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[1]
Email: mvsineg@gmail.com

Nadezhda A. Orlova[1*]
Email: nobiol@gmail.com*

Sergey V. Kovnir[1]
Email: kovnir.serge@gmail.com

Lutsia K. Dayanova[2]
l.daianowa@yandex.ru

Ivan I Vorobiev[1,2]
Email: ptichman@gmail.com,

*Corresponding author
1 Laboratory of Mammalian Cell Bioengineering, Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences. 7, bld. 1, 60 let Oktjabrja pr-t, Moscow 117312, Russia
2 Laboratory of Glycoproteins Biotechnology, Institute of Bioorganic Chemistry of the Russian Academy of Sciences, 16/10, Miklukho-Maklaya str., Moscow 119971, Russia.

Keywords: SARS-CoV-2, Spike protein, Receptor binding domain, Chinese hamster ovary cells

Abbreviations: CHO, Chinese hamster ovary; DHFR, dihydrofolatereductase; EMCV, encephalomyocarditis virus; IRES, internal ribosome entry site; MTX, methotrexate; HT, hypoxanthine-thymidine; RBD, receptor-binding domain; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2;ACE2, angiotensin-converting enzyme 2; *EEF1A1*, Eukaryotic translation elongation factor-1 alpha gene; PNGase F, Peptide:N-glycosidase F

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 similar applications.

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-CoV2 are reviewed in [1].

At the beginning of the pandemic, PCR testing methods dominated since such test systems can be developed urgently, soon after the emergence of a new virus in the population. Among the disadvantages of PCR-tests is a high sensitivity to contamination and dependence on sampling's correctness, a high proportion of false-positive signals. Unlike PCR diagnostics, serological testing gives positive results long after the event of infection, at least for several months.

This testing method makes it possible to reliably determine whether a person is infected with the SARS-CoV-2, even in the absence of disease symptoms. We need serological tests, both in express format and screening tests based on ELISA. Serologic tests are also needed to detect convalescent plasma of therapeutic interest and assess emerging vaccines' effectiveness.

In order for serological testing to have a more significant predictive value, mapping of the epitopes to which neutralizing antibodies appear should be carried out, as was done for SARS-CoV1 [2], and convalescent or postvaccinal sera should be massively tested for the presence of neutralizing antibodies, for example, with a surrogate virus neutralization test based on antibody-mediated blockage of ACE2–spike protein-protein interaction [3] or another that can be carried out on a relatively large scale.

The use of highly specific and high-affinity viral antigens is already a big step towards improving diagnostic accuracy.

The immunodominant antigen of SARS-CoV2 is the RBD domain of the spike protein [4]. Another antigen widely used for diagnostics - the nucleocapsid (N) protein - combines high sensitivity and low specificity; therefore, it needs accurate antigen mutagenesis to remove highly conserved areas without compromising affinity. Cases are described for SARS-CoV when the results of testing with N-protein were clarified using two subunits of spike protein [5].

The coronaviruses' spike (S) protein forms large coronal-like protrusions on the virions surface, hence the name of the family *Coronaviridae*. The S protein plays a crucial role in receptor recognition, cell membrane fusion, internalization of viruses, and their exit from the endosomes. It is described in detail in the review [6]. It consists of S1 and S2 subunits and, in the case of the SARS-CoV-2 virus, has 1260 amino acids [7].

The S protein is co-translationally incorporated into the rough endoplasmic reticulum (ER) and is glycosylated by N-linked glycans. Glycosylation is essential for proper folding and transport of the S protein. The S protein trimer is transported from the ER. 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 [8].

During their intracellular processing, S proteins of many types of coronaviruses, including SARS-CoV-2 and 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. Apparently, most of the S protein copies on the membrane of SARS-CoV-2 viral particles are trimers of S2 subunits that are incapable of interacting with the receptor. The full-length S protein trimer on the viral particle's surface also undergoes complex conformational rearrangements during the formation of the RBD-receptor complex and the virus's penetration into the cell.

The S protein homotrimer binds to the ACE2 dimer, detailed study of this interaction is available here [9]. As part of the trimer, the spike protein's monomer "moves its head" - the S1 subunit can form the open or closed conformation; that is, it can have a raised fragment or a lowered RBD domain, this can influence the affinity of antibodies targeted to it.

The S protein of SARS-CoV-2 amino acid sequence is variable, with more than 200 and 18 relatively frequent S protein amino acid variations. 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 [10] 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.

It should be noted that not all of the 22 potential N-glycosylation sites are occupied, for S1 and S2 subunits, obtained from transiently transfected HEK293 cells [11] N-glycosylation events were experimentally confirmed only for 17 out of 22 sites, also at least one O-glycosylation site was experimentally found inside the RBD-domain area of the S1 subunit with the mucin-like structures. Non-vertebrate cells may 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, possibly blocking the interaction of antibodies with the folded S protein [12]. Computational modeling of the glycan shield, performed for the HEK293-derived S protein, revealed that in the case of human cells, around 40% of the protein’s surface is effectively shielded from IgG antibodies [13].

The use of full-length S protein for practical serological testing is nearly impossible due to its insolubility, caused by the presence of transmembrane domain. An artificial trimer of its ectodomain has been successfully used as an antigen in serological tests; however, such complex protein cannot be obtained in large quantities in mammalian cells, apparently due to the limitation on the folding of the trimerized abundantly glycosylated protein and subsequent difficulties in its isolation and purification. 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 [11] 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 may be proposed that the addition of N-glycans in these sites is needed for correct folding of the RBD in the ER of eukaryotic cells.

The SARS-CoV-2 S protein RBD, expressed in E.coli, also was detected only as inclusion bodies and was found to be unreactive even on blotting [16]. Hyperglycosylated yeast-derived SARS-CoV-2 RBD was obtained in reasonable quantities (50 mg/L in bioreactor culture) by the *P. pastoris* expression system and successfully used for mice immunization [17]. Unfortunately, yeast-derived glycosylated proteins contain immunogenic glycans and cannot be used for immune assays with human antibodies. Similarly, SARS-CoV-2 RBD may be produced in the *Nicotiana benthamiana* plant, resulting in non-vertebrate N-glycans addition, potentially reactive with human antibodies [18].

Most early preprints and peer-reviewed articles describing the SARS-CoV-2 S protein and its RBD domain production methods were focused on transient transfection of HEK293 cells [11] [19] and purification of small protein lots in a very short time. For example, D. Stadlbauer [20] reports more than 20 mg/L target protein titer in transiently transfected HEK-293 cells. Simultaneously, the scalability of transiently transfected cell lines cultivation is still questionable, and gram quantities of RBD, needed for large scale in vitro diagnostic activity, may 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 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, may 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 coding sequence was synthesized according to [26]. The DNA fragment encoding the RBDv1 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 oligo's, PCR reagents Plasmid Miniprep Purification kit and PCR Clean-Up System, were from Evrogen. The PCR product was restricted using *Abs*I (Sibenzyme, Novosibirsk, Russia) and *Nhe*I (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 # 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 *Nhe*I site in the vector backbone was destroyed – this provided an opportunity to use a unique *Nhe*I 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 *Nhe*I and *Xma*I (Sibenzyme, Novosibirsk, Russia) and cloned into pTM vector, restricted by *Nhe*I and *AsiG*I (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 [22, 24]. 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 [27]. Genomic DNA was purified by The Wizard® 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'-TTCCTTGGCTGTGGTGTTG-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 3-4*10^5 cells/ml.

Stably transfected cell lines generation and genomic amplification of target gene

The 50-80 μg of each plasmid were precipitated by the addition of 96% ethanol and 3M sodium acetate, washed with 70% ethanol, dried, and resuspended in 100 μl 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 μl of the plasmid solution described. Five micrograms of reporter plasmid pEGFP-C2 (Clontech), 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 μl 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 DHFR inhibitor methotrexate (MTX). 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 μM and 8 μM. 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 μM MTX until cell concentration exceeds 1-1.5 mln 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 3-4*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 µM 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 µl/well). Plates were left undisturbed for 14 days at 37°C, 5% CO2 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 µM 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^2 10 kDa MWCO PES tangential flow filtration cassette (Sartorius, Germany) and the ultrafiltration reservoir (Sartorius). Concentrated cell supernatant was diafiltered by 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^{2+} 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_4, 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^2). 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_4 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^2 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 diavolumes. 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.

Protein characterization.

SDS-PAGE analysis

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 [28], 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 [29] 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, Bazel, 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 Millenium software (Waters), the molecular mass of unknowns was calculated by linear regression of the elution time – log(MW) chart.

Peptide:N-glycosidase F treatement

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, 50 mM ammonium bicarbonate solution (ACN-ABC). Destained gel pieces were dried *in vacuo* and rehydrated by the 12.5 µg/ml trypsin (Promega, USA), 50 mM ABC, 5 mM $CaCl_2$ solution. Proteolytic digestion was carried out for 16 h at 37°C. Peptides were extracted from the gel with 25 mM ABC solution, following by the 80% ACN solution. Extracts were vacuum-dried and redissolved in the 0.5% trifluoroacetic acid (TFA), 3% ACN solution. 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 microcolumns (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 Ultraflextreme

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 GPMAW 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 [30].

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 [29] at the 100 ng per well antigens load. Antigens were applied on ELISA 96-well plates (Corning, USA) overnight at + 4oC, 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 1h 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 ^{o}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 and discussion

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 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 [26], 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, 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] 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 (Fig. S1). 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 IMAC chromatography step, utilizing the IDA-based resin Chelating Sepharose Fast Flow (Cytiva), Ni2+ ions, and step elution by increasing imidazole concentrations (Fig 1B, Fig 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 mln cells/ml (Fig S2), 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 [31] and envelope viral protein gp120 [32].
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] [33] obtained for the trimeric SARS-CoV-2 S protein ectodomain. Initially used 319 - 541 coordinates, described in the [26] 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 $C_{538}VNF_{541}$, 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 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], 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] is shown on Fig.2A.

CHO DG44 cells were transfected by the pTM-RBDv2 plasmid, stably transfected cell population was established at the 200 nM MTX selection pressure. 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 mg/L at 8 days of cultivation in the 2 L shake flask (Fig 3D, 3E). 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 μM MTX cell culture attained a higher cell density of 4.4 mln 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 μM 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 3B), and an acceptable cell doubling time of 48 h. This clonal cell line may be used to produce large quantities of RBDv2 with anticipated high uniformity in the pattern of post-translational modifications.

All cell populations, secreting RBD proteins, were analyzed by the quantitative PCR and it was found, that increased productivity of populations, adapted to higher concentrations of MTX corresponds to higher copy numbers of target gene (Fig 3C). Higher cell productivity in the case of RBDv2 protein was not due to higher target gene copy numbers, then in the case of 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.

Elution was performed by the 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 intact RBDv1 was determined as 35.3 kDa, deglycosylated RBDv1 – 26.1 kDa, theoretical molecular weight – 27647 Da. RBDv2 molecular weight was determined as 35.7 kDa for the intact protein, deglycosylated protein – 28.5 kDa, theoretical molecular mass – 27459 Da (Fig 4A). Both protein variants possess two distinct forms of intramolecular disulfide bonds sets, visible as two closely adjacent bands in non-reducing conditions and complete absence of such band pattern in reducing conditions. 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 major monomer form's apparent molecular weight was determined as 32.4 kDa (Fig S3), admixtures peaks apparent molecular masses 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 (Fig S4); these mass shifts correspond well to the removal of two N-linked oligosaccharide groups. The molecular masses of deglycosylated RBDv1 and RBDv2 were determined as 28781 Da and 28123 Da, significantly higher than theoretical molecular masses – 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_{339}QPTESIVR_{347} 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 [11]. 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 (Table S5). These structures were directly detected in the V_{339}QPTESIVR_{347} 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, 4C), except for N-terminal peptides, found only in the form of glycopeptides and subsequent

F_{348}PNITNLCPFGEVFNATR$_{365}$ peptide, containing two N-glycosylation sites in positions N350 and N362 (Fig S5, S6). This long peptide was completely absent in both spectra of de-glycosylated proteins (Table S1 – S4). 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

Here we describe a method of generating stably transfected CHO cell lines, secreting large quantities of monomeric SARS-CoV-2 RBD, suitable for serological assays. At present, serological assays for detection of seroconversion upon SARS-CoV-2 infection are mostly based on two viral antigens – nucleoprotein (NP) and S protein or fragments of the S protein, including the RBD. There are various reports on the specificity and sensitivity of assays based on these two antigens. In some cases, the sensitivity of clinically approved NP-based assays was challenged by direct re-testing of NP-negative serum samples by the RBD-based assays [34]. Other studies question the specificity of NP-based ELISA tests, demonstrating a significant level of false-positive results for the full-length SARS-CoV-2 NP [35]. It may be proposed that testing of serum samples with both SARS-CoV-2 antigens will produce the most accurate results, as was done, for example, in the South-East England population study [36]; this 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 then 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 [4]. 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, may be important for IVD test manufacturers in securing the sources of RBD antigen with highly predictable 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 may be considered as the viable candidate for a simple subunit vaccine. It demonstrated the significant protective immune response development in rodents, without signs of ADE effect [41] and some RBD-based protein subunit vaccine have advanced to Phase II clinical trials. Cultured CHO cells are the reliable source of RBD protein for this kind of vaccines; at the productivity level achieved in our study, only 30 m^3 of cell culture supernatant will provide enough antigen material for 100 mln of typical 10 μg/vial vaccine doses.

Data and Materials Availability: The recombinant RBD antigen from the spike protein described in this study is available under a standard MTA with the Federal Research Center "Fundamentals of Biotechnology" Russian Academy of Sciences. Plasmids are available from Addgene: p1.1-Tr2-eGFP # 162782; pTM # 162783; pTM-RBDv2 # 162785.

Acknowledgements

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 on the equipment of 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.

References

1. Kilic, T., R. Weissleder, and H. Lee, *Molecular and Immunological Diagnostic Tests of COVID-19: Current Status and Challenges.* iScience, 2020. **23**(8): p. 101406.
2. He, Y., et al., *Antigenic and immunogenic characterization of recombinant baculovirus-expressed severe acute respiratory syndrome coronavirus spike protein: implication for vaccine design.* J Virol, 2006. **80**(12): p. 5757-67.
3. Tan, C.W., 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): p. 1073-1078.
4. Premkumar, L., 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).
5. Maache, M., 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): p. 409-14.
6. Li, F., *Structure, Function, and Evolution of Coronavirus Spike Proteins.* Annu Rev Virol, 2016. **3**(1): p. 237-261.
7. Huang, Y., et al., *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19.* Acta Pharmacol Sin, 2020. **41**(9): p. 1141-1149.
8. Burrell, C.J., C.R. Howard, and F.A. Murphy, *Fenner and White's medical virology*. Fifth edition. ed. 2017, Amsterdam ; Boston: Elsevier/AP, Academic Press is an imprint of Elsevier. xix, 583 pages.
9. Yang, J., et al., *Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor.* Nat Commun, 2020. **11**(1): p. 4541.
10. Xu, W., et al., *Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19.* Front Immunol, 2020. **11**: p. 565278.
11. Shajahan, A., et al., *Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2.* Glycobiology, 2020.
12. Zhang, Y., et al., *Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins.* Mol Cell Proteomics, 2020.
13. Grant, O.C., et al., *Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition.* Sci Rep, 2020. **10**(1): p. 14991.
14. Nyon, M.P., et al., *Engineering a stable CHO cell line for the expression of a MERS-coronavirus vaccine antigen.* Vaccine, 2018. **36**(14): p. 1853-1862.
15. Chen, W.H., 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): p. 648-58.

16. Jegouic, S.M., et al., *Recombinant SARS-CoV-2 spike proteins for sero-surveillance and epitope mapping.* bioRxiv, 2020: p. 2020.05.21.109298.
17. Arbeitman, C.R., et al., *Structural and Functional Comparison of SARS-CoV-2-Spike Receptor Binding Domain Produced in Pichia pastoris and Mammalian Cells.* bioRxiv, 2020: p. 2020.09.17.300335.
18. Rattanapisit, K., 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): p. 17698.
19. Tee, K.L., et al., *Purification of recombinant SARS-CoV-2 spike, its receptor binding domain, and CR3022 mAb for serological assay.* bioRxiv, 2020: p. 2020.07.31.231282.
20. Stadlbauer, D., 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): p. e100.
21. Orlova, N.A., et al., *Improved elongation factor-1 alpha-based vectors for stable high-level expression of heterologous proteins in Chinese hamster ovary cells.* BMC Biotechnol, 2014. **14**: p. 56.
22. Orlova, N.A., et al., *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): p. 33.
23. Kovnir, S.V., et al., *A Highly Productive CHO Cell Line Secreting Human Blood Clotting Factor IX.* Acta Naturae, 2018. **10**(1): p. 51-65.
24. Orlova, N.A., 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): p. e0219434.
25. Du, L., 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): p. 211-9.
26. Amanat, F., et al., *A serological assay to detect SARS-CoV-2 seroconversion in humans.* Nat Med, 2020. **26**(7): p. 1033-1036.
27. Gregory, T.R., et al., *Eukaryotic genome size databases.* Nucleic Acids Res, 2007. **35**(Database issue): p. D332-8.
28. Kang, D., M.-K. Suh, and □□□, *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): p. 1511-1512.
29. Greenfield, E.A., *Antibodies : a laboratory manual*. Second edition. ed. 2014, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. xxi, 847 pages.
30. Cooper, C.A., E. Gasteiger, and N.H. Packer, *GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data.* Proteomics, 2001. **1**(2): p. 340-9.

31. Wang, S., 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): p. 1906-10.
32. Plotnik, D., et al., *Extracellular Matrix Proteins Mediate HIV-1 gp120 Interactions with alpha4beta7.* J Virol, 2017. **91**(21).
33. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281-292 e6.
34. Rosadas, C., et al., *Testing for responses to the wrong SARS-CoV-2 antigen?* Lancet, 2020. **396**(10252): p. e23.
35. Yamaoka, Y., et al., *Whole Nucleocapsid Protein of Severe Acute Respiratory Syndrome Coronavirus 2 May Cause False-Positive Results in Serological Assays.* Clinical Infectious Diseases, 2020.
36. Wells, P.M., et al., *Estimates of the rate of infection and asymptomatic COVID-19 disease in a population sample from SE England.* J Infect, 2020.
37. de Assis, R.R., et al., *Analysis of SARS-CoV-2 Antibodies in COVID-19 Convalescent Blood using a Coronavirus Antigen Microarray.* bioRxiv, 2020: p. 2020.04.15.043364.
38. Jiang, H.W., et al., *SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses.* Nat Commun, 2020. **11**(1): p. 3581.
39. database. *Foundation for Innovative New Diagnostics. SARS-CoV-2 diagnostic pipeline* 2020 November 2020; Available from: https://www.finddx.org/covid-19/pipeline/.
40. Dong, Y., et al., *A systematic review of SARS-CoV-2 vaccine candidates.* Signal Transduct Target Ther, 2020. **5**(1): p. 237.
41. Quinlan, B.D., et al., *The SARS-CoV-2 receptor-binding domain elicits a potent neutralizing response without antibody-dependent enhancement.* bioRxiv, 2020: p. 2020.04.10.036418.

Tables and figures:

Table 1 Primers used for molecular cloning

Primers for RBDv1 cloning, restriction sites are underlined	
AD-COV-AbsF	AA**CCTCGAGG**CCGCCACCATGTTCATGCCTTCTT
AD-RBD-myc6HNheR	**GCTAGC**CTAATGGTGATGGTGATGATGACCGGTATGCATATTCAGATCCTCTTCTGAGATGAGTTTTTGTTCGAAGTTCACGCATTTGTT
Primers for pTM construction, sticky ends of annealed pairs are underlined	
AS-Myc6H-AbsF	**TCGA**GGCCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTCTCGGCTA
AS-Myc6H-AbsR	**GCGC**TAGCCGAGACGAAGACTGCTCCACACAGCAGCAGCACACAGCAGAGCCCTCTCTTCATTGCATCCATGGTGGCGGCC
AS-Myc6H-SpeF	**GCGC**TACCGGTGAACAAAAACTGATCAGCGAAGAGGATCTGTCTGCAGGCGGTCATCACCATCACCATCACCATCACCATCA
AS-Myc6H-SpeR	**CTAG**TGATGGTGATGGTGATGGTGATGGTGATGACCGCCTGCAGACAGATCCTCTTCGCTGATCAGTTTTTGTTCACCGGTA
Primers for RBDv2 cloning, restriction sites are underlined	
AD-SFR2-NheF	**GCTAGC**GTGCAGCCCACCGAATCC
AD-SFR2-XmaR	**CCCGGG**TTTGTTCTTCACGAGATTGGT
Sequencing primers	
SQ-5CH6-F	GCCGCTGCTTCCTGTGAC
IRESA rev	AGGTTTCCGGGCCCTCACATTG
SQ-MycH-R	GATGACCGCCTGCAGAC

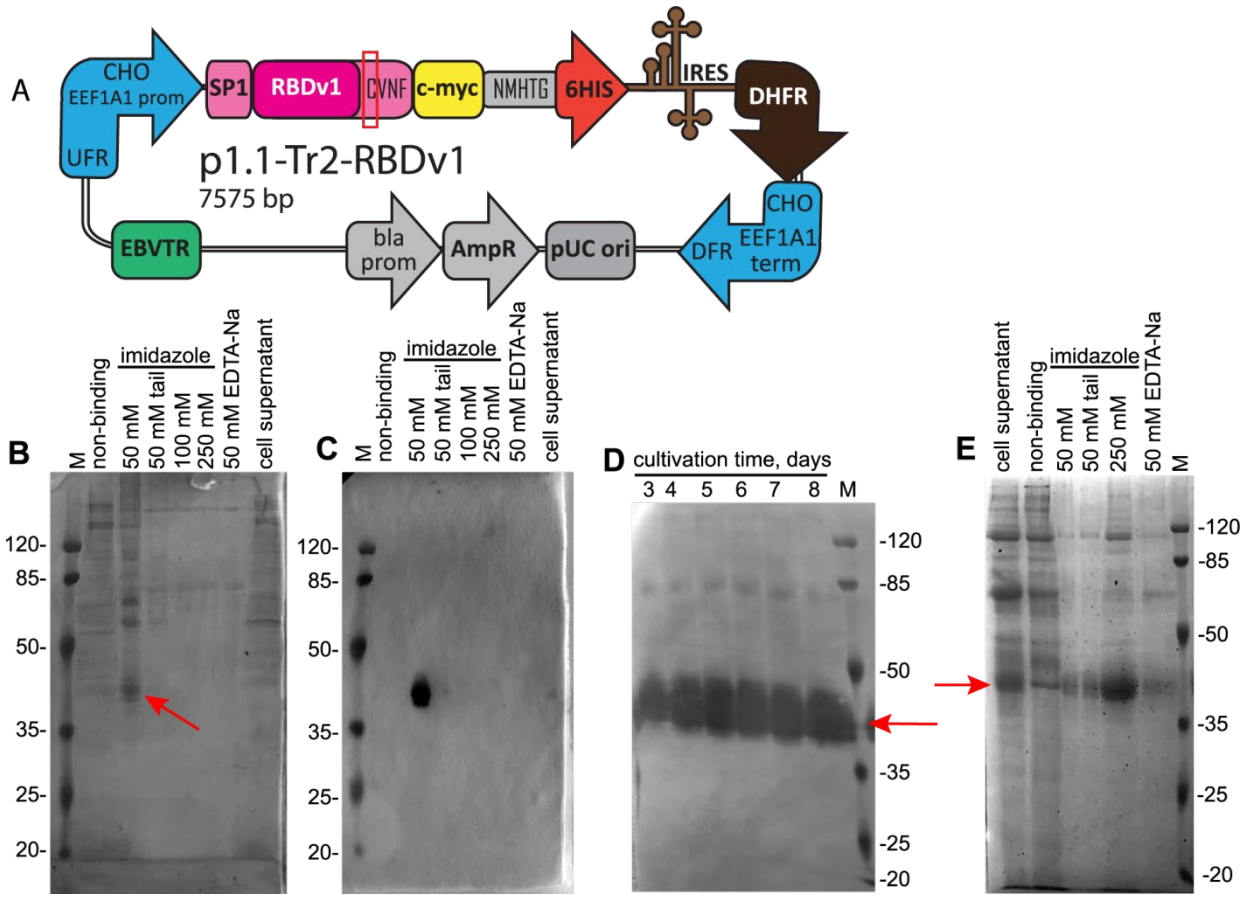

Figure 1. Map of the p1.1-Tr2-RBDv1 plasmid, purification of the RBDv1 protein it's extended batch cultivation.

A - Scheme of p1.1-Tr2-RBDv1 expression plasmid. CHO EEF1A1 UFR – upstream flanking area of the *EEF1A1* gene (elongation factor-1 alpha gene promoter, flanked with 5' UTR), DFR – downstream flanking area (elongation factor-1 alpha gene terminator and polyadenylation signal, flanked with 3' UTR); 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 µM MTX selection pressure, cell supernatant and non-binding proteins fractions applied without ultrafiltration, 10 µl/lane, column elution fractions applied as 10 µl 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 µM MTX cell population, 2 µl of cell supernatant/lane.

E - SDS-PAGE analysis of the RBDv1 purification by the Ni-NTA chromatography resin. Lane loads same to panel B.

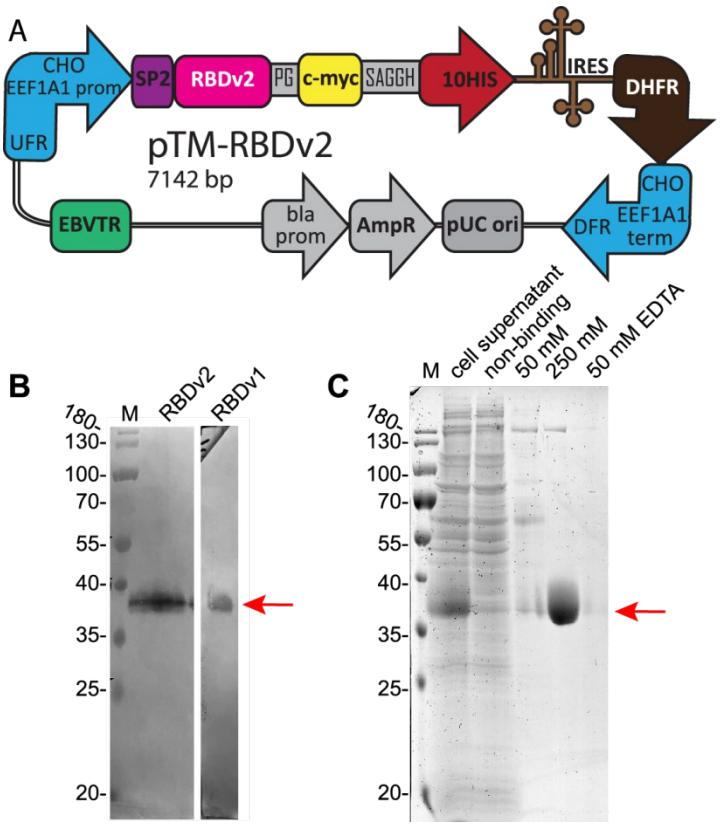

Figure 2. Map of the pTM-RBDv2 plasmid and purification of the RBDv2.
A - Scheme of pTM- RBDv2 expression plasmid. CHO EEF1A1 UFR – upstream flanking area of the *EEF1A1* gene (elongation factor-1 alpha gene promoter, flanked with 5' UTR), DFR – downstream flanking area (elongation factor-1 alpha gene terminator and polyadenylation signal, flanked with 3' UTR); 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 µl of cell supernatant were obtained from the RBDv2 2 µM 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.

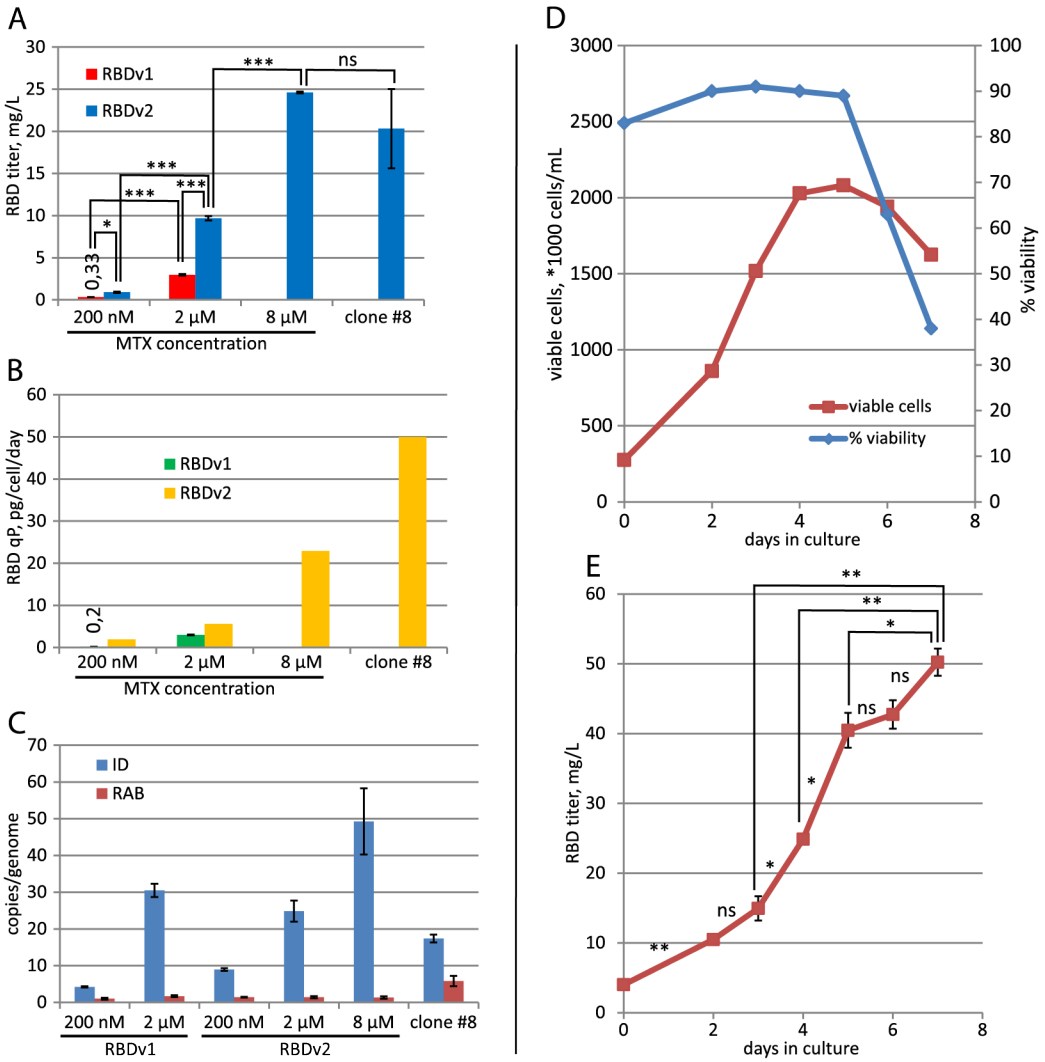

Figure 3. Productivity dynamics of cell populations and clonal lines, secreting RBD proteins.
A – Concentration of target proteins in cell supernatant by ELISA. 3-day cultures, all seeded at 3-4 *105 cells/ml. Data are shown as mean, n=2. * - P < 0.05, t-test; ** - <0.01, *** - <0.001, ns – p >= 0.05.
B – Specific productivities of cell populations. FSH secretion level in the stably transformed cell populations was determined by ELISA.
C – 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. Error bars represent standard deviation, n=4. All values for ID primer pair are significantly different from each other, t-test.
D – Cell culture growth curve for the preparative cultivation in 2 L Erlenmeyer flask, RBDv2 8 µM MTX cell population.
E – RBDv2 titer dynamics by ELISA, preparative culture in 2 L Erlenmeyer flask, RBDv2 8 µM MTX cell population. N=2, t-test for pairs of cell supernatants, taken on two subsequent days.

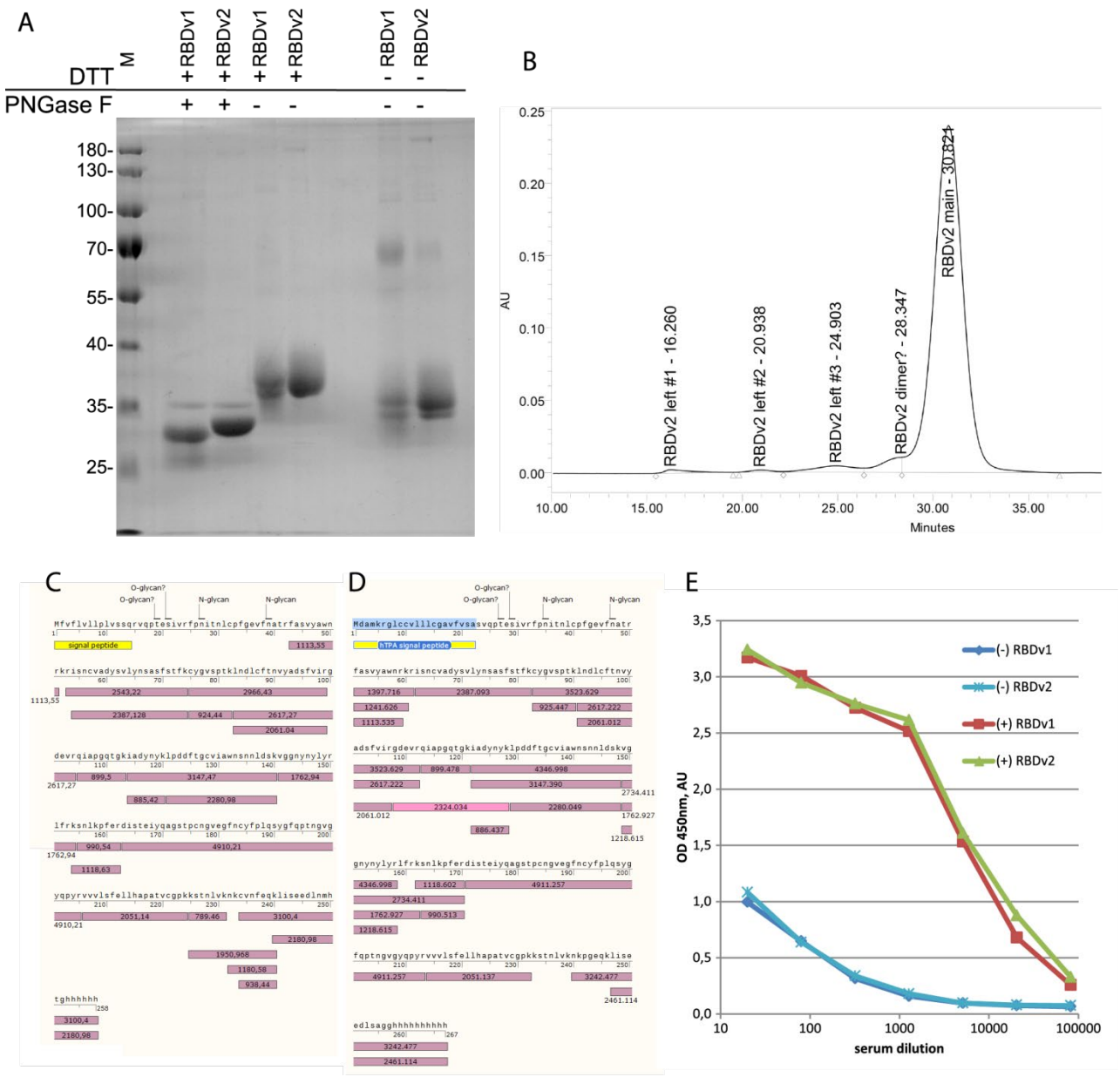

Figure 4. Analysis of purified RBDv1 and RBDv2 proteins.
A – SDS-PAGE for PNGase F – treated RBDv1 and RBDv2 proteins and non-reduced proteins.
B – Size exclusion chromatography trace for the RBDv2 protein.
C, D – protein sequence coverage by tryptic peptides, MALDI-TOF analysis. Glycosylated peptides found are not pictured, signal peptides are yellow, detected tryptic peptides – violet, experimentally obtained masses, [M+H]+, 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.

Supporting figures and tables.

Supporting Figure S1. Cell growth and viability dynamics of initial selection and MTX-driven target gene amplification.

Supporting Figure S2. Cell growth curve for the extended batch cultivation of RBDv1 and RBDv2 – producing cell populations, 2 uM MTX selection pressure.

Supporting Figure S3. Size exclusion chromatography trace of molecular mass calibrators and molecular mass calibration curve.

Supporting Figure S4. MALDI-TOF spectra traces of intact proteins in glycosylated and deglycosylated forms.

Supporting Figure S5. MALDI-TOF spectra traces of tryptic peptides mxtures from intact and deglycosylated RBDv1.

Supporting Figure S6. MALDI-TOF spectra traces of tryptic peptides mxtures from intact and deglycosylated RBDv2.

Supporting Table S1. Peptides mass list of the RBDv1 intact protein, in-gel digestion, reduced protein.

Supporting Table S2. Peptides mass list of the RBDv2 intact protein, in-gel digestion, reduced protein.

Supporting Table S3. Peptides mass list of the RBDv1 protein, treated by PNGase F, in-gel digestion, reduced protein.

Supporting Table S4. Peptides mass list of the RBDv2 protein, treated by PNGase F, in-gel digestion, reduced protein.

Supporting Table S 5. List of identified glycopeptides.