Synthetic protein antigens for COVID-19 diagnostics

Catherine H. Schein, MSc, PhD\textsuperscript{a,b,*}, Corri B. Levine, MS\textsuperscript{c}, Susan L F McLellan, MD, MPH\textsuperscript{d}, Surendra S. Negi\textsuperscript{a,e}, PhD, Werner Braun, PhD\textsuperscript{a,c}, Stephen C. Dreskin, MD, PhD\textsuperscript{f}, Elizabeth S. Anaya,BSc, Jurgen Schmidt, PhD\textsuperscript{g}

\textsuperscript{a}Department of Biochemistry and Molecular Biology, \textsuperscript{b}Institute for Human Infections and immunity (IHII), \textsuperscript{c}Institute for Translational Sciences, \textsuperscript{d}Department of Internal medicine – Infectious Diseases, \textsuperscript{e}Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, TX.
\textsuperscript{f}Division of Allergy and Clinical Immunology, Department of Medicine, University of Colorado Denver, Aurora, CO. 80045
\textsuperscript{g}B-11 Bioenergy and Biome Sciences, Bioscience Division Los Alamos National Laboratory, Los Alamos, NM 87545
*to whom correspondence should be addressed, chschein@utmb.edu

Orcid ID: 0000-0002-8290-2109

Abstract

There is an urgent need for inexpensive, rapid and specific antigen-based assays to test for infection with SARS-CoV-2 and distinguish variants arising as the COVID-19 pandemic spreads. We have identified a small, synthetic protein (JS7), representing a region of maximum variability within the receptor binding domain (RBD), which binds antibodies in sera from nine patients with PCR-verified COVID-19 of varying severity. Antibodies binding to either JS7 or the SARS-CoV-2 recombinant RBD, as well as those that disrupt binding between a fragment of the ACE2 receptor and the RBD, are proportional to disease severity and clinical outcome. Binding to JS7 was inhibited by linear peptides
from the RBD interface with ACE2. Variants of JS7, such as N501Y, can be quickly synthesized in a pure form in large quantities by automated methods. JS7 and related synthetic antigens can provide a basis for specific diagnostics for SARS-CoV-2 infections.

**Keywords**

COVID-19 variants  
Structure based design  
Receptor binding domain  
S protein epitopes  
ACE2 interaction
Introduction

COVID-19, caused by a new coronavirus (SARS-CoV-2)\textsuperscript{5, 6} is closely related, in its sequence, structure\textsuperscript{8}, binding to the human ACE2 receptor\textsuperscript{10} and epitopes recognized by neutralizing antibodies isolated from survivors\textsuperscript{1-11} of the SARS-CoV-1 (SARS) 2002-2003 outbreak\textsuperscript{12, 13} and COVID-19\textsuperscript{14}. In addition, the previously identified SARS-specific human antibody, CR3022, binds the receptor binding domain (RBD) of SARS-Cov-2 with nanomolar affinity\textsuperscript{16}. Practical tools to distinguish these viruses and especially the recently identified variants of SARS-CoV-2\textsuperscript{15, 17} are needed. However, the viruses diverge from each other in that the case fatality rate of the newer virus appears lower. SARS-CoV-2 is much more communicable via inhaled aerosols\textsuperscript{6, 18}, perhaps due to insertion of a furin cleavage site around position 701\textsuperscript{19}. SARS-CoV-2 has continued to mutate during its path through humanity, accumulating changes in this furin site and the RBD that may eventually affect its phenotype, immune sensitivity and resistance to therapies\textsuperscript{4, 20-23}. Although both SARS viruses are quite distinct in sequence from the even deadlier β-CoV that causes MERS\textsuperscript{24}, variants or recombinants of these pathogens could result in a virus that combines the high transmissibility of SARS-CoV-2 with the high mortality of SARS-CoV-1 or MERS. In addition, while the origins of both SARS viruses are unclear, there is a high probability that similar viruses are circulating, re-combining and mutating in some animal reservoir. Cross-over of the human virus into other species could also lead to changes in its infectivity severity, and resistance to therapies\textsuperscript{25, 26}. It is thus essential to have sensitive and specific diagnostic tools for the public health community to detect currently circulating variants. Outbreaks with mutated forms of SARS-CoV-2 are continuing to emerge\textsuperscript{15, 22, 27}. Many assays for determining prior COVID-19 infection are based on detecting antibodies in sera to recombinant versions of the surface (S) protein, which was previously identified as the binding site for neutralizing antibodies produced in response to SARS-CoV-1. At least one SARS-CoV-1 neutralizing monoclonal antibody\textsuperscript{28, 29} also binds with less affinity to the S
protein SARS-CoV-2 \(^{30}\), which varies by about 20% throughout its length (Fig. S1). Fortunately, methods developed for expressing the SARS S protein could be adapted to express also that of SARS-CoV-2, and the RBD, despite the sequence and length (1255/1277 aa for SARS/SARS-CoV2) diversity \(^{7}\). However, the S proteins of future variants may not be as amenable to recombinant synthesis.

Recent advances allow chemical synthesis of small proteins in large amounts, supplying pure protein variants rapidly. We have thus concentrated on reducing the size of the protein needed to distinguish COVID-19 infections to the area of maximum sequence variation between SARS-CoV-1 and SARS-CoV-2 \(^{9,31-33}\). This area also coincides with a region where the epitopes of many COVID-19 neutralizing monoclonal antibodies cluster.\(^{34}\) As we show here, synthetic proteins representing this region bind antibodies from sera of patients with COVID-19 infections of varying severity. Modifications of this protein can be the basis for detecting variants that may affect treatment protocols.

**Results and Discussion**

We synthesized two proteins, of about 10 kD and 7 KD (JS10 and JS7, Fig. 1, see Supplementary Material for production and characterization of the latter). This area of the RBD structure, in complex with the ACE2 in a cryo-EM structure\(^{35}\), contains a region of antiparallel \(\beta\)-sheet (Fig. 1), but is otherwise flexibly structured with 4 tyrosines (Y453, Y489, Y449, Y505) mediating many contacts across the interface. The recent UK mutations, N501Y, could enhance this binding. The CD

![Fig.1. The JS7 synthetic protein (S443-Y505, blue line) represents the most variable region of SARS-CoV-2 S protein (relative to SARS-CoV-1 from 2003) which mediates its interaction with ACE2 Top) A cryo-EM structure (PDB entry 7KMB) of the complex shows how the JS7 segment (red) of the RBD (blue) lies at the intersection with the ACE2 cell receptor (gray). Sidechains are shown for the N- (S443) and C- (Y505) terminal residues of JS7 and the 4 Y residues forming hydrogen bonds across the interface. Bottom) Alignment showing three circulating human variants (red arrows, yellow side chains) at L452R (recent California), E484K\(^4\) and N501Y\(^{15}\).
spectrum of JS7 (Fig. S2) suggests it forms the flexible structure and β-strands that characterize the experimental structure.

We compared JS7 to four recombinant proteins obtained from other groups, including two versions of the RBD of the SARS-CoV-2. One is expressed in 293 cells (derived from human embryonic kidney cells) and the other in yeast (proteins 1 and 4, respectively in Fig. 2), one approximately full length S protein of SARS-CoV-1 (JSP-657, protein 2), and its corresponding RBD area (RBD 219-N1). The two versions of the SARS CoV-2 RBD differ in that (1) was purified from a mammalian cell line attached to a linker protein and the other (4) was purified in yeast cells (Fig. 2). An alignment of the RBD sequences (proteins 1,3,4) is included in supplementary materials.

Convalescent sera from 9 hospitalized patients with COVID-19 disease of varying severity, from “mild” to critical, were used for dotspots, ELISAs and an assay for inhibition of association of the RBD with its cellular receptor, ACE2. As shown in Table 1S, these patients had co-morbidities that have been found, in other studies, to be associated with hospitalization due to COVID-19. Of the 9, 5 were known to be diabetic, and 6 had elevated glucose levels at time of diagnosis. Four of the 6 patients who had severe or critical disease were obese (BMI>30). Notably, the one patient who succumbed did not have these risk factors except for age >70. As Fig. 3 shows, sera from patients with severe or critical disease had antibodies that interfered with the binding of the SARS-CoV-2 RBD (protein 4) to an ACE2
fragment. Sera from the two patients with mild or moderate disease did not inhibit in this assay. However, both patients had antibodies that recognized the RBD and/or the JS7 in dotspots or ELISA; this recognition was inhibited by peptides from the RBD/ACE2 interface (Fig. 4, 5; see supplementary material for peptide sequences).

A dotspot assay showed that all 9 patient sera contained antibodies that recognized JS7 (spot #6), while recognition of the SARS-CoV-2 recombinant RBD (Spots 1 and 4) varied (Table 1, Table S1). Binding to JS7 was similar, in ELISA assays, to that of the SARS-CoV-2 RBD (Fig. 4). While all patients had antibodies that recognized both SARS-CoV-2 RBD proteins, one patient (16) with mild disease did not recognize the full-length S protein of SARS-CoV-1 (spot 2), in two separate blood draws 5 weeks apart (Table 1).

Antibody binding, as measured by ELISA of either protein was approximately proportional to the severity of the infection (Fig. 4, 5). Antibody binding could be reduced, for some sera to the level of

**Table 1. Antibody binding to recombinant S protein and JS7 from sera of 3 COVID 19 patients and a control serum.** Methods and data for all 9 patients and controls are described in Supplementary material. Recombinant proteins (1-4) or synthetic proteins (5,6) (see Fig. 2) were bound to nitrocellulose for dotspots and reacted with sera diluted 1:100. ELISA, for binding to recombinant SAR-CoV-2 RBD (Protein 4) or JS7 (6) is the highest dilution factor of serum where significant OD450 is measured; maximum dilution in the assay was 3.3 x 10^5. ACE2 inhibition is the degree to which each serum inhibited the binding of an ACE2 fragment to SARS-CoV-2 RBD (protein 4). A negative number indicates inhibition, where protein 4 inhibition (of binding to itself) = -0.82. (average of triplicates is shown).
controls, by adding peptides from the RBD/ACE2 interface (Figure 4, 5 and Supplementary material), further emphasizing the importance of this region for binding to the cell receptor. In the case of JS7, most of the binding could be abolished by adding only one of the peptides, from the central portion of the interface with ACE2. Thus detecting evidence of prior or ongoing COVID-19 infection with the smaller synthetic protein was as sensitive as with the larger RBD, and was specific for antibodies recognizing an area essential for interacting with the ACE2. This suggests it to be an excellent candidate for point of care diagnostics that could be used, in addition to PCR, to determine immune system involvement.

Assays with sera from normal volunteers who participated in a study of chronic urticaria, collected in Denver, CO USA in spring, 2019 before widespread introductions of SARS-Cov-2 into the US (Table S2), illustrates one of the great advantages to using the small synthetic proteins for diagnostics. Larger recombinant proteins may need to be expressed with protein tags, for solubility and/or ease in purification. As the strong binding of both control and patient sera to protein 1, which still contains such tags, shows these can cause assay artifacts if left intact (Tables 1, S1, S2). Antibodies in all the control sera showed somewhat less recognition of the RBD fragment of both SARS-CoV-1 and SARS-CoV-2 produced in yeast (proteins 3,4; the sequences of all 3 forms of the RBD are compared in Supplementary...
Material). Two of the control sera bound three of the four recombinant proteins to the same extent as patients with only mild COVID-19 disease; this was confirmed by ELISA (Figs. 4, 5). This binding could be due to previous infections with other coronaviruses, or simply reflect germline antibody recognition. In support of the latter, others have found that neutralizing monoclonal antibodies against the SARS-CoV-2 S protein, isolated from many different donors, were from public clonotypes, bound autoantigens and contained relatively few somatic mutations from germline.

The greatest advantage of using JS7 for diagnostic purposes is that its sequence can be rapidly modified to reflect variants now emerging in patients throughout the world. Although the N501Y mutant did not alter neutralization by polyclonal antibodies in some test sera, other changes in this area at the interface of the RBD and its ACE2 receptor (Fig. 1B) may alter the usefulness of treatments, such as convalescent plasma, monoclonal antibodies and small molecules designed to disrupt this interaction.

Several lines of evidence indicate that neutralization by serum antibodies and especially monoclonal antibodies is limited by variation in the JS7 area of the SARS viruses. The affinity of the cross reactive SARS-CoV-1 antibody, CR3022, whose epitope is located in the conserved area just upstream of the JS7 site, was greatly reduced by the single amino acid change, P to A found in the SARS-CoV-2 protein. Further, while all 9 COVID-19 patients in this study had antibodies that recognize the RBD of SARS-CoV-2, one patient with mild illness did not recognize the whole S protein of SARS-CoV-1 (protein 2), despite the 80% overall sequence identity (Table 1, S1). Further, some variants in this area, which also arose as escape mutants of monoclonal antibodies, can render the virus insensitive to neutralization by convalescent sera. The rise of these variants in several different areas of the world suggests they do not prevent virus replication and may be selected for under immune pressure.

In this day of very specific treatments, assays which discriminate among variants of SARS related viruses and other coronaviruses are essential. It is thus important to have a panel of antigens
representing areas of maximum variation in the RBD that can be used to rapidly determine specific variants and support planning clinical treatments. Reagents and linkers needed for assays can be integrated into variants of JS7 during synthesis, while controlling its sequence, stereochemistry and disulfide patterns. The JS7 protein, further modified to display variants and to be serum stable, may also have a future as a vaccine additive or booster. For example, recent methods have been developed to generate one component-synthetic proteins with incorporated adjuvant.²²

In conclusion, we show here that short synthetic proteins, which can be produced quickly in large quantities, can be the basis of specific assays to detect antibodies in sera that that are produced in response to infection with SARS-CoV-2. Antibodies in sera of 9 patients who had PCR confirmed, COVID-19 of differing severity recognized the JS7 protein. The sera from 8 of these patients was taken up to 2 months after diagnosis, while the 9th patient died during hospitalization. JS7 and future variants should prove to be another tool for eventually defeating the current outbreak.

Acknowledgements: We thank Wendy S. Baker for expert technical assistance, and Wen-Hsian Chen and others in the group at Baylor College of Medicine and Daniel Wrapp (Dartmouth) and the group of Jason McLellan of the University of Texas at Austin for supplying the recombinant proteins (1-4 in Fig 2) used throughout this work. Peptide and protein syntheses at LANL were supported by LDRD ER funding.

Literature cited


