Novel inhibitors of surface layer processing in *Clostridium difficile*

T. H. Tam Dang [a], Robert P. Fagan [b], Neil F. Fairweather [b,c], Edward W. Tate [a,c,*]

[a] *Department of Chemistry, Imperial College London, London SW72AZ, United Kingdom*
[b] *Department of Life Sciences, Imperial College London, London SW72AZ, United Kingdom*
[c] *Institute of Chemical Biology, Imperial College London, London SW72AZ, United Kingdom*

A B S T R A C T

Clostridium difficile, a leading cause of hospital-acquired bacterial infection, is coated in a dense surface layer (S-layer) that is thought to provide both physicochemical protection and a scaffold for host-pathogen interactions. The key structural components of the S-layer are two proteins derived from a polypeptide precursor, SlpA, via proteolytic cleavage by the protease Cwp84. Here, we report the design, synthesis and in vivo characterization of a panel of protease inhibitors and activity-based probes (ABPs) designed to target S-layer processing in live *C. difficile* cells. Inhibitors based on substrate-mimetic peptides bearing a C-terminal Michael acceptor warhead were found to be promising candidates for further development.

1. Introduction

Clostridium difficile is an anaerobic Gram-positive bacterium and produces spores that can survive for extended periods in the environment. First isolated by Hall and O'Toole in 1935 and named *Bacillius difficilis* due to difficulties of isolation and culture,[1] by the late 1970s it was identified as the cause of diarrhoea and colitis occurring mostly as a complication of antibiotic therapy. During the past few years, there has been renewed interest in *C. difficile* due to the recognition that the disease is more common, more severe, and more resistant to standard treatment than previously thought.[2,3] *C. difficile* is the most frequent nosocomial (hospital acquired) pathogen, with a mortality and morbidity rate greater that due to better-known methicillin-resistant *Staphylococcus aureus* (MRSA) in many parts of the United States and Europe. It causes a wide range of gastrointestinal diseases ranging from mild diarrhoea to more life threatening pseudomembranous colitis.[4]

The main virulence factors of *C. difficile* are toxins A and B,[5–8] but other virulence factors have also been described, such as adhesins and hydrolytic enzymes.[9–11] Among these, the surface layer proteins (SLPs) are proposed to aid the colonization of the host, and induce both inflammatory and antibody responses in the host. These proteins may adhere directly to the host cell or enhance competition with host flora. As in several other bacteria, the surface layer proteins are the predominant surface proteins in *C. difficile*. However, in contrast to most other species, where the S-layer is composed of one main structural protein, the S-layer of *C. difficile* is constructed from two distinct proteins: a low
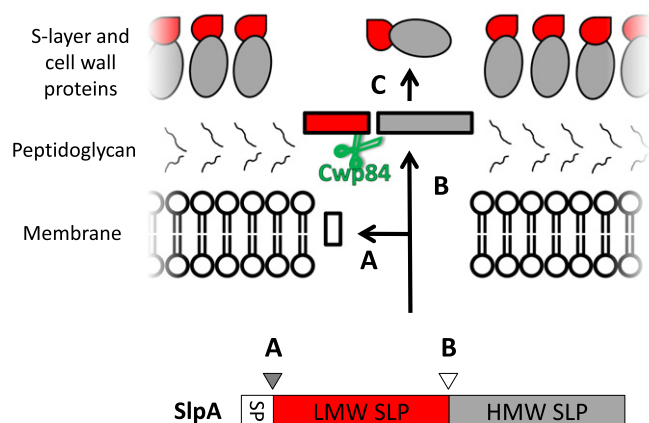
molecular weight (LMW) SLP and high molecular weight (HMW) SLP.[12–14] Both the high and low molecular weight subunits are derived from a single gene, *slpA*.[15,16]

This gene is strongly transcribed during the entire growth phase[17] and up to 400 molecules of SlpA per cell per second are produced, translocated, and cleaved during exponential phase. The translated gene product SlpA is known to undergo at least two rounds of post-translational cleavage: firstly near the N-terminus to remove the signal sequence that targets the nascent chain for translocation across the cytoplasmic membrane, and then internally to release the two mature SLPs, LMW–SLP and HMW–SLP (Fig. 1).[18–20] It is presumed that removal of the signal peptide occurs prior to or during translocation across the membrane, while the second cleavage takes place either within the membrane or within the cell wall. After cleavage, the mature proteins form a complex and reassemble to build the S-layer. We recently reported the structure of the resultant LMW/HMW–SLP complex using a combination of genetic and structural techniques.[21]

As noted above, proteases are of interest as potential virulence factors in *C. difficile*. The MEROPS database (Sanger Institute, UK) lists 139 known and putative peptidases in *C. difficile* including known and unknown functions, and in addition 63 non-peptidase homologues. Given the difficulty of identifying bacterial proteases by homology, this number may considerably underestimate the real number of proteases. However, with the notable exception of the self-processing toxins, only a few protease activities have been reported to date in *C. difficile*, despite its importance as a pathogen. Proteolytic activity has been observed in ten strains of *C. difficile*, with gelatinase, collagenase, hyalurodinase and chondroitin sulfatase activity observed.[22,23] However, the proteases responsible are not characterized and their identities remain

* Corresponding author.
  E-mail address: e.tate@imperial.ac.uk (E.W. Tate).

Figure 1. Post-translational processing of SlpA. The high- and low-molecular weight (HMW and LMW) surface layer proteins (SLPs) are synthesized in *C. difficile* as a 'full-length' precursor, SlpA. (A) During the process of translocation across the membrane the signal peptide (SP) is removed, presumably by a signal peptidase; (B) further processing by the papain class cysteine protease Cwp84 generates the mature S-layer proteins; (C) LMW and HMW SLPs self-assemble to form the S-layer.

unknown. This situation is exacerbated by limited genetic tools in *C. difficile*; the generation of unconditional knockout mutants in *C. difficile* is generally feasible,[24] but it is currently extremely difficult to construct either multiple or conditional knockouts.[6] This has led to notable problems, for example, the difficulty encountered in establishing the essentiality of the major *C. difficile* toxins for virulence.[25] Alternative chemistry-based approaches such as activity-based protein profiling (ABPP) thus have particular interest for characterization of enzymes and post-translational protein processing in such challenging and medically important organisms,[26–29] with a view to identifying novel drug targets for future intervention.[30–33]

In the first reported application of ABPP in Clostridia, we developed activity-based probes (ABPs) against the protease responsible for processing SlpA in *C. difficile*.[34] Using these probes, the protease was identified in vivo as Cwp84, a papain class cysteine protease, and confirmed in a Cwp84 knockout strain.[35] In parallel, the Shen, Garcia and Bogyo labs have exploited a different class of ABPs to characterize the proteolytic activation pathway of the *C. difficile* toxins,[36–39] focusing on toxin in isolation rather than identification of proteases de novo in live bacteria. Here, we report the design and synthesis of a broader range of potential Cwp84 inhibitors, and characterize their ability to inhibit SlpA processing in live bacteria.

2. Results and discussion

2.1. Probe design

Previously-reported epoxysuccinyl inhibitors and probes for S-layer processing are based on the structure of E-64 (Scheme 1), and the design of our new probes took a similar approach of combining a specificity element that binds at the same site as the natural substrate (SlpA), and a reactive electrophilic warhead to trap irreversibly the nucleophilic active center of the protease. The SlpA cleavage site is well-characterized in *C. difficile* reference strain 630, occurring at Ser355 in the sequence ...LETKS-ANDT..., with the LMW and HMW SLPs lying to the N-terminal and C-terminal sides of this cleavage site, respectively.[40] Interestingly, tyrosine is present in place of lysine at P2 in several strains of SlpA. Peptide-based irreversible (or 'suicide') Cwp84 inhibitors reported to date have been proposed to bind the protease with the amide backbone reversed relative to the natural substrate, the most common mode

for inhibition of papain cysteine proteases by E-64 derivatives. In this model, the warhead would sit in the Ser355 (P1) position and the peptide residues would make interactions predominantly via their side chains in the pocket that accommodates the P2 Lys in the majority of strains, or P2 Tyr in some strains. In our previous work, this hypothesis was explored through a structure–activity study that matched this inhibitor binding mode to the P2 position, whilst there was no clear preference shown for residues at P3 or beyond.[34] Inhibitors containing Lys, Arg and Tyr were effective, whilst most other residues were not tolerated with the exception of Leu, which is present in the parent inhibitor E-64 (Scheme 1). In the present study, we explored the influence of inhibitor reactivity on inhibition of SlpA processing across a selection of warheads at the N-terminus of the specificity element, including halo- and acyloxyacetamides. We also examined the influence of backbone direction by placing a Michael acceptor motif at either the N- or C-termini.
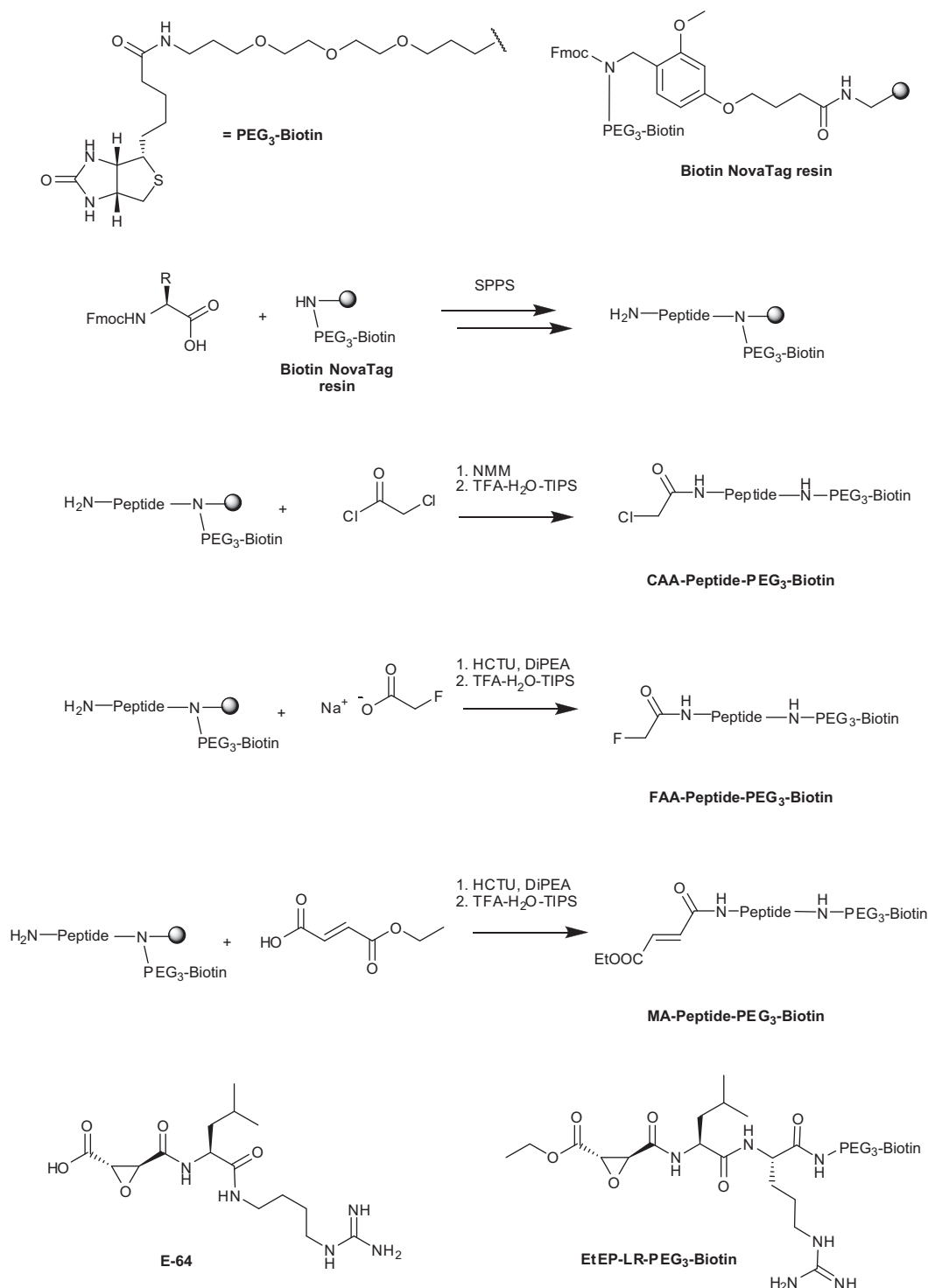
2.2. Probe synthesis

A flexible route was designed to provide ready access to N-terminal warhead combinations (Scheme 1), based on our solid phase route to E-64 analogues.[34] Starting from Biotin NovaTag resin, the specificity element was built up by standard Fmoc/^{t}Bu solid phase peptide synthesis (SPPS), and the warhead moiety coupled directly to the N-terminal amine prior to cleavage from the resin as a C-terminal PEG-Biotin conjugate. N-Terminal chloroacetamide (CAA), fluoroacetamide (FAA) and Michael acceptor (MA) probes were readily synthesized in acceptable overall yields, with a range of specificity elements derived from our previous work on N-terminal epoxysuccinamides (EP).

A solid phase synthesis route was also developed for the synthesis of a prototype C-terminal Michael acceptor probe (Scheme 2). In this approach, the specificity element was anchored to a polystyrene resin via a side chain, enabling the C-terminal acid to be replaced. For this purpose, ethyl-3-[Fmoc-L-Ser]-(*E*)-propenoate, a protected serine analogue bearing a Michael acceptor in place of the acid was synthesised starting from Fmoc-Ser(^{t}Bu)-OH. Formation of the aldehyde Fmoc-Ser(^{t}Bu)-H via reduction of the intermediate benzyl thioester was followed by Wittig olefination and deprotection of the *tert*-butyl ether to render the required analogue in good overall yield. After loading on chlorotrityl resin probes could be constructed containing a C-terminal serine, as found at Ser355 in the SlpA cleavage site. We postulated that this would favor adoption of a backbone comparable to the natural substrate, placing the Cwp84 active site cysteine thiolate in close proximity to the Michael acceptor.

2.3. In vivo inhibition of S-layer processing

We next tested each of the probes synthesized above in an in vivo assay for inhibition of S-layer processing described in our previous work.[34] In brief, bacteria were cultured for 16 h in the presence of an inhibitor (or DMSO control), and surface layer proteins extracted and analyzed by Western blot against an antibody reactive for SlpA. In the presence of an effective inhibitor of S-layer processing SlpA remains unprocessed, and can be detected as a single band at 74 kDa. Amongst the N-terminal warheads, the chloroacetamides (CAA) showed inhibition activity against S-layer processing at 100 μM (overnight culture), and this appears to be directed by the selectivity element (Fig. 2A).

Arg and Leu-Arg sequences were found to have activity whilst Ala and Ser did not, which correlates with the structure–activity pattern in previous studies.[34] The N-terminal FAA and MA probes showed no activity at 100 μM in this assay (data not

Scheme 1. Synthesis of N-terminal chloroacetamide (CAA), fluoroacetamide (FAA), and Michael acceptor (MA) probes. For each warhead, Peptide = -Arg-(R), -Leu-Arg-(LR), -Ala-(A) or -Ser-(S).The specificity element was synthesized on PEG-Biotin-Novatag™ resin using standard SPPS with subsequent coupling of the desired warhead. CAA warhead: ClCOCH$_2$Cl (1.2 equiv), NMM (2.5 equiv), 1 h, rt, overall yield 4–37%. FAA warhead: NaCOOCH$_2$F (5 equiv), HCTU (5 equiv) and DiPEA (6 equiv), 1 h, rt, overall yield 6–15%. *MA warhead*: HOOCCH=CHCOOEt (5 equiv), HCTU (5 equiv) and DiPEA (6 equiv), 1 h, rt, overall yield 10–39%.

shown), implying that their reactivity is insufficient, or that they make poor contacts in the enzyme active site. The activity of the structurally similar CAA and EtEP warheads suggests that the lack of activity against Cwp84 is more likely the result of reduced chemical reactivity in the warhead. In contrast to the N-terminal Michael acceptors, the C-terminal MA inhibitor

Ac-LETKS-MA showed good activity at 100 µM against S-layer processing, comparable to that of our previous E-64-based N-terminal epoxides (Fig. 2A). This activity was maintained in two ABPs, Ac-K(Biotin)YS-MA and Ac-K(Biotin)LS-MA, supporting the hypothesis that these inhibitors bind with the backbone in the 'native' conformation, and tolerate significant divergence at P3.
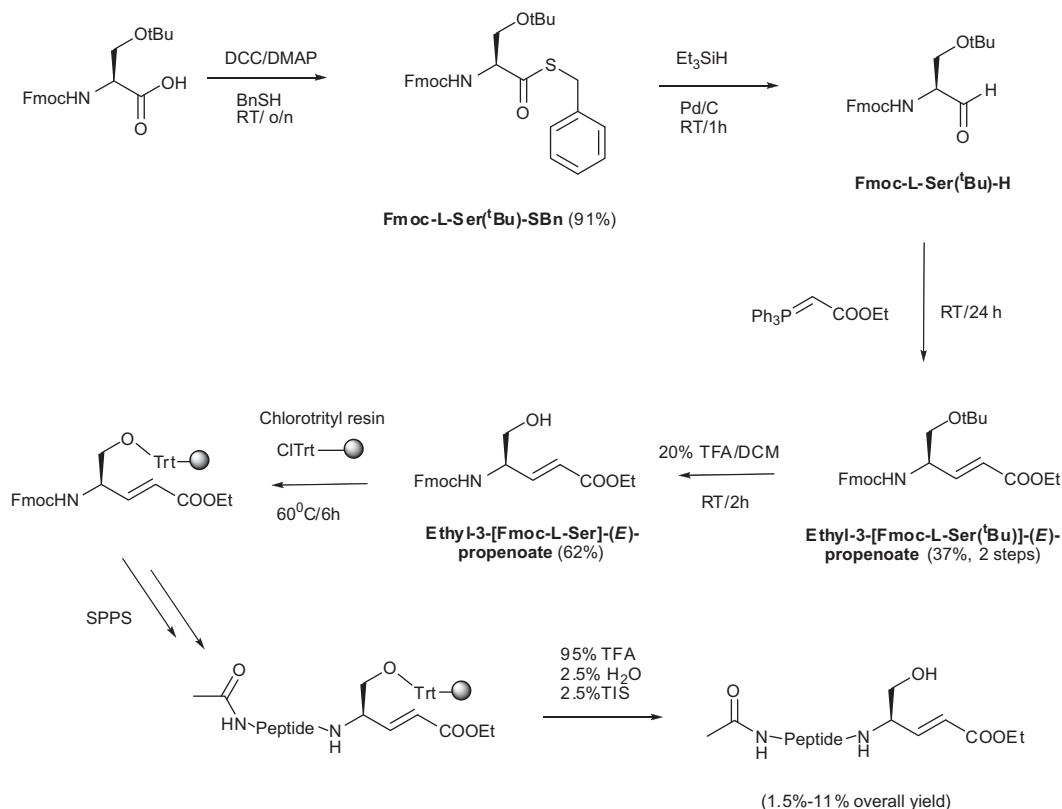
Scheme 2. Synthesis of C-terminal Michael acceptor (MA) probes. Peptide = -Leu-Thr-Glu-Lys-(LETK), -Lys(ε-Biotin)-Leu-(K(Biotin)L), or -Lys(ε-Biotin)-Tyr-(K(Biotin)Y).
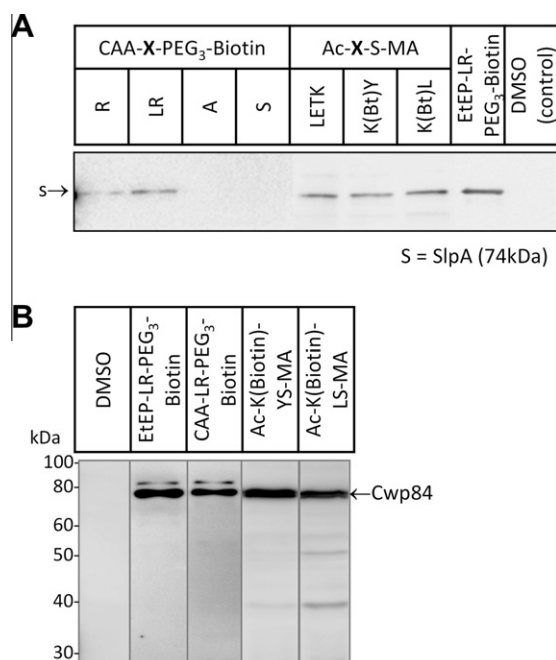
Figure 2. (A) Testing N-terminal chloroacetamide (CAA) and C-terminal Michael acceptor (MA)warheads in vivo against S-layer processing; K(Bt) = ε-biotinyl lysine; structures of compounds are shown in Schemes 1 and 2. C. difficile strain 630 was cultured overnight under anaerobic conditions in brain-heart infusion (BHI) medium, and in the presence of 100 μM of each compound or DMSO vehicle, after which surface layer proteins were isolated, run on SDS–PAGE, and Western blotted against primary anti-LMW SLP antibody followed by horseradish peroxidase (HRP) conjugated secondary antibody, with visualization by chemiluminescence. (B) Labeling activity of selected N-terminal CAA and C-terminal MA activity-based probes in vivo. Inhibited cell lysates were Western blotted against Neutravidin-HRP, and visualized by chemiluminescence.

2.4. In vivo labeling of Cwp84

Finally, we examined the ability of selected active ABPs to label proteins in the S-layer of C. difficile. Following treatment of live cells with the ABPs under the same inhibition conditions as above, the cell lysates were subjected to Western blotting against Neutravidin-horseradish peroxidase (HRP). Interestingly, the chloroacetamide CAA-LR-PEG3-Biotin provided the same labeling pattern as our previously-reported epoxide ABP EtEP-LR-PEG3-Biotin, giving a strong band at 74 kDa and a weaker band at slightly higher apparent molecular weight (Fig. 2B). Previous studies strongly suggest that Cwp84 is processed from an inactive zymogen to an active form in the S-layer,[34,41] and we have previously demonstrated that these bands correspond to Cwp84, likely in pre-processed and processed forms. The two C-terminal Michael acceptor probes Ac-K(Biotin)YS-MA and Ac-K(Biotin)LS-MA both strongly label the lower band, but do not appear to label the higher molecular weight form of Cwp84. Both classes of new probe are remarkably selective, with only minor biotinylated bands seen in addition to Cwp84.

3. Conclusion

In summary, we report the discovery of two new classes of inhibitor for S-layer processing in C. difficile based on N-terminally linked chloroacetamide (CAA) or C-terminal Michael acceptors (MA). The structure–activity relationships for the CAA class follow those of our previously reported N-terminal epoxy succinimidyl esters, and show similarly highly selective irreversible labeling of Cwp84, the protease responsible for SlpA processing. Taken together, these data suggest acommon binding mode for the CAAs, and a mode of action that encompasses aspects of the processing pathway for Cwp84. These experiments also demonstrate that the warhead chemotype

is an important determinant for activity, supporting the hypothesis that inhibition is partly dependent on warhead electrophilicity.

The C-terminal electrophile MA class shows activity comparable to both the epoxide and CAA inhibitors, despite a markedly different configuration. In contrast to previously reported inhibitors, this class has the potential to mimic a well-conserved serine at P1 in SlpA and furthermore presents the remaining upstream residues (P2–P5) without the need to invoke inversion of the main chain. It is interesting to note that analogous inhibitors with an MA warhead at the N-terminus areinactive, suggesting that correct orientation of the backbone and/or the presentation of a serine in P1 contributes significantly to binding and activity. The lack of labeling of the higher molecular weight form of Cwp84 may have its origins in greater selectivity for a putative activated form of the enzyme over an inactive form. An alternative explanation is that the other classes of inhibitor hit an off-target protease that can process and activate Cwp84 in *trans*, and thus the unprocessed form is present only when this putative off-target is also inhibited. Our parallel experiments exploring Cwp84 processing are reported elsewhere,[42] however, both scenarios are consistent with the C-terminal MA possessing greater selectivity, perhaps as a result of a more substrate-mimetic design.

In future it will be interesting to further explore the structure–activity relationships within these new inhibitor classes. For example, the apparent selectivity of the Michael acceptors for the processed form of Cwp84 is significant, as it suggests a chemical approach for probing these processes in vivo. Further screening of C-terminal electrophilic peptide inhibitors against *C. difficile* strains with diverse SlpA cleavage sites could probe the determinants for binding in a systematic fashion, and thus provide a sound structural basis for the design of improved or strain-selective Cwp84 inhibitors. In the present study, we have focused on the surface layer proteins, but it will also be of interest to explore potential membrane and intracellular targets of these and other ABPs in *C. difficile*. More generally, the activity-based labeling provided by these classes of inhibitor adds to a growing range of powerful chemical tools for investigating biological systems.[43–46] They also suggest potential applications in imaging the process of S-layer biogenesis at the level of catalytic activity of the enzymes that mediate post-translational modification of SlpA.

4. Experimental section

All general laboratory chemicals obtained from chemical suppliers (Novabiochem UK or Aldrich Chemical Co.) were used without further purification. Flash chromatography was performed on Screening Devices Silica Gel 60 (0.04–0.063 mm). TLC-analysis was conducted on DC-alufolien (Merck, Kiesel gel 60, F254) with detection by UV-absorption (254 nm). ^{1}H and ^{13}C NMR spectra were recorded on a Brüker AV-400 (400/100 MHz) spectrometer. Chemical shifts (δ) are given in ppm relative to tetramethylsilane and chloroform, H_2O or DMSO residual solvent peak was used as internal standard. Coupling constants are given in Hertz (Hz). Infrared spectra were obtained on PerkinElmer Spectrum 100 FT-IR spectrophotometer, from a thin film deposited onto a sodium chloride plate from dichloromethane. The IR spectra were then analyzed with Spectrum Express, Version 1.01.00 (PerkinElmer). Mass spectra and accurate mass data were obtained by J. Barton at the Chemistry Department Mass Spectrometry Service (Imperial College London) by electrospray ionisation or chemical ionisation techniques.

Solid phase resins: Rink-amide resin and Biotin-PEG-Novatag resin were purchased from Novabiochem UK. The resin substitution ratios were as follows: Rink-amide resin: 0.71 mmol/g; Biotin-PEG-Novatag resin: 0.48 mmol/g; Universal-PEG-Novatag resin: 0.33 mmol/g.

Amino acids: N-α-9-Fluorenylmethoxycarbonyl (N-α-Fmoc) protected amino acids were obtained from Novabiochem UK with the following side chain protecting groups: Arg(Pbf), Lys(Boc), Tyr(*t*Bu), Thr(*t*Bu), Glu(*t*Bu), Ser(*t*Bu).

Other reagents: Peptide synthesis grade dimethylformamide (DMF) was purchased from National Diagnostics, UK. Benzotriazole-1-yloxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBop), N-hydroxybenzotriazole (HOBt), diisopropylethylamine (DIPEA) and 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HBTU) were purchased from Novabiochem UK and Sigma Aldrich. Dry tetrahydrofuran (THF) and DMF, chloroform, dimethylsulfoxide (DMSO), dichloromethane (DCM), methanol, piperidine, thioanisole, *tert*-butylmethylether (TBME), trifluoroacetic acid (TFA), ethanedithiol (EDT), iodoacetonitrile, 3-mercaptoethyl propionate, sodium thiophenolate, monobasic and dibasic sodium phosphate, trimethylsilylisopropane (TIPS) were regent grade from Sigma Aldrich. A 'deprotection mixture' of TFA–H_2O–TIPS (97:2.5:1) was used for cleavage and deprotection reactions.

4.1. General procedures

4.1.1. Solid phase peptide synthesis (SPPS)

Automated solid phase synthesis was performed using an Advanced ChemTech Apex 396 multiple peptide synthesizer (Advanced ChemTech Europe, Cambridge, UK). Peptide synthesis was carried out in peptide synthesis grade DMF. Resin (25 μmol per well) was swelled in DMF for 60 min before coupling each cycle, N-α-amino Fmoc group was treated with 20% (v/v) piperidine in DMF for 15 min, with two further repetitions. A fivefold excess of amino acid over resin reactive groups (125 μmol, 250 μL of 0.5 M solution) was used for each coupling. In situ activation and coupling was carried out for 45 min with a mixture of HBTU/HOBt (125 μmol, 250 μL of 0.5 M solution) and DIPEA (125 μmol, 250 μL of 0.5 M solution). The peptide was washed with DMF (3 × 1 mL) between each deprotection and coupling step. A typical cycle consisted of deprotection, DMF wash, coupling and a further DMF wash. The N-α-Fmoc protection at the final residue was removed at the end of the synthesis under the usual conditions. After synthesis was complete a preactivated solution of (2S,3S)-3-(ethoxycarbonyl)oxirane-2-carboxylic acid (12.01 mg, 3 equiv), HCTU (51.7 mg, 3 equiv) and DiPEA (19.3 mg, 6 equiv) in DMF (1 mL) was added to each well and agitated for 1 h. The peptidyl resin was removed from the synthesizer, washed several times (3 × 2 mL DMF, 3 × 2 mL DCM, 3 × 2 mL methanol, 3 × 2 mL diethylether) and dried in vacuo.

4.1.2. Deprotection and cleavage from resin

Cleavage and deprotection of peptides was achieved by adding 1.5 mL of deprotection mixture to dry peptidyl resin (25 μmol). The mixture was agitated on an orbital shaker for up to 3 h and then filtered. The resin was washed twice with a small volume of TFA and the combined washings and filtrate were precipitated with 10 mL ice cold *tert*-butylmethylether (*t*BME). The mixture was centrifuged at 5000 rpm for 15 min at 0 °C, the supernatant discarded and the remaining peptide washed with a fresh aliquot of tBME. The process was repeated three times to ensure complete removal of all organic impurities. The crude peptide was dried in a desiccator over silica gel to yield an off-white solid.

4.1.3. Peptide purification

Peptide analysis and purification was performed either on Gilson RP-HPLC system or Waters LC–MS system.Purification of crude peptides was performed on a Gilson semi-preparative RP-HPLC system (Anachem Ltd, Luton, UK) equipped with 306

pumps and a Gilson 155 UV/Vis detector. Analytical RP-HPLC was performed on a Gilson analytical HPLC system (Anachem Ltd, Luton, UK) equipped with a Gilson 151 UV/Vis detector and Gilson 234 auto injector. For both HPLC systems, the peptide bond absorption was detected at 223 nm. The following elution methodwas used: 2–98% MeCN in H_2O over 40 min; all solvents used were supplemented with 0.1% TFAand were degassed with helium. LC–MS analysis was performed on Waters HPLC system (Waters 2767 autosampler for samples injection and collection; Waters 515 HPLC pump to deliver the mobile phase to the source; XBridge C18 column (Waters, 4.6 mm D × 100 mm L for analytical and 19 mm D × 100 mm L for preparative); Waters 3100 mass spectrometer with ESI and Waters 2998 Photodiode Array (detection at 200–600 nm)). The following elution method-was used: 5–95% MeOH in H_2O over 15 min. All solvents used were supplemented with 0.1% formic acid and were degassed with helium.

4.2. Solution-phase synthesis and characterization

4.2.1. Fmoc-L-Ser(^{t}Bu)-SBn

DMAP (0.159 g, 1.30 mmol, 0.1 equiv) and DCC (2.820 g, 13.69 mmol, 1.05 equiv) were added sequentially to a solution of Fmoc-L-Ser(tBu)-OH (5.000 g, 13.04 mmol, 1 equiv) and benzyl mercaptan (3.06 mL, 26.08 mmol, 2 equiv) in THF (65 mL) at rt. The cloudy reaction mixture was stirred at rt for 18 h and then filtered. After concentrating the filtrate it was partitioned between EtOAc (70 mL) and 1.0 M HCl (40 mL). The organic layer was dried over Na_2SO_4 and concentrated. Subsequent addition of Et_2O (15 mL) resulted in a white precipitate that was collected by vacuum filtration and washed with petroleum ether. The product was obtained as a yellow solid and used without further purification (5.826 g, 8.18 mmol, 91% yield). R_f = 0.66 (Hexane–EtOAc, 2:1); mp 163–165 °C; δ_H/ppm (400 MHz; $CDCl_3$): 7.79 (2H, d, *J* = 7.50, Fmoc-4 and 5), 7.66–7.63 (2H, m, Fmoc-1 and 8), 7.41 (2H, app td, *J* = 2.72, 7.38, 7.35, Fmoc-3 and 6), 7.36–7.30 (7H, m, Bn, Fmoc-2 and 7), 5.78 (1H, d, *J* = 8.94, NH), 4.55 (2H, dd, *J* = 6.94, 10.40, Fmoc-CH_2), 4.42–4.35 (1H, m, Sα), 4.29 (1H, t, *J* = 6.99, Fmoc-9), 4.16 (2H, dd, *J* = 13.76, 35.67, Sβ), 3.96 (1H, dd, *J* = 2.50, 9.02, Bn-CH_2), 3.57 (1H, dd, *J* = 3.44, 9.03, Bn-CH_2), 1.15 (9H, s, *t*Bu). ν_{max} (neat)/cm^{-1}: 3322, 2954, 2936, 2908, 2853, 1724 (CO), 1684, 1625, 1554, 1509, 1392, 1331, 1234, 1174, 1071, 965, 698. HRMS (ESI): *m/z* = found 490.2039 ([M+H]$^+$), 512.1859 ([M+Na]$^+$), (required 490.2052; $C_{29}H_{31}NO_4S$ [M+H]$^+$).

4.2.2. Fmoc-L-Ser(^{t}Bu)-H

Triethylsilane (6.50 mL, 40.90 mmol, 5 equiv) was slowly added to a suspension of Fmoc-L-(OtBu)Ser-SBn (4 g, 8.18 mmol, 1 equiv) and Pd/C (10%, 2.20 g) in acetone (200 mL) at rt previously de-gassed with N_2. The reaction mixture was stirred for 15 min and then filtered through Celite. The filtrate was concentrated to afford 6.21 g of brown oily crude, which was used without further purification. R_f = 0.57 (Hexane–EtOAc, 3:2); δ_H/ppm (400 MHz; $CDCl_3$): 9.66 (1H, s, CHO), 7.81 (2H, d, *J* = 7.34, Fmoc-4 and 5), 7.65 (2H, dt, *J* = 5.74, 11.49, Fmoc-1 and 8), 7.44 (2H, t, *J* = 7.29, Fmoc-3 and 6), 7.36 (7H, dd, *J* = 4.98, 9.52, Fmoc-2 and 7), 5.70 (1H, d, *J* = 7.30, NH), 4.46 (2H, dd, *J* = 1.41, 6.38, Fmoc-CH_2), 4.41 (2H, dd, *J* = 5.84, 10.17, Sα), 4.29 (1H, t, *J* = 6.86, 13.98, Fmoc-9), 4.00 (1H, dd, *J* = 3.10, 9.30, Sβ), 3.67 (1H, dd, *J* = 4.20, 9.38, Sβ), 1.20 (9H, s, *t*Bu). ESI-MS: *m/z* = found 390.1681 [M+Na]$^+$, (required 390.1681; $C_{22}H_{25}NO_4Na$).

4.2.3. Ethyl-3-[Fmoc-L-Ser(^{t}Bu)]-(*E*)-propenoate

(Carbethoxymethylene)-triphenylphosphane (2.137 g, 6.135 mmol, 1.5 equiv) was added to a solution of crude Fmoc-L-

(O^{t}Bu)Ser-H (6.210 g) in THF (60 mL) at rt. The resulting mixture was stirred for 24 h and then concentrated under reduced pressure. The residue was purified by silica gel flash column chromatography (Hexane–EtOAc, 9:1) to afford a white solid (0.662 g, 37% yield from Fmoc-L-(O^{t}Bu)Ser-SBn). R_f = 0.51 (Hexane–EtOAc, 2:1); mp 50–53 °C; δ_H/ppm (400 MHz; $CDCl_3$): 7.80 (2H, d, *J* = 7.49, Fmoc-4 and 5), 7.63 (2H, d, *J* = 7.12, Fmoc-1 and 8), 7.43 (2H, t, *J* = 7.40, Fmoc-3 and 6), 7.35 (2H, t, *J* = 7.11, Fmoc-2 and 7), 6.96 (1H, dd, *J* = 4.64, 15.63, Sα-C*H*), 6.00 (1H, d, *J* = 15.66, C*H*COOEt), 5.31 (1H, d, *J* = 7.85, NH), 4.49 (1H, s, Sα), 4.45 (2H, d, *J* = 6.88, Fmoc-CH_2), 4.28 (1H, t, *J* = 6.92, Fmoc-9), 4.23 (2H, q, *J* = 7.08, COOCH_2CH$_3$), 3.54–3.47 (2H, m, Sβ), 1.33 (3H, t, *J* = 7.15, COOCH$_2$CH_3), 1.21 (9H, s, *t*Bu). ν_{max} (neat)/cm^{-1}: 3342, 1725 (CO), 1697, 1541, 1182, 1032, 881, 738. ESI-MS: *m/z* = found 438.2276 ([M+H]$^+$), 460.2100 ([M+Na]$^+$), (required 460.2100, $C_{26}H_{31}NO_5Na$).

4.2.4. Ethyl-3-Fmoc-L-Ser-(*E*)-propenoate

TFA (2 mL) was added to a solution of ethyl-3-[Fmoc-L-Ser(^{t}Bu)]-(*E*)-propenoate (0.318 g, 0.833 mmol) in DCM (8 mL) at rt. The reaction mixture was stirred for 2 h, and then the volatiles were removed under reduced pressure. The residue was triturated with a 1:1 mixture of Et_2O and hexanes (3 mL), and the resulting white solid was collected by filtration and washed thoroughly with cold Et_2O–Hexane (1:1) to afford the product (0.197 g, 0.518 mmol, 62% yield). R_f = 0.51 (Hexane–EtOAc, 2:1); mp 110–113 °C; δ_H/ppm (400 MHz; $CDCl_3$): 7.78 (2H, d, *J* = 7.50, Fmoc-4 and 5), 7.60 (2H, d, *J* = 7.05, Fmoc-1 and 8), 7.41 (2H, t, *J* = 7.40, Fmoc-3 and 6), 7.32 (2H, t, *J* = 7.26, Fmoc-2 and 7), 6.92 (1H, dd, *J* = 4.19, 15.67, Sα-C*H*), 6.00 (1H, d, *J* = 15.70, C*H*COOEt), 5.50 (1H, d, *J* = 7.80, NH), 4.50–4.36 (3H, m, Fmoc-CH_2, Sα), 4.29–4.09 (3H, m,CH$_2$CH$_3$, Fmoc-9), 3.75 (2H, s, Sβ), 1.31 (3H, t, *J* = 7.14, COOCH$_2$CH_3). δ_C/ppm (100 MHz; $CDCl_3$): 166.26, 156.23 (CO), 145.10 (CH-Fmoc), 143.72, 141.33 (*C*-Fmoc), 127.78, 127.12, 125.00, 122.67, 120.03 (CH-Fmoc, CHCH), 66.91, 64.03, 60.76 (CH$_2$), 53.80 (CH-Fmoc), 47.19 (Cα), 14.21 (CH$_3$). ν_{max} (neat)/cm^{-1}: 3320, 2974, 2878, 2839, 1716 (CO), 1692, 1535, 1461, 1377, 1287, 1212, 1016, 736. ESI-MS: *m/z* = 382.1650 ([M+H]$^+$), 405.1526 ([M+Na]$^+$), (required 382.1576, $C_{22}H_{24}NO_5$).

4.2.5. Ethyl-3-[Fmoc-L-Ser(ClTrt resin)]-(*E*)-propenoate

2-Chlorotrityl chloride resin (0.199 g, 1 equiv) was added to a solution of ethyl-3-Fmoc-L-Ser-(*E*)-propenoate (0.197 g, 0.518 mmol, 2 equiv) and pyridine (83.73 μL, 1.035 mmol, 4 equiv) in dry THF (2 mL). The reaction mixture was heated at 60 °C with stirring for 6 h. The resin was filtered and washed with DCM–MeOH–DiPEA (17:2:1) and 3 × DCM, 2 × DMF, 2 × DCM. Finally, the resin was dried in vacuo over KOH. The yield of the loading was 72% as determined by the Fmoc-test.

4.2.6. Ethyl-3-[-L-Ser-Lys-Thr-Glu-Leu-Ac]-(*E*)-propenoate (Ac-LETKS-MA)

Ac-LETKS-MA was synthesised by standard SPPS, starting from the resin above. δ_H/ppm (400 MHz; D_2O): 6.84 (1H, ddd, *J* = 3.41, 4.81, 15.79, CHC*H*COOEt), 5.90 (1H, ddd, *J* = 1.72, 11.57, 15.86, C*H*CHCOOEt), 4.88 (2H, d, *J* = 3.76, OH), 4.55 (1H, dt, *J* = 3.12, 6.34, NHC*H*CO), 4.49 (1H, dd, *J* = 5.03, 9.04, NHC*H*CO), 4.36 (2H, dd, *J* = 5.39, 8.99, NHC*H*CO), 4.30 (1H, dd, *J* = 5.55, 8.79, NHC*H*CO), 4.25–3.98 (6H, m, CH$_2$(OEt), 4 × NHC*H*CO), 3.70–3.56 (2H, m, C*H*(Thr), NHC*H*CH), 2.91 (2H, t, *J* = 7.33, CH$_2$), 2.78 (1H, t, *J* = 6.26, CH$_2$(Ser)), 2.60–2.55 (1H, m, CH$_2$(Ser)), 2.44–2.25 (2H, m, CH$_2$), 2.13–2.00 (1H, m, C*H*(Leu)), 1.94 (3H, s, CH$_3$(Ac)), 1.81–1.29 (10H, m, 5 × CH$_2$), 1.19 (3H, t, *J* = 7.18, CH$_3$(Thr)), 1.16–1.09 (3H, m, CH$_3$(OEt)), 0.84 (3H, d, *J* = 6.24, CH$_3$(Leu)), 0.80 (3H, d, *J* = 6.24, CH$_3$(Leu)). R_t: 5.26.

4.3. Peptide characterization data

Name	ESI-MS [M+H$^+$]		R_t	Yield (%)
	Calculated	found		
CAA-LR-PEG$_3$-Biotin	792.4209 ($C_{34}H_{62}ClN_9O_8S$)	792.4240	9.87	28
CAA-A-PEG$_3$-Biotin	594.2728 ($C_{25}H_{45}ClN_5O_7S$)	594.2730	9.45	5
CAA-R-PEG$_3$-Biotin	679.3337 ($C_{28}H_{52}ClN_8O_7S$)	679.3362	8.71	37
CAA-S-PEG$_3$-Biotin	610.2677 ($C_{25}H_{45}ClN_5O_8S$)	610.2681	9.16	7
FAA-LR-PEG$_3$-Biotin	776.4504 ($C_{34}H_{63}FN_9O_8S$)	776.4523	9.59	7
FAA-A-PEG$_3$-Biotin	578.3024 ($C_{25}H_{45}FN_5O_7S$)	578.3052	9.18	15
FAA-R-PEG$_3$-Biotin	663.3664 ($C_{28}H_{52}FN_8O_7S$)	663.3661	8.46	9
FAA-S-PEG$_3$-Biotin	594.2973 ($C_{25}H_{45}FN_5O_8S$)	594.2974	8.91	1
MA-LR-PEG$_3$-Biotin	842.4810 ($C_{38}H_{68}N_9O_{10}S$)	842.4819	10.39	10
MA-A-PEG$_3$-Biotin	644.3329 ($C_{29}H_{50}N_5O_9S$)	644.3331	10.18	17
MA-R-PEG$_3$-Biotin	729.3969 ($C_{32}H_{57}N_8O_9S$)	729.3961	9.33	39
MA-S-PEG$_3$-Biotin	660.3278 ($C_{29}H_{50}N_5O_{10}S$)	660.3292	9.92	25
Ac-LTEKS-MA	673.3772 ($C_{30}H_{53}N_6O_{11}$)	673.3775	5.26	1.5
Ac-K(Biotin)Y-MA	832.4279 ($C_{40}H_{62}N_7O_{10}S$)	832.4247	10.35	11
Ac-K(Biotin)L-MA	782.4486 ($C_{37}H_{64}N_7O_9S$)	782.4479	10.55	4

Note: all peptides were synthesized by standard SPPS (see above); % overall yield is given relative to total resin load prior to loading of the first amino acid.

4.4. Bacterial culture and biochemical methods

4.4.1. *C. difficile* culture conditions

C. difficile 630 was provided by Dr. Peter Mullany, Eastman Dental Institute, London, and has been fully sequenced by the Wellcome Trust Sanger Institute.[47] *C. difficile* was routinely cultured either on blood agar base II (Oxoid Ltd, Basingstoke, UK) supplemented with 7% horse blood (TCS Biosciences, BotolphClaydon, UK); or in brain-heart infusion (BHI) broth (Oxoid). Culture was undertaken in an anaerobic cabinet (Don Whitley Scientific, Shipley, UK) at 37 °C in a reducing anaerobic atmosphere (10% CO_2, 10% H_2, 80% N_2). To obtain cells during exponential phase, an overnight (ca. 16–18 h) liquid culture was inoculated 1:20 v/v into BHI that had been pre-equilibrated (at least 2 h) to the temperature and atmosphere of the anaerobic cabinet, and the culture was shaken vigorously for 15 s to mix thoroughly. During growth, the culture liquid was agitated gently using the shaker at 200 rpm to allow cells to disperse. In order to monitor the growth of *C. difficile*, the OD_{600} of a sample was determined using WPA biowave CD8000 cell density meter (Isogen Life Science, UK). To extract surface layer proteins *C. difficile* bacteria from a 16–18 h culture in BHI were harvested by centrifugation (15 min at 3500*g*) and resuspended in 0.04 volumes (relative to culture volume) of 0.2 M glycine–HCl pH 2.2. After rotating incubation for 30 min at room temperature, the mixture was centrifuged to remove bacteria (5 min at 10,000 *g*) and the supernatant neutralised by adding 2 M Tris portion-wise (5 μL at a time) until the desired pH was obtained.

4.4.2. SDS–PAGE analysis

Standard glycine–Tris gels at 10% or 12% acrylamide were prepared using mini-protean kits (Bio-Rad Laboratories). The separating gel (10 or 12% 39:1 acrylamide: bisacrylamide, 375 mM Tris pH 8.8, 0.1% SDS, 0.5% APS, 0.07% TEMED) was poured and ethanol carefully layered on top. Once the gel was set, the ethanol was discarded and the stacking gel (5% 39:1 acrylamide:bisacrylamide, 62.5 mM Tris pH 6.8, 0.1% SDS, 0.1% APS, 0.24% TEMED) was poured and the comb put in place to form the wells. Once set, the comb was removed and the gel was immersed in the running buffer. Protein samples were diluted in sample loading buffer. These, along with a 10 μL aliquot of Broad Range Protein Marker (NEB), and a 3.5 μL aliquot of biotinylated protein ladder (Cell Signalling Technology) were heated to 100 °C for 5 min. Samples were loaded into the wells of the gel and a constant voltage of 200 V applied for 40 min. For general visualisation of proteins, gels were stained in Coomassie blue staining buffer for 4–16 h, followed by destaining. Gels were image scanned and then rehydrated in 10% acetic acid and dried between two cellulose membranes.

4.4.3. Western blotting

Gels were transferred to PVDF membrane for immunoblotting. SDS–polyacrylamide gels were transferred to Trans-blot™ transfer medium (Bio-Rad Laboratories) PVDF membrane using a semi-dry protocol. The membrane was prepared for transfer by incubation sequentially in methanol (1–2 s), H_2O (1–2 min) and finally in transfer buffer (2 min). Gel and filter paper were pre-soaked in transfer buffer. Filter paper, membrane and gel were stacked in the Trans-blot™ SD semi-dry transfer cell (Bio-Rad) as a sandwich in the following order: filter paper-membrane-gel-filter paper. The transfer was performed at 15 V for 15 min. After transfer, the membrane was soaked in methanol (5 mL) and dried for 20 min, which avoids the additional blocking step typically required for minimising non-specific binding or back ground signal. The membrane was incubated with primary antibody diluted to the specified ratio v/v in 10 mL PBS-T +3% (w/v) non-fat powdered milk for at least 1 h with gentle rocking. After three washes in 10 mL PBS-T, secondary antibody diluted to the specified ratio v/v in 10 mL PBS-T+3% (w/v) non-fat powdered milk was added and the membrane rocked gently for at least 30 min. The membrane was washed three times in PBS-T. Signal was detected using ECL reagents (Amersham Biosciences) or SuperSignal® West Pico Chemiluminescent Substrate (Pierce) followed by image capture using a LAS-3000 Image Reader (Fujifilm), with appropriate sizing carried out using Microsoft Office PowerPoint 2007. A rabbit α-LMW-SLP antibody, raised against 630 LMW SLP, was used at 1/200,000 dilution to visualize SlpA as reported in our previous work.[34] Goat α-rabbit-HRP was used as secondary (DakoCytomation), at 1/2000 dilution.

4.4.4. NeutrAvidin™–HRP immunoblotting

SDS–polyacrylamide gels were transferred using the method describe above, and the membranes were blocked for 1 h in 10 mL PBS-T containing 5% BSA (w/v) before washing for 3×10 min in 10 mL PBS-T. The blots were probed using NeutrAvidin™-HRP at 1:5000 dilution in 10 mL PBS-T for 1 h at room temperature. The membranes were then washed with 10 mL PBS-T for 3×5 min before detection as described above.

References and notes

1. Hall IC, O. T. E. *Am. J. Dis. Child.* **1935**, *49*, 390.
2. Kelly, C. P.; LaMont, J. T. *N. Eng. J. Med.* **2008**, *359*, 1932.
3. Rupnik, M.; Wilcox, M. H.; Gerding, D. N. *Nat. Rev. Microbiol.* **2009**, *7*, 526.
4. Bartlett, J. G. *J. Clin. Gastroenterol.* **2007**, *41*, S24.
5. Thelestam, M.; Chaves-Olarte, E. *Cytotoxic effects of the Clostridium difficile toxins*, 2000.
6. Kuehne, S. A.; Cartman, S. T.; Heap, J. T.; Kelly, M. L.; Cockayne, A.; Minton, N. P. *Nature* **2010**, *467*, 711.
7. Jank, T.; Giesemann, T.; Aktories, K. *Glycobiology* **2007**, *17*, 15R.
8. Pothoulakis, C. *Ann. N.Y. Acad. Sci.* **2000**, *915*, 347.
9. Sullivan, N. M.; Pellett, S.; Wilkins, T. D. *Infect. Immun.* **1982**, *35*, 1032.
10. Borriello, S. P.; Davies, H. A.; Kamiya, S.; Reed, P. J.; Seddon, S. *Rev. Infect. Dis.* **1990**, *12*, S185.
11. Karjalainen, T.; Barc, M. C.; Collignon, A.; Trolle, S.; Boureau, H.; Cottelaffitte, J.; Bourlioux, P. *Infect. Immun.* **1994**, *62*, 4347.
12. Cerquetti, M.; Molinari, A.; Sebastianelli, A.; Diociaiuti, M.; Petruzzelli, R.; Capo, C.; Mastrantonio, P. *Microb. Pathog.* **2000**, *28*, 363.
13. Kawata, T.; Takeoka, A.; Takumi, K.; Masuda, K. *FEMS Microbiol. Lett.* **1984**, *24*, 323.
14. Takeoka, A.; Takumi, K.; Koga, T.; Kawata, T. *J. Gen. Microbiol.* **1991**, *137*, 261.
15. Calabi, E.; Ward, S.; Wren, B.; Paxton, T.; Panico, M.; Morris, H.; Dell, A.; Dougan, G.; Fairweather, N. *Mol. Microbiol.* **2001**, *40*, 1187.
16. Waligora, A.-J.; Hennequin, C.; Mullany, P.; Bourlioux, P.; Collignon, A.; Karjalainen, T. *Infect. Immun.* **2001**, *69*, 2144.
17. Savariau-Lacomme, M. P.; Lebarbier, C.; Karjalainen, T.; Collignon, A.; Janoir, C. *J. Bacteriol.* **2003**, *185*, 4461.
18. Calabi, E.; Calabi, F.; Phillips, A. D.; Fairweather, N. F. *Infect. Immun.* **2002**, *70*, 5770.
19. Wright, A.; Drudy, D.; Kyne, L.; Brown, K.; Fairweather, N. F. *J. Med. Microbiol.* **2008**, *57*, 750.
20. Wright, A.; Wait, R.; Begum, S.; Crossett, B.; Nagy, J.; Brown, K.; Fairweather, N. *Proteomics* **2005**, *5*, 2443.
21. Fagan, R. P.; Albesa-Jove, D.; Qazi, O.; Svergun, D. I.; Brown, K. A.; Fairweather, N. F. *Mol. Microbiol.* **2009**, *71*, 1308.
22. Poilane, I.; Karjalainen, T.; Barc, M. C.; Bourlioux, P.; Collignon, A. *Can. J. Microbiol.* **1998**, *44*, 157.
23. Steffen, E. K.; Hentges, D. J. *J. Clin. Microbiol.* **1981**, *14*, 153.
24. Heap, J. T.; Pennington, O. J.; Cartman, S. T.; Carter, G. P.; Minton, N. P. *J. Microbiol. Methods* **2007**, *70*, 452.
25. Lyras, D.; O'Connor, J. R.; Howarth, P. M.; Sambol, S. P.; Carter, G. P.; Phumoonna, T.; Poon, R.; Adams, V.; Vedantam, G.; Johnson, S.; Gerding, D. N.; Rood, J. I. *Nature* **2009**, *458*, 1176.
26. Berry, A. F.; Heal, W. P.; Tarafder, A. K.; Tolmachova, T.; Baron, R. A.; Seabra, M. C.; Tate, E. W. *ChemBioChem* **2010**, *11*, 771.
27. Heal, W. P.; Dang, T. H.; Tate, E. W. *Chem. Soc. Rev.* **2011**, *40*, 246.
28. Heal, W. P.; Wickramasinghe, S. R.; Tate, E. W. *Curr. Drug Discov. Technol.* **2008**, *5*, 200.
29. Thomas, J. C.; Green, J. L.; Howson, R. I.; Simpson, P.; Moss, D. K.; Martin, S. R.; Holder, A. A.; Cota, E.; Tate, E. W. *Mol. Biosyst.* **2010**, *6*, 494.
30. Thongyoo, P.; Tate, E. W.; Leatherbarrow, R. J. *Chem. Commun.* **2006**, 2848.
31. Thongyoo, P.; Jaulent, A. M.; Tate, E. W.; Leatherbarrow, R. J. *ChemBioChem* **2007**, *8*, 1107.
32. Thongyoo, P.; Roque-Rosell, N.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.* **2008**, *6*, 1462.
33. Thongyoo, P.; Bonomelli, C.; Leatherbarrow, R. J.; Tate, E. W. *J. Med. Chem.* **2009**, *52*, 6197.
34. Dang, T. H.; de la Riva, L.; Fagan, R. P.; Storck, E. M.; Heal, W. P.; Janoir, C.; Fairweather, N. F.; Tate, E. W. *ACS Chem. Biol.* **2010**, *5*, 279.
35. Kirby, J. M.; Ahern, H.; Roberts, A. K.; Kumar, V.; Freeman, Z.; Acharya, K. R.; Shone, C. C. *J. Biol. Chem.* **2009**, *284*, 34666.
36. Lupardus, P. J.; Shen, A.; Bogyo, M.; Garcia, K. C. *Science* **2008**, *322*, 265.
37. Puri, A. W.; Lupardus, P. J.; Deu, E.; Albrow, V. E.; Garcia, K. C.; Bogyo, M.; Shen, A. *Chem. Biol.* **2011**, *17*, 1201.
38. Shen, A.; Lupardus, P. J.; Albrow, V. E.; Guzzetta, A.; Powers, J. C.; Garcia, K. C.; Bogyo, M. *Nat. Chem. Biol.* **2009**, *5*, 469.
39. Shen, A.; Lupardus, P. J.; Gersch, M. M.; Puri, A. W.; Albrow, V. E.; Garcia, K. C.; Bogyo, M. *Nat. Struct. Mol. Biol.* **2011**, *18*, 364.
40. Calabi, E.; Fairweather, N. *J. Bacteriol.* **2002**, *184*, 3886.
41. Janoir, C.; Pechine, S.; Grosdidier, C.; Collignon, A. *J. Bacteriol.* **2007**, *189*, 7174.
42. De la Riva, L.; Willing, S.; Tate, E. W.; Fairweather, N. *J. Bacteriol.* **2011**, *193*, 3276.
43. Heal, W. P.; Wickramasinghe, S. R.; Bowyer, P. W.; Holder, A. A.; Smith, D. F.; Leatherbarrow, R. J.; Tate, E. W. *Chem. Commun.* **2008**, 480.
44. Heal, W. P.; Wickramasinghe, S. R.; Leatherbarrow, R. J.; Tate, E. W. *Org. Biomol. Chem.* **2008**, *6*, 2308.
45. Heal, W. P.; Tate, E. W. *Org. Biomol. Chem.* **2010**, *8*, 731.
46. Heal, W. P.; Jovanovic, B.; Bessin, S.; Wright, M. H.; Magee, A. I.; Tate, E. W. *Chem. Commun.* **2011**, *47*, 4081.
47. Sebaihia, M.; Wren, B. W.; Mullany, P.; Fairweather, N. F.; Minton, N.; Stabler, R.; Thomson, N. R.; Roberts, A. P.; Cerdeno-Tarraga, A. M.; Wang, H.; Holden, M. T.; Wright, A.; Churcher, C.; Quail, M. A.; Baker, S.; Bason, N.; Brooks, K.; Chillingworth, T.; Cronin, A.; Davis, P.; Dowd, L.; Fraser, A.; Feltwell, T.; Hance, Z.; Holroyd, S.; Jagels, K.; Moule, S.; Mungall, K.; Price, C.; Rabbinowitsch, E.; Sharp, S.; Simmonds, M.; Stevens, K.; Unwin, L.; Whithead, S.; Dupuy, B.; Dougan, G.; Barrell, B.; Parkhill, J. *Nat. Genet.* **2006**, *38*, 779.