1 *Escherichia coli* recombinant expression of SARS-CoV-2 protein fragments.

2

3

4 Bailey E. McGuire^a, Julia E. Mela^a, Vanessa C. Thompson^a, Logan R Cucksey^a, Claire E. Stevens^a, Ralph L.

5 McWhinnie^a, Dirk F. H. Winkler^b, Steven Pelech^{b,c}, and Francis E. Nano^{a#}

6

7 ^aDepartment of Biochemistry and Microbiology, University of Victoria, Victoria, B.C. Canada

8 ^bKinexus Bioinformatics Corporation, Vancouver, B.C. Canada

9 ^cDivision of Neurology, Department of Medicine, University of British Columbia, Vancouver, B.C. Canada

10

11 #Corresponding author

12 Department of Biochemistry and Microbiology

13 PO Box 1700 STN CSC

14 University of Victoria

15 Victoria, B.C. V8W 2Y2

16 fnano@uvic.ca

17

18 Key words: SARS-CoV-2, COVID-19, recombinant spike protein, carbohydrate-binding module, CBM9,

19 serological assay, low cost, *Escherichia coli*.

20

21 Running title: Low cost expression of SARS-CoV-2 protein fragments

22

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 used 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 fusion proteins were expressed in *E. coli* at about 0.1 g/L, and could be purified with a single affinity binding step using inexpensive cellulose powder. Three purified CBM9-SARS-CoV-2 fusion proteins were tested and found to bind antibody directed to the appropriate SARS-CoV-2 antigenic region. The largest intact CBM9 fusion protein incorporates spike protein amino acids 540-588, which is a conserved region immediately C-terminal to the receptor binding domain that is widely recognized by human convalescent sera and contains a putative protective epitope.

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 COVID-19.

Some studies of anti-SARS-CoV-2 antibody response examine antibody reactivity with linear epitopes using 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. In a similar fashion, our work explores the utility of expressing SARS-CoV-2 epitopes in *E. coli*, using the under-appreciated protein carrier family 9 carbohydrate-binding module (CBM9) from the *Thermotoga maritima* enzyme xylanase 10A (19, 20) that promotes high level protein expression and uses inexpensive reagents for protein purification (21, 22).

MATERIALS AND METHODS

Recombinant techniques. Plasmid pRSET5A was used as the backbone for all expression plasmid constructs (23). 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 DNA fragment was designed with the native codon composition, since this was known to function in *E. coli* from previous work (20) and this fragment was amplified with primers F-nCBD and R-nCBD which have BsaI sites on their ends. Since the native *CBM9* gene has an internal Esp3I site and pRSET5A has internal BsaI sites, the plasmid and CBM9 amplicons were separately digested with Esp3I and BsaI and then joined by a standard DNA ligation. The CBM9-C, F and H3 DNA fragments were codon optimized (24) 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 the products were joined to pRSET5A using a simultaneous cutting and ligation reaction (25) 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 Supplemental Table S1. A color-coded example of a CBM9 fusion clone is shown in Supplemental Fig. S1.

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) (26); 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 2 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 x g. The cell pellet from each 50 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). 3 g of glass beads (≤106 μm, Sigma), and 5 μL Benzonase® nuclease (Sigma) were added and the cell suspension was vortexed vigorously for two 1 min intervals, and then subjected to centrifugation for 10 min at 16,000 x g. The supernatant was added to 0.5 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 15 mL conical centrifuge tube. The tube was rocked for 1 h at room temperature, and the cellulose powder was pelleted by centrifugation for 2 min at 4,000 x g. The cellulose resin was washed three times by adding 12 mL of solution A, rocking for 15 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 elution buffer (1 M glucose, 15 mM NaCl, 10 mM Tris, pH 7.6), and rocked for 15 min. Centrifugation was applied to pellet the cellulose, and the supernatant was removed and added to a Vivaspın® 6, 10 kDa MW cut off (Sartorius) protein concentrator to change the buffer to 150 mM NaCl, 10 mM Tris, pH 7.6.

For SDS-PAGE analysis of the CBM9 fusion protein-expressing cultures, 1 mL of cultures were centrifuged and the cell pellet suspended in 150 μL of Bug Buster (EMD Millipore), supplemented with lysozyme and Benzonase® nuclease as described above. Cell extracts were analyzed by SDS-PAGE using MOPS-SDS running buffer and TruPAGE™ gels (Sigma).

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 the Supplemental Materials.

Commercial recombinant proteins and antibodies. The following recombinant SARS-CoV-2 proteins (18) that were expressed in *E. coli* were procured from the MRC Protein Phosphorylation and Ubiquitination Unit Reagents and Services at the University of Dundee (Dundee, Scotland; <https://mrcppu-covid.bio/>): MBP-Spike RBD amino acids (aa) 319-541 (DU 67753); GST-NSP1 aa 1-180 (DU 66413); GST-NSP2 aa 1-638 (DU 66414); NSP13 aa 1-601 (DU 66417); GST-NSP14 aa 1-527 (DU 66418); GST-NSP15 aa 1-527 (DU 66419); GST-Membrane Protein aa 1-222 (DU 67699); and GST-Nucleocapsid Protein aa 1-419 (DU 67726).

Rabbit polyclonal antibodies (Kinexus Bioinformatics) directed against synthetic peptides based on SARS-CoV-2 proteins were used in dot blots and are as follows: Spike aa 333-353 (NNCOV2S-5); Spike aa 450-467 (NNCOV2S-6); Spike aa 480-494 (NNCOV2S-7); Spike aa 505-524 (NNCOV2S-1); Spike aa 566-581 (NNCOV2S-9); Spike aa 574-588 (NNCOV2S-10); Membrane aa 3-23 (NNCOV2M-1); Nucleocapsid aa 156-170 (NNCOV2N-1); ORF1a aa 151-168 – NSP1 (NNCOV2-1A-1); ORF1a aa 735-750 – NSP2 (NNCOV2-1A-2); ORF1b aa 5606-5619 – NSP13 (NNCOV2-1B-1); ORF1b aa 6053-6071 – NSP14 (NNCOV2-1B-2) and ORF1b aa 6713-6735 – NSP15 (NNCOV2-1B-3).

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 16567-09:39:354-06-2020. 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”). For recovery of the serum, the collected blood was allowed to clot at room temperature. After about 2 h, possible clots sticking on the walls of the tubes were released by scraping with pipette tips around the inside walls of the collection tubes before the samples were centrifuged at 2000 x g for at least 15 min. When larger blood volumes (up to 10 mL) were processed, the top layer containing the serum was transferred into a syringe and sterile filtered into storage tubes. Sera from small blood samples (less than 500 μ L) were transferred directly into the storage tubes without filtration. The recovered serum was stored at -80 °C.

All of the preparations of recombinant proteins were robotically spotted as 0.24 μ L of a ~ 8 μ M final concentration (except GST-NSP2 SARS CoV2 [DU 66414] in Spot B1 was printed at a 6.5 μ M 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 for 30 min. After washing the membranes twice with T-TBS (TBS with 0.05% Tween 20), the arrays were incubated with either affinity-purified rabbit polyclonal antibodies against SARS-CoV-2 protein sequences at 2 μ g/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 °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 0.125 M 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.

Accession numbers and availability of recombinant clones. The nucleotide sequences for all of the clones described in this work were deposited with GenBank (CBM9-(PT)₄P, MZ322548; CBM9-ID-A, MZ322549; CBM9-ID-B, MZ322550; CBM9-ID-C, MZ322551; CBM9-ID-D, MZ322552; CBM9-ID-E, MZ322553; CBM9-ID-F, MZ322554; CBM9-ID-G, MZ322555; CBM9-ID-H1, MZ322556; CBM9-N, MZ322557). The recombinant clones expressing CBM9-(PT)₄P, CBM9-ID-F, CBM9-ID-H1 and CBM9-N have been deposited with AddGene (<https://www.addgene.org/>).

RESULTS AND DISCUSSION

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

Pure proteins or fragments of proteins from the SARS-CoV-2 virus can be used to test for the presence of anti-viral antibodies in serological assays. The amount of antigen needed can be minimized if the fragments of the viral proteins that elicit an antibody response are known and these specific fragments are used in the serological assay. 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 has been recombinantly produced in a number of hosts, including engineered HEK-293 cells, insect cells, and insect larvae (17, 27, 28). However, these expression systems require growth media and bioreactors that are orders of magnitude more expensive than the material and equipment used for microbial expression systems. Therefore, we felt that there was value in developing an inexpensive microbial expression system that can produce 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. 1 A-C). Kavosi et al. (21) showed that the CBM9 module expresses at high levels, even when a protein was fused to the C-terminus.

In a separate work (22), these same researchers showed that linking CBM9 to a protein with a proline-threonine rich linker ([PT]₄P-IEGR) resulted in a fusion protein that was resistant to protease attack by endogenous *E. coli* proteases (22). Thus, we adopted the use of (PT)₄P as a linker between the CBM9 and SARS-CoV-2 spike protein fragments, and an example of the gene organization is shown in Supplemental Fig. S1.

The bulk of the studies on the antibody response 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 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 (Supplemental 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 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.

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 (29), 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. 2A and the purified products in Fig. 2B; 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 about 100 mg/L upon IPTG-induction, and this is consistent with the 200 mg/L estimates of Kavoosi et al. (21). These experiments were performed using standard research growth flasks at an A₆₀₀ of less than 10, and presumably the levels of recombinant protein that are produced could be significantly increased using an optimized fed-batch bioreactor protocol.

The recombinant fusion proteins were isolated by binding to cellulose powder, using batch purification (Fig. 2B). Following this, the CBM9 and the CBM9-ID-F samples were heated to 70 °C for 10 minutes 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. 2B). As well, samples were stored at -20 °C in 50% glycerol (Fig. 2A 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 (30).

We constructed a number of clones of the ID-H region (see 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. 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. 3A). It is possible that the CBM9-H1 recombinant product is resistant to proteases, while shorter than other CBM9 fusions that are susceptible, because the ID-H1 clone encodes a potential self-folding protein domain (Fig. 3B). 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. 3C).

Interestingly, the region encompassed by the H1 clone does not encompass any of the amino acid changes found among any of the variants of concern (12), but it does contain a putative protective epitope. 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. Further, when the neutralizing sera was depleted of reactivity against the peptide representing amino acids 562-579 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 (Fig. 4) and human sera (Fig. 5 and Fig. 6). 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. 4F, 4G), 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. 4C), 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) (Fig. 5 and Fig. 6). 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 (31).

In this work, we have demonstrated that fragments of the SARS-CoV-2 virus, fused to the CBM9 module through a flexible linker, could be produced at high levels - about 0.1 g/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 1,000-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 cleavage site. Indeed, using this approach Kavoosi et al. (21) found that a CBM9-GFP fusion protein remained largely intact even in the absence of protease inhibitors, unless cleaved with factor Xa whose cleavage site was incorporated into the linker. While specific proteases may be found that function with a

specific CBM9 fusion, with or without addition of the cleavage site into the linker, it is also possible to use chemical treatments that cleave with an acceptable level of specificity. For example, in the case of recombinant product CBM9-ID-H1, 45 amino acids of the 49 amino acid SARS-CoV-2 fragment could be released by treatment with hydroxylamine which attacks asparaginyl-glycine bonds (32) as CBM9 lacks an arginine-glycine sequence. 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.

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.

Funding. This work was supported by grants from Natural Sciences and Engineering Research Council of Canada (ALLRP 553524 – 20 and RGPIN-2018-03747).

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

REFERENCES

1. Yoshimoto FK. 2020. The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J* 39:198–216. <https://doi.org/10.1007/s10930-020-09901-4>.
2. Zamecnik CR, Rajan J V., Yamauchi KA, Mann SA, Loudermilk RP, Sowa GM, Zorn KC, Alvarenga BD, Gaebler C, Caskey M, Stone M, Norris PJ, Gu W, Chiu CY, Ng D, Byrnes JR, Zhou XX, Wells JA, Robbiani DF, Nussenzweig MC, DeRisi JL, Wilson MR. 2020. ReScan, a multiplex diagnostic pipeline, pans human sera for SARS-CoV-2 antigens. *Cell Reports Med* 1. <https://doi.org/10.1016/j.xcrm.2020.100123>.

3. Amrun SN, Lee CYP, Lee B, Fong SW, Young BE, Chee RSL, Yeo NKW, Torres-Ruesta A, Carissimo G, Poh CM, Chang ZW, Tay MZ, Chan YH, Chen MIC, Low JGH, Tambyah PA, Kalimuddin S, Pada S, Tan SY, Sun LJ, Leo YS, Lye DC, Renia L, Ng LFP. 2020. Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBioMedicine* 58. <https://doi.org/10.1016/j.ebiom.2020.102911>.
4. Holenya P, Lange PJ, Reimer U, Woltersdorf W, Panterodt T, Glas M, Wasner M, Eckey M, Drosch M, Hollidt JM, Naumann M, Kern F, Wenschuh H, Lange R, Schnatbaum K, Bier FF. 2021. Peptide microarray-based analysis of antibody responses to SARS-CoV-2 identifies unique epitopes with potential for diagnostic test development. *Eur J Immunol*. <https://doi.org/10.1002/eji.202049101>.
5. Qi H, Ma M liang, Jiang H wei, Ling J ya, Chen L yun, Zhang H nan, Lai D yun, Li Y, Guo Z wen, Hu C sheng, Guo S juan, Meng Q feng, Ren Y, Yang X, Wang W, Zhou J, Zhao X dong, Li H, Tao S ce. 2021. Systematic profiling of SARS-CoV-2-specific IgG epitopes at amino acid resolution. *Cell Mol Immunol* 18:1067–1069. <https://doi.org/10.1038/s41423-021-00654-3>.
6. Mishra N, Huang X, Joshi S, Guo C, Ng J, Thakkar R, Wu Y, Dong X, Li Q, Pinapati RS, Sullivan E, Caciula A, Tokarz R, Briese T, Lu J, Lipkin WI. 2021. Immunoreactive peptide maps of SARS-CoV-2. *Commun Biol* 4:1–7. <https://doi.org/10.1038/s42003-021-01743-9>.
7. Shrock E, Fujimura E, Kula T, Timms RT, Lee IH, Leng Y, Robinson ML, Sie BM, Li MZ, Chen Y, Logue J, Zuiani A, McCulloch D, Lelis FJN, Henson S, Monaco DR, Travers M, Habibi S, Clarke WA, Caturegli P, Laeyendecker O, Piechocka-Trocha A, Li JZ, Khatri A, Chu HY, Villani AC, Kays K, Goldberg MB, Hacohen N, Filbin MR, Yu XG, Walker BD, Wesemann DR, Larman HB, Lederer JA, Elledge SJ. 2020. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* 370. <https://doi.org/10.1126/science.abd4250>.
8. Heffron AS, McIlwain SJ, Amjadi MF, Baker DA. 2020. The landscape of antibody binding in SARS-CoV-2 infection. *bioRxiv*. <https://doi.org/10.1101/2020.10.10.334292>.
9. Poh CM, Carissimo G, Wang B, Amrun SN, Lee CYP, Chee RSL, Fong SW, Yeo NKW, Lee WH,

- 361 Torres-Ruesta A, Leo YS, Chen MIC, Tan SY, Chai LYA, Kalimuddin S, Kheng SSG, Thien SY, Young
362 BE, Lye DC, Hanson BJ, Wang CI, Renia L, Ng LFP. 2020. Two linear epitopes on the SARS-CoV-2
363 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* 11.
364 <https://doi.org/10.1038/s41467-020-16638-2>.
- 365 10. Zhang B, Hu Y, Chen L, Yau T, Tong Y, Hu J, Cai J, Chan K-H, Dou Y, Deng J, Wang X, Hung IF-N,
366 To KK-W, Yuen KY, Huang J-D. 2020. Mining of epitopes on spike protein of SARS-CoV-2 from
367 COVID-19 patients. *Cell Res* 30:702–704. <https://doi.org/10.1038/s41422-020-0366-x>.
- 368 11. Li Y, Lai D yun, Zhang H nan, Jiang H wei, Tian X, Ma M liang, Qi H, Meng Q feng, Guo S juan, Wu
369 Y, Wang W, Yang X, Shi D wei, Dai J biao, Ying T, Zhou J, Tao S ce. 2020. Linear epitopes of SARS-
370 CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol* 17:1095–
371 1097. <https://doi.org/10.1038/s41423-020-00523-5>.
- 372 12. Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R,
373 Rambaut A, COVID-19 Genomics UK (COG-UK) Consortium, Peacock SJ, Robertson DL. 2021.
374 SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 19.
375 <https://doi.org/10.1038/s41579-021-00573-0>.
- 376 13. Yang J, Wang W, Chen Z, Lu S, Yang F, Bi Z, Bao L, Mo F, Li X, Huang Y, Hong W, Yang Y, Zhao Y,
377 Ye F, Lin S, Deng W, Chen H, Lei H, Zhang Z, Luo M, Gao H, Zheng Y, Gong Y, Jiang X, Xu Y, Lv Q,
378 Li D, Wang M, Li F, Wang S, Wang G, Yu P, Qu Y, Yang L, Deng H, Tong A, Li J, Wang Z, Yang J,
379 Shen G, Zhao Z, Li Y, Luo J, Liu H, Yu W, Yang M, Xu J, Wang J, Li H, Wang H, Kuang D, Lin P, Hu
380 Z, Guo W, Cheng W, He Y, Song X, Chen C, Xue Z, Yao S, Chen L, Ma X, Chen S, Gou M, Huang W,
381 Wang Y, Fan C, Tian Z, Shi M, Wang FS, Dai L, Wu M, Li G, Wang G, Peng Y, Qian Z, Huang C, Lau
382 JYN, Yang Z, Wei Y, Cen X, Peng X, Qin C, Zhang K, Lu G, Wei X. 2020. A vaccine targeting the RBD
383 of the S protein of SARS-CoV-2 induces protective immunity. *Nature* 586:572–577.
384 <https://doi.org/10.1038/s41586-020-2599-8>.
- 385 14. Tortorici MA, Beltramello M, Lempp FA, Pinto D, Dang H V., Rosen LE, McCallum M, Bowen J,

- Minola A, Jaconi S, Zatta F, De Marco A, Guarino B, Bianchi S, Lauron EJ, Tucker H, Zhou J, Peter A, Havenar-Daughton C, Wojcechowskyj JA, Case JB, Chen RE, Kaiser H, Montiel-Ruiz M, Meury M, Czudnochowski N, Spreafico R, Dillen J, Ng C, Sprugasci N, Culap K, Benigni F, Abdelnabi R, Foo SYC, Schmid MA, Cameroni E, Riva A, Gabrieli A, Galli M, Pizzuto MS, Neyts J, Diamond MS, Virgin HW, Snell G, Corti D, Fink K, Veessler D. 2020. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* 370:950–957. <https://doi.org/10.1126/science.abe3354>.
15. Brochet E, Demey B, Touzé A, Belouzard S, Dubuisson J, Schmit JL, Duverlie G, Francois C, Castelain S, Helle F. 2020. Anti-spike, anti-nucleocapsid and neutralizing antibodies in SARS-CoV-2 inpatients and asymptomatic individuals. *Front Microbiol* 11:1–8. <https://doi.org/10.3389/fmicb.2020.584251>.
16. Stadlbauer D, Amanat F, Chromikova V, Jiang K, Strohmeier S, Arunkumar GA, Tan J, Bhavsar D, Capuano C, Kirkpatrick E, Meade P, Brito RN, Teo C, McMahon M, Simon V, Krammer F. 2020. SARS-CoV-2 seroconversion in humans: A detailed protocol for a serological assay, antigen production, and test setup. *Curr Protoc Microbiol* 57. <https://doi.org/10.1002/cpmc.100>.
17. Fujita R, Hino M, Ebihara T, Nagasato T, Masuda A, Lee JM, Fujii T, Mon H, Kakino K, Nagai R, Tanaka M, Tonooka Y, Moriyama T, Kusakabe T. 2020. Efficient production of recombinant SARS-CoV-2 spike protein using the baculovirus-silkworm system. *Biochem Biophys Res Commun* 529:257–262. <https://doi.org/10.1016/j.bbrc.2020.06.020>.
18. Rihn SJ, Merits A, Bakshi S, Turnbull ML, Wickenhagen A, Alexander AJT, Baillie C, Brennan B, Brown F, Brunker K, Bryden SR, Burness KA, Carmichael S, Cole SJ, Cowton VM, Davies P, Davis C, de Lorenzo G, Donald CL, Dorward M, Dunlop JJ, Elliott M, Fares M, da Silva Filipe A, Freitas JR, Furnon W, Gestuveo RJ, Geyer A, Giesel D, Goldfarb DM, Goodman N, Gunson R, James Hastie C, Herder V, Hughes J, Johnson C, Johnson N, Kohl A, Kerr K, Leech H, Lello LS, Li K, Lieber G, Liu X, Lingala R, Loney C, Mair D, McElwee MJ, McFarlane S, Nichols J, Nomikou K, Orr A, Orton RJ, Palmarini M, Parr YA, Pinto RM, Raggett S, Reid E, Robertson DL, Royle J, Cameron-Ruiz N, Shepherd

- JG, Smollett K, Stewart DG, Stewart M, Sugrue E, Szemiel AM, Taggart A, Thomson EC, Tong L, Torrie LS, Toth R, Varjak M, Wang S, Wilkinson SG, Wyatt PG, Zusinaite E, Alessi DR, Patel AH, Zaid A, Wilson SJ, Mahalingam S. 2021. A plasmid DNA-launched SARS-CoV-2 reverse genetics system and coronavirus toolkit for COVID-19 research. PLoS Biol 19:1–22. <https://doi.org/10.1371/journal.pbio.3001091>.
19. Notenboom V, Boraston AB, Kilburn DG, Rose DR. 2001. Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10a in native and ligand-bound forms. Biochemistry 40:6248–6256. <https://doi.org/10.1021/bi0101704>.
20. Boraston AB, Creagh AL, Alam MM, Kormos JM, Tomme P, Haynes CA, Warren RAJ, Kilburn DG. 2001. Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. Biochemistry 40:6240–6247. <https://doi.org/10.1021/bi0101695>.
21. Kavoosi M, Meijer J, Kwan E, Creagh AL, Kilburn DG, Haynes CA. 2004. 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 807:87–94. <https://doi.org/10.1016/j.jchromb.2004.03.031>.
22. Kavoosi M, Creagh ALL, Kilburn DGDG, Haynes CA. 2007. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in *Escherichia coli*. Biotechnol Bioeng 98:599–610. <https://doi.org/10.1002/bit.21396>.
23. Schoepfer R. 1993. The pRSET family of T7 promoter expression vectors for *Escherichia coli*. Gene 124:83–85. [https://doi.org/10.1016/0378-1119\(93\)90764-T](https://doi.org/10.1016/0378-1119(93)90764-T).
24. Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, Jahn D. 2005. JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res 33. <https://doi.org/10.1093/nar/gki376>.
25. Engler C, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One 3:1–7. <https://doi.org/10.1371/journal.pone.0003647>.

26. Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41:207–234. <https://doi.org/10.1016/j.pep.2005.01.016>.
27. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, Guo L, Guo R, Chen T, Hu J, Xiang Z, Mu Z, Chen X, Chen J, Hu K, Jin Q, Wang J, Qian Z. 2020. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun* 11. <https://doi.org/10.1038/s41467-020-15562-9>.
28. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581:215–220. <https://doi.org/10.1038/s41586-020-2180-5>.
29. Marino G, Eckhard U, Overall CM. 2015. Protein termini and their modifications revealed by positional proteomics. *ACS Chem Biol* 10:1754–1764. <https://doi.org/10.1021/acscchembio.5b00189>.
30. Vieille C, Zeikus GJ. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 65:1–43. <https://doi.org/10.1128/mmbr.65.1.1-43.2001>.
31. Majdoubi A, Michalski C, O’Connell SE, Dada S, Narpala S, Gelinas J, Mehta D, Cheung C, Winkler DFH, Basappa M, Liu AC, Görges M, Barakauskas VE, Irvine M, Mehalko J, Esposito D, Sekirov I, Jassem AN, Goldfarb DM, Pelech S, Douek DC, McDermott AB, Lavoie PM. 2021. A majority of uninfected adults show preexisting antibody reactivity against SARS-CoV-2. *JCI Insight* 6:2–12. <https://doi.org/10.1172/jci.insight.146316>.
32. Bornstein P, Balian G. 1977. Cleavage at Asn-Gly bonds with hydroxylamine. *Methods Enzymol* 47:132–145. [https://doi.org/10.1016/0076-6879\(77\)47016-2](https://doi.org/10.1016/0076-6879(77)47016-2).
33. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181:281–292. <https://doi.org/10.1016/j.cell.2020.02.058>.
34. Schrödinger L. 2020. The PyMOL Molecular Graphics System, Version 2.4.0.

Table 1. Amino acid sequence of B-cell epitopes cloned as fusions to CBM9, and location of 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)	<u>NLTTRTQLPPAYTNSFTRGVYYPDKV</u> F /RSSVLHSTQ	
	Calculated: 26076.08	Observed: 26075.1
ID-B (spike 221-245)	<u>RDLPQGFSALEPLVDLPIGINIT</u> R /FQTLLALHRSYLTPGD	
	Calculated: 25560.93	Observed: 25559.9
ID-C (spike 261-285)	<u>SSGWTAGAAAY</u> Y /YVGYLQPRTFLLKYNENGTTDAVD	
	Calculated: 23967.94	Observed: 23966.9
ID-D (spike 330-349)	<u>ESIVRFPNITN</u> N /LCPFGEVFNATRFASVYAWNRKR	
	Calculated: 24216.16	Observed: 24215.2
ID-E (spike 375-394)	<u>NSASFSTFK</u> K /CYGVSPKLNLDLCFTNVYAD	
	Calculated: 23914.95	Observed: 23913.9
ID-F (spike 450-469)	<u>DSKVGGNYYLYRLFRKSNLKPFERDISTE</u> IIYQ	
	Calculated: 26940.52	Observed: 26939.5
ID-G (spike 480-499)	<u>QAGSTPCNGVEGFNCYFPLQ</u> Q /SYGFQPTNGVGYQ	
	Calculated: 23967.94	Observed: 23966.9
ID-H1 (spike 540-588)	<u>NFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQ</u> TLEILDIT	
	Calculated: 28384.23	Observed: 28383.2
N (nucleocapsid 386-403)	<u>QKKQQTVTLLPAADLDDF</u>	
	Calculated: 24957.55	Observed: 24956.57
CBM9-(TP) ₄ P	Not applicable:	
	Calculated: 22945.49	Observed: 22944.51

*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. Underlined sequence corresponds to synthetic peptide found to bind antibody as determined by Zhang, et al. (10). The highlighted region with the slash indicates the putative protease cleavage site, by endogenous *E. coli* proteases.

Figure 1. Strategies used to clone SARS-CoV-2 epitopes as fusions to CBM9. A. Initial clones were made by amplifying pRSET5A by inverse PCR, and ligating the plasmid amplicon to synthetic DNA encoding CBM9 with a linker (abbreviated to save space) 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. In Panels A-C the bent arrow indicates T7 promoter region; half circle indicates T7 ribosome binding site; “T” symbol represents transcriptional terminator.

Figure 2. A. 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. B. Samples of CBM9 fusion proteins purified by batch absorption to cellulose powder, after storage at 4 $^{\circ}$ C for a minimum of two weeks.

Figure 3. A. Surface topology view of an electron microscopy structure of the Spike protein trimer in the open state (PDB: 6VYB) (33), created with PyMOL (34). The S protein subunit in the up position is colored in blue and the subunits in the down position are colored in yellow and pink. The RBD of the blue subunit is shown in dark blue and the H1 region in light blue, though it is important to note that some residues of the RBD are missing in the electron microscopy structure. B. Cartoon representation of the H1 region of the S protein (PDB: 6VYB) (33), created with PyMOL (34). C. Immunodominant region of SARS-CoV-2 spike protein. Several groups have identified the region that encompasses approximately amino acids 540-600 as a region that elicits antibody response following infection.

Figure 4. Dot blotting analysis of SARS-CoV-2 recombinant proteins with rabbit polyclonal antibodies for diverse SARS-CoV-2 proteins. Expected target positions of SARS-CoV-2 proteins for each antibody are circled. Identification and location of each recombinant protein is shown in Panel A.

Figure 5. Dot blot of proteins with sera from pre-COVID and healthy controls. Panels B-M, “Control” samples correspond to healthy individuals whose serum samples were collected in 2020. Panels 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.

Figure 6. Dot blot of proteins with sera from COVID-19 and sick individuals. Panels B-H, “COVID” samples correspond to individuals that PCR-tested positive for SARS-CoV-2 RNA and whose serum samples were collected in 2020. Panels 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.