

## RESEARCH ARTICLE

# High-level expression of the monomeric SARS-CoV-2 S protein RBD 320-537 in stably transfected CHO cells by the *EEF1A1*-based plasmid vector

Maria V. Sinegubova<sup>1</sup>, Nadezhda A. Orlova<sup>1\*</sup>, Sergey V. Kovnir<sup>1</sup>, Lutsia K. Dayanova<sup>2</sup>, Ivan I. Vorobiev<sup>1,2</sup>

**1** Laboratory of Mammalian Cell Bioengineering, Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia, **2** Laboratory of Glycoproteins Biotechnology, Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Moscow, Russia

\* [nobiol@gmail.com](mailto:nobiol@gmail.com)



## OPEN ACCESS

**Citation:** Sinegubova MV, Orlova NA, Kovnir SV, Dayanova LK, Vorobiev II (2021) High-level expression of the monomeric SARS-CoV-2 S protein RBD 320-537 in stably transfected CHO cells by the *EEF1A1*-based plasmid vector. PLoS ONE 16(2): e0242890. <https://doi.org/10.1371/journal.pone.0242890>

**Editor:** Paulo Lee Ho, Instituto Butantan, BRAZIL

**Received:** November 19, 2020

**Accepted:** January 13, 2021

**Published:** February 2, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0242890>

**Copyright:** © 2021 Sinegubova et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files. Plasmids are deposited in GenBank - p1.1-Tr2-eGFP [GenBank: MW187857];

## Abstract

The spike (S) protein is one of the three proteins forming the coronaviruses' viral envelope. The S protein of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has a spatial structure similar to the S proteins of other mammalian coronaviruses, except for a unique receptor-binding domain (RBD), which is a significant inducer of host immune response. Recombinant SARS-CoV-2 RBD is widely used as a highly specific minimal antigen for serological tests. Correct exposure of antigenic determinants has a significant impact on the accuracy of such tests—the antigen has to be correctly folded, contain no potentially antigenic non-vertebrate glycans, and, preferably, should have a glycosylation pattern similar to the native S protein. Based on the previously developed p1.1 vector, containing the regulatory sequences of the Eukaryotic translation elongation factor 1 alpha gene (*EEF1A1*) from Chinese hamster, we created two expression constructs encoding SARS-CoV-2 RBD with C-terminal c-myc and polyhistidine tags. RBDv1 contained a native viral signal peptide, RBDv2—human tPA signal peptide. We transfected a CHO DG44 cell line, selected stably transfected cells, and performed a few rounds of methotrexate-driven amplification of the genetic cassette in the genome. For the RBDv2 variant, a high-yield clonal producer cell line was obtained. We developed a simple purification scheme that consistently yielded up to 30 mg of RBD protein per liter of the simple shake flask cell culture. Purified proteins were analyzed by polyacrylamide gel electrophoresis in reducing and non-reducing conditions and gel filtration; for RBDv2 protein, the monomeric form content exceeded 90% for several series. Deglycosylation with PNGase F and mass spectrometry confirmed the presence of N-glycosylation. The antigen produced by the described technique is suitable for serological tests and subunit vaccine studies.

p1.1-Tr2-RBDv1 [GenBank: MW187858]; pTM [GenBank: MW187855]; pTM-RBDv2 [GenBank: MW187856]; and available from Addgene: p1.1-Tr2-eGFP - #162782; pTM - #162783; pTM-RBDv2 - #162785.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** MVS and LKD declare that they have no competing interests. SVK, NAO and IIV are inventors of the patent RU2488633, which covers the use of the p1.1 plasmid; a derivative of this plasmid is used in current research. The existence of the patent does not alter our adherence to PLOS ONE policies on sharing data and materials.

**Abbreviations:** ABC, ammonium bicarbonate solution; ACE2, angiotensin-converting enzyme 2; ACN, acetonitrile; CHO, Chinese hamster ovary; DHFR, dihydrofolatereductase; EEF1A1, Eukaryotic translation elongation factor-1 alpha gene; EMCV, encephalomyocarditis virus; HT, hypoxanthine - thymidine; IDA, iminodiacetic acid; IRES, internal ribosome entry site; MTX, methotrexate; NTA, nitrilotriacetic acid; ORF, open reading frame; PNGase F, peptide:N-glycosidase F; RBD, receptor-binding domain; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; TFA, trifluoroacetic acid.

## Introduction

Humanity is faced with an unprecedented challenge—the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes a severe respiratory illness—coronavirus disease 2019 (COVID-19) pandemic. Countries were sent to lockdown; people could not make informed decisions about the possibility of social contacts; the need for diagnostic tests is very high. Existing tests for SARS-CoV-2 are reviewed in [1].

At the beginning of the pandemic, polymerase chain reaction (PCR) testing methods dominated since such test systems can be developed urgently, soon after the emergence of a new virus in the population. Serological testing makes it possible to reliably determine whether a person is infected with the SARS-CoV-2, even in the absence of disease symptoms, or long after the event of infection. Serologic tests are also needed to detect convalescent plasma of therapeutic interest and assess emerging vaccines' effectiveness.

The immunodominant antigen of SARS-CoV-2 is the receptor-binding domain (RBD) of the spike (S) protein [2]. Another antigen widely used for diagnostics—the nucleoprotein (N) protein—combines high sensitivity and low specificity. Cases are described for SARS-CoV when the results of testing with N-protein were clarified using two subunits of S-protein [3]. S-protein and its fragments can be used not only for routine testing for anti-SARS-CoV-2 antibodies presence but also for mapping of neutralizing antibodies epitopes, as was done for SARS-CoV [4] and for affordable surrogate virus neutralization tests, based, for example, on antibody-mediated blockage of ACE2–S interaction [5].

The S protein plays a crucial role in receptor recognition, cell membrane fusion, internalization of viruses, and their exit from the endosomes [6]. The S protein is co-translationally incorporated into the rough endoplasmic reticulum and is glycosylated by N-linked glycans. Interacting with the M and E proteins S protein trimer is transported to the virus's assembly site. S protein is required for cell entry but not necessary for virus assembly [7].

During their intracellular processing, S proteins of many types of coronaviruses, including SARS-CoV-2 and the Middle East respiratory syndrome-coronavirus (MERS-CoV), but not SARS-CoV, undergo partial proteolytic degradation at the furin signal protease recognition site with the formation of two subunits S1 and S2. The S protein homotrimer binds to the ACE2 dimer; a detailed study of this interaction is available [8].

A glycan shield is formed by N-linked glycans on the S protein surface, which is likely to help viral immune escape. In a comparative study of genome-wide sequencing data of natural isolates of SARS-CoV-2 [9] for the detected 228 variants of the S protein, all 22 potential N-glycosylation sites within the S protein's ectodomain were completely conserved, which confirms the importance of each of these sites for maintaining the integrity of the S protein oligosaccharide envelope.

Not all of these 22 potential N-glycosylation sites are occupied, for S1 and S2 subunits, obtained separately by the transiently transfected Human Embryonic Kidney cells 293 (HEK293) [10] N-glycosylation events were experimentally confirmed only for 17 out of 22 sites. At least one O-glycosylation site was experimentally found inside the RBD-domain area of the S1 subunit with the mucin-like oligosaccharide structures. Non-vertebrate cells can be used to produce the S protein or its fragments; in this case, N-glycans are present mostly in the form of bulky high mannose or paucimannose structures, excessively blocking the interaction of antibodies with the folded S protein [11]. Computational modeling of the glycan shield, performed for the HEK293-derived S protein, revealed that around 40% of the protein's surface is effectively shielded from IgG antibodies [12].

The use of full-length S protein for routine serological testing is nearly impossible due to its insolubility, caused by the transmembrane domain. An artificial trimer of its ectodomains was

used in serological tests [13]; however, such complex protein cannot be obtained in large quantities in mammalian cells. It is generally believed that the SARS-CoV-2 S protein receptor-binding domain is a minimal proteinaceous antigen, adequately resembling the immunogenicity of the whole spike protein. This domain contains only two occupied N-glycosylation sites [10] and 1–2 occupied O-glycosylation sites. It does not contribute to the trimer formation, and its surface is mostly unshielded.

Isolated RBD's of the S proteins of beta-coronaviruses were produced in various expression systems. Bacterial expression of the RBD from MERS-CoV produced no soluble target protein; refolding attempts also were unsuccessful [14]. Budding yeasts *Pichia pastoris* were the suitable host for the secretion of MERS-CoV RBD with at least two (from three) N-linked glycosylation sites present. Similar data were obtained for the RBD from SARS-CoV virus—removal of all N-glycosylation sites resulted in the sharp drop of protein secretion rate in the *P. pastoris* yeast; in the case of full RBD domain (residues 318–536), secretion of the un-glycosylated target protein was stopped completely [15]. It can be assumed that N-glycosylation is essential for correct folding of the RBD.

The SARS-CoV-2 S protein RBD was also completely insoluble upon *Escherichia coli* expression; solubilized inclusion bodies were unreactive even on blotting [16]. Hyperglycosylated yeast-derived SARS-CoV-2 RBD was obtained in reasonable quantities (50 mg/L in bioreactor culture) and successfully used for mice immunization [17]. Unfortunately, yeast-derived glycosylated proteins are not suitable for serological testing since they contain immunogenic glycans. Similarly, SARS-CoV-2 RBD produced in the *Nicotiana benthamiana* plant contains non-vertebrate N-glycans, potentially reactive with human antibodies [18].

Most of the early publications on the SARS-CoV-2 S protein and its RBD domain production in mammalian cells describe the transient transfection of HEK293 cells [10, 19] and purification of small protein lots in a very short time. For example, D. Stadlbauer [20] reports more than 20 mg/L titers in transiently transfected HEK-293 cells. The scalability of transiently transfected cell lines cultivation is still questionable; the gram quantities of RBD for large-scale serological testing can be produced only by stably transfected cell lines.

Previously we have developed the plasmid vector p1.1, containing large fragments of non-coding DNA from the *EEF1A1* gene of the Chinese hamster and fragment of the Epstein-Barr virus long terminal repeat concatemer [21] and employed it for unusually high-level expression of various proteins in Chinese hamster ovary (CHO) cells, including blood clotting factors VIII [22], IX [23], and heterodimeric follicle-stimulating hormone [24]. CHO cells were successfully used for transient SARS-CoV RBD expression at 10 mg/L secretion level [25]. We have proposed that SARS-CoV-2 RBD, suitable for in vitro diagnostics use, can be expressed in large quantities by stably transfected CHO cells, bearing the EEF1A1-based plasmid.

## Materials and methods

### Molecular cloning

p1.1-Tr2-RBDv1 construction. The RBD 319–541 coding sequence was synthesized according to [13], synthetic gene SARS\_CoV\_2RBD\_his [GenBank: MT380724.1]. The DNA fragment encoding the RBDv1 open reading frame (ORF) with Kozak consensus sequence and C-terminal c-myc and 6xHis tags were obtained by PCR using primers AD-COV-AbsF and AD-RBD-myc6HNheR (listed in Table 1) and Tersus polymerase mix (Evrogen, Moscow, Russia). Synthetic oligos, PCR reagents, Plasmid Miniprep Purification kit, PCR Clean-Up System were from Evrogen. The PCR product was restricted using *AbsI* (Sibenzyme, Novosibirsk, Russia) and *NheI* (Fermentas, Vilnius, Lithuania) enzymes and inserted into p1.1-Tr2 vector instead of eGFP (p1.1-Tr2-eGFP [GenBank: MW187857] is available from Addgene, plasmid

**Table 1. Primers used for molecular cloning.**

Primers for RBDv1 cloning, restriction sites are underlined	
AD-COV-AbsF	<u>AACTCGAGG</u> CCGCCACCATGTTTCATGCCTTCTT
AD-RBD-myc6HNheR	<u>GCTAGC</u> CTAATGGTGATGGTGATGATGACCGGTATGCATATTCAGATCCTCTTCTGAGATGAGTTTTTGGTTCGAAGTTCACGCATTTGTT
Primers for pTM construction, sticky ends of annealed pairs are underlined	
AS-Myc6H-AbsF	<u>TCGAG</u> GCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGTGTGGAGCAGTCTTCGTCTCGGCTA
AS-Myc6H-AbsR	<u>GCGC</u> TAGCCGAGACGAAGACTGCTCCACACAGCAGCAGCACACAGCAGAGCCCTCTCTTCATTGCATCCATGGTGGCGGCC
AS-Myc6H-SpeF	<u>GCGC</u> TACCGGTGAACAAAACTGATCAGCGAAGAGGATCTGTCTGCAGGCGGTTCATCACCATCACCATCACCATCACCATCA
AS-Myc6H-SpeR	<u>CTAG</u> TGATGGTGATGGTGATGGTGATGACCGCTGCAGACAGATCCTCTTCGCTGATCAGTTTTTGGTTCACCGGTA
Primers for RBDv2 cloning, restriction sites are underlined	
AD-SFR2-NheF	<u>GCTAGC</u> GTGCAGCCCACCGAATCC
AD-SFR2-XmaR	<u>CCCGGG</u> TTTGTCTTCACGAGATTGGT
Sequencing primers	
SQ-5CH6-F	GCCGCTGCTTCTGTGAC
IRESA rev	AGGTTTCCGGGCCCTCACATTG
SQ-MycH-R	GATGACCGCTGCAGAC

<https://doi.org/10.1371/journal.pone.0242890.t001>

#162782). The resulting plasmid p1.1-Tr2-RBDv1 was sequenced using SQ-5CH6-F and IRESA rev primers (Table 1).

**PTM vector construction.** A polylinker was constructed from synthetic oligos listed in Table 1. Oligos were 5' phosphorylated by manufacturer, annealed in pairs (AS-Myc6H-AbsF with AS-Myc6H-AbsR; AS-Myc6H-SpeF with AS-Myc6H-SpeR) in equimolar concentrations in a thermocycler (heated 95°C for 3 minutes, and gradually cooled down -2°C per cycle for 36 cycles), ligated by T4 and cloned to p1.1-Tr2-eGFP restricted by *AbsI-NheI* in place of the eGFP insert. Polylinker had *AbsI-SpeI* sticky overhangs, so *NheI* site in the vector backbone was destroyed—this provided an opportunity to use a unique *NheI* site inside the new polylinker. The Resulting pTM vector was sequenced as described above, available from Addgene, plasmid #162783.

**pTM-RBDv2 construction.** RBD ORF was amplified using adaptor primers AD-SFR2-NheF and AD-SFR2-XmaR restricted by *NheI* and *XmaI* (Sibenzyme, Novosibirsk, Russia) and cloned into pTM vector, restricted by *NheI* and *AsiGI* (Sibenzyme, Novosibirsk, Russia). The resulting construct was sequenced using SQ-5CH6-F and SQ-MycH-R primers. pTM-RBDv2 is available from Addgene, plasmid #162785.

Plasmids for cell transfections were purified by the Plasmid Midiprep kit (Evrogen, Moscow, Russia) and concentrated by ethanol precipitation in sterile conditions.

**Quantitative PCR, PCR.** The transgene copy number in the CHO genome was determined by the quantitative real-time-PCR (qPCR) as described in [21, 23]. Serial dilutions of p1.1-eGFP [21] or pGem-Rab1 plasmids were used for calibration curves generation. The weight of one CHO haploid genome was taken as 3 pg, according to [26]. Genomic DNA (gDNA) was purified by The Wizard<sup>®</sup> SV Genomic DNA Purification System (Promega, USA), diluted to 10ng/μl, 50 ng of gDNA used for one PCR reaction. For genome insert quantification primers RT-ID-F (5' -GCCACAAGATCTGCCACCATG-3') and RT-ID-R (5' -GTAGGTCTCCGTTCTTGCCAATC-3') were used, RT-Rab1-F (5' -GAGTCCTACGCTAATGTGAAAC-3') and RT-Rab1-R (5' -TTCCTTGGCTGTGGTGTG-3') were used for normalization. Primers and qPCRmix-HS SYBR reaction mixture (Evrogen) and iCycler iQ thermocycler (Bio-Rad, USA) were used. Calculations of threshold cycles, calibration curves, PCR efficiency and copy numbers were made by the iCycler Iq4 program.

## Cell culture

Chinese hamster ovary DG-44 cells (Thermo Fischer Scientific) were cultured in the ProCHO 5 medium (Lonza, Switzerland), supplemented by 4 mM glutamine, 4 mM alanyl-glutamine and hypoxanthine-thymidine supplement (HT) (PanEco, Moscow, Russia). Cells were grown as a suspension culture in sterile 125 ml Erlenmeyer flasks with vented caps, routinely passaged 3 to 4 days with centrifugation (300 g, 5 min) and seeding density  $3\text{--}4 \times 10^5$  cells/ml.

## Stably transfected cell lines generation and genomic amplification of target gene

The 50–80  $\mu\text{g}$  of each plasmid were precipitated by the addition of 96% ethanol and 3M sodium acetate, washed with 70% ethanol, dried, and resuspended in 100  $\mu\text{l}$  of sterile R-buffer, Neon transfection kit (Thermo Fischer Scientific). The cells were passaged 24 h before transfection. Thirty millions cells per transfection were pelleted by centrifugation, washed once with DPBS (Paneco), and resuspended in 100  $\mu\text{l}$  of the plasmid solution described. Five micrograms of reporter plasmid pEGFP-C2 (Takara Bio Inc., Kusatsu, Shiga, Japan), coding the green fluorescent protein, was added for each transfection. Cells were nucleofected with Neon Transfection system (Thermo Fischer Scientific) at 1700 V, single pulse, 20 ms, 100  $\mu\text{l}$  transfection tip and placed in 30 ml warm ProCHO 5 medium supplemented with 1xHT and 8 mM glutamine and cultured for 48 h in 125 ml Erlenmeyer shaking flask. The transfection efficiency was assessed 48 h post-transfection as the proportion of eGFP-positive live cells. Transfection efficiency was 9% in the case of RBDv1 and 24% in the case of RBDv2. The medium was changed for the selective one, lacking HT and containing 200 nM of methotrexate (MTX, Sandoz, Holzkirchen, Germany). Cells were passaged every 3–4 days until the viability was restored to 85% (20–25 days). After generating a stably transfected cell line up to two rounds of target gene amplification were performed by increasing MTX concentration from 200 nM to 2  $\mu\text{M}$  and 8  $\mu\text{M}$ . Cells were passaged every 3–4 days until the viability restored to 85%; the typical time for establishing genome-amplified cultures was 12–21 days.

## Preparative cell cultures

Seeding cell culture was grown in 125 ml Erlenmeyer shake flasks with 30 ml of Lonza ProCHO 5 medium, supplemented with 4 mM glutamine, 4 mM alanyl-glutamine and 2–8  $\mu\text{M}$  MTX until cell concentration exceeds  $1\text{--}1.5 \times 10^6$  cells/ml. Cell suspension was transferred to four 250 ml Erlenmeyer flasks, each containing 60 ml of culture medium, and grown to the same cell density. The entire cell suspension was transferred to a single 2 L Erlenmeyer flask with 1 L culture medium, final seeding density  $3\text{--}4 \times 10^5$  cells/ml. Cells were cultured for three days, on the fourth day of culture, daily glucose measurements were started. Glucose concentration in the cell supernatant was measured by the Accutrend Plus system (Roche, Switzerland); if glucose level was below 20 mM, it was added up to 50 mM as the sterile 45% solution. The culture in 2 L flask was grown for 6 to 8 days until the cell viability, measured by trypan blue exclusion, dropped below 50%.

## Clonal cell line generation

The clonal cell line was obtained by the limiting dilution method from the cell population, cultured in 8  $\mu\text{M}$  MTX. Methotrexate was omitted in the culture medium for two 3 d passage before cloning. Cells were additionally split by 1:1 dilution 24 hours before the cloning procedure. Cells were diluted in EXCELL-CHO (Merck, Germany) culture medium supplemented with 4 mM glutamine, 4 mM alanyl-glutamine, HT and 10% of untransfected CHO DG 44

conditioned medium resulting in seeding density 0.5 cell/well, and the suspension was seeded into 96-well plates (200  $\mu$ l/well). Plates were left undisturbed for 14 days at 37°C, 5% CO<sub>2</sub> atmosphere. Wells with single colonies were screened by microscopy; well grown colonies were detached by pipetting and transferred to the wells of 12-well plate, containing 4 ml of the EXCELL-CHO, supplemented as described above and grown for 7 days undisturbed. Product titer was measured by ELISA, as described below, 6 wells with highest RBDv2 titer were used for further cultivation. Best-producing clonal cell lines were transferred to 125 ml Erlenmeyer flasks with the ProCHO 5 culture medium supplemented with 4 mM glutamine, 4 mM alanyl-glutamine and 8  $\mu$ M MTX and after 5 days in suspension culture, the best producing clone was determined by measuring the product titer and cell concentration.

### Small-scale protein purification

Harvested cell suspension, up to 150 ml, was clarified by 5 minutes centrifugations at 2000 g to remove cells and subsequent 15 minutes centrifugation of supernatant at 20000 g. The clarified supernatant was concentrated by ultrafiltration to approximately 70 ml using one 200 cm<sup>2</sup> 10 kDa MWCO PES tangential flow filtration cassette (Sartorius, Germany) and the ultrafiltration reservoir (Sartorius). Concentrated cell supernatant was diafiltered with 500 ml of the PBS, then adjusted to 300 mM NaCl and used for chromatographic purification.

Chromatography was performed at room temperature by the Akta Explorer system (Cytiva, USA) and the Chelating Sepharose Fast Flow 1 ml disposable HiTrap column (Cytiva), charged by Ni<sup>2+</sup> ions or 1 ml Tricorn 5/10 column (Cytiva) packed by Ni-NTA resin (Thermo Fisher Scientific). Columns were equilibrated by the 50 mM Na-PO<sub>4</sub>, pH = 8.0; 500 mM NaCl solution, sample application was performed at 1 ml/min, columns were washed by the equilibration solution until stable baseline (approximately 10 column volumes) at 2 ml/min flow velocity and subjected to stepwise elution at 1 ml/min, utilizing equilibration solution with 50–250 mM imidazole-HCl, pH 8.0 added. Columns were stripped by the 50 mM EDTA-Na, pH 8.0 solution.

### Preparative protein purification

One liter of harvested cell suspension was clarified by 5 minutes centrifugation at 2000 g to remove cells and subsequent 15 minutes centrifugation of supernatant at 20000 g. The clarified supernatant was concentrated approximately fivefold by ultrafiltration using the QuixStand apparatus with the peristaltic pump (Cytiva), ultrafiltration cassettes holder (Sartorius) and three 5 kDa MWCO cassettes (PES, Millipore, 0.1 m<sup>2</sup>). Concentrated cell supernatant was buffer exchanged by diafiltration by 10 mM imidazole-HCl, pH 8.0 solution until constant filtrate conductivity was achieved, approximately 15 diafiltration volumes. Desalted culture supernatant was supplemented by sodium phosphate, pH 8.0 and sodium chloride solutions, final concentrations– 50 mM Na-PO<sub>4</sub> and 500 mM NaCl.

Chromatographic protein purification was performed at room temperature by the Akta Explorer system (Cytiva) and the XK 16/20 column (Cytiva), packed with 7 ml Ni-NTA Agarose resin (R90101, Thermo Fischer Scientific). Column was equilibrated by 50 mM sodium phosphate, pH 8.0; 500 mM NaCl and 10 mM imidazole-HCl, pH 8.0 solution. Desalted culture supernatant was applied to the column at 3.5 ml/min (2 minutes contact time). The column was washed by equilibration solution until a stable baseline at 7 ml/min flow (approximately 10 column volumes), then by the equilibration solution with 50 mM imidazole-HCl at the same flow velocity for 10 column volumes. The target protein was eluted by the equilibration solution with 300 mM imidazole-HCl at 3.5 ml/min; the eluate was collected as a single

fraction. The flow was paused for 5 minutes until the first 7 ml of eluate were collected for minimization of total elution volume.

Eluted purified RBD protein solutions were desalted by the 200 cm<sup>2</sup> disposable tangential flow filtration cassette, Hydrosart, 10 kDa MWCO (Sartorius), connected to the diafiltration reservoir (Sartorius) and the peristaltic pump (MasterFlex, USA). The diafiltration solution was 10 mM sodium phosphate, pH 7.4; diafiltration was carried out until constant conductivity of the filtrate, approximately 20 diafiltration volumes. Desalted RBD solutions were finally concentrated to the 3–8 mg/ml by ultrafiltration centrifugal concentrators, 10 kDa MWCO PES membranes (Sartorius); supplemented by NaCl, 140 mM final concentration, aliquoted and stored frozen at -70°C.

### **Sodium Dodecyl Sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)**

SDS-PAGE was performed with the 12.5% acrylamide in the separating gel, in reducing conditions, if not stated otherwise, with the PageRuler prestained marker, 5 µl/lane (ThermoFisher Scientific). Gels were stained by the colloidal Coomassie blue according to [27], scanned by the conventional flatbed scanner in the transparent mode as 16-bit grayscale images and analyzed by the TotalLab TL120 gel densitometry software (Nonlinear Dynamics, UK).

### **Western blot**

SDS-PAGE was performed as described above, protein transfer, blocking, hybridization and color development were done according to [28] using nitrocellulose transfer membrane (GVS Group, Bologna, Italy) and Towbin buffer with methanol. Primary anti-c-myc antibody (SCI store, Moscow, Russia #PSM103-100) was used at the 1:2000 dilution, anti-mouse-HRP conjugate (Abcam, Cambridge, UK, ab6789) was used at 1:2000 dilution; membrane was developed by the DAB-metal substrate and scanned by the flatbed scanner in the reflection mode.

### **Analytical size-exclusion chromatography**

Multimeric forms of the RBD were quantified by size exclusion chromatography, utilizing Waters Alliance 2695 chromatography apparatus (Waters, USA) with the PDA detector and the Superdex 200 GL 10/30 column (Cytiva). Mobile phase was 1xPBS + 0.02% sodium azide, flow 0.5 ml/min, injection volume– 30 µl. All samples were analyzed in duplicates. Molecular size calibrators used were chimeric IgG1 infliximab (Roche, Basel, Switzerland), bovine serum albumin (Merck), follicle stimulating hormone (produced in-house according to [24]) and RNase A (Merck). Absorbance was detected at 280 nm. Chromatography traces were analyzed by Waters Millennium software (Waters), the molecular mass of unknowns was calculated by linear regression of the elution time–log(MW) chart.

### **Peptide:N-glycosidase F treatment**

Ten micrograms of purified RBD proteins were treated by 500 units of PNGase F (New England Biolabs, Ipswich, USA #P07045) according to manufacturer's instruction, denaturing protocol. Five micrograms of deglycosylated proteins were analyzed by SDS-PAGE, the rest was used for mass-spectrometry analysis.

### **Mass spectrometry**

Tryptic peptides were obtained from RBDv1 and RBDv2 bands, excised from the SDS-PAGE gel and sliced into pieces. The gel pieces were de-stained with 2 washes by 200 µl 50% acetonitrile (ACN), 50 mM ammonium bicarbonate solution (ABC). Destained gel pieces were dried

*in vacuo* and rehydrated by the 12.5 µg/ml trypsin (Promega, USA), 50 mM ABC, 5 mM CaCl<sub>2</sub> solution. Proteolytic digestion was carried out for 16 h at 37°C. Peptides were extracted from the gel with 25 mM ABC, following by the 80% ACN. Extracts were vacuum-dried and redissolved in the 0.5% trifluoroacetic acid (TFA), 3% ACN. Prepared solutions were mixed at 3:1 ratio with 20% α-cyano-4-hydroxycinnamic acid (Merck) solution in 20% ACN, 0.5% TFA on the target plate.

Solutions of intact and deglycosylated proteins were passed through the ZipTip C18 micro-columns (Millipore), washed and eluted according to manufacturer protocol. One and a half µl of protein solutions were mixed on the target plate with 0.5 µl of the 20% 2,5-dihydroxybenzoic acid (Merck) solution in 20% ACN, 0.5% TFA. Mass spectra were obtained by the MALDI-TOF mass spectrometer Ultraflexxtreme Bruker (Germany) with the UV-laser (Nd), linear mode, positive ions. Spectra were obtained in the 500–5000 Da range for tryptic peptides mixtures, 5000–50000 Da range for intact proteins.

Mass lists for each sample were calculated by the Bruker Daltonics flexAnalysis software (Germany), peptides identification was performed by the GPMW 4.0 software (Lighthouse data, Denmark) and by the Mascot server (Matrix Science, Boston, USA). Glycopeptides mass assignment was performed by the GlycoMod online software tool [29].

### Enzyme-linked immunosorbent assay (ELISA)

Sandwich ELISA with anti-S protein antibodies was performed using a prototype of the SARS-CoV-2 antigen detection kit (Xema Co., Ltd., Moscow, Russia, a generous gift of Dr. Yuri Lebedin). Pre-COVID-19 normal human plasma sample (Renam, Moscow, Russia) was used for preparation of the SARS-CoV-2 negative serum sample. Control pooled serum samples obtained from patients with the PCR-confirmed SARS-CoV-2 infection and distributed by the Xema ltd were tested as positive sample.

Antibody capture ELISA with human serum samples was performed according to [28] at the 100 ng per well antigens load. Antigens were applied on ELISA 96-well plates (Corning, USA) overnight at +4°C, in PBS, 100 µl/well. Plates were washed by PBS–0.02% Tween (PBST) thrice and blocked by 250 µl/well of the 3% BSA in PBS solution, washed by PBST and used immediately. Test sera were inactivated by heating at 56°C for 30 min, diluted by the 1% BSA-PBS, applied as serial dilutions in the 1:20–1:82 000 range, and incubated for 1 h at 37°C. Wells were washed three times by PBST, secondary anti-human IgG antibody-HRP conjugate (Xema Co., Ltd., cat. T271X@1702) was used at the 1:20 000 dilution, the incubation time was 1 h at +37°C. Wells were washed five times by the PBST. The color was developed for exactly 10 minutes at room temperature (+25±2 °C), utilizing the ready-to-use TMB solution (Xema Co., Ltd.), 200 µl/well. The reaction was stopped by adding 100 µl of 5% orthophosphoric acid. Optical densities were measured by the Multiskan EX plate reader (Thermo Fischer Scientific) at 450 nm. All samples were tested in duplicates or triplicates.

### Statistical analysis

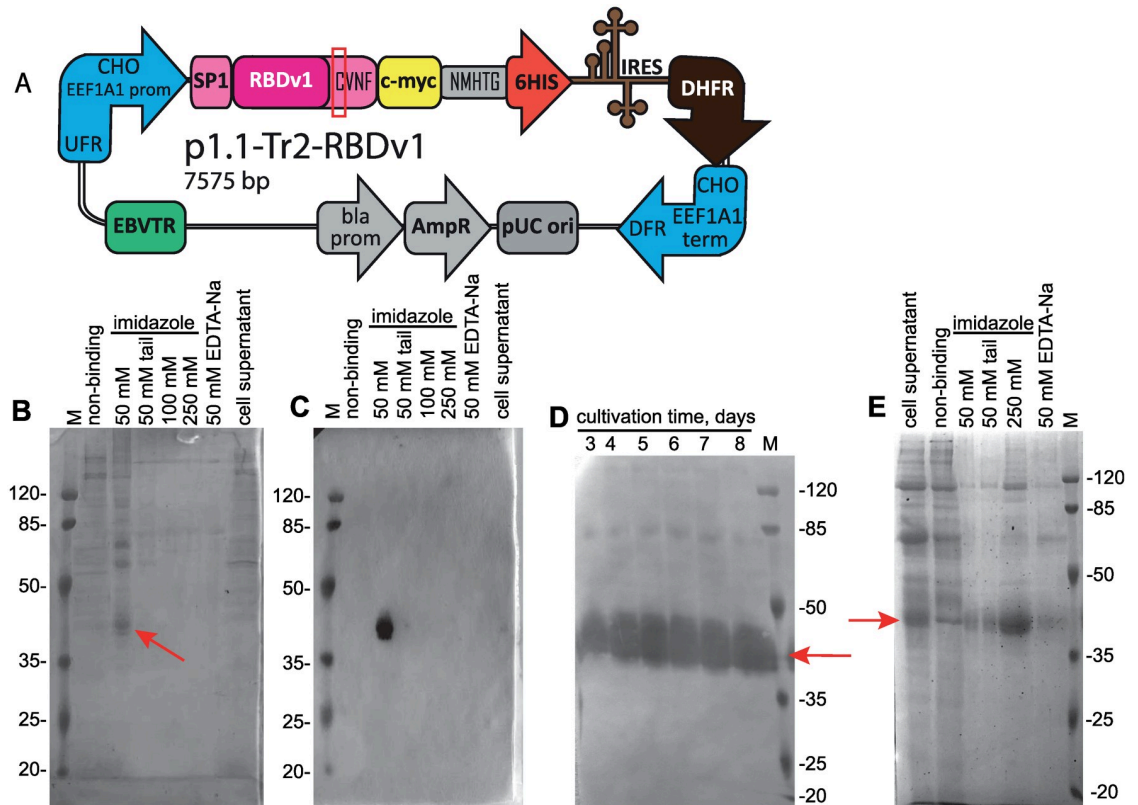
The t-test was performed using the GraphPad QuickCalcs Web site: <https://www.graphpad.com/quickcalcs/ttest1.cfm> (accessed November 2020).

## Results

### SARS-CoV-2—RBDv1

The native N-terminal signal peptide of SARS-CoV-2 S protein (amino acid sequence MFVFLVLLPLVSSQ) was fused to the RBD sequence (319–541), according to [NCBI





**Fig 1. Map of the p1.1-Tr2-RBDv1 plasmid, purification of the RBDv1 protein its extended batch cultivation.** A—Scheme of p1.1-Tr2-RBDv1 expression plasmid. CHO EEF1A1 UFR—upstream flanking area of the *EEF1A1* gene (*EEF1A1* gene promoter, flanked with 5' untranslated region), DFR—downstream flanking area (*EEF1A1* gene terminator and polyadenylation signal, flanked with 3' untranslated region); IRES—encephalomyocarditis virus internal ribosome entry site; DHFR—ORF of the dihydrofolate reductase gene of *Mus musculus*, pUC ori—replication origin; AmpR and bla prom—ampicillin resistance gene and the corresponding promoter; EBVTR—fragment of the long terminal repeat from the Epstein-Barr virus; SP1—signal peptide, RBDv1—RBD ORF, amino acids from natural Spike protein sequence (SP1 and CVNF) are shown in pink color; unpaired cysteine is surrounded by a red rectangle; c-myc and 6HIS—C-terminal fusion tags, NMHTG—linker peptide sequence. B—SDS-PAGE analysis of the RBDv1 purification by the IDA-Ni chromatography resin. Protein was obtained from the cell culture at 2  $\mu$ M MTX selection pressure, cell supernatant and non-binding proteins fractions applied without ultrafiltration, 10  $\mu$ l/lane, column elution fractions applied as 10  $\mu$ l from 5 ml total, concentration factor 30x. Position of the target protein band is depicted by arrow. Molecular masses are given in kDa. C—Western blot analysis of the RBDv1 purification by the IDA-Ni chromatography resin. Anti c-myc primary antibodies, lanes load same to panel B. D—Western blot analysis of the extended batch cultivation of the RBDv1 2  $\mu$ M MTX cell population, 2  $\mu$ l of cell supernatant/lane. E—SDS-PAGE analysis of the RBDv1 purification by the Ni-NTA chromatography resin. Lane loads same to panel B.

<https://doi.org/10.1371/journal.pone.0242890.g001>

Proteins ID: YP\_009724390.1)) and joined with a C-terminal c-myc epitope (EQKLISEEDL), short linker sequence, and hexahistidine tag. N-terminal part of the RBDv1 gene was constructed according to [13], utilizing the optimized codon usage gene structure. C-terminal tags were not optimized for codon usage frequencies. The resulting synthetic gene was cloned into the p1.1-Tr2 vector plasmid (S1 File), a shortened derivative of the p1.1 plasmid [21], and used for transfection of DHFR-deficient CHO DG44 cells. The resulting expression plasmid p1.1-Tr2-RBDv1 [GenBank: MW187858] (S2 File) is shown on Fig 1A. The stably transfected cell population was obtained by selection in the presence of 200 nM of DHFR inhibitor methotrexate, RBD titer 0.33 mg/L was detected for 3-days culture (S1 Fig in S1 Raw file). One-step target gene amplification was performed by increasing the MTX concentration tenfold and maintaining the cell culture for 17 days until cell viability restored to

more than 85%; the resulting polyclonal cell population could secrete up to 3.0 mg/L RBD in the 3-days culture. The target protein was purified by a single immobilized metal affinity chromatography (IMAC) step, utilizing the IDA-based resin Chelating Sepharose Fast Flow (Cytiva), Ni<sup>2+</sup> ions, and step elution by increasing imidazole concentrations (Fig 1B and 1C). The resulting protein production method was found to be sub-optimal due to unexpectedly low secretion rate, signs of cellular toxicity of the target gene—33 h cell duplication time, maximal cell density in shake flask of  $2.3 \times 10^6$  cells/ml (S2 Fig in S1 Raw file), and unacceptable level of contaminant proteins co-eluting with the RBDv1. At the same time, the RBDv1 protein was stable in the culture medium during the extended batch cultivation of cells for at least 7 days (Fig 1D), making the long-term feed batch cultivations a viable option for its production in large quantities.

We proposed that target protein secretion rate and its purity after one-step purification could be significantly improved by a simultaneous shift of the RBD domain boundaries, exchange of the SARS-CoV-2 S protein native signal peptide to the signal peptide of more abundantly expressed protein, two-step genome amplification and switch from IDA-based resin to the NTA-based one (Fig 1E).

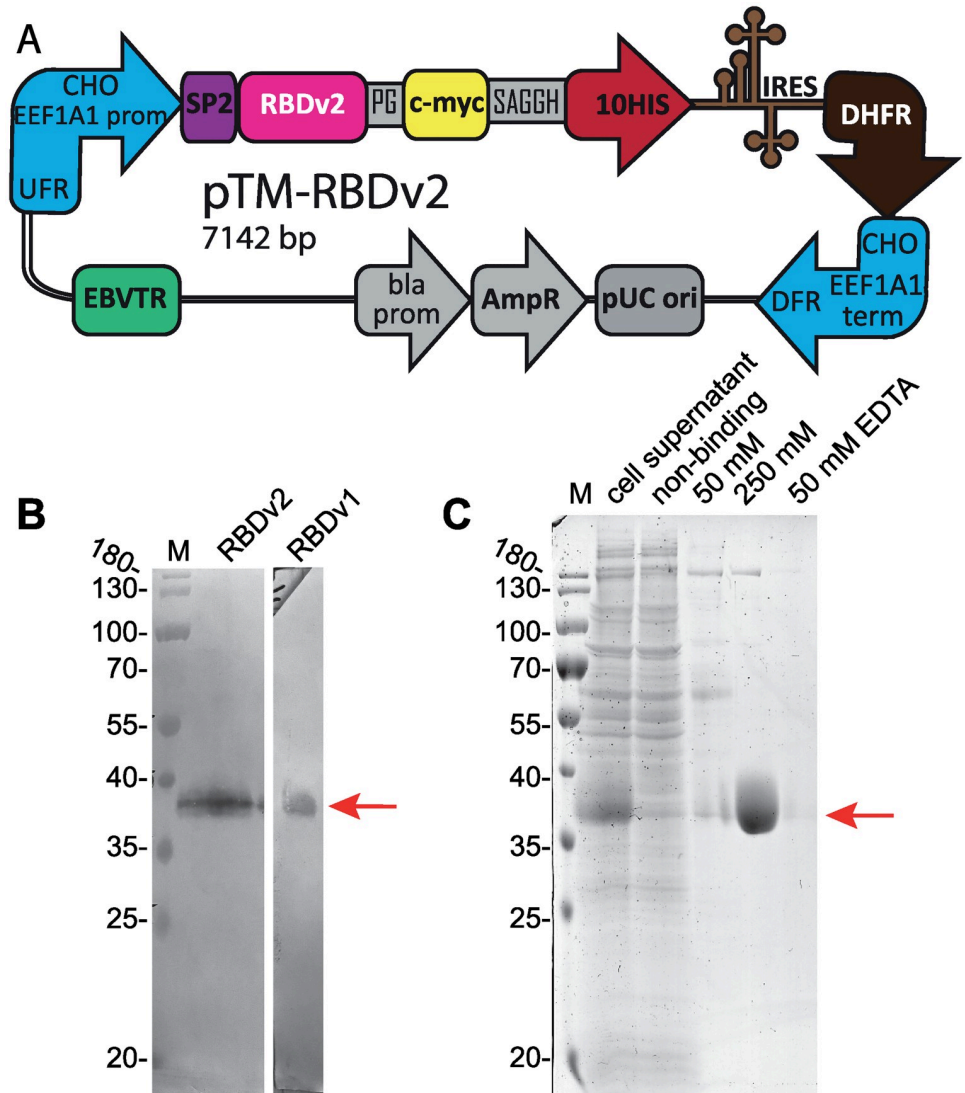
## SARS-CoV-2—RBDv2

Human tissue plasminogen activator signal peptide (hTPA SP, amino acid sequence MDAMKRGLCCVLLLCGAVFVSAS) is commonly used for heterologous protein expression in mammalian cells. It was successfully used for the expression of SARS-CoV S protein in the form of DNA vaccine [30] and envelope viral protein gp120 [31]. In the case of MERS-CoV S protein RBD–Fc fusion protein, various heterologous signal peptides modulate target protein secretion rate by the factor of two [14].

Corrected boundaries of the SARS-CoV-2 RBD were determined according to the cryo-EM data [PDB ID: 6VXX] [32] obtained for the trimeric SARS-CoV-2 S protein ectodomain. Initially used 319–541 coordinates, described in the [13] include one unpaired Cys residue originated from the N-terminal part of the next domain SD1 (structural domain 1), so we excluded Lys319 from the N-terminus of the mature RBD protein, aiming at the maximization of signal peptide processing, and removed C-terminal aminoacids C<sub>538</sub>VNF<sub>541</sub>, which form the structure of the SD1 domain. Both linker areas surrounding the folded RBD domain core remain present in the RBDv2 protein (320–537, according to the [Genbank: YP\_009724390.1]). Additionally, we redesigned C-terminal tags by introducing the Pro residue immediately upstream of the c-myc tag, adding the short linker sequence SAGG between the c-myc tag and polyhistidine tag, and extending the polyhistidine tag up to 10 residues. We expected this structure to expose the c-myc tag properly on the protein globule's surface and move the decahistidine tag away from possible masking negatively charged protein surface areas.

We constructed an expression vector pTM [GenBank: MW187855] (S3 File), where consensus Kozak sequence, hTPA SP and c-myc and 10-histidine tags are coded in the polylinker. RBD coding fragment was cloned in-frame, resulting pTM-RBDv2 expression plasmid [GenBank: MW187856] (S4 File) is shown on Fig 2A.

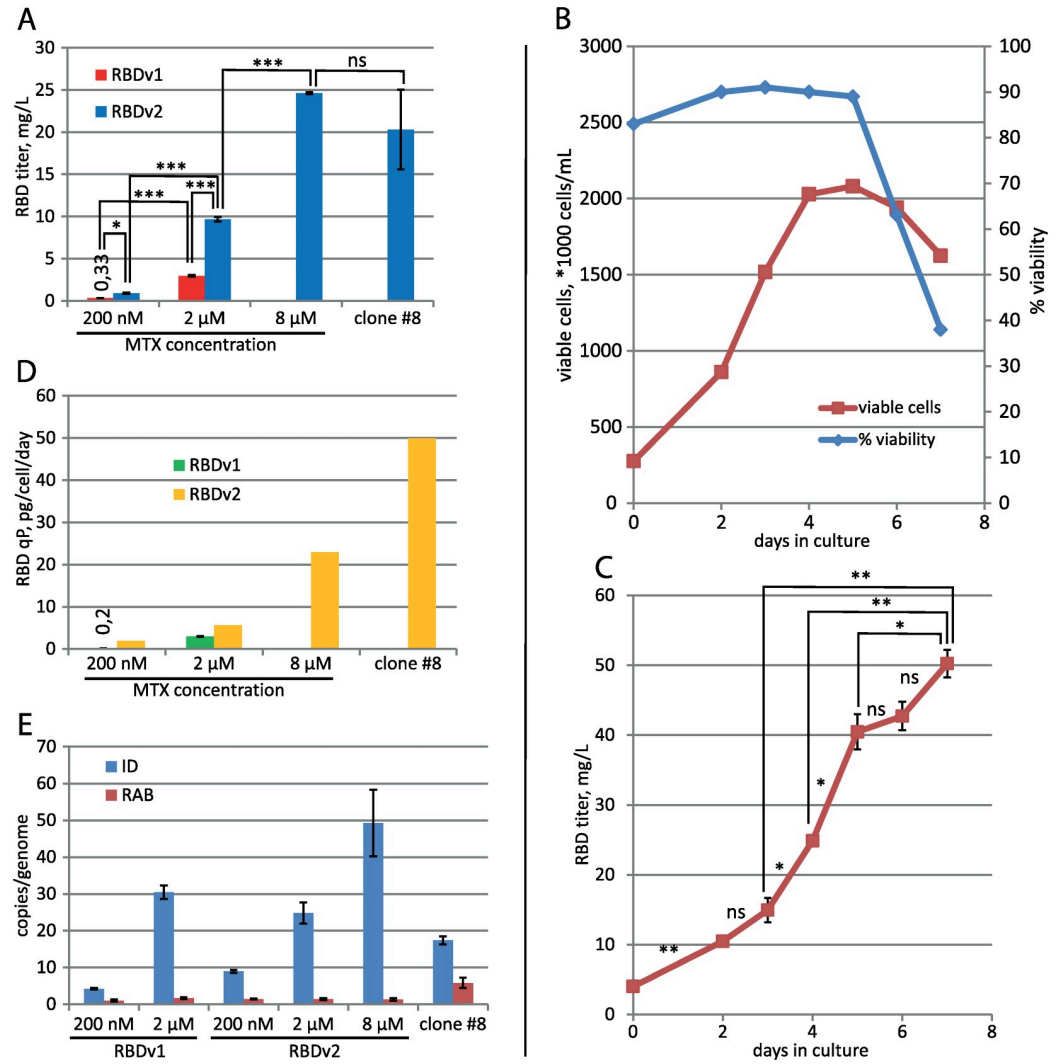
CHO DG44 cells were transfected by the pTM-RBDv2 plasmid; the stably transfected cell population was established at the 200 nM MTX. Target protein titer was similar to the previous plasmid design—0.9 mg/L for 3-days culture, but after one step of the MTX-driven genome amplification, it increased eleven-fold to 9.7 mg/L at 2 μM MTX (Fig 2B) and then increased by a factor of 2.5 after second amplification step at 8 μM MTX, resulting titer was 24.6 mg/L for 3-days culture (Fig 3A). A steady increase of the target protein titer was detected for the extended batch cultivation of polyclonal cell population obtained at 8 μM MTX, peaking at 50



**Fig 2. Map of the pTM-RBDv2 plasmid and purification of the RBDv2.** A—Scheme of pTM-RBDv2 expression plasmid. CHO EEF1A1 UFR—upstream flanking area of the EEF1A1 gene (EEF1A1 gene promoter, flanked with 5' untranslated region), DFR—downstream flanking area (EEF1A1 gene terminator and polyadenylation signal, flanked with 3' untranslated region); IRES—encephalomyocarditis virus internal ribosome entry site; DHFR—ORF of the dihydrofolate reductase gene of *M. musculus*, pUC ori—replication origin; AmpR and bla prom—ampicillin resistance gene and the corresponding promoter; EBVTR—fragment of the long terminal repeat from the Epstein-Barr virus; SP2—hTPA signal peptide, RBDv2—RBD ORF; c-myc and 10HIS—C-terminal fusion tags, PG and SAGGH—linker peptide sequence. B—Western blot analysis of the RBDv2 in the culture medium, 16  $\mu$ l of cell supernatant were obtained from the RBDv2 2  $\mu$ M MTX culture. Control RBDv1 protein was used in purified form, 100 ng/lane. C—SDS-PAGE analysis of the RBDv2 purification by the Ni-NTA chromatography resin. 50 mM, 250 mM—eluates, obtained at 50 mM and 250 mM imidazole concentrations.

<https://doi.org/10.1371/journal.pone.0242890.g002>

mg/L at 8 days of cultivation in the 2 L shake flask (Fig 3B and 3C). A similar ratio of product titer increase after multi-step MTX-driven genome amplification was described for the MERS-CoV RBD— 40-fold increase after 9 steps of consecutive increments of MTX concentration, overall amplification period length was 60 days [14]. Vector plasmid pTM, used in this study, allowed a much more rapid amplification course—a 27-fold titer increase in two steps, 33 days total. This RBD variant was less toxic for cells, as the 2  $\mu$ M MTX cell culture attained a



**Fig 3. Productivity dynamics of cell populations and clonal lines, secreting RBD proteins.** A—Concentration of target proteins in cell supernatant by ELISA. 3-day cultures, all seeded at  $3-4 \times 10^5$  cells/mL. Data are shown as mean  $\pm$  SD,  $n = 2$ . \*— $p < 0.05$ , t-test; \*\*— $< 0.01$ , \*\*\*— $< 0.001$ ,  $ns - p > = 0.05$ . B—Cell culture growth curve for the preparative cultivation in 2 L Erlenmeyer flask, RBDv2 8  $\mu$ M MTX cell population. C—RBDv2 titer dynamics by ELISA, preparative culture in 2 L Erlenmeyer flask, RBDv2 8  $\mu$ M MTX cell population. Data are shown as mean  $\pm$  SD,  $n = 2$ , t-test for pairs of cell supernatants, taken on two subsequent days. D—Specific productivities of cell populations. RBD secretion level in the stably transformed cell populations was determined by ELISA. E—qPCR analysis of the expression cassette copy numbers per haploid genome for RT-ID-F, RT-ID-R primer pair (ID); Rab1 copy numbers for the control sequence (RAB), supposedly unique to the CHO genome, RT-Rab1-F, RT-Rab1-R primer pair,  $n = 4$ . All values for ID primer pair are significantly different from each other, t-test.

<https://doi.org/10.1371/journal.pone.0242890.g003>

higher cell density of  $4.4 \times 10^6$  cells/ml in shake flask with only 25 h cell duplication time. The clonal cell line, secreting the RBDv2 protein, was obtained from the 8  $\mu$ M MTX-resistant polyclonal cell population by limiting dilution. Clone #8, one of the 12 clones screened, possessed a very high specific productivity, 50 pg/cell/day (Fig 3D), and an acceptable cell doubling time of 48 h. This clonal cell line can be used to produce large quantities of RBDv2 with anticipated high uniformity in the pattern of post-translational modifications.

The quantitative PCR analysis of all RBD-secreting cell populations showed that increased productivity of populations adapted to higher MTX concentrations corresponds to higher

copy numbers of the target gene (Fig 3E). At the same time, higher cell productivity for RBDv2 was achieved not due to an increase in gene copy number compared with RBDv1.

Cell culture medium Pro CHO5 (Lonza), utilized in this study, contains unknown components, blocking His-tagged RBD protein's interaction with the Ni-NTA chromatography resin. Clarified conditioned medium, used for protein purification, was concentrated approximately tenfold by tangential flow ultrafiltration on the 5 kDa MWCO cassettes and completely desalted by diafiltration, 20 diafiltration volumes of the 10 mM imidazole-HCl, pH 8.0 solution. RBDv1 and RBDv2 proteins were purified by IMAC utilizing Ni-NTA Agarose (Thermo Fischer Scientific, USA) in the same conditions. Desalted conditioned medium was applied onto the column in the presence of 10 mM imidazole; the column was washed by the solution containing 50 mM imidazole, it removed the majority of admixtures and only minor quantities of the target proteins. In the case of Ni-IDA ligand, interaction with the 6xHis tag on the RBDv1 protein was insufficient to retain the target protein on the column in 50 mM imidazole, resulting in premature elution of RBDv1 along with contaminating cell proteins. In the case of Ni-NTA ligand, both 6xHis and 10xHis tags were suitable for efficient removal of proteinaceous admixtures. This observation is in line with the general properties of immobilized metal affinity chromatography resins. [33]. Elution was performed by the 250–300 mM imidazole solution; further column strip by the 50 mM EDTA-Na solution revealed no detectable target protein RBDv2 in the eluate (Fig 2C). Purified proteins were desalted by another round of ultrafiltration/diafiltration on the centrifugal concentrators with 5 kDa MWCO membranes; diafiltration solution was PBS; final concentration 3–7 mg/ml. Purified proteins were flash-frozen in liquid nitrogen and stored frozen in aliquots. Overall protein yield for RBDv2 was 64%, 32 mg of purified RBDv2 were obtained from 1 L shake flask culture.

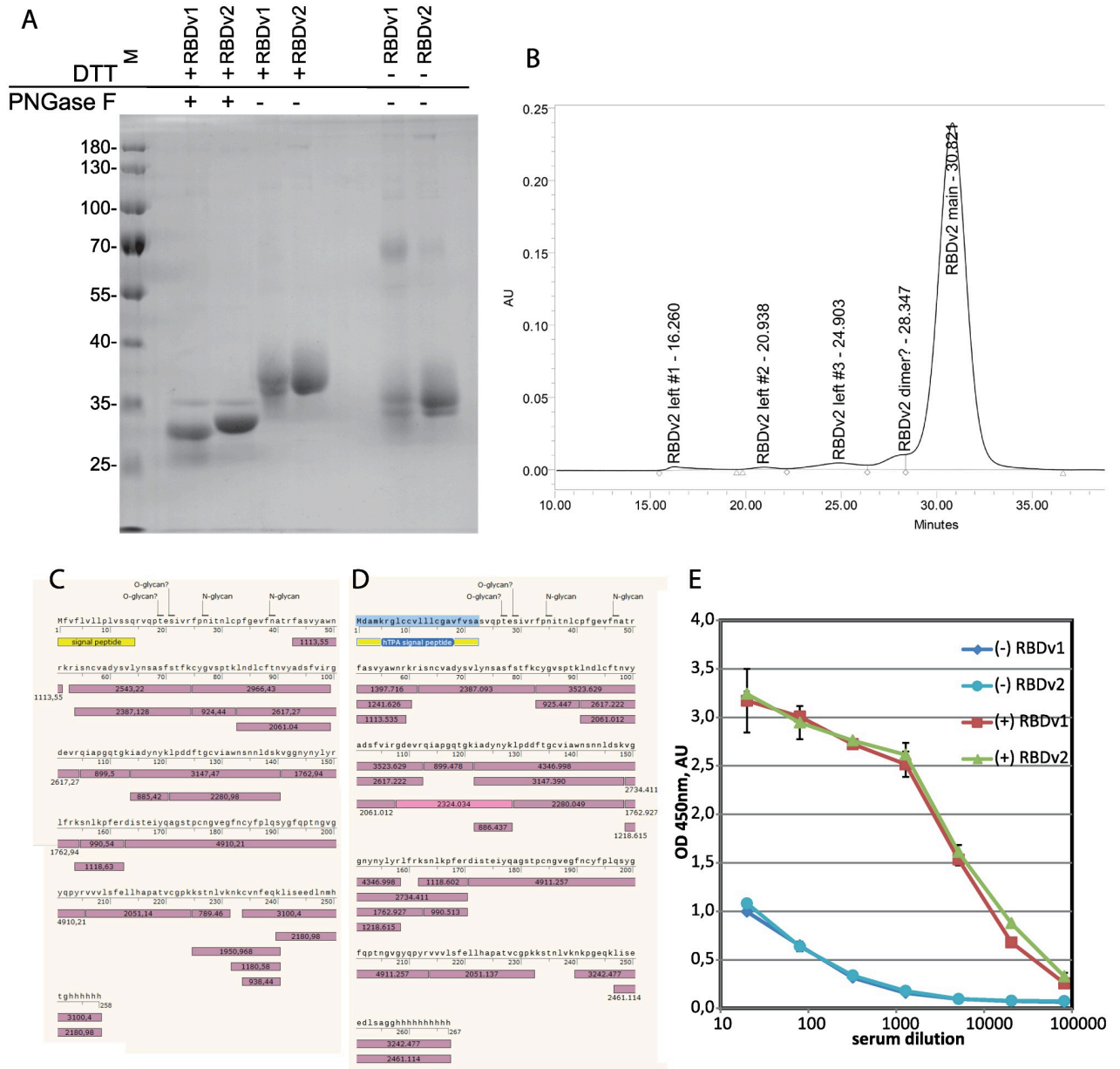
### Analysis of purified proteins

The apparent molecular weight of the RBDv1 estimated by the SDS-PAGE mobility was 35.3 kDa for the intact protein and 26.1 kDa for the deglycosylated RBDv1; its theoretical molecular weight is 27647 Da. For the RBDv2 the apparent molecular weight was 35.7 kDa for the intact protein and 28.5 kDa for the deglycosylated protein; its theoretical molecular weight is 27459 Da (Fig 4A). Both protein variants supposedly possess two distinct forms of intramolecular disulfide bonds sets. This variation is visible as pairs of closely adjacent protein bands in non-reducing conditions and the complete absence of such band pattern in reducing conditions for RBDv2 protein variant. Detailed analysis of this phenomenon will require the complex MS/MS analysis of Cys-linked tryptic peptides of the RBD variants and is beyond the limits of the current study.

Previously it was reported that SARS-CoV-2 RBD 319–541, expressed transiently in HEK-293 cells, tends to form a covalent dimer, around 30% from the total, visible as the 60 kDa band on the denaturing gel in non-reducing conditions [19]. We confirmed this observation; in the case of stably transfected CHO cells, covalent dimerization was also 31%, according to gel densitometry data. At the same time, it should be noted that the RBDv2 protein, redesigned explicitly for mitigation of this unwanted dimerization and containing an even number of Cys residues, still forms 6% of the covalent dimer.

Purified RBDv2 was tested by size exclusion chromatography. The apparent molecular weight of the major monomer form was 32.4 kDa (S3 Fig in S1 Raw file); the apparent molecular masses of the admixtures peaks corresponded well to RBD dimer, tetramer, and two high molecular mass oligomers accounting for 6% of all peak areas (Fig 4B).

Mass-spectrometry analysis of RBDv1 and RBDv2 revealed that both proteins' molecular masses diminished after PNGase F treatment by approximately 3200–3500 Da (S4 Fig in



**Fig 4. Analysis of purified RBDv1 and RBDv2 proteins.** A–SDS-PAGE for PNGase F–treated RBDv1 and RBDv2 proteins and non-reduced proteins. B–Size exclusion chromatography trace for the RBDv2 protein. C, D–protein sequence coverage by tryptic peptides, MALDI-TOF analysis. Glycosylated peptides found are not pictured, signal peptides are yellow, detected tryptic peptides–violet, experimentally obtained masses, [M+H]<sup>+</sup>, are stated in the boxes. E–Immunoreactivity of RBDv1 and RBDv2 by ELISA with pooled serum samples from PCR-positive patients(+), pre-COVID-19 pooled sera (-). All sera samples were analyzed in duplicates, data are mean ± SD.

<https://doi.org/10.1371/journal.pone.0242890.g004>

S1 Raw file); these mass shifts correspond to the removal of two N-linked oligosaccharide groups. The molecular masses of deglycosylated RBDv1 and RBDv2 were 28781 Da and 28123 Da, the theoretical molecular weights of polypeptides—27263 and 27548 Da. We suppose that both protein variants have O-linked glycans. This hypothesis is supported by the complete absence of the signal from the N-terminal tryptic peptide (S)V<sub>339</sub>QPTESIVR<sub>347</sub> for both protein variants. This peptide was found to contain O-glycans in the systemic LC-MS/MS scan of the full-length ectodomain of the S protein [10]. In silico search of possible glycopeptides with

O-linked glycans for RBDv1 and RBDv2 tryptic peptides mixtures revealed two pairs of candidate structures—(Hex)1 (HexNAc)1 and (Hex)1 (HexNAc)1 (NeuAc)1, presented in both mass spectra (S5 Table in [S1 Raw file](#)). These structures were directly detected in the V<sub>339</sub>QPTEsIVR<sub>347</sub> peptide by the LC-MS/MS analysis of the S protein.

Both RBD protein variants were fully covered by tryptic peptides obtained by the in-gel digestion ([Fig 4B and 4C](#)), except for N-terminal peptides, found only in the form of glycopeptides and subsequent F<sub>348</sub>PNITNLCPFGVEFNATR<sub>365</sub> peptide, containing two N-glycosylation sites in positions N350 and N362 (S5 and S6 Figs in [S1 Raw file](#)). This long peptide was completely absent in both spectra of deglycosylated proteins (S1–S4 Tables in [S1 Raw file](#)). A more detailed analysis of this area of the RBD protein may be of some interest for the S protein structure-function investigation but is out of scope for the present study.

Purified RBD variants were used as antigens for microplates coating and subsequent direct ELISA with pooled sera obtained from patients with the RT-PCR-confirmed COVID-19 diagnosis, and serum sample obtained from healthy donors before December 2019 ([Fig 4E](#)). Both RBD variants perform equally—all serum samples produce highly similar OD readings for all dilutions tested with both antigens.

## Discussion

We describe a method of generating stably transfected CHO cell lines, secreting large quantities of monomeric SARS-CoV-2 RBD, suitable for serological assays. Serological assays for detecting seroconversion upon SARS-CoV-2 infection are mostly based on two viral antigens—N and S proteins, or fragments of the S protein, including the RBD. In some cases, the sensitivity of clinically approved N-based assays was challenged by direct re-testing of N-negative serum samples by the RBD-based assays [34]. Other studies question N-based ELISA tests' specificity, demonstrating a significant level of false-positive results for the full-length SARS-CoV-2 N-antigen [35]. The most accurate results can be presumably achieved by testing of the serum samples with both SARS-CoV-2 antigens, as was done, for example, in the South-East England population study [36]; the same conclusion was made in the microarray study of a limited number of patients serum samples [37].

It is unclear yet, which part of the S protein is the optimal antigen for serological assays; microarray analysis revealed that S2 fragment generates more false-positive results than S1 or RBD antigen variants [37] in the case of IgG detection; at the same time, the RBD protein generated much lower signals on COVID-19 patients serum samples than S1 or S1+S2 antigens. In another microarray study it was found that IgG response toward the RBD domain in the convalescent plasma samples correlates well with the response toward full-length soluble S protein [38]. In the conventional ELISA test format, RBD demonstrated nearly 100% specificity and sensitivity on a limited number of SARS-CoV-2 patients and control serum samples [2]. As of 26.10.20, at least 104 various immunoassays for SARS-CoV-2 antibodies were authorized for *in vitro* diagnostic use in the EU [39], many of them use RBD as the antigen. A simple ELISA screening test with the 96-well microplate will consume around 10 µg of the RBD antigen for 40 test samples, so even one million tests will require 250 mg of the purified RBD protein, making the antigen supply a critical step in the production of such tests. Method of the generation of highly productive stably transfected CHO cell line, secreting the RBD protein, can be useful for manufacturers of diagnostic tests, who need a source of antigen with constant properties.

Although the RBD fragment of the S-protein from SARS-CoV-2 is not the most popular antigen variant in the current efforts of anti-SARS-CoV-2 vaccine development [40], it can be considered as the viable candidate for a simple subunit vaccine. It demonstrated the significant

protective immune response development in rodents, without signs of antibody-dependent enhancement effect [41], and some RBD-based protein subunit vaccines have advanced to Phase II clinical trials. Cultured CHO cells are the reliable source of RBD protein for this kind of vaccine; at the productivity level achieved in our study, only 30 m<sup>3</sup> of cell culture supernatant will provide enough antigen material for 100 mln of typical 10 mg/vial vaccine doses.

The described method for obtaining the 320–537 RBD fragment of the SARS-CoV-2 S-protein can be immediately employed to produce a highly consistent RBD reagent for serological assays. The same cell line creation strategy will be applied to obtain RBD 320–537 protein without C-terminal tag, potentially suitable for subunit vaccine studies. This variant of the RBD protein will be characterized in much more detail, including direct analysis of N-linked and O-linked glycans, peptide mapping with S-S bridges positions quantification and surface plasmon resonance studies of interaction with the ACE2 receptor and known neutralizing antibodies. Highly productive RBD-secreting CHO cell lines are already mentioned in the SARS-CoV-2 RBD dimer vaccine candidate studies but without any details on the cell line generation methodology [42].

Other S-protein domains and multi-domain subunits, suitable for serological assays or vaccine studies, can be obtained in stably transfected CHO cells using the vector plasmid pTM described in this study. Although we found no visible changes in the properties of the RBD 319–541 and RBD 320–537 variants used as ELISA antigens, there is a sharp difference in the antigenic properties of the RBD monomer and dimer, described in [42]—dimeric beta-CoV RBD produced much higher titers of neutralizing antibodies due to unknown reasons. Thus, SARS-CoV-2 RBD preparations with a high content of the covalent dimer admixture, i.e., 319–541 variant, can produce incorrectly positive results in the subunit vaccine studies; this variant should be replaced by the shorter 320–537 RBD variant, containing much lower content of the covalent dimer admixture.

## Supporting information

**S1 File. p1.1-Tr2-eGFP [Genbank MW187857].** p1.1-Tr2-eGFP plasmid sequence and feature map.

(GB)

**S2 File. p1.1-Tr2-RBDv1 [Genbank MW187858].** p1.1-Tr2-RBDv1 plasmid sequence and feature map.

(GB)

**S3 File. pTM [Genbank MW187855].** pTM plasmid sequence and feature map.

(GB)

**S4 File. pTM-RBDv2 [Genbank MW187856].** pTM-RBDv2 plasmid sequence and feature map.

(GB)

**S1 Raw images.**

(DOCX)

**S1 Raw file.**

(DOCX)

## Acknowledgments

We thank Mr. Arthur Isaev (Genetico, Moscow, Russia) and Dr. Alexander Ivanov (Institute of Molecular biology Russian Academy of Sciences, Moscow, Russia) for valuable comments



and early access to the SARS-CoV-2 S protein sequence data, Dr. Yuri Lebedin, Eugenia Kostrikina and Xema Co., Ltd., for providing anti-RBD mAbs conjugates and control sera samples.

The measurements were carried out in the Shared-Access Equipment Centre "Industrial Biotechnology" the Research Center of Biotechnology of the Russian Academy of Sciences. DNA sequencing was carried out in the inter-institutional Center for collective use "GENOME" IMB RAS, organized with the support of the Russian Foundation of Basic Research. The authors would like to acknowledge all the doctors who diagnose and treat patients during the COVID-19 pandemic.

## Author Contributions

**Conceptualization:** Nadezhda A. Orlova, Ivan I. Vorobiev.

**Data curation:** Nadezhda A. Orlova.

**Formal analysis:** Maria V. Sinegubova, Nadezhda A. Orlova, Ivan I. Vorobiev.

**Investigation:** Maria V. Sinegubova, Nadezhda A. Orlova, Sergey V. Kovnir, Lutsia K. Dayanova, Ivan I. Vorobiev.

**Methodology:** Maria V. Sinegubova, Nadezhda A. Orlova, Sergey V. Kovnir, Ivan I. Vorobiev.

**Project administration:** Nadezhda A. Orlova.

**Supervision:** Ivan I. Vorobiev.

**Visualization:** Nadezhda A. Orlova.

**Writing – original draft:** Maria V. Sinegubova, Nadezhda A. Orlova, Ivan I. Vorobiev.

**Writing – review & editing:** Nadezhda A. Orlova, Ivan I. Vorobiev.

## References

1. Kilic T, Weissleder R, Lee H. Molecular and Immunological Diagnostic Tests of COVID-19: Current Status and Challenges. *iScience*. 2020; 23(8):101406. Epub 2020/08/11. <https://doi.org/10.1016/j.isci.2020.101406> PMID: 32771976.
2. Premkumar L, Segovia-Chumbez B, Jadi R, Martinez DR, Raut R, Markmann A, et al. The receptor binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Sci Immunol*. 2020; 5(48). Epub 2020/06/13. <https://doi.org/10.1126/sciimmunol.abc8413> PMID: 32527802.
3. Maache M, Komurian-Pradel F, Rajoharison A, Perret M, Berland JL, Pouzol S, et al. False-positive results in a recombinant severe acute respiratory syndrome-associated coronavirus (SARS-CoV) nucleocapsid-based western blot assay were rectified by the use of two subunits (S1 and S2) of spike for detection of antibody to SARS-CoV. *Clin Vaccine Immunol*. 2006; 13(3):409–14. Epub 2006/03/09. <https://doi.org/10.1128/CVI.13.3.409-414.2006> PMID: 16522785.
4. He Y, Li J, Heck S, Lustigman S, Jiang S. Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design. *J Virol*. 2006; 80(12):5757–67. Epub 2006/05/30. <https://doi.org/10.1128/JVI.00083-06> PMID: 16731915.
5. Tan CW, Chia WN, Qin X, Liu P, Chen MI, Tiu C, et al. A SARS-CoV-2 surrogate virus neutralization test based on antibody-mediated blockage of ACE2-spike protein-protein interaction. *Nat Biotechnol*. 2020; 38(9):1073–8. Epub 2020/07/25. <https://doi.org/10.1038/s41587-020-0631-z> PMID: 32704169.
6. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*. 2016; 3(1):237–61. Epub 2016/09/01. <https://doi.org/10.1146/annurev-virology-110615-042301> PMID: 27578435.
7. Burrell CJ, Howard CR, Murphy FA. Fenner and White's medical virology. Fifth edition. ed. Amsterdam; Boston: Elsevier/AP, Academic Press is an imprint of Elsevier; 2017. xix, 583 pages p.

8. Yang J, Petitjean SJL, Koehler M, Zhang Q, Dumitru AC, Chen W, et al. Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. *Nat Commun.* 2020; 11(1):4541. Epub 2020/09/13. <https://doi.org/10.1038/s41467-020-18319-6> PMID: 32917884.
9. Xu W, Wang M, Yu D, Zhang X. Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. *Front Immunol.* 2020; 11:565278. Epub 2020/10/06. <https://doi.org/10.3389/fimmu.2020.565278> PMID: 33013929.
10. Shajahan A, Supekar NT, Gleinich AS, Azadi P. Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology.* 2020. Epub 2020/05/05. <https://doi.org/10.1093/glycob/cwaa042> PMID: 32363391.
11. Zhang Y, Zhao W, Mao Y, Chen Y, Wang S, Zhong Y, et al. Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins. *Mol Cell Proteomics.* 2020. Epub 2020/10/21. PMID: 33077685.
12. 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. Epub 2020/09/16. <https://doi.org/10.1038/s41598-020-71748-7> PMID: 32929138.
13. Amanat F, Stadlbauer D, Strohmeier S, Nguyen THO, Chromikova V, McMahon M, et al. A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat Med.* 2020; 26(7):1033–6. Epub 2020/05/14. <https://doi.org/10.1038/s41591-020-0913-5> PMID: 32398876.
14. Nyon MP, Du L, Tseng CK, Seid CA, Pollet J, Naceanceno KS, et al. Engineering a stable CHO cell line for the expression of a MERS-coronavirus vaccine antigen. *Vaccine.* 2018; 36(14):1853–62. Epub 2018/03/03. <https://doi.org/10.1016/j.vaccine.2018.02.065> PMID: 29496347.
15. Chen WH, Du L, Chag SM, Ma C, Tricoche N, Tao X, et al. Yeast-expressed recombinant protein of the receptor-binding domain in SARS-CoV spike protein with deglycosylated forms as a SARS vaccine candidate. *Hum Vaccin Immunother.* 2014; 10(3):648–58. Epub 2013/12/21. <https://doi.org/10.4161/hv.27464> PMID: 24355931.
16. Jegouic SM, Loureiro S, Thom M, Paliwal D, Jones IM. Recombinant SARS-CoV-2 spike proteins for sero-surveillance and epitope mapping. *bioRxiv.* 2020:2020.05.21.109298. <https://doi.org/10.1101/2020.05.21.109298>
17. Arbeitman CR, Auge G, Blaustein M, Bredeston L, Corapi ES, Craig PO, et al. Structural and Functional Comparison of SARS-CoV-2-Spike Receptor Binding Domain Produced in *Pichia pastoris* and Mammalian Cells. *bioRxiv.* 2020:2020.09.17.300335. <https://doi.org/10.1038/s41598-020-78711-6> PMID: 33311634
18. Rattanapit K, Shanmugaraj B, Manopwisedjaroen S, Purwono PB, Siri wattananon K, Khorattanakulchai N, et al. Rapid production of SARS-CoV-2 receptor binding domain (RBD) and spike specific monoclonal antibody CR3022 in *Nicotiana benthamiana*. *Sci Rep.* 2020; 10(1):17698. Epub 2020/10/21. <https://doi.org/10.1038/s41598-020-74904-1> PMID: 33077899.
19. Tee KL, Jackson PJ, Scarrott JM, Jaffe SRP, Johnson AO, Johari Y, et al. Purification of recombinant SARS-CoV-2 spike, its receptor binding domain, and CR3022 mAb for serological assay. *bioRxiv.* 2020:2020.07.31.231282. <https://doi.org/10.1101/2020.07.31.231282>
20. Stadlbauer D, Amanat F, Chromikova V, Jiang K, Strohmeier S, Arunkumar GA, et al. SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr Protoc Microbiol.* 2020; 57(1):e100. Epub 2020/04/18. <https://doi.org/10.1002/cpmc.100> PMID: 32302069.
21. Orlova NA, Kovnir SV, Hodak JA, Vorobiev II, Gabibov AG, Skryabin KG. Improved elongation factor-1 alpha-based vectors for stable high-level expression of heterologous proteins in Chinese hamster ovary cells. *BMC Biotechnol.* 2014; 14:56. Epub 2014/06/16. <https://doi.org/10.1186/1472-6750-14-56> PMID: 24929670.
22. Orlova NA, Kovnir SV, Gabibov AG, Vorobiev II. Stable high-level expression of factor VIII in Chinese hamster ovary cells in improved elongation factor-1 alpha-based system. *BMC Biotechnol.* 2017; 17(1):33. Epub 2017/03/28. <https://doi.org/10.1186/s12896-017-0353-6> PMID: 28340620.
23. Kovnir SV, Orlova NA, Shakhparonov MI, Skryabin KG, Gabibov AG, Vorobiev II. A Highly Productive CHO Cell Line Secreting Human Blood Clotting Factor IX. *Acta Naturae.* 2018; 10(1):51–65. Epub 2018/05/02. PMID: 29713519.
24. Orlova NA, Kovnir SV, Khodak YA, Polzikov MA, Nikitina VA, Skryabin KG, et al. High-level expression of biologically active human follicle stimulating hormone in the Chinese hamster ovary cell line by a pair of tricistronic and monocistronic vectors. *PLoS One.* 2019; 14(7):e0219434. Epub 2019/07/06. <https://doi.org/10.1371/journal.pone.0219434> PMID: 31276557
25. Du L, Zhao G, Chan CC, Li L, He Y, Zhou Y, et al. A 219-mer CHO-expressing receptor-binding domain of SARS-CoV S protein induces potent immune responses and protective immunity. *Viral Immunol.* 2010; 23(2):211–9. Epub 2010/04/09. <https://doi.org/10.1089/vim.2009.0090> PMID: 20374001.

26. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, et al. Eukaryotic genome size databases. *Nucleic Acids Res.* 2007; 35(Database issue):D332–8. Epub 2006/11/09. <https://doi.org/10.1093/nar/gkl828> PMID: 17090588.
27. Kang D, Suh M-K, 강철훈. Highly Sensitive and Fast Protein Detection with Coomassie Brilliant Blue in Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis. *Bulletin of the Korean Chemical Society.* 2002; 23(11):1511–2. <https://doi.org/10.5012/BKCS.2002.23.11.1511>
28. Greenfield EA. *Antibodies: a laboratory manual.* Second edition. ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2014. xxi, 847 pages p.
29. Cooper CA, Gasteiger E, Packer NH. GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics.* 2001; 1(2):340–9. Epub 2001/10/30. PMID: 11680880.
30. Wang S, Chou TH, Sakhatskyy PV, Huang S, Lawrence JM, Cao H, et al. Identification of two neutralizing regions on the severe acute respiratory syndrome coronavirus spike glycoprotein produced from the mammalian expression system. *J Virol.* 2005; 79(3):1906–10. Epub 2005/01/15. <https://doi.org/10.1128/JVI.79.3.1906-1910.2005> PMID: 15650214.
31. Plotnik D, Guo W, Cleveland B, von Haller P, Eng JK, Guttman M, et al. Extracellular Matrix Proteins Mediate HIV-1 gp120 Interactions with alpha4beta7. *J Virol.* 2017; 91(21). Epub 2017/08/18. <https://doi.org/10.1128/JVI.01005-17> PMID: 28814519.
32. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020; 181(2):281–92 e6. Epub 2020/03/11. <https://doi.org/10.1016/j.cell.2020.02.058> PMID: 32155444.
33. Block H, Spriestersbach A, Brinker N, Kubicek J, Fabis R, Labahn J, Schäfer F. Immobilized-Metal Affinity Chromatography (IMAC): A Review. In: Richard R. Burgess MPD, editor. *Methods in Enzymology.* 463. San Diego: Academic Press; 2009. p. 439–73. [https://doi.org/10.1016/S0076-6879\(09\)63027-5](https://doi.org/10.1016/S0076-6879(09)63027-5) PMID: 19892187
34. Rosadas C, Randell P, Khan M, McClure MO, Tedder RS. Testing for responses to the wrong SARS-CoV-2 antigen? *Lancet.* 2020; 396(10252):e23. Epub 2020/09/01. [https://doi.org/10.1016/S0140-6736\(20\)31830-4](https://doi.org/10.1016/S0140-6736(20)31830-4) PMID: 32866429.
35. Yamaoka Y, Jeremiah SS, Miyakawa K, Saji R, Nishii M, Takeuchi I, et al. Whole Nucleocapsid Protein of Severe Acute Respiratory Syndrome Coronavirus 2 May Cause False-Positive Results in Serological Assays. *Clinical Infectious Diseases.* 2020. <https://doi.org/10.1093/cid/ciaa637> PMID: 32445559
36. Wells PM, Doores KJ, Couvreur S, Nunez RM, Seow J, Graham C, et al. Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England. *J Infect.* 2020. Epub 2020/10/18. <https://doi.org/10.1016/j.jinf.2020.10.011> PMID: 33068628.
37. de Assis RR, Jain A, Nakajima R, Jasinskas A, Felgner J, Obiero JM, et al. Analysis of SARS-CoV-2 Antibodies in COVID-19 Convalescent Blood using a Coronavirus Antigen Microarray. *bioRxiv.* 2020:2020.04.15.043364. <https://doi.org/10.1101/2020.04.15.043364> PMID: 32511302
38. Jiang HW, Li Y, Zhang HN, Wang W, Yang X, Qi H, et al. SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat Commun.* 2020; 11(1):3581. Epub 2020/07/16. <https://doi.org/10.1038/s41467-020-17488-8> PMID: 32665645.
39. database. Foundation for Innovative New Diagnostics. SARS-CoV-2 diagnostic pipeline 2020 [updated November 2020]. <https://www.finddx.org/covid-19/pipeline/>.
40. Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F. A systematic review of SARS-CoV-2 vaccine candidates. *Signal Transduct Target Ther.* 2020; 5(1):237. Epub 2020/10/15. <https://doi.org/10.1038/s41392-020-00352-y> PMID: 33051445.
41. Quinlan BD, Mou H, Zhang L, Guo Y, He W, Ojha A, et al. The SARS-CoV-2 receptor-binding domain elicits a potent neutralizing response without antibody-dependent enhancement. *bioRxiv.* 2020:2020.04.10.036418. <https://doi.org/10.1101/2020.04.10.036418>
42. Dai L, Zheng T, Xu K, Han Y, Xu L, Huang E, et al. A Universal Design of Betacoronavirus Vaccines against COVID-19, MERS, and SARS. *Cell.* 2020; 182(3):722–33 e11. <https://doi.org/10.1016/j.cell.2020.06.035> PMID: 32645327.