

Development of a Platform for Producing Recombinant Protein Components of Epitope Vaccines for the Prevention of COVID-19

Anna S. Karyagina^{1,2,3,a*}, Alexander V. Gromov¹, Tatyana M. Grunina^{1,2}, Alexander M. Lyaschuk¹,
Alexander V. Grishin¹, Natalia V. Strukova¹, Maria S. Generalova¹, Zoya M. Galushkina¹,
Lyubov' A. Soboleva¹, Olga Yu. Dobrinina¹, Tatyana N. Bolshakova¹, Marina E. Subbotina^{1,2,b*},
Ekaterina A. Romanovskaya-Romanko⁴, Igor V. Krasilnikov⁵, Nikita B. Polyakov^{1,6},
Andrey I. Solovyev¹, Daniil A. Grumov¹, Vladimir G. Zhukhovitsky¹,
Ekaterina I. Ryabova¹, Vladimir V. Prokofiev¹, and Vladimir G. Lunin^{1,2}

¹Gamaleya National Research Center of Epidemiology and Microbiology,
Ministry of Health of the Russian Federation, 123098 Moscow, Russia

²All-Russia Research Institute of Agricultural Biotechnology, 127550 Moscow, Russia

³Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, 119992 Moscow, Russia

⁴Institute of Influenza, Ministry of Health of the Russian Federation, 197376 St. Petersburg, Russia

⁵Saint Petersburg Institute of Vaccines and Sera, FMBA, 198320 St. Petersburg, Russia

⁶Vernadsky Institute of Geochemistry and Analytical Chemistry, Russian Academy of Sciences, 119991 Moscow, Russia

^ae-mail: akaryagina@gmail.com

^be-mail: subbotinam@gmail.com

Received June 8, 2021

Revised July 11, 2021

Accepted July 28, 2021

Abstract—A new platform for creating anti-coronavirus epitope vaccines has been developed. Two loop-like epitopes with lengths of 22 and 42 amino acid residues were selected from the receptor-binding motif of the Spike protein from the SARS-CoV-2 virus that participate in a large number of protein-protein interactions in the complexes with ACE2 and neutralizing antibodies. Two types of hybrid proteins, including one of the two selected epitopes, were constructed. To fix conformation of the selected epitopes, an approach using protein scaffolds was used. The homologue of Rop protein from the *Escherichia coli* ColE1 plasmid containing helix-turn-helix motif was used as an epitope scaffold for the convergence of C- and N-termini of the loop-like epitopes. Loop epitopes were inserted into the turn region. The conformation was additionally fixed by a disulfide bond formed between the cysteine residues present within the epitopes. For the purpose of multimerization, either aldolase from *Thermotoga maritima*, which forms a trimer in solution, or alpha-helical trimerizer of the Spike protein from SARS-CoV-2, was attached to the epitopes incorporated into the Rop-like protein. To enable purification on the heparin-containing sorbents, a short fragment from the heparin-binding hemagglutinin of *Mycobacterium tuberculosis* was inserted at the C-terminus of the hybrid proteins. All the obtained proteins demonstrated high level of immunogenicity after triplicate parenteral administration to mice. Sera from the mice immunized with both aldolase-based hybrid proteins and the Spike protein SARS-CoV-2 trimerizer-based protein with a longer epitope interacted with both the inactivated SARS-CoV-2 virus and the Spike protein receptor-binding domain at high titers.

DOI: 10.1134/S0006297921100096

Keywords: SARS-CoV-2, Spike protein, RBD, RBM, epitope vaccine, epitope, epitope scaffold, aldolase, hybrid protein

INTRODUCTION

If COVID-19 becomes a seasonal disease, the world's population will need periodic vaccinations, just as

it happens in the framework of influenza prevention, which makes development of new variants of highly effective, safe, and inexpensive vaccines for the prevention of COVID-19 an extremely urgent task. One of the prob-

Abbreviations: ALD, aldolase; BCA, bicinchoninic acid; COVID-19, coronavirus infection 2019 (COroNaVirus Disease 2019); DTT, dithiothreitol; HBD, heparin binding domain; IAA, iodoacetamide; IPTG, isopropyl- β -D-thiogalactopyranoside; MALDI, matrix assisted laser desorption/ionization; RBD, receptor binding domain; RBM, receptor binding motif; Tri, α -helix mediating trimerization of the SARS-CoV-2 Spike protein.

* To whom correspondence should be addressed.

lems of the whole-virion vaccines (protein vaccines containing full-length viral proteins) and vector vaccines (RNA vaccines leading to the production of viral proteins in the body) is that not all antibodies they induce are protective [1]. Moreover, under certain circumstances, antibodies can be formed that contribute to a more severe course of the disease upon infection after vaccination (the phenomenon of antibody-dependent enhancement of infection) [2-4]. One of the promising approaches that allow to knowingly avoid several problems, including the ones listed above, is the development of the so-called epitope vaccines – vaccines based on the individual small protein epitopes of an infectious agent. They do not have the disadvantages typical of live vaccines (reversion of pathogenic properties, residual virulence, incomplete inactivation, etc.). In addition, such vaccines are highly standardized, have weak reactogenicity, and can be used to avoid both development of autoimmune processes during immunization and formation of the non-protective antibodies and antibodies that contribute to the development of antibody-dependent enhancement of virus infection.

The basis for successful functioning of epitope vaccines is the correct selection of epitopes – surface regions of viral proteins that can effectively induce formation of the desired spectrum of antibodies. An adequate selection of epitopes can direct immune system to produce only virus-neutralizing antibodies. In the case of the SARS-CoV-2 virus, the choice of epitopes can be based directly on the structural data of complexes of the surface Spike protein, which is responsible for cell penetration, with the human virus-neutralizing antibodies. Such opportunity is provided by the rapid accumulation of the necessary structural data: for example, in the spring and summer of 2020, the first data describing structure of the complexes were published [5], and at the end of May 2021 200 structures of Spike protein complexes with antibodies have already been deposited in the Protein Data Bank database (PDB, <https://www.rcsb.org/>). Neutralizing antibodies that interact with Spike protein can be divided into several classes according to the region of interaction. A large class of neutralizing antibodies directly inhibit interaction of the Spike protein with angiotensin-converting enzyme 2 (ACE2), which initiates viral entry into the cell. These antibodies are in contact with amino acid residues (aa) located in the so-called receptor-binding motif (RBM, 437-508 aa); therefore, it seems appropriate to choose determinants for epitope vaccines within this motif.

Another important aspect of the success of epitope vaccines is solving the problem of low immunogenicity by ensuring optimal exposure of epitopes and their multimerization. The most promising modern strategies in this area are associated with the use of approaches based on the so-called epitope scaffolds. Epitope scaffolds are usually determined as carrier proteins to which peptide epitopes are attached by means of genetic engineering or

others [6, 7]. Proteins forming trimeric, tetrameric, or hexameric structures are often used as carrier proteins. In this case, short linear determinants can induce production of the protective antibodies when located at the N-terminus of the scaffold protein apparently due to spontaneous formation of the native conformation [6]. Sufficiently extended linear epitopes of viral proteins, which form numerous contacts in complexes with antibodies, can be represented in the structure of the viral protein by the loop-like structures with N- and C-ends close to each other. In this case, the scaffold protein can be used for fixing the ends of the loop-like protein regions close to each other, similarly to the work done by Juraja et al. [8]. It can be expected that the conformational correspondence of the epitope to its native conformation in the composition of the full-length antigen will be largely ensured with this approach.

The aim of this work was to develop a new platform for creating protein components of the epitope anticoronavirus vaccines based on either epitope scaffold for fixing conformation, or on the domain for multimerization (trimerization) of epitopes.

MATERIALS AND METHODS

Strain and vectors. Bacterial strain *Escherichia coli* BL21 (DE3) (*E. coli* B F⁻ *dcm ompT hsdS*(r_Bm_B) *gal λ*(DE3) (Agilent Technologies, USA), modified plasmid vector pQE6 (Qiagen, USA), in which the T5 promoter is replaced by T7, and plasmid pREP4 from *E. coli* M15 [pREP4] (Qiagen) were used.

Construction of gene encoding the Rop-D2-Rop-Tri-HBD protein. The Rop-D2-Rop-Tri-HBD protein includes amino acid sequence of the α -helix of the Rop-like protein from *Methylococcus capsulatus* (2-34 aa, PDB-code: 2JS5_A, RefSeq ID: WP_010959602), sequence of the D2 determinant of the receptor-binding motif (RBM) of the Spike protein of the virus SARS-CoV-2 (470-490 aa, UniProtKB ID: locus SPIKE_SARS2, accession P0DTC2), sequence of the second α -helix of the Rop-like protein (34-65 aa), α -helix sequence mediating trimerization of the SARS-CoV-2 Spike protein (Tri) (958-991 aa, UniProtKB ID: locus SPIKE_SARS2, accession P0DTC2), and a short region of heparin-binding hemagglutinin HBHA from *Mycobacterium tuberculosis*, heparin-binding domain (HBD, heparin-binding domain) (160-174 aa, identifier UniProtKB/Swiss-Prot: A1KFU9), as well as additional residues representing linker sequences and residues corresponding to the nucleotide sequences with introduced restriction sites. Nucleotide sequence of the corresponding chimeric gene was designed in such a way that the NcoI and BamHI restriction sites were located at its 5'-end, and the BglII and Kpn2I sites at the 3'-end. The region encoding the D2 determinant was flanked by the

AgeI and Eco8II sites. Optimization of the codon composition of the synthetic gene was carried out using the JCat program (<http://www.jcat.de/>), correction of the transcribed RNA secondary structure – using the DINAMelt web server (<http://www.unafold.org/Dinamelt/applications/two-state-melting-folding.php>). The synthetic gene was inserted into the modified plasmid vector pQE6 at the NcoI and Kpn2I sites, resulting in the plasmid pL1003 encoding the Rop-D2-Rop-Tri-HBD protein. All synthetic genes used in this work were obtained from Evrogen (Russia).

Construction of gene encoding the Rop-D3-Rop-Tri-HBD protein. A DNA fragment encoding the determinant D3 of the receptor binding motif (RBM) of the Spike protein of the SARS-CoV-2 (453-494 aa, UniProtKB ID: locus SPIKE_SARS2, accession P0DTC2) flanked by the AgeI and Eco8II sites was optimized for codon composition and secondary structure of the mRNA (synthesized by Evrogen) and cloned into the plasmid pL1003 at sites AgeI and Eco8II. As a result, the plasmid pL1008 was obtained, encoding the Rop-D3-Rop-Tri-HBD protein, similar in structure to the Rop-D2-Rop-Tri-HBD protein, except for the region of the antigenic determinant of the Spike protein of the SARS-CoV-2.

Construction of gene encoding the Rop-D2-Rop-ALD-HBD protein. Rop-D2-Rop-ALD-HBD protein includes the same regions of amino acid sequences of various proteins as the Rop-D2-Rop-Tri-HBD protein, except that the α -helix sequence mediating trimerization of the SARS-CoV-2 Spike protein (Tri), was replaced by the sequence of aldolase from *Thermotoga maritima* (2-201 aa, PDB code: 1WA3). Nucleotide sequence of the gene was optimized in terms of the codon composition and secondary structure of mRNA and was designed so that the NcoI and BamHI sites were located at its 5'-end, and the BglII and Kpn2I sites at the 3'-end. The region encoding the D2 determinant was flanked by the AgeI and Eco8II sites. The synthetic gene was inserted into the modified plasmid pQE6 at the NcoI and Kpn2I sites. As a result, the plasmid pL989 encoding the Rop-D2-Rop-ALD-HBD protein was obtained.

Construction of gene encoding the Rop-D3-Rop-ALD-HBD protein. The fragment encoding D3 determinant from the plasmid pL1008 was cloned into the plasmid pL989 at sites AgeI and Eco8II. The resulting plasmid pL990 encodes the Rop-D3-Rop-ALD-HBD protein, which is structurally similar to the Rop-D2-Rop-ALD-HBD protein, except for the region of antigenic determinant of the Spike protein of the SARS-CoV-2 virus.

Construction of gene encoding the Rop-RBM-Rop protein. The Rop-RBM-Rop protein includes two amino acid sequences of the Rop-like protein from *M. capsulatus* (2-34 and 35-66 aa, PDB code: 2JS5_A, RefSeq ID: WP_010959602), a region that includes the RBM protein of the SARS-CoV-2 (433-511 aa, UniProtKB ID: locus SPIKE_SARS2, accession P0DTC2), as well as addition-

al residues representing linker sequences and residues corresponding to the nucleotide sequences with introduced restriction sites. The sites NcoI and BamHI were inserted at the 5'-end of the nucleotide sequence, and BglII and Kpn2I at the 3'-end. The synthetic gene was inserted into a modified plasmid vector pQE6 carrying the T7 promoter instead of the T5 promoter at the NcoI and Kpn2I sites. As a result, the plasmid pL926 encoding the Rop-RBM-Rop protein was obtained.

In all cases, correct assembly of the sequences was confirmed by Sanger sequencing.

Cultivation of producer strains. Cultures of transformed *E. coli* cells were grown in LB media supplemented with kanamycin (25 mg/liter) and ampicillin (200 mg/liter) on a shaker at 180 rpm and 37°C until OD₆₀₀ reached 1.0-1.2. Protein synthesis was induced by addition of 0.5 mM IPTG (isopropyl- β -D-thiogalactopyranoside), then the cultures were grown for another 4 h at 180 rpm. Producers of insoluble proteins were cultured at 37°C, and of soluble ones – at 30°C. The resulting cell cultures were centrifuged for 30 min at 5,000g and 10°C. The biomass was stored at -20°C.

Disruption of *E. coli* cells. Thawed biomass of bacterial cells (1 g) was suspended in a lysis buffer [20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1% Triton X 100, 1 mM phenylmethylsulfonyl fluoride (PMSF)] at a ratio of not less than 1 : 10 (w/v), next 100 μ g/ml of lysozyme was added, incubated for 20 min at room temperature, and cells were disrupted with a Vibra Cell VCX750 ultrasonic device (Sonics, USA) at 40% amplitude for 2.5 min (3-s sonication pulses with 2-s intervals) on ice. The mixture was centrifuged for 30 min at 20,000g and 10°C, as a result, proteins of the soluble fraction remained in the supernatant, and insoluble fraction in the form of inclusion bodies was in the obtained sediment. Both proteins with the α -helix-trimerizer (Rop-D2-Rop-Tri-HBD and Rop-D3-Rop-Tri-HBD), as well as the Rop-RBM-Rop protein were obtained in form of inclusion bodies, proteins with aldolase (Rop-D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD) were predominantly in a soluble form, so they remained in the supernatant.

Preparation for chromatography of proteins in the form of inclusion bodies. The inclusion bodies were washed twice with lysis buffer and once with 20 mM Tris-HCl (pH 8.0). Washed inclusion bodies containing proteins Rop-D2-Rop-Tri-HBD and Rop-D3-Rop-Tri-HBD were dissolved in 8 M urea and 10 mM Tris-HCl (pH 8.0) in a volume equal to the volume used for lysis of cell biomass, centrifuged for 30 min at 9,000g.

Preparation for chromatography of soluble proteins. The supernatant fractions of the Rop-D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD proteins after lysis were diluted with a buffer (20 mM Tris-HCl (pH 8.0), 1% Triton X 100) to a concentration of ~1 mg/ml and heated at 65°C for 15 min on a MR 3001 magnetic stirrer with heating (Heidolph, Germany). Next, the obtained sam-

ple was centrifuged at 9,000g for 30 min producing the target protein mainly in the supernatant, and the contaminating bacterial proteins mainly in the sediment.

Column chromatography. Column chromatography was performed using a low-pressure chromatographic unit ÄKTA start (GE Healthcare Life Sciences, USA). Protein concentration was determined by absorbance at 280 nm using the Bicinchoninic acid (BCA) Protein Assay (AppliChem, Germany). To prepare protein samples for polyacrylamide gel electrophoresis (PAGE) according to Laemmli, we used a buffer containing 2 mM EDTA, 125 mM Tris-HCl (pH 6.8), 20% glycerol, 4% sodium dodecyl sulfate and 0.015% bromophenol blue supplemented with a reducing agent (0.2 M dithiothreitol, DTT) or without it. Protein samples were diluted with a buffer at a 1 : 1 ratio, heated for 5 min at 95°C, and centrifuged for 5 min at 16,000g; the supernatant was applied to PAGE. Electrophoresis was performed using an SE 260 Mighty Small II device (Amersham, USA) and a set of accessories from GE Healthcare. After electrophoresis, the gels were stained with Coomassie brilliant blue R250 (Bio-Rad, USA). To estimate apparent molecular weights, set of 14–97 kDa protein markers (Bio-Rad) was used. The gels were imaged using a Gel Doc™ XR + gel documentation system (Bio-Rad). Densitometric evaluation of bands on the gel was performed using a Quantity One software (Bio-Rad).

Chromatography of proteins from inclusion bodies on WorkBeads 40S sorbent. The proteins Rop-D2-Rop-Tri-HBD, Rop-D3-Rop-Tri-HBD, and Rop-RBM-Rop prepared for chromatography were loaded onto a column with 6 ml of WorkBeads 40S sorbent (Bio-Works, Sweden) equilibrated with a solution of 8 M urea in 10 mM Tris-HCl (pH 8.0), at a rate of 1 ml/min, column washed at a rate of 2 ml/min until the absorption at 280 nm reached a plateau. Elution was carried out at the same rate with a solution of NaCl (a linear concentration gradient of 0–1 M in a volume of 60 ml) in the same buffer. The eluate from the column was collected in 5-ml fractions, then the fractions of the eluate with maximum absorbance at 280 nm were combined and dialyzed against a solution of 4 M urea in 25 mM Tris-HCl buffer (pH 8.0) for 24 h at 4°C and centrifuged in for 30 min at 9,000g. The volume ratio of the dialyzed sample to the dialysis buffer was 1 : 10. Protein solutions (hereinafter) were dialyzed using a Zellu Trans/ROTH 3,5 E657.1 cellulose dialysis membrane (Carl Roth, Germany) with 25 µm thickness and molecular weight cut-off 3500 Da. After dialysis, electrophoresis of the pooled fractions was performed, and protein concentration was estimated from the absorbance of the solution at 280 nm using the BCA Protein Assay.

Chromatography of soluble proteins on WorkBeads 40S sorbent. The proteins Rop-D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD prepared for chromatography were filtered with a syringe through a polyethersulfone

(PES) membrane filter with a pore diameter of 0.45 µm (Corning, USA). The filtrate was applied to a column with 10 ml of WorkBeads 40S sorbent, equilibrated with 20 mM Tris-HCl, pH 8.0 at a rate of 1 ml/min, washed at a rate of 2 ml/min until the absorption at 280 nm reached a plateau. The elution was carried out at the same rate with a solution of NaCl (with a linear concentration gradient of 0–1 M in a volume of 100 ml) in the same buffer, the eluate from the column was collected in 4-ml fractions, then the fractions corresponding to maximum absorption at 280 nm were combined and dialyzed against a buffer containing 20 mM Tris-HCl (pH 8.0) for 24 h, centrifuged for 30 min at 9,000g. The volume ratio of the dialyzed sample to the dialysis buffer was 1 : 10. Then, the supernatant was taken, and electrophoresis of the obtained proteins was carried out and protein concentration was estimated by the absorbance of the solution at 280 nm using the BCA Protein Assay.

Affinity chromatography of Rop-D2-Rop-Tri-HBD and Rop-D3-Rop-Tri-HBD proteins on heparin-sepharose CL 6B. The resulting protein solution after dialysis was subjected to affinity chromatography on heparin-sepharose CL 6B column (GE Healthcare). The column with 10 ml of sorbent was equilibrated with a solution of 4 M urea in 20 mM Tris-HCl buffer (pH 8.0), protein binding and column washing were carried out in the same buffer, all operations were performed at a flow rate of 1 ml/min. The protein was eluted with a 0–1 M NaCl concentration gradient in a 4 M urea solution in Tris-HCl buffer (pH 8.0) at a flow rate of 1 ml/min. Protein fractions of a total volume of 3 ml were dialyzed against 25 mM Tris-HCl buffer (pH 6.8), gradually decreasing the urea concentration in the dialysis buffer (0.5 M steps) for 24 h at 4°C. The ratio of the volume of the dialyzed sample to the dialysis buffer was 1 : 10 at all stages of dialysis. After dialysis, concentration and total amount of protein were determined based on absorbance of the solution at 280 nm using the BCA Protein Assay. The obtained protein preparations were lyophilized and used for further studies.

Affinity chromatography of soluble proteins on heparin-sepharose CL 6B. The resulting solution of Rop-D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD proteins after dialysis was loaded onto columns with 10 ml of heparin-sepharose sorbent equilibrated with 20 mM Tris-HCl buffer (pH 8.0). Protein binding and column washing were performed in the same buffer; all operations were performed at a flow rate of 1 ml/min. The protein was eluted with a 0–1 M NaCl concentration gradient in 20 mM Tris-HCl buffer (pH 8.0) at a rate of 1 ml/min. Three 2-ml fractions were collected. Protein fractions with a total volume of 6 ml were pooled and dialyzed against a 20 mM Tris-HCl buffer (pH 8.0), 50 mM NaCl at a volume ratio of 1 : 10 for 24 h at 4°C with three buffer changes. After dialysis, concentration and total amount of protein were determined based on absorbance

of the solution at 280 nm using the BCA Protein Assay. The obtained protein preparations were lyophilized and used for further studies.

Chromatography of Rop-RBM-Rop protein on WorkBeads 40 DEAE sorbent. A solution of the Rop-RBM-Rop protein in a buffer containing 4 M urea and 20 mM Tris-HCl (pH 8.0), obtained after dialysis of the fractions collected during chromatography on WorkBeads 40S, was loaded onto a column with 5 ml of WorkBeads 40 DEAE sorbent (Bio-Works) pre-equilibrated with a buffer containing 4 M urea, 20 mM Tris-HCl (pH 8.0). All column manipulations were performed at a flow rate of 1 ml/min. The sorbent with immobilized protein was washed with the same buffer (50 ml) until the UV detector signal dropped and stabilized. Elution of the target protein was performed with a 0-1 M NaCl concentration gradient in 4 M urea in the same buffer by five column volumes. Pure Rop-RBM-Rop protein was contained in the unbound protein fraction. When eluted from the sorbent, impurity proteins were also eluted along with the target protein. The pooled protein fractions (14 ml) of unbound protein were incubated in the elution buffer for a day, after which the protein solution was dialyzed against 1 liter of 20 mM sodium phosphate buffer (pH 5.5) at 4°C with two buffer changes.

Hydrolysis of the Rop-RBM-Rop protein with trypsin in electrophoresis gel before mass spectrometry. Proteins separated by PAGE were hydrolyzed inside the gel fragments according to the method proposed by Shevchenko et al. [9] with and without addition of DTT and iodoacetamide (IAA). Tryptic peptides were extracted from the gel according to the method we published earlier [10]. The obtained extract was dried to dryness using a Savant SPD121P vacuum concentrator (Thermo Scientific, USA) and dissolved in 20 µl of 0.1% formic acid. The samples were desalted on self-made columns with a reverse-phase membrane (analogue C18); the sample was eluted with 90% solution of acetonitrile in water. For further use, 4 µl of a sample in 90% acetonitrile was taken and mixed with 1 µl of a saturated solution of 2.5 dihydroxybenzoic acid in a 30% acetonitrile solution containing 0.5% glacial acetic acid (v/v) on a steel target following by air-drying at room temperature.

MALDI mass spectrometry analysis. Analysis of the peptides was carried out with an Ultraflextreme time-of-flight MALDI mass spectrometer (Bruker Daltonik GmbH, Germany). Reflectron mode was used for detection of positive ions at the following voltages: IS1 – 20.12 kV, IS2 – 17.82 kV; Lens – 7.47 kV, Ref1 – 21.07 kV, Ref2 – 10.80 kV.

Ions were detected in the m/z range 650-5000. Peaks of autolytic fragments of trypsin and keratin were used as internal standards, which were then excluded from the final lists of detected masses.

Fragmentation of individual ions was carried out under conditions of dissociation induced by the collision

of ions with argon molecules, in the detection mode of positive ions at the following voltages: IS1 – 7.50 kV, IS2 – 6.7 kV; Lens – 3.50 kV, Ref1 – 29.50 kV, Ref2 – 14.00 kV; Lift1 – 19 kV, Lift2 – 2.80 kV. Laser power in the process of recording mass spectra was varied to achieve best fragmentation of the ions under study.

Analysis of mass spectrometry data. Mass spectra were processed using the Flex Analysis 3.4 software (Bruker Daltonik GmbH). Mass spectra were smoothed according to the Savitzky – Golay algorithm (width 0.2 m/z , 1 cycle) and baseline was subtracted according to the TopHat algorithm. The following parameters of peak detection were used: peak detection algorithm was SNAP2, signal-to-noise ratio was 3, maximum number of peaks in the spectrum was 500. Search and identification of the proteins by the “peptide mass fingerprint” method in the database was carried out using the software complex MASCOT (local version 2.1.03, Matrix Science, Great Britain). The following search parameters were used: accuracy in determining the mass – 50 ppm, possible post-translational modifications – oxidation of methionine, histidine, tryptophan, Glu pyro Glu, Gln → pyro Gln. Proteins with a probability parameter MOWSE score >32 were reliably identified (95%, $p < 0.05$). Additional criteria for reliability of determination of the primary structure of a protein was coincidence of its molecular weight with the value determined experimentally. Interpretation of mass spectra, in particular of fragmentation, was additionally carried out using the programs Biotoools (version 3.2) and Proteinscape (version 4.1) (Bruker Daltonik GmbH).

Obtaining the receptor binding domain (RBD) of the surface Spike protein of the SARS-CoV-2. Amino acid sequence of the receptor binding domain (RBD) of the surface Spike protein of the SARS-CoV-2 (319-541 aa, UniProtKB ID: locus SPIKE_SARS2, accession P0DTC2) was modified at the N-terminus with the alkaline phosphatase signal peptide SEAP (MLLLLLLLGLRLQLSLGI) and at the C-terminus with a glycine serine linker and histidine tag sequence (GSHHHHHH-HHHH). Nucleotide sequence of the obtained polypeptide was synthesized by the Evrogen company and cloned into the pCEP plasmid at the XbaI and HindIII restriction sites, thus obtaining the pCEP RBD plasmid. Then, the CHO S cell culture (Thermo Fisher Scientific) was transiently transfected with the pCEP RBD plasmid using the CHOgro system (Mirus Bio, USA) in accordance with the manufacturer’s protocol. The cells were cultured in Erlenmeyer flasks at 125 rpm, 5% CO₂, 80% humidity, 37°C, after 24 h the temperature was lowered to 32°C and culturing was continued for 10 days. Starting from the 3rd day, Cell boosts 7a (2%), 7b (0.2%) (HyClone, USA) and 0.5% CHO Bioreactor Feed (Sigma, USA) were added once a day. After 10 days, the culture fluid was clarified by centrifugation at 5,000g. The receptor-binding domain was purified by metal affinity chromatography on a

ÄKTA start system (GE Healthcare Life Sciences) using HisTrap FF 5 ml columns (GE Healthcare Life Sciences) according to the manufacturer's protocol. Additional purification and buffer exchange for 20 mM sodium phosphate, 500 mM sodium chloride (pH 7.2) was performed on a XK 26/100 column (GE Healthcare Life Sciences) packed with Superdex 200 pg sorbent (GE Healthcare Life Sciences).

Preparation of recombinant proteins for immunization. On the basis of each of the four obtained recombinant proteins, immunogenic compositions were prepared containing 1 nmol of protein in a single dose of 150 μ l (Rop-D2-Rop-Tri-HBD – 18.7 μ g, Rop-D3-Rop-Tri-HBD – 21.4 μ g, Rop-D2-Rop-ALD-HBD – 36.1 μ g, Rop-D3-Rop-ALD-HBD – 38.7 μ g), as well as diethylaminoethyl (DEAE) dextran 500 – 1 mg, montanide ISA 201 (Seppic, France) – 75 μ l, retinol palmitate (100,000 IU/ml) produced by JSC Retinoids (Russia) – 1.5 μ l.

Immunization of laboratory animals. We used female mice of the syngeneic inbred BALB/c line at the age of 5–6 weeks, weighing 18–20 g ($n = 35$, NF Gamaleya Research Center for Epidemiology and Microbiology, Ministry of Health of Russia). Mice divided into four experimental groups ($n = 7$) were injected three times with immunogenic compositions containing recombinant proteins, with an interval of two weeks. The injection was performed with a disposable sterile insulin syringe subcutaneously at the withers of the animal. The volume of injection solution was 150 μ l. The animals of the control group ($n = 7$) were injected with physiological saline according to the described scheme. Immune response was investigated two weeks after the last injection. Throughout the experiment, constant monitoring of physiological state of the experimental animals was carried out. Two weeks after the last injection, laboratory animals of the experimental and control groups were anesthetized by inhalation of isoflurane (Piramal Enterprises, India) and blood samples were collected from the cardiac cavity to determine antibody titer in the blood serum. After collecting all individual blood samples in vacuum tubes, they were left at room temperature for 20 min, then centrifuged for 30 min at 1500g without refrigeration. Supernatant was transferred into sterile 1.5-ml tubes and frozen at -80°C .

Enzyme-linked immunosorbent assay. Presence of the specific immunoglobulins of class G (IgG) was determined in serum samples of mice after three-injection immunization. The proteins used for immunization, the formalin-inactivated SARS-CoV-2, and the RBD Spike protein of the SARS-CoV-2 were adsorbed onto the plates as antigens. Virus preparation was obtained by propagation of SARS-CoV-2 in a Vero cell culture (ATCC bank, CCL 81 line) for 4 days. Then the supernatant was collected and centrifuged at room temperature at 3500 rpm (1900g) for 10 min. To inactivate the virus,

formalin (37% formaldehyde) was added to the supernatant at a ratio of 1 : 1000. After incubation of the virus with formalin for 96 h, ultracentrifugation was performed in a JA 30.50 rotor at 28,700 rpm ($\sim 100,000g$) for 3 h at 4°C through a 30% sucrose pad. Supernatant was removed, and pellet was resuspended in Dulbecco's phosphate buffered saline (DPBS). Protein concentration was measured with a Qubit fluorometer (Invitrogen, USA) using the Quant-iT Protein Assay Kit (Invitrogen) according to the manufacturer's recommendations. For ELISA, a preparation of the formalin-inactivated virus with protein concentration of 1 $\mu\text{g/ml}$ was used.

Sorption of antigens diluted in DPBS to a concentration of 1.0 $\mu\text{g/ml}$ was carried out overnight at 4°C in 96 well Maxisorp plates (Thermo Scientific, Denmark) (100 μ l per well). Then the plates were washed twice with a solution of 0.1% Tween 20 (Serva, Germany) in sodium phosphate buffer (PBS) using an ELx50 automatic washer (BioTek, USA), 500 μ l per well. After washing, the plates were incubated with blocking buffer containing PBS and 5% dry milk powder (TF Ditol, Russia) for 2 h at room temperature, then washed again. After that, a series of tenfold dilutions of experimental mice sera were added to the plate. Plates with serum were incubated for 1 h at room temperature, then washed 3 times. Goat anti-Mouse IgG (H + L) labeled with horseradish peroxidase (Invitrogen) were used as secondary antibodies at a dilution of 1 : 1200, 100 μ l per well. After incubation with secondary antibodies, the plates were washed 4 times. A solution of tetramethylbenzidine (TMB) (Bioservice, Russia) was added as a chromogenic substrate, 100 μ l per well, followed by incubation for 10 min at room temperature, then the reaction was stopped by adding 1 N H_2SO_4 . Optical density at a wavelength of 450 nm was measured with a CLARIOstar microplate reader equipped with LVF-monochromators (based on linear variable filters) (BMG Labtech, USA).

Statistical analysis. Statistical analyses were performed using the Statistica 12.0 software package (Statsoft, USA). Data were presented as mean \pm standard deviation (SD) or as geometric mean \pm geometric standard deviation of the logarithm 2 base of the antibody titer. Distribution normality was tested using the Kolmogorov–Smirnov test. To assess statistical significance of the differences in the data, a two-way analysis of variance (one-way ANOVA) with Tukey's post hoc analysis was used. Differences were considered significant at $p < 0.05$.

RESULTS

Recombinant protein design. Four genetically engineered constructs encoding chimeric proteins Rop-D2-Rop-Tri-HBD, Rop-D3-Rop-Tri-HBD, Rop-D2-Rop-ALD-HBD, and Rop-D3-Rop-ALD-HBD were

```

S455-469          S475-499 E484K          N501Y
433 VIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVV 511
      * * * * *          * * *          * * * * *
# ##### ## #####          #####          ##### # #####
D2 TEIYQAGSTPCNGVEGFNCYFP
D3 YRLFRKSNLKPSERDISTEIYQAGSTPCNGVEGFNCYFPLQS

```

Fig. 1. RBM sequence of the Spike protein of the SARS-CoV-2 and the sequence of determinants D2 and D3. Bold font denotes residues that are included in the linear epitopes S455-469 and S475-499 (from the article by Lu et al. [7]). Variable amino acid residues E484 and N501 are in italic. Asterisks mark amino acid residues interacting with ACE2, and hash signs indicate amino acid residues interacting with at least one of 13 neutralizing antibodies with a known three-dimensional structure (in accordance with Fig. 2 from Lee et al. [11]). Interaction criterion – the distance between residues is not more than 4.5 Å.

obtained, immunogenic properties of which were planned to be investigated. The proteins contain the selected regions D2 or D3 of the RBM Spike protein of the SARS-CoV-2 that have many amino acid residues in their composition forming contacts with ACE2 and neutralizing antibodies (Fig. 1). We will further call these regions antigenic determinants or epitopes D2 and D3. The sequences of the selected determinants overlap: D2 corresponds to 470-490 aa, and the D3 sequence corresponds to 453-494 aa of the Spike protein of the SARS-CoV-2 (UniProtKB ID: locus SPIKE_SARS2, accession P0DTC2) (Fig. 1). In the D3 sequence the hydrophobic phenylalanine residue at position 464 in the native structure of the Spike protein, which has its hydrophobic radical facing towards interior of the molecule and not forming contacts with antibodies, was replaced by a hydrophilic serine residue in order to prevent possible formation of the aggregates of vaccine proteins with D3 due to hydrophobic interactions. Glutamic acid residue at position 484 in many strains of the SARS-CoV-2 virus, in particular in the South African line B.1.351 was replaced by a lysine residue. Within the selected determinants, there are two cysteine residues, at positions 480 and 488, which form a disulfide bond in the native structure of the Spike protein.

Both determinants were chosen in such a way that their N- and C-terminus were brought close together in the three-dimensional structure of the Spike protein (Fig. 2), and this position could be fixed due to additional scaffold structure attached to them. The epitope scaffold in all four proteins is represented by two α -helices of the Rop-like protein from *M. capsulatus* contacting with each other (2-34 and 35-66 aa, PDB code: 2JS5_A, RefSeq ID WP_010959602). Between 34 and 35 aa of the Rop-like protein, the D2 or D3 sequences were inserted. Thus, in the protein sequences, both determinants are flanked by α -helices of the Rop-like protein. In addition to the proximity of the ends, the loop-like conformation of both determinants should also be maintained by a disulfide bond between the cysteine residues 480 and 488. The presence of an additional disulfide bond allows expecting good correspondence of the native conformation of the regions of the selected determinants to the

native structure of these regions in the SARS-CoV-2 Spike protein.

Domains providing trimerization are attached to the C-terminus of the Rop-D2/D3-Rop hybrid molecules through the linker sequence, or α -helix sequence mediating trimerization of the SARS-CoV-2 (Tri) Spike protein (958-991 aa, identifier UniProtKB: locus SPIKE_SARS2, accession P0DTC2), or the sequence of aldolase from *T. maritima* (2-201 aa, PDB code: 1WA3). At the C-terminus of all four proteins a short region of the heparin-binding domain (HBD) of the Heparin-binding hemagglutinin (HBHA) protein from *M. tuberculosis* (160-174 aa, UniProtKB/Swiss-Prot ID: A1KFU9) is inserted, which allows purification of the obtained proteins on the heparin-containing sorbents. Quaternary structure of the obtained proteins was not characterized, but, presumably, they should spontaneously form homotrimers in solution, each with three identical loop-shaped structural epitopes in the N-terminal part. Schematic diagram of the presumed structure of the trimers of the obtained fusion proteins is shown in Fig. 2.

Molecular weight of the monomeric proteins – from 18.7 to 38.7 kDa – is already considered sufficient for effective antibody production. Increase in molecular weight during trimerization should further enhance immune response to the hybrid antigens.

To confirm existence of a disulfide bond in the proteins containing loop-like Spike protein fragments with Cys480 and Cys488 and the proteins obtained by synthesis in *E. coli*, a protein containing a full-length RBM flanked by sequences from the Rop-like protein was designed. The receptor-binding motif in the native Spike protein forms a loop-like structure that includes loop-like epitopes D2 and D3 at the top of the hairpin. The cysteines at positions 480 and 488 are included in all three fragments – RBM, D2, and D3.

Preparation and purification of recombinant proteins.

Cloning of the genes of the recombinant proteins was carried out according to the “Materials and Methods” section. All engineered constructs provided high level of gene expression. The proteins Rop-D2-Rop-Tri-HBD, Rop-D3-Rop-Tri-HBD, and Rop-RBM-Rop were produced in the form of inclusion bodies, while the proteins Rop-

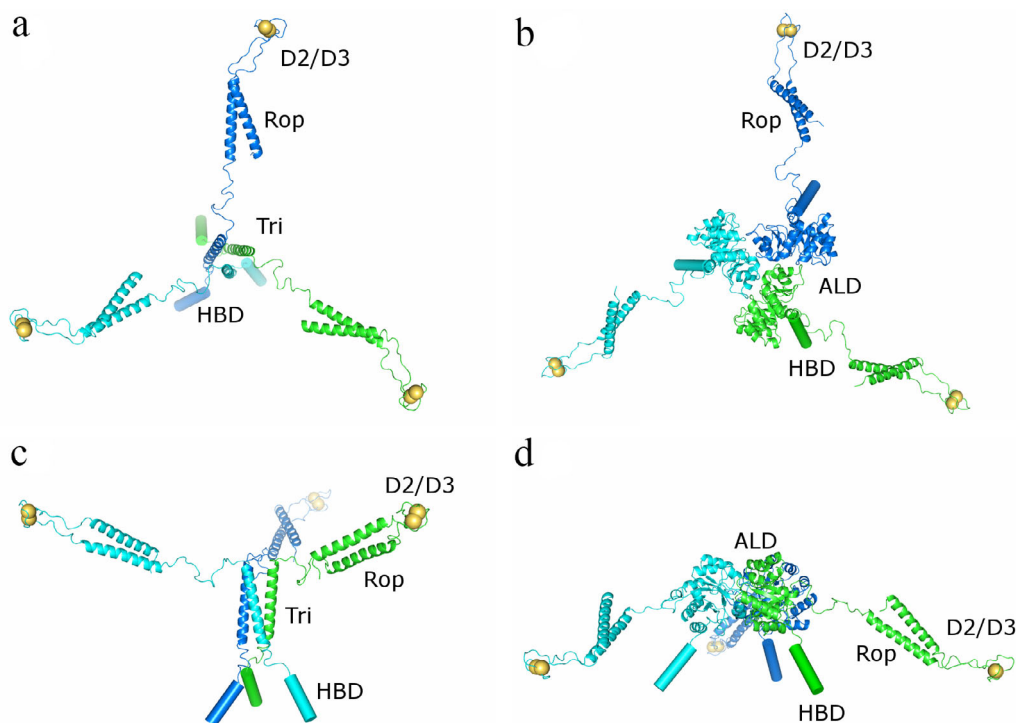


Fig. 2. Three-dimensional structure of Rop-D2/D3-Rop-Tri-HBD (a and c) and Rop-D2/D3-Rop-ALD-HBD (b and d) proteins. 3D Models of the structure of recombinant proteins are constructed in the PyMOL program (<https://pymol.org/2/>), both top view (a and b) and side view (c and d) are presented. To construct the models, we used the structures of a Rop-like protein from *M. capsulatus* (PDB code: 2JS5_A, 2-66 aa), RBD of the Spike protein of SARS-CoV-2 (PDB code: 6VW1, loop D2 is shown (470-490 aa); loop D3 (453-494 aa), which has an elongated loop-like conformation, is not shown), α -helix of the trimerizer from the SARS-CoV-2 (Tri) Spike protein (PDB code: 1ZVB, 2-34 aa), and aldolase *T. maritima* (PDB code: 1WA3, 2-201 aa). Proteins are depicted as ribbon models; cysteines forming disulfide bonds – in the form of spheres; heparin-binding domain (HBD) of the heparin-binding hemagglutinin protein from *M. tuberculosis* (160-174 aa, UniProtKB/Swiss-Prot ID: A1KFU9) – in the form of cylinders (not to scale). Different monomers are shown in different colors.

D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD were synthesized in a soluble form. Accordingly, isolation schemes for these proteins differed: proteins with a trimerizer from SARS-CoV-2 and Rop-RBM-Rop were isolated under denaturing conditions in the presence of urea, while proteins with aldolase were isolated under non-denaturing conditions (Fig. 3). WorkBeads 40S sorbent was used as the first stage of purification for all proteins. At the second stage, four proteins containing the heparin-binding domain were purified on the heparin sepharose and further used for immunization.

The second stage of Rop-RBM-Rop protein purification was carried out using WorkBeads 40 DEAE, most of the impurity proteins were strongly bound to the sorbent in the process, and major amount of the target protein did not bind to the sorbent and, after passing through the column, came out in a highly purified and sufficiently concentrated form (Fig. 3f, lane 2). The yield of proteins was 4.8-8.0 mg per 1 g of raw biomass, and the degree of purification was 97-98%.

Characteristics of proteins, including experimental data and theoretical values of some parameters obtained

using the Internet resource <https://web.expasy.org/protparam/>, are presented in Table 1.

Identification of the Rop-RBM-Rop protein and confirmation of the presence of a disulfide bond in the Rop-RBM-Rop and other recombinant proteins. To identify the protein under study, the bands of interest from PAGE were subjected to trypsinolysis by adding and not adding the reducing agent DTT and iodoacetamide (IAA) before the protein proteolysis reaction in the gel. In the presence and absence of DTT and IAA, the S-S bonds in the protein should, respectively, be restored (destroyed) or retained. The MALDI mass spectrum of the obtained peptides is shown in Fig. S1, b and c in the Supplement.

Protein preparation treated with DTT and IAA produced mass spectrum (Fig. S1c in the Supplement) that contained several peaks corresponding to tryptic and chymotryptic cysteine-containing peptides of the Rop-RBM-Rop protein, while cysteines in these peptides were present only in the form of S-carbamidomethylcysteines. Thus, the peak with m/z 2145.0365 corresponds to the peptide DISTEIQAGSTPCNGVEGF (77-96 aa), the peak with m/z 2582.0867 corresponds to DIS-

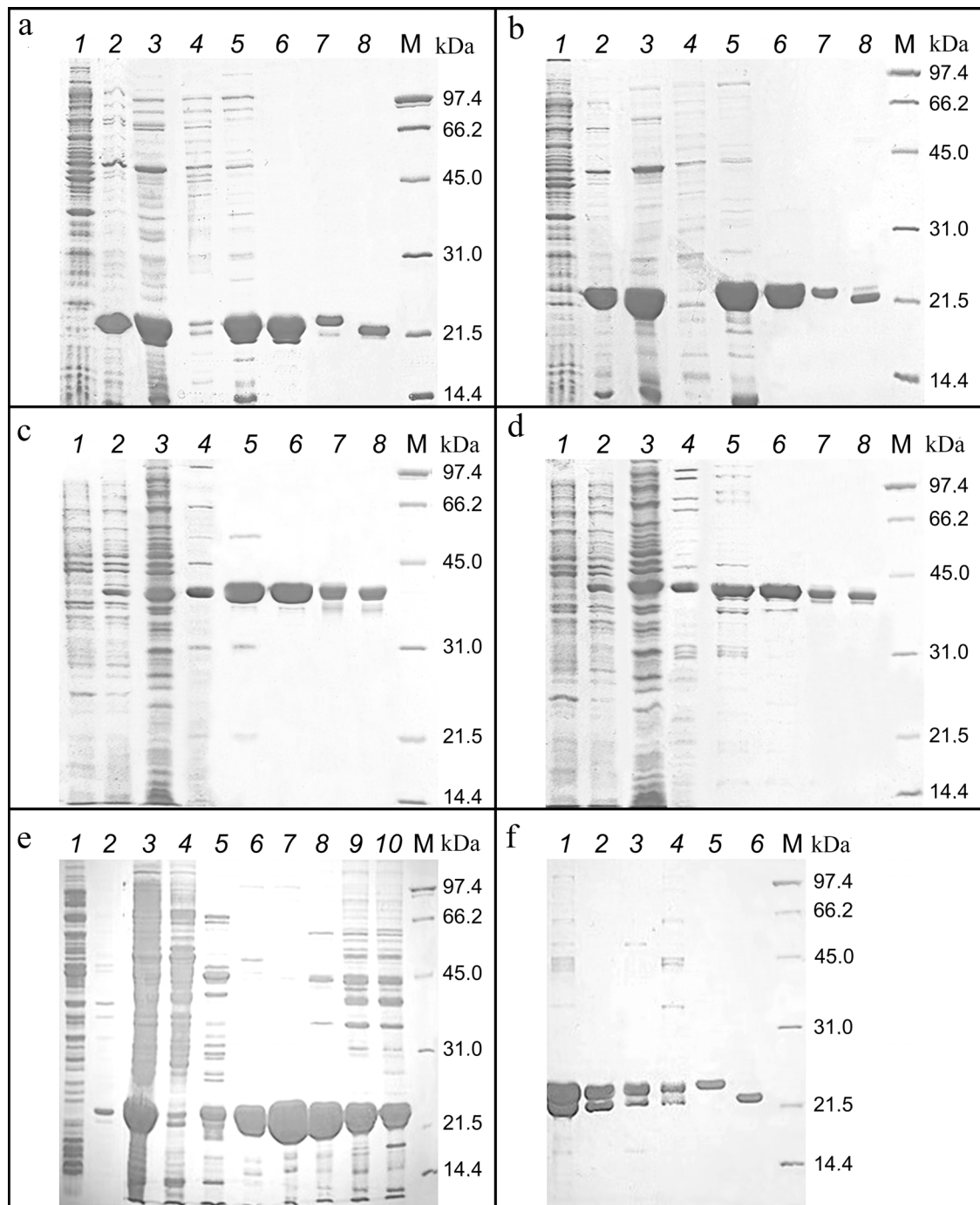


Fig. 3. Electrophoregrams of recombinant protein preparations separated by 12% PAGE. a and b) Purification and analysis of Rop-D2-Rop-Tri-HBD and Rop-D3-Rop-Tri-HBD proteins. Lanes: 1) cell proteins before IPTG induction; 2) cell proteins after IPTG induction; 3) inclusion bodies after lysis of biomass of producer strains; 4) supernatant after lysis of biomass of producer strains; 5) combined fractions obtained by elution from the WorkBeads 40S sorbent; 6) pooled fractions obtained by elution with heparin-sepharose; 7) protein samples in the presence of a reducing agent DTT; 8) protein samples without DTT. c and d) Purification and analysis of Rop-D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD proteins. Lanes: 1) cell proteins before IPTG induction; 2) cell proteins after IPTG induction; 3) supernatant after lysis of biomass of producer strains; 4) heated supernatant after biomass lysis; 5) combined fractions obtained by elution from the WorkBeads 40S sorbent; 6) pooled fractions obtained by elution with heparin-sepharose; 7) protein samples in the presence of DTT; 8) protein samples without DTT. e) Purification and analysis of the Rop-RBM-Rop protein. Lanes: 1) supernatant after lysis of the biomass of the producer strain; 2) sediment after lysis of the biomass of the producer strain; 3) proteins from the sediment dissolved in 8 M urea; 4) proteins from the sediment not bound to the WorkBeads 40S sorbent; 5-10) fractions after elution. f) Purification and analysis of the Rop-RBM-Rop protein. Lanes: 1) pooled fractions after purification on WorkBeads 40S; 2) protein not bound to WorkBeads 40 DEAE; 3 and 4) fractions after elution; 5) protein in the presence of DTT; 6) protein without DTT; M – marker of molecular weight 14.4-97.4 kDa (Bio-Rad, USA).

Table 1. Characteristics of the recombinant proteins obtained in the work

Parameter	Rop-D2-Rop-Tri-HBD	Rop-D3-Rop-Tri-HBD	Rop-D2-Rop-ALD-HBD	Rop-D3-Rop-ALD-HBD	Rop-RBM-Rop
Molecular weight, Da	18,714	21,359	36,102	38,747	20,550
Protein localization after destruction of the biomass	inclusion bodies	inclusion bodies	soluble fraction	soluble fraction	inclusion bodies
Theoretical isoelectric point (pI)	8.83	9.39	8.58	9.1	7.73
Absorption of an aqueous protein solution (1 mg/ml) at 280 nm for the fully oxidized protein form	0.325	0.355	0.486	0.491	1.071
Absorption of an aqueous protein solution (1 mg/ml) at 280 nm for the reduced protein form	0.318	0.349	0.483	0.488	1.065

Table 2. Geometric mean titer of serum antibodies to various antigens in mouse sera after triple immunization (1/log₂ titer)

Group of mice	Protein	Antigen used in immunization	Formalin-inactivated SARS-CoV-2	RBD
1	Rop-D2-Rop-Tri-HBD	14.21 ± 1.04	8.64 ± 1.00	10.73 ± 1.12
2	Rop-D3-Rop-Tri-HBD	13.48 ± 1.19	9.67 ± 1.18	11.44 ± 1.19
3	Rop-D2-Rop-ALD-HBD	> 15.64	9.88 ± 1.12	11.36 ± 1.17
4	Rop-D3-Rop-ALD-HBD	> 15.64	9.75 ± 1.10	12.61 ± 1.08
5	Saline	< 7.0	< 7.0	< 7.0

TEIQAGSTPCNGVEGFNCY (77-99 aa, this peptide contains two cysteine residues), peak with m/z 1191.5626 – NCYFPLQSY (97-105 aa), peak with m/z 1323.5494 – QAGSTPCNGVEGF (84-96 aa). The sequences of peptides (84-96 aa) and (96-105 aa) were additionally confirmed by collision-induced dissociation (CID) (data not shown). In general, coverage of the sequence with peptides (Fig. S1a in the Supplement) was 100%, which is a good result and fully confirms the structure of the studied protein.

When comparing mass spectra of the hydrolysates of the Rop-RBM-Rop samples (Fig. S1, b and c in the Supplement), it was shown that in the sample analyzed under non-denaturing conditions, no peaks corresponding to peptides containing cysteine residues in the form of S-carbamidomethylcysteines, S-propionamidocysteines or peptides with free unmodified cysteine residues were detected, while there were peaks confirming the presence of a disulfide bond in the protein under study. Detailed analysis of the peaks corresponding to the cysteine-containing peptides identified in the Rop-RBM-

Rop-DTT-IAA protein with a disulfide bond formed between the cysteines is presented in Table S1 in the Supplement.

The peaks with m/z 2631.1467; 3791.486; 2561.1583 and 1645.6688 correspond to two possible structures of peptides, however, they have the same empirical formula and differ only in the position at which the cleavage of the polypeptide chain by trypsin took place (italicized in Table S1 in the Supplement). The structure of ions with m/z 2561.044; 2865.261 and 3791.486 was additionally confirmed by the tandem mass spectrometry. The collision-induced dissociation (CID) mass spectra (Fig. S2 in the Supplement) contain peaks corresponding to the a-, y-, and b-series of fragments of the corresponding parent ions, the structures of which were assumed based on the analysis of the mass spectrum of the tryptic hydrolysate (Fig. S1c; Table S1 in the Supplement) – this fact additionally confirms correctness of the determined structure of these ions.

Thus, mass spectrometry showed the presence of S–S bond between the cysteines 90 and 98 (480 and 488)

in the Rop-RBM-Rop protein preparation obtained by synthesis in *E. coli* cells in the form of inclusion bodies, followed by chromatographic purification under denaturing conditions and refolding by dialysis.

The rest of the proteins were not analyzed with mass spectrometry. Formation of the disulfide bond in the isolated proteins was detected from the difference in electrophoretic mobility of the bands corresponding to the reduced (heated at 95°C in the presence of the reducing agent DTT) and unreduced forms of the protein. In the Rop-RBM-Rop protein, for which the presence of disulfide bond was shown by mass spectrometry, the reduced form has lower mobility in the gel (Fig. 3). Similar behavior is typical for all four proteins with determinants D2 and D3 (Fig. 3). This confirms the presence of disulfide bond in the isolated proteins.

Antibody formation after immunization. Immunogenic compositions based on four proteins were used to immunize BALB/c inbred mice. Two weeks after the last of three injections administered with 2-week interval, presence of the specific immunoglobulins of class G (IgG) in the serum samples of mice was determined. The antigens used were the same proteins that were used for immunization, the SARS-CoV-2 inactivated by formalin, and the RBD of the Spike protein of the SARS-CoV-2. The results of immunization are presented in Table 2.

As can be seen from the data presented in Table 2, all four proteins induced formation of IgG antibodies in mice at a sufficiently high titer interacting both with the antigen itself used for immunization, with the inactivated virus, and with RBD. The most immunogenic proteins are Rop-D2-Rop-ALD-HBD and Rop-D3-Rop-ALD-HBD. The least immunogenic is the Rop-D3-Rop-Tri-HBD protein. The most pronounced response to RBD was observed in the case of the Rop-D3-Rop-ALD-HBD protein. The most pronounced response to the inactivated SARS-CoV-2 was observed in the group of animals immunized with the Rop-D2-ALD-HBD protein. The described comparisons can be considered as observed trends, since the sample size used in the experiment (seven animals per group) is insufficient to provide statistical significance (two-way ANOVA using the Tukey post-hoc, $p < 0.05$).

DISCUSSION

High titers of antibodies interacting with RBD obtained in a eukaryotic producer, as well as with inactivated SARS-CoV-2, indicate that the recombinant proteins obtained in this work, containing short RBM fragments corresponding to conformational epitopes D2 and D3, can effectively induce production of the specific antibodies. One of the important circumstances, which probably largely determines efficiency of antibody production, is that there are no glycosylation sites within the selected

determinants in the Spike protein, therefore, the virus surface regions and RBD in the region of the D2 and D3 determinants are not modified by sugar residues similar to the epitopes D2 and D3 in the proteins obtained in our work and synthesized in bacterial cells. In our case, attachment of the N- and C-terminus of the D2 and D3 determinants to the α -helices of the Rop-like scaffold protein, apparently, provides high conformational similarity between the D2 and D3 loops in the recombinant proteins and in the native Spike protein. This is also evidenced by the formation of characteristic disulfide bond, directly demonstrated for the Rop-RBM-Rop protein using mass spectrometry and indirectly for all four proteins used for immunization (from the reduced mobility of the reduced form during PAGE). In the future, it is planned to obtain other recombinant proteins based on the platform developed in this work that include promising viral epitopes that are not necessarily loop-shaped. If these epitopes have any other stable conformation, for example, α -helical, or represent β -sheet regions, flexible glycine-containing linkers will be used to attach them to the Rop-like scaffold protein. Conformation of the epitopes can be additionally stabilized by replacing amino acid residues that do not interact with antibodies and the receptor with cysteine residues, which will ensure proximity of their protein chains via formation of disulfide bonds. This approach was implemented in the work by Zuniga et al. [12] during development of the epitope vaccine based on short protein fragments of the respiratory syncytial virus (RSV) conjugated with synthetic nanoparticles.

The presence of trimerization domains in proteins probably leads to additional enhancement of the immune response due to increased size of the antigen. The maximum effect is observed in the case of proteins with aldolase (Fig. 2). The ability of aldolases to spontaneously multimerize is often used to obtain candidate vaccines based on protein nanoparticles, in particular, vaccines based on the RBD of coronaviruses [13-15]. Aldolase used in our work is a rather large protein compared to viral determinants, so it can be assumed that among the antibodies produced, a significant part of antibodies will be produced specifically against aldolase and other additional domains of the recombinant proteins. Nevertheless, in numerous works on the creation of vaccines based on the viral and other proteins capable of self-assembly that form huge protein complexes (for example, consisting of 24 or 60 subunits), the formation of cross-specific or autoreactive antibodies to the protein components of the carrier has not been reported [16]. However, in this work, in addition to aldolase, we also used another trimerizer – α -helix from the Spike protein of the SARS-CoV-2, which is much smaller in size and, therefore, should “pull” the “antibody response” to itself to a lesser extent, besides, it originates from the same virus as the target antigens. A similar trimerizer (27 aa foldon domain

of fibrin from bacteriophage T4) has been used for a long time to obtain candidate vaccines based on the viral proteins, in particular, it is used in the previously developed vaccines against SARS-CoV and MERS [17, 18]. These vaccines and other protein vaccines against coronaviruses [13, 15] contain large fragments or the whole Spike protein of coronaviruses synthesized in eukaryotic expression systems as an antigenic component. All of them have high immunogenicity and, to a greater or lesser extent, cause production of the virus-neutralizing antibodies. The aim of this work was to obtain short determinants of the SARS-CoV-2 Spike protein (epitopes) with a native conformation in *E. coli*, to study their immunogenicity and ability of antibodies formed in response to them to interact with RBD protein and inactivated virus. Successful implementation of this stage of the study allows us to proceed to the study of the virus neutralization effect of the obtained antibodies, expanding the list of studied determinants, and to the subsequent stages of developing of epitope vaccine.

It should be noted that within both selected determinants in the RBD sequence there is a glutamic acid residue E484, which, when replaced with lysine, provides lower binding efficiency with monoclonal antibodies and sera of patients who have been infected and vaccinated [19-21]. This mutation is present in new variants of the SARS-CoV-2, derived from the B.1.1.28 lineage, first described in Brazil and Japan (P.2 and P.1), as well as in the strains of the B.1.351 line, also known as 501Y.V2, first identified in South Africa, and now widespread [22]. Based on the platform developed in the article, it is technically easy to obtain variants of immunogenic recombinant proteins containing the E484K substitution. Introduction of both variants of proteins into the vaccine composition should ensure production of the antibodies that bind to both the original and mutant variants of the virus.

The approach closest to the one proposed in our work was implemented in the study by Lu et al. [7], within which the surface B-cell epitopes of the Spike protein were predicted and synthesized, and vaccines based on the virus-like particles (VLPs) of the hepatitis B virus core protein (HBc) were created using the SpyCatcher/SpyTag system [7]. Immunogenicity of the epitopes was tested using sera from mice 10 days after three-injection immunization with two-week intervals. For epitope S455-469, which is part of the D3 sequence, and epitope S475-499, which is part of both D2 and D3 (Fig. 1), high immunogenicity was shown in the ELISA experiments using peptides as antigens (antibody titer in sera of mice $>10^4$).

Lu et al. [7] used short linear synthetic peptides, their density on the virus-like particles and size of the particles themselves are probably higher in comparison with the trimeric hybrid proteins obtained in our work. Nevertheless, when using our approach, the antibody titer

in the sera of mice also exceeds 10^4 . In our study, activity of the mouse sera against RBM and inactivated virus was also shown, which indicated native conformation of the loop-shaped epitopes obtained by us. In the study by Lu et al. [7] neutralizing activity of the generated antibodies against pseudoviruses for epitopes S455-469 and S475-499 of variants D614 and G614 was also shown. This suggests that immunization with proteins with epitopes D2 and D3 will also lead to the formation of virus-neutralizing antibodies against both variants of viruses.

Thus, on the basis of fusion proteins, including the Rop-like protein from *M. capsulatus* and trimerization domains, aldolase from *T. maritima* or α -helix of the Spike protein of SARS-CoV-2, a platform has been created for obtaining immunogenic constructs containing conformational epitopes with continuous amino acid sequence, in particular, the epitopes of the SARS-CoV-2 proteins, which can be used to create new variants of epitope vaccines for prevention of COVID-19. Future work would involve expansion of the spectrum of the investigated determinants and characterization of their ability to induce formation of the neutralizing antibodies. Production of the proteins via microbiological synthesis is an inexpensive alternative to the production of proteins by synthesis in eukaryotic cells. The choice of epitopes, significant part of which are exposed on the protein surface, including amino acid residues involved in interaction with neutralizing antibodies and not containing glycosylation sites, seems to be the optimal strategy.

Funding. This work was supported by the Russian Foundation for Basic Research (grant no. 20-04-60373\20 “Development of approaches for design of a new type of preventive vaccine against COVID-19 based on the use of recombinant SARS-CoV-2 antigens as a part of chimeric protein nanoparticles”).

Ethics declarations. The authors declare no conflict of interest in financial or any other sphere. All applicable international, national and/or institutional guidelines for the care and use of animals have been followed.

Supplementary information. The online version contains supplementary material available at <https://doi.org/10.1134/S0006297921100096>.

REFERENCES

1. Burton, D. R. (2002) Antibodies, viruses and vaccines, *Nat. Rev. Immunol.*, **9**, 706-713, doi: 10.1038/nri891.
2. Neu, K. E., Henry Dunand, C. J., and Wilson, P. C. (2016) Heads, stalks and everything else: how can antibodies eradicate influenza as a human disease? *Curr. Opin. Immunol.*, **42**, 48-55, doi: 10.1016/j.coi.2016.05.012.
3. Sok, D., and Burton, D. R. (2018) Recent progress in broadly neutralizing antibodies to HIV, *Nat. Immunol.*, **19**, 1179-1188, doi: 10.1038/s41590-018-0235-7.

4. Dejnirattisai, W., Supasa, P., Wongwiwat, W., Rouvinski, A., Barba-Spaeth, G., et al. (2016) Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with zika virus, *Nat. Immunol.*, **17**, 1102-1108, doi: 10.1038/ni.3515.
5. Yuan, M., Liu, H., Wu, N. C., Lee, C. D., Zhu, X., et al. (2020) Structural basis of a shared antibody response to SARS-CoV-2, *Science*, **369**, 1119-1123, doi: 10.1126/science.abd2321.
6. Xu, K., Acharya, P., Kong, R., Cheng, C., Chuang, G. Y., et al. (2018) Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1, *Nat. Med.*, **24**, 857-867, doi: 10.1038/s41591-018-0042-6.
7. Lu, S., Xie, X. X., Zhao, L., Wang, B., Zhu, J., et al. (2021) The immunodominant and neutralization linear epitopes for SARS-CoV-2, *Cell Rep.*, **34**, 108666, doi: 10.1016/j.celrep.2020.108666.
8. Juraja, S. M., Mulhern, T. D., Hudson, P. J., Hattarki, M. K., Carmichael, J. A., and Nuttall, S. D. (2006) Engineering of the *Escherichia coli* Im7 immunity protein as a loop display scaffold, *Protein Eng. Des. Sel.*, **19**, 231-244, doi: 10.1093/protein/gzl005.
9. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes, *Nat. Protoc.*, **1**, 2856-2860, doi: 10.1038/nprot.2006.468.
10. Grunina, T. M., Demidenko, A. V., Lyaschuk, A. M., Poponova, M. S., Galushkina, Z. M., et al. (2017) Recombinant human erythropoietin with additional processable protein domains: purification of protein synthesized in *Escherichia coli* heterologous expression system, *Biochemistry (Moscow)*, **82**, 1285-1294, doi: 10.1134/S0006297917110062.
11. Lee, S.-Y., Kim, C., Ryu, D.-K., Lee, J., Kim, Y.-I., et al. (2020) A novel neutralizing antibody targeting receptor binding domain of SARS-CoV-2, *Res. Square*, preprint, doi: 10.21203/rs.3.rs-59639/v1.
12. Zuniga, A., Rassek, O., Vrohings, M., Marrero-Nodarse, A., Moehle, K., et al. (2021) An epitope-specific chemically defined nanoparticle vaccine for respiratory syncytial virus, *Vaccines*, **6**, 85, doi: 10.1038/s41541-021-00347-y.
13. Cohen, A. A., Gnanapragasam, P., Lee, Y. E., Hoffman, P. R., Ou, S., et al. (2021) Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice, *Science*, **371**, 735-741, doi: 10.1126/science.abf6840.
14. Kang, Y. F., Sun, C., Zhuang, Z., Yuan, R. Y., Zheng, Q., et al. (2021) Rapid Development of SARS-CoV-2 spike protein receptor-binding domain self-assembled nanoparticle vaccine candidates, *ACS Nano*, **15**, 2738-2752, doi: 10.1021/acsnano.0c08379.
15. Tan, T. K., Rijal, P., Rahikainen, R., Keeble, A. H., Schimanski, L., et al. (2021) A COVID-19 vaccine candidate using SpyCatcher multimerization of the SARS-CoV-2 spike protein receptor-binding domain induces potent neutralising antibody responses, *Nat. Commun.*, **12**, 542, doi: 10.1038/s41467-020-20654-7.
16. Nguyen, Q. D., Kikuchi, K., Maity, B., and Ueno, T. (2021) The versatile manipulations of self-assembled proteins in vaccine design, *Int. J. Mol. Sci.*, **22**, 1934, doi: 10.3390/ijms22041934.
17. Li, J., Ulitzky, L., Silberstein, E., Taylor, D. R., and Viscidi, R. (2013) Immunogenicity and protection efficacy of monomeric and trimeric recombinant SARS coronavirus spike protein subunit vaccine candidates, *Viral Immunol.*, **26**, 126-132, doi: 10.1089/vim.2012.0076.
18. Tai, W., Zhao, G., Sun, S., Guo, Y., Wang, Y., et al. (2016) A recombinant receptor-binding domain of MERS-CoV in trimeric form protects human dipeptidyl peptidase 4 (hDPP4) transgenic mice from MERS-CoV infection, *Virology*, **499**, 375-382, doi: 10.1016/j.virol.2016.10.005.
19. Baum, A., Fulton, B. O., Wloga, E., Copin, R., Pascal, K. E., et al. (2020) Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies, *Science*, **369**, 1014-1018, doi: 10.1126/science.abd0831.
20. Greaney, A. J., Starr, T. N., Gilchuk, P., Zost, S. J., Binshtein, E., et al. (2021) Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition, *Cell Host Microbe*, **29**, 44-57.e9, doi: 10.1016/j.chom.2020.11.007.
21. Greaney, A. J., Loes, A. N., Crawford, K. H. D., Starr, T. N., Malone, K. D., et al. (2021) Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies, *Cell Host Microbe*, **29**, 463-476.e6, doi: 10.1016/j.chom.2021.02.003.
22. Tegally, H., Wilkinson, E., Lessells, R. J., Giandhari, J., Pillay, S., et al. (2021) Sixteen novel lineages of SARS-CoV-2 in South Africa, *Nat. Med.*, **27**, 440-446, doi: 10.1038/s41591-021-01255-3.