Species-specific detection of *C. difficile* using targeted antibody design

Lawry, B.M[1#].,Johnson, C.L[2#]., Flanagan, K[3]., Spoors, J.A.[2]., McNeil, C.J[2]., Wipat, A[3]., Keegan, N[2*].

[1] School of Biomedical Sciences, Faculty of Medical Sciences, Newcastle University, Newcastle-Upon-Tyne, NE2 4HH, UK

[2] Diagnostic and Therapeutic Technologies, Institute of Cellular Medicine, Newcastle University, Newcastle-Upon-Tyne, NE2 4HH, UK

[3] School of Computing, Urban Sciences Building, Newcastle University, Newcastle-Upon-Tyne, NE4 5TG, UK

[#] Both authors contributed equally to this work

* Corresponding Author, Email: Neil.Keegan@ncl.ac.uk

Abstract

Clostridium difficile is a Gram-positive, spore-forming bacterium that continues to present a world-wide problem in healthcare settings. The bacterium causes disease, the symptoms of which include diarrhoea, fever, nausea, abdominal pain and even death. Despite the prevalence of the disease, the diagnosis of *C. difficile* infection is still challenging, with a variety of methods available, each varying in their effectiveness. In this work we sought to identify a new biomarker for *C. difficile,* develop affinity reagents and design a diagnostic assay for *C. difficile* infection which could be used in a typical two-step testing algorithm. Initially a bioinformatics pipeline was developed that identified a surface associated biomarker "AKDGSTKEDQLVDALA" present in all *C. difficile* strains sequenced to-date and unique to the *C. difficile* species. Monoclonal antibodies were subsequently raised against peptides corresponding to the biomarker sequence. During characterisation studies, monoclonal antibody 521 (mAb521) was shown to bind all known *C. difficile* surface layer types, but not closely related strains. Surface plasmon resonance measurements were used to calculate an apparent equilibrium dissociation constant of 36.5 nM between the purified protein target containing the biomarker (surface layer protein A) and mAb521. We demonstrate a limit of detection of 12.4 ng/ml against surface layer protein A and 1.7 x 10^6 cells/ml in minimally processed *C. difficile* cultures. The utility of this computational approach to antibody design for diagnostic tests is the ability to produce antibodies which can act as universal species identifiers whilst mitigating the likelihood of false-positive detection by intelligently screening potential biomarkers against RefSeq data for other non-target bacteria.

Introduction

Clostridium difficile infection (CDI) continues to be a significant economic burden, particularly in healthcare settings, with the spectrum of clinical disease ranging from mild diarrhoea to death. *C. difficile* includes both pathogenic (toxin-producing) and non-pathogenic strains, with

both strains able to colonise their hosts, however only toxin producing strains are associated with disease [1].

The choice of laboratory test for *C. difficile* is a contentious one with a multitude of tests available. One commonly used test is an enzyme immunoassay (EIA) which detects glutamate dehydrogenase (GDH), a secreted enzyme found in stool samples during CDI which converts L-glutamate into α-ketoglutarate. This pan-*C. difficile* antigen can be tested for in a few hours using the glutamate dehydrogenase enzyme immunoassay (GDH EIA), however it does not determine whether the *C. difficile* strain is actively producing toxin [2]. To elucidate this, the toxins associated with CDI, toxin A/B, are typically assayed using a second EIA, again typically taking a few hours to process [3]. More recently, nucleic acid amplification tests (NAAT) have been increasingly used for toxin A/B testing [4]. However, when relying solely on molecular methods for CDI diagnosis, it has been suggested this results in the over-diagnosis of CDI through the detection of asymptomatic carriage, resulting in overtreatment and increased healthcare costs [5]. The current protocol for testing and diagnosing CDI in the UK is based on guidelines from the Department of Health, who advise that organisations adhere to a two-step testing algorithm which consists of a GDH EIA (or NAAT) measurement to screen samples, followed by a toxin A/B EIA [6]. If the first test in the algorithm is negative, the second test does not need to be performed. The authors of the guidelines acknowledge that no test or combination of tests is infallible and the clinical condition of the patient should always be taken into consideration when making management and treatment choices.

Rates and severity of CDI in hospitals in Europe and North America have increased since 2000 and correlate with the dissemination of an epidemic strain, 027, characterised by higher than usual toxin A/B production in addition to demonstrating resistance to fluoroquinolone and cephalosporin antibiotics [7]. Typing studies from 2008 revealed that PCR ribotype 027 was the most common strain isolated from symptomatic patients, accounting for over 41.3 % of isolates in English hospitals, followed by type 106 (20.2 %) and type 001 (7.8 %) [8]. A mixture of 44 other PCR ribotypes accounted for the remaining 28.9 % of isolates [8]. More recent data from the *C. difficile* Ribotyping Network indicate that 027 prevalence has decreased in the United Kingdom, potentially because of efforts to reduce fluoroquinolone and cephalosporin usage [9].

Although a relatively simple and inexpensive assay, concerns have been raised about the sensitivity of the commonly used GDH EIA, particularly for non-027 strains [10]. This is not because of strain-dependent difference in GDH expression but most likely because 027 strains tend to attain higher organism burdens, which means that GDH is easier to detect using the GDH EIA as it is present at a greater multiplicity [11]. Although the sensitivity of the GDH EIA is comparable to PCR for 027 strains, it has been demonstrated that PCR is significantly more sensitive than the GDH EIA for detecting strains of ribotypes other than 027 [12]. Another disadvantage of the GDH EIA is that antiserum against *C. difficile* GDH has the potential to cross-react with GDH from other anaerobes [3]. Despite the sensitivity and specificity of the GDH EIA being inferior to PCR it is still widely used in clinical diagnostic algorithms, because it is cheaper and quicker than standard PCR.

Bearing in mind the current limitations outlined above, we sought to create a new diagnostic test for CDI which could act as an alternative to the GDH EIA. A major hurdle in the development of any new medical diagnostic test for infectious disease is the development of suitable affinity reagents, such as antibodies, that bind specifically to the target organism. To address this challenge, and to find a new specific biomarker for *C. difficile*, a bioinformatics method was developed building on a previously described approach [13]. This approach is able to predict biomarkers for a given group of bacteria from their genome sequence data. In this work we demonstrate the utility of a computational approach for generating species-specific *C. difficile* monoclonal antibodies (mAbs) against a biomarker ubiquitous to all *C. difficile* strains sequenced to-date.

Materials and Methods

Biomarker identification in *C. difficile* using a bioinformatics approach

There are several idealised requirements for pathogenic biomarkers in a diagnostic setting. Firstly, the biomarker must exist in all strains that the diagnostic test is required to identify. Secondly, the biomarker must *not* occur in any other organism; i.e. the biomarker must be globally unique to the group of interest in order to avoid false-positive results. Our team inserted a third project specific requirement; the biomarker must encode an epitope on the surface of the cell in order to be accessible with minimal sample preparation. This rationale was inserted, as the biomarker will be more amenable to point of care scenarios if limited sample processing is required.

A custom bioinformatics pipeline building on principles first described by Flanagan and co-workers was exploited in order to produce a set of putative biomarkers in *C. difficile* that conform to the specific selection criteria outlined above [13]. For a more detailed description of the biomarker identification process see Supporting Information.

Bacterial storage, growth and strains used in this work

All of the following cultures were grown anaerobically at 37°C using anaerobic jars and Anaerocult A gas packs (Merck). *C. difficile* strains were plated on *C. difficile* agar base (Sigma), with 7 % sheep's blood and *C. difficile* C.D.M.N. – selective supplement (Oxoid) or cultured in brain heart infusion (BHI) (Sigma) or cooked meat broth (Sigma). *Clostridium sordellii*, *Clostridium perfringens*, *Peptostreptococcus anaerobius* and *Enterococcus faecalis* were cultured in BHI or on soy-tryptone agar (15 g/l tryptone, 5 g/l soytone, 5 g/l NaCl, 15 g/l agar) supplemented with 7 % sheep's blood. *Clostridium hiranonis* was cultured in PY medium (tryptone 5 g/l, peptone (pepsin digested) 5 g/l, yeast 10 g/l, L-cysteine 0.5 g/l, D-glucose 5 g/l) + 40 ml salt solution ($CaCl_2.H_2O$ 0.25 g/l, $MgSO_4$ x 7 H_2O 0.5 g/l, K_2HPO_4 1 g/l, KH_2PO_4 1 g/l, $NaHCO_3$ 10 g/l, NaCl 2 g/l). *Escherichia coli*, *Pseudomonas aeruginosa*, *Citrobacter freundii*, *Proteus mirabilis* and *Klebsiella pneumoniae* strains were all cultured aerobically at 37°C in LB media (bacto-tryptone 10 g/l, yeast 5 g/l, NaCl 10 g/l).

Strains used to probe for interaction with mAb521 are shown in Table S-1. The *C. difficile* Oxford strains (Ox) used in this study are representative members from each of the 12 surface layer types (SLTs) plus an additional strain, Ox575, which contains a hybrid cassette

comprising the *cwp66* gene of cassette 2 and the *secA2* and *slpA* genes of cassette 6. The Ox strains were all kindly donated by Dr Kate Dingle, Nuffield Department of Medicine, Oxford. UK.

Monoclonal antibody production

Based on the identified biomarker amino acid sequence, linear peptides were synthesised to produce mouse mAbs through the hybridoma technique (Abmart).

Hybridoma culture and production of mAb521

All cell culture reagents were purchased from Sigma. Hybridoma cells specific for mAb521 production were grown at 37°C in a 5 % CO_2 incubator, in basic media (RPMI 1640, 10 % foetal calf serum (FCS) (v/v), 100 mg/l penicillin-streptomycin and 2 mM (w/v) glutamine), initially with the addition of 10 % (v/v) Condimed to aid recovery. Fully confluent cells were then grown in media with reducing FCS content until they were able to grow efficiently in serum free media (10 %, 5 %, 2 %, 0.5 %, serum free). Large scale growth of the cells was performed in specialist media, EX-CELL® 610-HSF Serum-Free Medium for Hybridoma Cells, low-protein (11 mg/l), with L-glutamine, to which 100 mg/l penicillin-streptomycin was added. Cells were left in this media for 8-10 days until the majority were dead, after which the supernatant was recovered by centrifugation at 13,000 *g* for 10 min. mAb521 was purified from the supernatant using a HiTrap Protein G HP prepacked column (GE Healthcare Life Sciences). The column was equilibrated with 10 column volumes buffer A (20 mM sodium phosphate pH 7.2). Sodium phosphate, 100 mM pH 7.2, was added to the culture supernatant at a ratio of 1:5 (final concentration 20 mM sodium phosphate pH 7.2). A minimum of 260 ml cell culture supernatant was loaded on the column before the application of 20 column volumes of buffer A. Finally, mAb521 was eluted with 100 mM glycine/HCl, pH 2.3 in 1 ml fractions collected into 100 µl 1 M Tris-HCl pH 8.0 to immediately neutralise the fraction.

Polyclonal antibody production against purified surface layer protein A

Surface layer protein A (SlpA) was purified to homogeneity (see Supporting Information) to raise polyclonal antibodies in mice (Cambridge Research Biochemicals). IgG antibody fractions were recovered from serum using a Protein A column pre-equilibrated in buffer A. Briefly, serum was diluted 1 in 5 using 100 mM sodium phosphate pH 7.2 before loading onto a pre-equilibrated Protein A column. The column was then washed with 10 column volumes of buffer A before elution using 100 mM glycine/HCl, pH 2.3, collecting 1 ml fractions into 100 µl 1 M Tris-HCl pH 8.0. The resulting IgG polyclonal antibody pool was designated pAbSlpA. pAbSlpA was conjugated to HRP using EZ-link Plus Activated HRP (Thermo Fisher) via primary amine groups on the antibody according to manufacturer's instructions. Briefly, 1 mg pAbSlpA in 50 mM carbonate buffer pH 9.6 was added to lyophilised activated HRP and incubated for 1 h at room temperature (RT). Thereafter, 10 µl 5 M sodium cyanoborohydride solution (Sigma) was added to the enzyme reaction mix and incubated for 15 min at RT followed by the addition of 60 µl 1 M ethanolamine pH 8.5 for 15 min at RT to quench any unreacted material. pAbSlpA-HRP was purified by size exclusion using a HiLoad 16/600 Superdex 200 column, fractions were collected and assessed for antibody binding and HRP activity.

SPR measurements

Prior to SPR, mAb521 and SlpA were dialysed extensively into PBS. All analyses of interactions between mAb521 and SlpA were performed on a BIAcore X100 system equipped with a CM5 chip (BIAcore GE Healthcare) at a flow rate of 10 µl/min. For immobilization of SlpA, the chip was activated with NHS/EDC (1:1) before injection of SlpA in 10 mM acetate, pH 4.0 to attain 255 Resonance Units (RU) of binding. This is designated the FC2 sample channel. In contrast the FC1 reference channel was NHS/EDC treated and then blocked using ethanolamine. Monoclonal antibody 521 (250 nM – 0.65 nM) was injected over both reference and sample surfaces for 300 s before a 600 s wait time followed by a 12 s regeneration using 10 mM glycine pH 2.0. Injection of PBS alone served as a negative control. All measurements were performed in duplicate at 25°C in PBS. The results derived from the reference channel FC1 were subtracted from the FC2 sample channel. SPR data was exported and plotted using OriginPro 8. The responses for each of the different antibody concentrations at 380 s were taken and fitted to a sigmoidal fitting function to derive the apparent K_D.

Results and Discussion

Biomarker identification in *C. difficile* using a custom bioinformatics pipeline

The custom bioinformatics pipeline identified 28 biomarker containing proteins that were positive for all surface associated localisation tools (see Supporting Information). Therefore, a total of 28 biomarker containing proteins were predicted to be exposed on the surface of *C. difficile*. After manual curation and analysis of the 28 biomarker containing proteins, SlpA was judged to be the most promising candidate. The surface layer (S-layer), as the name suggests, typically forms the outermost structure in the bacteria in which they are produced. Functions of the S-layer include cell shape determination, molecular sieving and host cell adhesion/invasion [14]. The proteinaceous *C. difficile* S-layer is encoded by the *slpA* gene, located within the cell wall protein (*cwp*) gene cluster, which also includes *cwp66* (adhesin) and *secA2* (secretory translocase) [15]. Whole genome sequencing and phylogenetic analysis has shown these genes formed a 10-kb cassette, of which 12 distinct surface layer types (SLTs) can be identified [16]. Unlike most bacteria whose S-layers are composed of a single protein subunit, the S-layer of *C. difficile* is constructed using two protein subunits, High Molecular Weight Surface Layer Protein A (HMW SlpA) and Low Molecular Weight Surface Layer Protein A (LMW SlpA) (Figure 1a). The two subunits are derived from a single polypeptide precursor, Surface layer Protein A (SlpA), containing an N-terminal secretion signal [17,18]. The precursor, SlpA, is translocated across the cytoplasmic membrane via the accessory Sec system [19], which directs the signal peptide across the membrane, where the protein is cleaved by CWP84 (Figure 1b) into the HMW and LMW subunits [20]. The HMW and LMW SlpA form a 1:1 heterodimer complex through non-covalent interactions between highly conserved sequences [21] in a Ca^{2+} dependent manner [22]. HMW SlpA contains three tandem cell wall binding 2 motifs (PF04122) which are suggested to be important in mediating interaction with the cell wall, acting as an anchor[18]. These domains interact with anionic polymer PSII, highly conserved in *C. difficile* strains, to direct the protein to the cell wall [23]. It is estimated that there are 590,000 S-layer subunits per cell, requiring synthesis, export and assembly of 164

subunits per second during exponential growth [18,21]. The S-layer appears to be essential, as evidenced by an inability to generate transposon-mediated insertional mutants within the *slpA* gene, consistent with the fact that S-layer proteins have been detected in all *C. difficile* strains to-date [21]. The SlpA biomarker - comprises of 16 amino acids 'AKDGSTKEDQLVDALA' within the HMW subunit. SlpA is the most abundant protein in *C. difficile*, accounting for 10-15 % of the total cellular protein [19]. Additionally the *slpA* gene is strongly transcribed during the entire growth phase [24]. These two characteristics suggested SlpA is an ideal diagnostic candidate.

The bioinformatics prediction that the biomarker is surface associated is consistent with the wider literature. In the absence of a high-resolution structure of the *C. difficile* S-layer, it is challenging to predict the precise topology of the S-layer, or the surface accessibility of the HMW biomarker under native anaerobic conditions. LMW SlpA demonstrates higher inter-strain variability compared with the HMW subunit [25], exhibiting on average 32 % sequence identity between any two strains [21]. This suggests a possible role in immune evasion and host cell adhesion. Accordingly, in bacterial adherence studies, testing *C. difficile* adherence to enterocytes, the LMW subunit was found to be more crucial in host-cell attachment than the HMW subunit, suggesting LMW is more surface-exposed [26]. This is also consistent with a HMW: LMW SAXS structure, suggesting HMW is involved in peptidoglycan anchoring [21]. Conversely, evidence also exists that demonstrates the importance of HMW SlpA in adhesion to enterocytes. FACS analysis has shown that antiserum to recombinant HMW completely eliminated *C. difficile* binding to HEp-2 cells and similarly human colon tissue sections incubated with LMW and HMW subunits revealed the HMW subunit showed higher levels of binding [27].

Regarding the specific 3D location of the biomarker in the HMW subunit, in the absence of a defined crystal structure a suite of bioinformatics tools were used to predict that the biomarker was solvent exposed based on amino acid polarity and secondary structure of the biomarker. A total of 8 mAbs were produced based on the SlpA biomarker sequence and were screened by ELISA for binding to the biomarker peptide sequence and purified SlpA. Two antibodies, mAb521 and mAb652, demonstrated binding to both the peptide sequence and purified SlpA, – mAb521 showed the greatest binding, in all screening assays performed (data not shown) and was taken forward for characterisation.

To be able to perform appropriate laboratory controls to assess the specificity of mAb521 produced against the SlpA biomarker, it was important to identify the species which have amino acid sequences which most closely resemble the SlpA biomarker. BLASTP searches using the SlpA biomarker sequence identified *C. hiranonis* and *P. anaerobius* as having sequences that were the most closely related to SlpA (at the time the BlastP search was performed). However, the avalanche of next generation sequencing data deposited in RefSeq, on a daily basis, makes bioinformatics results highly dynamic and subject to change on a daily basis. A more recent analysis (July 2018) reveals *Clostridium argentinense* and *Clostridium senegalense* as the closest SlpA homologs in the form of uncharacterised cell-wall binding repeat containing proteins. That said, the two species chosen are very closely related and still act as exemplars of the reagent specificity. In addition to investigating the similarity of the SlpA sequence to other proteins, the similarity of full genome sequences to *C. difficile* was also explored. The Pathosystems Resource Integration Centre (PATRIC) was used to identify that *C. sordellii* was the species most closely related to *C. difficile* [28]. Thus, *C. hiranonis*, *P. anaerobius* and *C. sordellii* served as negative controls for mAb521 screening where

appropriate. In addition and possibly more importantly, other bacteria commonly found in stool samples were also screened against the affinity reagents as negative controls.

Western blot analysis against whole cell lysates

Currently, all sequenced *C. difficile* strains fall into 12 SLTs, which therefore represent the known S-layer diversity within the *C. difficile* species [16]. Whole cell lysates of each of the 12 SLTs plus *C. difficile* 630 were probed using mAb521 in a Western blot (see Supporting information). Additionally, for completeness, Ox575 was probed for interaction with mAb521 in the event that the hybrid displayed differential binding to mAb521 than any of the 12 SLTs. When analysed by SDS-PAGE, each *C. difficile* cell lysate displayed the characteristic dual banding pattern associated with the presence of HMW SlpA and LMW SlpA (Figure 2). This is in agreement with previous studies which have shown variability in the mobility of the two SLPs between different strains [29,30]. mAb521 (5 µg/ml) binds to HMW SlpA within all 12 SLTs and not to any other proteins within the negative control strains. Furthermore, the positive binding occurs in one defined band, demonstrating unique specificity of mAb521 to HMW SlpA.

Whole cell direct ELISA binding of mAb521 against all known *C. difficile* SLTs

Whole cells of *C. difficile* 630, the 12 SLTs and Ox575 were tested for interaction with mAb521 (100 ng/ml) using a whole cell ELISA format (See Supporting Information). The OD600 nm of the cells was kept constant throughout, enabling semi-quantitative comparative analysis of mAb521 binding. The antibody was shown to recognise all 12 SLTs plus Ox575 whilst displaying negligible cross-reactivity to the closely related species, *C. hiranonis*, *P. anaerobius* and *C. sordellii* (Figure 3). Additionally, the binding displayed across the *C. difficile* strains was relatively uniform, with a maximal variation of 35 % between the highest (Ox1437a) and the lowest (Ox1396) absorbance value. If Ox1396 is not considered, the binding will be highly comparable between all remaining SLTs with a variability of < 5 %.

Whole cell direct ELISA testing mAb521 against bacteria commonly found in stool samples

In order to validate that mAb521 did not cross react with other bacteria typically found in stool samples, a selection of common faecal bacteria were selected and probed for binding to mAb521 using a whole cell ELISA (see supporting information). Agreeably, mAb521 showed limited cross reactivity to all the faecal bacteria tested under these conditions. In fact, the absorbance values were similar to negative controls in which PBS was added instead of cells (Figure 4). *C. difficile* 630 was used as a positive control and displayed > 90 % binding compared with the highest negative control value of *E. faecalis.*

SPR analysis of mAb521 – SlpA interaction

To calculate the affinity of mAb521 for its target HMW SlpA, SlpA was purified to homogeneity (See Supporting Information) before being tethered to a CM5 chip. The kinetics of the interaction could not be reliably calculated due to mass transport effects which were consistent under all binding conditions tested (data not shown). As such, the equilibrium dissociation constant for mAb521 binding to SlpA was calculated through duplicate titrations of mAb521 (250 nM – 0.65 nM) over SlpA (Figure 5). The equilibrium dissociation constant for mAb521 binding to SlpA was calculated to be 36.5 nM (Figure 5 inset). This demonstrates that mAb521 can recognise the biomarker when it undertakes its native structure as part of

SlpA and further validates the whole cell binding assays, which demonstrated universal binding to native protein in all the SLTs.

Sandwich ELISA using mAb521 and pAbSlpA

Initially we planned to use a sandwich ELISA, using mAb521 for both the capture and detection of whole *C. difficile* cells on the premise that the biomarker is surface exposed, based on the observations from whole cell direct ELISA data (Figures 3 & 4). Provided the biomarker is indeed surface exposed, whole *C. difficile* cells should be multivalent with respect to the biomarker and therefore should be amenable to capture and detection using mAb521. However, when using a sandwich ELISA format, instead of a whole cell direct ELISA we were unable to detect *C. difficile* cells (data not shown). The simplest explanation for this is that the HMW SlpA biomarker is not surface exposed in anaerobically grown lab cultures of *C. difficile* 630 and that the whole cell direct ELISAs used to characterise mAb521 binding are detecting SlpA which is exposed or released from *C. difficile* cells during the overnight non-specific electrostatic adherence to the ELISA plate. However, as the assay conditions are aerobic this could also be a contributing factor in releasing SlpA. Therefore, in order to chemically expose the SlpA biomarker we modified the published glycine extraction method into a rapid 10 min two-step extraction (See Supporting Information). The revised method was applied to various *C. difficile* cell numbers (10^4/ml- 10^9/ml) and purified SlpA before assaying the extract in a sandwich ELISA, using mAb521 for capture and a polyclonal antibody raised against purified SlpA (pAbSlpA-HRP) for detection (see Supporting Information). Using this sandwich ELISA set-up we were able to derive a LOD of 12.4 ng/ml against purified SlpA (Figure 6a) and detect SlpA released from a minimum of 1.7 x 10^6 cells/ml (Figure 6b).

Conclusions

This work describes the computational identification of a unique biomarker for *C. difficile*. The biomarker identified by the bioinformatics pipeline was predicted to be conserved in all *C. difficile* strains. This was validated by screening a representative member of each of the known *C. difficile* SLTs against mAb521 in a direct ELISA format in order to assess the specificity of mAb521 independently. This demonstrated mAb521 could, as predicted, bind each of the 12 SLTs of *C. difficile* and, by extrapolation of this representative model data, all *C. difficile* species. Additionally the biomarker was predicted to be unique to *C. difficile* whilst not being represented in other bacteria. In order to test this,mAb521 was screened using a direct ELISA against a subset of bacteria commonly found in stool samples. As predicted by the bioinformatics pipeline, mAb521 showed minimal cross-reactivity to these faecal bacteria under these conditions. Furthermore, mAb521 displayed negligible cross reactivity to species which have biomarker signatures most closely resembling that of SlpA, *C. hiranonis* and *P. anaerobius*, as well as the most closely related species to *C. difficile*, *C. sordellii.*

This species-specific recognition of *C. difficile* by mAb521 makes it an ideal candidate for a diagnostic antibody. It has been shown on average ~ 10-fold more vegetative cells (~4.75 x 10^6 cells/ml) than spores are typically found in stool samples of patients with CDI [31]. The LOD of 1.7 x 10^6 cells/ml, calculated against *C. difficile* cells using our mAb521-pAbSlpA sandwich ELISA demonstrate that it has the capacity to detect *C. difficile* in clinically relevant levels found in stool samples. In order to achieve these clinically relevant levels of detection against

C. difficile cells, a sample pre-treatment using low pH glycine was required, taking 10 min. The sample pre-treatment releases SlpA from the S-layer, exposing the SlpA biomarker such that it can interact with mAb521 in its active conformation of a 1:1 heterodimer composed of HMW SlpA : LMW SlpA. Although the HMW biomarker is clearly not accessible on the surface of the bacteria under these aerobic diagnostic assay conditions, there remains the possibility the biomarker is surface exposed under the truly native anaerobic environment of the gastrointestinal tract, so the current data does not completely exclude the possibility of the biomarker being a suitable vaccine or drug target.

The biomarker and antibodies developed are clearly appropriate for centralised laboratory analysis using conventional ELISA formats. The main area of future work will to be apply the biomarker and mAb521 to new assay formats, which are amenable for use at the point of care, such as lateral flow assays enhanced with nanozymes [32]. In addition, we believe it is important to validate the biomarker accessibility in a truly anaerobic model system. Future work will assess mab521 binding to whole *C. difficile* cells under anaerobic conditions. If mAb521 can inhibit *C. difficile* adherence to host enterocytes by neutralising SlpA, the antibody may have potential as a neutralising therapeutic option. Additionally, if the HMW biomarker is found to be surface accessible under anaerobic conditions, this gives the biomarker utility as a vaccination target.

This work demonstrates the bioinformatics approach employed has the potential to locate epitopes ubiquitously expressed across an entire group of interest, whilst mitigating the likelihood of false positive detection by intelligently screening potential biomarkers against RefSeq data for "other non-target bacteria". From this work it is clear the biomarker "AKDGSTKEDQLVDALA" is a universal species identifier, which was successfully used to guide the production of a universal antibody, mAb521, against the species, *C. difficile*. In addition, mAb521, in conjunction with a polyclonal partner has both high sensitivity and specify to the biomarker. A clinical study is planned in order to evaluate mAb521 detection of SlpA, as a direct comparator to the GDH EIA, in a two-step testing algorithm for *C. difficile.*

The computational approach described here for biomarker identification could also be deployed against any Gram-positive or Gram-negative bacteria with sequences deposited in RefSeq. The team are pursuing further confirmatory targets in both categories.

Associated Content

The Supporting Information is available free of charge on the

ACS Publications website at DOI:

Computational and experimental methods, bacterial strains used in this work and bioinformatics approach to biomarker identification.

Acknowledgements

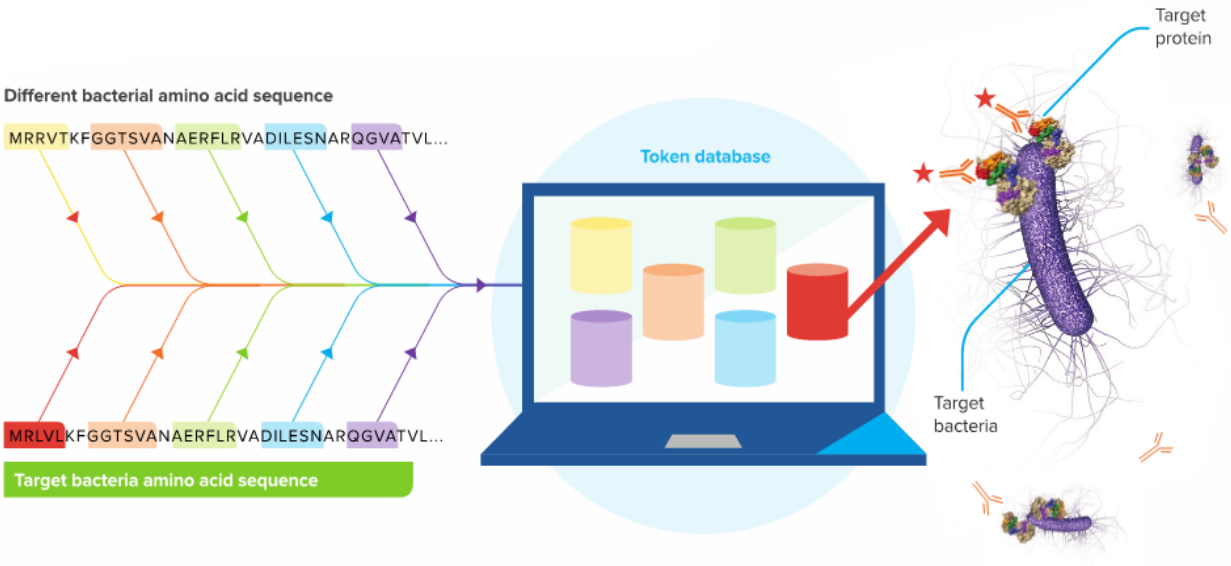
This research was funded by the EPSRC IRC in Early-Warning Sensing Systems for Infectious Diseases (i-sense) EP/K031953/1. We would like to thank Dr Kate Dingle for the kind donation of the Oxford *C. difficile* strains.

Citations

(1) Warny, M.; Pepin, J.; Fang, A.; Killgore, G.; Thompson, A.; Brazier, J.; Frost, E.; McDonald, L. C. *Lancet* **2005**, *366*, 1079-1084.
(2) Sp Moore, L.; Donaldson, H. *Br. J. Hosp. Med. (Lond.)* **2013**, *74 Suppl 10*, C146-149.
(3) Wilkins, T. D.; Lyerly, D. M. *J. Clin. Microbiol.* **2003**, *41*, 531-534.
(4) Casari, E.; De Luca, C.; Calabro, M.; Scuderi, C.; Daleno, C.; Ferrario, A. *Antimicrob. Resist. Infec. Control.* **2018**, *7*, 40.
(5) Polage, C. R.; Gyorke, C. E.; Kennedy, M. A.; Leslie, J. L.; Chin, D. L.; Wang, S.; Nguyen, H. H.; Huang, B.; Tang, Y. W.; Lee, L. W.; Kim, K.; Taylor, S.; Romano, P. S.; Panacek, E. A.; Goodell, P. B.; Solnick, J. V.; Cohen, S. H. *JAMA Intern. Med.* **2015**, *175*, 1792-1801.
(6) DOH. *Updated guidance on the diagnosis and reporting of clostridium difficile,* **2012**.
(7) O'Connor, J. R.; Johnson, S.; Gerding, D. N. *Gastroenterology* **2009**, *136*, 1913-1924.
(8) Brazier, J. S.; Raybould, R.; Patel, B.; Duckworth, G.; Pearson, A.; Charlett, A.; Duerden, B. I. *Euro Surveill.* **2008**, *13*.
(9) Wilcox, M. H.; Shetty, N.; Fawley, W. N.; Shemko, M.; Coen, P.; Birtles, A.; Cairns, M.; Curran, M. D.; Dodgson, K. J.; Green, S. M.; Hardy, K. J.; Hawkey, P. M.; Magee, J. G.; Sails, A. D.; Wren, M. W. *Clin. Infect. Dis.* **2012**, *55*, 1056-1063.
(10) Tenover, F. C.; Baron, E. J.; Peterson, L. R.; Persing, D. H. *J. Mol. Diagn.* **2011**, *13*, 573-582.
(11) Fang, F. C.; Polage, C. R.; Wilcox, M. H. *J. Clin. Microbiol.* **2017**, *55*, 670-680.
(12) Tenover, F. C.; Novak-Weekley, S.; Woods, C. W.; Peterson, L. R.; Davis, T.; Schreckenberger, P.; Fang, F. C.; Dascal, A.; Gerding, D. N.; Nomura, J. H.; Goering, R. V.; Akerlund, T.; Weissfeld, A. S.; Baron, E. J.; Wong, E.; Marlowe, E. M.; Whitmore, J.; Persing, D. H. *J. Clin. Microbiol.* **2010**, *48*, 3719-3724.
(13) Journal of integrative bioinformaticsFlanagan, K.; Cockell, S.; Harwood, C.; Hallinan, J.; Nakjang, S.; Lawry, B.; Wipat, A. *J. Integr. Bioinform.* **2014**, *11*, 242.
(14) Sara, M.; Sleytr, U. B. *J. Bacteriol.* **2000**, *182*, 859-868.
(15) 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., et al. *Nat. Genet.* **2006**, *38*, 779-786.
(16) Dingle, K. E.; Didelot, X.; Ansari, M. A.; Eyre, D. W.; Vaughan, A.; Griffiths, D.; Ip, C. L.; Batty, E. M.; Golubchik, T.; Bowden, R.; Jolley, K. A.; Hood, D. W.; Fawley, W. N.; Walker, A. S.; Peto, T. E.; Wilcox, M. H.; Crook, D. W. *J. Infect. Dis.* **2013**, *207*, 675-686.
(17) Calabi, E.; Ward, S.; Wren, B.; Paxton, T.; Panico, M.; Morris, H.; Dell, A.; Dougan, G.; Fairweather, N. *Mol. Microbiol.* **2001**, *40*, 1187-1199.
(18) Kirk, J. A.; Banerji, O.; Fagan, R. P. *Microb. Biotechnol.* **2017**, *10*, 76-90.
(19) Fagan, R. P.; Fairweather, N. F. *J. Biol. Chem.* **2011**, *286*, 27483-27493.
(20) Kirby, J. M.; Ahern, H.; Roberts, A. K.; Kumar, V.; Freeman, Z.; Acharya, K. R.; Shone, C. C. *J. Biol. Chem.* **2009**, *284*, 34666-34673.
(21) Fagan, R. P.; Albesa-Jove, D.; Qazi, O.; Svergun, D. I.; Brown, K. A.; Fairweather, N. F. *Mol. Microbiol.* **2009**, *71*, 1308-1322.

(22) Takumi, K.; Endo, Y.; Koga, T.; Oka, T.; Natori, Y. *Tokushima J. Exp. Med.* **1992**, *39*, 95-100.
(23) Ganeshapillai, J.; Vinogradov, E.; Rousseau, J.; Weese, J. S.; Monteiro, M. A. *Carbohydr. Res.* **2008**, *343*, 703-710.
(24) Savariau-Lacomme, M. P.; Lebarbier, C.; Karjalainen, T.; Collignon, A.; Janoir, C. *J. Bacteriol.* **2003**, *185*, 4461-4470.
(25) Calabi, E.; Fairweather, N. *J. Bacteriol.* **2002**, *184*, 3886-3897.
(26) Merrigan, M. M.; Venugopal, A.; Roxas, J. L.; Anwar, F.; Mallozzi, M. J.; Roxas, B. A.; Gerding, D. N.; Viswanathan, V. K.; Vedantam, G. *PLoS One* **2013**, *8*, e78404.
(27) Calabi, E.; Calabi, F.; Phillips, A. D.; Fairweather, N. F. *Infect. Immun.* **2002**, *70*, 5770-5778.
(28) Snyder, E. E.; Kampanya, N.; Lu, J.; Nordberg, E. K.; Karur, H. R.; Shukla, M.; Soneja, J.; Tian, Y.; Xue, T.; Yoo, H.; Zhang, F.; Dharmanolla, C.; Dongre, N. V.; Gillespie, J. J.; Hamelius, J.; Hance, M.; Huntington, K. I.; Jukneliene, D.; Koziski, J.; Mackasmiel, L., et al. *Nucleic Acids Res.* **2007**, *35*, D401-406.
(29) Br J Hosp Med (Lond)Cerquetti, M.; Molinari, A.; Sebastianelli, A.; Diociaiuti, M.; Petruzzelli, R.; Capo, C.; Mastrantonio, P. *Microb. Pathog.* **2000**, *28*, 363-372.
(30) Takeoka, A.; Takumi, K.; Koga, T.; Kawata, T. *J. Gen. Microbiol.* **1991**, *137*, 261-267.
(31) Jump, R. L.; Pultz, M. J.; Donskey, C. J. *Antimicrob. Agents Chemother.* **2007**, *51*, 2883-2887.
(32) Loynachan, C. N.; Thomas, M. R.; Gray, E. R.; Richards, D. A.; Kim, J.; Miller, B. S.; Brookes, J. C.; Agarwal, S.; Chudasama, V.; McKendry, R. A.; Stevens, M. M. *ACS Nano* **2018**, *12*, 279-288.

TOC graphic



Figures

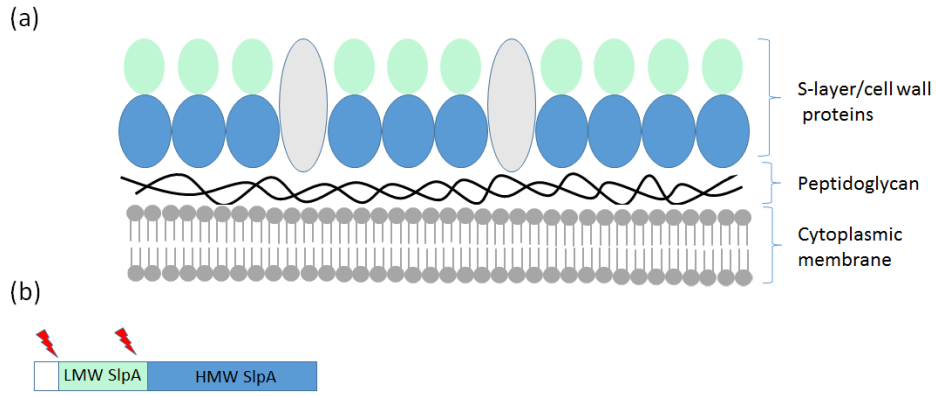


Figure 1: Schematic representation of the *C. difficile* cell wall and SlpA processing. (A) *C. difficile* cell wall - the two SLPs are shown, HMW SlpA (blue) and LMW SlpA (green). Other minor cell wall proteins are shown as ovals (grey). (b) The precursor protein SlpA showing the cleavage sites (lightning bolt) generating the signal peptide (white rectangle), the mature LMW SlpA (green rectangle) and HMW SlpA (blue rectangle).

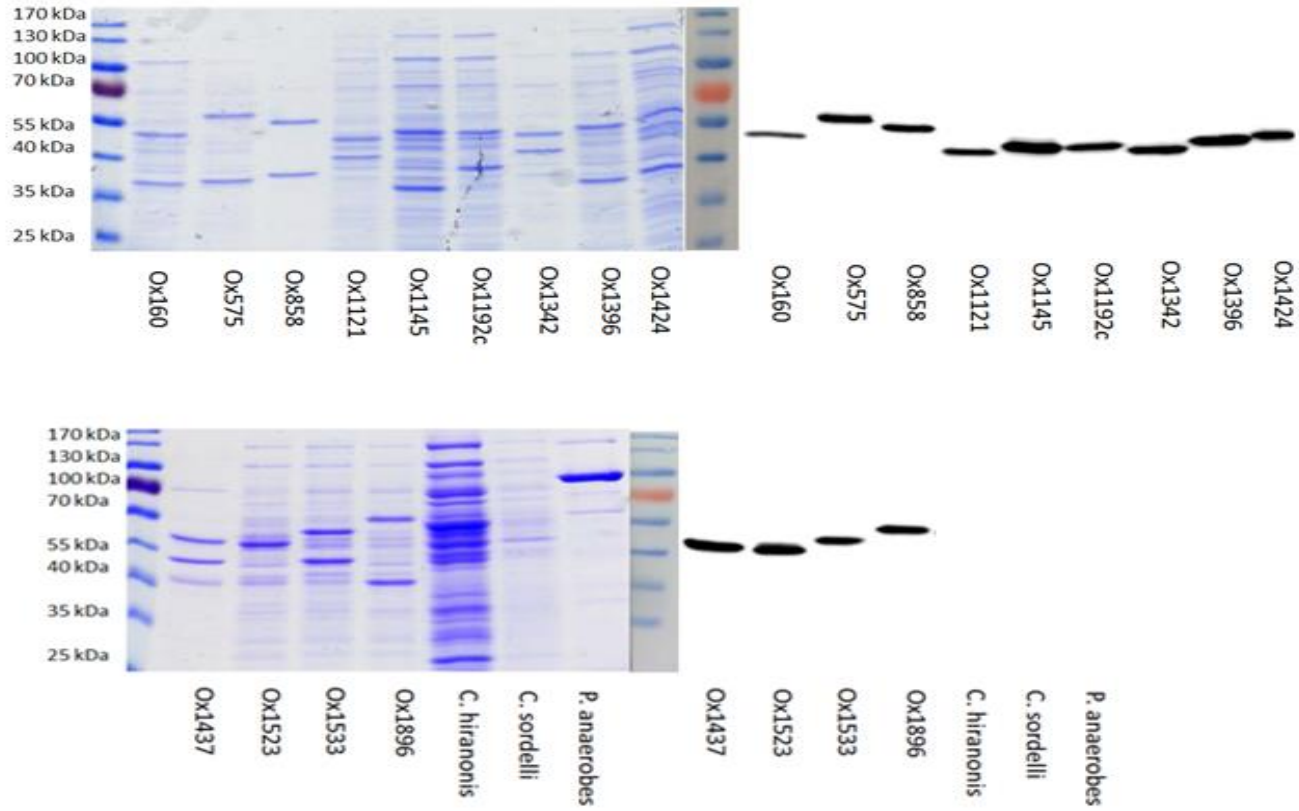


Figure 2: Western blot analysis against whole cell lysates: mAb521 (5 µg/ml) was tested for binding against whole cell lysates from the 12 SLTs, Ox575 and the closely related species C. *hiranonis, C. sordellii,* and *P. anaerobius.*

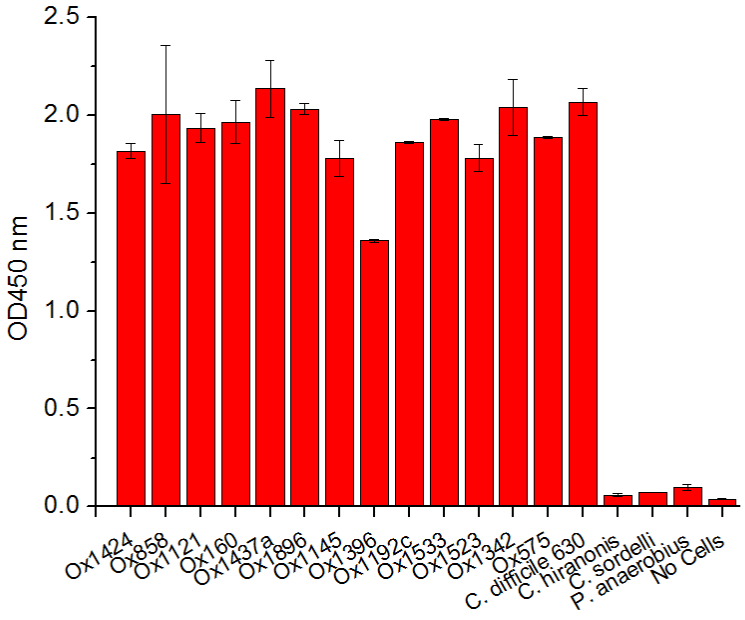


Figure 3: Whole cell direct ELISA using mAb521 against all SLTs: mAb521 (100 ng/ml) was probed for binding against the 12 SLTs, Ox575 and the closely related species *C. hiranonis*, *C. sordellii*, and *P. anaerobius*. The error bars show the standard deviation seen between the two replicates of each sample.

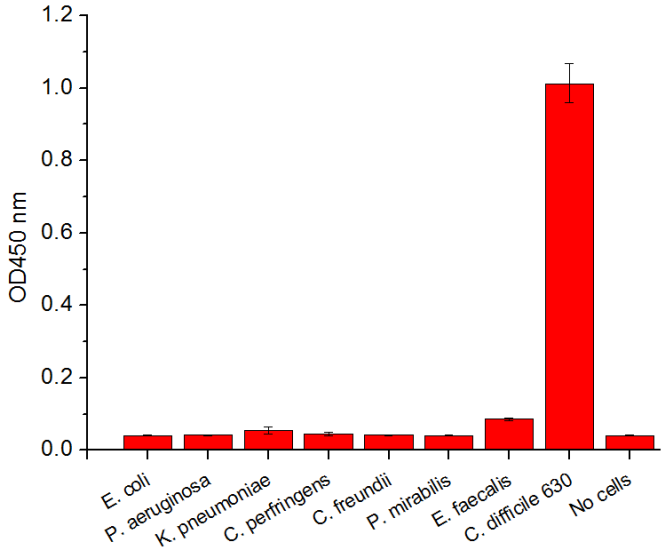


Figure 4: Whole cell direct ELISA using mAb521 against faecal bacteria. mAb521 (100 ng/ml) was probed for binding against the common faecal bacteria *Escherichia coli*, *P. aeruginosa*, *K. pneumoniae, C. perfringens, C. freundii, P. mirabilis* and *E. faecalis*. *C. difficile* 630 was used as a positive control and PBS as a negative control. The error bars show the standard deviation seen between the two replicates of each sample.

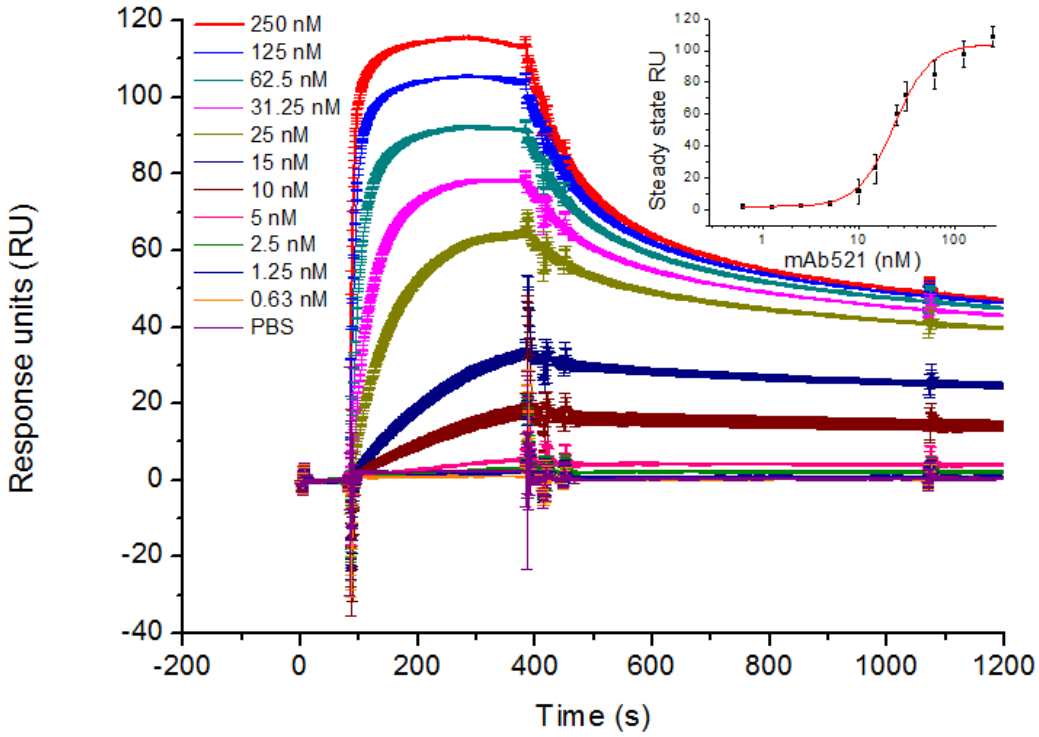


Figure 5: Using SPR to characterise the binding of mAb521 against purified SlpA. The apparent equilibrium dissociation constant (K_D) for mAb521 binding to SlpA was calculated at 380 s. $R^2 = 0.99$ (inset) through duplicate titrations of mAb521 (250 nM – 0.65 nM) over SlpA (main figure). Duplicate injections of PBS over SlpA served as negative controls. The error bars show the standard deviation seen between the two replicates of each sample.

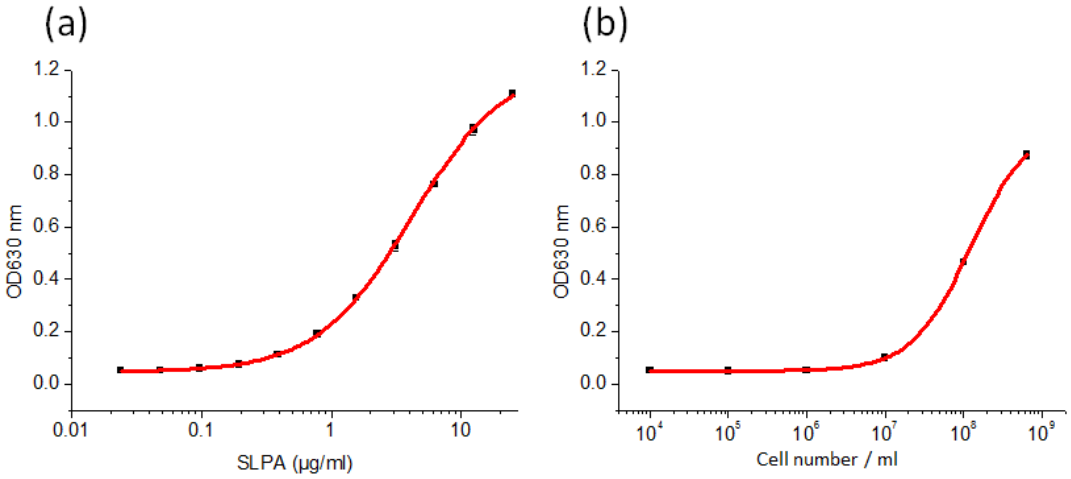


Figure 6: Determining the LOD of the sandwich ELISA. (a) Sandwich ELISA against purified SlpA using mAb521/pAbSlpA. $R^2 = 0.99$. (b) Sandwich ELISA against rapid glycine extracted cells (10^9-10^4/ml), extracted for 10 min. $R^2 = 0.99$. The error bars show the standard deviation seen between the two replicates of each sample.