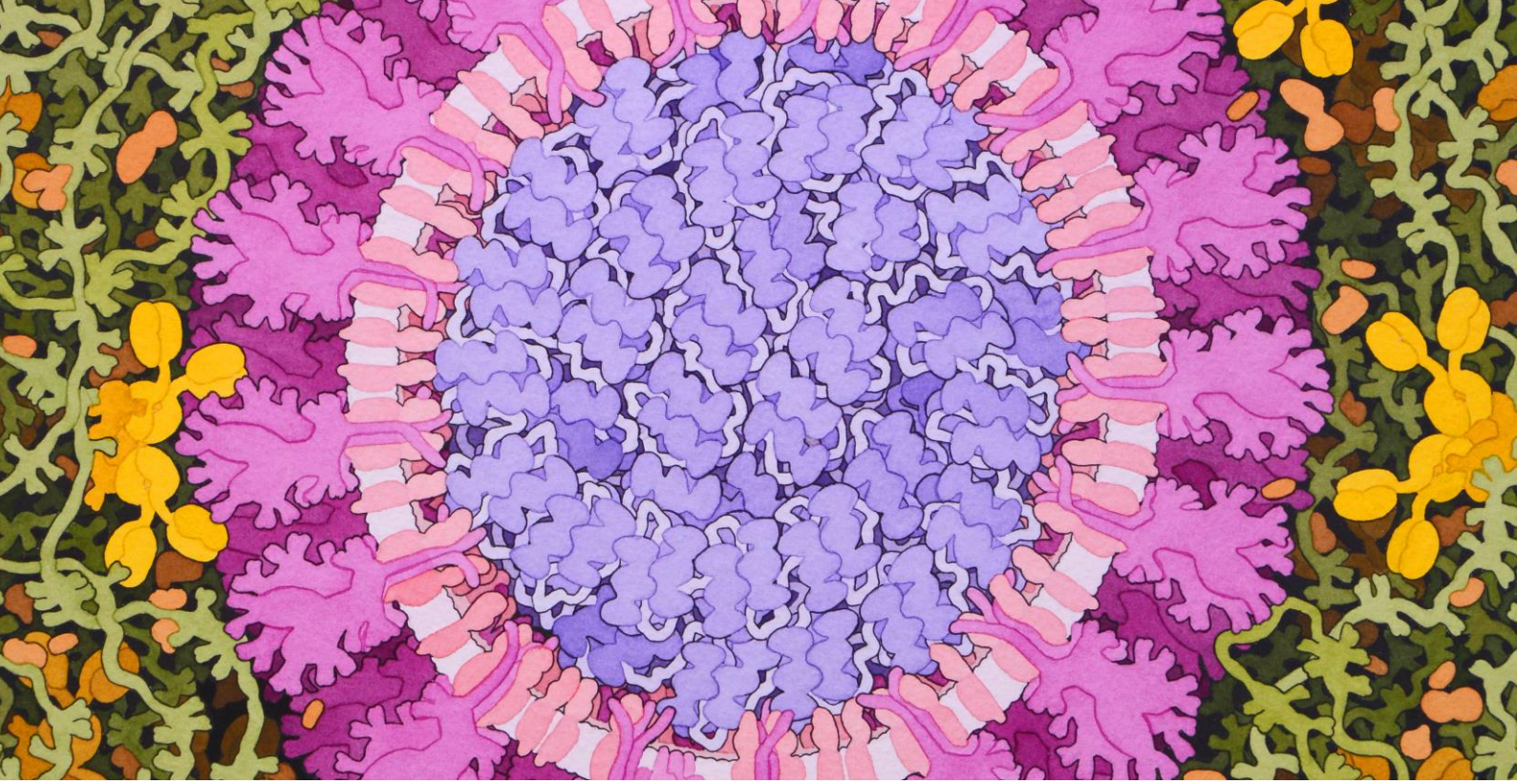


Technical Reports

Nucleoprotein and spike glycoprotein, a combination of two quite different antigens for COVID-19 *in vitro* diagnostic.

THE AIM OF THE RESEARCH PRESENTED IN THIS TECHNICAL REPORT IS INTRODUCING OUR NEW RECOMBINANT ANTIGENS FOR COVID-19: NP (CTD), WHICH IS A PROTEIN WITH A NON-VERY COMPLICATED CONFORMATION, PRODUCED IN *ESCHERICHIA COLI* AND ONE OF THE MAJOR ANTIGENS OF SARS-COV 2. ON THE OTHER SIDE, IS THE RECEPTOR BINDING DOMAIN (RBD) OF THE SPIKE GLYCOPROTEIN, WITH MULTIPLE CONFORMATION-DEPENDENT EPITOPES, THEREFORE, IT SHOULD BE PRODUCED IN A EUKARYOTIC HOST. WE HAVE PRODUCED THIS PROTEIN IN *PICHIA PASTORIS*. BOTH ANTIGENS SHOWED VERY PROMISING RESULTS IN AN ELISA ASSAY AND ARE READY TO BE USED AS RAW MATERIAL IN YOUR IVD TEST.

On 31 December 2019, WHO was informed of a cluster of cases of pneumonia of unknown cause detected in Wuhan City, Hubei Province of China. The coronavirus disease (COVID-2019) was identified as the causative virus by Chinese authorities on 7 January. As part of WHO's response to the outbreak, the R&D has been activated to accelerate diagnostics, vaccines, and therapeutics for this novel coronavirus.

Regarding Ab detection, the new coronavirus SARS-CoV 2, cause of the pandemic COVID-19, is mainly diagnosed by point of care diagnostics by using mainly two different biomarkers. Let´s have a look at these specific antigens for COVID-19 which are extensively used in the rapid serodiagnosis Ab tests. Also, let me introduce you to Rekom antigens NP and S1(RBD).

NUCLEOCAPSID

The nucleocapsid phosphoprotein (NP) is a highly phosphorylated protein which not only is responsible for the construction of the ribonucleoprotein complex by interacting with the viral genome and regulating the synthesis of viral RNA and protein but also serves as a potent immunogen that induces humoral and cellular immunity. The nucleoprotein is one of the major antigens of the SARS-CoV, and there are abundant antigenic sites predicted in this protein. Compared with other viral structural proteins, the low variable rate of the N protein with regards to its size, suggests its importance to the survival of the virus (figure 1).

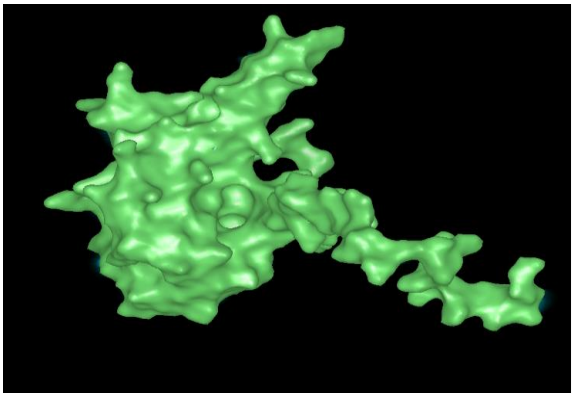


Figure 1. Structure of the N-terminal RNA binding domain of the SARS-CoV nucleocapsid protein (PDB code: 1SSK).

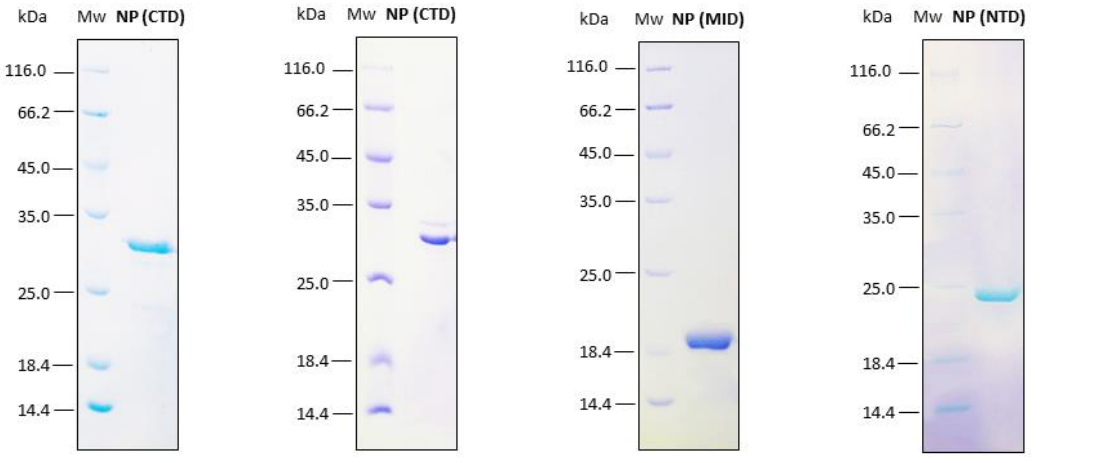
Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

If we perform an alignment of the NP amino acid sequences of both related coronaviruses (SARS-CoV 1 and SARS-CoV 2), the high identity between both, predict a clear cross-reactivity in a serological test (figure 2).

```
COVID-19          MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHG
SARSCoV(2003)     MSDNGPQNQRSAPRITFGGPSDSSDNSKNGERNGARPKQRRPQGLPNNTASWFTALTQHG
                  **********.***********:...:****.***.**********************

COVID-19          KEDLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAG
SARSCoV(2003)     KENLTFPRGQGVPINTNSSKDDQIGYYRRATRRIRGGDGKMKELSPRWYFYYLGTGPEAG
                  **:*.************** **********************:*****************

COVID-19          LPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
SARSCoV(2003)     LPYGANKEGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
                  *******:****************************************************

COVID-19          QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQ
SARSCoV(2003)     QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDTALALLLLDRLNQLENKVSGKGQQ
                  ************************************:**************.*:******

COVID-19          QQGQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
SARSCoV(2003)     QQGQTVTKKSAAEASKKPRQKRTATKQYNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKH
                  ************************** *********************************

COVID-19          WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAY
SARSCoV(2003)     WPQIAQFAPSASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDAY
                  ********************************* **********:***:***********

COVID-19          KTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQQSMSSADSTQA
SARSCoV(2003)     KTFPPTEPKKDKKKKADELQALPQRQKKQQTVTLLPAADLDEFSKQLQQSMSGTDSTQA
                  ****************** **********************:**********.:*****
```

Figure 2. Alignment of the two NP proteins from SARS-CoV 1 and SARS-CoV 2.

This protein is highly identical between SARS-CoV 1 (2003) and SARS-CoV 2 (COVID-19). None of the diagnostic IVD test currently in the market and based on this protein can discriminate between these two viruses. At Rekom Biotech, we have developed the C-terminus of the NP from SARS-CoV 2, (C-terminal domain, RAG0071), and three antigens from the nucleocapsid of the SARS-CoV (outbreak 2003): NP-NTD (N-terminal domain, RAG0082); NP-MID (middle domain, RAG0081) and NP-CTD (C-terminal domain, RAG0080) (figure 3).



NP (CTD) SARS-CoV 2 RAG0071 | **NP (CTD) SARS-CoV 1 RAG0080** | **NP (MID) SARS-CoV 1 RAG0081** | **NP (NTD) SARS-CoV 1 RAG0082**

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

To study the location of different antigenic determinants throughout this protein, several ELISA assays were performed with a serum sample panel of 28 pre-validated IgG-positive specimen sera obtained from the Biobank of Andalusian Public health System and 4 negative pre-COVID specimen sera from the general population.

The nucleocapsid protein from SARS-CoV (outbreak 2003) was truncated in three different fragments to increase the specificity of the resulted biomarker and to analyze the location of the antigenic determinants. These fragments correspond to the following amino acid sequences (table 1).

NAME	REFERENCE	AMINO ACIDS
NP(NTD)	RAG0082	2-132
NP(MID)	RAG0081	110-219
NP(CTD)	RAG0080	213-398

The comparative ELISA assays with positive specimen sera for the three recombinant NP fragments showed that the most prevalent antigenic determinants seem to be located at the C-terminus of this protein (figure 4).

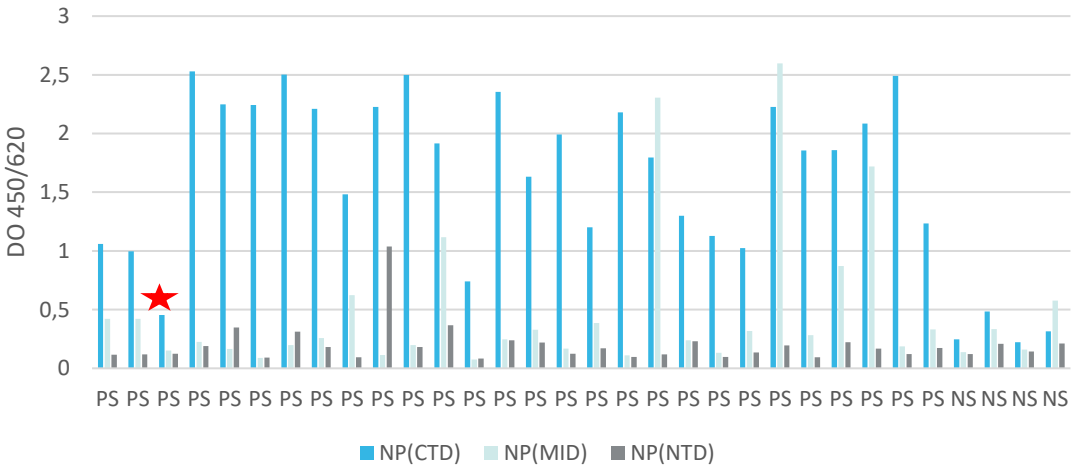


Figure 4. Indirect IgG ELISA assay performed with the different fragments produced as recombinant proteins from nucleocapsid protein. PS means positive serum and NS means negative serum. The red star is pointing the positive serum lost by the NP(CTD) fragment.

These ELISA experiments show that the antigenic determinants of the nucleocapsid are located mainly in the C-terminus of the protein. The only positive serum (Nº A289SUEA002) not detected by the NP(CTD) protein (red star), is not recovered by the other fragments. Therefore, on a preliminary basis, it seems that the MID and NTD fragments would not be able to recover the 3.57% of loss of sensitivity detected with the CTD fragment (table 2), being this serum, a positive serum lost by the complete NP.

NAME	REFERENCE	AMINO ACIDS	SENSITIVITY (%)
NP(NTD)	RAG0082	2-132	3.57%
NP(MID)	RAG0081	110-219	21.43%
NP(CTD)	RAG0080	213-398	96.43%

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

As both NP are nearly identical, we assumed that the more prevalent epitopes were also at the C-terminus of the SARS-CoV 2 NP protein, and we produced its C-terminal fragment as a recombinant protein (aa 212-417). Both proteins were compared and analyzed with the 28 positive COVID-19 specimen sera to detect possible differences in the reactivity of these two proteins (figure 5).

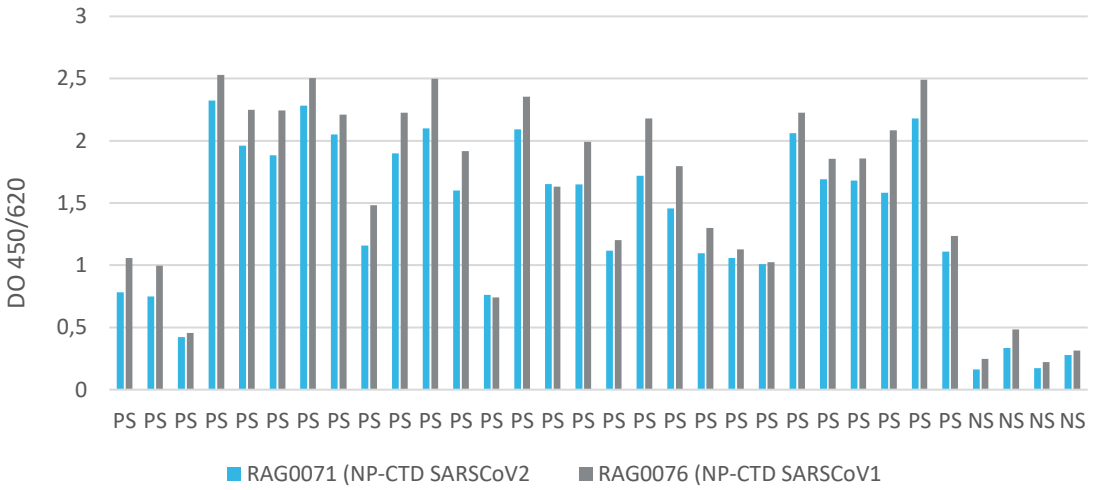


Figure 5. An indirect IgG ELISA assay was performed with both NP(CTD) proteins: SARS-CoV 1 and SARS-CoV 2 NPs. PS means positive serum and NS means negative serum.

The obtained results seem to show that both proteins contain the same antigenic determinants and the same antigenic characteristics (sensitivity 96.5% and specificity 100%). These results are coherent to the high identity (94%) and similarity (97%) in their amino acid sequence. This identity makes both proteins be able to bind the same antibodies in the sera specimen samples.

SPIKE PROTEIN

The Spike (S) protein is a class I viral fusion protein that is synthesized as a single-chain precursor of ~1,300 amino acids and trimerizes after folding. It is a surface glycoprotein located on the envelope of the virions. It is cleaved into S1 and S2 subunits. The S1 subunit is responsible for host-receptor binding while the S2 subunit contains the membrane-fusion machinery (figure 6). There are two domains in coronavirus S1: N-terminal domain (S1-NTD) and C-terminal domain (S1-CTD). The receptor-binding domain (S1 RBD) is also referred to as the S1 CTD or domain B. This domain has been described as a promising antigen for detecting CoV specific antibodies in human sera.

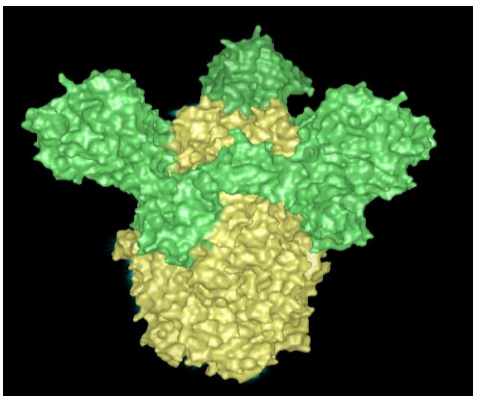
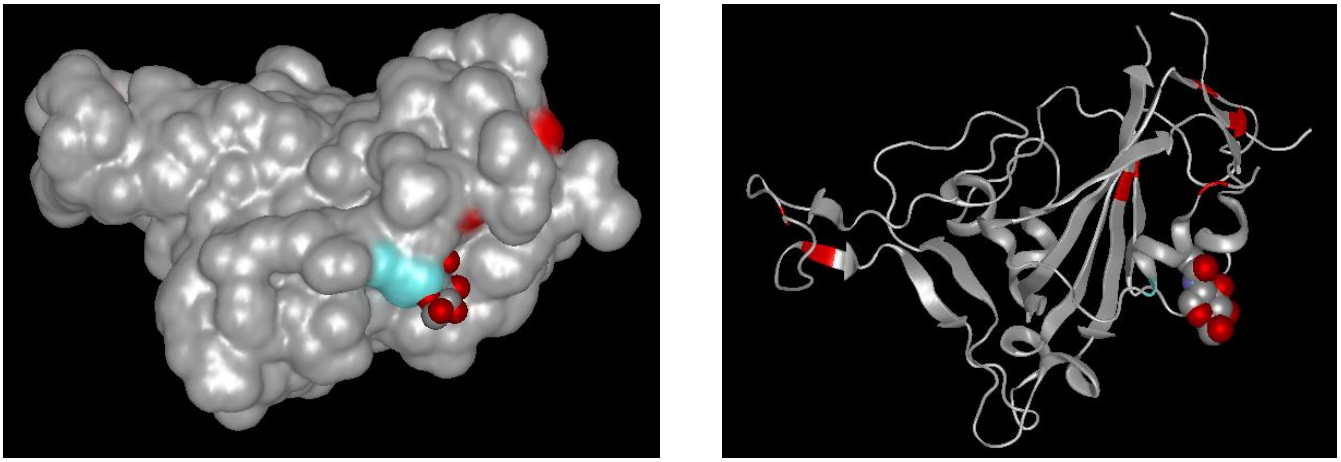


Figure 6. Structure of the trimer of the spike protein (PDB code: 6CRV). In green is shown the S1 domain and in yellow the S2 domain.

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

The Spike (S) glycoprotein alone can mediate the membrane fusion required for virus entry and cell fusion. It is a major immunogen, and target for entry inhibitors. Since the RBD of SARS-CoV S protein contains multiple conformation-dependent epitopes, it must be produced in a eukaryotic expression system. This should be an expression system able to produce post-translational modifications such as disulfide bonds, which have a high impact on the conformation. The spike glycoprotein contains a high number of cysteines. Cys disulfides are often key points in stabilizing protein structure and conformation. Cysteine and cystine groups are relatively hydrophobic and usually can be found within the core of a protein. At the RBD core, we can find 4 internal disulfide bonds, maintaining the specific conformation of this part of the spike glycoprotein (figure 7).



(a) (b)

Figure 7. Structure of the RBD of the spike protein (PDB code: 7BZ5). The disulfide bonds are in the core of the protein and the N-glycosylation bound to a glycan is showed in blue at the solvent accessible protein surface (a). In the ribbon cartoon structure (b), the four disulfide bonds appear in red.

Initially, we tried to express this heterologous gene in *E. coli*. Unfortunately, this recombinant antigen over-produced in *E. coli* produced insoluble aggregate forms in the absence of chaotropic agents and lost antigenicity due to misfolding. Although the material could be partially refolded, the procedure was extremely inefficient, and the protein was not able to recover full antigenicity.

However, this same antigen produced in the yeast *Pichia pastoris*, resulted in a very soluble and stable secreted protein, maintaining its integrity, conformation, and antigenicity. The methylotrophic yeast *Pichia pastoris* has been widely used for heterogeneous protein expression for pharmaceutical and vaccine applications owing to its ability to i) produce large amounts of protein in defined media absent animal-derived growth factors; and ii) offer easy scale-up at low cost. Also, as a eukaryotic expression system, Pichia is capable of post-translational modifications, such as proteolytic processing, folding, disulfide bond formation and glycosylation, which may be necessary for the functions of the proteins.

This protein is the new biomarker produced at Rekom Biotech, under reference RAG0074 (figure 8.A). The protein contains two N-glycosylation in amino acids N331 and N343. These N-glycosylations were eliminated after digestion with the enzyme Peptide-N-Glycosidase F (PNGase). Due to these heterogeneous N-glycosylation, the protein appears in the SDS-PAGE with a smear appearance (figure 8.B). This appearance disappears after PNGase digestion, and a noticeably clear band appears in the gel corresponding with the Mw of the protein. However, as the emerald-green reaction is maintained, although very low, maybe this protein contains one O-glycosylation also (figure 8.C).

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

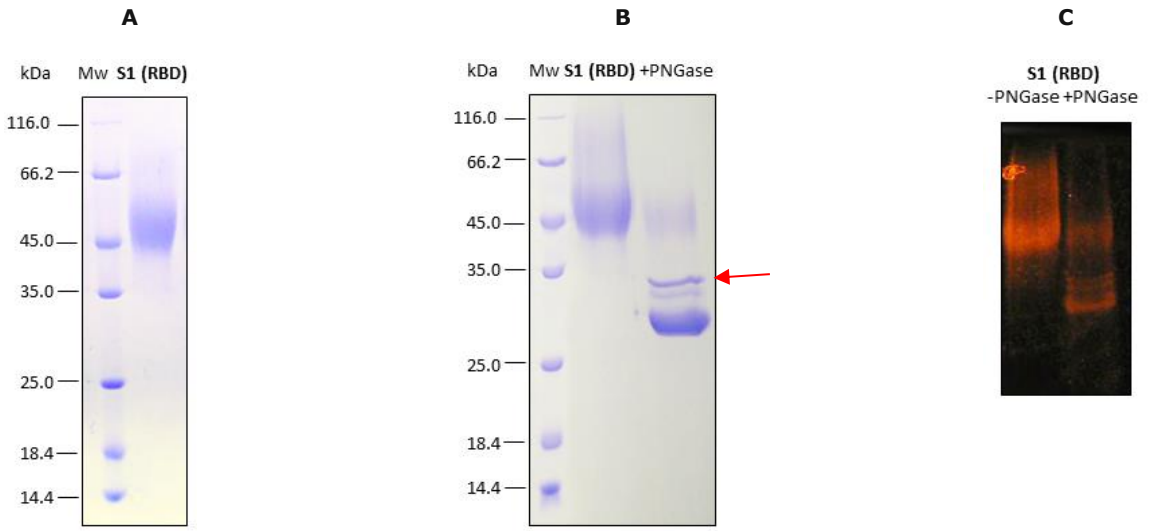


Figure 8. A) SDS-PAGE analysis of the S (RBD) from SARS-CoV 2 produced in *Pichia pastoris*. B) De-N-glycosylation of S1(RBD) by PNGase. A red arrow shows the PNGase enzyme. SDS-PAGE analysis of glycosylated and de-N-glycosylated S1(RBD) expressed in *P. pastoris*. Lane S1(RBD): the protein domain expressed before de-N-glycosylation; lane + PNGase: the protein domain after treating with PNGase F. The red arrow is marking the PNGase. C) SDS-PAGE analysis and stained with emerald-green to detect glycosylation (without and with a previous treatment with PNGase), the de-N-glycosylated S1(RBD) has reduced the intensity after de-N-glycosilation.

This protein was analyzed also with the serum sample panel of 28 pre-validated IgG-positive specimen sera, obtaining the results indicated on figure 9.

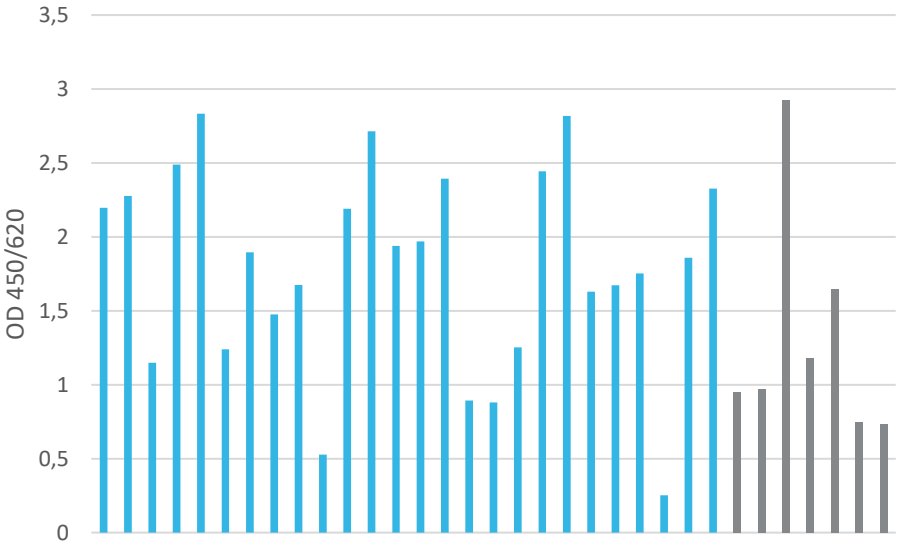


Figure 10. An indirect IgG ELISA assay performed with 26 IgG positive COVID-19 sera and 7 IgG negative COVID-19 sera. PS (positive serum) are in blue and NS (negative serum) are in grey colour.

As it is showed in the figure, several false positives were obtained. Negative sera showed extremely high reactivities.

Studying the bibliography, several unspecific interactions have been described with antibodies against carbohydrate determinants (CD) when glycoproteins have been used in immunodiagnostic. Several yeast species adapted to colonize human tissue (for instance, the endo-saprophytic yeast *Candida albicans*) contain β-linked mannose

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

structures which have been established to be potent antigens for the adaptive immune response and can elicit specific infection protective antibodies (Shibata *et al*., 1985).

Among these virulence attributes, is the yeast cell wall containing large amounts of carbohydrates and carbohydrates covalently linked to a noncarbohydrate moiety classified as glycoconjugates, either glycoproteins or glycolipids. Despite similarities in the early steps of processing, glycans' mature structure differs substantially between yeasts and mammals. Depending on the species, fungal high mannose glycans contain distinctive modifications, such as the addition of mannosyl phosphate and β-linked mannose (Shibata *et al*., 1985). These IgG antibodies against the β-linked mannose are frequently present among healthy individuals without regard to age, race, or gender (Lehmann and Reiss, 1980) (figure 11).

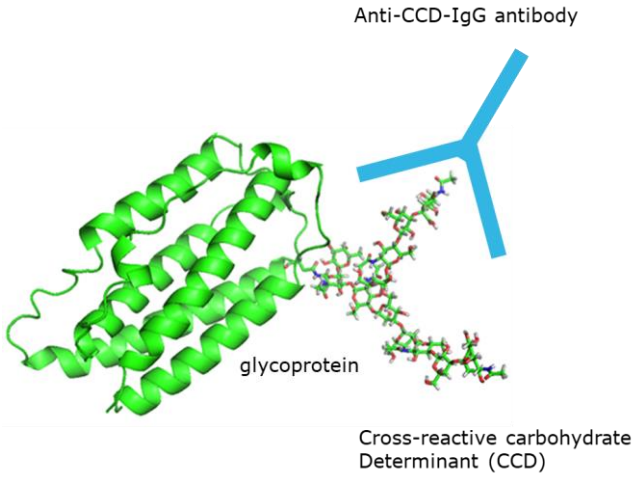


Figure 11. These antibodies present in normal human sera are able to react against the carbohydrates structure of the glycoproteins containing these β-linked mannose.

The presence of these fungal-like glycans, such as those containing β-mannose (Man) linkages, has been described also in *Pichia* (Mille *et al*., 2008). The IgG antibodies present in the normal human sera can cross-react to glycoproteins containing these β-1,2-linked Man residues in their glycan structure, as in the glycoproteins produced in *Pichia pastoris*. To eliminate these unspecific interactions, a sorbent was prepared to kidnap the anti-β-mannose antibodies and tested with the S1 (RBD) protein (figure 12).

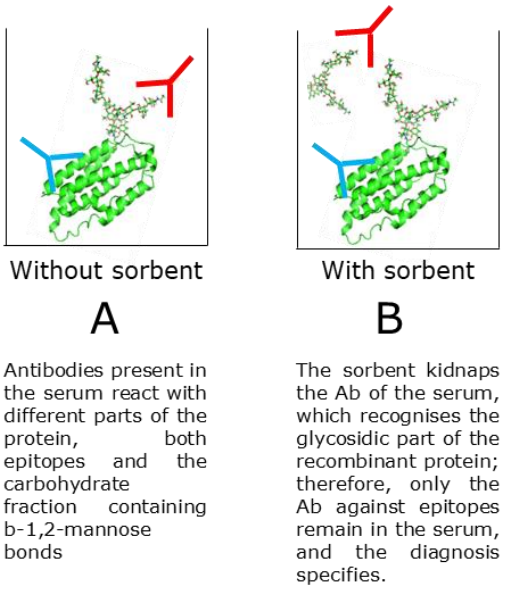


Figure 12. The sorbent kidnaps the antibodies of the serum directed against the coated glycoprotein's carbohydrate structures, specifying the diagnosis as finally, only the antibodies directed against epitopes of the protein are detected.

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

Again, the serum sample panel of 26 pre-validated IgG-positive specimen sera, and several pre-covid negative sera were tested with and without the presence of this sorbent (figure 13)

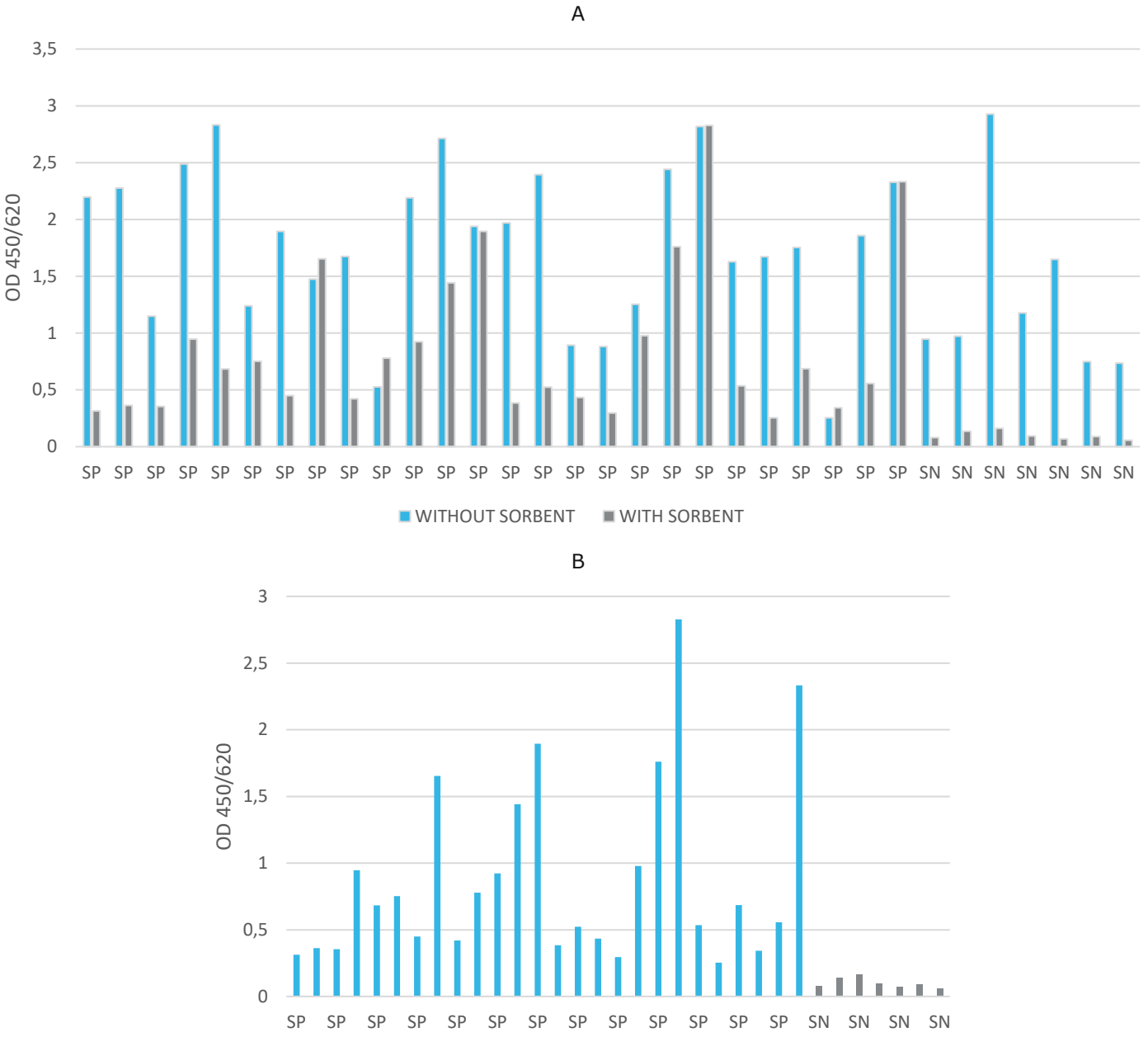


Figure 13. ELISA assay of the S1(RBD) with 26 positive IgG sera and 7 negative IgG sera for COVID-19. The assays were performed with and without sorbent. A) comparison of both results and B) results only with sorbent.

The presence of the sorbent can reduce the reactivity observed in the negative sera. Also, some positive sera also reduce their own signal; only 26% of the positive sera maintain the original signal determined without sorbent. It should also be emphasized that the serum Nº A289SUEA002, lost by the NP(CTD) protein, is recovered by the S1 (RBD) protein.

To perform this ELISA assay, the protein was coated at 0.5 μg/ml in plates and the sorbent was added to the diluted sera (25 μg per serum) just 10 min before adding the serum on the plate (figure 14), mixing at RT.

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

Reconstitute with 1 ml of water, shake at RT for 5 min to obtain a homogeneous solution, considering that part of the cake can be on the container's glass walls. If the complete content is not immediately used, it can be aliquoted and stored at -20ºC.

Add 0.025 ml of the reconstituted sorbent to the serum when it is diluted for the assay. MIx at RT for 10 min.

Figure 14. In any assay performed, the sera used should be pre-incubated just 10 min at RT with the sorbent to eliminate all the antibodies directed against CCD.

In summary, the carbohydrate structures (mannoproteins) present in pathogenic fungus (like Candida) glycoproteins present in the gastrointestinal tract of most healthy adults are sometimes present in Pichia-derived material. Thus they can be recognized by the antibodies raised against these glycoproteins. A similar situation has been described in allergy; the IgE binding to CCDs in vitro diagnostic tests causes false-positive results. The key elements of CCDs are the core α 1,3-fucose occurring in insect glycoproteins, or this fucose plus β 1,2-xylose in plants glycoproteins. Anti-CCD IgE was found in 15-30% of sera patients with allergy. In this case, it is difficult to know in which % of the population these antibodies are present in the sera, as Candida species inhabit the gastrointestinal tract of most healthy adults, seemingly living there as a harmless commensal.

CONCLUSSION

In summary, for COVID-19 IVD, two antigens are the main principal characters. Both are quite different proteins: NP, the nucleocapsid, shows the determinant antigenic located at the C-terminus. It is a protein with no Cys in its sequence, which can be produced easily in *E. coli* without loss of antigenicity. S glycoprotein (RBD) is, such as the complete protein, a highly conformational antigen, with 4 disulfide bonds into its core and two N-glycosylations. This last protein requires a eukaryotic expression system for its production. The cost of this protein produced in insect cells and mammalian cells is extremely high. At Rekom Biotech, we have produced the RBD of SARS-CoV S in Pichia pastoris. As a eukaryotic expression system, Pichia is capable of post-translational modifications like other eukaryotic systems, but with the advantage that it offers an easy scale-up at low cost. Therefore, you could get an antigen produced in a eukaryotic system at the price of an antigen produced in a prokaryotic one!.

Please, look at our new COVID-19 recombinant biomarkers and ask for a quotation! The references are shown in table 3:

REFERENCES	DESCRIPTION	EXPRESSION SYSTEM
RAG0071	NP(CTD) SARS-CoV 2	*E. coli*
RAG0074	S1(RBD) SARS-CoV 2	*P. pastoris*
RAG0080	NP(CTD) SARS-CoV 1	*E. coli*
RAG0081	NP(MID) SARS-CoV 1	*E. coli*
RAG0082	NP(NTD) SARS-CoV 1	*E. coli*

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com

Please, find more interesting information regarding Rekom Biotech products, by visiting our web site in www.rekombiotech.com.

Rekom Biotech S.L. – BIC-Granada, Avda. Innovación, 1 – 18016, Granada (Spain) – Tel: +34 958 63 70 85
E-mail: info@rekombiotech.com – Web: www.rekombiotech.com