Original Article

Reactivity of human antisera to codon optimized SARS-CoV2 viral proteins expressed in Escherichia coli

Yee-Huan Toh†, Yu-Weng Huang‡, Yo-Chen Chang§, Yi-Ting Chen‖, Ya-Ting Hsu*, Guang-Huey Lin†*‡*

Introduction

The current coronavirus disease 2019 (COVID-19) pandemic caused by the novel SARS-CoV2 coronavirus was first detected in China during late 2019 and has rapidly spread around the world since then. The Tzu Chi Foundation has around 100 service branches worldwide with more than 10 million volunteers, and these volunteers are exposed to high risk as they work to provide relief to those in need during the current pandemic. To provide safe working environments, accurate and rapid testing kits to identify uninfected, infected, and recovered populations are necessary, and here we describe the establishment of an Escherichia coli-based viral protein expression and purification platform that can produce codon-optimized SARS-CoV2 viral proteins for use in rapid antibody/antigen diagnostic kits.

The genome size of coronaviruses ranges between 26 and 32 kb, with a variable number of open reading frames (ORFs) [1]. The SARS-CoV2 genome is about 30 kb in size, and has 14 ORFs that encode a total of 29 known proteins [2]. The first ORF, which covers 67% of the entire genome, encodes 16 non-structural proteins, while the other encode structural and accessory proteins [2]. The S protein plays an important role in angiotensin-converting enzyme 2 (ACE2) receptor binding [3] and includes an N-terminal domain (NTD) of 271 amino acids (aa), receptor-binding domain (RBD) of 103 aa, and S2 domain of 600 aa. The RBD domain was expressed in amounts large enough to support large-scale production. The N-terminal domain, receptor-binding domain (RBD), Region 3, and the S2 domain were expressed in small but sufficient amounts for experiments. Immunoblotting results showed that anti-N IgG and anti-N IgM antibodies were detected in most patient sera, but only 60% of samples reacted with the recombinant RBD and S2 domain expressed by E. coli. Conclusion: The results indicated that codon-optimized SARS-CoV2 viral proteins can be expressed in E. coli and purified for rapid antibody detection kit preparation, with the codon-optimized N protein, RBD, and S2 protein demonstrating the most potential.

Keywords: Codon optimization, COVID-19, Escherichia coli expression system, N protein, SARS-CoV2 antisera

Abstract

Objective: The coronavirus disease 2019 (COVID-19) pandemic caused by the SARS-CoV2 virus continues to pose a serious threat to public health worldwide. The development of rapid diagnostic kits can assist the Tzu Chi Foundation in supporting global volunteers working to provide relief during the current pandemic. Materials and Methods: In this study, nucleotide sequences derived from publicly available viral genome data for several domains of the SARS-CoV2 spike and nucleocapsid (N) proteins were chemically synthesized, with codon optimization for Escherichia coli protein expression. No actual viral particles were involved in these experiments. The synthesized sequences were cloned into an E. coli expression system based on pQE80L, and expressed viral proteins were subsequently purified using Ni-affinity chromatography. Western blotting was conducted using human antiviral sera to assess the response of codon-modified viral proteins to COVID-19 patient sera. Results: N protein was expressed in amounts large enough to support large-scale production. The N-terminal domain, receptor-binding domain (RBD), Region 3, and the S2 domain were expressed in small but sufficient amounts for experiments. Immunoblotting results indicated that anti-N IgG and anti-N IgM antibodies were detected in most patient sera, but only 60% of samples reacted with the recombinant N protein, RBD, and S2 protein demonstrating the most potential.

Keywords: Codon optimization, COVID-19, Escherichia coli expression system, N protein, SARS-CoV2 antisera

*Address for correspondence: Prof. Guang-Huey Lin, Master Program in Microbiology and Immunology, School of Medicine, Tzu Chi University, 701, Section 3, Chung-Yang Road, Hualien, Taiwan. E-mail: veronica@gms.tcu.edu.tw

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKL RPMedknow_reprints@wolterskluwer.com

How to cite this article: Toh YH, HuangYW, Chang YC, Chen YT, Hsu YT, Lin GH. Reactivity of human antisera to codon optimized SARS-Cov2 viral proteins expressed in Escherichia coli. Tzu Chi Med J 2021; 33(2): 146-53.
key functional role in ACE2 binding and is also important for neutralizing host antibody production as well. The N protein, which binds the RNA of the viral genome, is the most abundantly produced viral protein and is highly immunogenic during viral infection [2].

*E. coli* expression systems are distinguished for their ease of use and convenience, but have a key limitation in that posttranslational modification is absent, and this can impact the expression of proteins that require posttranslational modification. Certainly, protein expression systems that can perform posttranslational modification have been established in different eukaryotic cells, including *Saccharomyces cerevisiae* [4,5]. Chinese hamster ovary cells [6], and baculovirus expression systems [7]. These systems have an advantage in that protein phosphorylation and glycosylation are achievable, but *E. coli* expression systems are superior in terms of the ability to rapidly produce large amounts of viral proteins that can serve as antigens for diagnostic tests.

Comparative analysis of 13 different coronaviruses regarding preferred nucleotides, codon bias, nucleotide changes at the third position (NT3s), synonymous codon usage, and relative synonymous codon usage has revealed that 18 amino acids have preferred codons, of which eight of these were over-biased [8]. Since different organisms have their own bias regarding usage of the 61 available codons, when *E. coli* systems are forced to express proteins with “rare” codons under heterologous gene expression conditions, this can result in low expression levels or truncated protein products due to premature termination of protein translation. This issue of codon bias can be overcome through two possible approaches: the first involves cotransformation with a plasmid that encodes the transfer RNAs corresponding to the “rare” codons, while the second requires chemical synthesis of genetic sequences in which the majority of “rare” codons have been replaced with *E. coli* preferred codons [9].

SARS-CoV2 viral sequences have been published by research teams from several countries [10-12], and as of July 18, 2020, research teams from Taiwan alone have published 32 SARS-CoV2 genome sequences. In this study, we referenced the N protein and S protein coding sequences from this publicly available viral genomic data and proceeded to modify the codons based on *E. coli* preference. The original coding sequence for the SARS-CoV2 N protein and the codon-modified coding sequence of the S protein were chemically synthesized by Genomics (New Taipei City, Taiwan) and then cloned into an *E. coli* expression system. *E. coli* with modified N protein coding sequences were obtained from pET15b-NE cloned by the Academia Sinica. The viral protein production capacity of cloned *E. coli* expression systems was subsequently tested, and the reactivity of these purified viral proteins with patient sera was also assessed. The results indicated that *E. coli* expression systems were capable of expressing sufficient levels of viral proteins that could be used to develop rapid and accurate diagnostic tests.

### Materials and Methods

#### Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are described in [Table 1].

#### Primers

The primers used in this study are listed in [Table 2].

#### Polymerase chain reaction and expression clone construction

DH5α/pUC57-SE [Table 1] served as the template for the amplification of codon-modified S protein domain DNA fragments by colony polymerase chain reaction (PCR). DH5α/pET15b-NE [Table 1] served as the template for amplification of codon-modified N protein coding sequences. The PCR products of different coding sequences were generated with equal amounts of *Pfu* and *Taq* DNA polymerase to reduce mutation rates following amplification. The PCR products of each DNA fragment were cloned into the respective restriction enzyme sites in pQE80L (Qiagen, Hilden, Germany) to generate the expression plasmids, pNE, pSE, pSE-NTD, pSE-RBD, pSE-R3, and pSE-S2 [Table 2]. After transformation, potential clones were evaluated with a previously described rapid screening method [13], followed by sequencing analysis (Genomics).

#### Recombinant protein expression and purification

To express viral proteins, transformants of each respective recombinant plasmid were expanded at 37°C by inoculating a 0.5-mL overnight culture into 50 mL of LB medium containing ampicillin. Incubation was continued at 37°C until the culture reached an OD<sub>600</sub> of 0.6. Protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM or 1 mM. After incubating for 3 h at 37°C, cells were harvested by centrifugation at 4,000 × g for 15 min. The cells were then stored at −80°C until use.

Protein purification was performed by a method described previously [14]. Frozen cells overexpressing target proteins were suspended in 10 mL of 1× binding buffer containing 5 mM imidazole, 0.5 mM NaCl, and 20 mM Tris-HCl (pH 8.0), and subjected to 3 cycles of freezing and thawing at −80°C and room temperature, respectively. The thawed cells were disrupted by sonication at 4°C using a Microson ultrasonic cell disruptor (Misonix, Farmingdale, NY, USA). Cell extracts were separated from the cell debris by centrifugation at 17,000× g for 30 min at 4°C (Avanti J-25 Centrifuge, JA25.5 rotor; Beckman Coulter, Brea CA, USA). Recombinant proteins were then purified from the cell extract using Ni-affinity chromatography (Novagen, Madison WI, USA). Purified fractions were analyzed with different concentrations of sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), based on the different molecular weights of each recombinant protein, and stained with Coomassie Brilliant Blue G-250.

#### Immunoblotting

Total proteins from bacteria cultured under different conditions were collected and separated by SDS-PAGE,
and then transferred to a polyvinylidene difluoride (PVDF) membrane, after which specific antibodies were applied to detect protein expression. Human anti-SARS-CoV2 polyclonal antisera was obtained from Access Biologies, USA (Mitek, COVID-19 Panel, 2X Panel 1.1, PW: 56435). Human subjects including any gender with age ranged from 18 to 90 years old. The confirmed infection with COVID-19 determined by a positive Roche swab results. Sample collected <10 days and more than 30 days after positive results (https://accessbiologicals.com/news/covid-19/). Goat anti-human IgG Fe (HRP) ab97225 and goat anti-human IgM mu chain (HRP) ab97205 were obtained from Abcam (Cambridge, UK). Anti-His monoclonal antibody was obtained from Santa Cruz Biotechnology (Dallas TX, USA). After blocking the transferred membrane with 5% skim milk in 1×TTBS (0.25% Tween-20, 0.05 M Tris-HCl [pH 7.6], 0.9% NaCl), antiserum at 1:1000 dilution in 1×TTBS was added, followed by incubation at room temperature for 1
h. The membrane was washed six times for 10 min with 
×TTBS. Goat anti-human IgG Fe (HRP) ab97225 and Goat 
anti-human IgM mu chain (HRP) ab97205 was added at 
1:10,000 dilution in 1×TTBS, followed by incubation at room 
temperature for 1 h. The PVDF membrane was then washed 
five times with 1×TTBS, and then rinsed with 1×TBS (0.05 
M Tris-HCl [pH 7.6], 0.9% NaCl) before enhanced 
chemiluminescence (ECL) reagent (Pierce ECL Western 
Blotting Substrate, Thermo Fisher Scientific, Waltham 
MA, USA) was applied. Autoradiography was obtained by UVP 
BioSpectrum 810 (Thermo Fisher Scientific).

RESULTS

In silico analysis of SARS-CoV2 genetic organization

The SARS-CoV2 genome contains about 29,000–30,000 
bases, with four primary structural proteins [SEM, and N; 
Figure 1a]. These structural proteins are likely to be the 
main targets of immune responses, and we further broke 
down larger structural proteins into functional domains to 
better assess immunogenicity. Based on in silico functional 
domain analysis and a review of the literature, the S protein 
coding sequence (3,822 bases) was subdivided into the NTD, 
RBD, and S2 domain [Figure 1b]. Previous research has 
also shown that Region 3 (R3) of the RBD is functionally 
important for ACE2 receptor binding [15], and thus the R3 
domain (comprising 135 amino acids) was also examined in 
this study.

Codon optimization

Previous studies have shown that viral sequences may 
not be properly expressed in E. coli due to codon usage 
differences with eukaryotic cells [9]. To resolve this issue, 
we designed sequences for chemical synthesis that replaced 
the original codons with E. coli preferred codons that still 
code for the same amino acid. In silico analysis of the 
codon-optimized NTD coding sequence showed that it shared 
72% identity with the original viral sequence [Figure S1a], 
while the codon-optimized RBD, S2 domain, and N protein 
coding sequences respectively shared 71.1% [Figure S1b], 
72.9% [Figure S1c], and 72.5% [Figure S1d] identity with 
the original viral sequences. The final translated amino 
acid sequences of all codon-optimized coding sequences 
shared 100% identity with the original viral amino acid 
sequences [Figure S1e and f].

Primer design and cloning

In this study, the Qiagen pQE80L expression system 
was used to express codon-optimized viral proteins, which 
were then purified with Ni-affinity chromatography. For the 
amplication of coding sequences for the S protein and related 
domains, pUC57-SE was used as a template for colony PCR; 
while for the amplification of N protein coding sequences, 
PET15b-NE was used as the template. Primers were designed 
according to the restriction enzyme sites for cloning the 
codon-optimized coding sequences of viral proteins into 
pQE80L [Table 2], and PCR was performed with Pfu and Taq 
DNA polymerases to ensure that the correct DNA sequences 
were amplified. PCR conditions were set at the respective 
melting temperatures for the pair of primers corresponding to 
each cloned coding sequence.

The sizes of the DNA coding sequences for the 
NTD, RBD, R3 domain, S2 domain, and N protein 
are respectively 813 base pairs (bp), 819 bp, 405 bp, 
1,800 bp and 1,260 bp [Figure 2]. After restriction enzyme 
digestion and ligation with pQE80L, the resulting plasmids 
with insert were checked by electrophoresis following 
quick screening buffer treatment and colony PCR. Results 
showed that one transformant (lane 11) of the NTD coding 
sequence had the correct insertion size [Figure 3a], and this 
was subsequently confirmed by sequencing analysis. Three 
transformants carrying the correct insert size for the RBD 
coding sequence were identified [Figure 3b], and all were 
confirmed by sequencing analysis. Seven transformants 
carrying the correct insert size for the S2 domain coding 
sequence were identified [Figure 3c], of which three clones 
(lane 2, 5, and 7) were confirmed by sequencing 
analysis. Four transformants of the R3 domain coding 
sequence were identified via colony PCR [Figure 3e], 
of which three (lane 4, 6 and 11) were confirmed by 
sequencing analysis. For transformants of the original viral 
coding sequence of the N protein, only three transformants 
were identified out of more than 700 transformants

Figure 1: Schematic representation of the genome organization of SARS-CoV2 and functional domains of the S protein. (a) The single-strand RNA genome of SARS-CoV2 is mostly taken up by open reading frame 1a and open reading frame 1b, and the structural protein open reading frames that encode the spike (S), envelope (E), membrane (M) 
and nucleocapsid (N) proteins are clustered near the 3’ end. The structural proteins are common to all coronaviruses. (b) The S protein primarily consists of the S1 and S2 subunit (aa 671–1270). The S1 subunit can be further divided into an N-terminal domain (aa 27–297) and receptor binding domain (aa 310–583), and the latter includes the Region 3 domain (aa 391-525)
screened, and the insert sequences still differed from the original sequences, perhaps due to codon bias in *E. coli*. One transformant (lane 13) was obtained and confirmed by sequencing analysis for the codon-optimized N protein coding sequence [Figure 3d].

**Protein expression and purification**

Confirmed clones were induced by 1 mM IPTG for 3 h, and at least 100 mL of bacterial culture was collected for protein purification by Ni-affinity chromatography. Subsequent SDS-PAGE results showed that the NTD, with size of about 30.4 kiloDaltons (kDa), was found in the induction fraction [Figure 4a, lane 2] and purified in lower amounts [Figure 4a, lane 3]. RBD (with size of about 30.6 kDa) was also found in the induction fraction [Figure 4a, lane 4, 5] and was successfully purified [Figure 4a, lane 6]. R3 (with size of about 15.4 kDa) expression could be induced after treatment with 1 mM IPTG for 3 h [Figure 4a, lane 8] and was also successfully purified [Figure 4a, lane 9]. Expression of the largest domain of the S protein, S2, was induced after IPTG treatment [Figure 4b, lane 2], and was successfully purified [Figure 4b, lane 3] even though the domain contains at least two transmembrane sections. Interestingly, despite the many difficulties encountered in cloning and deriving transformants with the N protein coding sequence, transformants with the codon-optimized N protein coding sequence demonstrated abundant expression after IPTG induction, with N protein representing about 30% of total protein expressed [Figure 4c, lane 2], and which was successfully purified [Figure 4c, lane 3]. Previously, a team at the Academia Sinica in Taiwan reported that they used the pET15b plasmid of the pET system (Novagen) to express N protein under induction conditions at 16°C for 10–12 h (personal communication); however, our results show that it is possible to obtain codon-optimized N protein using the pQE80L system (Qiagen) following induction at 37°C for 3 h, suggesting that expression efficiency can be enhanced through codon optimization.

**Patient serum response to N protein and receptor-binding domain**

A total of 0.5 µg of purified N protein, 1.5 µg of purified RBD, and 8.5 µg of purified S2 domain was transferred to a PVDF membrane, which was divided into 18 strips for immunoblotting. With anti-human IgG used as secondary anti-serum, the results demonstrated that all tested serum samples responded to N protein, except those for patients 9011 [Figure 5a, lane 8] and 9013 [Figure 5a, lane 10]. Anti-N protein IgM was not found in the serum of patients 9011 [Figure 5b, lane 8], 9013 [Figure 5b, lane 10], and 9014 [Figure 5b, lane 11]. These results show that more than 80% of patient sera responded to codon-optimized N protein produced in *E. coli*, suggesting that this may serve...
as a suitable antigen for rapid antibody detection tests. For RBD serum tests, anti-RBD IgG was detected in only seven samples out of 15, and anti-RBD IgM was detected in nine samples [Figure 6]. Interestingly, serum samples from patients 9011 and 9013 did not respond to N protein, but contained anti-RBD IgM [Table 3]. For S2 serum tests, 60% of samples contained anti-S2 IgG, but only patient 9099 had a high concentration of anti-S2 IgM [Figure 7]. Taken together, these results suggest that N protein and RBD would constitute a better combination for the construction of rapid antibody detection kit with greater accuracy and efficiency.

**DISCUSSION**

To date, many COVID-19 diagnostic kits have been developed for commercial production in other countries around the world, but as there is no commercial benefit for diagnostic kit production in Taiwan due to the low number of cases and excellent control of the pandemic, very few local companies have devoted resources to the development of such kits. However, since the start of the pandemic, the Tzu Chi Foundation has provided financial and healthcare support for many people in different countries, and the development of rapid antibody-based testing kits would not only help to protect Foundation volunteers working to provide relief, but could also be used in areas that lack quantitative reverse transcription PCR facilities to quickly identify the infected and enable the timely quarantine of patients.

The use of an *E. coli* expression system can enable the rapid production of SARS-CoV2 viral proteins in sufficient quantities to support the development of testing kits that can be used by the Tzu Chi Foundation for nonprofit purposes.

Although RT-PCR detection of viral RNA remains the most accurate and sensitive testing method currently available, not all
areas have access to the advanced facilities and equipment required to conduct such tests, and rapid antibody or viral protein detection kits would be a better choice in such cases. Previous research has shown that 100% of patients tested positive for antiviral IgG within 19 days, in a study that used magnetic chemiluminescence enzyme immunoassay (MCLIA) for viral-specific antibody detection in 285 patients with COVID-19 [16]. Serological detection will also be very helpful for the diagnosis of putative patients with negative RT-PCR results and for identifying asymptomatic infections and recovered infections. According to our immunoblotting results, codon-optimized N protein demonstrated an 80% positive response in serum samples against which it was tested, and the RBD of the S protein also exhibited a positive response in 50% of tested serum, indicating that these viral proteins produced in E. coli expression systems have strong potential to be used in serological testing kits.

To obtain large amounts of proteins in a short timescale, E. coli expression systems have a strong advantage over other expression systems. Cloning and validation can be completed within two weeks for an E. coli expression system [17]. However, the absence of posttranslational modification represents a major advantage and defect of E. coli expression systems [18-20], and codon bias may lead to truncated products because of premature termination of protein translation [21,22]. To overcome this issue, we prepared chemically synthesized viral protein coding sequences in which E. coli preferred codons were used. The results showed that these codon-optimized viral proteins were successfully produced and were capable of reacting with patient serum samples.

Previous studies have shown that the N protein of most coronaviruses is expressed abundantly during infection and is highly immunogenic [23,24]. High levels of IgG antiserum against N protein have been detected in patients with SARS infection [25], and SARS-specific T-cell proliferation and cytotoxic activity triggered by the N protein were typical T-cell responses in a vaccine developed for SARS [26]. IgM antiserum against the N protein of SARS-CoV2 was detected in 80% of serum samples in this study [Table 3], and most previous studies agreed with our observation that the N protein.

| Lane Numbers | RBD (IgG) | RBD (IgM) | S2 (IgG) | S2 (IgM) | N (IgG) | N (IgM) |
|--------------|-----------|-----------|----------|----------|---------|---------|
| 1            | +         | +         | +        | +        | +       | +       |
| 2            | +         | +         | +        | +        | +       | +       |
| 3            | +         | +         | +        | +        | +       | +       |
| 4            | +         | +         | ++       | +        | +       | +       |
| 5            | +         | +         | +        | +        | +       | +       |
| 6            | +         | +         | +        | +        | +       | +       |
| 7            | +         | +         | ++       | +        | +       | +       |
| 8            | +         | +         | +        | +        | +       | +       |
| 9            | +         | +         | +        | +        | +       | +       |
| 10           | +         | +         | +        | +        | +       | +       |
| 11           | +         | +         | +        | +        | +       | +       |
| 12           | ++        | +         | ++       | +        | +       | +       |
| 13           | +         | +         | ++       | +        | +       | +       |
| 14           | +         | +         | +        | +        | +       | +       |
| 15           | +         | +         | ++       | +        | +       | +       |
| 16           | +         | +         | +        | +        | +       | +       |
| 17           | ND        | ND        | ND       | ND       | ND      | ND      |
| 18           | ND        | ND        | ND       | ND       | ND      | ND      |

Percentage: 46.6%, 66.6%, 60%, 6%, 86.6%, 80%

+: interaction with antibodies, ++: Intense interaction with antibodies. NC: negative control anti-serum. ND: Not determined.
protein represents a good candidate for antibody detection [8]. Indeed, a previous study has used an N-terminal truncated N protein to produce specific antibodies against SARS N protein with no cross-reactivity with human coronaviruses OC43 and 229E [27]. In the future, we will also seek to generate N-terminal truncated N protein to reduce the possibility of cross-reactivity and improve the accuracy of antibody detection kits.

**Conclusion**

In conclusion, here we describe an *E. coli* expression system capable of rapidly and abundantly producing SARS-CoV2 viral proteins from chemically synthesized and codon-optimized sequences. Our results showed that the resulting codon-optimized N protein, RBD, and S2 domain were capable of reacting with patient sera, and have strong potential to be used in rapid and accurate serological tests for COVID-19.

**Acknowledgements**

The authors would like to thank Dr. Shun-Ping Huang for providing Human anti-SARS-CoV2 polyclonal antiserum was obtained from Access Biologicals (Mitek, COVID-19 Panel, 2X Panel 1.1, PW: 56435) and Professor Ingrid Y Liu for valuable discussions.

**Financial support and sponsorship**

This study was funded by Grant No. TCIRP109001-02 from Tzu Chi University to Guang-Huey Lin. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

**Conflicts of interest**

There are no conflicts of interest.

**REFERENCES**

1. Song Z, Xu Y, Bao L, Zhang L, Yu P, Qu Y, et al. From SARS to MERS, thrusting coronaviruses into the spotlight. Viruses 2019;11:59.
2. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 2020;583:459-68.
3. 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.
4. Nguyen NL, Kim JM, Park JA, Park SM, Jang YS, Yang MS, et al. Expression and purification of an immunogenic dengue virus epitope using a synthetic consensus sequence of envelope domain III and *Saccharomyces cerevisiae*. Protein Expr Purif 2013;88:235-42.
5. Nguyen NL, So KK, Kim JM, Kim SH, Jang YS, Yang MS, et al. Expression and characterization of an M cell-specific ligand-fused dengue virus tetravalent epitope using *Saccharomyces cerevisiae*. J Biosci Bioeng 2015;119:19-27.
6. Chaudhary G, Luo R, George M, Tescione L, Khetan A, Lin H. Understanding the effect of high gas entrance velocity on Chinese hamster ovary (CHO) cell culture performance and its implications on bioreactor scale-up and sparger design. Biotechnol Bioeng 2020;117:1684-95.
7. Tao Y, Li G, Zheng W, Shu J, Chen J, Yang F, et al. Development of a combined genetic engineering vaccine for porcine circovirus type 2 and “Mycoplasma hypopneumoniae” by a baculovirus expression system. Int J Mol Sci 2019;20: 4425-40.
8. Sheikh A, Al-Taher A, Al-Nazawi M, Al-Mubarak AI, Kandeel M. Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. J Virol Methods 2020;277:113806.
9. Kaur J, Kumar A, Kaur J. Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. Int J Biol Macromol 2018;106:803-22.
10. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, et al. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe 2020;27:325-8.
11. Saha S, Malaker R, Sajib MSI, Hasanuzzaman M, Rahman H, Ahmed ZB, et al. Complete genome sequence of a novel coronavirus (SARS-CoV-2) isolate from Bangladesh. Microbiol Resour Announc 2020;9: e00568-20.
12. Paraskeviss D, Kostaki EG, Magiorkini G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel coronavirus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol 2020;79:104212.
13. Kado CI, Liu ST. Rapid procedure for detection and isolation of large and small plasmids. J Bacteriol 1981;145:1365-73.
14. Lin HR, Shu HY, Lin GH. Biological roles of indole-3-acetic acid in *Acinetobacter baumannii*. Microbiol Res 2018;216:30-9.
15. Ibrahim IM, Abdelmalek DH, Elshahat ME, Elflyi AA. COVID-19 spike-host cell receptor GRP78 binding site prediction. J Infect 2020;80:554-62.
16. Long QX, Tang XJ, Shi QL, Li Q, Deng HJ, Yuan J, et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat Med 2020;26:1200-4.
17. Chuan YP, Lua LH, Middelberg AP. High-level expression of soluble viral structural protein in *Escherichia coli*. J Biotechnol 2008;134:64-71.
18. Ahmad I, Nawaz N, Darwesh NM, Ur Rahman S, Mustafa MZ, Khan SB, et al. Overcoming challenges for amplified expression of recombinant proteins using *Escherichia coli*. Protein Expr Purif 2018;144:12-8.
19. Hayat SMG, Farahani N, Golichenari B, Sahelkhar A. Recombinant protein expression in *Escherichia coli* (*E. coli*): What we need to know. Curr Pharm Des 2018;24:718-25.
20. Smolskaya S, Logashina YA, Andreev YA. *Escherichia coli* extract-based cell-free free expression system as an alternative for difficult-to-obtain protein biosynthesis. Int J Mol Sci 2020;21:928-49.
21. Ejdebäck M, Young S, Samuelsson A, Karlsson BG. Effects of codon usage and vector-host combinations on the expression of spinach plastocyanin in *Escherichia coli*. Protein Expr Purif 1997;11:17-25.
22. Chen GF, Inouye M. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. Nucleic Acids Res 1990;18:1465-73.
23. Pan Y, Li X, Yang G, Fan J, Tang Y, Zhao J, et al. Serological immunochromatographic approach in diagnosis with SARS-CoV-2 infected COVID-19 patients. J Infect 2020;81:e28-32.
24. Cong Y, Ulasli M, Schepers H, Mauthe M, V’kovski P, Kriegenburg F, et al. Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. J Virol 2020;94: e01925-19.
25. Leung DT, Tam FC, Ma CH, Chan PK, Cheung JL, Niu H, et al. Antibody response of patients with severe acute respiratory syndrome (SARS) targets the viral nucleocapsid. J Infect Dis 2004;190:379-86.
26. Gao W, Tamin A, Soloff A, D’Aiuto L, Nwangebo E, Robbins PD, et al. Effects of a SARS-associated coronavirus vaccine in monkeys. Lancet 2003;362:1895-6.
27. Lee HK, Lee BH, Seok SH, Baek MW, Lee HY, Kim DJ, et al. Production of specific antibodies against SARS-coronavirus nucleocapsid protein without cross reactivity with human coronaviruses 229E and OC43. J Vet Sci 2010;11:165-7.
Figure S1: (a) Nucleic acid alignment of NTD of Spike protein encoding genes. Identity between NTD-SO and NTD-SE is 72%. O is original sequence. E is E. coli modified sequences. (b) Nucleic acid alignment of RBD of Spike protein encoding genes. Identity between RBD-SO and RBD-SE is 71.1%. O is original sequence. E is E. coli modified sequences. (c) Nucleic acid alignment of S2 of Spike protein encoding genes. Identity between S2-SO and S2-SE is 72.9%. O is original sequence. E is E. coli modified sequences. (d) Nucleic acid alignment of Nucleocapsid protein encoding genes. Identity between NO and NE is 72.5%. O is original sequence. E is E. coli modified sequences. (e) Amino acid sequences alignment of nucleocapsid protein encoding genes. Identity between NO and NE is 100%. O is original sequence. E is E. coli modified sequences. (f) Amino acid sequences alignment of spike protein encoding genes. Identity between SO and SE is 100%. O is original sequence. E is E. coli modified sequences.