

## EXPERIMENTAL WORKS

UDC 577.112.6:578.834.1

doi: <https://doi.org/10.15407/ubj93.06.031>

### PRODUCTION OF RECOMBINANT SARS-COV-2 PROTEINS AND DIPHTHERIA TOXOID CRM197-BASED FUSION

O. I. KRYNINA<sup>1</sup>, S. I. ROMANIUK<sup>1</sup>, O. B. GORBATIUK<sup>1,2</sup>,  
O. H. KORCHYNSKYI<sup>1,3,4</sup>, A. V. REBRIIEV<sup>1</sup>, Ya. S. KULYK<sup>1</sup>,  
Ye. O. KOZADAIEVA<sup>1</sup>, A. A. SIROMOLOT<sup>1,5</sup>, M. M. GUZYK<sup>1</sup>,  
D. V. KOLYBO<sup>1</sup>✉, S. V. KOMISARENKO<sup>1</sup>

<sup>1</sup>Palladin Institute of Biochemistry, National Academy of Sciences of Ukraine, Kyiv;

<sup>2</sup>State Institute of Genetic and Regenerative Medicine, National Academy of Medical Sciences of Ukraine, Kyiv;

<sup>3</sup>Centre for Innovative Research in Medical and Natural Sciences, Faculty of Medicine, University of Rzeszow, Rzeszow, Poland;

<sup>4</sup>S. Gzhytskyi National University of Veterinary Medicine and Biotechnologies, Lviv, Ukraine;

<sup>5</sup>ESC "Institute of Biology and Medicine", Taras Shevchenko National University of Kyiv, Ukraine;

✉e-mail: kolibo@biochem.kiev.ua

**Received:** 10 October 2021; **Accepted:** 12 November 2021

The quickly emerged global COVID-19 pandemic raised a desperate need in the development of protecting vaccines targeting this disease. Therefore, a generation of effective producers of recombinant SARS-CoV-2 proteins became an urgent task. Its resolving contributes to the study of functional SARS-CoV-2 properties, as well as will allow developing the domestic COVID-19 vaccine in Ukraine, thus playing an important strategic role in tackling the pandemics. The aim of the study was to generate prokaryotic and eukaryotic producers of recombinant SARS-CoV-2 proteins and to isolate nucleocapsid (N) protein, receptor-binding domain (RBD) of spike (S) protein, as well as RBD fused to the carrier – diphtheria toxoid CRM197. For this purpose, appropriate genetic constructs, in particular, replication deficient recombinant AdvC5-based adenoviral vectors expressing the SARS-CoV-2 proteins and CRM197-fused conjugate were created through methods of molecular biology and genetic engineering. Restriction analysis and/or DNA sequencing confirmed that we created the correct constructs. Immobilized metal affinity chromatography was used to purify the recombinant proteins. Compliance of their properties was confirmed by the results from polyacrylamide gel electrophoresis, Western blotting, immunoenzymatic assay and MALDI-TOF mass spectrometry. As a result, we generated *E. coli* Rosetta (DE3) bacterial strain and HEK293 cell line producing recombinant SARS-CoV-2 proteins and CRM197-based fusion. In addition, pure N protein, RBD of S protein and RBD-CRM197 fusion protein were isolated. The obtained recombinant SARS-CoV-2 proteins can be used to study immunogenic and antigenic properties of the SARS-CoV-2 proteins. Cells producing recombinant SARS-CoV-2 proteins and RBD-CRM197 fusion protein are able to provide cheap and safe synthesis of the antigenic substances for domestic development and production of immunodiagnostics for COVID-19 and COVID-19 vaccines in Ukraine.

**Key words:** SARS-CoV-2, recombinant proteins, fusion protein, CRM197, immunodiagnostics, vaccine, COVID-19.

In 2020, Ukraine faced the biggest challenge to its biosafety and health care system – the pandemic of the dangerous SARS-CoV-2 virus, which has caused the acute respiratory disease COVID-19 [1]. The number of recorded deaths from COVID-19 worldwide has hit five million, less than two years since the pandemic began. Almost 80,000 Ukrainians have fallen victim to the disease [2]. The world scientific community has focused efforts on the SARS-CoV-2 study to find approaches to overcome the pandemic. However, no effective specific antiviral drug has been found to reduce mortality. Therefore, clinicians have high hopes above all for COVID-19 vaccines.

Now 323 foreign companies are developing COVID-19 vaccines, at that 129 companies are in clinical-development phases, and 35% of them are testing vaccines based on recombinant SARS-CoV-2 proteins [3]. Several pharmaceutical companies had successfully completed clinical trials and released to the market the COVID-19 vaccines based on mRNA, adenoviral vectors or inactivated SARS-CoV-2 virus particles. The first to enter the European market was the COVID-19 vaccine produced by Pfizer-BioNTech (USA/Germany) and approved for use in the United Kingdom on December 2, 2020 [4]. Three more COVID-19 vaccines: from Moderna (USA), Oxford-AstraZeneca (UK/Sweden) and Johnson & Johnson (USA/Belgium) were followed by it in 2021. These vaccines are based on adenoviruses or mRNA. Chinese vaccines (from CanSino, Sinopharm, Sinovac, Sinopharm-Wuhan), Russian vaccines (from Gamaleya, Vector Institute), and Indian vaccines (from Bharat Biotech) have been approved for use in some countries. The first COVID-19 vaccine based on the recombinant SARS-CoV-2 proteins from Novavax (USA) is currently receiving approval to use [5]. This vaccine is based on the recombinant SARS-CoV-2 spike protein nanoparticles with saponin adjuvant, and according to preliminary data, it showed an effectiveness of 96% against SARS-CoV-2 [6].

Spike (S) protein and especially receptor-binding domain (RBD) of S protein are considered the main proteins used as a target in COVID-19 vaccines, because antibodies against these proteins are able to block the interaction between RBD and ACE2 (angiotensin-converting enzyme 2) cell surface protein [7] and to prevent viral infection of cells. A promising target of COVID-19 vaccine also may be nucleocapsid (N) protein – the main struc-

tural component of SARS-CoV-2 virion, which protects the viral RNA and is able to induce an intense immune response [8].

Recombinant protein-based vaccines are known to be unable to cause a strong immune response, so they require additional immunostimulatory molecules and multiple immunizations. We propose to use highly immunogenic carriers – derivatives of diphtheria toxin, in particular diphtheria toxoid CRM197, to enhance the immunogenicity of the recombinant SARS-CoV-2 proteins. CRM197 is a non-toxic diphtheria toxin mutant protein that contains a single amino acid substitution (G52E) [9]. It is widely used as a carrier in conjugated vaccines against meningococcal, pneumococcal, and hemophilic infections [10]. In these vaccines, low-immunogenic bacterial oligosaccharides are chemically conjugated to the surface of the carrier molecule. CRM197 can also be used to enhance the immune response to peptide antigens attached to its molecule by chemical methods [11].

Various scientific approaches have been used to develop COVID-19 vaccines, but it is unknown exactly what approach will be the most effective. The first pharmaceutical corporations succeeded in the developing COVID-19 vaccines have captured a huge market; they are at an advantage and can dictate prices when demand is high. Launched vaccination campaigns, first of all, cover the population of countries producing COVID-19 vaccines. There are humanitarian programs to provide developing countries, including Ukraine, with COVID-19 vaccines, but it is unknown, how regular the supplies of these vaccines will be in the future. Immunity against coronaviruses is not long-lasting, and the demand for COVID-19 vaccines is unlikely to disappear over time. Therefore, an important task is to develop an effective domestic vaccine to combat SARS-CoV-2 in Ukraine.

The availability of cells producing the recombinant SARS-CoV-2 proteins can significantly accelerate the emergence of the domestic COVID-19 vaccine in Ukraine that is of great strategic importance. Therefore, the aim of the study was to generate prokaryotic and eukaryotic producers of the recombinant SARS-CoV-2 proteins and to isolate nucleocapsid (N) protein, receptor-binding domain (RBD) of spike (S) protein and conjugate based on the RBD fused to the diphtheria toxoid CRM197.

## Materials and Methods

**Isolation of plasmid DNA.** Plasmid DNA was isolated using the alkaline lysis procedure [12]. Additional purification and concentration of plasmid DNA were performed using DNA Extraction Kit (Thermo Fisher Scientific, Waltham, MA, USA).

**PCR and DNA electrophoresis.** PCR Thermal Cycler 2720 (Applied Biosystems, Waltham, MA, USA) was used for DNA amplification. PrimeSTAR DNA polymerase (Takara Bio, Kusatsu, Shiga, Japan) was used to amplify the insert for cloning, and DreamTaq polymerase (Thermo Fisher Scientific, Waltham, MA, USA) was used for PCR screening of bacterial colonies.

DNA electrophoresis was performed in 1% or 3% agarose gel in TAE buffer solution (0.4 M Tris-acetate, pH 8.3; 0.01 M EDTA). Samples were dissolved in 6X Loading Buffer (Thermo Fisher Scientific, Waltham, MA, USA). After the run, gels were stained with 0.005% ethidium bromide solution (Sigma-Aldrich, St. Louis, MO, USA) for 5 min and washed with distilled water. Transilluminator Vilber Lourmat TFX-20.LM and gel documentation system Vilber Lourmat DP-001.FDC (Vilber Lourmat, Collegien, France) has been used for DNA visualization in the gel.

**Creation of genetic constructs containing SARS-CoV-2 genes.** Genetic constructs for expression of recombinant SARS-CoV-2 N, RBD proteins and the protein chimera RBD-CRM197 in bacterial hosts were created on the base of *pET-24a(+)* and

*pET-28a(+)* vectors (Merck, Darmstadt, Germany) using *2019-nCoV\_N\_Positive Control* (Integrated DNA Technologies, Coralville, IA, USA), *pcDNA3-SARS-CoV-2-S-RBD-8xHis* (briefly – *pcDNA3-RBD-8His*) (Addgene, Watertown, MA, USA), as well as *pET24a-CRM197* vector, which has been obtained earlier [13]. The *pET-24a(+)* vector was modified by introducing the enterokinase cleavage site before polyhistidine tag (His-tag) using NotI and XhoI restriction enzyme sites resulting in the *pET24a-entero* expression vector.

Genetic constructs for expression of recombinant RBD domain of S protein and RBD-CRM197 fusion protein in the eukaryotic hosts were created on the base of *pcDNA3* and *pShuttle-CMV* vectors, as well as adenoviral *pEasy1-CMV* vector using plasmids mentioned above and created genetic constructs for expression in prokaryotic cells.

The cDNA of SARS-CoV-2 N and RBD genes were amplified by PCR using specific primer pairs (shown in Table 1), *2019-nCoV\_N\_Positive Control* and *pcDNA3-RBD-8His* vectors as templates.

PCR-mediated overlap extension (PCR-OE) was used to obtain the DNA sequence of RBD-CRM197 fusion protein [14]. To do this, when performing PCR1, modified DNA sequences were amplified using specific primers listed in Table 2.

*pcDNA3-RBD-8His* and *pET24a-CRM197* expression vectors were used as templates for amplification of RBD-CRM197 DNA sequence fragments. When PCR2 was performed using PCR1 products

*Table 1. Nucleotide sequences of primers used to amplify the DNA sequences of SARS-CoV-2 N and RBD proteins*

RBD sense (NdeI)	gatcatatgaagaccatcatcgcc
RBD antisense (NotI)	caagcggccgcacttttgggcccaaac
N-prot sense (NheI)	catgctagctctgataatggacccaaa
N-prot antisense	gtagcggccgcggcctgagttgagtcag
pET antisense	gctagttattgctcagcgg

*Table 2. Nucleotide sequences of primers used to construct the DNA sequence of RBD-CRM197*

RBD sense (NdeI)	gatcatatgaagaccatcatcgcc
RBD antisense fusion	aggtcttctgcaagaccctgatcccttttgggcc
CRM197 sense fusion	gtcttcgcagaagacctggctgccggcgtgatgatgtt
CRM197 antisense fusion	gtggcggccgcagctcttgattcaaaaaatag
pET antisense	gctagttattgctcagcgg

and flanking primers, the complete RBD-CRM197 DNA sequence was synthesized.

The obtained nucleotide sequences of N, RBD and RBD-CRM197 proteins, as well as vectors for insertion were treated with certain restriction enzymes (NdeI, NheI, NotI, BamHI, XhoI, XbaI) and, in some cases, Klenow fragment (Thermo Fisher Scientific, Waltham, MA, USA). After separation of the hydrolyzed DNA in 1% agarose gel, the required fragments were isolated using DNA Extraction Kit (Thermo Fisher Scientific, USA). Joining of DNA inserts to the plasmid vectors was performed using T4 DNA ligase (Thermo Fisher Scientific, Waltham, MA, USA). Endo-Free Maxi-prep kit (Qiagen, Hilden, Germany) was used to purify recombinant adenoviruses. The main stages of different genetic constructs creation are considered in detail in the "Results and Discussion" section.

*Cultivation of eukaryotic and prokaryotic producers of recombinant proteins.* HEK293 cells were obtained from the Bank of cell lines of R.E. Kavetsky Institute of Experimental Pathology, Oncology and Radiobiology, NAS of Ukraine. Eukaryotic cells were cultured in nutrient medium RPMI-1640 (Sigma-Aldrich, St. Louis, MO, USA) with L-glutamine, containing 5% fetal bovine serum (Sigma-Aldrich, St. Louis, MO, USA), streptomycin (100 mg/l), penicillin (10,000 U) and amphotericin B (250 µg/l) in a humidified atmosphere of 5% CO<sub>2</sub> at 37°C (antibiotics are produced by Thermo Fisher Scientific, Waltham, MA, USA).

Cloning strain *E. coli* DH10B (Thermo Fisher Scientific, Waltham, MA, USA) and expression strain *E. coli* Rosetta (DE3) (Merck, Darmstadt, Germany) cells were cultured in LB Broth medium (Sigma-Aldrich, St. Louis, MO, USA), containing 50 µg/ml kanamycin (Arterium, Kyiv, Ukraine), to a density at which A<sub>600</sub> = 0.3-0.5.

Expression of recombinant proteins was induced by incubation for 3-4 h at 30°C with active aeration in the presence of 1 mM isopropyl-β-D-1-thiogalactopyranoside (Thermo Fisher Scientific, Waltham, MA, USA). Also, we used a modified autoinduction protocol, according to which the bacterial culture was grown overnight at 37°C in 2 ml of 2xYT medium (Sigma-Aldrich, St. Louis, MO, USA), containing 50 µg/ml kanamycin, 1% glucose (Arterium, Kyiv, Ukraine); and then the culture was grown for 18-24 h at 37°C in a dilution of 1:1000 in fresh 2xYT medium, containing 50 µg/ml kanamycin, 25 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 50 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM

Na<sub>2</sub>HPO<sub>4</sub>, 1mM MgSO<sub>4</sub>, 0.05% glucose, 0.2% α-lactose and 0.5% glycerol (AppliChem, Darmstadt, Germany).

*Isolation of recombinant proteins.* Bacterial cells were disrupted by 1 mg/ml lysozyme (Sigma-Aldrich, USA) in PBS, freezing/thawing procedures and action of Labsonic M ultrasonic homogenizer (Sartorius, Germany) for 3 min with amplitude of 220 µm and intensity of 600 W/cm<sup>2</sup>. Soluble and insoluble fractions were separated by centrifugation for 5 min at 13200 g. Isolation of recombinant proteins from insoluble fraction was performed according to the protocol of extraction from bacterial inclusion bodies [15].

Recombinant proteins were purified by immobilized metal affinity chromatography (IMAC) on 26/20 XK column (GE Healthcare, USA) with Ni<sup>2+</sup>-NTA Superflow agarose (Qiagen, Germany). The wash buffer (pH 8.0) contained 100 mM Na<sub>2</sub>HPO<sub>4</sub>, 10 mM Tris · Cl, 500 mM NaCl, and 8 M urea (Thermo Fisher Scientific, USA) with a gradual decrease in urea concentration. Target proteins were eluted with wash buffer containing 410 mM imidazole (AppliChem, Germany).

Protein N refolding was carried out by dialysis against PBS. The gradual dilution method [16] was performed for the renaturation of other recombinant proteins. The renaturing buffer contained oxidized and reduced glutathione (AppliChem, Germany) to ensure the formation of disulfide bonds, as well as 0.1 M L-arginine (AppliChem, Germany) and 0.2% Tween 20 (Sigma-Aldrich, USA) to prevent protein aggregation.

*Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).* Denaturing SDS-PAGE was performed according to the modified procedure of Schagger H. [17] with the addition of 12% SDS (Sigma-Aldrich, USA). The anode buffer contained 0.2 M Tris (pH 8.9), the cathode buffer contained 0.1 M Tricine (Sigma-Aldrich, USA), 0.1 M Tris and 0.1% SDS. Before gel loading, the samples were heated at 85°C for 5 min with sample buffer containing 6 M urea, 4% SDS, 20% glycerol, 5% β-mercaptoethanol, and bromophenol blue (Sigma-Aldrich,). Roti-Mark Tricolor protein marker (Carl Roth, Germany), PageRuler Prestained and Unstained Protein Ladders (Thermo Fisher Scientific, USA) were used to determine the molecular weight of recombinant proteins. Analysis of electrophoregrams was performed using TotalLab TL120 software (TotalLab Ltd, United Kingdom).

**Chemiluminescent Western blotting.** Wet transfer of proteins from polyacrylamide gel to nitrocellulose membrane (Merck Millipore, USA) was performed using Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, USA). The transfer buffer contained 25 mM Tris, 0.1% SDS, 20% methanol (Miranda C, Ukraine), 192 mM glycine (Sigma-Aldrich, USA). After transfer, the membrane was incubated for 2 h at 25°C in 5% skim milk (Sigma-Aldrich, USA) in PBS, and overnight at 4°C in PBS-T buffer (0.04% Tween-20 in PBS) with the addition of 1% skim milk and rabbit anti-RBD antibodies, which were kindly provided by Private Enterprise Geron (Ukraine). After washing with PBS-T, the membrane was incubated with goat anti-rabbit Ig horseradish polymer peroxidase conjugate (Sigma-Aldrich, USA) at dilution of 1:10,000 in PBS-T with the addition of 1% bovine serum albumin (Sigma-Aldrich, USA). Substrate solution containing 100 mM Tris-HCl, 0.025% luminol, 0.003% coumaric acid (Sigma-Aldrich, USA) and 0.035% H<sub>2</sub>O<sub>2</sub> was used to detect peroxidase. Luminescence image was captured on an X-ray film using developing and fixing solutions (Agfa-Gevaert, Belgium).

**MALDI-TOF mass spectrometry.** To remove coomassie brilliant blue, the protein-containing gel fragment was washed with a mixture of 200 mM ammonium bicarbonate and acetonitrile (Sigma-Aldrich, USA) in a ratio of 1:1 (by volume), dehydrated in acetonitrile and dried using Eppendorf Concentrator plus (Eppendorf, Germany). Protein hydrolysis was performed for 18 h at 37°C after adding 10 µl TPCk-treated bovine trypsin (Applied Biosystems, USA). Next, 50 µl of 0.1% trifluoroacetic acid (Sigma-Aldrich, USA) was added, and ultrasound irradiation was performed to extract tryptic peptides from the gel and concentrate the solution to a volume of 5-10 µl using a vacuum concentrator. Samples containing 0.8 µl tryptic peptide solution and 1 µl α-cyano-4-hydroxycinnamic acid (Sigma-Aldrich, USA) solution (1 mg/ml) in a mixture (1:1 by volume) of acetonitrile and 0.1% trifluoroacetic acid (Sigma-Aldrich, USA) applied to a plate, dried and tested using Voyager DE PRO MALDI-TOF spectrometer (Applied Biosystems, USA). Mass spectra analysis was performed using the Data Explorer Version 4.1 processing software (Applied Biosystems, USA). The amino acid sequence of tryptic peptides was determined using online Peptide Mass Calculator [18].

**Indirect solid-phase enzyme-linked immunosorbent assay (ELISA).** At each step of the assay,

solutions were added to a 96-well plate (Deltalab, Spain) at 100 µl/well, incubated for 1 h at 37°C and washed three times with PBS-T buffer. At first step, recombinant protein solutions (3 µg/ml) in carbonate-bicarbonate buffer (0.035 M NaHCO<sub>3</sub>, 0.015 M Na<sub>2</sub>CO<sub>3</sub>, pH 9.2) were added in plate wells. Then, 1% skim milk (powder) in PBS was used to block non-specific binding sites in the plate. Rabbit anti-spike RBD antibodies (Private Enterprise Geron, Ukraine) in PBS-T were used as primary antibodies; goat anti-rabbit Ig horseradish polymer peroxidase conjugate (dilution 1:10,000 in PBS-T) was used as secondary antibodies. Substrate solution contained 1 part of 1 mg/ml 3,3',5,5'-tetramethylbenzidine in DMSO (Sigma-Aldrich, USA), 9 parts of 0.05 M citrate-phosphate buffer (pH 5.0) and 0.007% H<sub>2</sub>O<sub>2</sub>. The plate was incubated with the substrate in the dark for 10-30 min at 37°C, and 1N H<sub>2</sub>SO<sub>4</sub> was added at 50 µl/well to stop the reaction. The absorbance of each well of the plate was determined at 450 nm using BioTek ELx800 microplate absorbance reader (BioTek Instruments, USA).

**Statistical processing of the obtained results.** Normality of data distribution was tested with Shapiro-Wilk test. The obtained data were presented as  $M \pm SE$ , where M is the average value, and SE is the standard error of the mean. Student's t-test was used for direct comparisons of means between groups:  $P \leq 0.05$  was considered statistically significant. Data Explorer 4.1 (Applied Biosystems, USA) and Origin 9.0 (OriginLab Corporation, USA) software were used for data analysis and presentation.

## Results and Discussion

**Creating the genetic constructs for expression in prokaryotic cells.** Genetic constructs for production of the recombinant SARS-CoV-2 proteins in prokaryotic cells were created on the base of the *pET24a-entero* vector with introduced enterokinase cleavage site before His-tag. This feature allows using His-tag at the C-terminus of a protein molecule for purification of recombinant proteins by immobilized metal affinity chromatography and then to remove His-tag from protein by site-specific proteolysis with enterokinase.

Three genetic constructs containing the nucleotide sequences of SARS-CoV-2 N, RBD proteins or RBD-CRM197 fusion protein were created on the base of the *pET24a-entero* expression vector. Schematic diagram of the overall organization of genetic constructs used in this study for expression in prokaryotic cells is shown in Fig. 1, A.

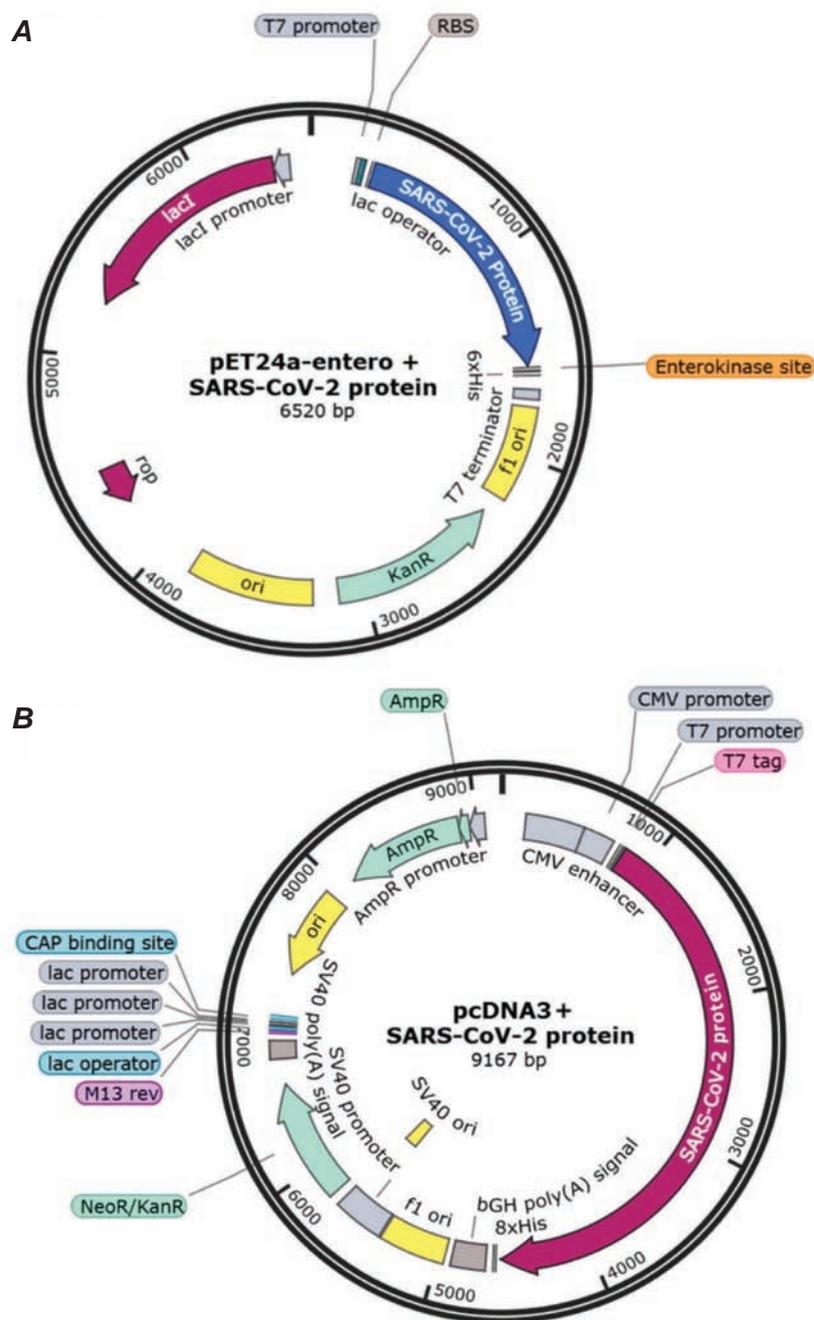


Fig. 1. Generalized schematic representation of genetic constructs based on pET24(a)-entero vector (A) and pcDNA3 vector (B) encoding SARS-CoV-2 proteins. Made using SnapGene software (GSL Biotech LLC, USA)

Combination in one open reading frame the DNA sequences of SARS-CoV-2 spike RBD and diphtheria toxoid CRM197 was performed using the PCR-mediated overlap extension (PCR-OE), the principle of which is schematically shown in Fig. 2, A. In the first stage, PCR1 was carried out using the primer pair (b and c) complementary to the RBD DNA sequence at the 3'-end and CRM197 DNA se-

quence at the 5'-end (Fig. 2, C). As a result, intermediate AB and CD PCR1 products that contained complementary to each other overlapping regions were obtained.

In the second stage, AB and CD products were applied as DNA templates in PCR2 using flanking primers a and d. As a result, RBD and CRM197 DNA sequences were combined into a single nucleo-

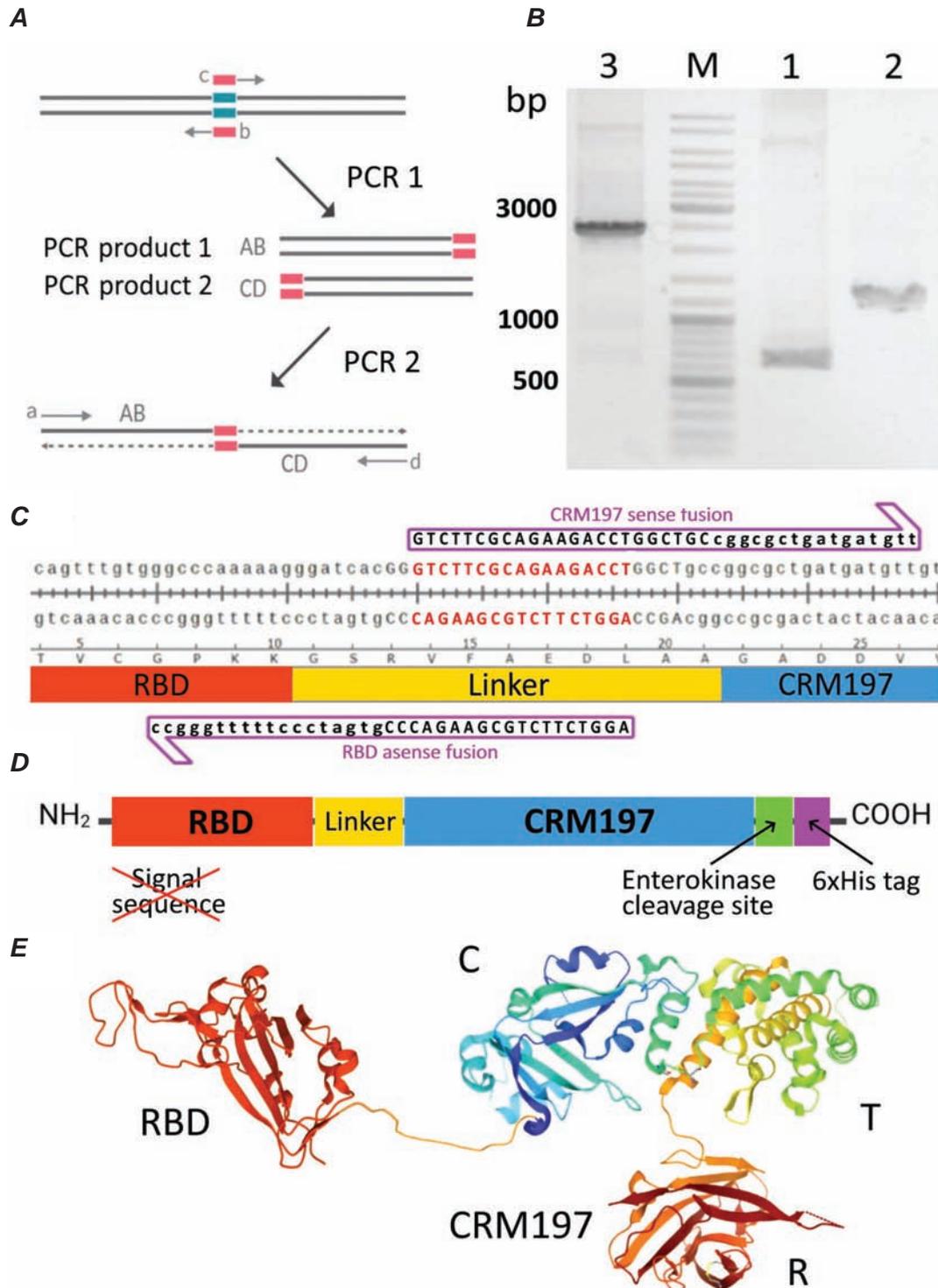


Fig. 2. Main stages of creating RBD-CRM197 fusion protein. (A) Scheme of PCR-mediated overlap extension (PCR-OE). (B) DNA electrophoregram of PCR products obtained to create RBD-CRM197: 1 and 2 – DNA fragments synthesized in stage I, 3 – final product, M – DNA markers. (C) Nucleotide sequences of primers and regions of RBD-CRM197 components joining through a linker. (D) Scheme of RBD-CRM197 fusion protein structure after design optimization. (E) Tertiary structure model of RBD-CRM197 chimera reproduced according to the crystallography of SARS-CoV-2 spike RBD (PDB 6M0J) and diphtheria toxin mutant CRM197 (PDB 4AE0) [19]. The picture shows that the CRM197 molecule consists of three domains: catalytic domain (C), transmembrane domain (T), and receptor-binding domain (R)

tide sequence due to overlapping regions (Fig. 2, B). Linker, encoded by the DNA sequence within the overlap, ensured free movement of the fusion protein components and preservation of their structure and function.

The amplified DNA sequences of N, RBD and RBD-CRM197 proteins were successively treated with restriction endonucleases: NdeI (RBD and RBD-CRM197 proteins), NheI (N protein), and NotI (all proteins). Prepared DNA inserts and pET24a-entero expression vector were isolated, joined and ligated using “sticky” ends, resulting in the corresponding genetic constructs pET24a-entero-N-protein, pET24a-entero-RBD and pET24a-entero-RBD-CRM197.

Subsequently, the genetic construct for the expression of RBD-CRM197 fusion protein in prokaryotic cells was modified. Firstly, fragment encoding the influenza virus hemagglutinin signal sequence (MKTIIALSYIFCLVFA) [20], which provides protein secretion directly to the extracellular medium, was removed from the RBD-CRM197 DNA sequence. Secondly, the DNA sequence of the linker sensitive to the cellular proteases was replaced with the DNA sequence of the more stable and flexible linker (Gly-Gly-Gly-Gly-Ser)<sub>3</sub> [21]. The changed RBD-CRM197 DNA sequence was combined with the prepared pET28a (+) expression vector resulting in the pET28-RBD-CRM197 genetic construct. Structure diagram of RBD-CRM197 fusion protein after optimization is presented in Fig. 2, D, and its spatial structure model – in Fig. 2, E.

Synthesis of genetic constructs for expression in eukaryotic cells. Recombinant SARS-CoV-2 proteins expressed in prokaryotic cells are not subject to post-translational modifications, in particular, glycosylation that may be one of the reasons for their structural and functional differences from natural SARS-CoV-2 proteins. Primarily, this applies to SARS-CoV-2 S protein, in which glycans shield approximately 40% of the protein surface that determine the availability for recognition of certain parts of a protein molecule by antibodies [22]. Therefore, two genetic constructs based on the pcDNA3 expression vector and containing RBD or RBD-CRM197 DNA sequences were used to obtain the RBD domain of S protein and RBD-CRM197 fusion protein in the eukaryotic expression system (generalized scheme of the overall organization of these genetic constructs is presented in Fig. 1, B).

Two approaches were tested to create pcDNA3-RBD-CRM197 genetic construct. Firstly, the RBD-

CRM197 DNA sequence was amplified in PCR using pET24a-entero-RBD-CRM197 as a template. The resulting product was treated with NotI restriction enzyme, pcDNA3-RBD-8His expression vector was treated with BamHI restriction enzyme. Then, the insert DNA and vector were sequentially treated with Klenow fragment and NheI restriction enzyme. After purification, the insert DNA was joined with prepared pcDNA3-8His vector and ligated using “sticky” ends. Secondly, CRM197 gene was excised by BamHI and XhoI enzymes from pET24a-CRM197 and cloned into pcDNA3-RBD-8His expression vector. Restriction analysis and DNA sequencing using a primer complementary to the DNA sequence within CMV promoter region confirmed that the resulting pcDNA3-RBD-CRM197 construct is correct.

In addition, genetic constructs based on pShuttle-CMV vector and adenoviral pEasy1-CMV vector was created for the expression of spike RBD and RBD-CRM197 fusion protein in eukaryotic cells. These genetic constructs can be used for the development of recombinant adenovirus type-5 (Ad5) vectored COVID-19 vaccine.

The RBD DNA was excised by NheI and XbaI restriction enzymes from pcDNA3-RBD-8His and cloned into pShuttle-CMV vector (linearized by XbaI restriction enzyme and dephosphorylated), resulting in the pShuttle-CMV-RBD-8His construct. Restriction analysis using enzymes XhoI and HindIII/EcoRV confirmed that the insert is present in this construct in the correct orientation (data not presented).

To create pEasy1-CMV-RBD-8His genetic construct based on the Adv5 adenovirus genome, pShuttle-CMV-RBD-8His vector (linearized by PmeI restriction enzyme) was fused with ΔE1a-ΔE3 fragment of adenovirus genome by homologous recombination in BJ1586 cells. Restriction analysis using PacI and BamHI enzymes confirmed that the pEasy1-CMV-RBD-8His expression vector is constructed correctly (Fig. 3).

Generation of recombinant protein producers and isolation of recombinant SARS-CoV-2 proteins and RBD-CRM197 fusion protein. Created genetic constructs containing DNA sequences of SARS-CoV-2 N and RBD proteins, as well as RBD-CRM197 fusion protein sequence, were used for transfection of prokaryotic and eukaryotic cells (*E. coli* cloning strain DH10B, *E. coli* expression strain Rosetta (DE3) and human embryonic kidney

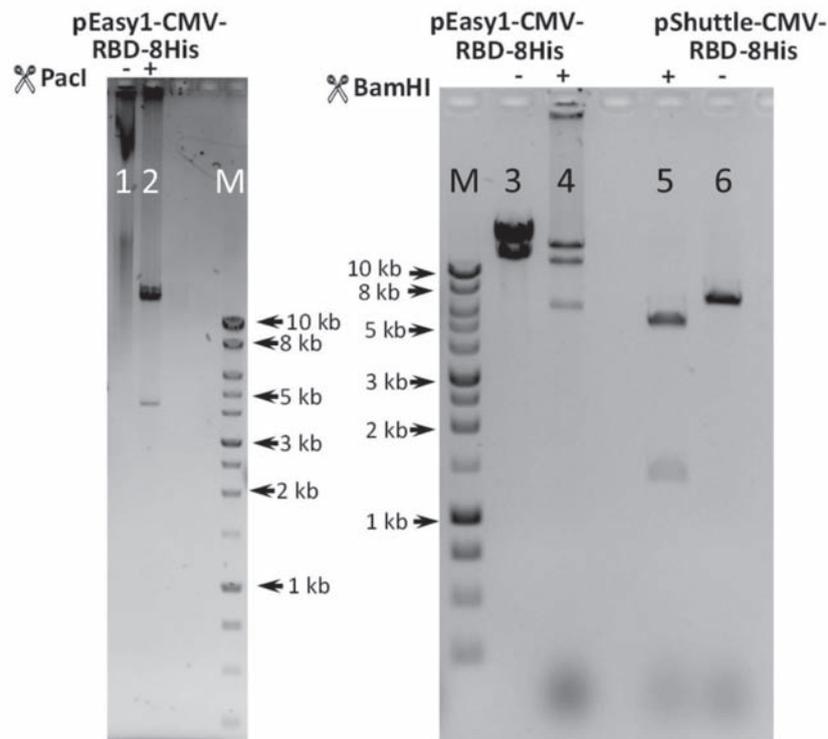


Fig. 3. Restriction analysis of pEasy1-CMV-RBD-8His (1-4) and pShuttle-CMV-RBD-8His (5, 6) genetic constructs: before (1) and after (2) treatment with PaclI restriction enzyme; before (3, 6) and after (4, 5) treatment with BamHI restriction enzyme; M – DNA ladders

cells HEK293). The presence of genetic constructs in the cells was confirmed by PCR test results.

Genetic constructs based on the adenoviral pEasy1-CMV vector after treatment with PaclI restriction enzyme were also transfected into the HEK-293 cells. After reproduction, the primary isolate of recombinant virus was purified by centrifugation in CsCl gradient and used to infect the HEK293 cells.

Purification of the recombinant proteins was performed separately from a soluble fraction of bacterial lysates, inclusion bodies, and supernatants of eukaryotic cells by immobilized metal affinity chromatography using Ni<sup>2+</sup>-NTA agarose with refolding. Purification quality, yield and molecular weights of the obtained recombinant proteins were evaluated by SDS-PAGE. It should be noted that the yield of recombinant proteins expressed in eukaryotic cells was low.

Purified fractions of the recombinant proteins expressed in prokaryotic cells contained a minimal amount of contaminants, fractions of the recombinant N and RBD proteins had a high concentration of the target proteins (Fig. 4). The molecular weights of the obtained recombinant proteins were deter-

mined by SDS-PAGE and were as follows: 48 kDa – protein N, 25 kDa – RBD of S protein, 83 kDa – RBD-CRM197 fusion protein that corresponds to the calculated molecular weights. The molecular weight of SARS-CoV-2 N protein can vary in the range of 47-60 kDa depending on the presence of post-translational modifications (phosphorylation and glycosylation) [23].

Study of antigenic properties of the recombinant proteins expressed in prokaryotic cells by ELISA and Western blotting. Antigenic properties of the recombinant RBD and RBD-CRM197 were studied by indirect ELISA using rabbit polyclonal antibodies against SARS-CoV-2 RBD domain of S protein and biotinylated Fab fragments of equine anti-diphtheria toxoid antibodies.

The results of ELISA using rabbit polyclonal anti-spike RBD antibodies confirmed RBD presence in the recombinant RBD (Fig. 5, A) and RBD-CRM197 (Fig. 5, B) proteins.

The results of ELISA using biotinylated Fab fragments of equine anti-diphtheria toxoid antibodies confirmed CRM197 presence in the recombinant RBD-CRM197 fusion protein (Fig. 5, C). As a posi-

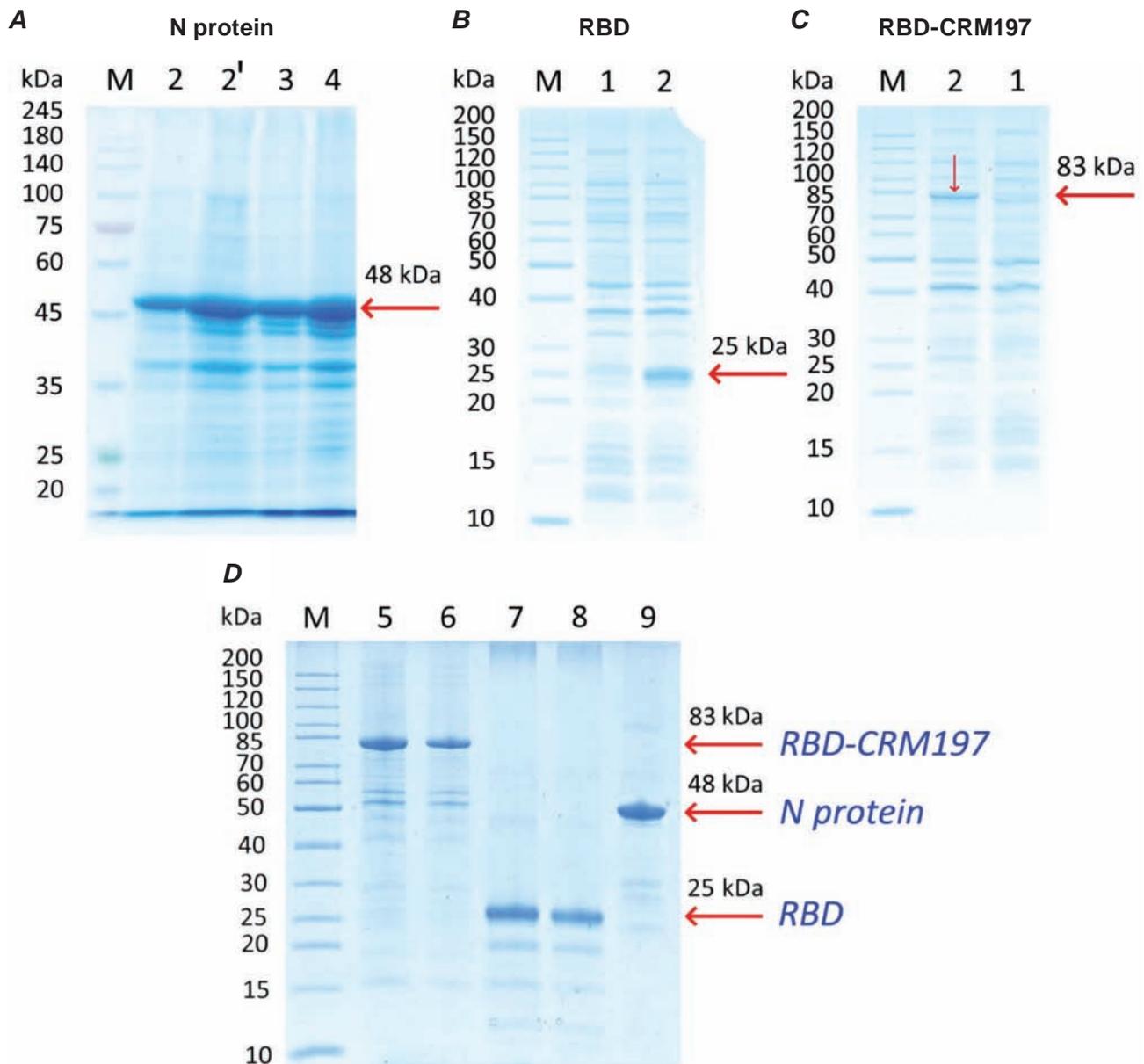


Fig. 4. Electrophoregrams of the recombinant SARS-CoV-2 proteins (expressed in prokaryotic cells) at different stages of protein isolation and purification: (above) bacterial cell lysates, prepared from producers of N protein (A), S protein RBD (B) and RBD-CRM197 fusion protein (C) (1 – before induction of protein expression, 2 (2') – after induction of protein expression; 3-4 – selected inclusion bodies); (bottom) recombinant proteins purified by immobilized metal affinity chromatography, IMAC (D) (5-6 – RBD-CRM197, 7-8 – RBD, 9 – N protein); M – protein molecular weight markers

tive control, we used the recombinant CRM197 obtained earlier [13].

The Western blotting results also showed that rabbit polyclonal anti-spike RBD antibodies recognized the recombinant RBD and RBD-CRM197 proteins (Fig. 5, D) that confirmed RBD presence in these proteins.

Identification of amino acid sequence of the recombinant proteins by MALDI-TOF mass spectrometry. For MALDI-TOF analysis, tryptic protein hydrolysates containing peptides of different lengths from the recombinant N and RBD-CRM197 proteins were used. The mass spectrum of tryptic hydrolysate of the recombinant SARS-CoV-2 N protein is presented in Fig. 6.

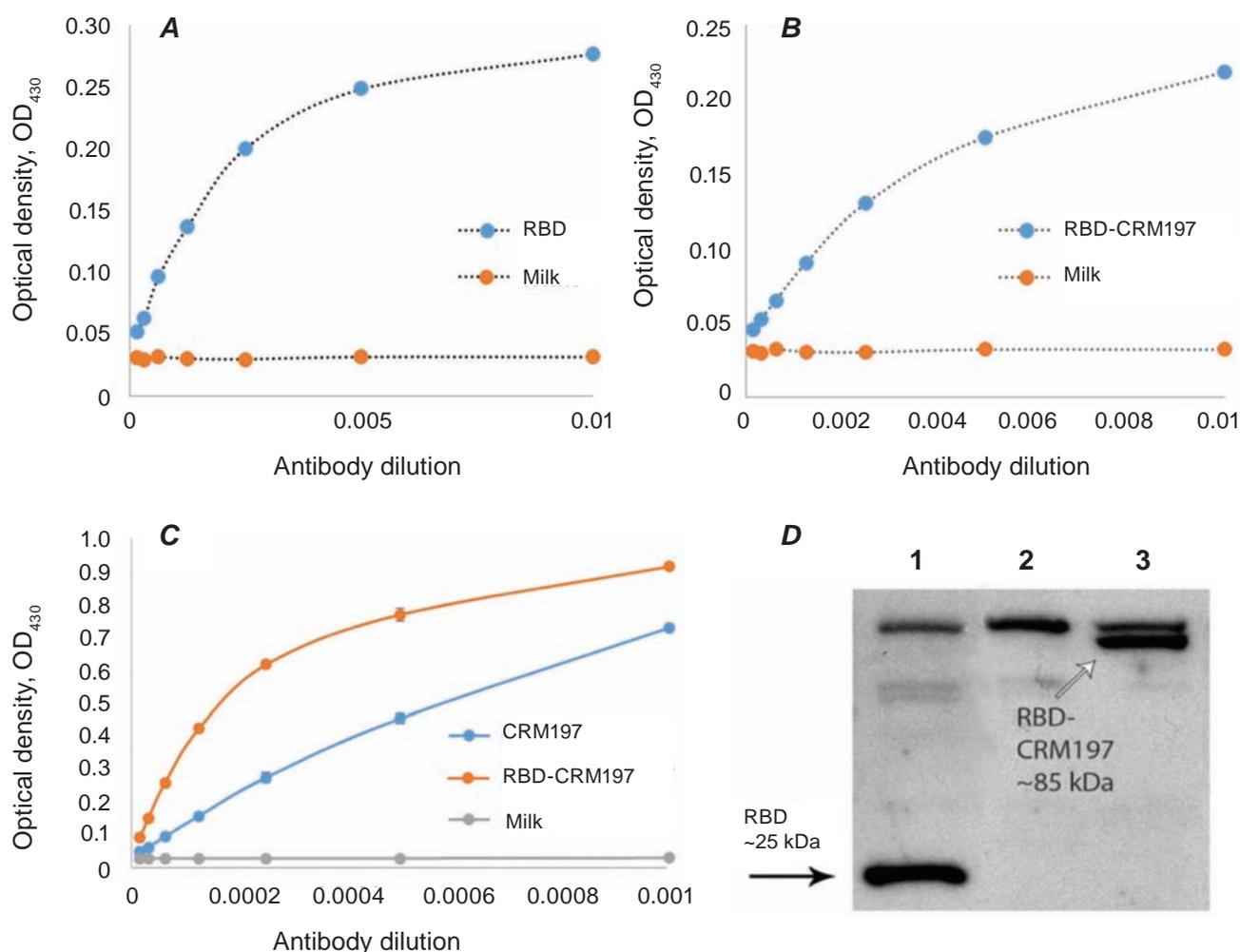


Fig. 5. Antigenic properties of the recombinant RBD and RBD-CRM197 (expressed in prokaryotic cells) studied by immunochemical methods: enzyme-linked immunosorbent assay of binding RBD (A), RBD-CRM197 (B) and control antigen (milk proteins) (A, B) with anti-RBD antibodies; binding of CRM197, RBD-CRM197 and control antigen (milk proteins) with anti-diphtheria toxoid antibodies (C); Western blot analysis (D) of interactions between bacterial lysates from RBD (1), RBD-CRM197 (3) cell producers and anti-RBD antibodies in comparison with negative control (2) – bacterial cell lysate before induction of protein expression

Amino acid sequences of peptides corresponding to individual peaks of the mass spectrum are shown in Table 3. As can be seen from the table, the overlapping peptides represent the complete amino acid sequence of the SARS-CoV-2 N protein that confirms the identity of the recombinant and natural protein N sequences.

MALDI-TOF mass spectrum analysis also confirmed the sequence identity of the recombinant RBD-CRM197 fusion protein and natural (RBD, CRM197) proteins (data not presented).

As follows, molecular weights, antigenic properties and amino acid sequences of the obtained recombinant proteins and natural SARS-CoV-2

proteins are identical that was confirmed by SDS-PAGE, Western blotting, ELISA and MALDI-TOF mass spectrometry. In addition, the identity of the recombinant N, RBD and RBD-CRM197 proteins was also confirmed by DNA sequencing.

The obtained recombinant SARS-CoV-2 N and RBD proteins, and RBD-CRM197 fusion protein can be used to study antigenic and immunogenic properties of SARS-CoV-2 proteins, as well as to develop new immunodiagnosics and COVID-19 vaccines.

The use of RBD-CRM197 fusion protein in COVID-19 vaccines will probably increase the immunogenicity of RBD. This is supported by the fact

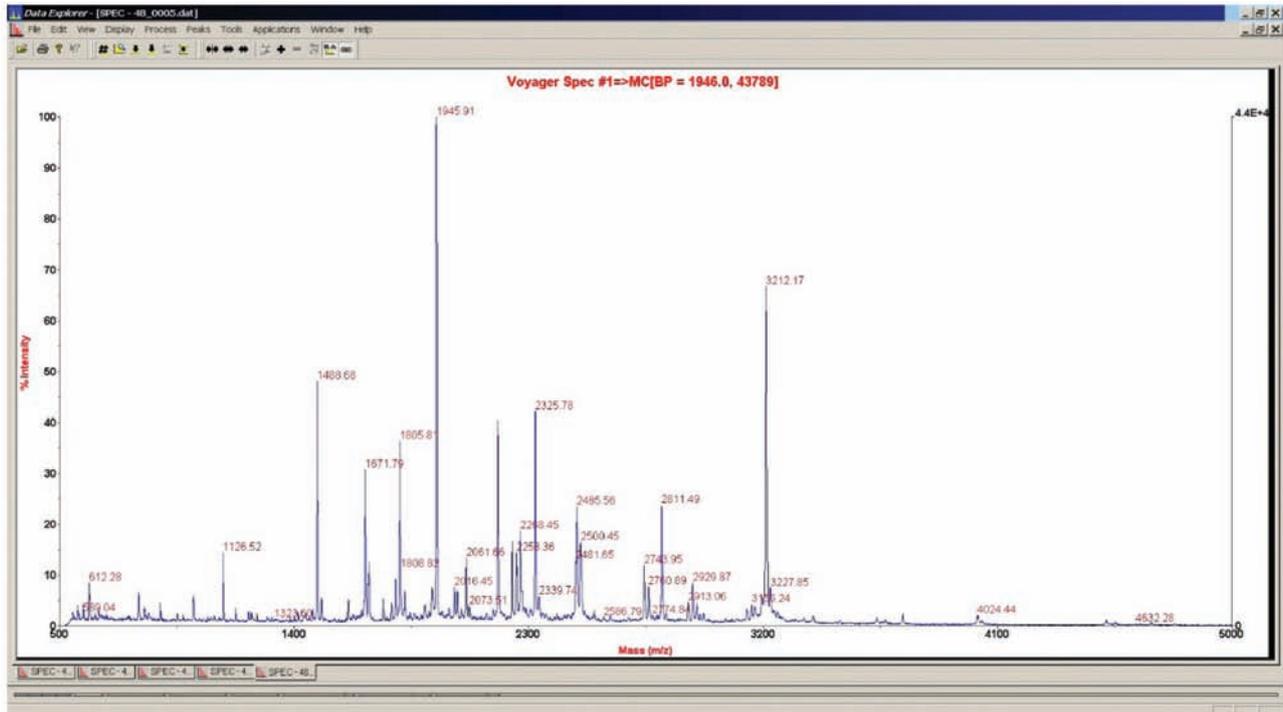


Fig. 6. MALDI-TOF mass spectrum of recombinant SARS-CoV-2 N protein tryptic peptides

Table 3. Identification of recombinant SARS-CoV-2 N protein tryptic peptides by MALDI-TOF mass spectrometry

No	Mass spectrum peak	Amino acid sequence	No	Mass spectrum peak	Amino acid sequence	No	Mass spectrum peak	Amino acid sequence
1	2928.3886	296-321*	16	2195.0071	17-38*	31	1256.6593	377-387*
2	2927.5373	152-179*	17	2181.0207	71-90	32	1234.646	229-239*
3	2809.4379	43-67*	18	2060.1498	152-171	33	1202.6124	240-250
4	2759.2451	408-432	19	2025.0361	322-340	34	1126.564	269-278
5	2739.49	146-171*	20	1990.0491	390-407*	35	1044.5796	252-261*
6	2608.3491	41-63*	21	1944.9522	279-295*	36	1044.5724	364-372*
7	2581.2429	68-90*	22	1861.9541	391-407	37	904.4887	64-70*
8	2496.269	322-344*	23	1823.8154	17-34	38	886.4053	172-179
9	2483.3286	212-235*	24	1788.8511	280-295	39	846.4502	103-109*
10	2481.164	280-301*	25	1742.7987	1-16*	40	805.3799	198-205
11	2364.2306	130-151*	26	1684.8904	130-145	41	679.3482	188-193*
12	2337.1218	71-91*	27	1670.8894	212-228	42	656.4202	259-263*
13	2324.1894	43-63	28	1643.8427	358-371*	43	605.3365	35-40*
14	2267.0807	110-129	29	1605.8013	236-250*	44	588.31	206-211
15	2236.0756	302-321	30	1330.7073	240-251*	45	587.3147	105-109

\*Peptides corresponding to amino acid sequences in SARS-CoV-2 N protein

that immunization of mice with chemical conjugate of CRM197 with several synthetic peptides from SARS-CoV-2 S protein stimulated high titers of anti-S protein IgG antibodies with neutralizing activity against SARS-CoV-2 [24]. In addition, another CRM197-based fusion protein, which was recently obtained by replacing the R-domain of CRM197 with S protein RBD [25], is proposed to use as a component of COVID-19 vaccines.

It should be noted that due to the presence of R and T domains of CRM197, the recombinant RBD-CRM197 fusion protein is able to bind to the proHB-EGF receptor and to be translocated across the endosome membrane directly into the cytosol of immune cells. This may promote the presentation of S protein RBD antigenic peptides on the surface of immune cells for recognition and, thanks to that, stimulate the immune response against RBD.

Thus, the production of the recombinant SARS-CoV-2 N and RBD proteins and RBD-CRM197 fusion is the first step towards the emergence of the domestic COVID-19 vaccine in Ukraine. The undoubted advantage of the domestic COVID-19 vaccine developing is the ability to regulate the production volume and deliver the required vaccine quantity timely to the population of Ukraine, regardless of foreign policy situation and international agreements.

*Conflict of interest.* Authors have completed the Unified Conflicts of Interest form at [http://ukr-biochemjournal.org/wp-content/uploads/2018/12/coi\\_disclosure.pdf](http://ukr-biochemjournal.org/wp-content/uploads/2018/12/coi_disclosure.pdf) and declare no conflict of interest.

*Funding.* This work was supported by a collective grant from the National Research Foundation of Ukraine (2020.01/0322 “Study of coronavirus SARS-CoV-2 protein immunogenic properties as a basis for COVID19 vaccine development”; NRFU competition “Science for the safety of human and society”).

## **ОДЕРЖАННЯ РЕКОМБІНАНТНИХ ПРОТЕЇНІВ КОРОНАВІРУСУ SARS-COV-2 ТА ЗЛИТОГО КОН'ЮГАТУ ІЗ ДИФТЕРІЙНИМ ТОКСОЇДОМ CRM197**

*O. I. Krynina<sup>1</sup>, S. I. Romaniuk<sup>1</sup>,  
O. B. Gorbatiuk<sup>1,2</sup>, O. G. Korchinskyi<sup>1,3,4</sup>,  
A. V. Rebriv<sup>1</sup>, Ya. S. Kulik<sup>1</sup>, S. O.  
Kozadaeva<sup>1</sup>, A. A. Siromolot<sup>1,5</sup>, M. M. Guzik<sup>1</sup>,  
D. V. Kolibo<sup>1✉</sup>, S. V. Komisarenko<sup>1</sup>*

<sup>1</sup>Інститут біохімії ім. О. В. Палладіна  
НАН України, Київ;

<sup>2</sup>ДУ «Інститут генетичної та регенеративної  
медицини НАМН України», Київ;

<sup>3</sup>Центр інноваційних біомедичних досліджень  
і медичний факультет Жешувського  
університету, Жешув, Польща;

<sup>4</sup>Національний університет ветеринарної медицини  
та біотехнологій ім. С. З. Гжицького, Львів, Україна;

<sup>5</sup>ННЦ «Інститут біології та медицини»,  
Київський національний університет  
імені Тараса Шевченка, Україна;  
✉e-mail: kolibo@biochem.kiev.ua

Швидке поширення у світі пандемії COVID-19, спричиненої коронавірусом SARS-CoV-2, виявило нагальну потребу у розробці захисних вакцин для боротьби з цим захворюванням. Тому створення ефективних продуцентів рекомбінантних протеїнів коронавірусу SARS-CoV-2 стало актуальним завданням, вирішення якого сприятиме вивченню функціональних властивостей SARS-CoV-2, а також появи в Україні вакцини власного виробництва проти COVID-19, що має важливе стратегічне значення для боротьби з пандемією. Метою роботи було створити прокаріотичні та евкаріотичні продуценти рекомбінантних протеїнів коронавірусу SARS-CoV-2 і виділити протеїн нуклеокапсиду (N), рецепторзв'язувальний домен (RBD) протеїну «шипа» (S) і злитий кон'югат RBD-домену

із носієм – дифтерійним токсодом CRM197. Для цього за допомогою методів молекулярної біології та генної інженерії створено відповідні генетичні конструкції, зокрема нездатні до самостійної реплікації аденовірусні вектори на основі AdvC5, що експресують протеїни SARS-CoV-2 і злитий кон'югат. Правильність будови конструкцій підтверджено результатами рестрикційного аналізу та/або секвенування ДНК. Рекombінантні протеїни було виділено за допомогою металоафінної хроматографії, відповідність їх властивостей підтверджено результатами електрофорезу в поліакриламідному гелі, вестерн-блотингу, імуноензимного аналізу та MALDI-TOF мас-спектрометрії. У результаті було створено продуценти рекombінантних протеїнів коронавірусу SARS-CoV-2 та злитого кон'югату на основі бактеріального штаму *E. coli* Rosetta (DE3) та клітинної лінії HEK293, а також виділено у чистому вигляді протеїн N, RBD-домен протеїну S і злитий протеїн RBD-CRM197. Одержані рекombінантні протеїни можуть бути використані для вивчення імуногенних і антигенних властивостей протеїнів коронавірусу SARS-CoV-2, а створені продуценти цих протеїнів здатні забезпечити дешевий і безпечний синтез антигенних субстанцій для розробки та виробництва вітчизняних імунодіагностикумів і вакцин проти COVID-19.

**Ключові слова:** SARS-CoV-2, рекombінантні протеїни, злитий кон'югат, CRM197, імунодіагностикум, вакцина, COVID-19.

### References

1. Komisarenko S. The global coronavirus crisis. K.: LAT&K, 2020. 120 p. (In Ukrainian).
2. COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available at <https://coronavirus.jhu.edu/map.html> (accessed, November 7, 2021).
3. World Health Organization. COVID-19 vaccine tracker and landscape. Available at <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines> (accessed, November 5, 2021).
4. PFIZER and BIONTECH achieve first authorization in the world for a vaccine to combat COVID19. Available at <https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-achieve-first-authorization-world> (accessed, December 2, 2020).
5. Zimmer C, Corum J, Wee S-L. Coronavirus vaccine tracker. Available at <https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html> (accessed, November 4, 2021).
6. Corum J, Zimmer C. How the Novavax vaccine works. Available at <https://www.nytimes.com/interactive/2020/health/novavax-covid-19-vaccine.html> (accessed, May 7, 2021).
7. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020; 367(6483): 1260-1263.
8. Ahlén G, Frelin L, Nikouyan N, Weber F, Höglund U, Larsson O, Westman M, Tuvešson O, Gidlund EK, Cadossi M, Appelberg S, Mirazimi A, Sällberg M. The SARS-CoV-2 N Protein Is a Good Component in a Vaccine. *J Virol*. 2020; 94(18): e01279-20.
9. Malito E, Bursulaya B, Chen C, Lo Surdo P, Picchianti M, Balducci E, Biancucci M, Brock A, Berti F, Bottomley MJ, Nissum M, Costantino P, Rappuoli R, Spraggon G. Structural basis for lack of toxicity of the diphtheria toxin mutant CRM197. *Proc Natl Acad Sci USA*. 2012; 109(14): 5229-5234.
10. Bröker M, Berti F, Schneider J, Vojtek I. Polysaccharide conjugate vaccine protein carriers as a "neglected valency" - Potential and limitations. *Vaccine*. 2017; 35(25): 3286-3294.
11. Jaffe J, Wucherer K, Sperry J, Zou Q, Chang Q, Massa MA, Bhattacharya K, Kumar S, Caparon M, Stead D, Wright P, Dirksen A, Francis MB. Effects of Conformational Changes in Peptide-CRM 197 Conjugate Vaccines. *Bioconjug Chem*. 2019; 30(1): 47-53.
12. Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res*. 1979; 7(6): 1513-1523.
13. Labyntsev AJ, Korotkevych NV, Manoilov KJ, Kaberniuk AA, Kolybo DV, Komisarenko SV. Recombinant fluorescent models for studying the diphtheria toxin. *Russ J Bioorganic Chem*. 2014; 40(4): 401-409.
14. Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nat Protoc*. 2007; 2(4): 924-932.

15. Rudolph R, Lilie H. In vitro folding of inclusion body proteins. *FASEB J*. 1996; 10(1): 49-56.
16. Basu A, Li X, Leong SSJ. Refolding of proteins from inclusion bodies: rational design and recipes. *Appl Microbiol Biotechnol*. 2011; 92(2): 241-251.
17. Schägger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*. 1987; 166(2): 368-379.
18. ExPasy Swiss Bioinformatics Resource Portal. <https://www.expasy.org> (accessed, November 5, 2021).
19. Protein Data Bank. <https://www.rcsb.org> (accessed, November 5, 2021).
20. Qwitterer U, Pohl A, Langer A, Koller S, Abdalla S. A cleavable signal peptide enhances cell surface delivery and heterodimerization of Cerulean-tagged angiotensin II AT1 and bradykinin B2 receptor. *Biochem Biophys Res Commun*. 2011; 409(3): 544-549.
21. Yu K, Liu C, Kim BG, Lee DY. Synthetic fusion protein design and applications. *Biotechnol Adv*. 2015; 33(1): 155-164.
22. Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci Rep*. 2020; 10(1): 14991.
23. Supekar NT, Shajahan A, Gleinich AS, Rouhani DS, Heiss C, Chapla DG, Moremen KW, Azadi P. Variable posttranslational modifications of severe acute respiratory syndrome coronavirus 2 nucleocapsid protein. *Glycobiology*. 2021; 31(9): 1080-1092.
24. Zhao Q, Gao Y, Xiao M, Huang X, Wu X. Synthesis and immunological evaluation of synthetic peptide based anti-SARS-CoV-2 vaccine candidates. *Chem Commun (Camb)*. 2021; 57(12): 1474-1477.
25. Bellone ML, Puglisi A, Dal Piaz F, Hochkoeppler A. Production in *Escherichia coli* of recombinant COVID-19 spike protein fragments fused to CRM197. *Biochem Biophys Res Commun*. 2021; 558: 79-85.