

RESEARCH

Open Access



Escherichia coli recombinant expression of SARS-CoV-2 protein fragments

Bailey E. McGuire¹, Julia E. Mela¹, Vanessa C. Thompson¹, Logan R. Cucksey¹, Claire E. Stevens¹, Ralph L. McWhinnie¹, Dirk F. H. Winkler², Steven Pelech^{2,3} and Francis E. Nano^{1*} 

Abstract

We have developed a method for the inexpensive, high-level expression of antigenic protein fragments of SARS-CoV-2 proteins in *Escherichia coli*. Our approach uses the thermophilic family 9 carbohydrate-binding module (CBM9) as an N-terminal carrier protein and affinity tag. The CBM9 module was joined to SARS-CoV-2 protein fragments via a flexible proline–threonine linker, which proved to be resistant to *E. coli* proteases. Two CBM9–spike protein fragment fusion proteins and one CBM9–nucleocapsid fragment fusion protein largely resisted protease degradation, while most of the CBM9 fusion proteins were degraded at some site in the SARS-CoV-2 protein fragment. All of the fusion proteins were highly expressed in *E. coli* and the CBM9-ID-H1 fusion protein was shown to yield 122 mg/L of purified product. Three purified CBM9–SARS-CoV-2 fusion proteins were tested and found to bind antibodies directed to the appropriate SARS-CoV-2 antigenic regions. The largest intact CBM9 fusion protein, CBM9-ID-H1, incorporates spike protein amino acids 540–588, which is a conserved region overlapping and C-terminal to the receptor binding domain that is widely recognized by human convalescent sera and contains a putative protective epitope.

Keywords: SARS-CoV-2, COVID-19, Recombinant spike protein, Carbohydrate-binding module, CBM9, *Escherichia coli*

Introduction

One of the public health surveillance tools needed to respond to the coronavirus disease 2019 (COVID-19) pandemic is the ability to detect seroconversion to antigens of the SARS-CoV-2 virus, the causative agent of COVID-19. The ability to detect antibodies that are specific to SARS-CoV-2 allows an assessment of the level of probable immunity to COVID-19 in a population. At the individual level, the ability to detect anti-SARS-CoV-2 antibodies can help one assess their personal level of vulnerability to COVID-19 due to immunity generated by either vaccination or from an indeterminate or asymptomatic infection.

The human antibody response to SARS-CoV-2 infection can include the development of antibodies reactive with any of the 29 proteins encoded in the viral genome [1], including 16 non-structural proteins (NSP's) encoded by the ORF1a/b gene. However most studies have focused on studying the antibody response to the abundant spike and nucleocapsid proteins [2–11]. Among the SARS-CoV-2 proteins, the spike protein varies the most between coronaviruses [12]; using it allows for the greatest specificity in an antibody assay as well as the ability to differentiate reaction with SARS-CoV-2 from other common coronaviruses. As well, some antibodies reactive with the spike protein, especially those that react within or near its receptor binding domain (RBD), are neutralizing [9, 11, 13–15]. Thus, detection of anti-spike protein antibodies may indicate a level of immunity to SARS-CoV-2.

Some studies of anti-SARS-CoV-2 antibody response examine antibody reactivity with linear epitopes using

*Correspondence: fnano@uvic.ca

¹ Department of Biochemistry and Microbiology, University of Victoria,

STN CSC, PO Box 1700, Victoria, BC V8W 2Y2, Canada

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

synthetic peptides that correspond to the primary structure of viral proteins [2–10]. While this type of assay misses many antibody responses against conformational and topographically assembled epitopes, it is the most practical, both technically and economically. Although synthetic peptides are far less expensive than full-length recombinant SARS-CoV-2 spike protein, their production cost can still present barriers in resource-poor health systems or when large quantities are needed.

The SARS-CoV-2 spike protein is the most common antigen used to test for seroconversion, and recombinant spike protein, the RBD fragment or synthetic peptides corresponding to spike sequences are commonly used to test sera. Typically, human derived HEK293 cells are used to express spike protein, and Stadlbauer et al. [16] reported expression levels of about 20 mg/L for the RBD fragment and about 5 mg/L of the trimer form of the full-length spike protein. However, other recombinant expression systems have been used to produce the SARS-CoV-2 spike protein. For example, Yang et al. [13] expressed the spike protein RBD in the Sf9 insect cell line (*Spodoptera frugiperda*) using a BAC-to-BAC expression system, and Fujita et al. [17] expressed full-length spike protein in silk worm larvae at a level of about 10 mg/L of larval serum. Recently Rihn et al. [18] described the construction of glutathione S-transferase (GST) and maltose binding protein (MBP) fusions to all of the ORFs of SARS-CoV-2, as part of an expansive effort to develop molecular tools to study SARS-CoV-2. These fusion proteins are expressed in *E. coli*, and while their properties and yields are not reported, these recombinants represent significant tools for obtaining SARS-CoV-2 antigenic material. Other groups have expressed the RBD fragment in *E. coli*, though it is largely insoluble and therefore requires solubilization and refolding, perhaps due to the incompatibility between the four disulphide bonds in the native RBD structure and the reducing environment of the *E. coli* cytoplasm [19–21]. As well, the RBD has four confirmed glycans which are absent when the fragment is produced in *E. coli*, leading to differences in antibody

binding between human-expressed and *E. coli*-expressed RBD fragments but maintenance of ACE2 receptor binding [19–22]. Maeda and Tian et al. [23] expressed peptides from the spike proteins of the betacoronavirus SARS-CoV-2 and alphacoronavirus porcine epidemic diarrhea virus (PEDV) on the surfaces of genome-reduced *E. coli*, and used those strains as killed whole-cell vaccines for protection against PEDV. The peptide fragments they used, corresponding to SARS-CoV-2 spike protein residues 812–833 and PEDV spike protein residues 884–909, are highly conserved between coronaviruses and vaccination and boosting with both killed whole-cell vaccines provided protection against PEDV. In a similar fashion, our work explores the utility of expressing SARS-CoV-2 peptide epitopes in *E. coli*, using the under-appreciated protein carrier family 9 carbohydrate-binding module (CBM9) from the *Thermotoga maritima* enzyme xylanase 10A [24, 25] that promotes soluble high level protein expression and uses inexpensive reagents for protein purification [26, 27].

Materials and methods

Recombinant techniques

Plasmid pRSET5A was used as the backbone for all expression plasmid constructs [28]. All of the synthetic DNA regions designed to encode CBM9-SARS-CoV-2 spike protein fusions were made by Twist Biosciences. To initially test the expression of CBM9 peptide fusions we cloned synthetic DNA encoding CBM9, CBM9-ID-C, CBM9-ID-F and CBM9-H3 (Fig. 1A). Plasmid pRSET5A was amplified by inverse PCR using primers F-R5A and R-R5A, which have Esp3I sites added to the ends that upon digestion yield 5'-overhangs compatible with the overhangs generated for the PCR amplicons of the synthetic DNA fragments. The CBM9-C, F and H3 DNA fragments were codon optimized [29] for *E. coli* and designed to lack an internal Esp3I site. These fragments were amplified with primers F-CBD (forward primer for all fragments) and R-CBD-IDc, R-CBD-IDf and R-CBD-h3 as the reverse primers. After amplification

(See figure on next page.)

Fig. 1 CBM9-SARS-CoV-2 epitope cloning strategies and recombinant fusion protein expression. **A** Initial clones were made by amplifying pRSET5A by inverse PCR, and ligating the plasmid amplicon to synthetic DNA encoding CBM9 with a linker fused to spike protein epitope ID-C, ID-F, or ID-H3. **B** To create fusion clones of ID-A, B, D, E, G, H1 and I, synthetic DNA encoding just the epitope regions replaced the ID-C encoding region. **C** To create the clones CBM9-(PT)₄P, and N, primers with long overhang regions were used in an inverse PCR reaction using pRSET5A::CBM9-*id-a* to exchange the epitopes fused to CBM9. **D** Representation of linear ID-A through ID-I regions with the amino acid numbers of the SARS-CoV-2 spike protein recognized by antibody from COVID-19 convalescent sera, as described by Zhang et al. [10]; RBD is the receptor binding domain. **E** Diversity of amino acid residues in the SARS-CoV-2 spike protein. Spike protein entropy/diversity data from the Nextstrain SARS-CoV-2 latest global analysis open data was downloaded on January 6th 2022 and plotted using GraphPad Prism 9 [37]. Diversity/entropy values are reported from 0 to 1, where higher numbers represent greater diversity/uncertainty in the amino acid at that position. **F** Cell extracts, equivalent to 33 μ L of detergent-soluble material, of recombinants encoding CBM9 fusion proteins with and without IPTG induction. For comparison 1.3 μ g of purified CBM9-N is shown in the last lane on the right. **G** Samples of CBM9 fusion proteins purified by batch absorption to cellulose powder, after storage at 4 $^{\circ}$ C for a minimum of two weeks. In Panels A-C the bent arrow indicates T7 promoter region; half circle indicates T7 ribosome binding site; "T" symbol represents transcriptional terminator

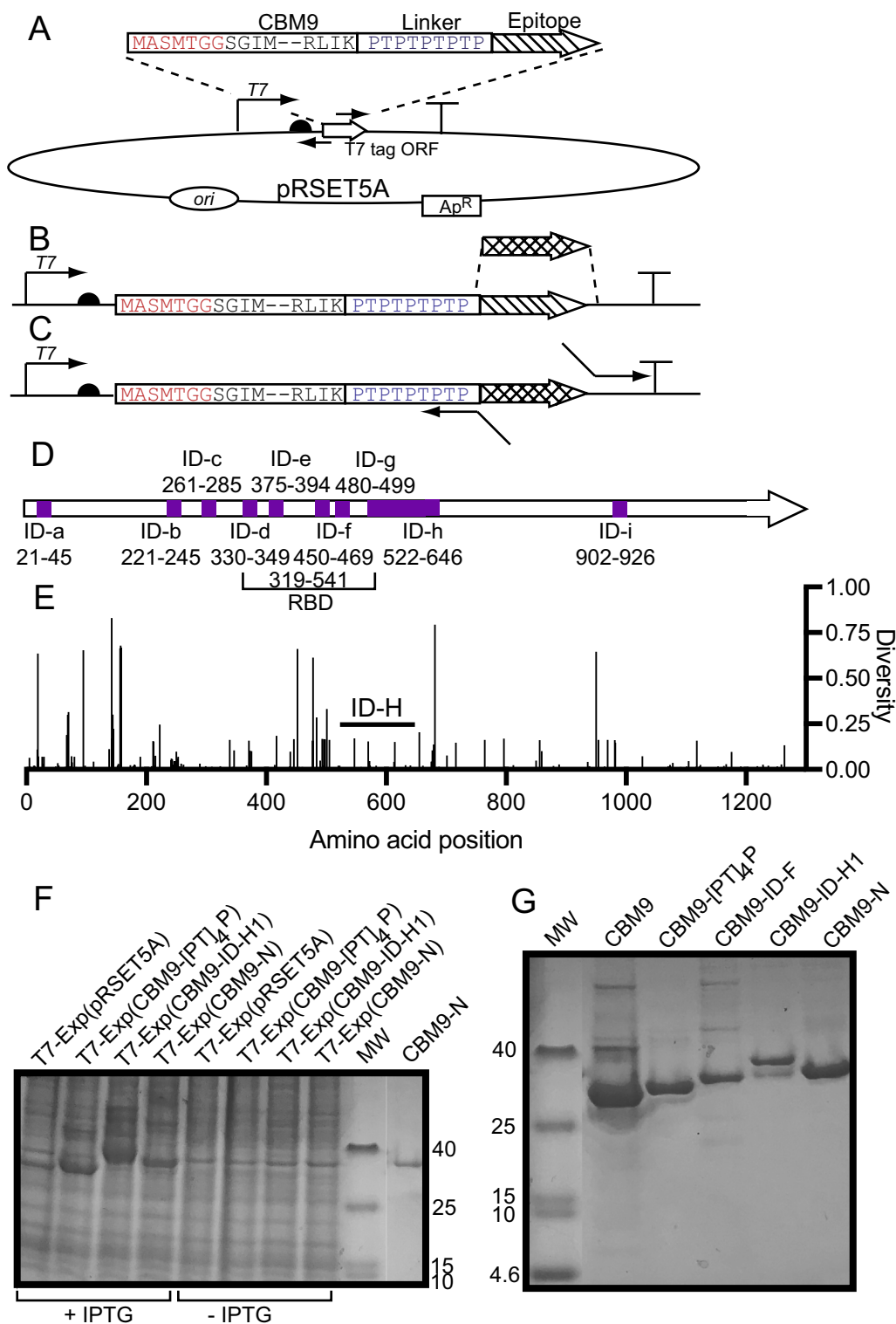


Fig. 1 (See legend on previous page.)

the products were joined to pRSET5A using a simultaneous cutting and ligation reaction [30] using Esp3I as the restriction enzyme. Briefly, 30 cycles of 5 min at 37 °C and 5 min at 16 °C were followed by 10 min at 65 °C. Ligated DNA was transformed into T7 Express *lysY/T^q* *E. coli* (NEB) and selected on LB agar (per liter, 5 g yeast extract; 10 g tryptone, 5 g NaCl; 15 g agar) supplemented with chloramphenicol (10 µg/mL) and carbenicillin (250 µg/mL). Once initial clones were sequence verified and shown to produce the appropriate protein product, further recombinants were constructed using the pRSET5A::CBM9-*id-c* clone as the backbone. This plasmid was amplified by inverse PCR using primers Fb-R5A and R-R5AidC so as to remove the SARS-CoV-2 spike protein-encoding fragment of DNA and replace it with another fragment of synthetic DNA (ID-a, b, d, g, h, h1, h2, i; Twist Biosciences) (Fig. 1B) using a cutting-ligation reaction as described above. To make a plasmid encoding just the CBM9-(TP)₄P (no SARS-CoV-2 fragment) or CBM9-N (containing a nucleocapsid epitope), plasmid pRSET5A::CBM9-*id-a* was amplified with primers nF2-R5A-CBD and nR-R5A-Flex; or F-nucl-ep and R-nucl-ep and Esp3I digested and ligated with a strategy depicted in Fig. 1C. The primers used to make all of the constructs are listed in Additional file 1: Table S1. A color-coded example of a CBM9 fusion clone is shown in Additional file 1: Fig. S1. All CBM9-SARS-CoV-2 recombinants expressing protein fusion constructs were sequence verified. The GenBank accession numbers and the availability of recombinant clones is described in Additional file 1: Materials and Methods.

Fusion protein isolation and analysis

Frozen aliquots of seed cultures were prepared and 0.4 mL of the stocks were added to 20 mL of modified auto-induction ZYM-5052 media that lacks an added carbon source (1% N-Z amine, 1% yeast extract, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 2 mM MgSO₄, 100 µM FeCl₃) [31]; the media was supplemented with carbenicillin (250 µg/mL) and chloramphenicol (10 µg/mL). These cultures were grown at 24 °C for 16 h. The A₆₀₀ was determined and the culture was added to 100 mL fresh broth with antibiotics in a 1 L flask to give an A₆₀₀ of 2. IPTG was added to a final concentration of 4 mM, and the culture was incubated with shaking (250 RPM) for 3 h at 30 °C. The cultures were cooled on ice for 15 min, and then subjected to centrifugation for 10 min at 16,000×g. The cell pellet from each 25 mL of culture was suspended in 5 mL of solution B (500 mM NaCl, 10 mM MgCl₂, 0.5% CHAPS, 50 mM potassium phosphate, pH 7.0, 100 µg/mL lysozyme). 1.5 g of glass beads (≤ 106 µm, Sigma), and 5 µL Benzonase[®] nuclease (Sigma) were added and the cell suspension was

vortexed vigorously for four 1 min intervals. The beads and insoluble debris were pelleted by centrifugation at 16,000×g for 10 min and the supernatant was transferred to a new tube. A second 5 mL of solution B was added to the beads, the vortexing and pelleting steps were repeated and the supernatant was combined with the supernatant from the previous lysis step. The pooled supernatant was added to 2.5 g of cellulose powder (Sigma, cat # 435236) that had previously been equilibrated with solution A (500 mM NaCl, 50 mM potassium phosphate, pH 7.0) in a 50 mL conical centrifuge tube. The tube was rocked for 16 h at room temperature, and the cellulose powder was pelleted by centrifugation for 2 min at 4000×g. The cellulose resin was washed three times by adding 12 mL of solution A, rocking for 2 min, and separation of the cellulose from the washing solution. This was followed by three washes using solution C (150 mM NaCl, 50 mM potassium phosphate, pH 7.0). Following the last wash with solution C, the cellulose was suspended in 5 mL elution buffer (1 M glucose, 15 mM NaCl, 10 mM Tris, pH 7.6), and rocked for 30 min. Centrifugation was applied to pellet the cellulose, the supernatant was removed, the elution step was repeated twice more and eluates were pooled. Pooled eluates were added to a Vivaspin[®] 6, 10 kDa MW cut off (Sartorius) protein concentrator to change the buffer to 150 mM NaCl, 10 mM Tris, pH 7.6.

Protein concentrations of aqueous samples were determined using the Qubit protein assay (ThermoFisher). Following SDS-PAGE separation a LI-COR Odyssey[®] CLx Imager was used to measure the fluorescence following excitation at 700 nm of Coomassie blue stained protein bands. The laser scan data was processed with Image Studio v5.2 software which was used to determine the relative amount of specific recombinant protein bands to the total protein in a sample.

Mass spectroscopy analysis of recombinant proteins

Liquid chromatography–mass spectrometry for the determination of intact molecular weight of isolated CBM9 fusion proteins was conducted by the University of Victoria/Genome BC Proteomics Centre. Details of the methods are presented in Additional file 1: Materials and Methods.

Commercial recombinant proteins and antibodies

The following recombinant SARS-CoV-2 proteins [18] were procured from the MRC Protein Phosphorylation and Ubiquitination Unit Reagents and Services at the University of Dundee (Dundee, Scotland; <https://mrcppu-covid.bio/>), and are listed in Additional file 1. Rabbit polyclonal antibodies (Kinexus Bioinformatics) directed against synthetic peptides based on

SARS-CoV-2 proteins were used in dot blots, and are listed in Additional file 1.

Human serum samples

The process of the use of human blood products for the investigation of COVID-19 received an ethical approval by Veritas Independent Review Board Inc. (Saint-Laurent, QC, Canada), IRB Protocol 16,567–09:39:354–06-2020. The majority of blood serum samples were generated from blood donations of volunteers. A small number of serum or plasma samples from positive tested donors and all pre-COVID-19 samples were acquired by purchase (Precision for Medicine, Norton, MA, USA; Innovative Research, Novi, MI, USA; AllCells, Alameda, CA, USA) or by donation (CureImmune, Vancouver, BC, Canada).

The blood samples were collected from persons that had been confirmed as positive using a PCR genetic test (designated as “COVID”), those that showed symptoms similar to COVID-19 but were not tested (designated as “sick”), those that were healthy and asymptomatic (designated as “Control”) and those that were healthy and donated prior to April 2019 (“pre-COVID”).

All of the preparations of recombinant proteins were robotically spotted as 0.24 μL of a $\sim 8 \mu\text{M}$ final concentration (except GST-NSP2 SARS CoV2 [DU 66414] in Spot B1 was printed at a 6.5 μM concentration) on to nitrocellulose membranes. The blots were washed three times with TBS (aqueous solution of 20 mM Tris-base and 250 mM NaCl; pH 7.5). These dot blot arrays were blocked with 2.5% BSA in T-TBS (TBS with 0.05% Tween 20) for 30 min. After washing the membranes twice with T-TBS, the arrays were incubated with either affinity-purified rabbit polyclonal antibodies against SARS-CoV-2 protein sequences at 2 $\mu\text{g}/\text{mL}$ or serum from recovered COVID-19 patients and healthy controls at 1:200 dilution in T-TBS. The incubation was carried out at 4 $^{\circ}\text{C}$ overnight. To detect the bound antibodies, the arrays were washed with T-TBS three times followed by the incubation with goat anti-human IgG + IgA + IgM pAb or donkey-anti-rabbit IgG (HRP conjugates, 1:20,000 dilution; Jackson ImmunoResearch, West Grove, Pennsylvania, USA). After 30 min incubation, the arrays were washed six times with T-TBS, once with 125 mM NaCl and rinsed twice with water. The bound secondary antibody was visualized by enhanced chemiluminescence (ECL) on a Bio-Rad FluorS-Max scanner. The ECL scan was performed at a scanning time of 300 s. Eight images of the scanned array were generated during that scanning time.

Results and discussion

Expression of SARS-CoV-2 protein fragments as fusions to the CBM9 module

The goal of our research was to test if SARS-CoV-2 protein fragments known to elicit a human antibody response could be produced inexpensively using a universally available microbial expression system. The SARS-CoV-2 spike protein or its RBD have been recombinantly produced in a number of hosts, including engineered HEK-293 cells, insect cells, insect larvae and *Escherichia coli* [17, 19–21, 32–36]. However, these eukaryotic expression systems require growth media and bioreactors that are orders of magnitude more expensive than the material and equipment used for microbial expression systems, and *E. coli*-produced RBD fragments are insoluble, and require refolding. Therefore, we felt that there was value in developing an inexpensive microbial expression system that can produce soluble SARS-CoV-2 protein fragments in substantial amounts.

Our strategy was to use a standard *E. coli* T7 RNA polymerase approach to drive high levels of mRNA transcription, and use a proven carrier protein module, a thermophilic family 9 carbohydrate-binding module (CBM9), to carry the SARS-CoV-2 fragment at the C-terminus of the fusion protein (Fig. 1A–C). Kavooosi et al. [26] showed that the CBM9 module expresses at high levels, even when a protein was fused to the C-terminus. In a separate work [27], these same researchers showed that linking CBM9 to a protein with a proline-threonine rich linker ($[(\text{PT})_4\text{P-IEGR}]$) resulted in a fusion protein that was resistant to protease attack by endogenous *E. coli* proteases [27]. Thus, we adopted the use of $(\text{PT})_4\text{P}$ as a linker between the CBM9 and SARS-CoV-2 spike protein fragments, and an example of the gene organization is shown in Additional file 1: Fig. S1.

The bulk of the studies on the antibody responses to SARS-CoV-2 have been conducted using overlapping synthetic peptides corresponding to SARS-CoV-2 proteins, primarily the spike protein, but sometimes several proteins or the entire proteome. At the time that we initiated our studies, the work of Zhang et al. [10] was one of the more comprehensive analyses of the human antibody response in COVID-19. These researchers identified nine immunodominant amino acid regions in the spike protein, designated ID-A through ID-I (Fig. 1D), which were recognized by antibodies from COVID-19 convalescent sera, and we chose these regions as well as one nucleocapsid protein epitope to clone and express. An SDS-PAGE analysis of the expressed CBM9-spike protein fusions (Additional file 1: Fig. S2) indicated which fusion proteins were most resistant to degradation by *E. coli* and we chose a sub-group of the clones to study further (abandoning ID-H, ID-H2, ID-H3 and ID-I). Table 1

Table 1 Amino acid sequence of B-cell epitopes fused to CBM9, and putative protease cleavage sites

Epitope (amino acid numbers)	Amino acid sequence and calculated MW of processed protein*, and observed MW of CBM9 fusion protein
ID-A (spike 21–45)	NLTTR <u>T</u> QLPAYTNSFTRGVVYPDKVF/RSSVLHSTQ Calculated: 26,076.08 Observed: 26,075.1
ID-B (spike 221–245)	RDLPQGFSALEPLVDLPIGINITR/FQTLALHRSYLTPGD Calculated: 25,560.93 Observed: 25,559.9
ID-C (spike 261–285)	SSGWTAGAAAY/YVGYLQPRTEFLLYKYNENGTITDAVD Calculated: 23,967.94 Observed: 23,966.9
ID-D (spike 330–349)	ESIVRFPNITN/LCPFGEVFNATRFASVYAWNRRK Calculated: 24,216.16 Observed: 24,215.2
ID-E (spike 375–394)	NSASFSTFK/CYGVSPKLNLDLCTNVYAD Calculated: 23,914.95 Observed: 23,913.9
ID-F (spike 450–469)	DSKVGGNYNLYRLFRKSNLKPFFERDISTEIQ Calculated: 26,940.52 Observed: 26,939.5
ID-G (spike 480–499)	QAGSTPCNGVEGFNCYFPLO/SYGFQPTNGVGYQ Calculated: 23,967.94 Observed: 23,966.9
ID-H1 (spike 540–588)	NFNENGLTGTGLTESNKKFLPFOQFGRDIADTTDAVRDPQTLLEILDIT Calculated: 28,384.23 Observed: 28,383.2
N (nucleocapsid 386–403)	QKKQQTVTLPAADLDDF Calculated: 24,957.55 Observed: 24,956.57
CBM9-(PT) ₄ P	Not applicable: Calculated: 22,945.49 Observed: 22,944.51

*The calculated molecular weight was based on sequences assuming that the amino terminal methionine was removed. The C-terminal amino acid was chosen to best match the molecular weight of the dominant form found in the MS experiments; in most cases the C-terminal amino acid lies before the slash (“/”). Underlined sequence corresponds to synthetic peptide found to bind antibody as determined by Zhang et al. [10]. The region surrounding the slash indicates the putative protease cleavage site, by endogenous *E. coli* proteases

lists the amino acid sequences of the encoded spike (and one nucleocapsid [“N”]) protein regions in the clones we constructed and chose to study further, and the specific epitope regions of the spike protein identified by Zhang and co-workers [10] are underlined. As the COVID-19 pandemic has progressed many amino acid changes have been observed in the SARS-CoV-2 spike protein, and spike protein amino acid diversity data, obtained from Nextstrain.org [37], is shown in Fig. 1E.

Purification and mass spectroscopy analysis of CBM9 fusion proteins.

From the recombinant CBM9 fusion clones that we chose, we expressed and isolated the recombinant protein using powdered cellulose in a batch purification. The resulting purified proteins were subjected to mass spectroscopy analysis to determine the molecular weight of the dominant purified product. The results (Table 1) indicated, as expected, that all products had the N-terminal methionine removed. Most cloned products were processed, presumably by endogenous *E. coli* proteases, so that some portion of the C-terminal end was removed, effectively removing a few to several amino acids of the spike protein fragment. However, clones expressing

CBM9-(PT)₄P, CBM9-ID-F, CBM9-H1 and CBM9-N produced, as the dominant purified product, proteins that were 1 Dalton smaller than the predicted monoisotopic product. Since the CBM9-(PT)₄P dominant product was 1 Dalton smaller than predicted, we interpreted this to mean that an unidentified chemical modification occurs, many of which are documented [38], on the CBM9 module to remove one atomic mass unit. The clone expressing CBM9-ID-A was processed so as to remove only two amino acids from the B-cell epitope identified by Zhang et al. [10].

For further work we chose *E. coli* clones that expressed intact fusion proteins of CBM9 and fragments of SARS-CoV-2 proteins as the dominant recombinant products. The expression of clones CBM9-(PT)₄P, CBM9-ID-H1 and CBM9-N are shown in Fig. 1F and the purified products in Fig. 1G; for comparison, the carrier protein module CBM9 is also shown. By comparing the staining intensity of the protein band in the cell extracts to the band of purified CBM9-N, we estimated that the clones expressed recombinant product at levels of at least 100 mg/L upon IPTG-induction, and this is consistent with the 200 mg/L estimates of Kavooosi et al. [26]. We examined the purification yield of clone CBM9-ID-H1

(Table 2 and Additional file 1: Fig. S3). Because the CBM9-ID-H1 lacks any enzymatic activity that can be used to quantify its specific activity we used the inherent fluorescence of Coomassie blue stained proteins with 700 nm excitation to measure the relative amount of CBM9-ID-H1 in a protein mixture of known protein content. After 6 h of IPTG induction this clone yielded recombinant protein at about 27% of the total soluble protein, or 756 mg/L. The final recombinant protein yield of the specific protein band confirmed by mass spectroscopy to be the complete, intact CBM9-ID-H1 fusion protein was 16%, or 122 mg/L. A slightly smaller protein co-purified with CBM9-ID-H1, and presumably this is a proteolytic fragment of CBM9-ID-H1. This lesser band represents 33% of the total protein of the final purified protein preparation, and, thus, the CBM 9-ID-H1 is about 67% pure. That is, the final purified protein preparation yielded 122 mg/L of CBM9-ID-H1, and 60 mg/L of a fragment of CBM9-ID-H1. The contaminating band could be removed using a further purification step, but this may be unnecessary in an antibody detection assay as the contaminating protein is unlikely to interfere in the assay any more than the CBM9 portion of the CBM9-ID-H1 fusion protein. These experiments were performed using standard research growth flasks at an A_{600} of less than 10, and it is likely that the levels of recombinant protein produced could be significantly increased using an optimized fed-batch bioreactor protocol.

Different treatments were tested for their effect on the stability of purified CBM9 fusion protein (Fig. 1G). The CBM9 and the CBM9-ID-F samples were heated to 70 °C for 10 min and the CBM9-(PT)₄P, CBM9-ID-H1 and CBM9-N samples were filtered sterilized, all before storing at 4 °C for at least two weeks (Fig. 1G). As well, samples were stored at -20 °C in 50% glycerol (Fig. 1F last lane on right, CBM9-N). All storage conditions preserved the integrity of the sample. However, heating to 70 °C seemed to generate small amounts of multimers of the protein, consistent with previously reported observations for hyperthermophilic enzymes [39].

We constructed a number of clones of the ID-H region (see Additional file 1: Fig. S2C), and it was fortuitous that the CBM9-ID-H1 clone highly expressed a product that

was largely resistant to *E. coli* proteases. The ID-H region (residues 522–646) partially overlaps with the RBD (residues 319–541) of the SARS-CoV-2 spike protein, is only missing one confirmed glycosylation at N616 when produced in *E. coli* [22], and contains no cysteine pairs involved in disulphide bonds (though it does contain C525 which forms a disulphide bond with C391 in the RBD) [20]. In the 3D-structure oriented with the RBDs at the top, the ID-H1 region (residues 540–588) slightly overlaps with and lies below the RBD (Fig. 2A). It is possible that the CBM9-H1 recombinant product is resistant to proteases, while shorter CBM9 fusions are susceptible, because the ID-H1 clone encodes a potential self-folding protein domain (Fig. 2B). The region encompassed by CBM9-ID-H1 includes amino acid sequences identified by several groups as B-cell epitopes, as defined by synthetic peptides that are recognized by convalescent sera from COVID-19 patients (Fig. 2C).

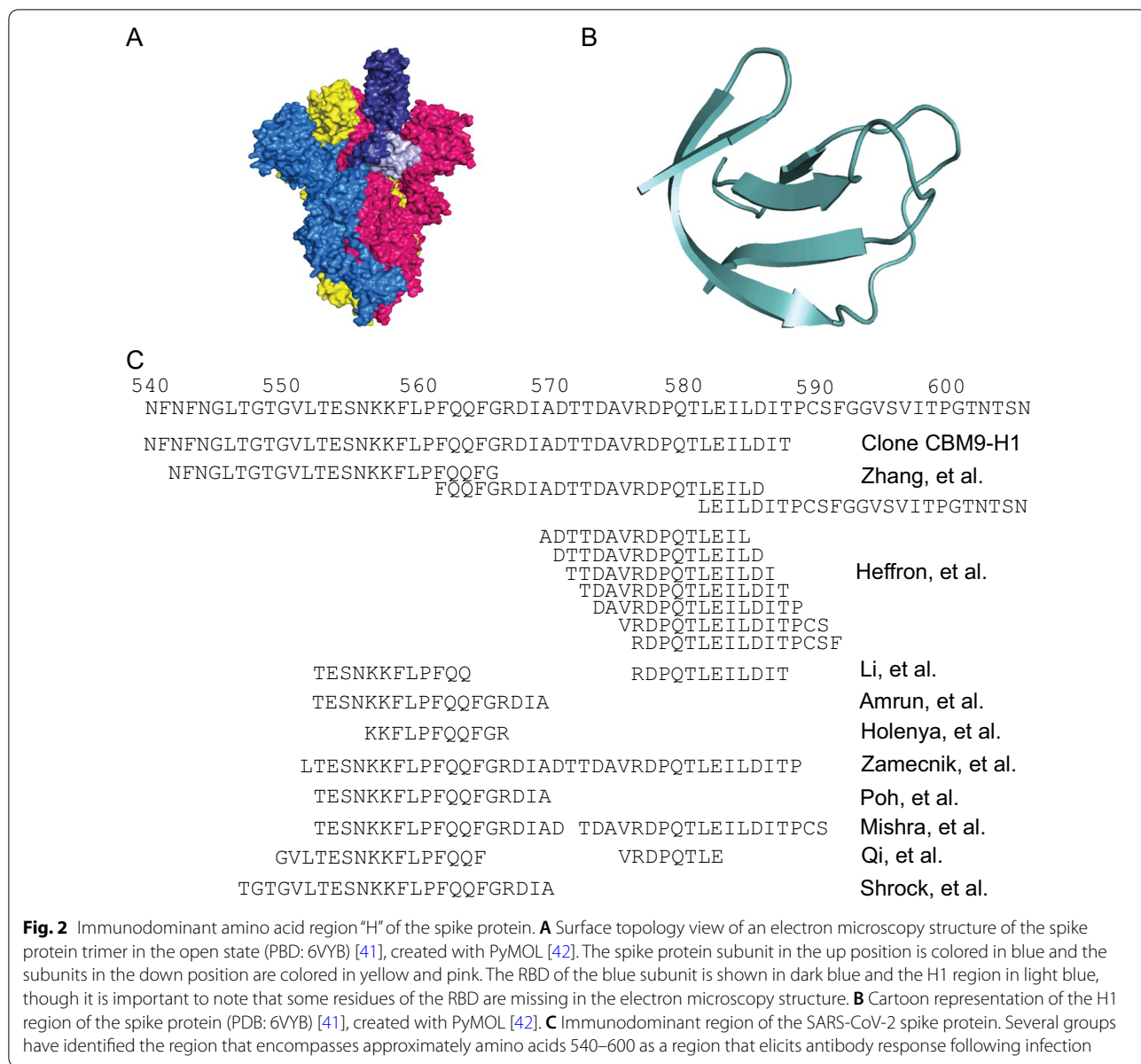
With the exception of the Alpha variant mutation A570D and the Omicron variant mutation T547K, all of the amino acid changes in the spike protein found in SARS-CoV-2 variants of concern lie outside of the region encompassed by the H1 clone [12, 37]. Additionally, the A570D mutation has decreased in frequency over time and is now exclusively an alanine, meaning that there is only one mutation in the H1 region at this time. Globally over time, the ID-H1 region appears to have lower diversity or higher conservation than other regions of the spike protein (Fig. 1E), but due to the variation in diversity at each amino acid position this observation is not statistically significant. As well, Poh et al. [9] found that titer of antibody in sera from COVID-19 convalescent patients that reacted with a peptide corresponding to amino acids 562–579 of the spike protein correlated with the amount of in vitro pseudovirus neutralization. When the neutralizing sera was depleted of reactivity against this peptide the neutralization activity fell sharply. Such evidence indicates that a strategy to elicit antibodies against this region may be an effective way of protecting against variants with amino acid changes in the RBD of the SARS-CoV-2 spike protein.

CBM9-SARS-CoV-2 fusion proteins react with rabbit anti-spike protein sera and human sera

As stated above, we used proline-threonine flexible linker regions to join the CBM9 module to SARS-CoV-2 spike protein and nucleocapsid protein regions. In this research the purpose of the linker was to allow the SARS-CoV-2 protein fragment to be accessible to antibody binding. To determine if the linker accomplished this, we reacted purified CBM9-(PT)₄P, CBM9-ID-F, CBM9-ID-H1 and CBM9-N with purified rabbit antibodies that had been raised to different portions of SARS-CoV-2 proteins

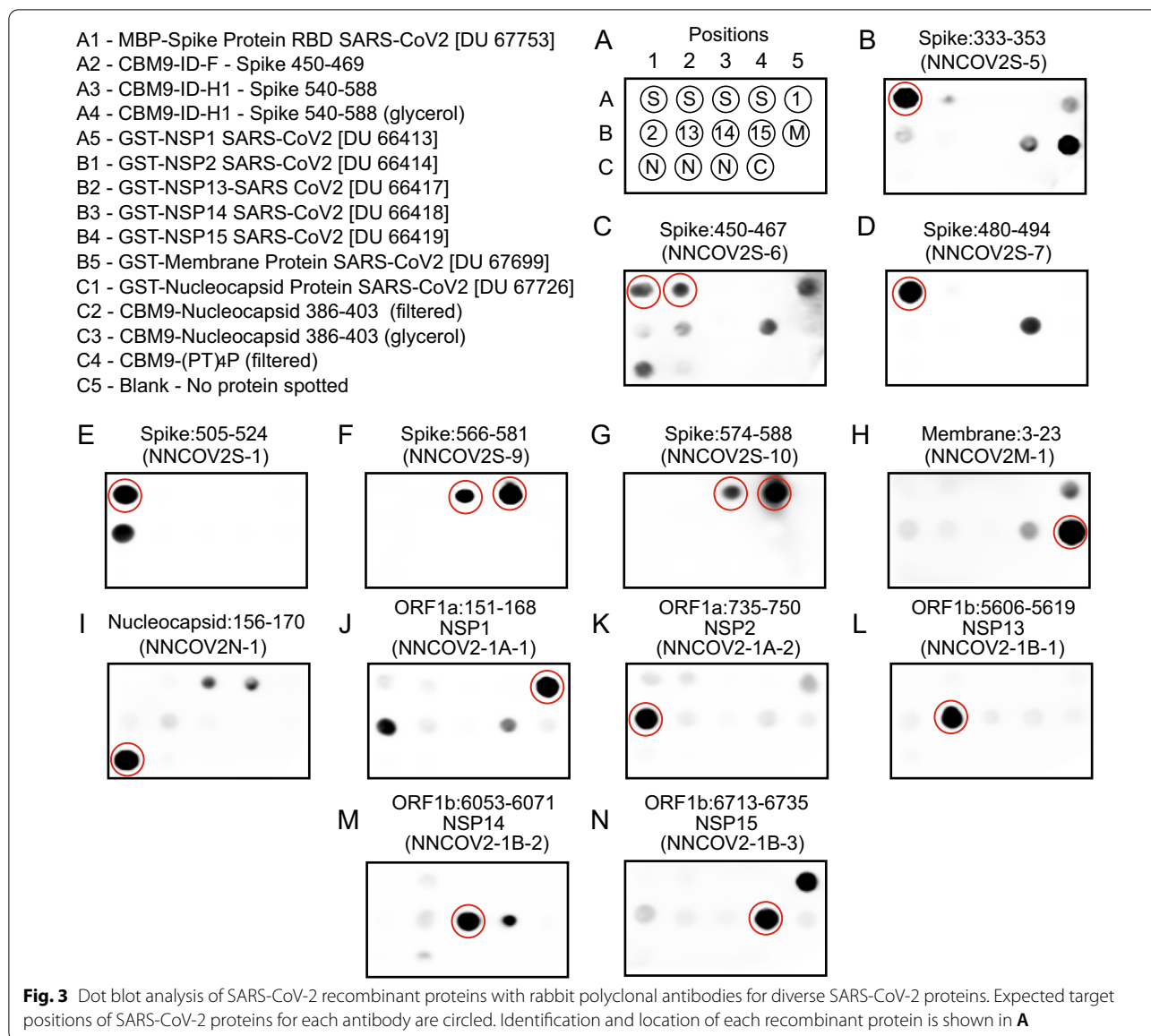
Table 2 Purification steps for CBM9-ID-H1 isolation

Purification step	Total protein (mg/L)	CBM9-H1 (mg/L)	Yield (%)	Fold purification
Soluble cell extract	2800	756	100	1
Eluate from cellulose resin	298	242	32	3
After buffer exchange	182	122	16	2.5



(Fig. 3) and human sera (Figs. 4, 5). In a semi-quantitative dot blot assay, we found that rabbit antibodies raised against the appropriate fragments of the SARS-CoV-2 spike protein reacted strongly with CBM9-H1 (Fig. 3E, G), but only weakly, or not at all, with antibodies directed against other regions of the spike protein. Reaction of the appropriate antibodies with CBM9-ID-F was moderately strong (Fig. 3C), and likewise poor or not detectable with the other antibodies. We had access to a small sampling of sera from COVID-19 confirmed (n=7), COVID-19 suspected (n=13), and healthy individuals (n=20) (Figs. 4, 5). While this small sample set size and the dot blot assay cannot provide an epidemiological story, the

results did show that human sera clearly reacted with the CBM9 fusions carrying ID-F and the nucleocapsid epitope. Many sera samples from both sick and healthy individuals reacted strongly with ID-F, indicating that this region may be similar in other coronaviruses or may be similar to another commonly encountered antigen. Surprisingly, the sera from several individuals, both healthy and ill, reacted apparently more strongly with CBM9-ID-F (spike amino acids 450–469) than with the MBP-RBD fusion protein (spike amino acids 319–541), even though the latter encompasses the ID-F region. These results may reflect a property of the antigen, such as accessibility of the SARS-CoV-2 portion to the antibody; or it may be



that the ID-F region is an especially immunogenic region of the RBD. Overall, with this small sample of sera there was little difference between the patterns of reactivity of the sera from the sick and healthy groups. This high detection of anti-spike and anti-nucleocapsid immunoreactivity in serum samples from healthy individuals is consistent with previous studies using two different serological tests developed by Mesoscale Devices and Kinexus [40].

In this work, we have demonstrated that soluble fragments of the SARS-CoV-2 virus, fused to the CBM9 module through a flexible linker, could be produced at high levels—over 100 mg/L—using universally available equipment with inexpensive materials. The costs

of cultivating *E. coli* are 10- to 100-fold less expensive than the costs of growing non-microbial eukaryotic cells, which are the usual hosts for expressing SARS-CoV-2 antigens. Further, the cost of using cellulose powder for affinity purification of CBM9 fusion proteins is about 100- to 1000-fold less than using the conventional immobilized nickel resin or a combination of traditional protein purification columns, such as ion-exchange with size exclusion resins. Lastly, while we described the production and isolation of CBM9 fusion proteins, there may be applications that require the separation of the CBM9 module from the SARS-CoV-2 fragment. Often the components of a fusion protein are separated using a highly specific protease

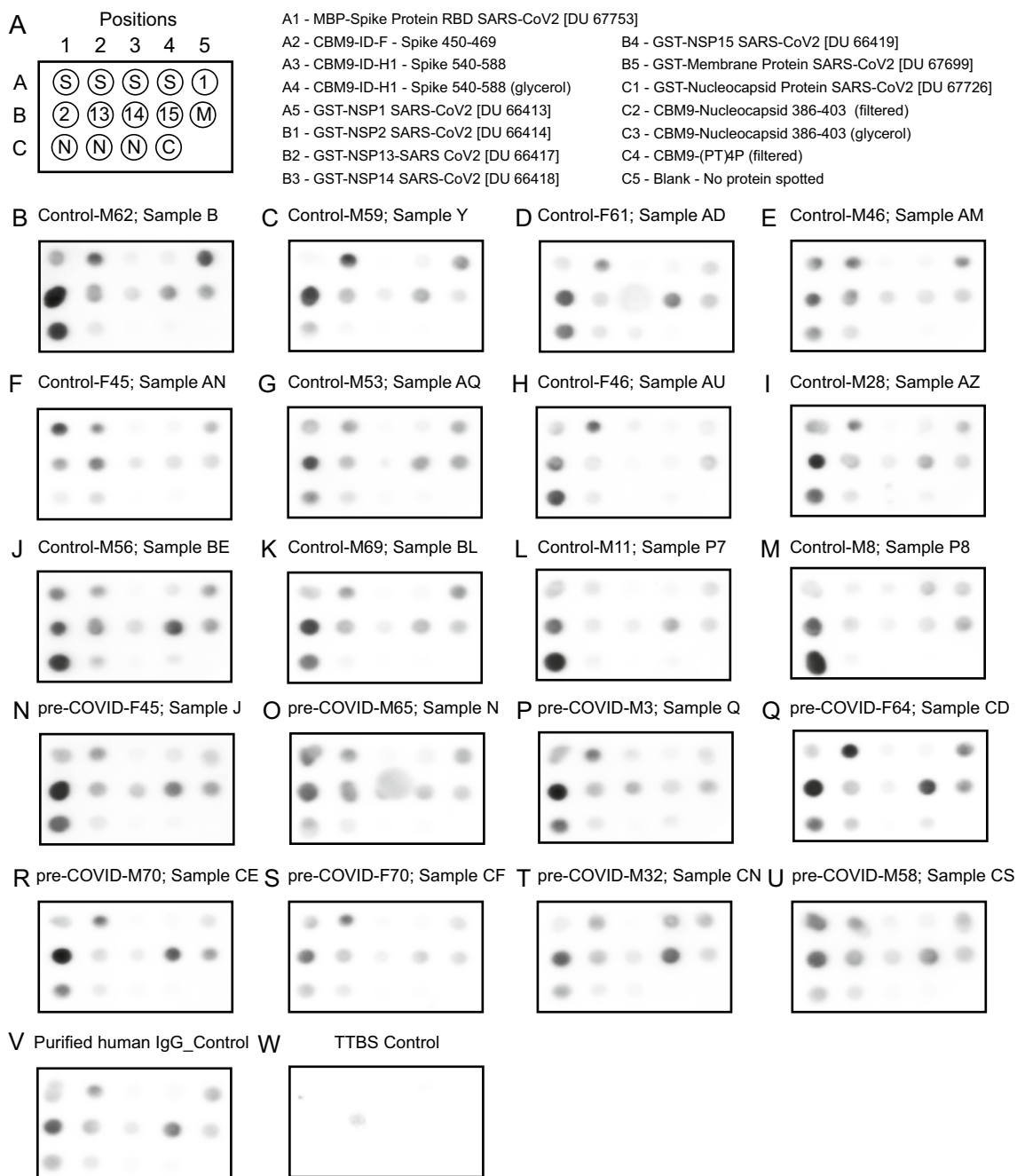


Fig. 4 Dot blot of proteins with sera from pre-COVID and healthy controls. **B–M** “Control” samples correspond to healthy individuals whose serum samples were collected in 2020. **N–U** “pre-COVID” samples were from healthy individuals whose serum samples were retrieved prior to April 2019. Individuals are also identified by sex (M for male and F for female) followed by age in years. The MBP-spike RBD protein (spot position A1) includes amino acids 319–541 of the spike protein

cleavage site. Indeed, using this approach Kavooosi et al. [26] found that a CBM9-GFP fusion protein remained largely intact even in the absence of protease inhibitors, unless cleaved with factor Xa when a cleavage site was incorporated into the linker. Thus, the use of

CBM9-SARS-CoV-2 protein fragment fusions allows for the economical production of antigens to be used for a variety of purposes, including in COVID-19 serological assays.

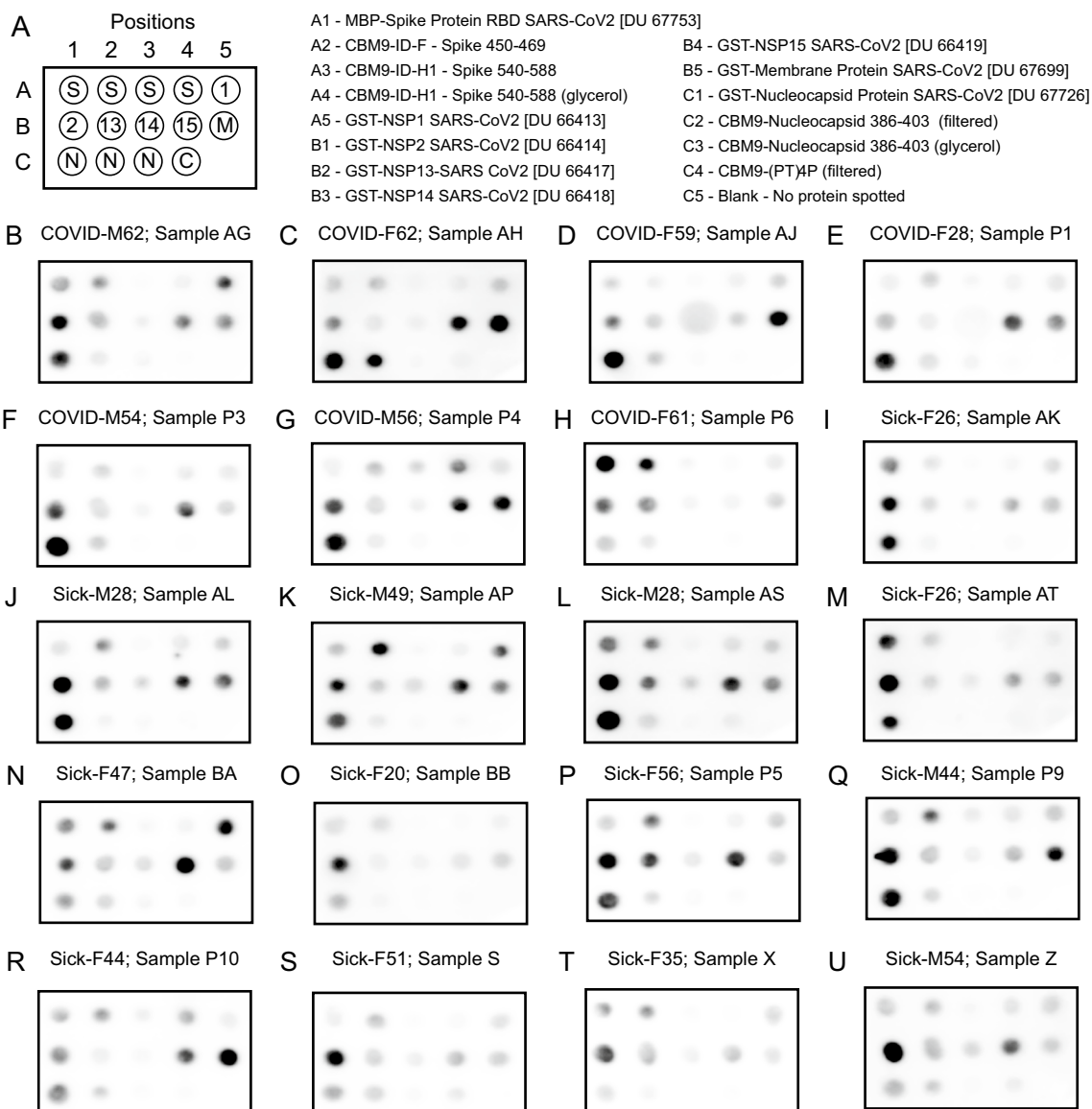


Fig. 5 Dot blot of proteins with sera from COVID-19 and sick individuals. **B–H** “COVID” samples correspond to individuals that PCR-tested positive for SARS-CoV-2 RNA and whose serum samples were collected in 2020. **I–U** “Sick” samples were from individuals who had COVID-19 symptoms and whose serum samples were collected in 2020. Individuals are also identified by sex (M for male and F for female) followed by age in years. The MBP-spike RBD protein (spot position A1) includes amino acids 319–541 of the spike protein

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12934-022-01753-0>.

Additional file 1. Proteomics methods, Source of commercial recombinant proteins, Sources of rabbit anti-peptide antibodies, Accession number of recombinant clones. **Table S1.** Primers used in this study. **Figure S1.** Example sequences of a CBM9-ID clone. **Figure S2.** SDS-PAGE of all recombinant clones. **Figure S3.** SDS-PAGE of CBM9-ID-H1 isolation and purification.

Acknowledgements

We thank Jeff Wang for aid in the drawing of blood samples from the study participants. Our thanks go also to Akshra Atrey for the probing of the dot blot arrays, and to Andrew Rodd for scanning SDS-PAGE gels. This work was supported by Grants from the Natural Sciences and Engineering Research Council of Canada (ALLRP 553524-20 and RGPIN-2018-03747).

Authors' contributions

BM: data analysis and manuscript preparation; JM: bench research; VT: bench research, experimental design, data analysis, manuscript preparation; LC: bench research; CS: bench research; RM: conceptualization, data analysis; DW: bench research, concept development, data analysis, manuscript preparation;

SP: concept development, experimental design, data analysis, manuscript preparation; FN: bench research, concept development, experimental design, data analysis, manuscript preparation. All of the authors consented to publication of this material. All authors read and approved the final manuscript.

Data and materials availability

All of the data relevant to the conclusion are presented in the manuscript or in the material presented in the additional file. The accession numbers for the nucleotide sequences of the recombinant clones and the source for obtaining clones is described in Additional file 1.

Declarations

Ethics approval and consent to participate

Ethics approval for obtaining the human blood samples is described under "Human Serum Samples" in the Materials and Methods section of this manuscript.

Competing interests

SP is the majority shareholder of Kinexus Bioinformatics Corporation. All other authors declare no conflicts.

Author details

¹Department of Biochemistry and Microbiology, University of Victoria, STN CSC, PO Box 1700, Victoria, BC V8W 2Y2, Canada. ²Kinexus Bioinformatics Corporation, Vancouver, BC, Canada. ³Division of Neurology, Department of Medicine, University of British Columbia, Vancouver, BC, Canada.

Received: 6 October 2021 Accepted: 28 January 2022

Published online: 05 February 2022

References

- Yoshimoto FK. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2 or n-CoV19), the cause of COVID-19. *Protein J.* 2020;39:198–216.
- Zamecnik CR, Rajan JV, Yamauchi KA, Mann SA, Loudermilk RP, Sowa GM, et al. ReScan, a multiplex diagnostic pipeline, pans human sera for SARS-CoV-2 antigens. *Cell Reports Med.* 2020;1:100123.
- Amrun SN, Lee CYP, Lee B, Fong SW, Young BE, Chee RSL, et al. Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBioMedicine.* 2020;58:9.
- Holenya P, Lange PJ, Reimer U, Woltersdorf W, Panterodt T, Glas M, et al. Peptide microarray-based analysis of antibody responses to SARS-CoV-2 identifies unique epitopes with potential for diagnostic test development. *Eur J Immunol.* 2021;51:1839–49.
- Qi H, Ma M, Jiang H, Ling J, Chen L, Zhang H, et al. Systematic profiling of SARS-CoV-2-specific IgG epitopes at amino acid resolution. *Cell Mol Immunol.* 2021;18:1067–9.
- Mishra N, Huang X, Joshi S, Guo C, Ng J, Thakkar R, et al. Immunoreactive peptide maps of SARS-CoV-2. *Commun Biol.* 2021;4:1–7.
- Shrock E, Fujimura E, Kula T, Timms RT, Lee IH, Leng Y, et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science.* 2020;370:8.
- Heffron AS, McIlwain SJ, Amjadi MF, Baker DA. The landscape of antibody binding in SARS-CoV-2 infection. *PLoS Biol.* 2021;19:e3001265.
- Poh CM, Carissimo G, Wang B, Amrun SN, Lee CYP, Chee RSL, et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun.* 2020;11:2806.
- Zhang B, Hu Y, Chen L, Yau T, Tong Y, Hu J, et al. Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Res.* 2020;30:702–4.
- Li Y, Lai D, Zhang H, Jiang H, Tian X, Ma M, et al. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol.* 2020;17:1095–7.
- Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol.* 2021;19:409–24.
- Yang J, Wang W, Chen Z, Lu S, Yang F, Bi Z, et al. A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature.* 2020;586:572–7.
- Tortorici MA, Beltramello M, Lempp FA, Pinto D, Dang HV, Rosen LE, et al. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science.* 2020;370:950–7.
- Brochot E, Demeys B, Touzé A, Belouzard S, Dubuisson J, Schmit JL, et al. Anti-spike, anti-nucleocapsid and neutralizing antibodies in SARS-CoV-2 inpatients and asymptomatic individuals. *Front Microbiol.* 2020;11:1–8.
- Stadlbauer D, Amanat F, Chromikova V, Jiang K, Strohmaier S, Arunkumar GA, et al. SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup. *Curr Protoc Microbiol.* 2020;57:e100.
- Fujita R, Hino M, Ebihara T, Nagasato T, Masuda A, Lee JM, et al. Efficient production of recombinant SARS-CoV-2 spike protein using the baculovirus-silkworm system. *Biochem Biophys Res Commun.* 2020;529:257–62.
- Rihn SJ, Merits A, Bakshi S, Turnbull ML, Wickenhagen A, Alexander AJT, et al. A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. *PLoS Biol.* 2021;19:1–22.
- Fitzgerald GA, Komarov A, Kaznadzey A, Mazo J, Kireeva ML. Expression of SARS-CoV-2 surface glycoprotein fragment 319–640 in *E. coli*, and its refolding and purification. *Protein Expr Purif.* 2021;183:105861.
- Bellone ML, Puglisi A, Dal Piaz F, Hochkoeppler A. Production in *Escherichia coli* of recombinant COVID-19 spike protein fragments fused to CRM197. *Biochem Biophys Res Commun.* 2021;558:79–85.
- Gao X, Peng S, Mei S, Liang K, Khan MSI, Yong EG, et al. Expression and functional identification of recombinant SARS-CoV-2 receptor binding domain (RBD) from *E. coli* system. *Prep Biochem Biotechnol.* 2021;1:7.
- Zhao X, Chen H, Wang H. Glycans of SARS-cov-2 spike protein in virus infection and antibody production. *Front Mol Biosci.* 2021;8:1–10.
- Maeda DLNF, Tian D, Yu H, Dar N, Rajasekaran V, Meng S, et al. Killed whole-genome reduced-bacteria surface-expressed coronavirus fusion peptide vaccines protect against disease in a porcine model. *Proc Natl Acad Sci U S A.* 2021;118:8.
- Notenboom V, Boraston AB, Kilburn DG, Rose DR. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10a in native and ligand-bound forms. *Biochemistry.* 2001;40:6248–56.
- Boraston AB, Creagh AL, Alam MM, Kormos JM, Tomme P, Haynes CA, et al. Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry.* 2001;40:6240–7.
- Kavoosi M, Meijer J, Kwan E, Creagh AL, Kilburn DG, Haynes CA. Inexpensive one-step purification of polypeptides expressed in *Escherichia coli* as fusions with the family 9 carbohydrate-binding module of xylanase 10A from *T. maritima*. *J Chromatogr B Anal Technol Biomed Life Sci.* 2004;807:87–94.
- Kavoosi M, Creagh ALL, Kilburn DGDG, Haynes CA. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli*. *Biotechnol Bioeng.* 2007;98:599–610.
- Schoepfer R. The pRSET family of T7 promoter expression vectors for *Escherichia coli*. *Gene.* 1993;124:83–5.
- Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* 2005;33:W526–531.
- Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE.* 2008;3:1–7.
- Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif.* 2005;41:207–34.
- Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun.* 2020;11:1620.
- Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581:215–20.
- Kim J-S, Liu L, Fitzsimmons LF, Wang Y, Crawford MA, Mastrogianni M, et al. DksA–DnaJ redox interactions provide a signal for the activation of bacterial RNA polymerase. *Proc Natl Acad Sci U S A.* 2018;115:E11780.
- Pribis JP, Garcia-Villada L, Zhai Y, Lewin-Epstein O, Wang AZ, Liu J, et al. Gamblers: an antibiotic-induced evolvable cell subpopulation

differentiated by reactive-oxygen-induced general stress response. *Mol Cell*. 2019;74:785–800.

36. Ishihama A, Kori A, Koshio E, Yamada K, Maeda H, Shimada T, et al. Intracellular concentrations of 65 species of transcription factors with known regulatory functions in *Escherichia coli*. *J Bacteriol*. 2014;196:2718–27.
37. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34:4121–3.
38. Marino G, Eckhard U, Overall CM. Protein termini and their modifications revealed by positional proteomics. *ACS Chem Biol*. 2015;10:1754–64.
39. Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev*. 2001;65:1–43.
40. Majdoubi A, Michalski C, O'Connell SE, Dada S, Narpala S, Gelinias J, et al. A majority of uninfected adults show preexisting antibody reactivity against SARS-CoV-2. *JCI Insight*. 2021;6:2–12.
41. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Velesler D. Structure, function, and antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020;181:281–92.
42. Schrödinger L. The PyMOL molecular graphics system, Version 2.4.0. 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

