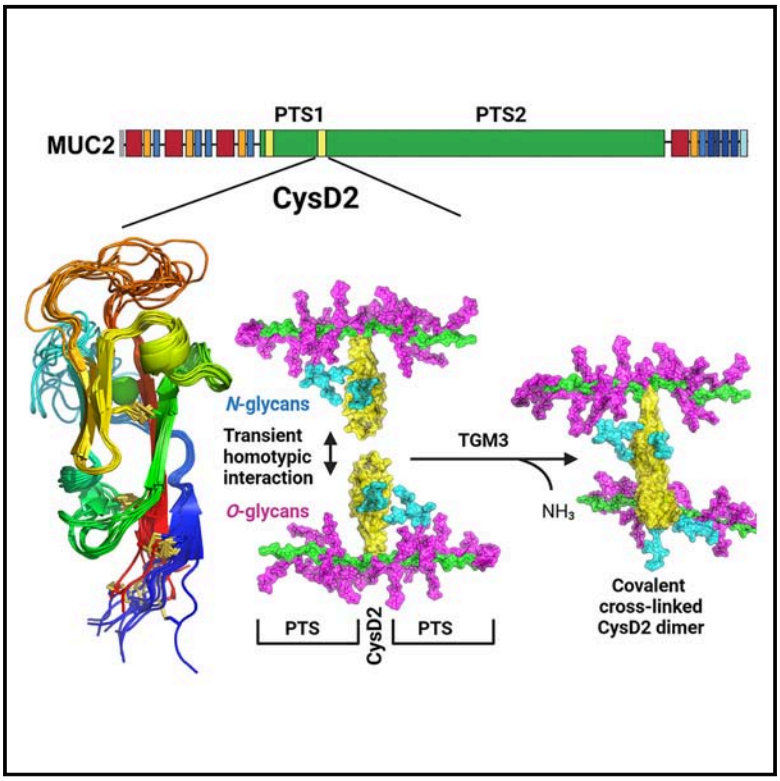
Cell Reports

Article

The structure of the second CysD domain of MUC2 and role in mucin organization by transglutaminase-based cross-linking

Graphical abstract



Highlights

- The MUC2 mucin CysD2 core structure forms a stable stalk resistant to intestinal proteases
- Three flexible surface loops are exposed outside of glycosylated domains
- CysD2 forms transient pH-dependent non-covalent dimers
- The dimers are stabilized by transglutaminase-catalyzed covalent cross-links in colon mucus

Authors

Christian V. Recktenwald, Göran Karlsson, Maria-Jose Garcia-Bonete, ..., Malin E.V. Johansson, Gunnar C. Hansson, Sergio Trillo-Muyo

Correspondence

christian.recktenwald@medkem.gu.se (C.V.R.), gunnar.hansson@medkem.gu.se (G.C.H.), sergio.muyo@medkem.gu.se (S.T.-M.)

In brief

Recktenwald et al. characterize the structure of the second CysD domain of the MUC2 mucin using NMR spectroscopy. The domain shows weak homotypic interactions at physiological pH that can be stabilized by transglutaminase 3 forming isopeptide-bond stabilized dimers in the mucus of the large intestine.

Recktenwald et al., 2024, Cell Reports *43*, 114207
May 28, 2024 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.celrep.2024.114207

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

The structure of the second CysD domain of MUC2 and role in mucin organization by transglutaminase-based cross-linking

Christian V. Recktenwald,[1,*] Göran Karlsson,[2] Maria-Jose Garcia-Bonete,[1] Gergely Katona,[3] Maja Jensen,[3] Richard Lymer,[4] Malin Bäckström,[4] Malin E.V. Johansson,[1] Gunnar C. Hansson,[1,5,*] and Sergio Trillo-Muyo[1,*]
[1]Department of Medical Biochemistry and Cell Biology, University of Gothenburg, 405 30 Gothenburg, Sweden
[2]Swedish NMR Centre, University of Gothenburg, 405 30 Gothenburg, Sweden
[3]Department of Chemistry and Molecular Biology, University of Gothenburg, 405 30 Gothenburg, Sweden
[4]Mammalian Protein Expression core facility, University of Gothenburg, 405 30 Gothenburg, Sweden
[5]Lead contact
*Correspondence: christian.recktenwald@medkem.gu.se (C.V.R.), gunnar.hansson@medkem.gu.se (G.C.H.), sergio.muyo@medkem.gu.se (S.T.-M.)
https://doi.org/10.1016/j.celrep.2024.114207

SUMMARY

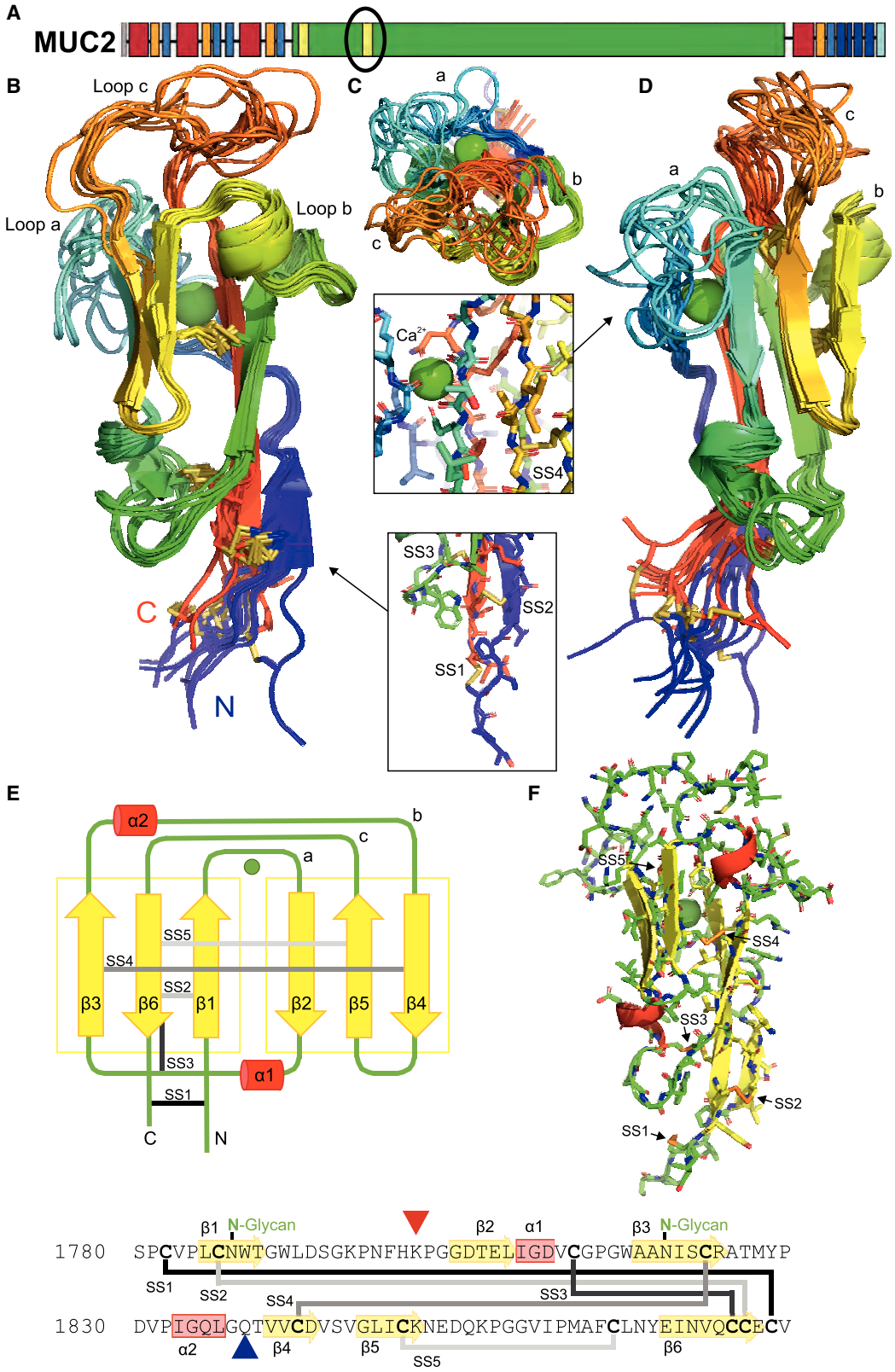
The MUC2 mucin protects the colonic epithelium by a two-layered mucus with an inner attached bacteria-free layer and an outer layer harboring commensal bacteria. CysD domains are 100 amino-acid-long sequences containing 10 cysteines that separate highly *O*-glycosylated proline, threonine, serine (PTS) regions in mucins. The structure of the second CysD, CysD2, of MUC2 is now solved by nuclear magnetic resonance. CysD2 shows a stable stalk region predicted to be partly covered by adjacent *O*-glycans attached to neighboring PTS sequences, whereas the CysD2 tip with three flexible loops is suggested to be well exposed. It shows transient dimer interactions at acidic pH, weakened at physiological pH. This transient interaction can be stabilized *in vitro* and *in vivo* by transglutaminase 3-catalyzed isopeptide bonds, preferring a specific glutamine residue on one flexible loop. This covalent dimer is modeled suggesting that CysD domains act as connecting hubs for covalent stabilization of mucins to form a protective mucus.

INTRODUCTION

Gel-forming mucins are large glycoproteins consisting of several N- and C-terminal von Willebrand D (vWD) assemblies and vWC domains separated by central proline, threonine, serine (PTS)-rich sequences that after glycosylation become mucin domains.[1] Except for the MUC6 mucin, these mucin domains are disrupted by one or several Cys-rich domains, the CysD domains. In the intestine, the MUC2 mucin is the major mucin secreted by goblet cells and thereby forms the mucus that covers the epithelium. This 5,130-amino-acid-long glycoprotein becomes heavily *O*-glycosylated during its transport through the Golgi, thereby raising the molecular mass of the primary translational product from 650 kDa to around 2.5 MDa. The molecule forms disulfide-based C-terminal dimers in the endoplasmic reticulum. Recently, it could be shown that MUC2 also forms N-terminal dimers during its way through the regulatory secretory pathway,[2] although earlier results also suggested trimers.[3] Therefore, polymers with molecular masses well above 10 MDa arise. During biosynthesis, the formation of isopeptide-based cross-links renders MUC2 insoluble.[4,5] Additional isopeptide-based cross-links are formed after secretion, required to build a protective mucus layer that is able to resist disease challenges.[6] Isopeptide bonds can be formed by transglutaminases (TGM, protein-glutamine γ-glutamyltransferases; E. C. 2.3.2.13), a family of Ca^{2+}-dependent enzymes catalyzing the deamidation or transamidation of glutamines. The latter reaction can lead to the cross-linking of proteins via the side chains of glutamine and lysine.[7] In the large intestine, TGM3 is the major isozyme and contributes to the homeostasis and stabilization of the two-layered mucus gel network.[6]

Structural analysis of mucins is difficult due to their extensive glycosylation and size. Domains of these types of molecules have been characterized by nuclear magnetic resonance (NMR) spectroscopy.[8,9] However, some of the domain types found in mucins have been revealed in the related von Willebrand factor (vWF) and in the transmembrane mucins MUC1 and MUC16 by X-ray crystallography.[10–13] Since the introduction of cryoelectron microscopy (cryo-EM) for structural studies, the situation has improved. Javitt and coworkers have shown the structure of the N-terminal part of MUC2[2] and we could recently show the C-terminal structure of MUC2.[14] At the low pH, resembling that of the Golgi apparatus, the first CysD (CysD1) domain is required for folding and packing of the filamentous structures of the MUC2 N termini.[2] Recently, the seventh CysD domain of the MUC5AC mucin was described after its structure was determined by X-ray crystallography.[15] However, the role/function of the second CysD domain (CysD2) that separates the two large

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



(legend on next page)

CellPress
OPEN ACCESS

mucin domains of MUC2 and the function of the numerous CysD domains in the other mucins are not well understood.

We have now deciphered the structure of CysD2 by NMR spectroscopy, small-angle X-ray scattering (SAXS), size exclusion chromatography, and microscale thermophoresis (MST). The results reveal that CysD2 forms dimers with lower affinities at physiological pH 7.5, but stronger at acidic pH corresponding to the intracellular granule storage conditions. In addition, we show that CysD2 serves as acyl-donor and -acceptor substrate in transglutaminase-catalyzed cross-linking reactions both *in vitro* and *in vivo*. Taken together, our observations suggest a model where the CysD domain forms transient extracellular interactions allowing mucin organization and dynamics before becoming stabilized by covalent cross-linking in the colonic mucus.

RESULTS

Initial biochemical characterization of the CysD domain

The CysD2 domain of human MUC2 consists of 97 amino acids at position 1,782 to 1,878 and separates the two large PTS sequences (Figures 1A and S1A). The domain contains 10 cysteine residues that are conserved in all CysD domains of gel-forming mucins. CysD2 was expressed as a fusion protein with N-terminal myc-Tag and a C terminus containing PTS sequences and an enterokinase cleavable immunoglobulin (Ig)G-Fc domain (Figure S1A). The recombinant protein was expressed in Lec3.2.8.1 Chinese hamster ovary (CHO) cells giving high-mannose *N*-glycans and GalNAc as the only *O*-glycan. The protein was purified by affinity chromatography using protein G, the IgG-Fc-tag cleaved off, and analyzed by gel electrophoresis (Figure S1B). Analyses of the free thiol content shows that all Cys residues are oxidized and form intramolecular disulfides (Figures S1B and S1C). The oxidized state of the Cys residues was also confirmed by NMR spectroscopy where the chemical shift values of the 10 cysteine Cβ signals ranged between 37 and 51 ppm. To characterize the different proteoforms of CysD2, intact mass spectrometry (MS) analysis of the protein was performed. The analysis detected 160 proteoforms from which 78% could be annotated to different glycosylation states of the protein (Figure S1D). The most abundant proteoform contained two *N*-glycans of the high-mannose type as well as three oxidized methionines together with an acetylated N terminus resulting in a molecular mass of 17,714 Da (Figure S1D). The mass of the different proteoforms ranged from approximately 16,000 to 19,000 Da thereby reflecting the migration behavior during SDS-PAGE (Figure S1B). However, in some proteoforms, only one of the two *N*-glycosylation sites was occupied. A common feature of all CysD domains is the sequence motif WXXW, suggested to be C-mannosylated on the first tryptophan.[16–18] However, the mass spectrometric analyses could not detect this modification.

Structure of the second CysD domain of MUC2 as revealed by NMR spectroscopy

The structure of the CysD2 domain was determined using solution NMR techniques (Table S1). Attempts to obtain isotopically labeled CysD2 domains using cell-free methods or in *E. coli,* intracellular, or in periplasm, were unsuccessful. Instead, expression and isotope labeling were performed in CHO cell cultures. Three separate samples were used for structure determination and biophysical characterization; a CysD2 form with complete $^{13}C/^{15}N$-label, one with normal media supplemented with ^{15}N-labeled valine and cysteine, and without any isotope label.

The superposition of the 10 lowest-energy structures of CysD2 revealed by NMR is shown in two side views and from the top in Figures 1B–1D. The approximately 100-amino-acids-long CysD2 domain forms a well-defined slightly elongated structure (Figure 1). The core structure consists of six β-strands (β1–6) arranged in anti-parallel β-sheets. The β-sheets are connected by three extended loops (a, b, and c) in one end of the domain (Figures 1B–1E). Loop a connects β1-β2, loop b β3-β4, and loop c β5-β6. Amide proton signals were very weak or not visible in loop regions corresponding to residues 1,796–1,804 (in loop a), 1,826–1,828 (in loop b), 1,853–1,859, and 1,863–1,865 (in loop c) (Figures 1B and 1E). Two short helices (α1 and α2) are located in the stretch connecting β2 and β3 and just before loop b. The binding site for one Ca^{2+}-ion is located in loop a.

The structure of the CysD2 domain is stabilized by five disulfide bonds, two (SS1 and SS2) of them are located in the immediate vicinity of the PTS regions. The NMR analyses showed that β1 and β6 are connected by hydrogen bonds and covalently by the disulfide bridge SS2 between Cys1786 and Cys1876. In addition, SS4 (between β3 Cys1823 and β4 Cys1842) and SS5 (between Cys1850 of β5 and Cys1866 of β6) both form cross-links between the two β-sheets. The disulfide bond SS3 (Cys1813-Cys1875) connects the region between β2 and β3 with β6. The remaining disulfide bridge (SS1) between Cys1782 and Cys1878 is connecting the N- and C-terminal ends of the domain (Figure 1E). The disulfide bonds SS1 and

Figure 1. Structure of the CysD2 domain as deciphered from NMR analyses

(A) Schematic sketch of the domains of MUC2 mucin with the two CysD domains (yellow) surrounded by PTS sequences densely *O*-glycosylated to form mucin domains (green). The analyzed CysD2 is encircled.

(B) Structure of CysD2 inside view as revealed from NMR spectroscopy, 10 isoform conformations, from N terminus (blue) to C terminus (red). The Ca^{2+} ion is shown as a green sphere. The structure is shown without sidechains. The localizations of three disulfide bonds stabilizing the stalk are shown as an inset including the amino-acid mainchain with sidechains.

(C) CysD2, top view.

(D) (B) turned anticlockwise by 90° on the y axis. The localization of the calcium ion is shown as an insert together with the SS4 disulfide bond.

(E) Schematic view of the arrangements of the six anti-parallel beta-strands (β1–β6, yellow), the two short alpha helices (α1–α2, red). The arrangement of the five disulfide bonds (SS1–SS5) as marked in the amino-acid sequence are shown at the bottom. The loops a, b, and c are marked. Gln1838 is marked by blue arrow and Lys1801 by a red arrow.

(F) Structure of CysD2 including amino-acid mainchain, the calcium ion (green sphere), and with the marked disulfide bonds (orange). β-strands are shown in yellow, α-helices in red, and loops in green.

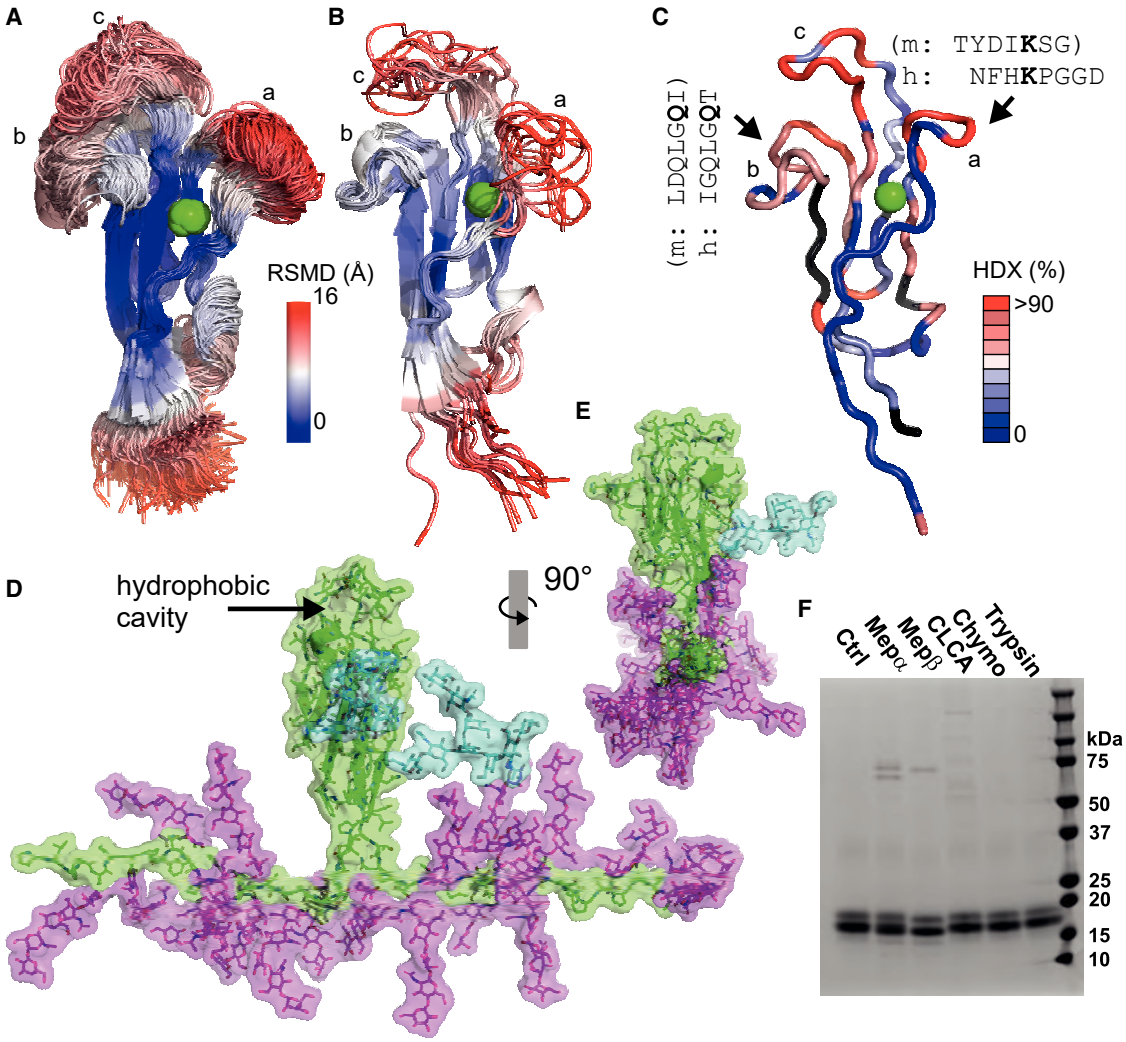


Figure 2. CysD2 has three flexible loops at its top extending outside of the mucin domain glycans

(A) Molecular dynamics of one NMR structure suggesting flexibility of the three loops a, b, and c. Superposition of 500 models generated by GROMACS, one per 100 ps, colored by RMSD (red for maximum, white for average and blue for minimum; calcium in green).

(B) Alignment of the 10 lowest-energy structures of CysD2 solved by NMR colored by RMSD as in (A).

(C) The 10 averaged NMR structures colored after analyzed by hydrogen-deuterium exchange (HDX) mass spectrometry experiments (red for maximum exchange, white for average and blue for minimum; black for no data and calcium in green).

(D) The structure of the CysD2 with neighboring PTS repeats at the N and C terminus, the protein backbone is shown in green together with the two *N*-glycans (cyan) on CysD and the *O*-glycans (magenta) on the PTS. The hydrophobic cavity is marked.

(E) (D) turned anticlockwise by 90° on y axis.

(F) CysD2 was subjected to proteolytic digestion by meprin α (Mepα), meprin β (Mepβ), CLCA1 (CLCA), chymotrypsin (Chymo), and trypsin followed by gel electrophoresis under reducing conditions. Meprin β cleaved off the myc-Tag, but this did not affect the CysD2.

SS2 are particularly important because they prevent the pulling forces of the N- and C-terminal PTS domains from separating the CysD2 core.

CysD2 has top flexible loops and extends from the mucin domain glycans

The deduced NMR structures suggest three loop regions at the top of the molecule (loops a, b, and c in Figure 1B). One of the structures was used as the starting point of molecular dynamics simulations shown in Figure 2A represented as maximum (red color), average (white), and minimum (blue) flexibility. The three loops were suggested to be flexible (red), especially loops a and c, whereas the stalk was more static (blue). The high and low flexibility regions agree well with the positional uncertainty of the 10 lowest-energy CysD2 structures as solved by NMR and as shown in Figure 2B colored similarly to the molecular dynamic simulation.

Hydrogen-deuterium exchange-mass spectrometry (HDX-MS) analysis was performed to further investigate the proton accessibility and flexibility of CysD2 (Figure S2). This shows

CellPress
OPEN ACCESS

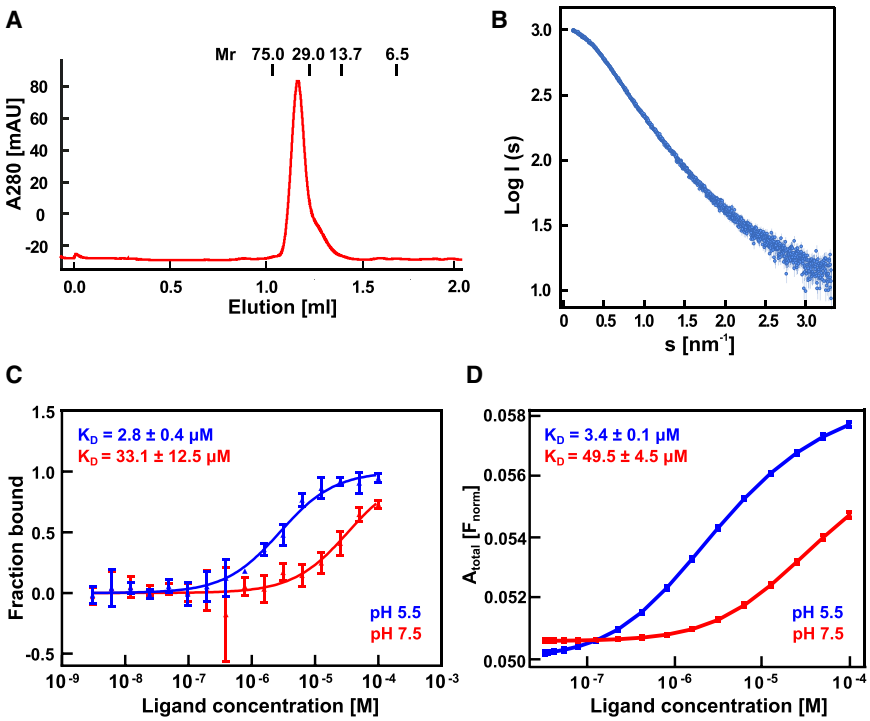


Figure 3. CysD2 can appear as a monomer or dimer depending on pH and concentration

(A) Size exclusion chromatography of CysD2 with an elution volume corresponding to an apparent molecular mass of 32 kDa, $n = 3$.

(B) Small-angle X-ray scattering analysis of CysD2, $n = 3$.

(C) Thermophoresis interaction analysis of CysD2 at pH 7.5 and 5.5, $n = 3$.

(D) Probabilistic modeling of thermophoresis progress curves in (C) by interaction analysis of CysD2 at pH 7.5 and 5.5. The median of $A_{total,n}$ posterior distribution is displayed, with error bars representing the 95% highest density interval of the distribution.

that the amide protons of the β-sheets have almost no H/D exchange as these are stabilized by 20 predicted hydrogen bridges between amide hydrogens and carbonyl groups in the adjacent β-strand (blue in Figure 2C). Nevertheless, the protons of the three top flexible loops show a high exchange rate (a, b, c in red Figure 2C), especially loop c forming a lid on the top and loop a carrying the HKP-sequence (Figure 2B). Interestingly, loop b carrying the GQT-sequence shows less proton exchange compared with loops a and c, what might be caused by the partial covering by loop c.

According to the mass spectrometric analysis (Figure S1), there were two *N*-glycans on the CysD2 for the most abundant isoform. These were localized on the asparagine residues 1,787 and 1,820 (Figure 1E). When these were modeled as di-antennary *N*-glycans without terminal sialic acids and added to the CysD2 structure, they both appear on the same side of the stalk region (Figures 2D and 2E). Furthermore, CysD2 is located between the two densely *O*-glycosylated PTS sequences in the native MUC2. To understand how much of the CysD2 domain might be covered by these, we modeled the PTS sequences with 20 amino acids on each side in an extended conformation with extended Core 3 *O*-glycans (Gal-GlcNAc-Gal-GlcNAc-3GalNAc-) on every serine and threonine (Figures 2D and 2E). The length of the *O*-glycans is as found in the human intestine[19] and this model shows that the glycans extend less than half-way up on CysD2 covering most of its stalk, but still leaving its head and the flexible loops available for interactions.

CysD is protected against proteolytic attack

The MUC2 mucin is required to be relatively stable to digestive and bacterial proteases present in the intestine. This is accomplished by its decoration with protecting glycans and by numerous disulfide bonds making the molecule compact with limited flexible regions. The structural organization of the CysD2 core, stabilized with disulfide bonds, suggests this part to be resistant, but its flexible loops should be vulnerable. To test the resistance of the domain against intestinal host proteases, CysD2 was treated with meprin α, meprin β, CLCA1, trypsin, and chymotrypsin. None of these host proteases cleaved in CysD2 (Figure 2F). These results suggest that the structural arrangement of the CysD domain and the amino acid sequences of the loops do not allow proteolytic hydrolysis by the host.

CysD2 shows transient homotypic interactions

We previously suggested that CysD2 can form homodimers based on its elution profile on a size exclusion column.[20] This experiment was repeated on a Superdex75 column providing a more suitable resolution for its molecular mass. Again, the CysD2 domain eluted at an apparent molecular mass around 32 kDa, suggesting a dimer (Figure 3A). However, complementing small-angle X-ray scattering (SEC-SAXS) with size exclusion chromatography showed a molecular mass in the range of 17.1–19.0 kDa, suggesting that the CysD2 molecule was in its monomeric state at pH 7.2 (Figures 3B; Table S2). Furthermore, the NMR model showed that the molecule has more of an elongated than a globular structure. Since molecular mass estimations at size exclusion chromatography are only applicable to globular proteins, we applied the results from the SAXS experiments (radius of gyration) to correct the gel filtration estimations. This showed that CysD2 in fact eluted as a monomer when analyzed by size exclusion chromatography.

However, microscale thermophoresis (MST) studies of CysD2 detected homotypic interactions. At acidic conditions (pH 5.5), reflecting the environment in goblet cell granules during the storage of the MUC2 mucin, the dissociation constant (K_D) was estimated to be around 3 μM (Figure 3C). This suggests that the CysD2 domain could support and organize the packing and storage of the MUC2 mucin in the storage granules. When the pH was raised

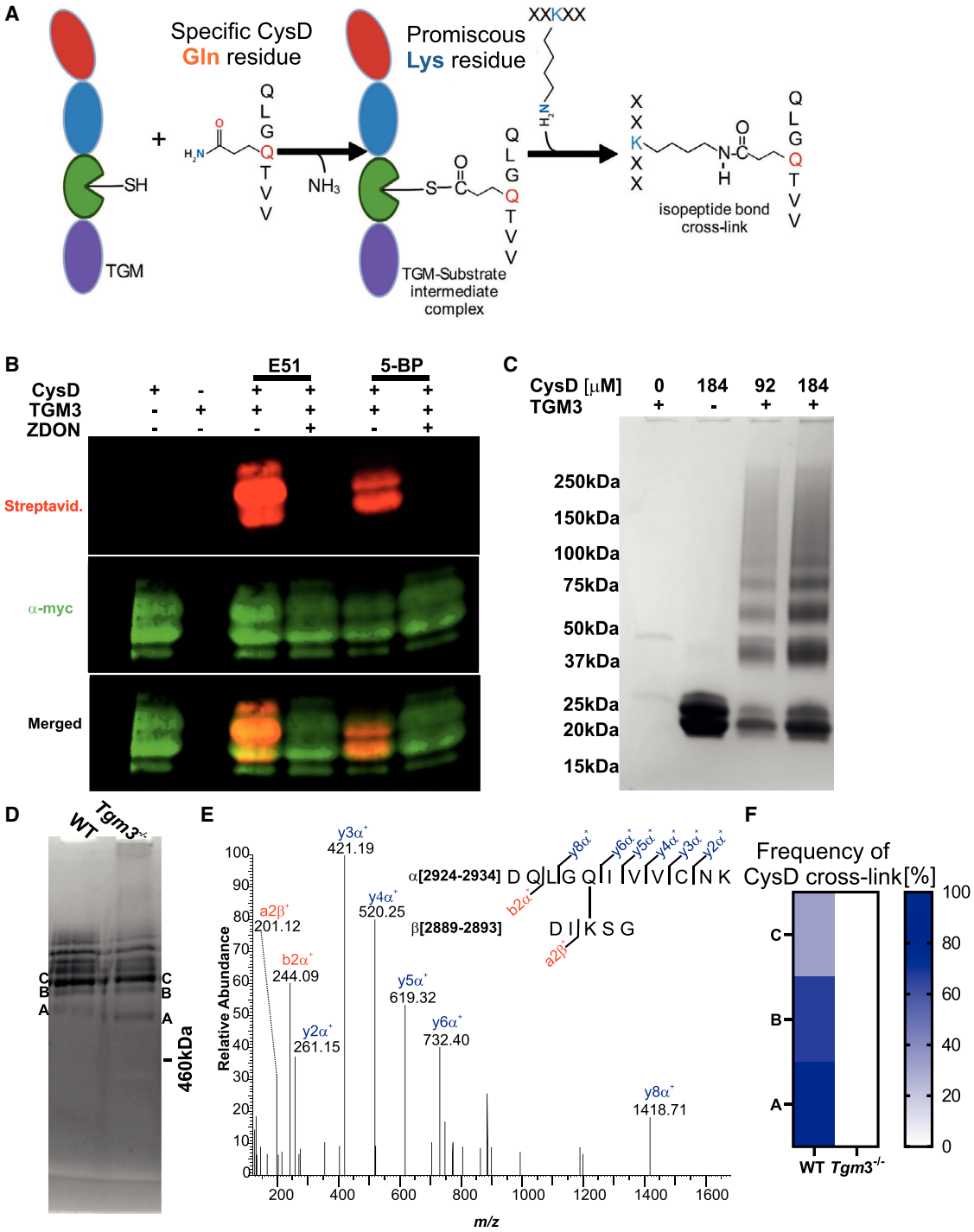


Figure 4. CysD2 as a transglutaminase substrate and presence of isopeptide bonds formed *in vitro* and *in vivo*

(A) Schematic figure explaining the two-step reaction catalyzed by transglutaminases for the formation of isopeptide bonds.

(B) CysD2 acts as both acyl- and amine-donor for the transglutaminase TGM3. CysD2 was incubated with the biotinylated acyl-donor peptide E51 or the primary amine biotinyl-cadaverine with TGM3 in the presence and absence of the transglutaminase-inhibitor ZDON, $n = 3$.

(C) TGM3 was incubated with CysD2 (with a partial PTS region and *myc*-Tag) showing the formation of multimers ($n = 3$, see Figure S3 for mass spectra).

(D) Composite Ag-PAGE analysis from the mouse colon mucus showing Muc2 from WT and Tgm3-deficient mice, respectively, $n = 3$. Muc2 was visualized by Alcian blue staining. The bands marked A, B, and C correspond to oligomeric forms of MUC2.

(legend continued on next page)

to more physiological conditions (pH 7.5),[21] the interaction was about 10 times weaker as reflected by a K_D of around 33 μM. This latter estimate is less accurate since saturation was not reached. This default target-ligand model does, however, not accurately represent equilibrium concentrations of protein dimerization experiments, especially as the maximum concentration of target-ligand complexes is constrained by the initial target concentration. To overcome this, we employed the Bayesian modeling for dimerization as implemented for MST progress curves.[22,23] The results, illustrated in Figure 3D and Table S3 confirm approximately one order of magnitude lower K_D at pH 5.5 compared with pH 7.5, but also an increased K_D to 49.5 μM at pH 7.5.

Dimerization of CysD2 at pH 7.5 was further supported by NMR analysis. Using an amino acid-specific labeled sample, ^{15}N T_1 (longitudinal) and ^{15}N T_2 (transverse) relaxation times for valines and cysteines were determined (Table S4). The rotational correlation time of the entire domain (τ_c) is related to its molecular size and shape. A τ_c of 7.9 ns was calculated from the average ratio of ^{15}N T_1 and ^{15}N T_2 relaxation times at 318 K. The corresponding molecular mass, 30 kDa, is consistent with the formation of a protein dimer. The concentration of CysD2 during NMR spectroscopy was 330 μM, 7-fold higher than the K_D measured by thermophoresis. The results from thermophoresis and NMR together suggest that CysD2 can form weak transient interactions with itself at physiological pH (7.5).

The flexible loops of CysD2 exhibit a different dynamic mobility. The ^{15}N NOE results reflect a rigid core, but residues in or close to the loops (Val1812, Cys1813, Val1840, Val1844, Val1846, Cys1850, and Cys1866) exhibit a lower ^{15}N T_2 relaxation consistent with slow dynamic exchange processes. Furthermore, Val1860, at the tip of the loop c, has a very weak signal in the ^{15}N HSQC spectrum, suggesting exposure to solvent exchange. The low ^{15}N NOE value of 0.71 (only the very N- and C-terminal residues exhibit similar low NOE values) and side chain chemical shift values are in perfect agreement with a random coil, confirming the c-loop flexibility.

CysD2 is a transglutaminase substrate resulting in covalent cross-linking of transient dimers

Transglutaminases (TGMs) are enzymes that are able to form isopeptide bonds. The reaction is initiated by the substrate interaction with the transglutaminase resulting in the formation of a thioester intermediate between a specific glutamine residue and the enzyme, followed by a second step with a nucleophilic attack of an exposed promiscuous lysine (Figure 4A).[7] The CysD2 domain has been shown to act as an acyl-donor and -acceptor substrate in a TGM2-catalyzed reaction thereby producing isopeptide bond-cross-linked CysD2 homodimers.[5] Recently, it was shown that TGM3 is the dominant transglutaminase in the large intestine and the lack of Tgm3 in the knockout mice increases the susceptibility to experimental colitis.[6] Therefore, it was tempting to speculate that CysD2 could also be a substrate for TGM3. By using the biotinylated TGM3-specific glutamine-donor substrate E51[24] and the primary amine biotinyl-cadaverine (5-BP), respectively, it was shown that CysD2 can also act as acyl-donor and -acceptor for TGM3 (Figure 4B).

Therefore, the domain should be able to form TGM3-catalyzed homodimers or homo-oligomers. In order to test this hypothesis, the protein was treated with recombinant TGM3. In this experimental set up, TGM3 catalyzed the formation of homodimers and a homo-oligomer ladder (Figure 4C). Mass spectrometric analyses of the oligomers detected several Gln-Lys isopeptide linkages (Figure S3). The preferred acyl-donor of CysD2 in most dipeptides was Gln1838 (Figure S3) that was cross-linked with alternate lysines (K1856, K1851). As the *myc*-Tag could also provide glutamine and lysine residues for cross-linking, dipeptides including the *myc*-Tag were also observed. The abundance of the different TGM3 reaction products were in similar intensity range (Figure S3).

To fully understand the importance of CysD2 cross-linking in colonic mucus, reduced and alkylated MUC2 from wild-type (WT) and *Tgm3*$^{-/-}$ mice was separated via composite agarose-PAGE (Figure 4D). As observed previously, several non-reducible bands were observed. Mass spectrometric analyses of WT-MUC2 identified a CysD2-CysD2 isopeptide cross-link between the side chains of Gln2928 and Lys2891 (Figure 4E). Gln2928 in mouse CysD2 corresponds to Gln1838 in the human sequence and Lys2891 to the human Lys1801 as marked in Figure 2C. CysD2 cross-links were always present in the faster moving WT-MUC2 bands whereas this cross-link was totally absent in the *Tgm3*$^{-/-}$ animals (Figure 4F). These results show that CysD2 functions as a substrate for TGM3 both *in vitro* and *in vivo* suggesting that this domain can act as a cross-linking hub for the organization of the bacteria-impenetrable colonic mucus layer.

According to the structure, Gln1835 is well exposed. However, MS results from both mouse colon mucus and human *in vitro* experiments suggest Gln 1838 as the preferred amino acid, although also that Gln1835 can be utilized cannot be excluded. The Gln1838 residue (LGQT sequence) in the loop b (Figures 1A and 2C) is buried and shielded by the loop c, something that should prevent the accessibility of the side chain to the catalytic cleft of TGM3. Molecular dynamics simulations (Figure 2A) showed that loop c could move to a perpendicular position, thereby exposing this residue in loop c. The flexibility of loop c was also confirmed by NMR analyses (Figure 2B). Furthermore, a shielding effect of loop b was suggested by the lower proton exchange of loop b (Figure 2C). Together, these observations suggest that this CysD2 loop and its conformation could influence the accessibility of the glutamine used for transamidation.

In theory, the Gln2928-Lys2891 (human Gln1838-Lys1801) cross-link could be intramolecular. However, the involved glutamine and the lysine side chains are too far apart in the structural model to allow such a reaction. Thus, this isopeptide cross-link must be intermolecular suggesting that the fast-moving MUC2

(E) Mass spectrum (MS2) of the isopeptide cross-link found in Muc2 from WT animals, but undetected in the *Tgm3*$^{-/-}$, n = 3. Gln2928 and Lys2891 of mouse Muc2 are the acyl-donor and acceptor sites, respectively. These amino acids correspond to the ones in humans as indicated in Figure 2C. B-ions are marked in red and y-ions are marked in blue. The parent ion showed an m/z $[M+2H]^{2+}$ of 887.96. Cleaved by trypsin/AspN.

(F) Frequency of isopeptide cross-links involving CysD2 in the different oligomeric forms of Muc2 in the mouse colon mucus from WT and *Tgm3*$^{-/-}$ mice (n = 3). The A, B and C mark refer to the bands in D.

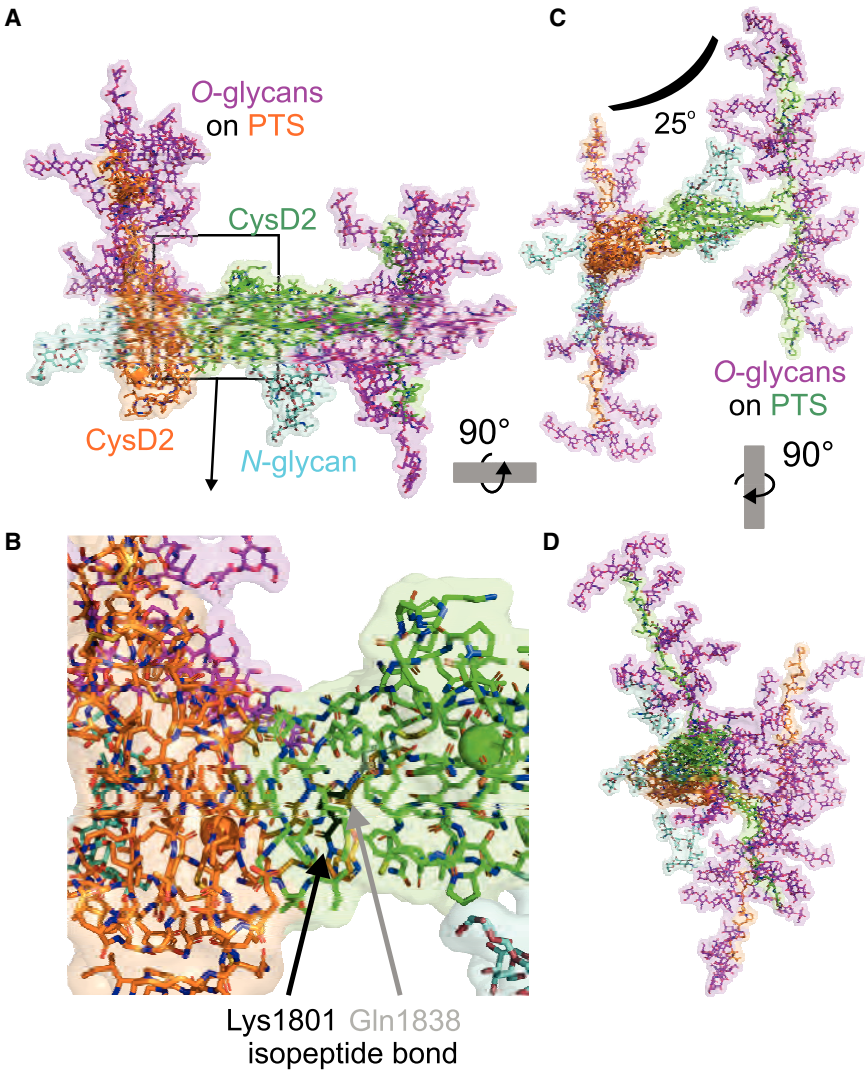


Figure 5. *In silico* model of CysD2 dimer stabilized by an isopeptide bond between Gln1838 and Lys1801

(A) Two CysD2 domains with two PTS repeats at N and C terminus (one protein orange and one protein green) with *O*-glycans (magenta) and *N*-glycans (cyan).

(B) Model with dimer interface enlarged. The isopeptide bond (black) is between Lys1801 (black) and Gln1838 (gray), indicated by black and gray arrows.

(C) (A) turned anticlockwise by 90° on the x axis. The tilt of 25° between the two molecules is indicated by an arch.

(D) (C) turned clockwise by 90° on the y axis.

band containing this linkage corresponds at least to a dimer (Figures 4D–4F).

Next, we tried to understand how the CysD2 dimer with a Gln1838-Lys1801 isopeptide bond should look and how it will arrange the flanking PTS1 and PTS2 domains. The orientation of the generated MUC2 homodimer will have important implications for how one should envision the organization of the mucus network. Using the molecular *in silico* docking programs PatchDock and FireDock, a model of the covalent homodimer could be established (Figure 5A).[25,26] The best model was found between the NMR solution number one for the glutamine donor and NMR solution number five for the lysine. The interaction surface spans ~650 Å^2, corresponding to 11% of the total surface of a non-glycosylated CysD2. The interaction occurs between loop a and the external side of β4 and β5 of the NMR model one and the loops a and c of model five. This interaction involves 18 interfacing residues from each monomer, resulting in the formation of numerous hydrophobic interactions, eight hydrogen bonds and the described isopeptide bond. These interactions are magnified in Figure 5B where the isopeptide bond between Gln1838 and Lys1801 is marked in black. Although the stalk continuation in the PTS sequences is likely very flexible, the mucin domains of the two MUC2 molecules are suggested to arrange nearly parallel with an angle of around 25° (Figures 5C and 5D).

Together, these results suggest that the second CysD domain of MUC2 is a substrate for the transglutaminase TGM3 *in vitro* and *in vivo*, thereby transforming a weak interaction into a covalent bond. This suggests an important role of the CysD2 domain in organizing and stabilizing the MUC2 mucins to form a protective colonic mucus.

DISCUSSION

CysD domains are a common characteristic of gel-forming mucins, except for the MUC6 mucin. However, their functional role has been poorly investigated so far. Here, we have studied the structural arrangement of the second CysD domain, CysD2, from the MUC2 mucin by MS, SAXS, thermophoresis, and NMR. These analyses showed that the core region of this domain is built by six anti-parallel β-strands and two short α-helices. This arrangement is further stabilized by five disulfide bridges at the top and bottom of the molecule (Figure 1). This structural arrangement is similar to the previously published crystallographic structures of the first CysD domain of MUC2[2] and the seventh CysD domain of MUC5AC recently published.[15] The main structural difference between CysD2 and the other two structures is the position of loop c that folds back over the loop b, thereby creating a hydrophobic cavity (marked in Figure 2D). CysD2 of MUC2 also resembles the CysD7 of MUC5AC as both contain only a single Ca^{2+} ion binding side and an extra α-helix (α1). Nevertheless, α1 is shorter in CysD2 and did not protrude as it does in MUC5AC CysD7, but it is located in a very flexible region. NMR and molecular dynamics showed that the lack of the second Ca^{2+} ion in the

loop a of CysD2 makes this loop highly flexible as it was suggested for MUC5AC CysD7.

In the present work, the structure of a CysD domain was determined by NMR spectroscopy, revealing the multiple conformations of the three flexible loops. This further illustrates what could also be suggested from the crystallographic model of the CysD7 in MUC5AC.[15] The comparison between NMR and Alphafold[27] CysD2 models reveals substantial differences at these loops (Figure S4). Although the models overlie almost perfectly at the stem region and partially at loop b and c, Alphafold prediction of loop a places it outside the area covered by the 10 NMR isoforms. The prediction of a short α-helix at the beginning of loop c together with abundant hydrophobic interactions between loop a and loop c, as well as loop c and loop b, produces a compact molecule giving a misleading impression about its flexibility. Understanding the flexibility of the three loops, especially loop c, is important, as this seems to regulate both interaction and accessibility of transglutaminase substrate sites.

The MUC2 CysD1 domain in isolation was revealed by X-ray crystallography and together with the full N-terminal end of MUC2 at low pH by cryo-EM, resembling the storing conditions in the goblet cell granule. CysD1 was then shown to be required for the molecular packing of the mucin. CysD1 was binding back to the von Willebrand domains (vWD), a similar arrangement as for the von Willebrand A domain of the vWF.[2,28] It is, however, difficult to envisage that the second CysD in MUC2 as well as the 7 and 9 CysDs in MUC5B and MUC5AC, respectively, could have the same role as CysD1 during intracellular packing. Furthermore, the role of the CysD domains after mucin secretion when the pH is raised is not understood.

CysD2 is surrounded by PTS sequences that are heavily *O*-glycosylated on the serines and threonines to form mucin domains. Modeling of these mucin domains together with the CysD2 shows that more than 50% of the CysD2's surface is exposed, as only the stalk will be buried (Figures 2D and 2E). This means that the top part with the three flexible loops and the calcium ion is exposed and available for interactions.

One common feature of CysD domains is the WXXW sequence that has been believed to be a motif for C-mannosylation of the first tryptophan, as suggested by several studies.[16–18] Mannosylation of this motif is claimed to be necessary for the correct maturation and secretion of mucins, but none of these studies has directly demonstrated the presence of the glycosylated tryptophan by MS or other methods. Our current and previous studies have also been unable to detect this modification on CysD2 of MUC2.[20] In addition, Javitt et al. demonstrated that the WXXW-sequence motif is not mannosylated in CysD1 of MUC2 or CysD7 of MUC5AC.[2,15] However, this could be due to protein expression in CHO or HEK cells, but we have never observed any mannosylation in native mouse colon mucus. Interestingly, in CysD2, but not CysD1, the putative C-mannosylation site is overlapping with a recognition motif for *N*-glycosylation (NWTGW). Since this asparagine residue is glycosylated, it is tempting to speculate that the *N*-glycosylation that takes place co-translationally should prevent mannosylation of this residue in CysD2. However, this cannot be the reason for lack of mannosylation in CysD1.

The formation of CysD homo- and/or heterodimers of gel-forming mucins has been directly and indirectly suggested by several authors. Perez-Vilar and coworkers demonstrated latent interactions of CysD domains from MUC5B and MUC5AC by chemical cross-linking.[17] In addition, Oikosins from protochordates are repeats of CysD domains and used to build cages, called "houses," for trapping food, further suggesting their molecular role as interacting domains.[29] Similarly, a transgenic mouse strain expressing a poly-CysD protein increased the colon mucus barrier.[30] We have previously also suggested that CysD2 forms a non-covalent dimer based on gel electrophoresis and its elution profile from a size exclusion chromatography.[20] However, as shown here, the elongated shape of CysD led to the misinterpretation of the actual monomer as a dimer.

Nevertheless, homotypic interactions were detected when CysD2 was analyzed by thermophoresis. Here, an apparent dissociation constant of around 3 μM at acidic pH, resembling the one of goblet cell granules, suggests a role of CysD2 for the packing and storage of the MUC2 mucin before secretion. The idea of CysD domains being important for intracellular storage and packing at low pH is further supported by the work of Javitt et al. showing that the MUC2 N terminus lacking the CysD1 domain is unable to assemble in filament structures.[2]

Interestingly, CysD2 also showed interactions at pH 7.5 with a dissociation constant between 33.1 and 49.5 μM as measured by thermophoresis. This dimerization was also supported by the rotational correlation time corresponding to a dimer during NMR analyses, performed at protein concentration higher than the K_D. Furthermore, NMR studies showing signals from Cβ carbon in the two different NMR experiments, HNCACB and HN(CO)CACB, respectively, were generally weak or not observable, even at elevated temperatures, as the NMR signal intensity relaxes faster with increasing molecular mass. The weak C-beta signals are an additional observation in line with dimerization and consistent with a CysD2 dimer under these conditions (pH 7.4, 300 μM, 45°C). As expected, at lower concentrations CysD2 was monomeric during size exclusion chromatography as detected by either UV or SAXS.

The experimentally determined K_D of 33.1 μM was further analyzed by probabilistic modeling suggesting a K_D of 49.5 μM, which in conjunction with the obtained NMR observations further supports transient protein-protein interactions. As the structure of CysD2 does not show any charged patches, this allows the assumption that the dimerization is a diffusion-controlled process with an expected k_{on} value of approximately 10^8 $M^{-1}s^{-1}$. The corresponding k_{off} value of around 300 s^{-1} is fast and should allow the observation of chemical shift changes in self-titration NMR experiments. However, no distinct chemical shift changes were observed in dilution experiments down to 10 μM CysD2. It should be kept in mind that the CysD2 domain is structurally heterogeneous, both due to the *N*-linked glycosylation, but also due to the variation in the oxidation state of the methionine residues. Interestingly, two methionine residues are in loop regions, and neither of them could be sequentially assigned during the NMR studies.

Together these results suggest that CysD2 could form complexes in solution, likely important for the mucin flexibility during the formation of an organized mucus. It can be hypothesized that

weaker interactions are important to enable mucus expansion where the mucin shall expand, maybe up to 1,000 times from the intracellular packed form to the mature secreted mucus.[31] For the MUC2 mucins to form a well-structured mucus excluding bacteria as in the colon,[32] the mucin interactions should be based on specific interacting domains. Weak interactions as shown here, should be ideal for a first rapid organization of the mucus.

We have shown that the MUC2 CysD2 domain is a substrate for tissue transglutaminases and that it binds to the TGM2 enzyme with a dissociation constant K_D of around 110 nM.[5] However, CysD2 seems to be a better substrate for TGM3, the most abundant isozyme in the colon, as demonstrated here by the catalysis of homo-, di-, and oligomers. The formation of oligo- and multimers is in line with a model of CysD multimers that was predicted earlier.[5] When the cross-linking reaction was catalyzed *in vitro* by TGM3, several transamidated peptides were detected. The most abundant dipeptides were between Gln1838 and either Lys1851 or Lys1856. According to the CysD2 structure, the side chain of Gln1838 is buried in the molecule and shielded by loop c. However, molecular dynamic studies showed that this loop can change its position and thereby expose Gln1838, something that could be forced by the binding of TGM3 and its reaction with the sidechain of this Gln. The Gln1835 is more exposed on CysD2 and our results cannot exclude that this Gln can also be utilized. That the Gln1838 is preferred is also suggested by the isopeptide bond between Gln2928, corresponding to Gln1838, and Lys2891 in MUC2 from colon mucus of live animals. We did not observe an isopeptide bond involving the first of the two Gln, Gln2925, in mouse MUC2. The lysine residues utilized differ between mouse and human CysD2 and might reflect sequence and conformation differences and the known promiscuity for the amine in TGM-catalyzed reactions.

In silico docking experiments suggest that two CysD2 domains could bind head-to-head, making it possible for the CysD2 to interact in a way that avoids interference from the numerous *N*- and *O*-glycans. The docking model shows that Gln1838 and Lys1801, representing the corresponding residues detected in colonic mouse mucus, become closely located to each other and that an isopeptide bond between their side chains does not cause major steric penalties. This CysD2-dimer is shown in Figure 5 where it is suggested that two CysD2 from two separate MUC2 mucins place the mucin domains relatively parallel, but with a tilt of 25° to each other. The interaction surface is large and includes 11% of the total surface of a non-glycosylated CysD2. This interaction involves the three flexible loops and the distinct hydrophobic cavity located between loops b and c.

As demonstrated by the lack of cleavage by normal intestinal proteases and two bacterial proteases (GluC and AspN, not shown), the structural organization of CysD2 provides resistance against protease-catalyzed hydrolysis. The protection of this domain to proteolysis seems to be important for the integrity of the mucus and the MUC2 mucin in the harsh environment of the gut where endogenous and exogeneous microbial proteases are highly abundant. This is reflected in previous results showing that CysD2 and the surrounding mucin domains were the only MUC2 domains not degraded in a mouse strain lacking the natural cross-linking enzyme TGM3.[6] Under steady state conditions, this mouse strain behaves normally and shows no obvious inflammation, but is more susceptible to dextran-sulfate-triggered colitis. As CysD2 is connecting the two mucin domains of MUC2, it can be assumed that cleavage of CysD2 would result in increased mucus instability, bacterial infiltration, and inflammation.

Taken together, this study elucidates the structure of the second CysD domain of MUC2 and suggests a dual physiological function. The weak and likely transient homotypic interactions at luminal pH will allow the organization of the MUC2 mucins into the observed well-organized mucus. This is especially important as the mucins come from a very densely packed intracellular form, something that requires quick expansion at the same time as it organizes into the mature mucus. In a second step, these transient CysD2 dimers can be stabilized by cross-linking through covalent bonds by the enzyme TGM3 in the colon. This will limit the pore sizes of the MUC2 network, important for excluding bacteria from entering and to make the inner colon mucus more resistant to mechanical and other forces. Thus, we suggest a model where the mucins are first highly flexible to organize and form a dynamic mucus that is then locked and stabilized in an organized way. As the MUC5B and MUC5AC mucins contain seven or nine CysD domains, it is tempting to suggest that these two mucins, abundant in the lungs, are also involved in organizing respiratory tract mucus in health and disease.[33]

Limitations of the study

The assignment of which Gln is used for cross-linking CysD2 is solely based on MS. Analyzing dipeptides with MS has its limitations, as the fragmentation ion intensities from the respective α and β peptide can be highly variable. Analysis of human CysD2 cross-linked in the test tube indicates that Gln1838 is involved, although Gln1835 might also be involved as the spectra might be a mixture of both. The NMR structure shows that the Gln1838 is less solvent accessible than Gln1835, questioning the involvement of Gln1838. However, we do not know how the interaction with TGM3 will affect the CysD2 conformation and utilization of either Gln. When we analyzed the native cross-link in mouse colonic mucus, the mass spectra strongly support the role of Gln2928, corresponding to Gln1838 in human CysD2, *in vivo*. Both Gln1835 and Gln1838 may be utilized for cross-linking, but not both at the same time. However, the specific identity of involved Gln residue is of less biological importance as any of these will transform a weaker CysD2 interaction into a covalent cross-link.

The MUC2 mucin in intestinal mucus is insoluble also in chaotropic salt, limiting the possibilities to study the effects of CysD interactions. This limits the possibilities to understand how and when the CysD cross-links appear and how these contribute to the mature mucus and its properties.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animals
  - Cell culture and protein production
- METHOD DETAILS
  - Reagents
  - Protein purification
  - NMR spectroscopy
  - Transglutaminase-catalyzed cross-linking of CysD2
  - Mass spectrometry and data analysis
  - Small-angle X-Ray scattering
  - Molecular dynamics
  - In-silico docking
  - MicroScale thermophoresis (MST)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - SAXS data analysis
  - Probabilistic modeling of MST progress curves

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.114207.

ACKNOWLEDGMENTS

This work was supported by the The Knut and Alice Wallenberg Foundation (2017.0028), European Research Council (694181), Swedish Research Council (2017-00958, 2015-06099, 2022-00646), IngaBritt and Arne Lundberg Foundation (2018-0117), Sahlgren's University Hospital (ALFGBG-440741), Bill and Melinda Gates Foundation (OPP1202459), Wilhelm and Martina Lundgren's Foundation, and EU Horizon 2020 (no. 964203). We thank the use and dedicated staff of beamline BM29 at the European Synchrotron Radiation Facility and the EMBL beamline P12, PetraIII (Instruct-ERIC, PID 5906), for their expertise during the SAXS analyses. We thank Simon Ekström (University of Lund) from the structural proteomics infrastructure unit of the SciLifeLab, Lund, for his expertise in performing the HDX-MS analysis. We thank Daniel Ambort for his input at the initiation of this project.

AUTHOR CONTRIBUTIONS

C.V.R., G.C.H., and M.E.V.J. initiated and planned the research; C.V.R. designed, performed, and analyzed experiments; G. Karlsson performed and analyzed the NMR studies; S.T.-M. and M.-J.G.-B. analyzed the structure; S.T.-M. performed docking and *in silico* studies; M.-J.G.-B. and C.V.R. prepared samples and performed and analyzed the MST experiments; G. Katona performed the Bayesian analysis of the MST data; M.-J.G.-B., M.J., and G. Katona performed and analyzed SAXS. R.L. and M.B. cultivated the cells; G.C.H. and G. Karlsson obtained funding; and C.V.R., G.C.H., and S.T.-M. wrote the manuscript with input from all authors who also approved the final text.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 22, 2023
Revised: February 15, 2024
Accepted: April 23, 2024

REFERENCES

1. Hansson, G.C. (2020). Mucins and the Microbiome. Annu. Rev. Biochem. *89*, 769–793. https://doi.org/10.1146/annurev-biochem-011520-105053.
2. Javitt, G., Khmelnitsky, L., Albert, L., Bigman, L.S., Elad, N., Morgenstern, D., Ilani, T., Levy, Y., Diskin, R., and Fass, D. (2020). Assembly Mechanism of Mucin and von Willebrand Factor Polymers. Cell *183*, 717–729.e16. https://doi.org/10.1016/j.cell.2020.09.021.
3. Ambort, D., Johansson, M.E.V., Gustafsson, J.K., Nilsson, H.E., Ermund, A., Johansson, B.R., Koeck, P.J.B., Hebert, H., and Hansson, G.C. (2012). Calcium and pH-dependent Packing and Release of the Gel-forming MUC2 Mucin. Proc. Natl. Acad. Sci. USA *109*, 5645–5650. https://doi.org/10.1073/pnas.1120269109.
4. Axelsson, M.A., Asker, N., and Hansson, G.C. (1998). O-glycosylated MUC2 monomer and dimer from LS 174T cells are water-soluble, whereas larger MUC2 species formed early during biosynthesis are insoluble and contain nonreducible intermolecular bonds. J. Biol. Chem. *273*, 18864–18870. https://doi.org/10.1074/jbc.273.30.18864.
5. Recktenwald, C.V., and Hansson, G.C. (2016). The reduction-insensitive bonds of the MUC2 mucin are isopeptide bonds. J. Biol. Chem. *291*, 13580–13590. https://doi.org/10.1074/jbc.M116.726406.
6. Sharpen, J.D.A., Dolan, B., Nyström, E.E.L., Birchenough, G.M.H., Arike, L., Martinez-Abad, B., Johansson, M.E.V., Hansson, G.C., and Recktenwald, C.V. (2022). Transglutaminase 3 crosslinks the secreted MUC2 mucin and stabilizes the colonic mucus layer. Nat. Commun. *13*, 45. https://doi.org/10.1038/s41467-021-27743-1.
7. Lorand, L., and Graham, R.M. (2003). Transglutaminases: crosslinking enzymes with pleiotropic functions. Nat. Rev. Mol. Cell Biol. *4*, 140–156. https://doi.org/10.1038/nrm1014.
8. Macao, B., Johansson, D.G.A., Hansson, G.C., and Härd, T. (2006). Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. Nat. Struct. Mol. Biol. *13*, 71–76. https://doi.org/10.1038/nsmb1035.
9. O'Leary, J.M., Hamilton, J.M., Deane, C.M., Valeyev, N.V., Sandell, L.J., and Downing, A.K. (2004). Solution Structure and Dynamics of a Prototypical Chordin-like Cysteine-rich Repeat (von Willebrand Factor Type C Module) from Collagen IIA. J. Biol. Chem. *279*, 53857–53866. https://doi.org/10.1074/jbc.M409225200.
10. Zhou, Y.F., and Springer, T.A. (2014). Highly reinforced structure of a C-terminal dimerization domain in von Willebrand factor. Blood *123*, 1785–1793. https://doi.org/10.1182/blood-2013-11-523639.
11. Dong, X., Leksa, N.C., Chhabra, E.S., Arndt, J.W., Lu, Q., Knockenhauer, K.E., Peters, R.T., and Springer, T.A. (2019). The von Willebrand factor D'D3 assembly and structural principles for factor VIII binding and concatemer biogenesis. Blood *133*, 1523–1533. https://doi.org/10.1182/blood-2013-11-523639.
12. Noguera, M.E., Jakoncic, J., and Ermácora, M.R. (2020). High-resolution structure of intramolecularly proteolyzed human mucin-1 SEA domain. Biochim. Biophys. Acta. Proteins Proteom. *1868*, 140361. https://doi.org/10.1016/j.bbapap.2020.140361.
13. White, B., Patterson, M., Karnwal, S., and Brooks, C.L. (2022). Crystal structure of a human MUC16 SEA domain reveals insight into the nature of the CA125 tumor marker. Proteins *90*, 1210–1218. https://doi.org/10.1002/prot.26303.
14. Gallego, P., Garcia-Bonete, M.J., Trillo-Muyo, S., Recktenwald, C.V., Johansson, M.E.V., and Hansson, G.C. (2023). The intestinal MUC2 mucin C-terminus is stabilized by an extra disulfide bond in comparison to von Willebrand factor and other gel-forming mucins. Nat. Commun. *14*, 1969. https://doi.org/10.1038/s41467-023-37666-8.
15. Khmelnitsky, L., Milo, A., Dym, O., and Fass, D. (2023). Diversity of CysD domains in gel-forming mucins. FEBS J. *290*, 5196–5203. https://doi.org/10.1111/febs.16918.
16. Gouyer, V., Demouveaux, B., Lacroix, G., Valque, H., Gottrand, F., and Desseyn, J.L. (2018). Non-C-mannosylable mucin CYS domains hindered proper folding and secretion of mucin. Biochem. Biophys. Res. Commun. *506*, 812–818. https://doi.org/10.1016/j.bbrc.2018.10.138.

17. Perez-Vilar, J., Randell, S.H., and Boucher, R.C. (2004). C-mannosylation of MUC5AC and MUC5B Cys subdomains. Glycobiology *14*, 325–337. https://doi.org/10.1093/glycob/cwh041.

18. Zanetta, J.P., Pons, A., Richet, C., Huet, G., Timmerman, P., Leroy, Y., Bohin, A., Bohin, J.P., Trinel, P.A., Poulain, D., and Hofsteenge, J. (2004). Quantitative gas chromatography/mass spectrometry determination of C-mannosylation of tryptophan residues in glycoproteins. Anal. Biochem. *329*, 199–206. https://doi.org/10.1016/j.ab.2004.02.033.

19. Larsson, J.M.H., Karlsson, H., Sjövall, H., and Hansson, G.C. (2009). A complex, but uniform *O*-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MS^n. Glycobiology *19*, 756–766. https://doi.org/10.1093/glycob/cwp048.

20. Ambort, D., van der Post, S., Johansson, M.E.V., MacKenzie, J., Thomsson, E., Krengel, U., and Hansson, G.C. (2011). Function of the CysD domain of the gel-forming MUC2 mucin. Biochem. J. *436*, 61–70. https://doi.org/10.1042/BJ20102066.

21. Shalon, D., Culver, R.N., Grembi, J.A., Folz, J., Treit, P.V., Shi, H., Rosenberger, F.A., Dethlefsen, L., Meng, X., and Yaffe, E. (2023). Profiling the human intestinal environment under physiological conditions. Nature *617*, 581–591. https://doi.org/10.1038/s41586-023-05989-7.

22. Anindya, A.L., Garcia-Bonete, M.-J., Jensen, M., Recktenwald, C.V., Bokarewa I, M., and Katona, G. (2022). Bayesian progress curve analysis of MicroScale thermophoresis data. Digital Discovery *1*, 325–332. https://doi.org/10.1039/D1DD00026H.

23. Jensen, M., Chandrasekaran, V., García-Bonete, M.J., Li, S., Anindya, A.L., Andersson, K., Erlandsson, M.C., Oparina, N.Y., Burmann, B.M., and Brath, U. (2023). Survivin prevents the polycomb repressor complex 2 from methylating histone 3 lysine 27. iScience *26*, 106976. https://doi.org/10.1016/j.isci.2023.106976.

24. Yamane, A., Fukui, M., Sugimura, Y., Itoh, M., Alea, M.P., Thomas, V., El Alaoui, S., Akiyama, M., and Hitomi, K. (2010). Identification of a preferred substrate peptide for transglutaminase 3 and detection of in-situ activity in skin and hair follicles. FEBS J. *277*, 3564–3574. https://doi.org/10.1111/j.1742-4658.2010.07765.x.

25. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. (2005). PatchDock and SymmDock: servers for rigid and symmetric docking. Nucleic Acids Res. *33*, W363–W367. https://doi.org/10.1093/nar/gki481.

26. Mashiach, E., Schneidman-Duhovny, D., Andrusier, N., Nussinov, R., and Wolfson, H.J. (2008). FireDock: a web server for fast interaction refinement in molecular docking. Nucleic Acids Res. *36*, W229–W232. https://doi.org/10.1093/nar/gkn186.

27. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., and otapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

28. Javitt, G., and Fass, D. (2022). Helical self-assembly of a mucin segment suggests an evolutionary origin for von Willebrand factor tubules. Proc. Natl. Acad. Sci. USA *119*, e2116790119. https://doi.org/10.1073/pnas.2116790119.

29. Spada, F., Steen, H., Troedsson, C., Kallesoe, T., Spriet, E., Mann, M., and Thompson, E.M. (2001). Molecular Patterning of the Oikoplastic Epithelium of the Larvacean Tunicate Oikopleura dioica. J. Biol. Chem. *276*, 20624–20632. https://doi.org/10.1074/jbc.M100438200.

30. Gouyer, V., Dubuquoy, L., Robbe-Masselot, C., Neut, C., Singer, E., Plet, S., Geboes, K., Desreumaux, P., Gottrand, F., and Desseyn, J.-L. (2015). Delivery of a mucin domain enriched in cysteine residues strengthens the intestinal mucous barrier. Sci. Rep. *5*, 9577. https://doi.org/10.1038/srep09577.

31. Verdugo, P. (1991). Mucin Exocytosis. Am. Rev. Respir. Dis. *144*, S33–S37. https://doi.org/10.1164/ajrccm/144.3_pt_2.S33.

32. Johansson, M.E.V., Phillipson, M., Petersson, J., Velcich, A., Holm, L., and Hansson, G.C. (2008). The inner of the two Muc2 mucin dependent mucus layers in colon is devoid of bacteria. Proc. Natl. Acad. Sci. USA *105*, 15064–15069. https://doi.org/10.1073/pnas.0803124105.

33. Fernandez-Blanco, J.A., Fakih, D., Arike, L., Rodriguez-Pineiro, A., Martinez-Abad, B., Skansebo, E., Jackson, S., Root, J., Singh, D., and McCrae, C. (2018). Attached stratified mucus separates bacteria from the epithelial cells in COPD lungs. J. Clin. Invest. Insights *3*, e120994. https://doi.org/10.1172/jci.insight.120994.

34. Stanley, P. (1989). Chinese Hamster Ovary cell mutants with multiple glycosylation defects for production of glycoproteins with minimal carbohydrate heterogeneity. Mol. Cell Biol. *9*, 377–383. https://doi.org/10.1164/ajrccm/144.3_pt_2.S33.

35. Gotze, M., Pettelkau, J., Fritzsche, R., Ihling, C.H., Schafer, M., and Sinz, A. (2015). Automated assignment of MS/MS cleavable cross-links in protein 3D-structure analysis. J. Am. Soc. Mass Spectrom. *26*, 83–97. https://doi.org/10.1007/s13361-014-1001-1.

36. Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., and Brutscher, B. (2013). BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J. Biomol. NMR *55*, 311–321. https://doi.org/10.1007/s10858-013-9715-0.

37. Isaksson, L., Mayzel, M., Saline, M., Pedersen, A., Rosenlöw, J., Brutscher, B., Karlsson, B.G., and Orekhov, V.Y. (2013). Highly Efficient NMR Assignment of Intrinsically Disordered Proteins: Application to B- and T Cell Receptor Domains. PLoS One *8*, e62947. https://doi.org/10.1371/journal.pone.0062947.

38. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins *59*, 687–696. https://doi.org/10.1002/prot.20449.

39. Gottstein, D., Kirchner, D.K., and Güntert, P. (2012). Simultaneous single-structure and bundle representation of protein NMR structures in torsion angle space. J. Biomol. NMR *52*, 351–364. https://doi.org/10.1007/s10858-012-9615-8.

40. Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., and Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins *77*, 114–122. https://doi.org/10.1002/prot.22570.

41. Salvatier, J., Wiecki, T.V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. PeerJ Comput. Sci. *2*, e55. https://doi.org/10.7717/peerj-cs.55.

42. Schutte, A., Ermund, A., Becker-Pauly, C., Johansson, M.E.V., Rodriguez-Pineiro, A.M., Backhed, F., Muller, S., Lottaz, D., Bond, J.S., and Hansson, G.C. (2014). Microbial Induced Meprin beta Cleavage in MUC2 Mucin and Functional CFTR Channel are Required to Release Anchored Small Intestinal Mucus. Proc. Natl. Acad. Sci. USA *111*, 12396–12401. https://doi.org/10.1073/pnas.1407597111.

43. Nystrom, E.E.L., Arike, L., Ehrencrona, E., Hansson, G.C., and Johansson, M.E.V. (2019). CLCA1 forms non-covalent oligomers in colonic mucus and has MUC2-processing properties. J. Biol. Chem. *294*, 17075–17089. https://doi.org/10.1074/jbc.RA119.009940.

44. Svensson, F., Lang, T., Johansson, M.E.V., and Hansson, G.C. (2018). The central exons of the human MUC2 and MUC6 mucins are highly repetitive and variable in sequence between individuals. Sci. Rep. *8*, 17503. https://doi.org/10.1038/s41598-018-35499-w.

45. Zhu, G., Xia, Y., Nicholson, L.K., and Sze, K.H. (2000). Protein Dynamics Measurements by TROSY-Based NMR Experiments. J. Magn. Reson. *143*, 423–426. https://doi.org/10.1006/jmre.2000.2022.

46. Garcia-Bonete, M.J., Jensen, M., Recktenwald, C.V., Rocha, S., Stadler, V., Bokarewa, M., and Katona, G. (2017). Bayesian Analysis of MicroScale Thermophoresis Data to Quantify Affinity of Protein:Protein Interactions with Human Survivin. Sci. Rep. *7*, 16816. https://doi.org/10.1038/s41598-017-17071-0.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-myc-tag	Abcam	Cat# ab9106, RRID:AB_307014
Goat-*anti*-rabbit-IgG-secondary antibody, IRDye800CW	LI-COR Biosciences	Cat# 926–32211, RRID:AB_621843
Chemicals, peptides, and recombinant proteins		
CF-488A dye protein labeling kit	Biotium, Fremont, CA	Cat#92213
MST capillaries	Nanotemper Technologies	Cat#MO-K022
Streptavidin-IR680RD	LI-COR Biosciences	Cat#926-68079
Human epidermal transglutaminase (TGM3)	Zedira GmbH	Cat#T013
Epidermal-transglutaminase-(TGM3)-substrate-peptide-E51	Zedira GmbH	Cat#B009
N-(Biotinyl)cadaverine	Zedira GmbH	Cat# B002
Trypsin	Promega	Cat#V5111
Chymotrypsin	Promega	Cat#V1061
ProCHO4 culture medium	Lonza	BE12-029Q
Geneticin	G418, Invitrogen	Cat#10131027
HiTrap Protein G HP column (1 mL)	Cytiva	Cat#29048581
Enterokinase EKMax	ThermoFisher	Cat# E18001
Mono Q column 5/50	Cytiva	Cat#17516601
$^{15}N/^{13}C$-labelled cysteine	Cambrige Isotope Laboratories	CNLM-3871
$^{15}N/^{13}C$-labelled valine	Cambrige Isotope Laboratories	CNLM-2395
$^{15}N/^{13}C$-labelled proline	Cambrige Isotope Laboratories	CNLM-436
Dialyzed FBS	Gibco	10270–106
BioExpress-6000	Cambrige Isotope Laboratories	CGM-6000-CN
Vivaspin 3K	VWR	28-9322-18
Deposited data		
SAXS results	SASBDB	SASDSG5
Mass spectrometry data	Proteome Xchange consortium	PXD029071
Ensemble of 10 best NMR structures	PDBe, www.ebi.ac.uk	8S03
Experimental models: Cell lines		
Lec3.2.8.1	P. Stanley	Stanley[34]
Recombinant DNA		
pSMCysD-IgG2a	lead contact	Ambort et al.[20]
pSMCysD(2TR)-IgG2a/His	lead contact	Ambort et al.[20]
Software and algorithms		
ATSAS package for SAXS analysis	EMBL	Version 3.0.0
GROMACS	https://doi.org/10.5281/zenodo.10017686.	Version 2022.4
PatchDock	Schneidman-Duhovny et al.[25]	N/A
FireDock	Mashiach et al.[26]	N/A
Monolith NT.115 instrument/software	NanoTemper Technologies	N/A
PyMol, The PyMOL Molecular Graphics System	Schrödinger, LLC	Version 2.0
ThermoBayes	Anindya et al.[22]	https://github.com/Katona-lab/ThermoBayes
StavroX	Gotze et al.[35]	https://stavrox.com/
Best-TROSY suite	Solyom et al.[36] and Isaksson et al., 2013[37]	N/A
NUS module in TopSpin	3.6	Bruker, Switzerland

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CCPN 2.4.2	Vranken et al.[38]	N/A
Cyana	Gottstein et al.[39]	N/A
Yasara	Krieger et al.[40]	N/A
PyMC3	Salvatier et al.[41]	https://www.pymc.io/projects/docs/learn.html
Other		
Recombinant Meprin α	Christoph Becker-Pauly, Christian-Albrechts-University zu Kiel, Germany	Schutte et al.[42]
Recombinant Meprin β	Christoph Becker-Pauly, Christian-Albrechts-University zu Kiel, Germany	Schutte et al.[42]
Recombinant CLCA1	Elisabeth Nyström, University of Gothenburg, Sweden	Nystrom et al.[43]
Human MUC2 peptide sequence	Mucin Database, www.medkem.gu.se/mucinbiology/databases	v.2.2, Svensson et al.[44]
Human MUC2 peptide sequence	www.uniprot.com	Q02817 (230222v3)
Mouse Muc2 peptide sequence	www.uniprot.com	Q80Z19 (230222v3)
900 MHz NMR Spectrometer +3 mm HCN cryoprobe	Bruker Corporation, USA	Avance III HD, Topspin 3.6
800 MHz NMR spectrometer + 3mm HCN cryoprobe,	Bruker Corporation, USA	Avance III HD Topspin 3.6

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gunnar C. Hansson (gunnar.hansson@medkem.gu.se).

Materials availability

Plasmids used in this study are available upon request.

Data and code availability

- The NMR structures were deposited in the PDB (ID: 8S03), SAXS results were deposited to SASBDB (SASDSG5) and mass spectra to Proteome Xchange consortium (PXD029071). Other data is available upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

C57/Bl6N mice (Taconic) were from our in-house breeding program, Tgm3$^{-/-}$ mice were provided from University of Rome, Italy. All procedures were approved by the ethical permits are 2285/19 by the Göteborgs djurförsöksetiska nämnd, Kammarrätten i Göteborg, Sweden. Mice were maintained at 22°C with light/dark cycles of 12 h each at a humidity of 40–70%. Animals received a standard rodent diet and water was supplied *ad libitum*. The animals were sedated and sacrificed followed by removal of the large intestine and gentle scarping of the colon mucus.

Cell culture and protein production

Lec 3.2.8.1 cells[34] grown in 10% FBS (fetal bovine serum) IMDM (Iscove's modified Dulbecco's medium; Lonza) with 1% penicillin/streptomycin were transfected with pSMCysD(2TR)-IgG2a/His using Lipofectamine 2000 (Invitrogen) and stable clones were generated 2–3 days later by adding 250 μg/mL Geneticin (Invitrogen). Clones were selected based on high secretion of fusion protein into the culture medium. One high-level expression clone was selected, recloned and used for protein production.

Adherent CHO-K1 Lec3.2.8.1 cells producing CysD(2TR)–IgG were resuspended at 0.3×10^6 cells/ml in 2% FBS ProCHO-4 medium (Lonza) with 4 mM L-glutamine, 1×ProHT supplement and 250 μg/mL Geneticin (Gibco). Cells were adapted to serum-free suspension growth within 6 weeks at 90 rev./min at 37°C in 5% CO_2. A 1.5 L perfusion culture with spin filter separation (10 μm) was performed in a 3 L bioreactor (Applikon) controlled by an ADI 1010 Bio Controller and an ADI 1025 Bio Console at 37°C, pH 6.9, 40% dissolved O_2 and 150–300 rev./min. Oxygen and CO_2 were introduced by bubble-free aeration using 9 m of 1-mm thick silicone tubing. Perfusion rate was 0.3–0.8 dilutions/day of culture volume.

METHOD DETAILS

Reagents

If not otherwise stated chemicals were bought from Sigma Aldrich. Meprin α and β were kindly provided by Prof. Christoph Becker-Pauly (Christian-Albrechts-University zu Kiel, Germany). CLCA1 was kindly provided by Dr. Elisabeth Nyström (University of Gothenburg, Sweden).

Protein purification

Spent culture medium (9.5 L) containing CysD(TR)–IgG fusion protein was filtered (0.45 μm Mini Capsule, PALL) and concentrated by Tangential Flow Filtration (Pellicon-2 system, Millipore) with two 5 kDa PLCC filters (four times in 1 L of PBS reduced to 500 mL). For Protein G purification of CysD(2TR)–IgG fusion protein, serum-containing medium (130 mL) was dialyzed (Spectra/Por Dialysis Membrane, molecular-mass cut-off of 6–8 kDa, Spectrum Laboratories) three times against 20 mM sodium phosphate buffer (pH 7) and filtered (Durapore Membrane Filter, 0.22 μm GVWP, Millipore). The supernatant from the Lec3.2.8.1 clone was cleared from cellular debris by centrifugation (4,000 x g; 30 min; RT) and the protein purified via a Protein G column after equilibration with 10 column volumes (CV) 20 mM PBS pH 7.3. The supernatant containing the Fc-CysD fusion protein was loaded and washed with 10 CV PBS. The Fc-tag was removed by digestion with 15 U Enterokinase (ThermoFischer) on the column for 12 h. CysD2 was eluted with 20 mM Tris pH 8.0. Cleaved CysD2 was further purified by ion-exchange chromatography using a Mono Q 5/50 GL column (GE Healthcare) using a gradient from 0 to 450 mM NaCl in 20 mM Tris buffer pH 8.0.

NMR spectroscopy

For selective labeling, the stable clone of Lec3.2.8.1/CysD-IgG2a/His was cultured in 5-layer flasks (Corning) in 200 mL DMEM (Gibco), 5% FBS, 20 mg/L L-proline, 30 mg/L methionine, 250 mg/L G418 (Gibco) with 63 mg/L $^{15}N/^{13}C$-labelled cysteine and 250 mg/L $^{15}N/^{13}C$-labelled valine (Cambridge Isotope laboratories, UK). Medium was harvested after ten days and the buffer changed to PBS using tangential flow filtration (Pellicon, Millipore).

For full $^{15}N/^{13}C$-labelling of all amino acids, the same clone was expanded to confluency in DMEM, 5% dialyzed FBS (Gibco, 20 mg/L proline, 250 mg/L G418, and medium was then replaced with BioExpress-6000 (CGM-6000-CN, Mammalian, Cambridge Isotope Laboratories, UK), 13.6 mg/L $^{15}N/^{13}C$-labelled proline, 5% dialyzed FBS (Gibco), 250 mg/L G418, 25 mM HEPES (Lonza). Medium was harvested after eight days and buffer changed to PBS as above and then used for further purification. The protein secreted into the growth medium was purified in two steps; Protein G followed by ion-exchange chromatography and the buffer exchanged to 10 mM phosphate buffer and 100 mM NaCl, pH 7.4 using Vivaspin 3K centrifugal concentrators. The total yield from one liter of cell culture was concentrated to 110 μL and transferred to a 3 mm Shigemi tube for NMR experiments.

Three different samples were used for the NMR analysis. The fully $^{13}C/^{15}N$-labelled sample (330 μM) was used for sequential assignments, side chain assignments, and isotopic edited 3D NOESY experiments. The Cys/Val-labelled sample (600 μM) was used to determine T_1-and T_2-relaxation and 15 N NOE data for valine and cysteine residues. A third, non-labeled sample (700 μM) was used for 2D ^1H–^1H NOESY experiments.

All NMR experiments were performed on an 800 MHz or a 900 MHz Bruker AVANCE III-HD spectrometer equipped with a 3 mm cryo-enhanced TCI probe.

Sequential assignment was performed using the Best-TROSY suite of backbone experiments.[36,37] All NMR experiments were performed in 40% non-uniform-sampling (NUS) mode. The temperature was set to 318 K to reduce solvent viscosity and improve the transversal relaxation properties. 3D isotope-edited NOESY experiments with 100 ms mixing from the Bruker library were used to obtain a ^{15}N-edited NOESY-HSQC, a ^{13}C NOESY-HSQC of the aliphatic region, and a ^{13}C NOESY-HSQC of the aromatic region. A 3D H(C)CH-spectrum with 14 ms isotropic mixing time was used for the side chain assignment.

^{15}N NOE, T_1 and T_2 relaxation times were measured at 800 MHz using TROSY type experiments.[45] In total 16 relaxation delays, including three triplicates, spanning intervals of 10-2,000 ms and 10–160 ms were used in the T_1 and T_2 experiments, respectively. Each experiment was recorded for 24 h as a pseudo 3D dataset using 40% non-uniform sampling (NUS).

Data were processed using the NUS module in TopSpin 3.6 (Bruker, Switzerland). Resonance assignment and spectral analysis were done in CCPN 2.4.2 [38]. NOEs and backbone dihedral angles obtained from backbone chemical shift analysis were used for the structure calculations. Geometric restraints defining a calcium-binding site were obtained from the crystal structure of the MUC2 CysD1 domain (6TM6, PDB) and sequence comparison with the CysD2 domain. Hydrogen bonds, obtained from non-exchanging H(N) atoms and with hydrogen bond donors identified from structure calculations without considering hydrogen bonds, were used to calculate 200 structures using Cyana.[39] Following energy minimization in Yasara,[40] the 10 best structures were deposited

in the PDB (ID: 8S03). A summary of the NMR statistics for the CysD structure are shown in Table S1. The ^{15}N T_1 and T_2 relaxation times (Table S4) and standard errors were estimated in CCPN using an exponential fit function with two parameters, peak height, and the time constant T. ^{15}N T_1 and T_2 relaxation time constants and the ^{15}N NOE values are reported in Table S4.

Transglutaminase-catalyzed cross-linking of CysD2

Purified CysD2 was incubated in 50 mM HEPES; 150 mM NaCl; 10 mM $CaCl_2$ pH 8.0 with 1 U TGM3 for 30 min at 37°C on a thermomixer set to 400 rpm. The reaction was terminated by the addition of 2x SDS-PAGE loading buffer and heating to 95°C for 5 min. Reaction products were separated via 4–20% SDS-PAGE and stained with Coomassie brilliant blue G-250. Bands of interest were cut out, destained, reduced and alkylated with iodoacetamide as described.[5] For the analysis of CysD2 acyl-donor and -acceptor sites 2 μg CysD2 were incubated with either 10 μM biotinylated-TGM3-substrate-peptide E51 or N-(Biotinyl)cadaverine (5-BP) and processed as described above. After electrophoresis the proteins were transferred to PVDF membranes and substrate incorporation in CysD2 detected with Streptavidin-IR680RD (1:20,000) after labeling of CysD2 with an anti-myc-Tag-antibody (1:500) followed by staining with anti-rabbit-IgG-secondary antibody (1:20,000) coupled to IRDye800CW.

Mass spectrometry and data analysis

Dried protein bands of interest were digested with trypsin/AspN (mouse mucus) or chymotrypsin/AspN (human CysD2).[6] After peptide extraction with 50% acetonitirle/0.2% TFA and removal of buffer contaminants with in-house C18 stage tips, peptides were analyzed via nano-LC-MS/MS using a Q-Exactive HF mass spectrometer (Thermo).

For the analysis of CysD-CysD cross-links from mouse mucus a previously published dataset was mined.[6] For this and the *in vitro* human CysD2 cross-link experiments, the StavroX software tool (version 3.6.6) set to the following parameters: mass tolerance for the precursor ion of 2 ppm; tolerance for fragment ions 20 ppm; full specificity for trypsin/AspN or chymotrypsin with a maximum of three missed cleavages; Gln and Lys as cross-linking sites; composition of the cross-link minus NH_3; carbamido methylation as static modification and methionine oxidation as variable modification.

Small-angle X-Ray scattering

Small Angle X-ray Scattering (SAXS) measurements were recorded at 20°C with a momentum transfer range from 0.02 to 7.24 nm^{-1} and $\lambda = 0.124$ nm, at out at the P12 EMBL beamline at PETRA III synchrotron (DESY, Hamburg, Germany). SEC-SAXS was collected during all the elution in 20 mM phosphate buffer pH 7.2, 70 mM NaCl according to the parameters described in Table S1. Data reduction and 1D scattering intensities of the sample were done at the beamline P12. Frames selection, averaging, buffer subtraction, and extrapolation were performed using CHROMIXS and primusqt from the ATSAS 3.0.0 package. The SAXS results are deposited to SASBDB with accession SASDSG5.

Molecular dynamics

The molecular dynamics behavior of CysD2 was investigated with GROMACS 2022.4 at constant pressure and temperature (NPT). The protein topology was generated from the CHARMM27 force field and solvated with the TIP3P water model using the first NMR model. The system was built in a cubic box of 1.0 nm distance from the protein to the surface under periodic boundary conditions box, neutralized by adding three sodium ions and energy minimized. A constant temperature of 300 K for 100 ps and one bar pressure were applied for equilibration. The molecular dynamics simulation was run for 50 ns with 2 fs time step.

In-silico docking

The in-silico docking was performed based on shape complementary using PatchDock. All combinations between the 10 lowest energy structures of CysD2 solved by NMR were analyzed. A distance constraint of 2.5 Å between Gln1838 OE1 and Lys1801 NZ was applied. The highest scored complexes were further refined and evaluated using FireDock.

MicroScale thermophoresis (MST)

CysD2 was fluorescently labeled using CF-488A dye protein labeling kit (Biotium, Fremont, CA). CysD-CysD binding experiments were carried out at RT on a Monolith NT 115.1 microscale thermophoresis instrument (Nanotemper Monolith NT.115) equipped with a blue-green fluorescent reader using 20% LED and medium IR power. Fluorescently labeled CysD and non-labelled CysD were prepared in a buffer containing 50 mM Tris pH 7.5 or 50 mM MES, pH 5.5; 150 mM NaCl and 0.05% Tween 20. The dilution series ranged from 100 μM to 3.1 nM of non-fluorescent CysD2. Fluorescent-labelled CysD was added with a final concentration of 30 nM. Microscale thermophoresis traces were followed for 10 s after heat induction and the ratio difference was plotted against the concentration of the titrated interaction partner.

QUANTIFICATION AND STATISTICAL ANALYSIS

SAXS data analysis

Data reduction and 1D scattering intensities of the sample were done at the beamline P12. Frames selection, averaging, buffer subtraction, and extrapolation were performed using CHROMIXS and primusqt from the ATSAS 3.0.0 package.

Probabilistic modeling of MST progress curves

The MST data were analyzed as previously described[22,23] with the following modifications. The experiments modeled here had a preheating fluorescence recording of 3.6 s and total post-heating time trace of 19.9 s. A Bayesian model implemented in Python using the pymc3 library.[41] By using an updated model, we assume that the interaction between two labeled CysD2 molecules is the same as the interaction between labeled and unlabelled CysD2, and that labeled CysD2 is monomeric at a concentration of 30 nM. In this section, we summarize the reported model parameters and discuss the differences.

The reaction was assumed to occur between the monomers (M) and the dimer (D). Labeled and unlabelled CysD2 were assumed kinetically indistinguishable:

$$M + M = D$$

$$K_d = \frac{[M]^2}{[D]}$$

$$c_n = c_{n,unlabelled} + c_{labelled}$$

$$c_n = [M]_n + 2[D]_n$$

where c_n represents the total concentration of CysD2 molecules in the different capillaries. $[M]_n$ and $[D]_n$ represent the equilibrium concentrations. The degree of association (α) is obtained by solving the resulting quadratic equation and it is a function of c_n and K_d:

$$\alpha_n = \frac{4c_n + K_d - \sqrt{K_d^2 + 8c_nK_d}}{4c_n}$$

K_d is a global variable with the same *a priori* expectations as previously described.[46] U and B represent the total amplitude of the exponential processes for pure monomeric and pure dimeric CysD2, respectively.

$$p(K_d | lower = 10^{-9}, upper = 10^{-3}) = \frac{1}{upper - lower}$$

$$p(U|\alpha = 1, \beta = 1) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} U^{\alpha-1}(1-U)^{\beta-1}$$

$$p(B|\alpha = 1, \beta = 1) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} B^{\alpha-1}(1-B)^{\beta-1}$$

The *a priori* assumptions of U and B are identical and are described by a flat β distribution. We ensure that fluorescence cannot increase beyond the initial level due to exponential (T-jump and thermophoretic) processes by assuming positive amplitudes. At total CysD2 concentration c_n, not completely dimerized CysD2 has intermediate exponential amplitudes $A_{total,n}$ determined by the law of mass action, which are an entirely deterministic combination of the stochastic components listed above:

$$A_{total,n} = U + (B - U) \cdot \frac{4c_n + K_d - \sqrt{K_d^2 + 8c_nK_d}}{4c_n}$$

All capillaries with the same c_n were modeled with the same parameters. They were not capillary specific in the triplicates.

In our model, the experimental data is modeled as a part of a normal distribution, and the scale parameter ε is a single global variable with a lognormal *a priori* distribution. This choice is motivated by the belief that errors do not vary between capillaries.

$$p(\varepsilon|\mu = 0, \tau = 1) = \frac{1}{\varepsilon}\sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}(\ln \varepsilon - \mu)^2}$$

Each of the n thermophoretic progress curves is linked to a local random variable, and their models have two exponential components. Unlike in our previous study,[23] the linear component was not present:

$$E_{1,n}(t) = A_{1,n} e^{-\nu_{1,n} t}$$

$$E_{2,n}(t) = A_{2,n} e^{-\nu_{2,n} t}$$

Because $A_{total,n}=A_{1,n}+A_{2,n}$, the values can be linked together with a single, curve-associated random ratio parameter (R_n) varying ranging from 0 to 1. R_n can be conveniently represented by a β distribution, and a slightly asymmetric prior expectation ensures that

exponential processes with larger and smaller amplitudes are grouped together for comparison.

$$E_{1,n}(t) = R_n A_{total,n} e^{-\nu_{1,n} t}$$

$$E_{2,n}(t) = (1 - R_n) A_{total,n} e^{-\nu_{2,n} t}$$

$$p(R_n | \alpha = 2, \beta = 1) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} R_n^{\alpha-1} (1 - R_n)^{\beta-1}$$

Since the progress curves are normalized to 1 at t = −3.6 s, the final fluorescence level that the two exponential components approaches in an experiment is $I_n = 1 - A_{total,n}$.

The rate parameters were modeled as random variables using a lognormal distribution.

$$p(\upsilon_{1,n} | \mu = 0, \tau = 1) = \frac{1}{\nu_{1,n}} \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}\left(\ln \nu_{1,n} - \mu\right)^2}$$

$$p(\upsilon_{2,n} | \mu = 0, \tau = 1) = \frac{1}{\nu_{2,n}} \sqrt{\frac{\tau}{2\pi}} e^{-\frac{\tau}{2}\left(\ln \nu_{2,n} - \mu\right)^2}$$

Before t = 0 s progress curves are modeled as constant at 1:

$$p(P_n(t) | \mu = 1, \varepsilon) = \sqrt{\frac{1}{2\pi\varepsilon}} e^{-\frac{1}{2\varepsilon}(P_n(t) - \mu)^2}$$

Where μ and ε corresponds to the location and scale parameter of the normal distribution (one global scale parameter).

When $t \geq 0$ s the progress curves are modeled as:

$$p(P_n(t) | \mu = E_{1,n}(t) + E_{2,n}(t) + I_n, \varepsilon) = \sqrt{\frac{1}{2\pi\varepsilon}} e^{-\frac{1}{2\varepsilon}(P_n(t) - \mu)^2}$$

The model was scaled and the posterior parameter space was sampled as described previously,[22,23] and a total of 48000 samples were collected in four parallel chains. The last 2000 samples from the four chains were combined to yield 8000 samples for the final analysis.