



The Glycoprotease CpaA Secreted by Medically Relevant *Acinetobacter* Species Targets Multiple O-Linked Host Glycoproteins

M. Florencia Haurat,^a Nichollas E. Scott,^b Gisela Di Venanzio,^a Juvenal Lopez,^a Benjamin Pluvinae,^c Alisdair B. Boraston,^c Michael J. Ferracane,^d Mario F. Feldman^a

^aDepartment of Molecular Microbiology, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA

^bDepartment of Microbiology and Immunology, The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Parkville, VIC, Australia

^cDepartment of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada

^dDepartment of Chemistry, University of Redlands, Redlands, California, USA

ABSTRACT Glycans decorate proteins and affect their biological function, including protection against proteolytic degradation. However, pathogenic, and commensal bacteria have evolved specific glycoproteases that overcome the steric impediment posed by carbohydrates, cleaving glycoproteins precisely at their glycosylation site(s). Medically relevant *Acinetobacter* strains employ their type II secretion system (T2SS) to secrete the glycoprotease CpaA, which contributes to virulence. Previously, CpaA was shown to cleave two O-linked glycoproteins, factors V and XII, leading to reduced blood coagulation. In this work, we show that CpaA cleaves a broader range of O-linked human glycoproteins, including several glycoproteins involved in complement activation, such as CD55 and CD46. However, only CD55 was removed from the cell surface, while CD46 remained unaltered during the *Acinetobacter nosocomialis* infection assay. We show that CpaA has a unique consensus target sequence that consists of a glycosylated serine or threonine residue after a proline residue (P-S/T), and its activity is not affected by sialic acids. Molecular modeling and mutagenesis analysis of CpaA suggest that the indole ring of Trp493 and the ring of the Pro residue in the substrate form a key interaction that contributes to CpaA sequence selectivity. Similar bacterial glycoproteases have recently gained attention as tools for proteomic analysis of human glycoproteins, and CpaA appears to be a robust and attractive new component of the glycoproteomics toolbox. Combined, our work provides insight into the function and possible application of CpaA, a member of a widespread class of broad-spectrum bacterial glycoproteases involved in host-pathogen interactions.

IMPORTANCE CpaA is a glycoprotease expressed by members of the *Acinetobacter baumannii-calcoaceticus* complex, and it is the first bona fide secreted virulence factor identified in these species. Here, we show that CpaA cleaves multiple targets precisely at O-glycosylation sites preceded by a Pro residue. This feature, together with the observation that sialic acid does not impact CpaA activity, makes this enzyme an attractive tool for the analysis of O-linked human protein for biotechnical and diagnostic purposes. Previous work identified proteins involved in blood coagulation as targets of CpaA. Our work broadens the set of targets of CpaA, pointing toward additional roles in bacterium-host interactions. We propose that CpaA belongs to an expanding class of functionally defined glycoproteases that targets multiple O-linked host glycoproteins.

KEYWORDS *Acinetobacter*, CpaA, T2SS, glycobiology, glycoprotease

Citation Haurat MF, Scott NE, Di Venanzio G, Lopez J, Pluvinae B, Boraston AB, Ferracane MJ, Feldman MF. 2020. The glycoprotease CpaA secreted by medically relevant *Acinetobacter* species targets multiple O-linked host glycoproteins. mBio 11:e02033-20. <https://doi.org/10.1128/mBio.02033-20>.

Editor M. Stephen Trent, University of Georgia

Copyright © 2020 Haurat et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Mario F. Feldman, mariofeldman@wustl.edu.

Received 22 July 2020

Accepted 8 September 2020

Published 6 October 2020

Members of the *Acinetobacter baumannii-calcoaceticus* complex are a frequent cause of serious multidrug-resistant infections that are associated with high mortality, and they are a top priority for the research and development of new antimicrobial therapies (1, 2). However, the development of innovative therapies against these Gram-negative pathogens demands a better understanding of the virulence and resistance mechanisms critical to *Acinetobacter* infection. Previous work has implicated the type II secretion system (T2SS) in the pathogenesis of several *Acinetobacter* spp. in different infection models (3–8). The T2SS is a trans-envelope machine that mediates the transport of effector proteins from the periplasm to the extracellular milieu. First, T2SS effectors are translocated to the periplasmic space by either the general secretory pathway or the twin-arginine translocation (TAT) system (5). Once in the periplasm, the effectors fold into their tertiary/quaternary structure, a process occasionally facilitated by dedicated chaperones, and are subsequently targeted to the T2SS machinery for secretion (7). Although the T2SS machinery is conserved across *Acinetobacter* spp., the effector repertoire is diverse and varies from strain to strain (7, 9, 10).

Approximately 20 to 60 T2SS effectors are secreted by any given *Acinetobacter* strain, and these proteins are involved in lipid assimilation, serum resistance, colonization of various host tissues, antibiotic resistance, and biofilm formation (3, 4, 7, 11). Among these T2SS effectors, the zinc-metallo-endopeptidase CpaA is the best-characterized and most abundant effector secreted by several medically relevant *Acinetobacter* strains (7, 8, 12). We demonstrated that CpaA stability and secretion depend on the membrane-bound chaperone CpaB, which is encoded adjacent to CpaA (7, 8). The N-terminal transmembrane domain of CpaB anchors the protein to the membrane, whereas its C-terminal domain interacts directly with CpaA in the periplasm (8). Removal of the CpaB transmembrane domain results in secretion of the CpaA-CpaB complex (8). Structural studies confirmed that CpaA and CpaB strongly interact in a 1:1 ratio via a novel protease-chaperone arrangement in which CpaA surrounds CpaB (8, 13). This unusual configuration was not observed in other previously characterized T2SS chaperone/effector pairs (14). An additional interesting feature observed in the cocrystal structure is that the CpaB C-terminal tail blocks access to the CpaA catalytic site (13).

In *Acinetobacter nosocomialis* M2, deletion of *cpaA* results in virulence defects comparable to those observed in a T2SS mutant strain (8). Importantly, using a murine pneumonia model, we demonstrated that CpaA plays a crucial role in dissemination to the spleen (8). The link between CpaA and dissemination within the host is further supported by *in vitro* data showing that CpaA is able to cleave human factor V (fV) and factor XII (fXII), thus interfering with blood coagulation (12, 15). Interestingly, CpaA cleaves fXII at two O-glycosylation sites in its proline-rich region, between Pro₂₇₉Thr₂₈₀ and Pro₃₀₈Thr₃₀₉, with both Thr residues being O glycosylated, though glycans are known to protect proteins from degradation (15). Moreover, glycosylation is required for CpaA activity, as deglycosylation of fXII abrogates CpaA activity (15). However, the full scope of CpaA substrates and the basis for their recognition by CpaA remain poorly understood.

Few glycoprotein-targeting zinc metallo-endopeptidases have been structurally characterized. They belong to either the metzincin or gluzincin protease superfamilies (16, 17). Metzincin proteases have a characteristic extended zinc-binding motif (HEXX-HXXGXXH) and a conserved Met turn, whereas gluzincin proteases have a glutamate residue in the zinc-binding motif (HEXXHE) (18). We previously determined the X-ray crystal structure of CpaA (13). Key features of its catalytic domain (an $\alpha+\beta$ fold, an extended zinc-binding motif, and a conserved Met turn) revealed that CpaA belongs to the metzincin protease superfamily (13, 19). In addition to its catalytic domain, CpaA possesses four very similar β -sheet tandem repeats, all of which share a similar Ig-like folds and resemble glycan-binding domains (13, 15). Interestingly, these repeats share fold similarities to a domain present in the metzincin protease StcE, a secreted glycoprotease from enterohemorrhagic *Escherichia coli* (EHEC), though their detailed sequence and structural similarities are very limited (13, 20). StcE is also a T2SS-secreted

protease known to target highly *O*-glycosylated epithelial substrates, such as CD55 (16, 17, 21–23).

Given these reported structural similarities, we hypothesized that CpaA could target a broader number of glycosylated host proteins. In this work, we combined biochemistry, mass spectrometry (MS), molecular modeling and *in vivo* assays to demonstrate that CpaA targets multiple *O*-linked human glycoproteins. This work furthers our understanding of the interaction of CpaA with its targets and provides valuable insight into *Acinetobacter* pathobiology. Furthermore, we show that CpaA has a possible biotechnological application as a tool for glycoproteomics.

RESULTS

CpaA is a broad-spectrum glycoprotease. Previous reports indicated that CpaA targets coagulation-related glycoproteins, such as fV and fXII (12, 15). CpaA cleaves the fXII mucin-like region at two sites (15). Mucins are a family of high-molecular-weight glycoproteins composed of Pro/Thr/Ser-rich domains that are heavily decorated with long oligosaccharides (24, 25). While glycosylation is protein dependent, the average mucin is ~50% *O*-linked glycan by mass, and these glycans tend to be very heterogeneous as a result of unique extensions and branching of the core structure (24, 25). Glycoproteases targeting mucins show a wide variability in terms of substrate preferences, even among members of the same family, and therefore, prediction of their targets is not possible (16, 21–23, 26–29). To assess whether CpaA can cleave other glycoproteins, we expressed and purified CpaA and its catalytically inactive point mutant (E520A) as C-terminally His-tagged proteins in *A. nosocomialis* M2, as was previously done (8, 13). CpaA and CpaA_{E520A} were incubated overnight with several medically relevant and commercially available glycoproteins (CD55, CD46, TIM1, TIM4, and C1 esterase inhibitor [C1-INH]), for which their glycosylation is at least partially characterized. Mucins run as large and diffuse bands as a result of their heterogeneity in size and glycosylation. Remarkably, CpaA, but not catalytically inactive CpaA_{E520A}, cleaved all the tested mucin glycoproteins, as observed by gel shifts to lower molecular weight (Fig. 1A; degradation products are indicated with asterisks). We also tested the ability of CpaA to cleave two well-characterized recombinant therapeutic glycoproteins, etanercept (Enbrel) and abatacept (Orencia). These fusion proteins are highly glycosylated proteins purified from mammalian cells and contain mucin-like domains in their structures (30–32). Once again, CpaA, but not CpaA_{E520A}, cleaved both recombinant glycoproteins, as observed in Fig. 1B (degradation products are indicated with asterisks).

CpaA specifically targets *O*-linked glycoproteins, and its activity is unaffected by sialic acid. The activity of glycoproteases is differentially impacted by many factors, such as *O*-glycosylation density and glycan composition. For instance, StcE and TagA, a secreted glycoprotease from *Vibrio cholerae*, uniquely cleave heavily glycosylated proteins, such as mucins, but they cannot digest less *O*-glycosylated proteins, such as fetuin (Ft). Fetuin, also known as alpha-2-HS-glycoprotein, is decorated with up to five *O*-linked sialylated glycans and is often used as a model protein for glycosylation-related studies (22, 26, 33). In contrast, the activity of other glycoproteases, such as the gluzincin IMPa from *Pseudomonas aeruginosa*, is unaffected by the degree of *O* glycosylation (16). Like IMPa, CpaA cleaved fetuin (Fig. 1C) but not enzymatically deglycosylated fetuin (Fig. 2A). The requirement of glycosylation for CpaA activity is also supported by the observation that RNase A and bovine serum albumin (BSA) were not digested by CpaA (Fig. S1A). Furthermore, our results are in agreement with those of Waack et al., who showed that enzymatically deglycosylated fXII was not cleaved by CpaA (15).

The activity of various glycoproteases can be dependent on, inhibited by, or unaffected by the presence of sialic acid, a negatively charged 9-C sugar with key biological functions (34). For example, glycoproteases from *Mannheimia haemolytica* and *Clostridium perfringens* require sialic acid for activity, whereas other glycoproteases are inhibited by this sugar (16, 35, 36). StcE is an example of a glycoprotease that is

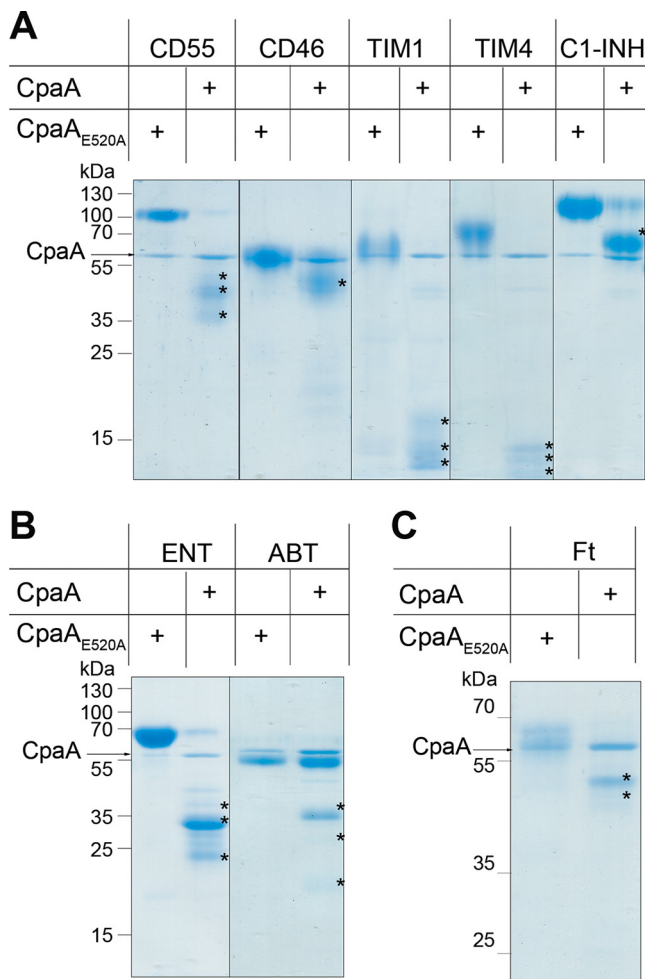


FIG 1 CpaA targets multiple proteins *in vitro*. Purified mucins (A), mucin-like proteins (B), and fetuin (C) were treated with purified CpaA and catalytically inactive CpaA (CpaA_{E520A}) (~55 kDa). Samples were separated by SDS-PAGE. Digestion products are indicated by asterisks. These data are representative of at least 3 independent experiments. C1-INH, C1 esterase inhibitor; TIM, T-cell immunoglobulin and mucin domain; ENT, etanercept; ABT, abatacept; Ft, bovine fetuin.

indifferent to sialic acid (22). Similarly to StcE, CpaA efficiently digested asialofetuin (AFt), a variant of fetuin lacking only the sialic moieties (Fig. 2B), indicating that CpaA glycoprotease activity is sialic acid independent.

All of the CpaA targets tested so far were simultaneously *N*- and *O*-glycosylated proteins. To test whether CpaA can cleave *N*-linked glycoproteins, we incubated CpaA with RNaseB, a glycoprotein modified only with *N*-linked glycans. No proteolytic activity was observed with RNaseB as a substrate (Fig. S1). Together, these results indicate that CpaA possesses broad substrate specificity, being able to cleave multiple *O*-glycosylated proteins in addition to fV and fXII, and that CpaA activity is unaffected by sialylation.

CpaA cleaves between Pro and a glycosylated Ser/Thr. We employed an MS approach to gain insight into the common molecular features that dictate CpaA substrate recognition. We excised the gel pieces containing the proteolysis products indicated with asterisks in Fig. 1, treated them with trypsin, and subsequently subjected them to MS analysis, as done previously (37). For most of the proteins, with the exceptions of TIM4 and C1-INH, we identified nontryptic peptides, resulting from CpaA activity (Fig. 3 and Fig. S2). These peptides contained an invariant C-terminal Pro residue (P1 position) (Fig. 3A and Fig. S2A, C, and D), which in the full-length protein sequences is always adjacent to a glycosylated Ser or Thr (S/T*, where the asterisk

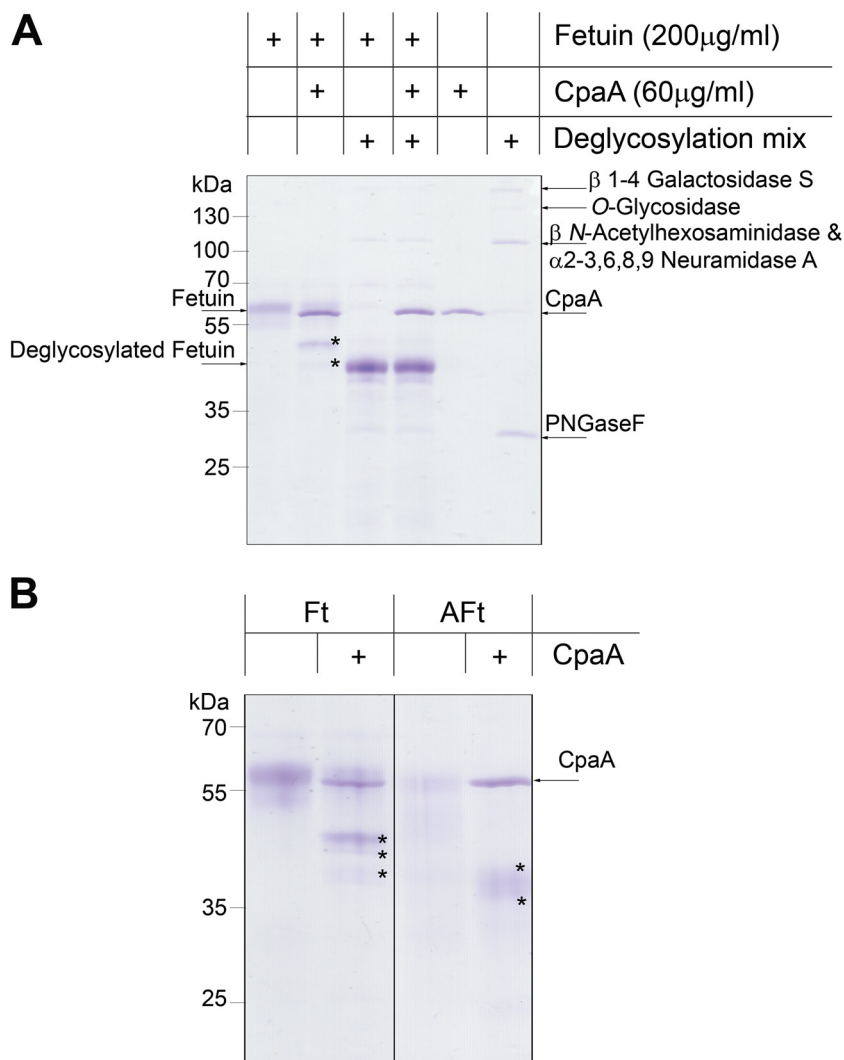


FIG 2 Fetuin glycosylation is required for CpaA activity. (A) Fetuin and fetuin treated with a deglycosylation mix were incubated with CpaA. Digestion products (indicated with asterisks) are observed only for fetuin. (B) The presence of sialic acid does not affect CpaA activity. Ft, bovine fetuin; Aft, asialofetuin.

indicates a glycosylated amino acid) (Fig. 3C). We also identified nontryptic glycopeptides derived from etanercept that contained a glycosylated N-terminal Ser (Fig. 3B and Fig. S2B). The peptides were glycosylated with either *N*-acetyl hexose-hexose-*N*-acetyl neuraminic acid (HexNAc-Hex-Neu) or HexNAc-Hex moieties (Fig. 3B and Fig. S2B), reinforcing the concept that CpaA activity is indifferent to the presence of sialic acid. The CpaA-dependent cleavage products, including those of fXII (previously reported), were used as WebLogo inputs (Fig. 3C) (15, 38). This analysis revealed that CpaA has a distinct peptide consensus sequence, P-S/T*, where cleavage occurred before the glycosylated Ser/Thr residue. As seen in Fig. 3C, the C-terminal Pro residue preceding the *O*-glycosylation site is likely a strict requirement for CpaA targeting. Human erythropoietin (EPO) contains the same Ser-bound oligosaccharide as fetuin but has an Ala residue preceding the glycosylation site. This protein was not cleaved by CpaA (Fig. S1B), further supporting the essentiality of the Pro residue preceding the *O*-glycosylation site.

Mammalian glycan array screening. The recombinant proteins tested in this study were all expressed and purified from HEK293 cells. This cell line generates proteins *O*-glycosylated with the disaccharide structure galactose β 1-3 *N*-acetylgalactosamine (Gal β 1-3GalNAc), also known as core I, as well as mono- and disialylated core I and core

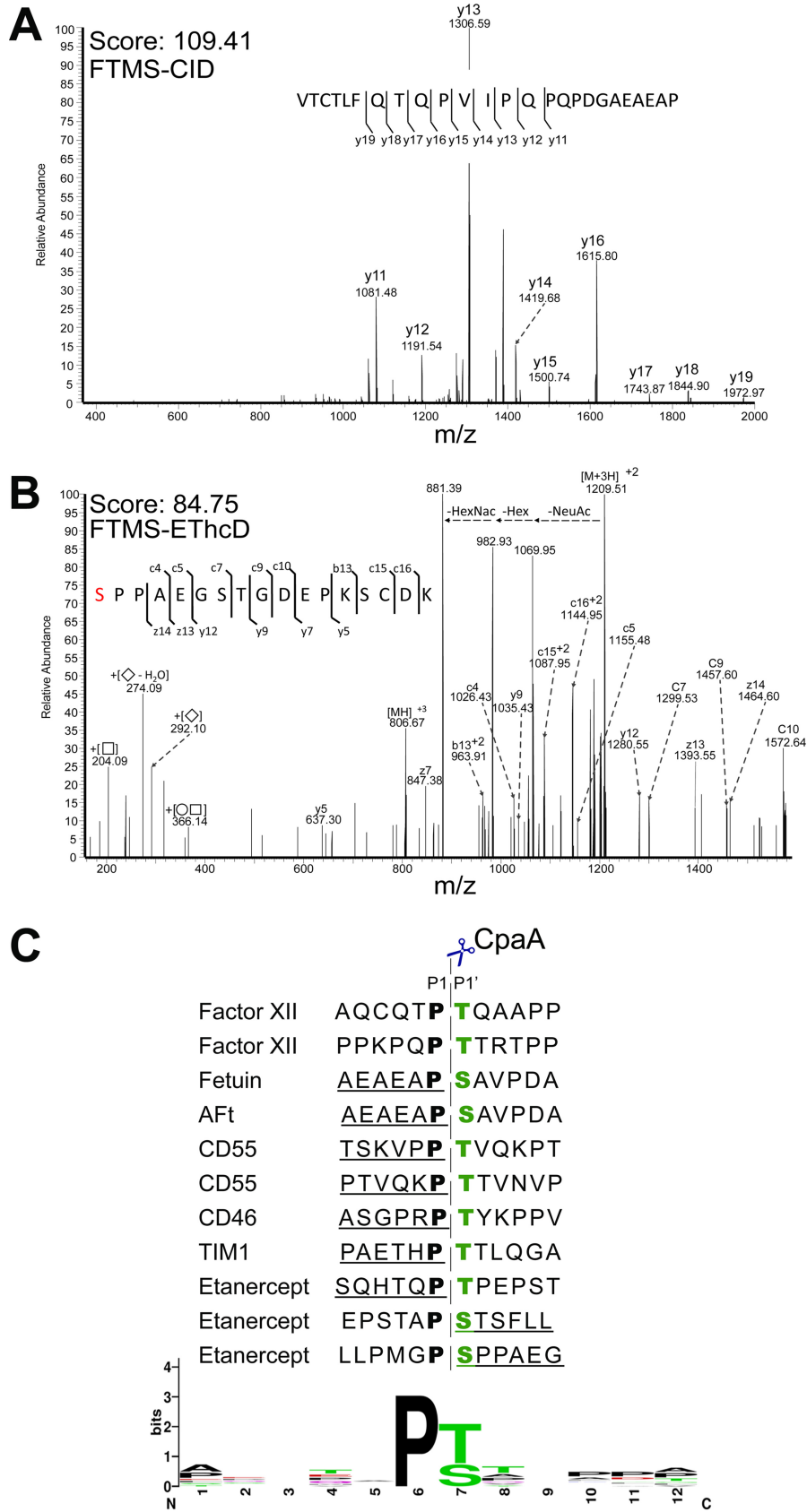


FIG 3 CpaA cleaves between Pro and a glycosylated Ser/Thr. Mass spectra of the nontryptic fragment of fetuin (A) and etanercept (B) are shown. The glycosylated Ser residue, indicated in red, is modified with (Continued on next page)

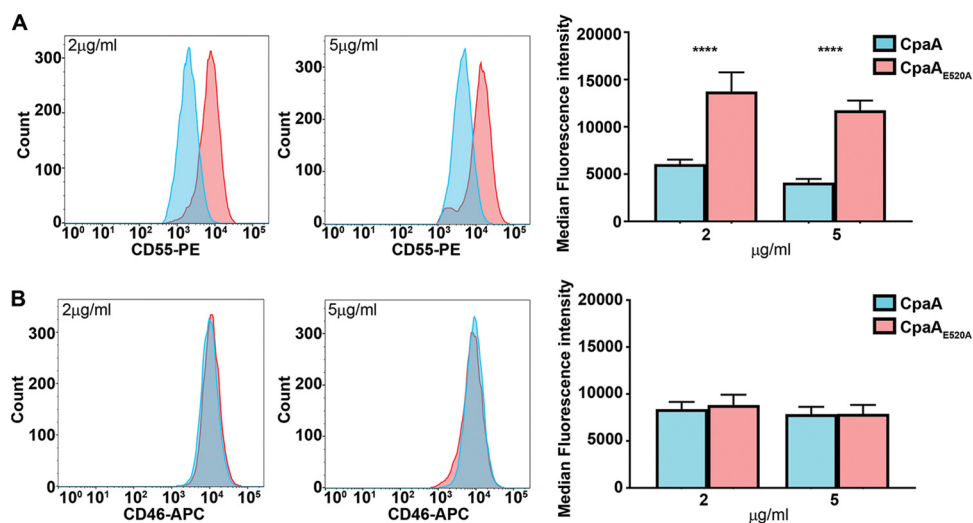


FIG 4 Purified CpaA cleaves CD55 but not CD46 from HeLa cell surfaces. Surface levels of CD55 (A) and CD46 (B) were measured by flow cytometry on HeLa cells. Cells were incubated with two different concentrations (2 µg/ml and 5 µg/ml) of purified CpaA and CpaA_{E520A} for 1 h. Left and middle panels are representative flow cytometry histograms. Right panels show the medians and standard deviations (SD) for three biological repeats. ****, $P < 0.0001$ (two-way ANOVA with Sidak's multiple-comparison test).

II (Gal β 1-3[GlcNAc β 1-6]GalNAc) structures (39). Other glycan cores have more complex (branched) structures that may or may not be accommodated by CpaA (40). To explore this further, we employed a mammalian glycan array screening assay (performed at the Consortium for Functional Glycomics) to test if CpaA can directly bind various glycans. However, we did not detect any significant peak indicative of binding for any glycan in the array when employing two protein concentrations (5 µg/ml and 50 µg/ml) of either CpaA or its catalytically inactive mutant CpaA_{E520A} (Fig. S3). A similar lack of binding has been reported for other glycoproteases, which may be due to low-affinity or transient interactions with the glycans (41). It is noteworthy that none of the glycans in the glycan array are conjugated to P-S/T sequences. Considering our previous results, it is likely that recognition by CpaA is dependent on a combination of both protein sequence and glycan structure adopted in the array.

CD55 is removed from epithelial surfaces by secreted CpaA. Our previous *in vitro* results expanded the known CpaA substrates to include human glycoproteins beyond those involved in blood coagulation (Fig. 1A). To gain insight about CpaA activity in the context of an infection, we tested whether CpaA directly cleaves surface exposed O-glycoproteins. Of all the glycoproteins tested *in vitro*, CD55 and CD46 are highly expressed cell surface O-glycoproteins in HeLa cells (42). Thus, HeLa cells were treated with purified CpaA and CpaA_{E520A} and the levels of CD55 and CD46 bound to the cell surface were quantified by flow cytometry (Fig. 4). At the two protein concentrations tested, cells treated with CpaA displayed a reduced amount of CD55 on their surfaces compared to those treated with CpaA_{E520A} (Fig. 4A). In contrast, although CpaA was able to cleave CD46 in our *in vitro* assay (Fig. 1A), levels of cell surface-exposed CD46 remained unaltered after CpaA treatment, independent of the protein concentration used (Fig. 4B).

Next, we infected HeLa cells with *A. nosocomialis* M2 expressing CpaA and CpaA_{E520A} at three different multiplicities of infection (MOI) (10, 100, and 1,000) and used flow cytometry analysis to quantify the levels of cell surface exposed CD55

FIG 3 Legend (Continued)

HexNAcHexNeuAc. (C) Sequences of the CpaA-dependent cleavage products were used as WebLogo inputs (weblogo.berkeley.edu). Factor XII cleavage sites were previously reported (15). The underlined peptide sequences were detected by mass spectrometry. The dashed line indicates the CpaA cleavage site. HexNAc, *N*-acetyl hexosamine; Hex, hexose; NeuAc, *N*-acetyl neuraminic acid.

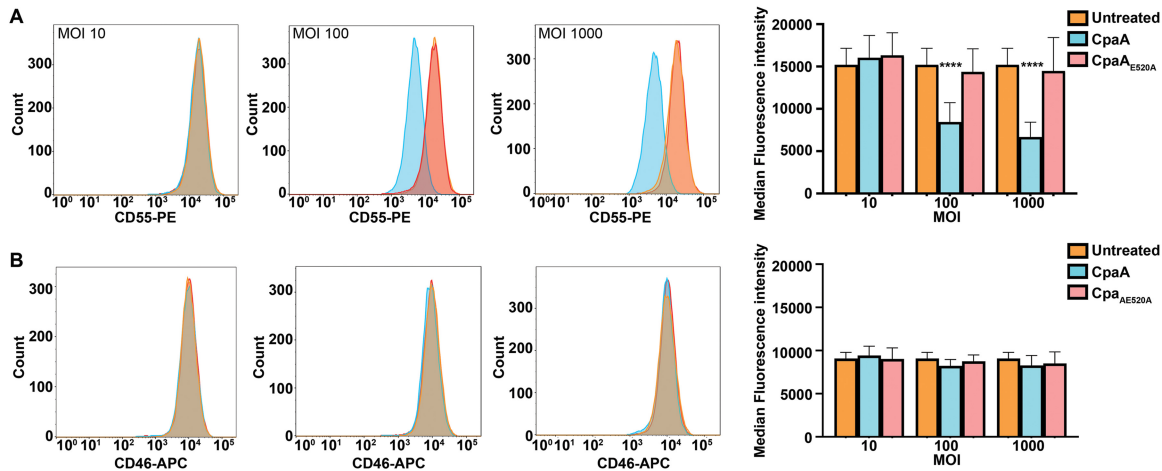


FIG 5 Secreted CpaA cleaves CD55 from HeLa cell surfaces. Surface levels of (A) CD55 and (B) CD46 were measured by flow cytometry on HeLa cells. Cells were incubated with three different MOI (10, 100, and 1,000) of *A. nosocomialis* M2 secreting either CpaA or CpaA_{E520A} for 2.5 h. Representative flow cytometry histograms are shown. Right panels show the medians and SD for three biological repeats. ****, $P < 0.0001$ (two-way ANOVA with Dunnett's multiple-comparison test).

and CD46 postinfection (Fig. 5). CD55 and CD46 levels remained unchanged when cells were infected with *A. nosocomialis* M2 secreting CpaA_{E520A}, indicating that *A. nosocomialis* M2 does not secrete any other proteases targeting either glycoprotein (Fig. 5). In agreement with our previous results, secreted CpaA cleaved CD55 but not CD46 from the cell surface (Fig. 5). The CD46 protein used in the *in vitro* assay (Fig. 1A) was expressed and purified from HEK293 cells. It is well known that glycosylation patterns differ between cell lines (43); thus, it is possible that different protein glycosylation patterns impact CpaA activity. To address this, we repeated the experiment infecting HEK293 cells. As observed with HeLa cells, secreted CpaA digested CD55 but not CD46 (Fig. S4), indicating that potential differences in glycosylation between these two cell lines do not account for these discrepancies. Importantly, an MOI of 100 was sufficient to detect cleavage of CD55 from HeLa cells, and increasing the MOI to 1,000 did not boost CD55 cleavage by CpaA (Fig. 5A). The remaining CD55 (and perhaps CD46) may be associated with proteins/ligands that prevent CpaA activity. We conclude that *A. nosocomialis* secretes physiological levels of CpaA that can digest host surface-exposed proteins during infection.

Molecular modeling identifies a putative mode of binding of glycopeptides to CpaA. We previously determined the X-ray cocrystal structure of the CpaA-CpaB complex (13), and others have obtained crystal structures of related metzincin enzymes with peptide- and peptidomimetic-based ligands (Table 1) (44, 45). X-ray structures of the related gluzincin glycopeptidases BT4244 (*Bacteroides thetaiotaomicron*), IMPa, and ZmpB (*Clostridium perfringens* ATCC 13124) have also been determined with bound glyco-amino acid and glycopeptide ligands (16). Notably, these gluzincin enzymes cleave fetuin, asialofetuin, and related synthetic glycopeptides to various degrees (Table 1). Thus, to better understand binding between CpaA and glycopeptide substrates, we performed docking experiments between a CpaA model and three different glycoforms of a fetuin-based fragment peptide (Ac-EAPSA-N-methyl [N-Me], where S is glycosylated). One of the glycoforms lacks the sialic acid moiety (Fig. 6 and Fig. S5A), while the other two are sialylated at two different position of the Gal β 1-3GalNAc core (Fig. S5B and C). The peptide portions of the docked species were able to contact the catalytic zinc ion while forming an antiparallel β -sheet with a β -strand of the active site (Fig. 6A), findings that are consistent with aforementioned metzincin crystal structures as well as docking experiments with StcE (16, 22). In the docked structures, the consensus P1 proline residue lies adjacent to W493 and is somewhat solvent exposed (Fig. 6B). An H- π interaction was observed between a proline beta hydrogen and the

TABLE 1 Summary of the biochemical and structural features of known O-glycoproteases

Glycoprotease	Source	Secretion	MEROPS	Family	Structural domain organization	Target mucins	Target fetuin	Sialic acid requirement	Recognition site	Reference(s)
CpaA	A spp.	T2SS	M72	Metzincin	IG1 IG2 IG3 IG4 CD	Yes	Yes	Indifferent	P-S/T*	This study; 13
StcE	EHEC	T2SS	M66	Metzincin	IG INS IG M-SD1 M-SD2 D1 D2 CD	Yes	No	Indifferent	S/T*X-S/T	17, 22, 33
BT_4244	Bt	Extracellular lipoprotein	M60	Gluzincin	IG1 CD IG2	Yes	No	No sialic acid	T/S-T/S*	16
IMPa	Pa	T2SS	M88	Gluzincin	$\alpha\beta$ H IG CD	Yes	Yes	Its removal enhances activity	ND	12, 29, 68
ZmpB	Cf	SP	M60	Gluzincin	IG1 CD IG2	Yes	Yes	Required	TE-T/ST/S	16
Pic	Ec	T5SS	S6	SPATE	Structure undetermined	Yes	No	Required	T/S-S/T*	51, 69, 70
ZmpC	Sp	Sortase	M26		Structure undetermined	Yes	Yes	Indifferent	T*XXX-X	51
AMUC_0627	Am	SP, extracellular, unknown	M60	Gluzincin	Structure undetermined	Yes	No	Its removal enhances activity	S/T*-S/T*	51
AMUC_0908	Am	SP, extracellular, unknown	M60	Gluzincin	Structure undetermined	Yes	No	Indifferent	T/S-T/S*	51
AMUC_1514	Am	Extracellular, unknown	M60	Gluzincin	Structure undetermined	Yes	No	No sialic acid	T/S-T/S*	51
TagA	Vc	T2SS	M66	Metzincin	Structure undetermined	Yes	No	ND	TE-T/ST/S	51
SsIE/YghJ	Ec/Sf	T2SS	M98	Gluzincin	Structure undetermined	Yes	ND	ND	ND	28, 71
O-sialoglycoprotease	Mh	Extracellular, unknown	M22	NA	Structure undetermined	Yes	No	Required	ND	70, 72

ND, not determined; NA, not applicable; CD, catalytic domain; IG, Ig-like domain; INS, insertion domain; M-SD, metalloprotease module; D, disordered region; $\alpha\beta$, α/β domain; H, helix bundle linker; A, *Acinetobacter*; Bt, *Bacteroides thetaiotaomicron*; Pa, *Pseudomonas aeruginosa*; Cf, *Clostridium perfringens*; Ec, *Escherichia coli*; Am, *Akkermansia muciniphila*; Vc, *Vibrio cholerae*; Sf, *Shigella flexneri*; Mh, *Mannheimia haemolytica*. Asterisk indicates that the amino acid residue is glycosylated.

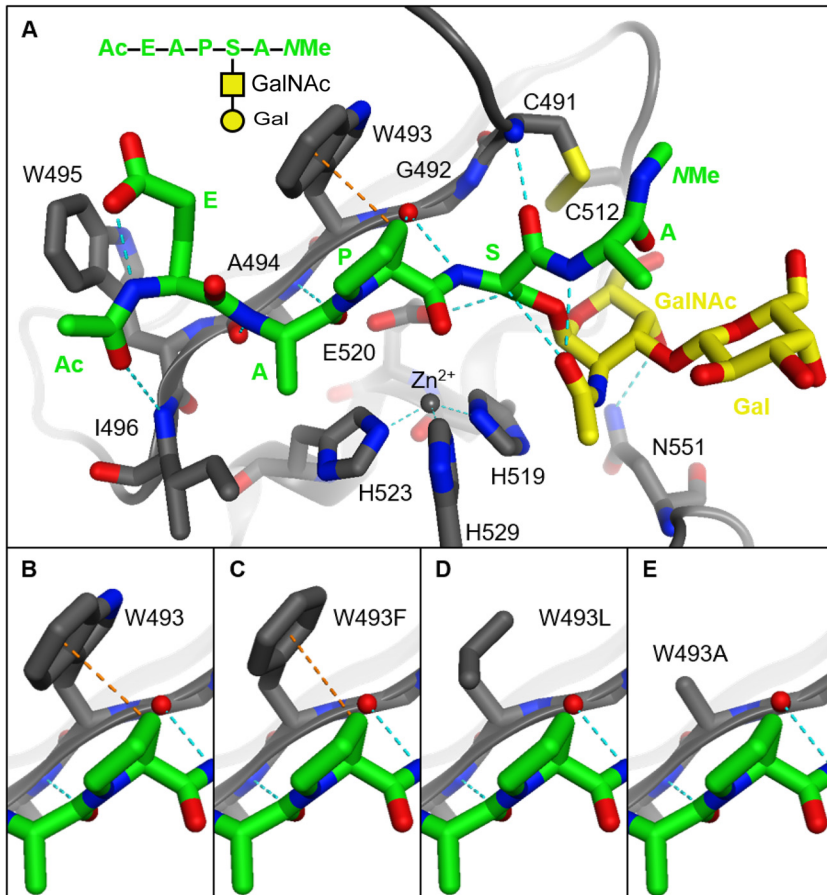


FIG 6 Structure of a fetuin-based model glycopeptide substrate docked in models of CpaA and related mutants. (A) The enzyme (gray) and the substrate's peptide portion (green) form hydrogen bonds (dashed blue lines) between their backbones similar to an antiparallel beta sheet. The catalytic histidine residues (gray sticks) and zinc ion (gray sphere) hydrolyze the amide bond between the substrate's proline (P1) and glycosylated serine (P1') residues. The GalNAc moiety (yellow sticks) forms hydrogen bonds with the substrate and the enzyme, whereas the Gal moiety (yellow sticks) is exposed to solvent. (B to E) The tryptophan residue of the native enzyme can form an H- π interaction (dashed orange lines) with the substrate. This interaction is weakened when this residue is mutated to phenylalanine and nonexistent with aliphatic residues.

tryptophan indole ring (Fig. 6B). These studies suggest that CpaA selectivity may be the result of W493 (i) forming a potential H- π interaction with the prolyl ring, (ii) minimizing the prolyl residue's exposure to solvent, and/or (iii) sterically holding the substrate in the active site.

In the docked structures, the glycan moieties were found to form few interactions with CpaA and the substrate peptide. The acetyl group of the GalNAc moieties all formed hydrogen bonds with the peptide backbone (Fig. 6A). Similar contacts were observed in docking studies using *O*-GalNAc-ylated peptides and StcE (22). Such contacts have been shown to affect the conformation of mucin-like glycopeptides (46), and they may bias substrates into an extended conformation that would be more easily recognized by the enzyme. We also found that the 4-OH of the GalNAc moieties formed hydrogen bonds with the side chain amide of N551 (Fig. 6A and Fig. S5A to C). This interaction would likely endow the enzyme with selectivity for peptides modified with GalNAc at P1'. Interestingly, the side chain of this residue occupies space similar to that of the side chains of tryptophan residues that are conserved in the related gluzincin enzymes BT4244, IMPa, and ZmpB (Fig. S5D and E). In the gluzincin enzymes, the indole nitrogens instead hydrogen bond to the acetyl groups of GalNAc ligands. Such a contact may be formed between CpaA and its substrates if minor conformational

changes occur. In either case, N551 appears to be the main residue recognizing the unique features of the GalNAc moiety, as the other residues flanking the glycans are mostly small and nonpolar. The related gluzincin enzymes, on the other hand, form several contacts between polar side chains and their GalNAc groups.

Larger, linear glycans project subsequent sugar residues into solvent; these groups are not predicted to interact with CpaA. Branched glycans are not well accommodated by the enzyme, as the disulfide bond adjacent to the active site limits the flexibility of the enzyme and the branched sugar's ability to bind. Conversely, both IMPa and ZmpB form interactions with additional sugars of their corresponding glycans. Still, only IMPa demonstrated binding activity in the mammalian glycan array screen (16), which may explain why no hits were found in the same screen with CpaA. Collectively, our docking studies indicate that CpaA interacts with both peptide and glycan components of the substrate and identify residue W493 as a potential mediator of the interaction between CpaA and the Pro residue of its substrate.

Effect of W493 on CpaA specific activity. Docking studies suggest an H- π interaction between the indole ring (of W493) of CpaA and a beta hydrogen of Pro residue (at the P1 position) of the substrate (Fig. 6B). While other substrate residues can form this contact with W493, the unique rigidity and geometry of Pro allow it to optimally present its beta hydrogen to the indole ring of W493. Thus, we hypothesized that W493 of CpaA plays a critical role in CpaA selectivity. Indeed, our molecular docking studies indicate a weaker H- π interaction when peptides were docked into a mutant W493F model (Fig. 6C). The modeled W493L and W493A mutants have aliphatic residues that are unable to form this interaction with the substrate (Fig. 6D and E). Moreover, all the mutant models formed fewer van der Waals contacts with the Pro residue and further exposed it to solvent.

To complement our molecular modeling experiments, we tested the effect of these mutations on CpaA activity. All CpaA point mutation variants were expressed and secreted at levels similar to those of CpaA and CpaA_{E520A}, and no degradation products of CpaA were observed in the whole-cell fractions of the CpaA variants (Fig. S6A and B). CpaA_{E520A} was included as a negative control for CpaA activity. We purified all His-tagged CpaA variants and determined their *in vitro* activities against various substrates (Fig. 7A and Fig. S6C and S7). The different mutations affected CpaA efficiency and site recognition in a substrate-specific manner. All mutants except CpaA_{E520A} were able to cleave fetuin, yielding a similar cleavage pattern (Fig. 7A). However, CpaAW493A cleaved fetuin with less efficiency. None of the CpaA variants were able to cleave EPO, further highlighting the essentiality of a Pro residue at P1 for targeting by CpaA (Fig. 7B). Treatment of CD55 and C1-INH with the CpaA mutants revealed that all variants are less active, as shown by an increase in the amounts of undigested substrate. Additional faint bands were observed, but we were unable to determine the cleavage site using MS analysis (Fig. S7).

We previously identified three cleavage sites for CpaA on the mid-region of etanercept, Pro₂₀₇Thr₂₀₈, Pro₂₁₅Ser₂₁₆, and Pro₂₂₅Ser₂₂₆ (Fig. 3B and Fig. S2). Cleavage by CpaA generates two fragments of similar molecular weights that comigrate as a single band of about ~36 kDa in SDS-PAGE (Fig. 1A and 7C). Notably, digestion of etanercept by CpaAW493A and W493L generated a major product of ~45 kDa instead (Fig. 7C). CpaAW493F produced three bands that migrate as ~36-, 40-, and 45-kDa fragments, respectively (Fig. 7C). MS analysis of these bands allowed the identification of the sites preferentially cleaved by the three CpaA variants (Fig. 7D). The unique ~45-kDa band results from cleavage at the Pro₁₈₃Thr₁₈₄ site, which is not a preferred site for wild-type CpaA. Unlike the other cleavage sites, the Pro₁₈₃Thr₁₈₄ site is in an area of low glycosylation of etanercept. Interestingly CpaAW493A and W493L variants were unable to cleave etanercept in the high-glycosylation-density region, which could explain the low activity against other mucin targets. Taken together, our data suggest that W493, although is not essential for activity, plays a role in CpaA substrate selectivity by interacting with the Pro residue of its target protein.

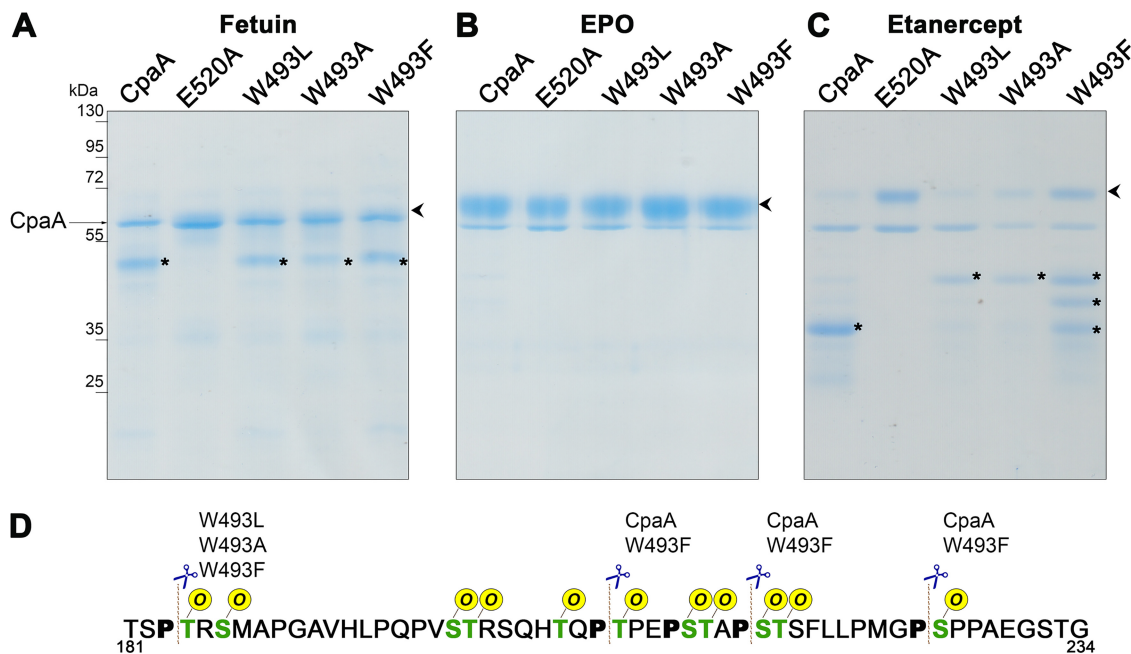


FIG 7 W493 affects CpaA activity. Fetuin (A), EPO (B), and etanercept (C) were incubated with purified CpaA and the different CpaA mutants (~55 kDa). Samples were separated by SDS-PAGE, and gels were stained with InstantBlue. Digestion products are indicated by asterisks. Undigested products are indicated by arrowheads. These data are representative of at least 3 independent experiments. (D) Fragment of the etanercept protein sequence. MS analysis of proteins bands allowed the identification of the sites preferentially cleaved by the four CpaA variants. All the known O-glycosylation sites are in green, and the Ps next to the glycosites are in bold. The dashed lines indicate the cleavage sites of CpaA and the different point mutants.

CpaA belongs to an expanding, functionally defined class of surface-exposed and secreted glycoproteases. Mounting evidence indicates that both commensal and pathogenic species produce and secrete glycoproteases to modulate adherence, penetrate the inner mucus layer, or evade the host immune response (18, 21, 47). For example, StcE contributes to immune evasion during EHEC infection by preventing immune cells from moving to the sites of infection (20, 23, 48, 49). Additionally, StcE activity against mucins promotes access of EHEC to epithelial cells, which assists host cell colonization (20, 50). *P. aeruginosa* produces IMPa, which cleaves the macrophage surface protein CD44, inhibiting phagocytosis (29). IMPa also cleaves P-selectin glycoprotein ligand 1 (PSGL-1), helping the bacterium to escape neutrophil attack (29). In *Vibrio cholerae*, secreted TagA targets host cell surface glycoproteins, modulating bacterial attachment during infection (26). These examples are indicative of the pivotal role of glycoproteases in modulating host-pathogen interactions by targeting various host proteins. Despite their relevance, only a relatively small number of bacterial glycoproteases have been biochemically characterized to various extents (Table 1). These enzymes are commonly encoded by bacteria isolated from mucin-rich environments, such as the human gut and lungs. These glycoproteases differ in their secretion mechanisms, protease class, domain organization, catalytic site, and recognized targets. Like IMPa, StcE, TagA, and SslE/YghJ, CpaA is secreted by a T2SS. CpaA, IMPa, and ZmpC display broad O-glycoprotease activity targeting mucins, regardless of their glycosylation density, as well as O-glycoproteins with low O-glycan chain density (fetuin). On the other hand, StcE can cleave only mucins with long mucin-like regions (such as CD43 and CD55) or short mucin-like regions (such as C1-INH). In contrast, other glycoproteases can target only a subset of mucins. For example, TagA requires more extensive densely glycosylated regions, whereas SslE/YghJ can digest major mucins such as MUC2 and MUC3, but it is inactive against mucins like CD43 and bovine submaxillary mucin (9, 23, 37, 51).

Due to low amino acid sequence conservation, it is not possible to differentiate proteases from glycoproteases solely on the basis of primary amino acid sequence.

However, structural analyses revealed the presence of Ig-fold domains in all these enzymes (Table 1, domains in yellow). Moreover, several factors impact substrate targeting by glycoprotease, including glycan chain identity and density, as well as amino acid composition (Table 1). Although some target motifs are known, the molecular bases for the recognition of specific target sequences remain poorly understood. Thus, even if structural analyses identify putative glycan-binding domains in a protein of interest, it is difficult to predict the specific substrates targeted by the putative glycoprotease. Together with the functional analysis, these studies define a functional class of secreted O-glycoproteases that mediate host-pathogen and host-commensal interactions.

DISCUSSION

Glycan chains decorate proteins to accomplish many different functions. One important role of glycosylation is the protection of proteins against proteolytic degradation. However, there is growing evidence that bacterial pathogens and commensals have evolved specific proteases that overcome the steric impediment posed by carbohydrates and indeed use glycans as recognition determinants to cleave glycoproteins right at the glycosylation site. In this work, we functionally characterized CpaA, a metzincin glycoprotease and T2SS-secreted virulence factor of several medically relevant *Acinetobacter* strains (4, 7, 8, 12). Previous work identified the blood coagulation proteins fV and fXII as targets of CpaA, suggesting a potential role in dissemination by interfering with the intrinsic coagulation pathway (4, 8, 12). The present work expands the known targets of CpaA and indicates that CpaA is a broad-spectrum enzyme with the ability to cleave various O-linked human glycoproteins. Our MS analysis of proteolytic fragments resulting from glycoprotein treatment with CpaA revealed that CpaA has a consensus target sequence consisting of a Pro residue followed by a glycosylated Ser or Thr (P-S/T), which is unprecedented for bacterial glycoproteases. Unlike other secreted glycoproteases, CpaA activity is not affected by sialic acid and is not restricted to highly O-glycosylated proteins (mucins). Indeed, CpaA also cleaves sparsely O-glycosylated proteins, such as fetuin. Although broad-spectrum secreted or surface-exposed glycoproteases appear to be widespread in bacteria, their identification cannot be assigned based on sequence homology, and biochemical and structural analyses are required to designate them as glycoproteases.

CpaA is composed of four very similar Ig-like domains and a catalytic domain (13). The catalytic domain located at the C terminus of CpaA exhibits all the canonical structural features of the metzincin superfamily. The four Ig-like domains are arranged in tandem, and they resemble the insertion domain of StcE, secreted by EHEC (13). These observations prompted us to further characterize CpaA activity. The StcE-specific motif S/T*-X-S/T (the asterisk denotes the glycosylation site) diverges from the P-S/T* motif recognized by CpaA (22). It is intriguing that despite recognizing different motifs, StcE and CpaA share the substrates C1-INH and CD55. Considering the different domain organization of the two proteases, it is not surprising that the proteins are classified in different metzincin subfamilies and target different proteins.

We previously showed that the substrate-binding cleft of CpaA is formed by residues from its four Ig-like domains and its catalytic domain (13). Thus, to be recognized by CpaA, the glycosylated substrate has to expose the P-S/T* peptide bond targeted for hydrolysis. It has been proposed that the interaction of the O-glycans with residues in the Ig-like domains (referred to as G sites by Noach et al. [16]) help to position the targeted peptide bond in the correct conformation to interact with the amino acids involved in catalysis (13, 16). Here, we show that CpaA activity is unaffected by the presence of sialic acid, indicating that the sialic acid can be accommodated inside the cleft but is not required for glycopeptide recognition. Our molecular modeling of CpaA with glycosylated substrates showed that linear glycans modified by sialic acid project this moiety away from the active site and into solvent, which supports our *in vitro* findings. Our modeling also revealed a possible interaction between the indole ring of W493 and the ring of the Pro residue in the targeted sequence. Although the

digestion of glycoproteins with W493 mutants displayed substrate-specific behaviors, overall, the mutants were less active and, in some cases, exhibited a shift in glycosylation site (glycosite) preference. We propose that CpaA selectivity may, at least in part, be the result of W493 forming a potential H- π interaction with the prolyl ring, minimizing the prolyl residue's exposure to solvent, and/or sterically holding the substrate in the active site. Further structural and biochemical studies are required to uncover the structural features that enable CpaA to target such a remarkably broad range of O-linked glycoproteins.

CpaA expression and secretion occur across several medically relevant *Acinetobacter* strains. Deletion of CpaA resulted in attenuation of *A. nosocomialis* M2 virulence in a respiratory murine infection model, playing a role in the dissemination from the lungs to the spleen (8). Previously, the only known substrates for CpaA were fV and fXII, proteins involved in blood coagulation. By digesting these proteins, CpaA increases the clotting time of human plasma (12, 15). We have now shown that CpaA can indeed cleave multiple proteins *in vitro*. Among these are several proteins involved in regulation of complement activation, including CD55 and CD46. However, only CD55 was removed from the cell surfaces, while CD46 remained unaltered during the *A. nosocomialis* infection assay. We hypothesize that CD46 interacts with another protein that blocks CpaA access. Thus, not all CpaA targets identified by our *in vitro* experiments are representative bona fide targets of CpaA *in vivo*. An additional role of CD55 is to act as an anti-adhesive molecule that regulates the release of neutrophils (52). Degradation of CD55 increases the retention of the neutrophils to the apical epithelial surface with the concomitant reduction of the amount of neutrophils that cross the epithelium (48, 49). CpaA expression and secretion are conserved across several medically relevant *Acinetobacter* strains isolated from diverse anatomical sites (8). We aligned the amino acid sequences of CpaA from *A. nosocomialis* M2 and several medically relevant *A. baumannii* strains (Fig. S8). We observed extremely high protein identity, suggesting that our findings can be extended to CpaA orthologs secreted by *A. baumannii*. Considering the broad-spectrum activity of CpaA and the abundance of glycosylated proteins in the human host, we propose that the physiological role of CpaA likely extends beyond interfering with the coagulation pathway or complement cascade. Further work will be required to understand the full extent of host immunomodulation by CpaA.

Host mucins and O-glycoproteins are major components of mucus, and they are ubiquitously expressed on cellular surfaces, where they act as physical barriers, receptor ligands, and mediators of intracellular signaling (53, 54). O-Glycoproteins and mucins lack a consensus sequence for O-glycosylation, and their O-linked glycans are highly heterogeneous in their glycan composition, numbers of residues, and linkages. However, aberrant mucin expression and glycosylation are also linked to various disease states, making mucins reliable biomarkers (54). For example, the mucin MUC1 is aberrantly expressed in the majority of cancers diagnosed each year in the United States. (53). Thus, the assessment of the mucin glycosylation status has high relevance for diagnosis of cancer and other diseases. Mucin domains are resistant to most commercially available proteases, which makes them difficult to analyze by traditional MS strategies. Bacterial glycoproteases have recently gained attention as tools for proteomic analysis of human glycoproteins (16, 22, 55). Our study shows that the broad-spectrum O-linked glycoprotease activity of CpaA is not affected by sialic acids. Moreover, it consistently digests any O-linked glycoprotein containing the P-S/T* sequence. These properties not only make CpaA a versatile enzyme modulating host-pathogen interactions but also highlight it as a robust and attractive new component of the glycoproteomics toolbox.

MATERIALS AND METHODS

Strains, plasmids, and growth conditions. Bacterial strains and plasmids used in this study can be found in Table S1. *E. coli* Stellar and *A. nosocomialis* M2 cells were grown in Lennox broth (LB) at 37°C. pWH1266-based plasmids were selected with tetracycline (5 μ g/ml).

Generation of CpaA point mutations. The PCR primers used for site-directed mutagenesis are listed in Table S1. To generate CpaAW493 variants, pWH-*cpaA-his-cpaB* was used as the template. PCR was carried out using Phusion DNA polymerase (Thermo Scientific). Site-directed mutagenesis was performed according to the method described by Fisher and Pei (56). Reaction products were transformed into *E. coli* Stellar cells and transformants were selected on LB agar supplemented with tetracycline. Mutagenesis was confirmed by DNA sequencing. The plasmids were then electroporated into the electrocompetent *A. nosocomialis* M2 Δ *cpaAB::frit* strain.

Protein purification. CpaA-His and its point mutation variants were purified from the supernatant of *A. nosocomialis* M2 (8). Briefly, *A. nosocomialis* M2 Δ *cpaAB::frit* strains carrying the pWH-based plasmids were grown in LB to mid-log phase. CpaA-His-tagged variants were purified by nickel affinity chromatography from cell-free filtered spent medium, as described before (8). Briefly, cells were pelleted at $8,000 \times g$ for 10 min. Cell-free supernatants were obtained by filtration using a Nalgene Rapid-Flow filter unit (pore size, 0.2 μ m) (Thermo Scientific), followed by concentration by approximately 3-fold using an Amicon Ultra-15 centrifugal filter unit (nominal molecular weight limit [NMWL], 10,000). Binding buffer (10 \times) was added to the cell-free supernatant prior to loading onto a nickel-nitrilotriacetic acid (Ni-NTA) agarose column (Gold Bio, St. Louis, MO) equilibrated with 1 \times binding buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole; pH 8.0). The column was washed with 20 column volumes (CV) of binding buffer and 10 CV of sashing buffer (50 mM NaH₂PO₄, 300 mM NaCl, 50 mM imidazole; pH 8.0). Proteins were eluted with elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 300 mM imidazole; pH 8.0). The purified proteins were concentrated, and buffer was exchanged for 20 mM HEPES, 150 mM NaCl, 50% glycerol (pH 7.4) using Amicon Ultra centrifugal filter units.

CpaA activity assay. Purified CpaA-His-tagged variants (60 μ g/ml) were assayed in reaction buffer (20 mM HEPES [pH 7.4], 150 mM NaCl, 1 mM ZnCl₂) containing the substrate proteins (200 μ g/ml) in a final volume of 15 μ l. All reaction mixtures were incubated at 37°C overnight. Reactions were monitored by SDS-PAGE, with staining with Coomassie blue or InstantBlue. Substrate proteins used in this study included CD55, CD46, TIM1, TIM4 C1-INH, erythropoietin (Sino Biological), asialofetuin (Sigma), fetuin and RNase B (New England Biolabs), etanercept (Enbrel), and abatacept (Orencia).

Immunoblotting. Bacterial whole-cell and supernatant samples were prepared as before (7, 8). Briefly, cultures were grown to an optical density at 600 nm (OD₆₀₀) of 0.5, and the OD 0.5 cultures were pelleted by centrifugation and resuspended in 50 μ l of Laemmli buffer for the whole-cell samples. Supernatant samples were obtained by trichloroacetic acid (TCA) precipitation of cell-free supernatants (7, 8). Protein samples were analyzed by SDS-PAGE, transferred to a nitrocellulose membrane, probed with polyclonal anti-CpaA (1:1,000) (8) and/or monoclonal anti-RNA polymerase (1:2,000; BioLegend). Western blots were probed with IRDye-conjugated secondary antibodies and visualized with an Odyssey CLx imaging system (Li-COR Biosciences, Lincoln, NE).

Fetuin deglycosylation. Fetuin was deglycosylated under denaturing conditions using a protein deglycosylation mix II kit (New England Biolabs) according to the manufacturer's protocol. Briefly, 100 μ g of fetuin was incubated in deglycosylation mix buffer 2 at 75°C for 10 min. After the mixture had cooled, protein deglycosylation mix II was added, with the exception of one tube to be used as a positive control for CpaA activity. The glycosylated and deglycosylated fetuin proteins were then used as substrates to test CpaA activity as described above.

Glycan array. The glycan array binding analysis was performed at the Consortium of Functional Glycomics (CFG). Two concentrations of purified CpaA (cfg_rRequest_3530) and CpaA_{E520A} (cfg_rRequest_3617) were screened with version 5.4 of the CFG printed array, which consists of 585 glycans in replicates of 6, according to the standard procedure of the CFG.

Flow cytometry. HeLa cells were maintained in Dulbecco's minimal essential medium (DMEM) with high glucose supplemented with 10% (vol/vol) heat-inactivated fetal bovine serum (FBS) (Gibco) at 37°C and 5% CO₂. HEK-293 cells were maintained in Eagle's minimal essential medium (EMEM) with 10% (vol/vol) heat-inactivated FBS (Gibco) at 37°C and 5% CO₂. Cells were seeded in 60-mm plates at a density of 1×10^6 cells per plate a day prior to each experiment. For infection assays, freshly transformed *A. nosocomialis* M2 cells with plasmids pWH-*cpaA-his-cpaB* and pMFH32 were incubated overnight in LB supplemented with tetracycline. Stationary-phase cultures were normalized to an OD of 1 and washed three times with phosphate-buffered saline (PBS) before addition to cells at the indicated multiplicity of infection (MOI) in the respective cell medium. Infected cells were incubated for 2.5 h at 37°C and 5% CO₂. Alternatively, cells were treated with 2 or 5 μ g/ml of purified CpaA or CpaA_{E520A} as indicated, and incubated for 1 h at 37°C and 5% CO₂. Following infection or treatment with purified protein, eukaryotic cells were washed with PBS (twice) and harvested using 0.05% trypsin (Corning). Cells were blocked with 2% (vol/vol) FBS in PBS and labeled with fluorophore-conjugated primary antibodies against CD55 (phycoerythrin [PE]; Sino Biological), and CD46 (allophycocyanin [APC]; Invitrogen). All incubations were carried out for 45 min on ice, and cells were washed three times with PBS. Labeled cells were fixed with 4% (vol/vol) paraformaldehyde in PBS for 10 min, washed with PBS, resuspended in PBS to a cell concentration of 1×10^6 cell/ml, and filtered prior to analysis. Flow cytometry was performed on a three-laser, 8-color FACSCanto II cytometer (BD Biosciences). Experiments were performed in technical triplicates, with 10,000 events analyzed per sample. Data were analyzed using FlowJo 10.6 (FlowJo LLC), and individual values from three independent experiments (each experiment has at least three technical replicates) were combined and analyzed by two-way analysis of variance (ANOVA) using GraphPad Prism.

Tryptic digest of gel-separated protein bands. CpaA digested proteins were separated using SDS-PAGE, fixed, and visualized with Coomassie blue or InstantBlue, as described above. Bands of interest were excised and processed as previously described (37). Briefly, gel bands were first destained in a

solution of 50 mM NH_4HCO_3 –50% ethanol for 20 min at room temperature with shaking at 750 rpm. Destained bands were dehydrated with 100% ethanol, vacuum dried for 20 min, and then rehydrated in 50 mM NH_4HCO_3 plus 10 mM dithiothreitol (DTT). Protein bands were reduced for 60 min at 56°C with shaking then washed twice in 100% ethanol for 10 min to remove residual DTT. Reduced ethanol washed samples were sequentially alkylated with 55 mM iodoacetamide in 50 mM NH_4HCO_3 in the dark for 45 min at room temperature. Alkylated samples were then washed with 50 mM NH_4HCO_3 followed by 100% ethanol twice for 5 min to remove residual iodoacetamide and then vacuum dried. Alkylated samples were then rehydrated with 12 ng/ μl trypsin (Promega) in 40 mM NH_4HCO_3 at 4°C for 1 h. Excess trypsin was removed, and gel pieces were covered in 40 mM NH_4HCO_3 and incubated overnight at 37°C. Peptides were concentrated and desalted using C_{18} stage tips (57, 58) before analysis by liquid chromatography (LC)-MS.

Identification of CpaA digestion products and cleavage sites using reverse-phase LC-MS.

Purified peptides prepared were resuspend in buffer A* (0.1% trifluoroacetic acid [TFA], 2% acetonitrile) and separated using a two-column chromatography setup composed of a PepMap100 C_{18} 20-mm by 75- μm trap and a PepMap C_{18} 500-mm by 75- μm analytical column (Thermo Fisher Scientific). Samples were concentrated on the trap column at 5 $\mu\text{l}/\text{min}$ for 5 min with buffer A (0.1% formic acid, 2% dimethyl sulfoxide [DMSO]) and infused into an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific) at 300 nl/min via the analytical column using a Dionex Ultimate 3000 ultraperformance liquid chromatograph (UPLC) (Thermo Fisher Scientific). Gradients (45 or 65 min) were run for each sample, altering the buffer composition from 1% buffer B (0.1% formic acid, 77.9% acetonitrile, 2% DMSO) to 28% B over 20 or 40 min, then from 28% B to 40% B over 5 min, and then from 40% B to 100% B over 2 min; the composition was held at 100% B for 3 min, dropped to 3% B over 5 min, and held at 3% B for another 10 min. For 45-min gradients, the Lumos mass spectrometer was operated in a data-dependent mode automatically switching between the acquisition of a single Orbitrap MS scan (240,000 resolution) every 3 s and MS2 events. For each ion selected, dissociation parameters were as follows. For collision-induced dissociation (CID), Fourier transform MS (FTMS) was at 15,000 resolution, maximum fill time was 100 ms, and automatic gain control (AGC) was 2×10^5 . For higher-energy collisional dissociation (HCD), FTMS was at 15,000 resolution, maximum fill time was 120 ms, normalized collision energy was 35, and AGC was 2×10^5 . For electron transfer–higher-energy collision dissociation (ETcD), FTMS was at 15,000 resolution, maximum fill time was 120 ms, supplementary activation was 15%, and AGC was 2×10^5 . For 65-min gradients, the Lumos mass spectrometer was operated in a data-dependent mode, automatically switching between the acquisition of a single Orbitrap MS scan (120,000 resolution) every 3 s and MS2 HCD scans of precursors (normalized collision energy [NCE], 30%; maximal injection time of 22 ms; AGC, 5×10^4 with a resolution of 15,000). HexNAc oxonium ion (204.087 m/z) product-dependent MS/MS analysis (59) was used to trigger two additional scans of potential glycopeptides, a CID scan (ion trap MS [ITMS]; maximum fill time, 100 ms, AGC, 5×10^4) and an ETcD scan (FTMS, 30,000 resolution; maximum fill time, 200 ms; supplementary activation, 15%; AGC, 2×10^5) scan.

Mass spectrometry data analysis. The assessment of the protein coverage within CpaA-digested bands and the identification of mucin glycopeptides was accomplished using MaxQuant (v1.5.3.30) (60). Searches were performed against the custom databases populated with the protein sequence of the recombinant proteins of interest with carbamidomethylation of cysteine set as a fixed modification. Searches were performed with semitrypsin cleavage specificity allowing 2 miscleavage events. For the identification of glycopeptide, multiple searches were performed on each sample, allowing oxidation of methionine and a maximum of three glycan variable modifications using (i) HexNAc (S/T), HexHexNAc (S/T), HexNAcHexNAc (S/T) or (ii) HexNAc (S/T), Hex(1)HexNAc(1)NeuAc(1) (S/T), Hex(2)HexNAc(1)NeuAc(1) (S/T). The precursor mass tolerance was set to 20 ppm for the first search and 10 ppm for the main search, with a maximum false discovery rate (FDR) of 1.0% set for protein and peptide identifications. The resulting protein group output was processed within the Perseus (v1.4.0.6) (60) analysis environment to remove reverse matches and common protein contaminants. Semitryptic peptides and semitryptic glycopeptides were manually inspected for correctness. Annotation of MS-MS provided within the supplementary data was undertaken using the Interactive Peptide Spectral Annotator (60).

Molecular modeling. All computational work was performed using the 2019 Molecular Operating Environment (MOE 2019.01) software suite (Chemical Computing Group ULC, Montreal, Canada) (61). The X-ray crystal structures of the metzincin enzymes CpaA (6038) (13) and serralyisin (3V11), as well as the gluzincin enzymes IMPa (5KDX, (16), ZmpB (5KDU and 5KDS) (16), and BT4244 (5KD8) (16), were overlaid using the residues of the highly conserved alpha helix within their active sites (Table S2).

Following a previously described protocol (22), the crystal structure of CpaA was prepared by adding unresolved side chains and hydrogens as well as capping termini with acetyl or *N*-methyl (*N*-Me) groups. Notably, the crystallographic chaperone protein CpaB was not included during this study. This protein arranges its C-terminal tail into the CpaA catalytic site, similar to zymogens of related metallopeptidases (62), and we assume that this portion of CpaB is displaced by substrate.

Peptidomimetic and peptidic ligands have been observed to bind in a similar conformation within the metzincin active site (44, 45), using the P2-P1' ligand residues to form an antiparallel β -sheet with a β -strand of the enzyme active site. Similar to previous work (22), the crystallographic peptide ligand bound to serralyisin (RPKPQQ) was used as a scaffold to construct the peptide portion (Ac–EAPSA–*N*-Me) of the fetuin substrate fragment bound to the CpaA model (Table S2). Related gluzincin enzymes were found to have glycan-containing ligands near the P1' site of their enzyme active sites (16). Thus, the crystallographic glycan bound to IMPa (Gal β 1-3GalNAc α 1) (5KDX) (16) and ZmpB (Gal β 1-3(Neu5Ac α 2-6)GalNAc α 1) (5KDU) (16) were grafted onto the fetuin peptide fragment to generate corresponding

glycoforms, and the Glycan Fragment Database (63) was used to identify a glycan (Neu5Ac α 2-3Gal β 1-3GalNAc α 1) (2CWG) (64) to generate the final glycoform (Table S2).

The glycopeptides were docked into the CpaA model in three steps: conformational search, virtual screen, and minimization. This process allowed exploration of all reasonable conformations of each glycopeptide prior to induced-fit docking with the CpaA model. Each glycopeptide underwent a conformational search using the Amber10:EHT force field (65) to generate a corresponding library of conformers. During this step, the side chains of the peptide and the pendant groups of each sugar were allowed to freely rotate; any sialic acid moieties were also allowed to move freely. All other atoms were fixed. This process generated small libraries of approximately 4,000 conformers for each glycopeptide species. The conformers of each glycopeptide library underwent virtual screening with the CpaA model, again using the Amber10:EHT force field. During this process, all atoms of the enzyme and each glycopeptide were fixed, and the docking score for each resulting complex was calculated using the GBVI/WSA dG scoring function (66). Finally, the top 10 complexes identified from the virtual screening were subsequently minimized using the Amber10:EHT force field. The glycopeptide substrate, the catalytic zinc ion, and residues of the CpaA model having atoms within 10 Å of the substrate were allowed to move; all other residues were fixed, and solvent molecules were omitted. The best, most consistent complexes are shown.

This screening and minimization process was repeated for the Gal β 1-3GalNAc α 1-modified glycopeptide conformer library with the mutant CpaA models CpaAW493F, CpaAW493L, and CpaAW493A. In all cases, the dihedral angles of the peptide backbones (ϕ , ψ) and side chains (χ) as well as the initial glycosidic linkages (ϕ , ψ) of the docked substrates were measured to ensure proper geometry.

Data availability. The mass spectrometry proteomics data have been deposited in the ProteomeX-change Consortium database via the PRIDE (67) partner repository with the data set identifier PXD019941.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 1.5 MB.

FIG S2, TIF file, 1.1 MB.

FIG S3, TIF file, 0.5 MB.

FIG S4, PDF file, 2.1 MB.

FIG S5, PDF file, 0.2 MB.

FIG S6, TIF file, 2.5 MB.

FIG S7, TIF file, 1.2 MB.

FIG S8, PDF file, 0.1 MB.

TABLE S1, PDF file, 0.1 MB.

TABLE S2, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

Flow cytometry experiments were performed at the Flow Cytometry & Fluorescence Activated Cell Sorting Core (Department of Pathology & Immunology, Washington University School of Medicine in St. Louis). Protein-Glycan Interaction Resource of the CFG (supporting grant R24 GM098791), and the National Center for Functional Glycomics (NCFG) at Beth Israel Deaconess Medical Center, Harvard Medical School (supporting grant P41 GM103694). We thank the Melbourne Mass Spectrometry and Proteomics Facility of The Bio21 Molecular Science and Biotechnology Institute for access to MS instrumentation. We thank Paula Magnelli (New England Biolabs) for generously providing us Etanercept and Abatacept. We thank the members of the Feldman lab for critical reading of the manuscript.

This work was supported by a National Health and Medical Research Council of Australia (NHMRC) project grant awarded to N.E.S. (APP1100164) and a Canadian Institutes of Health Research project grant to A.B.B. (PJT 159786). This work was also supported by grants from the National Institute of Allergy and Infectious Diseases grant R01AI144120 awarded to M.F.F.

We declare that we have no conflicts of interest with the contents of this article.

Conceptualization: M.F.H., M.F.F. Methodology: M.F.H., N.E.S., G.D.V., J.L., B.P., A.B.B., M.J.F.; Analysis: M.F.H., N.E.S., M.J.F., M.F.F.; Writing – Original Draft: M.F.H., M.F.F., Writing – Review & Editing: M.F.H., N.E.S., G.D.V., J.L., A.B.B., M.J.F., M.F.F.; Supervision: M.F.F.

REFERENCES

- Tacconelli E, Carrara E, Savoldi A, Kattula D, Burkert F. 2013. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. World Health Organization, Geneva, Switzerland.
- Giammanco A, Calà C, Fasciana T, Dowzicky MJ. 2017. Global assessment of the activity of tigecycline against multidrug-resistant Gram-negative pathogens between 2004 and 2014 as part of the tigecycline evaluation and surveillance trial. *mSphere* 2:e00310-16. <https://doi.org/10.1128/mSphere.00310-16>.
- Elhosseiny NM, Elhezawy NB, Attia AS. 2019. Comparative proteomics analyses of *Acinetobacter baumannii* strains ATCC 17978 and AB5075 reveal the differential role of type II secretion system secretomes in lung colonization and ciprofloxacin resistance. *Microb Pathog* 128:20–27. <https://doi.org/10.1016/j.micpath.2018.12.039>.
- Waack U, Johnson TL, Chedid K, Xi C, Simmons LA, Mobley HLT, Sandkvist M. 2017. Targeting the type II secretion system: development, optimization, and validation of a high-throughput screen for the identification of small molecule inhibitors. *Front Cell Infect Microbiol* 7:380. <https://doi.org/10.3389/fcimb.2017.00380>.
- Weber BS, Kinsella RL, Harding CM, Feldman MF. 2017. The secrets of *Acinetobacter* secretion. *Trends Microbiol* 25:532–545. <https://doi.org/10.1016/j.tim.2017.01.005>.
- Elhosseiny NM, El-Tayeb OM, Yassin AS, Lory S, Attia AS. 2016. The secretome of *Acinetobacter baumannii* ATCC 17978 type II secretion system reveals a novel plasmid encoded phospholipase that could be implicated in lung colonization. *Int J Med Microbiol* 306:633–641. <https://doi.org/10.1016/j.ijmm.2016.09.006>.
- Harding CM, Kinsella RL, Palmer LD, Skaar EP, Feldman MF. 2016. Medically relevant *Acinetobacter* species require a type II secretion system and specific membrane-associated chaperones for the export of multiple substrates and full virulence. *PLoS Pathog* 12:e1005391. <https://doi.org/10.1371/journal.ppat.1005391>.
- Kinsella RL, Lopez J, Palmer LD, Salinas ND, Skaar EP, Tolia NH, Feldman MF. 2017. Defining the interaction of the protease CpaA with its type II secretion chaperone CpaB and its contribution to virulence in *Acinetobacter* species. *J Biol Chem* 292:19628–19638. <https://doi.org/10.1074/jbc.M117.808394>.
- Eijkkelkamp BA, Stroehrer UH, Hassan KA, Paulsen IT, Brown MH. 2014. Comparative analysis of surface-exposed virulence factors of *Acinetobacter baumannii*. *BMC Genomics* 15:1020. <https://doi.org/10.1186/1471-2164-15-1020>.
- Wang N, Ozer EA, Mandel MJ, Hauser AR. 2014. Genome-wide identification of *Acinetobacter baumannii* genes necessary for persistence in the lung. *mBio* 5:e01163-14. <https://doi.org/10.1128/mBio.01163-14>.
- Johnson TL, Waack U, Smith S, Mobley H, Sandkvist M. 2015. *Acinetobacter baumannii* is dependent on the type II secretion system and its substrate LipA for lipid utilization and in vivo fitness. *J Bacteriol* 198:711–719. <https://doi.org/10.1128/JB.00622-15>.
- Tilley D, Law R, Warren S, Samis JA, Kumar A. 2014. CpaA a novel protease from *Acinetobacter baumannii* clinical isolates deregulates blood coagulation. *FEMS Microbiol Lett* 356:53–61. <https://doi.org/10.1111/1574-6968.12496>.
- Urusova DV, Kinsella RL, Salinas ND, Haurat MF, Feldman MF, Tolia NH. 2019. The structure of *Acinetobacter*-secreted protease CpaA complexed with its chaperone CpaB reveals a novel mode of a T2SS chaperone-substrate interaction. *J Biol Chem* 294:13344–13354. <https://doi.org/10.1074/jbc.RA119.009805>.
- Rosenau F, Tommassen J, Jaeger KE. 2004. Lipase-specific foldases. *ChemBiochem* 5:152–161. <https://doi.org/10.1002/cbic.200300761>.
- Waack U, Warnock M, Yee A, Huttinger Z, Smith S, Kumar A, Deroux A, Ginsburg D, Mobley HLT, Lawrence DA, Sandkvist M. 2018. CpaA is a glycan-specific adamalysin-like protease secreted by *Acinetobacter baumannii* that inactivates coagulation factor XII. *mBio* 9:e01606-18. <https://doi.org/10.1128/mBio.01606-18>.
- Noach I, Ficko-Blean E, Pluvinaige B, Stuart C, Jenkins ML, Brochu D, Buenbrazo N, Wakarchuk W, Burke JE, Gilbert M, Boraston AB. 2017. Recognition of protein-linked glycans as a determinant of peptidase activity. *Proc Natl Acad Sci U S A* 114:E679–E688. <https://doi.org/10.1073/pnas.1615141114>.
- Yu ACY, Worrall LJ, Strynadka NCJ. 2012. Structural insight into the bacterial mucinase StcE essential to adhesion and immune evasion during enterohemorrhagic *E. coli* infection. *Structure* 20:707–717. <https://doi.org/10.1016/j.str.2012.02.015>.
- Cerdà-Costa N, Gomis-Rüth FX. 2014. Architecture and function of metallopeptidase catalytic domains. *Protein Sci* 23:123–144. <https://doi.org/10.1002/pro.2400>.
- Gomis-Rüth FX. 2009. Catalytic domain architecture of metzincin metalloproteases. *J Biol Chem* 284:15353–15357. <https://doi.org/10.1074/jbc.R800069200>.
- Latham WW, Grys TE, Witowski SE, Torres AG, Kaper JB, Tarr PI, Welch RA. 2002. StcE, a metalloprotease secreted by *Escherichia coli* O157:H7, specifically cleaves C1 esterase inhibitor. *Mol Microbiol* 45:277–288. <https://doi.org/10.1046/j.1365-2958.2002.02997.x>.
- Nakjang S, Ndeh DA, Wipat A, Bolam DN, Hirt RP. 2012. A novel extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. *PLoS One* 7:e30287. <https://doi.org/10.1371/journal.pone.0030287>.
- Malaker SA, Pedram K, Ferracane MJ, Bensing BA, Krishnan V, Pett C, Yu J, Woods EC, Kramer JR, Westerlind U, Dorigo O, Bertozzi CR. 2019. The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins. *Proc Natl Acad Sci U S A* 116:7278–7287. <https://doi.org/10.1073/pnas.1813020116>.
- Szabady RL, Welch RA. 2013. StcE peptidase and the StcE-like metalloendopeptidases, p 1272–1280. *In* Rawlings ND, Salvesen G (ed), *Handbook of proteolytic enzymes*. Elsevier, Amsterdam, The Netherlands.
- Hang HC, Bertozzi CR. 2005. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem* 13:5021–5034. <https://doi.org/10.1016/j.bmc.2005.04.085>.
- Kesimer M, Sheehan JK. 2012. Mass spectrometric analysis of mucin core proteins. *Methods Mol Biol* 842:67–79. https://doi.org/10.1007/978-1-61779-513-8_4.
- Szabady RL, Yanta JH, Halladin DK, Schofield MJ, Welch RA. 2011. TagA is a secreted protease of *Vibrio cholerae* that specifically cleaves mucin glycoproteins. *Microbiology (Reading)* 157:516–525. <https://doi.org/10.1099/mic.0.044529-0>.
- Grys TE, Siegel MB, Latham WW, Welch RA. 2005. The StcE protease contributes to intimate adherence of enterohemorrhagic *Escherichia coli* O157:H7 to host cells. *Infect Immun* 73:1295–1303. <https://doi.org/10.1128/IAI.73.3.1295-1303.2005>.
- Nesta B, Valeri M, Spagnuolo A, Rosini R, Mora M, Donato P, Alteri CJ, Del Vecchio M, Buccato S, Pezzicoli A, Bertoldi I, Buzzigoli L, Tuscano G, Falduto M, Ripa V, Ashhab Y, Bensi G, Fontana MR, Seib KL, Mobley HLT, Pizza M, Soriani M, Serino L. 2014. SsE elicits functional antibodies that impair in vitro mucinase activity and in vivo colonization by both intestinal and extraintestinal *Escherichia coli* strains. *PLoS Pathog* 10:e1004124. <https://doi.org/10.1371/journal.ppat.1004124>.
- Bardoel BW, Hartsink D, Vughs MM, de Haas CJ, van Strijp JAG, van Kessel KPM. 2012. Identification of an immunomodulating metalloprotease of *Pseudomonas aeruginosa* (IMPa). *Cell Microbiol* 14:902–913. <https://doi.org/10.1111/j.1462-5822.2012.01765.x>.
- Peppel K, Poltorak A, Melhado I, Jirik F, Beutler B. 1993. Expression of a TNF inhibitor in transgenic mice. *J Immunol* 151:5699–5703.
- Moreland L, Bate G, Kirkpatrick P. 2006. Abatacept. *Nat Rev Drug Discov* 5:185–186. <https://doi.org/10.1038/nrd1989>.
- Peppel K, Crawford D, Beutler B. 1991. A tumor necrosis factor (TNF) receptor-IgG heavy chain chimeric protein as a bivalent antagonist of TNF activity. *J Exp Med* 174:1483–1489. <https://doi.org/10.1084/jem.174.6.1483>.
- Grys TE, Walters LL, Welch RA. 2006. Characterization of the StcE protease activity of *Escherichia coli* O157:H7. *J Bacteriol* 188:4646–4653. <https://doi.org/10.1128/JB.01806-05>.
- Varki A. 2008. Sialic acids in human health and disease. *Trends Mol Med* 14:351–360. <https://doi.org/10.1016/j.molmed.2008.06.002>.
- Abdullah KM, Udoh EA, Shewen PE, Mellors A. 1992. A neutral glycoprotease of *Pasteurella haemolytica* A1 specifically cleaves O-sialoglycoproteins. *Infect Immun* 60:56–62. <https://doi.org/10.1128/IAI.60.1.56-62.1992>.
- Yang W, Ao M, Hu Y, Li QK, Zhang H. 2018. Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides (EXO). *Mol Syst Biol* 14:e8486. <https://doi.org/10.15252/msb.20188486>.
- Musumeci MA, Hug I, Scott NE, Ielmini MV, Foster LJ, Wang PG, Feldman MF. 2013. In vitro activity of *Neisseria meningitidis* PgII

- O-oligosaccharyltransferase with diverse synthetic lipid donors and a UDP-activated sugar. *J Biol Chem* 288:10578–10587. <https://doi.org/10.1074/jbc.M112.432815>.
38. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <https://doi.org/10.1101/gr.849004>.
 39. Yang X, Tao S, Orlando R, Brockhausen I, Kan FWK. 2012. Structures and biosynthesis of the N- and O-glycans of recombinant human oviduct-specific glycoprotein expressed in human embryonic kidney cells. *Carbohydr Res* 358:47–55. <https://doi.org/10.1016/j.carres.2012.05.027>.
 40. Brockhausen I, Stanley P. 2017. O-GalNAc glycans. In Cummings VA, et al (ed), *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 41. Koropatkin N, Martens EC, Gordon JI, Smith TJ. 2009. Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont. *Biochemistry* 48:1532–1542. <https://doi.org/10.1021/bi801942a>.
 42. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson IM, Edlund K, Lundberg E, Navani S, Szijgyarto CAK, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H, Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, Von Feilitzen K, Forsberg M, Persson L, Johansson F, Zwahlen M, Von Heijne G, Nielsen J, Pontén F. 2015. Tissue-based map of the human proteome. *Science* 347:1260419–1260419. <https://doi.org/10.1126/science.1260419>.
 43. Croset A, Delafosse L, Gaudry JP, Arod C, Glez L, Losberger C, Begue D, Krstanovic A, Robert F, Vilbois F, Chevalet L, Antonsson B. 2012. Differences in the glycosylation of recombinant proteins expressed in HEK and CHO cells. *J Biotechnol* 161:336–348. <https://doi.org/10.1016/j.jbiotec.2012.06.038>.
 44. Grams F, Dive V, Yiotakis A, Yiallourous I, Vassiliou S, Zwilling R, Bode W, Stocker W. 1996. Structure of astacin with a transition-state analogue inhibitor. *Nat Struct Biol* 3:671–675. <https://doi.org/10.1038/nsb0896-671>.
 45. Oberholzer AE, Bumann M, Hege T, Russo S, Baumann U. 2009. Metzincin's canonical methionine is responsible for the structural integrity of the zinc-binding site. *Biol Chem* 390:875–881. <https://doi.org/10.1515/BC.2009.100>.
 46. Coltart DM, Royyuru AK, Williams LJ, Glunz PW, Sames D, Kuduk SD, Schwarz JB, Chen XT, Danishefsky SJ, Live DH. 2002. Principles of mucin architecture: structural studies on synthetic glycopeptides bearing clustered mono-, di-, tri-, and hexasaccharide glycodomains. *J Am Chem Soc* 124:9833–9844. <https://doi.org/10.1021/ja020208f>.
 47. Tapader R, Basu S, Pal A. 2019. Secreted proteases: a new insight in the pathogenesis of extraintestinal pathogenic *Escherichia coli*. *Int J Med Microbiol* 309:159–168. <https://doi.org/10.1016/j.ijmm.2019.03.002>.
 48. Furniss CRD, Low WW, Mavridou DAI, Dagley LF, Webb AI, Tate EW, Clements A. 2018. Plasma membrane profiling during enterohemorrhagic *E. coli* infection reveals that the metalloprotease StcE cleaves CD55 from host epithelial surfaces. *J Biol Chem* 293:17188–17199. <https://doi.org/10.1074/jbc.RA118.005114>.
 49. Szabady RL, Lokuta MA, Walters KB, Huttenlocher A, Welch RA. 2009. Modulation of neutrophil function by a secreted mucinase of *Escherichia coli* O157:H7. *PLoS Pathog* 5:e1000320. <https://doi.org/10.1371/journal.ppat.1000320>.
 50. Hews CL, Tran SL, Wegmann U, Brett B, Walsham ADS, Kavanaugh D, Ward NJ, Juge N, Schüller S. 2017. The StcE metalloprotease of enterohemorrhagic *Escherichia coli* reduces the inner mucus layer and promotes adherence to human colonic epithelium ex vivo. *Cell Microbiol* 19:e12717. <https://doi.org/10.1111/cmi.12717>.
 51. Shon DJ, Malaker SA, Pedram K, Yang E, Krishnan V, Dorigo O, Bertozzi CR. 2020. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins. *Proc Natl Acad Sci USA* 117:21299–21307. <https://doi.org/10.1073/pnas.2012196117>.
 52. Lawrence DW, Bruyninckx WJ, Louis NA, Lublin DM, Stahl GL, Parkos CA, Colgan SP. 2003. Antiadhesive role of apical decay-accelerating factor (CD55) in human neutrophil transmigration across mucosal epithelia. *J Exp Med* 198:999–1010. <https://doi.org/10.1084/jem.20030380>.
 53. Beatson R, Tajadura-Ortega V, Achkova D, Picco G, Tsoourouktsoglou TD, Klausung S, Hillier M, Maher J, Noll T, Crocker PR, Taylor-Papadimitriou J, Burchell JM. 2016. The mucin MUC1 modulates the tumor immunological microenvironment through engagement of the lectin Siglec-9. *Nat Immunol* 17:1273–1281. <https://doi.org/10.1038/ni.3552>.
 54. Pinho SS, Reis CA. 2015. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* 15:540–555. <https://doi.org/10.1038/nrc3982>.
 55. Zipfel PF, Skerka C. 2009. Complement regulators and inhibitory proteins. *Nat Rev Immunol* 9:729–740. <https://doi.org/10.1038/nri2620>.
 56. Fisher CL, Pei GK. 1997. Modification of a PCR-based site-directed mutagenesis method. *Biotechniques* 23:570–574. <https://doi.org/10.2144/97234bm01>.
 57. Rappsilber J, Ishihama Y, Mann M. 2003. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* 75:663–670. <https://doi.org/10.1021/ac026117i>.
 58. Rappsilber J, Mann M, Ishihama Y. 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* 2:1896–1906. <https://doi.org/10.1038/nprot.2007.261>.
 59. Saba J, Dutta S, Hemenway E, Viner R. 2012. Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. In *J Proteomics* 2012:560391. <https://doi.org/10.1155/2012/560391>.
 60. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26:1367–1372. <https://doi.org/10.1038/nbt.1511>.
 61. Chemical Computing Group ULC. 2019. Molecular Operating Environment (MOE).
 62. López-Pelegrín M, Cerdà-Costa N, Martínez-Jiménez F, Cintas-Pedrola A, Canals A, Peinado JR, Martí-Renom MA, López-Otín C, Arolas JL, Gomis-Rüth FX. 2013. A novel family of soluble minimal scaffolds provides structural insight into the catalytic domains of integral membrane metalloproteases. *J Biol Chem* 288:21279–21294. <https://doi.org/10.1074/jbc.M113.476580>.
 63. Jo S, Im W. 2013. Glycan fragment database: a database of PDB-based glycan 3D structures. *Nucleic Acids Res* 41:D470–D474. <https://doi.org/10.1093/nar/gks987>.
 64. Wright CS, Jaeger J. 1993. Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *J Mol Biol* 232:620–638. <https://doi.org/10.1006/jmbi.1993.1415>.
 65. Weiner SJ, Kollman PA, Singh UC, Case DA, Ghio C, Alagona G, Profeta S, Weiner P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 106:765–784. <https://doi.org/10.1021/ja00315a051>.
 66. Corbeil CR, Williams CI, Labute P. 2012. Variability in docking success rates due to dataset preparation. *J Comput Aided Mol Des* 26:775–786. <https://doi.org/10.1007/s10822-012-9570-1>.
 67. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, Pérez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz Ş, Tiwary S, Cox J, Audain E, Walzer M, Jarnuczak AF, Ternent T, Brazma A, Vizcaíno JA. 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47:D442–D450. <https://doi.org/10.1093/nar/gky1106>.
 68. Seo J, Brencic A, Darwin AJ. 2009. Analysis of secretin-induced stress in *Pseudomonas aeruginosa* suggests prevention rather than response and identifies a novel protein involved in secretin function. *J Bacteriol* 191:898–908. <https://doi.org/10.1128/JB.01443-08>.
 69. Gutiérrez-Jiménez J, Arciniega I, Navarro-García F. 2008. The serine protease motif of Pic mediates a dose-dependent mucolytic activity after binding to sugar constituents of the mucin substrate. *Microb Pathog* 45:115–123. <https://doi.org/10.1016/j.micpath.2008.04.006>.
 70. Jiang P, Mellors A. 2013. O-sialoglycoprotein endopeptidase, p 1664–1666. In Rawlings ND, Salvesen G (ed), *Handbook of proteolytic enzymes*. Elsevier Ltd., Amsterdam, The Netherlands.
 71. Tapader R, Bose D, Basu P, Mondal M, Mondal A, Chatterjee NS, Dutta P, Basu S, Bhadra RK, Pal A. 2016. Role in proinflammatory response of YghJ, a secreted metalloprotease from neonatal septicemic *Escherichia coli*. *Int J Med Microbiol* 306:554–565. <https://doi.org/10.1016/j.ijmm.2016.06.003>.
 72. Cladman WM, Watt MAV, Dini JP, Mellors A. 1996. The pasteurilla haemolytica O-sialoglycoprotein endopeptidase is inhibited by zinc ions and does not cleave fetuin. *Biochem Biophys Res Commun* 220:141–146. <https://doi.org/10.1006/bbrc.1996.0371>.