could be speculated that *Clca1*$^{-/-}$ mice might show increased mucus attachment and reduced mucus transport velocity in airways, as a consequence of this relationship. *In vitro* studies where recombinant proteins are mixed could show if FCGBP can directly interact with or be cleaved by CLCA1.

Essentially**, Paper III** is ready for submission, but it could benefit from more experiments that focus on the relationships between structure and function. Mixing recombinant FCGBP with growth medium from different anaerobic bacteria and analysing the protein processing in SDS-PAGE under non-reducing conditions would indicate if bacteria are able to break the disulphide bonds that hold the GDPH cleaved fragments together. This way, GDPH motifs could be linked to structural alterations. Furthermore, similar to what has been previously carried out in relation to the MUC1 mucin [124], using atomic force spectroscopy on recombinant proteins that are either reduced, non-reduced or have mutated GDPH motifs or cysteines, could show the axial forces needed to mechanically dissociate FCGBP. If a GDPH mutant is more stable than the non-mutated protein, then it would be shown that inter-fragment disulphide bridges are actually the weakest covalent links stabilising the FCGBP protein. With the murine Fcgbp missing the human FCGBP_N1 sequence [108] (**paper IV**), the cryo-EM structure of murine Fcgbp should be complemented with studies of recombinant human N-terminal protein. Some electrophoresis data from **paper IV** showing the migration of human and murine N-terminal recombinant proteins could instead be moved here, and

complemented with additional cryo-EM analysis. Alternatively, a truncated human protein lacking the first two of the central D-assembly triplets (D3-D8) could be cloned. This would likely be easy to express in CHO cells just like the murine variant of the full-length protein (**papers I-III**). The paper could be further complemented with experiments studying the function of the N-terminal sequence unique to human FCGBP. Using ELISA, the human recombinant N-terminus could be used in screening for binding to different types of exopolysaccharides or other molecules associated with biofilm formation, including biofilms from helical gliding bacteria.

Paper IV is in need of more work before it is ready to be submitted for publication. Here, it would be interesting to further study FCGBP in small intestinal mucus. *Ex vivo* mucus attachment in *Cftr*$^{-/-}$, WT and *Fcgbp*$^{-/-}$ mice could be compared, using previously described methods [53,82], also looking at protein cleavage by MEP1β and CLCA1, in addition to ion transport. It could be that not only MUC2 but also FCGBP is a mediator of mucus attachment. As *Cftr*$^{-/-}$ mice have more attached mucus *ex vivo* [53], it could be interesting to generate a *Fcgbp*$^{-/-}$ and *Cftr*$^{-/-}$ double knockout and and comparing the *ex vivo* mucus attachment of these with individual knockouts and WT mice, the double knockout model might reveal less mucus retention in the small intestine. Reduced attachment of mucus would then be mediate by a lack of Fcgbp. Creating such a strain should not be a problem since the *Fcgbp* gene is located on mouse chromosome 7, whereas the *Cftr* gene is found on chromosome 6. Calcium and pH-related molecular dynamics should also be further studied for recombinant proteins. This could be done *in vitro* by SEC, SAXS, cryo-EM, and *ex vivo* by adding bicarbonate to mucus on murine intestinal explants. As *Fcgbp*$^{-/-}$ mice thus far present few phenotypes (**paper II, IV**), it would be interesting to eventually challenge these with either DSS treatment [164] or *Citrobacter rodentium* infection [165] in order to further study the importance of FCGBP in inflammation and infection, as a link to inflammation was found in the airways (**paper II**).

8 ACKNOWLEDGEMENTS

8.1 To collegues and collaborators

Many people were involved in my PhD and I am very thankful for all of the input, guidance and asisstance that I received during this period.

Some people invested more time in my PhD and I would like to give some extra credit to them. Thank you **Pablo Gallego Alonso** and **Maria-José Garcia Bonete** for teaching me methodology, involving yourselves in the FCGBP project and giving input on my writing. Without your involvment in the structural biology, experiment planning, and data analysis, my thesis would have been much more difficult. Thank you very much **Frida Svensson** for involving yourself in the project in general but especially in planning and assisting in animal experiments. Hope you come to Värmland! Thank you very much **Ana-Maria Rodriguez Piñeiro** for involving yourself and teaching me how to do mass spec, but also for involvment in the the writing of **Paper I**. Good luck with your new job! Thank you **Christian Recktenwald** for all invaluable input especially on the biochemistry part of my project. I hope you will do more teaching in the future as you have a lot of knowledge to share. Thank you **Dalia Fakih** for involving me in lung research. You will have great future at Astra! Thanks also to **Beatriz Martinez-Abbad** for your involvment with the animals. Good luck with your expanding familly and I hope we can have more dinner nights even though I won't live in Gothenburg. Thank you **Sergio Trillo-Muyo** for your involvement in the structural and biochemical experiments. You have many interesting stories and anecdotes to tell! Thank you also **Sjoerd van Der Post** for your involvement in GDPH and MS-related experiments. I hope we meet at Sunne vattenpark! Thank you very much **Gunnar Hansson** and **Malin Johansson** for your supervision and giving me the opportunity to do my PhD in your prestigious mucin collective. I hope there will be many more good years of mucus studies!

There were **many others** involved in my PhD that were helpful and kind, teaching me methods and giving me input on my research. To **Liisa Arike**, I wish you and your expanding familly the best, hope the familly will keep getting bigger! Thank you **Elena Layunta** for fixing the Brady printer and showing me FISH. You are the best at mixing drinks! **Jenny Gustafsson** thank you for teaching me animal work and discussing my project. I wish you the best. Thank you **Anna Ermund** for the many discussions on FCGBP and the

many good laughs in the corridor. I wish you good fortune in the future! Thank you **Brendan Dolan** for teaching me methods, discussing experiments and proofreading my manuscripts. Good luck with your grand things! Thank you **Hannah Schneider** for helping in the lab. Have fun at Astra! Thanks to **Thaher Pelasayeed** for all the help in the lab and the many good hours of cleaning the cell culture lab. I hope you group will continue growing in the future! Thank you **Sofia Jäverfält** for the many good years. I am sure your thesis will turn out great! To **Grete Raba**, thank you for the many parties, hair discussions and secret fikas. I hope you will have a great future in Tallin! Thank you also **Jack Sharpen** for the many partys, I hope you wil have a great PhD! Thank you **Elisabeth Nyström** and **George Birchenough** for teaching me animal experiments and discussing the project. Hope you will publish even more prestigious papers in the future! Thank you **Joakim Bergström, Karolina Sjöberg Jabbar** and **Lisbeth Ekholm Eklund** for helping me early in the project. It was a pleasure working with you and I hope you are happy where you are now. Thank you also to **Malin Bäckström**, **Elizabeth Thomson**, and **Richard Lymer** from MPE, not only for producing nice batches of protein over the years, but also for teaching me methods and discussing results, especially early in my scientific career. It was nice working with you all! Thank you **Karin Ahlman** for your service in the lab and good maintenance of laboratory discipline. Very dangerous. I hope your retirement years are good! Thank you **Suzan Fares** for teaching me Danish culture, laws and language. Thank you **Anandi Raijan** for many years of wittiness and magic tricks. I hope you will buy a nice house in Malung! Thank you **Gustaf Hellsén** for being the strongest clinical connection maintained throughout my PhD years. You have both a great clinical and scientific career ahead of you. Thank you **Ana Luis** for the good music exchange in the lab. Good luck with your research! Thank you **Melania Giorgetti** for teaching me Italian culture, hand gestures and animal experiments. Hope you are enjoying a drink on the beach! To **Mahadevan Venkitasubramani**, I wish you a good, productive and long PhD! To **Franscesco Suriano**, I wish you a good time in Sweden and the mucin biology group! Thank you **Jenny Persson** for the good years. To **Sarah Thulin**, **Kristina Johansson**, **Jesper Magnusson** and **Gudrun Ragnarsdottir**, it was nice meeting you and I wish you the best for the future.

8.2 Figures

The icons cell-column-1 (**Fig. 1-8**), epithel-squamos-stratiedifed (**Fig. 1**), and gobelet-cell-2 (**Fig. 1, 7-8**) were altered for usage in the figures. They were taken from Servier Medical Arts (https://smart.servier.com/) and are licensed for free use under Creative Commons license (CC-BY 3.0 Unported https://creativecommons.org/licenses/by/3.0/).

REFERENCES

1 Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65, doi:nature08821 [pii];10.1038/nature08821 [doi] (2010).

2 Berg, R. D. The indigenous gastrointestinal microflora. *Trends in Microbiology* **4**, 430-435, doi:https://doi.org/10.1016/0966-842X(96)10057-3 (1996).

3 Birchenough, G. M., Johansson, M. E., Gustafsson, J. K., Bergström, J. H. & Hansson, G. C. New developments in goblet cell mucus secretion and function. *Mucosal Immunol* **8**, 712-719, doi:10.1038/mi.2015.32 (2015).

4 Dao DPD, L. P. *Histology, Goblet Cells.*, <https://www.ncbi.nlm.nih.gov/books/NBK553208/> (

5 Hansson, G. C. Mucins and the Microbiome. *Annu Rev Biochem* **89**, 769-793, doi:10.1146/annurev-biochem-011520-105053 (2020).

6 Lidell, M. E. *et al.* The recombinant C-terminus of the human MUC2 mucin forms dimers in CHO cells and heterodimers with full-length MUC2 in LS 174T cells. *Biochemical Journal* **372**, 335-345, doi: 10.1042/BJ20030003 (2003).

7 Javitt, G. *et al.* Assembly Mechanism of Mucin and von Willebrand Factor Polymers. *Cell* **183**, 717-729 e716, doi:10.1016/j.cell.2020.09.021 (2020).

8 Johansson, M. E. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U. S. A* **105**, 15064-15069, doi:10.4161/gmic.1.1.10470 (2008).

9 Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology* **10**, 323-335, doi:10.1038/nrmicro2746 (2012).

10 Bergstrom, K. S. B. & Xia, L. Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* **23**, 1026-1037, doi:10.1093/glycob/cwt045 (2013).

11 Brockhausen I, S. P. in *Essentials of Glycobiology* (ed Cummings RD Varki A, Esko JD, et al.) Ch. 10, (old Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015-2017, 2017).

12 Bäckström, M., Ambort, D., Thomsson, E., Johansson, M. E. V. & Hansson, G. C. Increased understanding of the biochemistry and biosynthesis of MUC2 and other gel-forming mucins through the recombinant expression of their protein domains. *Mol Biotechnol* **54**, 250-256, doi:10.1007/s12033-012-9562-3 (2013).

13 Hansson, G. C. Mucus and mucins in diseases of the intestinal and respiratory tracts. *J Intern Med* **285**, 479-490, doi:10.1111/joim.12910 (2019).

14 Agarwal S, K. K. *Histology, Skin*, <https://www.ncbi.nlm.nih.gov/books/NBK537325/> (

15 Murphrey MB, M. J., Zito PM. *Histology, Stratum Corneum.*, <https://www.ncbi.nlm.nih.gov/books/NBK513299/> (2021).

16 Herath, M., Hosie, S., Bornstein, J. C., Franks, A. E. & Hill-Yardin, E. L. The Role of the Gastrointestinal Mucus System in Intestinal Homeostasis: Implications for Neurological Disorders. *Frontiers in Cellular and Infection Microbiology* **10**, 248, doi:10.3389/fcimb.2020.00248 (2020).

17 Barker, N., Van de, W. M. & Clevers, H. The intestinal stem cell. *Genes Dev* **22**, 1856-1864 (2008).

18 Neutra, M. R., O'Malley, L. J. & Specian, R. D. Regulation of intestinal goblet cell secretion. II. A survey of potential secretagogues. *Am. J. Physiol* **242**, G380-G387, doi: 10.1152/ajpgi.1982.242.4.G380 (1982).

19 Brizuela M, W. R. *Histology, Oral Mucosa*, <https://www.ncbi.nlm.nih.gov/books/NBK572115/> (2021).

20 Widdicombe, J. H. & Wine, J. J. Airway Gland Structure and Function. *Physiol Rev* **95**, 1241-1319, doi:10.1152/physrev.00039.2014 (2015).

21 Wickström, C., Davies, J. R., Eriksen, G. V., Veerman, E. C. & Carlstedt, I. MUC5B is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and endocervix: identification of glycoforms and C-terminal cleavage. *Biochem J* **334 (Pt 3)**, 685-693, doi:10.1042/bj3340685 (1998).

22 Phillipson, M. *et al.* The gastric mucus layers: constituents and regulation of accumulation. *Am J Physiol Gastrointest Liver Physiol* **295**, G806-812, doi:10.1152/ajpgi.90252.2008 (2008).

23 Rao JN, W. J. in *Regulation of Gastrointestinal Mucosal Growth* (Morgan & Claypool Life Sciences, 2010).

24 Treuting, P. M., Arends, M. J. & Dintzis, S. M. in *Comparative Anatomy and Histology (Second Edition)* (eds Piper M. Treuting, Suzanne M. Dintzis, & Kathleen S. Montine) 191-211 (Academic Press, 2018).

25 Treuting, P. M., Arends, M. J. & Dintzis, S. M. in *Comparative Anatomy and Histology (Second Edition)* (eds Piper M. Treuting, Suzanne M. Dintzis, & Kathleen S. Montine) 213-228 (Academic Press, 2018).

26 de Almeida Pdel, V., Grégio, A. M., Machado, M. A., de Lima, A. A. & Azevedo, L. R. Saliva composition and functions: a comprehensive review. *J Contemp Dent Pract* **9**, 72-80 (2008).

27 Ligtenberg, A. J., Veerman, E. C., Nieuw Amerongen, A. V. & Mollenhauer, J. Salivary agglutinin/glycoprotein-340/DMBT1: a single molecule with variable composition and with different functions

in infection, inflammation and cancer. *Biol Chem* **388**, 1275-1289, doi:10.1515/bc.2007.158 (2007).
28 Madsen, J., Mollenhauer, J. & Holmskov, U. Review: Gp-340/DMBT1 in mucosal innate immunity. *Innate Immunity* **16**, 160-167, doi:10.1177/1753425910368447 (2010).
29 Rodriguez-Pineiro, A. M. *et al.* Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. *Am. J. Physiol Gastrointest. Liver Physiol* **305**, G348-G356, doi:ajpgi.00047.2013 [pii];10.1152/ajpgi.00047.2013 [doi] (2013).
30 Phillipson, M. Acid transport through gastric mucus. *Ups J Med Sci* **109**, 1-24, doi:10.3109/2000-1967-106 (2004).
31 Martinsen, T. C., Fossmark, R. & Waldum, H. L. The Phylogeny and Biological Function of Gastric Juice-Microbiological Consequences of Removing Gastric Acid. *Int J Mol Sci* **20**, doi:10.3390/ijms20236031 (2019).
32 Janssen, P. *et al.* Review article: the role of gastric motility in the control of food intake. *Aliment Pharmacol Ther* **33**, 880-894, doi:10.1111/j.1365-2036.2011.04609.x (2011).
33 Hsu M, S. A., Lui F. *Physiology, Stomach*, <https://www.ncbi.nlm.nih.gov/books/NBK535425/> (2021).
34 Case, R. M., Ishiguro, H. & Steward, M. C. in *Encyclopedia of Gastroenterology* (ed Leonard R. Johnson) 38-40 (Elsevier, 2004).
35 Heubi, J. E. in *Pediatric Gastrointestinal and Liver Disease (Sixth Edition)* (eds Robert Wyllie, Jeffrey S. Hyams, & Marsha Kay) 24-31.e22 (Elsevier, 2021).
36 Van der, S. M. *et al.* Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117-129, doi:10.1053/j.gastro.2006.04.020 (2006).
37 Velcich, A. *et al.* Colorectal Cancer in Mice Genetically Deficient in the Mucin Muc2. *Science* **295**, 1726-1729, doi: 10.1126/science.1069094 (2002).
38 Ambort, D. *et al.* Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proc Natl Acad Sci U S A* **109**, 5645-5650, doi:10.1073/pnas.1120269109 (2012).
39 Kobayashi, K., Hamada, Y., Blaser, M. J. & Brown, W. R. The molecular configuration and ultrastructural locations of an IgG Fc binding site in human colonic epithelium. *J. Immunol* **146**, 68-74 (1991).
40 Kobayashi, K., Blaser, M. J. & Brown, W. R. Identification of a unique IgG Fc binding site in human intestinal epithelium. *J. Immunol* **143**, 2567-2574 (1989).

41 Harada, N. *et al.* Human IgGFc binding protein (FcgammaBP) in colonic epithelial cells exhibits mucin-like structure. *J. Biol. Chem* **272**, 15232-15241, doi:10.1074/jbc.272.24.15232 (1997).

42 Kobayashi, K. *et al.* Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids. *Gut* **51**, 169-176, doi:10.1136/gut.51.2.169 (2002).

43 Albert, T. K. *et al.* Human intestinal TFF3 forms disulfide-linked heteromers with the mucus-associated FCGBP protein and is released by hydrogen sulfide. *J. Proteome. Res* **9**, 3108-3117, doi:10.1021/pr100020c [doi] (2010).

44 Johansson, M. E., Thomsson, K. A. & Hansson, G. C. Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the Fcgbp protein. *J. Proteome. Res* **8**, 3549-3557, doi:10.1021/pr9002504 (2009).

45 Sadler, J. E. von Willebrand factor: two sides of a coin. *J. Thromb. Haemost* **3**, 1702-1709, doi:10.1111/j.1365-2516.2010.02323.x (2005).

46 Lang, T., Hansson, G. C. & Samuelsson, T. Gel-forming mucins appeared early in metazoan evolution. *Proc. Natl. Acad. Sci. U. S. A* **104**, 16209-16214, doi:10.1073/pnas.0705984104. (2007).

47 Kamphuis, J. B. J., Mercier-Bonin, M., Eutamène, H. & Theodorou, V. Mucus organisation is shaped by colonic content; a new view. *Scientific Reports* **7**, 8527, doi:10.1038/s41598-017-08938-3 (2017).

48 Swidsinski, A., Loening-Baucke, V., Verstraelen, H., Osowska, S. & Doerffel, Y. Biostructure of fecal microbiota in healthy subjects and patients with chronic idiopathic diarrhea. *Gastroenterology* **135**, 568-579, doi:10.1053/j.gastro.2008.04.017 (2008).

49 Shimotoyodome, A., Meguro, S., Tokimitsu, I. & Sakata, T. Histochemical structure of the mucus gel layer coating the fecal surface of rodents, rabbits and humans. *J Nutr Sci Vitaminol (Tokyo)* **51**, 287-291, doi:10.3177/jnsv.51.287 (2005).

50 Ermund, A., Schutte, A., Johansson, M. E., Gustafsson, J. K. & Hansson, G. C. Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches. *Am. J. Physiol Gastrointest. Liver Physiol* **305**, G341-G347, doi:ajpgi.00046.2013 [pii];10.1152/ajpgi.00046.2013 [doi] (2013).

51 Gustafsson, J. K. *et al.* An ex vivo method for studying mucus formation, properties, and thickness in human colonic biopsies and mouse small and large intestinal explants. *Am. J. Physiol Gastrointest. Liver Physiol* **302**, G430-G438, doi:ajpgi.00405.2011 [pii];10.1152/ajpgi.00405.2011 [doi] (2012).

52 Bevins, C. L. & Salzman, N. H. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nature Reviews Microbiology* **9**, 356-368, doi:10.1038/nrmicro2546 (2011).
53 Schütte, A. *et al.* Microbial-induced meprin β cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus. *Proc Natl Acad Sci U S A* **111**, 12396-12401, doi:10.1073/pnas.1407597111 (2014).
54 Nyström, E. E. L. *et al.* Calcium-activated Chloride Channel Regulator 1 (CLCA1) Controls Mucus Expansion in Colon by Proteolytic Activity. *EBioMedicine* **33**, 134-143, doi:10.1016/j.ebiom.2018.05.031 (2018).
55 Nyström, E. E. L. *et al.* An intercrypt subpopulation of goblet cells is essential for colonic mucus barrier function. *Science* **372**, doi:10.1126/science.abb1590 (2021).
56 Schroeder, B. O. *et al.* Bifidobacteria or Fiber Protects against Diet-Induced Microbiota-Mediated Colonic Mucus Deterioration. *Cell Host Microbe* **23**, 27-40.e27, doi:https://doi.org/10.1016/j.chom.2017.11.004 (2018).
57 Birchenough, G. M., Nystrom, E. E., Johansson, M. E. & Hansson, G. C. A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion. *Science* **352**, 1535-1542, doi:352/6293/1535 [pii];10.1126/science.aaf7419 [doi] (2016).
58 Bergstrom, J. H. *et al.* Gram-positive bacteria are held at a distance in the colon mucus by the lectin-like protein ZG16. *Proc. Natl. Acad. Sci. U. S. A* **113**, 13833-13838, doi:1611400113 [pii];10.1073/pnas.1611400113 [doi] (2016).
59 Park, S. W. *et al.* The protein disulfide isomerase AGR2 is essential for production of intestinal mucus. *Proc. Natl. Acad. Sci. U. S. A* **106**, 6950-6955, doi: 10.1073/pnas.0808722106 (2009).
60 Bergstrom, J. H. *et al.* AGR2, an endoplasmic reticulum protein, is secreted into the gastrointestinal mucus. *PLoS. ONE* **9**, e104186, doi:10.1371/journal.pone.0104186 [doi];PONE-D-14-11658 [pii] (2014).
61 Mashimo, H., Wu, D. C., Podolsky, D. K. & Fishman, M. C. Impaired defense of intestinal mucosa in mice lacking intestinal trefoil factor. *Science* **274**, 262-265, doi:10.1126/science.274.5285.262 (1996).
62 Belle, N. M. *et al.* TFF3 interacts with LINGO2 to regulate EGFR activation for protection against colitis and gastrointestinal helminths. *Nature Communications* **10**, 4408, doi:10.1038/s41467-019-12315-1 (2019).
63 Houben, T., Harder, S., Schlüter, H., Kalbacher, H. & Hoffmann, W. Different Forms of TFF3 in the Human Saliva: Heterodimerization with IgG Fc Binding Protein (FCGBP). *Int J Mol Sci* **20**, doi:10.3390/ijms20205000 (2019).

64 Recktenwald, C. V. & Hansson, G. C. The reduction-insensitive bonds of the MUC2 mucin are isopeptide bonds. *J. Biol. Chem* **291**, 13580-13590, doi:M116.726406 [pii];10.1074/jbc.M116.726406 [doi] (2016).

65 Eckert, R. L., Sturniolo, M. T., Broome, A. M., Ruse, M. & Rorke, E. A. Transglutaminase function in epidermis. *J Invest Dermatol* **124**, 481-492, doi:10.1111/j.0022-202X.2005.23627.x (2005).

66 Brandtzaeg, P. Secretory IgA: Designed for Anti-Microbial Defense. *Frontiers in Immunology* **4**, doi:10.3389/fimmu.2013.00222 (2013).

67 Chen, K., Magri, G., Grasset, E. K. & Cerutti, A. Rethinking mucosal antibody responses: IgM, IgG and IgD join IgA. *Nature Reviews Immunology* **20**, 427-441, doi:10.1038/s41577-019-0261-1 (2020).

68 Andreas, N. J., Kampmann, B. & Mehring Le-Doare, K. Human breast milk: A review on its composition and bioactivity. *Early Human Development* **91**, 629-635, doi:https://doi.org/10.1016/j.earlhumdev.2015.08.013 (2015).

69 Justiz Vaillant AA, J. Z., Ramphul K. *Immunoglobulin*, <https://www.ncbi.nlm.nih.gov/books/NBK513460/> (2021).

70 Schwartz, J. L. Fcgbp - A Potential Viral Trap in RV144. *Open AIDS J* **8**, 21-24, doi:10.2174/1874613601408010021 (2014).

71 Gunn, B. M. *et al.* Enhanced binding of antibodies generated during chronic HIV infection to mucus component MUC16. *Mucosal Immunology* **9**, 1549-1558, doi:10.1038/mi.2016.8 (2016).

72 Johansson, M. E. *et al.* Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* **63**, 281-291, doi:10.1136/gutjnl-2012-303207 (2014).

73 van der Post, S. *et al.* Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. *Gut* **68**, 2142-2151, doi:10.1136/gutjnl-2018-317571 (2019).

74 Zhang, Y. Z. & Li, Y. Y. Inflammatory bowel disease: pathogenesis. *World J Gastroenterol* **20**, 91-99, doi:10.3748/wjg.v20.i1.91 (2014).

75 Yashiro, M. Ulcerative colitis-associated colorectal cancer. *World J Gastroenterol* **20**, 16389-16397, doi:10.3748/wjg.v20.i44.16389 (2014).

76 Luo, C., Cen, S., Ding, G. & Wu, W. Mucinous colorectal adenocarcinoma: clinical pathology and treatment options. *Cancer Communications* **39**, 13, doi:10.1186/s40880-019-0361-0 (2019).

77 Bodger, K. *et al.* Altered colonic glycoprotein expression in unaffected monozygotic twins of inflammatory bowel disease patients. *Gut* **55**, 973-977, doi:10.1136/gut.2005.086413 (2006).

78 Larsson, J. M. *et al.* Altered O-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm Bowel Dis* **17**, 2299-2307, doi:10.1002/ibd.21625 (2011).

79 Kobayashi, T. *et al.* Ulcerative colitis. *Nature Reviews Disease Primers* **6**, 74, doi:10.1038/s41572-020-0205-x (2020).
80 Magro, F. *et al.* European consensus on the histopathology of inflammatory bowel disease☆. *Journal of Crohn's and Colitis* **7**, 827-851, doi:10.1016/j.crohns.2013.06.001 (2013).
81 O'Sullivan, B. P. & Freedman, S. D. Cystic fibrosis. *Lancet* **373**, 1891-1904, doi:10.1016/s0140-6736(09)60327-5 (2009).
82 Gustafsson, J. K. *et al.* Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *J Exp Med* **209**, 1263-1272, doi:10.1084/jem.20120562 (2012).
83 van der Doef, H. P. *et al.* Association of the CLCA1 p.S357N variant with meconium ileus in European patients with cystic fibrosis. *J Pediatr Gastroenterol Nutr* **50**, 347-349, doi:10.1097/MPG.0b013e3181afce6c (2010).
84 Brouillard, F. *et al.* Blue native/SDS-PAGE analysis reveals reduced expression of the mClCA3 protein in cystic fibrosis knock-out mice. *Mol Cell Proteomics* **4**, 1762-1775, doi:10.1074/mcp.M500098-MCP200 (2005).
85 Young, F. D. *et al.* Amelioration of Cystic Fibrosis Intestinal Mucous Disease in Mice by Restoration of mCLCA3. *Gastroenterology* **133**, 1928-1937.e1923, doi:10.1053/j.gastro.2007.10.007 (2007).
86 Ermund, A. *et al.* The mucus bundles responsible for airway cleaning are retained in cystic fibrosis and by cholinergic stimulation. *Eur Respir J* **52**, doi:10.1183/13993003.00457-2018 (2018).
87 Ermund, A. *et al.* The normal trachea is cleaned by MUC5B mucin bundles from the submucosal glands coated with the MUC5AC mucin. *Biochem Biophys Res Commun* **492**, 331-337, doi:10.1016/j.bbrc.2017.08.113 (2017).
88 Fakih, D. *et al.* Normal murine respiratory tract has its mucus concentrated in clouds based on the Muc5b mucin. *Am J Physiol Lung Cell Mol Physiol* **318**, L1270-l1279, doi:10.1152/ajplung.00485.2019 (2020).
89 Bustamante-Marin, X. M. & Ostrowski, L. E. Cilia and Mucociliary Clearance. *Cold Spring Harb Perspect Biol* **9**, doi:10.1101/cshperspect.a028241 (2017).
90 MacNee, W. Pathology, pathogenesis, and pathophysiology. *BMJ* **332**, 1202-1204 (2006).
91 Barnes, Peter J. Cellular and molecular mechanisms of asthma and COPD. *Clinical Science* **131**, 1541-1558, doi:10.1042/cs20160487 (2017).
92 Fernandez-Blanco, J. A. *et al.* Attached stratified mucus separates bacteria from the epithelial cells in COPD lungs. *JCI Insight* **3**, doi:10.1172/jci.insight.120994 (2018).

93 Thornton, D. J., Rousseau, K. & McGuckin, M. A. Structure and Function of the Polymeric Mucins in Airways Mucus. *Annual Review of Physiology* **70**, 459-486, doi:10.1146/annurev.physiol.70.113006.100702 (2008).
94 Goetz, D. & Ren, C. L. Review of Cystic Fibrosis. *Pediatr Ann* **48**, e154-e161, doi:10.3928/19382359-20190327-01 (2019).
95 Elborn, J. S. Cystic fibrosis. *Lancet* **388**, 2519-2531, doi:10.1016/s0140-6736(16)00576-6 (2016).
96 Hoegger, M. J. *et al.* Impaired mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis. *Science* **345**, 818-822, doi:10.1126/science.1255825 (2014).
97 Ermund, A., Meiss, L. N., Gustafsson, J. K. & Hansson, G. C. Hyper-osmolarity and calcium chelation: Effects on cystic fibrosis mucus. *Eur J Pharmacol* **764**, 109-117, doi:10.1016/j.ejphar.2015.06.051 (2015).
98 Bonser, L. R. & Erle, D. J. Airway Mucus and Asthma: The Role of MUC5AC and MUC5B. *J Clin Med* **6**, doi:10.3390/jcm6120112 (2017).
99 Batruch, I. *et al.* Proteomic Analysis of Seminal Plasma from Normal Volunteers and Post-Vasectomy Patients Identifies over 2000 Proteins and Candidate Biomarkers of the Urogenital System. *Journal of Proteome Research* **10**, 941-953, doi:10.1021/pr100745u (2011).
100 Panicker, G., Ye, Y., Wang, D. & Unger, E. R. Characterization of the Human Cervical Mucous Proteome. *Clin Proteomics* **6**, 18-28, doi:10.1007/s12014-010-9042-3 (2010).
101 Lee, D.-C. *et al.* Protein profiling underscores immunological functions of uterine cervical mucus plug in human pregnancy. *Journal of Proteomics* **74**, 817-828, doi:https://doi.org/10.1016/j.jprot.2011.02.025 (2011).
102 Keats, B. J. & Berlin, C. I. Genomics and hearing impairment. *Genome Res* **9**, 7-16, doi:10.1101/gr.9.1.7 (1999).
103 Dubaissi, E. *et al.* A secretory cell type develops alongside multiciliated cells, ionocytes and goblet cells, and provides a protective, anti-infective function in the frog embryonic mucociliary epidermis. *Development* **141**, 1514-1525, doi:10.1242/dev.102426 (2014).
104 Dubaissi, E. *et al.* Functional characterization of the mucus barrier on the Xenopus tropicalis skin surface. *Proc Natl Acad Sci U S A* **115**, 726-731, doi:10.1073/pnas.1713539115 (2018).
105 Bi, M., Hickox, J. R., Winfrey, V. P., Olson, G. E. & Hardy, D. M. Processing, localization and binding activity of zonadhesin suggest a function in sperm adhesion to the zona pellucida during exocytosis of the acrosome. *Biochem J* **375**, 477-488, doi:10.1042/bj20030753 (2003).

106 Didier, R. *et al.* SCO-spondin and RF-GlyI: two designations for the same glycoprotein secreted by the subcommissural organ. *J. Neurosci. Res* **61**, 500-507, doi: 10.1002/1097-4547(20000901)61:5<500::AID-JNR4>3.0.CO;2-L (2000).
107 Cantaut-Belarif, Y., Sternberg, J. R., Thouvenin, O., Wyart, C. & Bardet, P. L. The Reissner Fiber in the Cerebrospinal Fluid Controls Morphogenesis of the Body Axis. *Curr Biol* **28**, 2479-2486.e2474, doi:10.1016/j.cub.2018.05.079 (2018).
108 Lang, T. *et al.* Searching the Evolutionary Origin of Epithelial Mucus Protein Components-Mucins and FCGBP. *Mol. Biol. Evol* **33**, 1921-1936, doi:msw066 [pii];10.1093/molbev/msw066 [doi] (2016).
109 Dong, X. *et al.* The von Willebrand factor D'D3 assembly and structural principles for factor VIII binding and concatemer biogenesis. *Blood*, doi:10.1182/blood-2018-10-876300 (2019).
110 Huang, R. H. *et al.* Assembly of Weibel-Palade body-like tubules from N-terminal domains of von Willebrand factor. *Proc. Natl. Acad. Sci. U. S. A* **105**, 482-487, doi:10.1073/pnas.0710079105. (2008).
111 Purvis, A. R. & Sadler, J. E. A covalent oxidoreductase intermediate in propeptide-dependent von Willebrand factor multimerization. *J. Biol. Chem* **279**, 49982-49988, doi:10.1074/jbc.M408727200. (2004).
112 Haberichter, S. L. von Willebrand factor propeptide: biology and clinical utility. *Blood* **126**, 1753-1761, doi:10.1182/blood-2015-04-512731 (2015).
113 Rana, A., Westein, E., Niego, B. e. & Hagemeyer, C. E. Shear-Dependent Platelet Aggregation: Mechanisms and Therapeutic Opportunities. *Frontiers in Cardiovascular Medicine* **6**, doi:10.3389/fcvm.2019.00141 (2019).
114 Han-Mou Tsai, I. I. S., Ronald L. Nagel. Shear Stress Enhances the Proteolysis of von Willebrand Factor in Normal Plasma. *Blood* **83**, 2171-2179, doi:https://doi.org/10.1182/blood.V83.8.2171.2171 (1994).
115 Sergey Okhota, I. M., Yuliya Avtaeva, Sergey Kozlov and Zufar Gabbasov,. Shear Stress-Induced Activation of von WillebrandFactor and Cardiovascular Pathology. *International Journal of Molecular Sciences* **21(20)**, doi:10.3390/ijms21207804 (2020).
116 Zhang, C., Kelkar, A. & Neelamegham, S. von Willebrand factor self-association is regulated by the shear-dependent unfolding of the A2 domain. *Blood Advances* **3**, 957-968, doi:10.1182/bloodadvances.2018030122 (2019).
117 Lenting, P. J., Christophe, O. D. & Denis, C. V. von Willebrand factor biosynthesis, secretion, and clearance: connecting the far ends. *Blood* **125**, 2019-2028, doi:10.1182/blood-2014-06-528406 (2015).
118 Katsumi, A., Tuley, E. A., Bodó, I. & Sadler, J. E. Localization of Disulfide Bonds in the Cystine Knot Domain of Human von

Willebrand Factor*. *Journal of Biological Chemistry* **275**, 25585-25594, doi:https://doi.org/10.1074/jbc.M002654200 (2000).

119 Lidell, M. E. & Hansson, G. C. Cleavage in the GDPH sequence of the C-terminal cysteine-rich part of the Human MUC5AC mucin. *Biochem. J* **399**, 121-129, doi:10.1042/BJ20060443. (2006).

120 Lidell, M. E., Johansson, M. E. V. & Hansson, G. C. An autocatalytic cleavage in the C-terminus of the human MUC2 mucin occurs at the low pH of the late secretory pathway. *J. Biol. Chem* **278**, 13944-13951, doi:10.1074/jbc.M210069200. (2003).

121 Soto, P., Zhang, J. & Carraway, K. L. Enzymatic cleavage as a processing step in the maturation of Muc4/sialomucin complex. *J Cell Biochem* **97**, 1267-1274, doi:10.1002/jcb.20718 (2006).

122 Macao, B., Johansson, D. G. A., Hansson, G. C. & Härd, T. Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nature Structural & Molecular Biology* **13**, 71-76, doi:10.1038/nsmb1035 (2006).

123 Levitin, F. *et al.* The MUC1 SEA Module Is a Self-cleaving Domain*. *Journal of Biological Chemistry* **280**, 33374-33386, doi:https://doi.org/10.1074/jbc.M506047200 (2005).

124 Pelaseyed, T. *et al.* Unfolding dynamics of the mucin SEA domain probed by force spectroscopy suggest that it acts as a cell-protective device. *Febs j* **280**, 1491-1501, doi:10.1111/febs.12144 (2013).

125 Kaczmarczyk, A., Thuveson, M. & Fries, E. Intracellular coupling of the heavy chain of pre-alpha-inhibitor to chondroitin sulfate. *J. Biol. Chem* **277**, 13578-13582, doi:10.1074/jbc.M200288200. (2002).

126 Thuveson, M. & Fries, E. The low pH in trans-Golgi triggers autocatalytic cleavage of pre-alpha -inhibitor heavy chain precursor. *J. Biol. Chem* **275**, 30996-31000, doi:10.1074/jbc.M002399200 (2000).

127 Schirrmeister, J. *et al.* Characterization of the self-cleaving effector protein NopE1 of Bradyrhizobium japonicum. *J Bacteriol* **193**, 3733-3739, doi:10.1128/JB.00437-11 (2011).

128 Osicka, R. *et al.* A novel "clip-and-link" activity of repeat in toxin (RTX) proteins from gram-negative pathogens. Covalent protein cross-linking by an Asp-Lys isopeptide bond upon calcium-dependent processing at an Asp-Pro bond. *J Biol Chem* **279**, 24944-24956, doi:10.1074/jbc.M314013200 (2004).

129 Siebold, C., Yamashita, T., Monnier, P. P., Mueller, B. K. & Pasterkamp, R. J. RGMs: Structural Insights, Molecular Regulation, and Downstream Signaling. *Trends in Cell Biology* **27**, 365-378, doi:https://doi.org/10.1016/j.tcb.2016.11.009 (2017).

130 Bell, C. H. *et al.* Structure of the repulsive guidance molecule (RGM)-neogenin signaling hub. *Science* **341**, 77-80, doi:10.1126/science.1232322 (2013).

131 Lockhart-Cairns, M. P. *et al.* Internal cleavage and synergy with twisted gastrulation enhance BMP inhibition by BMPER. *Matrix Biol* **77**, 73-86, doi:10.1016/j.matbio.2018.08.006 (2019).
132 Rehman, A. U. *et al.* Mutational Spectrum of MYO15A and the Molecular Mechanisms of DFNB3 Human Deafness. *Hum Mutat* **37**, 991-1003, doi:10.1002/humu.23042 (2016).
133 Patrick, M. E. & Egland, K. A. SUSD2 Proteolytic Cleavage Requires the GDPH Sequence and Inter-Fragment Disulfide Bonds for Surface Presentation of Galectin-1 on Breast Cancer Cells. *Int J Mol Sci* **20**, doi:10.3390/ijms20153814 (2019).
134 Johannes, L., Jacob, R. & Leffler, H. Galectins at a glance. *Journal of Cell Science* **131**, doi:10.1242/jcs.208884 (2018).
135 Kovács-Sólyom, F. *et al.* Mechanism of tumor cell-induced T-cell apoptosis mediated by galectin-1. *Immunology Letters* **127**, 108-118, doi:https://doi.org/10.1016/j.imlet.2009.10.003 (2010).
136 López-Mirabal, H. R. & Winther, J. R. Redox characteristics of the eukaryotic cytosol. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783**, 629-640, doi:https://doi.org/10.1016/j.bbamcr.2007.10.013 (2008).
137 Flajnik, M. F. Comparative analyses of immunoglobulin genes: surprises and portents. *Nature Reviews Immunology* **2**, 688-698, doi:10.1038/nri889 (2002).
138 Muñoz-Dorado, J., Marcos-Torres, F. J., García-Bravo, E., Moraleda-Muñoz, A. & Pérez, J. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. *Frontiers in Microbiology* **7**, doi:10.3389/fmicb.2016.00781 (2016).
139 Palsdottir, H. *et al.* Three-dimensional macromolecular organization of cryofixed Myxococcus xanthus biofilms as revealed by electron microscopic tomography. *J Bacteriol* **191**, 2077-2082, doi:10.1128/jb.01333-08 (2009).
140 Reichenbach, H. The ecology of the myxobacteria. *Environ Microbiol* **1**, 15-21, doi:10.1046/j.1462-2920.1999.00016.x (1999).
141 Lang, E. & Stackebrandt, E. Emended descriptions of the genera Myxococcus and Corallococcus, typification of the species Myxococcus stipitatus and Myxococcus macrosporus and a proposal that they be represented by neotype strains. Request for an Opinion. *Int J Syst Evol Microbiol* **59**, 2122-2128, doi:10.1099/ijs.0.003566-0 (2009).
142 Nakane, D., Sato, K., Wada, H., McBride, M. J. & Nakayama, K. Helical flow of surface protein required for bacterial gliding motility. *Proc Natl Acad Sci U S A* **110**, 11145-11150, doi:10.1073/pnas.1219753110 (2013).
143 Gharib, S. A. *et al.* Mapping the Lung Proteome in Cystic Fibrosis. *Journal of Proteome Research* **8**, 3020-3028, doi:10.1021/pr900093j (2009).

144 Kesimer, M. *et al.* Excess Secretion of Gel-Forming Mucins and Associated Innate Defense Proteins with Defective Mucin Un-Packaging Underpin Gallbladder Mucocele Formation in Dogs. *PLoS One* **10**, e0138988, doi:10.1371/journal.pone.0138988 (2015).
145 Yang, W. *et al.* Integrating proteomics and transcriptomics for the identification of potential targets in early colorectal cancer. *Int J Oncol* **55**, 439-450, doi:10.3892/ijo.2019.4833 (2019).
146 Griffith, O. L., Melck, A., Jones, S. J. & Wiseman, S. M. Meta-analysis and meta-review of thyroid cancer gene expression profiling studies identifies important diagnostic biomarkers. *J Clin Oncol* **24**, 5043-5051, doi:10.1200/jco.2006.06.7330 (2006).
147 Wang, Y. *et al.* FcGBP was upregulated by HPV infection and correlated to longer survival time of HNSCC patients. *Oncotarget* **8**, 86503-86514, doi:10.18632/oncotarget.21220 (2017).
148 Wang, K. *et al.* A bioinformatic analysis: the overexpression and clinical significance of FCGBP in ovarian cancer. *Aging (Albany NY)* **13**, 7416-7429, doi:10.18632/aging.202601 (2021).
149 Park, J. *et al.* Discovery and Validation of Biomarkers That Distinguish Mucinous and Nonmucinous Pancreatic Cysts. *Cancer Res* **75**, 3227-3235, doi:10.1158/0008-5472.Can-14-2896 (2015).
150 Laemmli, U. Slab gel electrophoresis: SDS–PAGE with discontinuous buffers. *Nature* **227**, 680-685 (1979).
151 Schägger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal Biochem* **199**, 223-231, doi:10.1016/0003-2697(91)90094-a (1991).
152 Lakbub, J. C., Shipman, J. T. & Desaire, H. Recent mass spectrometry-based techniques and considerations for disulfide bond characterization in proteins. *Anal Bioanal Chem* **410**, 2467-2484, doi:10.1007/s00216-017-0772-1 (2018).
153 Boldon, L., Laliberte, F. & Liu, L. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano Rev* **6**, 25661, doi:10.3402/nano.v6.25661 (2015).
154 Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1862**, 324-334, doi:https://doi.org/10.1016/j.bbagen.2017.07.020 (2018).
155 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
156 Nystrom, E. E. L. *et al.* Calcium-activated Chloride Channel Regulator 1 (CLCA1) Controls Mucus Expansion in Colon by Proteolytic Activity. *EBioMedicine* **33**, 134-143, doi:S2352-3964(18)30196-8 [pii];10.1016/j.ebiom.2018.05.031 [doi] (2018).

157 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol* **26**, 1367-1372, doi:nbt.1511 [pii];10.1038/nbt.1511 [doi] (2008).
158 Zhou, B., Baldus, I. B., Li, W., Edwards, S. A. & Gräter, F. Identification of allosteric disulfides from prestress analysis. *Biophys J* **107**, 672-681, doi:10.1016/j.bpj.2014.06.025 (2014).
159 Chivers, P. T., Prehoda, K. E. & Raines, R. T. The CXXC Motif: A Rheostat in the Active Site. *Biochemistry* **36**, 4061-4066, doi:10.1021/bi9628580 (1997).
160 Schroeder, B. O., Stange, E. F. & Wehkamp, J. Waking the wimp: redox-modulation activates human beta-defensin 1. *Gut Microbes* **2**, 262-266, doi:10.4161/gmic.2.4.17692 (2011).
161 Ijssennagger, N., van der Meer, R. & van Mil, S. W. C. Sulfide as a Mucus Barrier-Breaker in Inflammatory Bowel Disease? *Trends in Molecular Medicine* **22**, 190-199, doi:https://doi.org/10.1016/j.molmed.2016.01.002 (2016).
162 Ijssennagger, N. *et al.* Gut microbiota facilitates dietary heme-induced epithelial hyperproliferation by opening the mucus barrier in colon. *Proc Natl Acad Sci U S A* **112**, 10038-10043, doi:10.1073/pnas.1507645112 (2015).
163 De Lisle, R. C. & Borowitz, D. The cystic fibrosis intestine. *Cold Spring Harb Perspect Med* **3**, a009753, doi:10.1101/cshperspect.a009753 (2013).
164 Okayasu, I. *et al.* A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology* **98**, 694-702, doi:10.1016/0016-5085(90)90290-h (1990).
165 Higgins, L. M., Frankel, G., Douce, G., Dougan, G. & MacDonald, T. T. Citrobacter rodentium infection in mice elicits a mucosal Th1 cytokine response and lesions similar to those in murine inflammatory bowel disease. *Infect Immun* **67**, 3031-3039, doi:10.1128/iai.67.6.3031-3039.1999 (1999).