De novo **design of anti-variant COVID vaccine with T-cell memory**

Arpita Goswami[1]*, S. Madan Kumar[2], Milind M. Gore[3]

[1]Leo's Research Services and Suppliers, 171, Basaveshwaranagara, Hebbal, Mysuru 570016, India.

[2]Department of Chemistry-BMC Biochemistry, University of Uppsala, Husargatan 3, 75237, Uppsala, Sweden.

[3]5/1B, Krutika Co-Op Housing Society, Tejas Nagar, Kothrud, Pune 411039, India.

*Correspondence email: bt.arpita@gmail.com

Abstract:

Recent studies have shown the efficacy of hybrid COVID-19 vaccines using wild-type nucleocapsid (N) and Spike (S) protein. We upgrade this strategy by one step further using clinically proven spike protein (but with delta and post-delta omicron mutations) and nucleocapsid peptides conferring T-cell immunity. As per the latest research nucleocapsid peptides are perfect immunological replacement of nucleocapsid protein. Therefore, peptide linking strategy is pursued (economic for cellular biosynthesis than whole protein). One envelope peptide with potent T-cell response is also chosen. This peptide is also functionally indispensable for the virus. All these peptides were clustered in our designed cytoplasmic domain separated by non-immunogenic helical linkers. We also propose the idea of introduction of any T-cell peptide similar to other Human Corona Viruses (HuCoV) in these linker regions whenever required. In addition to COVID, the same approach can be applied for any emergency or even long-term unsolved outbreaks of Influenza, Dengue and West Nile Virus etc. In this era of novelty as presented by subunit and nucleic acid vaccines, multiepitope strategies like this can help to combat multiple diseases successfully in real time to give hope for better future.

Introduction:

COVID-19 has brought the world to an unprecedented standstill since last four years (1). The massacre and chaos created are prevailing still. As this infection demonstrates wide clinical spectrum, from asymptomatic, mild, moderate disease to even long tern positivity and mortality (2). Also, new variants are appearing regularly with mutations in anti-spike antibody binding sites (3) causing new panic. Although, mortality in infected patients is lower than earlier SARS, serious further complications are seen in reinfected and even in vaccinated populations. These include myocarditis, damage to CNS etc that too in 30-40 years' population (4, 5). Reinfection with emerging newer variants is also a cause of concern (6, 7).

In this scenario question arises on sustainability of vaccines based on only wild type spike. Spike protein which generates neutralizing antibodies (8) is essential of course for recovery. But problem is direct mucosal immune response is short term. Long-term memory in mucosal immune response is generated by T cell response (8, 9). In children older than 2 years, infections by related corona viruses (Common cold) generate mild and asymptotic response against OC43, HKU1, 229E, NL63 (HuCoV) (10). Several immune response studies on HuCoV infections have been carried out (10, 11, 12, 13) so far. By the age of 2-5 years, 80% children are immunized and have persistent neutralizing antibodies (12).

Even COVID-19 infection demonstrated that T cell response is important to generate memory response and lower severity in infected patients (14, 15, 16, 17). Other studies concluded that nucleocapsid protein has to be included in further vaccine strategy (18, 19, 20, 21). This can be a promising strategy to curb circulating Delta (22), Omicron variants (23) and hybrids (24) simultaneously. Therefore, it is

important to design a vaccine with delta-omicron spike (22, 23) protein and T cell epitopes (Majorly from patient studies in literature). For this, we have designed a construct using chimeric COVID spike and COVID T cell epitope sequences. Structural proteins with almost conserved sequences for different corona virus families were targeted for potential T-cell enticing candidates. Functional proteins were not considered since some of them are known to hijack MHC systems in human (25, 26, 27). There are numerous studies now for homology of nucleocapsid protein sequences between COVID and HuCoV strains (18, 28, 29, 30, 31). Also, T-cell response of COVID unexposed individuals to COVID was observed for at least one peptide from envelope protein (32) indicating potential effect from common cold immunity. Therefore, we exclusively focused on nucleocapsid protein and envelope protein for T cell epitopes and compared with convalescent patients' T cell response inducing sequences from literature. Also, more attention was given on CD4+ TCR epitopes than CD8+. Since the former sequences were found to be more stable than the other as per studies on patients (33). Finally, the designed sequence had 5 epitopes joined by linkers as cytoplasmic region. Among these, at least two epitopes were found folded after model building for the sequence without trans-membrane domain. As more well-formed structural epitopes can ensure more CD4+ T-cell activity (34) to cause persistent immunity. Thus we propose that this sequence should be considered for final vaccine construct against COVID and any other future variants/viruses. Because, the linker regions can also be used as potential sites for insertion of epitopes against new viruses in future.

Methods:

Clustal Omega analyses of SARS-1 and SARS-2 spike sequences and B-cell epitopes prediction: Clustal O alignment (35) of SARS-1 and SARS-2 (COVID) spike native sequences (from Uniprot) (36) was done using EBI server. The sequences were further subjected to B-cell epitope prediction in IEDB server (37). The predictions were based on Kolaskar method, surface accessibility and a hybrid mode of both. The predicted B-cell epitopes were colour coded according to the method used (Fig. 4). The individual protein domains were also mapped on the aligned sequences using information from both clustal alignment and Uniprot. The amino acid mutations corresponding to Delta and Omicron spikes were also taken from Uniprot. All these mutations are substituted in native COVID spike sequence to design hybrid spike. The native furin site (PRRAR) is kept since it is a hotspot for immune response (from Uniprot).

Selection of peptides from literature: T cell epitopes from COVID recovered convalescent patients' studies were selected from available vast literatures (14, 15, 17, 28, 29, 30, 32, 38). The search was more for CD4+ epitopes as it was found CD8+ epitopes were heavily mutated (causing nonpersistent immunity) in different patients' samples (33). We mostly selected potent epitopes with definite secondary structures as per PDBsum. Nucleocapsid protein peptides and envelope protein peptides were

given most importance including one memory peptide from the former (with both T-cell and B-cell inducing capabilities). The similarities with homologous proteins from common cold coronavirus families as per literature studies were given utmost importance. Peptides highly similar to human proteins and closely situated in neighbouring regions in individual open reading frames (ORFs) are excluded from our selection to avoid allergenicity (39). Finally, the secondary structures of the selected peptides are verified from PDBsum (40) or Phyre2 (41) (in absence of PDB structure) analysis to prevent misfolding in the design.

Sequence conservation studies on Corona spike, Nucleocapsid, Envelope and Membrane proteins: Clustal omega sequence alignments for these proteins were generated using Uniprot data on COVID, SARS and common cold coronavirus. Further, the alignment were refined by removing gapped sequences, non-human host harbouring coronavirus homologs and partial protein fragments. The final alignment and associated sequence similarity based neighbour joining tree (without distance corrections) were generated with maximum 5 iterations as possible in Clustal O. The guide tree was plotted in iTOL (interactive Tree of Life, ver. 6.5.8, EMBL) (42).

Design of vaccine sequence: The hybrid sequence had mutated amino acids from both Delta and Omicron spikes. The core β strand was deleted from S2 region to prevent pre-to post fusion transition as shown before for coronavirus spike transitions (Fig. 6, 43). Potential Antibody dependent cellular cytotoxicity/Antibody dependent enhancement (ADCC/ADE) region (44, 45, 46) mapped from Clustal O (Fig. 4) was further fine-tuned by structural alignment with SARS spike in PyMOL (Fig. 5). The minimum region destabilizing the same was deleted. The cytoplasmic domain was engineered with careful placement of N-epitopes interconnected by non-immunogenic EAAAK linker (47, 48, 49). These epitopes were kept according to their positions in native nucleocapsid. Last two N-epitopes actually interact with each other in PDB structure (Supplementary Fig. S6). Therefore, envelope peptide (15 residues) is placed between linkers among them to help in interaction (to cancel zero neighbour interaction effect from rigid EAAAK linker helices).

***de novo* model building:** The resulting sequence is first modelled in Robetta server (RoseTTAFold, BakerLab) using default settings (50). Further, the model buiding was done in AlphaFold in Uppsala server (UPPMAX, Uppsala Multidisciplinary Center for Advanced Computational Science, Uppsala University, Sweden) (51, 52). The analyses were done in alpha-viewer python package (Severin Dicks, IBSM Freiburg) and PyMOL (The PyMOL Molecular Graphics System, Open-Source version, Schrödinger, LLC). In parallel, models for hybrid spike without transmembrane domain and wild type COVID spike were also made by AlphaFold. For comparison, all three models were also made by monomer pTM (predicated template modelling) mode to obtain PAE (predicated aligned error) map. Finally, predicted local distance difference test plots (pLDDT) were calculated for all models.

Results:

Uniprot analyses:

Uniprot mining of COVID strcutural proteins (spike, nucleocapsid, membrane and envelope protein) were divided into 3 parts: a) Retrieval of COVID, SARS, common cold and MERS sequences, b) Clustal Omega alignment of each structural protein from Uniprot and c) Sequence similarity based neighbouring-joining tree generation (without distance values) in Clustal Omega. At first, sequences corresponding to coronaviruses infecting all host animals were considered. For nucleocapsid sequences, we found out inclusion of MERS nucleocapsid generates a distinct tree which slowly diverges from other corona including COVID (Fig 2A). Further, we generated circular tree by filtering out gapped sequences, protein fragments and non-human host infecting corona viruses. This confirmed that MERS nucleocapsid protein sequences indeed form a distinct tree, but by being more conserved as compared to others (Fig 2B). In fact, it did not even diverge much within MERS family (Fig. 2B). Usually, nucleocapsid protein is highly conserved than spike (53) and more immunogenic (53, 54). It is also essential for maintaining RNA structure, packaging etc inside viral particle (55). If random mutations arise in nucleocapsid, virus will not survive in long run. Taking this MERS scenario, introduction of recurrent random mutations in nucleocapsid protein (by immunological attack) may be beneficial for vaccine design. Although, COVID (SARS-2) nucleocapsid has more mutations than MERS, still the frequency is much low compared to spike (Fig. 2B, Supplementary Fig. S1). Alongside we also generated similar sequence-similarity trees with other structural proteins (Spike, Membrane and Envelope) (Supplementary Fig. S1, S2 and S3). The focus was on 2 conditions: 1) Conserved structural proteins among living corona viruses (MERS, COVID and common cold corona viruses; not SARS as it is already extinct by natural immunity); 2) How immunogenic these proteins are. There we found out that envelope protein also can be a suitable candidate like nucleocapsid. This protein sequences were highly conserved among COVID strains (Supplementary Fig. S3). Thus we have taken one non-allergenic (39) epitope from envelope protein which is important for binding to human cell junction proteins (56) and has robust immune response (32). Surprisingly, spike protein sequences were also found highly conserved among MERS (Supplementary Fig. S1). As already chosen in existing vaccine candidates including our design, not much study was done on this further. Membrane protein (M) was found highly conserved among extinct SARS strains (Supplementary Fig. S2) and immunologically less persistent (described below), therefore not chosen.

Literature hunting for potential T-cell epitopes for vaccine design:

Immunogenic sequences of COVID nucleocapsid protein and envelope protein were carefully shortlisted from literature studies. Initially, the focus were on the sequences with overlapping CD4+ and CD8+ epitopes. Similarity with common cold viral sequences were also taken into consideration. The data available from recovered patient and asymptomatic population were given highest importance. Final

chosen peptides were screened based on their structures (from PDBsum) or from secondary structure predicted by Phyre2 server (in absence of PDB for envelope protein; Supplementary Fig. S7). As known, SARS-2 CD8+ epitopes are hotspots of nonsynonymous mutations to cause short-lived immunity (33). Since CD8+ epitopes are mostly unstructured, special consideration was made to choose epitopes with definite structures to attract CD4 based immunity. Subsequently, two epitopes were found folded in Alpha Fold. For the rest, we found one epitope fully and another partially non-structural after validation in Alpha Fold (described below). They were kept in the final construct based on their ability to provoke instantaneous immune response due to high similarity with common cold (28, 29, 30) and strong memory response (18, 38). Membrane protein epitopes were found less persistent as per the literature study (32, 54) with exception of highly COVID-exposed hospital workers (57) who are more immune to COVID than any other population. The chosen peptides are described in Supplemental Table 1.

Model building in RoseTTAFold (Robetta server, Baker lab, HHMI) and further validation in AlphaFold (Deep Mind): The designed sequence had signal peptide, transmembrane domain and a cytoplasmic domain formed by T-cell epitopes interconnected by EAAAK linker. Major portion of ADCC/ADE region and core β-strand were removed. The positively charged Endoplasmic reticulum (ER) exit signal from COVID native spike was placed at the cytoplasmic end to facilitate easy exit from ER (58). This sequence was first modelled in RoseTTAFold and found to be structurally similar to native spike, all domains including cytoplasmic domain are distinctly well formed (Fig. 7). To validate further, we have modelled in Alpha Fold for two sequences: with and without transmembrane (ΔTM) region. It was seen the at least two epitopes (NFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSR and FYVYSRVKNLNSSRV) were folded in ΔTM sequence (Fig. 8). While first sequence is well folded in the TM+ (with transmembrane) construct, but not the other (Fig. 10). It appeared like that the first sequence is the major driver of folding for the designed cytoplasmic domain. Also inter-residual distance analysis in PyMOL showed almost double distance (16.2 Å) between this peptide and last sequence or memory epitope (QVILLNKHIDAYKTFPPTEPKKDKKKKADET) in TM+ construct than the other (Fig. 9). The other peptide (DLSPRWYFYYLGTGP) and memory epitope are fully unstructured in TM+ construct. Memory epitope is partially structured in ΔTM possibly due to closer distance (8.2 Å) (Fig. 11A) to folding driver peptide. Seemingly, this closeness is resulted from absence of transmembrane domain in ΔTM which otherwise pulled away (Fig. 9B) the cytoplasmic domain from main protein in TM+ model. In fact, both TM+ model and native spike (having transmembrane region) have shown unstructured cytoplasmic domain which is further validated by PAE scores similarities (Supplementary Fig. S9, S10) in this domain for these two models. For the main protein, the receptor binding doman (RBD) is found to be well folded and identical among all three models including native spike (Supplementary Fig. S4). Also, the core β-strand is replaced by two discontinuous stretches of β-strands which can form hydrogen bonds to the complementary strand in core β-sheet (Supplementary Fig. S5).

Thereby, the structural integrity of this region can be maintained. Although, the lengths of these strands are not enough to support pre- to post fusion transition of spike unlike native core β-strand.

Discussion:

This paper describes bioinformatics design of a molecule with potential to overcome COVID and any possible viral attack in future. Conserved structural proteins of corona viruses are targeted in our study. Functional proteins are avoided as some of them have reported adverse effects on immune system. Therefore, our vaccine constructs are designed with SARS-2 structural proteins' T-cell evoking epitopes screened from the literature. The hallmark of current COVID vaccination is COVID wild type spike protein because of its potent antigenicity. But presence of vaccine evading variants pushed us to add mutant spike in our design. In pre-vaccination era, there was deadly delta variant which was followed by milder omicron variants after mass vaccination. This massive paradigm shift predicts future variants will be more like these two and/or hybrids to evade vaccine generated immunity! Therefore, hybrid spike sequence with both delta and omicron mutations is used. Furthermore, spike protein's S2 segment has regions which do not generate neutralizing antibodies and do not have dominant T cell response. There is also a short cytoplasmic domain. We have replaced it by inserting COVID's T cell epitopes interconnected by non-immunogenic, helical & rigid linker sequence EAAAK (47, 48, 49). To incorporate all these, we also have removed a region known as core β-strand (buried deep inside spike core, structurally dispensable in pre-fusion spike and immunologically less significant) (Fig. 4, Fig. 6) from the S2 region of spike. This can help to add new segment to the whole designed molecule without making it too big. Additionally, the absence of this region can prevent the spike to go from pre- to post fusion conformation (Fig. 6). Thus, standard proline-proline (non-viral sequence) insertion which is used in other spike based vaccines (59) is not needed. We also have removed a putative region in S1 which caused ADE responses in SARS as per literature study and structural analysis (Fig. 5B). Although ADE response is diminished in current COVID strains, there is no warranty it will not resurge in incoming variants! Since there is already report of COVID vaccine induced ADE type symptoms (Vaccine-associated enhanced respiratory pathology) in hamsters (60)! Finally designed construct have two sequences: with and without transmembrane domain. Both are with engineered cytoplasmic regions replacing native domains.

For selection of T -epitopes to build cytoplasmic domain, non-allergenic peptides among nucleocapsid's numerous T cell epitopes (Fig. 3) were enlisted including memory peptide (18, 38) which helps strong immune responses in recovered and vaccinated patients (besides mouse models). These sequences are also highly similar to other common cold coronavirus (HuCoV) nucleocapsid protein sequences (Fig 1). As all humans have immune response against HuCoV from childhood driven by pulmonary immune memory including CD4+ epitopes. So using these epitopes has an advantage to generate long term and quick immune response (13, 21). This establishes their importance in persistent immunity. Next

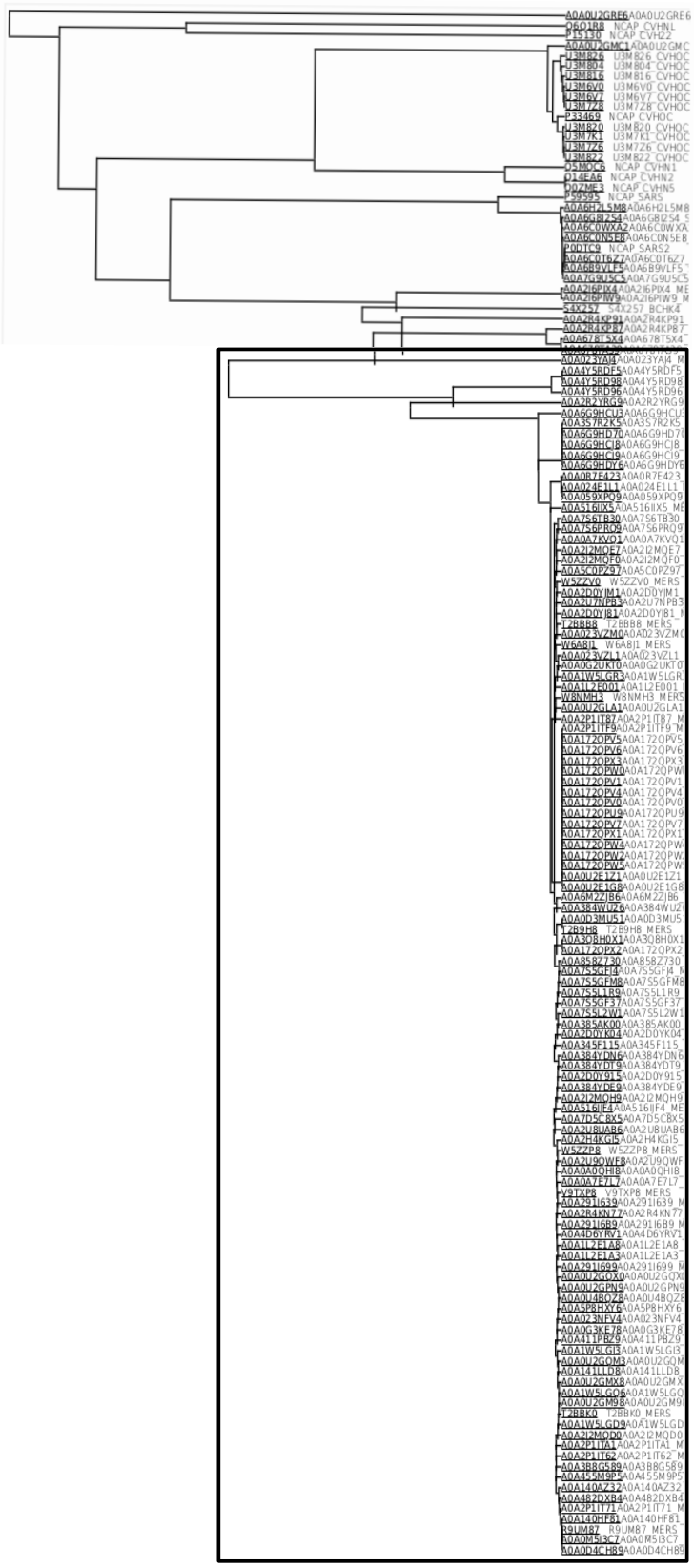
questions were, will these peptides cause misfolding in the generated construct? Subsequent PDBsum analysis helped to screen nucleocapsid peptides which were well formed with definite secondary structures (Fig 3B). For envelope peptide without any PDB structure, we have relied on Phyre2 result (Supplementary Fig. S7). Later model building in RosettaFold showed these peptides as alpha-helical in the designed cytoplasmic domain of the construct. Final Alpha Fold model building predicted that at least two peptides are folded while the other two epitopes are mostly unstructured in construct without transmembrane region. One of the folded epitopes can be the main player in folding for other subsequent peptides. The predicted LDDT values were >70 for this epitope in both ΔTM and +TM sequences (Fig. 8, Fig. 10). Also, we found out that the transmembrane domain pulls away the cytoplasmic domain from the other portions possibly due to its hydrophobicity in +TM construct. The resulting distance between last epitope and folding driver epitope doubles in +TM construct than ΔTM construct. Interestingly, the folding driver peptide was constructed by joining numerous CD4+ and CD8+ responsive N-peptides from immunological studies (Supplemental Table 1). In native Nucleocapsid protein, they are located as flanked to each other, parsed by immune system as numerous small peptides. Apart from immunological significance, they might play a role in folding of native nucleocapsid around viral RNA. Further studies on this can be done to prove this hypothesis.

As conclusion, we propose that this ΔTM construct with linker joined cytoplasmic epitopes can be a potential vaccine candidate against deadly COVID or any other virus! As other than providing support for the structure and immunogenic framework, there is also a future perspective for the linkers used in our construct. We keep the window open for inclusion of new epitopes replacing the spacers and/or obsolete (if happens to be) epitopes used in our construct. Post-COVID world is experiencing a lot. From human infecting ability of several zoonotic viruses like monkeypox virus (61), sarbecoviruses (62), lumpy skin disease viruses (63) etc to rise of deadly viruses like dengue etc (64). All these pinpoints to the need of a combinatorial strategy for pre-priming the immune system to combat COVID like pandemic and new infections of exotic and existing viruses thereafter. Also, currently there is an abyssal void for the treatment options for the children below 2 years, since most focus for current vaccination strategies are from adolescence. All these can be rectified with a readily available vaccine construct for all age groups with a potential for inclusion of epitopes in emergency.

```
sp|Q6Q1R8|NCAP_CVHNL          FPPPSFYMPLLVSSDKAPYRVIPRNLVPIGKGN-KDEQIGYWNVQER--WRMRRGQRVDL     72
sp|P15130|NCAP_CVH22          RIPYSLYSPLLVDS-EQPWKVIPRNLVPINKKD-KNKLIGYWNVQKR--FRTRKGKRVDL     74
sp|P59595|NCAP_SARS           NNTASWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATR-RVRGGDGKMKEL    105
tr|A0A6H2L5M8|A0A6H2L5M8_SARS2 NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRATR-RIRGGDGKMKDL    104
tr|A0A6G8I2S4|A0A6G8I2S4_SARS2 NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRATR-RIRGGDGKMKDL    104
tr|A0A6C0WXA2|A0A6C0WXA2_SARS2 NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRATR-RIRGGDGKMKDL    104
tr|A0A6C0N5E8|A0A6C0N5E8_SARS2 NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRATR-RIRGGDGKMKDL    104
sp|P0DTC9|NCAP_SARS2          NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRATR-RIRGGDGKMKDL    104
tr|A0A7G9U5C5|A0A7G9U5C5_SARS2 NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRATR-RIRGGDGKMKDL    104
tr|U3M826|U3M826_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M804|U3M804_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M816|U3M816_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M6V0|U3M6V0_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M6V7|U3M6V7_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M7Z8|U3M7Z8_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
sp|P33469|NCAP_CVHOC          VPYYSWFSGITQFQKGKEFEFVEGQGPPIAPGVPATEAKGYWYRHNRGSFKTADGNQRQL    119
tr|U3M820|U3M820_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M7K1|U3M7K1_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M7Z6|U3M7Z6_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
tr|U3M822|U3M822_CVHOC        VPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYRHNRRSFKTADGNQRQL    119
sp|Q5MQC6|NCAP_CVHN1          IPHYSWFSGITQFQKGRDFKFSDGQGVPIAFGVPPSEAKGYWYRHSRRSFKTADGQQKQL    118
sp|Q14EA6|NCAP_CVHN2          IPHYSWFSGITQFQKGRDFKFPDGQGVPIAYGIPPSEAKGYWYKHNRRSFKTADGQQKQL    118
                                  * :  :         ..   :  **       .  **:    *   :   *:  :*

sp|Q6Q1R8|NCAP_CVHNL          PPKVHFYYLGTGPHKDLKFRQRSDGVVWVAKEGAKTV-NTSLGNRKRNQKPLEPK---FS    128
sp|P15130|NCAP_CVH22          SPKLHFYYLGTGPHKDAKFRERVEGVVWVAVDGAKTE-PTGYGVRRKNSEPEIPH---FN    130
sp|P59595|NCAP_SARS           SPRWYFYYLGTGPEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNAATVLQLPQG    165
tr|A0A6H2L5M8|A0A6H2L5M8_SARS2 SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQG    164
tr|A0A6G8I2S4|A0A6G8I2S4_SARS2 SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQG    164
tr|A0A6C0WXA2|A0A6C0WXA2_SARS2 SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQG    164
tr|A0A6C0N5E8|A0A6C0N5E8_SARS2 SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQG    164
sp|P0DTC9|NCAP_SARS2          SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQG    164
tr|A0A7G9U5C5|A0A7G9U5C5_SARS2 SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQG    164
tr|U3M826|U3M826_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M804|U3M804_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M816|U3M816_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M6V0|U3M6V0_CVHOC        LPRWYFYYLGTGPHAKDQYGTDFDGVFWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M6V7|U3M6V7_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M7Z8|U3M7Z8_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
sp|P33469|NCAP_CVHOC          LPRWYFYYLGTGPHAKDQYGTDIDGVYWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M820|U3M820_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGIYWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M7K1|U3M7K1_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGIYWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M7Z6|U3M7Z6_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGIYWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
tr|U3M822|U3M822_CVHOC        LPRWYFYYLGTGPHAKDQYGTDIDGIYWVASNQADVNTPADIVDRDPSSDEAIPTRFPPG    179
sp|Q5MQC6|NCAP_CVHN1          LPRWYFYYLGTGPYANASYGESLEGVFWVANHQADTSTPSDVSSRDPTTQEAIPTRFPPG    178
sp|Q14EA6|NCAP_CVHN2          LPRWYFYYLGTGPYASSSYGDAHEGIFWVASHQADTSIPSDVSARDPTIQEAIPTRFSPG    178
                               *: :********     :    :*: *** . *          *    .         .

sp|Q6Q1R8|NCAP_CVHNL          KPRWKRVPTRE--ENVIQCFGPRDFNH---NMGDSDLVQNGVDAKGFPQLAELIPNQAAL    287
sp|P15130|NCAP_CVH22          KPRWKRQPNDDVTSNVTQCFGPRDLDH---NFGSAGVVANGVKAKGYPQFAELVPSTAAM    302
sp|P59595|NCAP_SARS           KPRQKRTATKQ--YNVTQAFGRRGPEQTQGNFGDQDLIRQGTDYKHWPQIAQFAPSASAF    315
tr|A0A6H2L5M8|A0A6H2L5M8_SARS2 KPRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAF    314
tr|A0A6G8I2S4|A0A6G8I2S4_SARS2 KPRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAF    314
tr|A0A6C0WXA2|A0A6C0WXA2_SARS2 KPRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAF    314
tr|A0A6C0N5E8|A0A6C0N5E8_SARS2 KPRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAF    314
sp|P0DTC9|NCAP_SARS2          KPRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAF    314
tr|A0A7G9U5C5|A0A7G9U5C5_SARS2 KPRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAF    314
tr|U3M826|U3M826_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M804|U3M804_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M816|U3M816_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M6V0|U3M6V0_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M6V7|U3M6V7_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M7Z8|U3M7Z8_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
sp|P33469|NCAP_CVHOC          KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M820|U3M820_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M7K1|U3M7K1_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M7Z6|U3M7Z6_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAF    323
tr|U3M822|U3M822_CVHOC        KPRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDSQFPILAELAPTAGAF    323
sp|Q5MQC6|NCAP_CVHN1          KPRQKRTPNKH--CNVQQCFGKRGPSQ---NFGNAEMLKLGTNDPQFPILAELAPTPGAF    321
sp|Q14EA6|NCAP_CVHN2          KPRQKRTPNKF--CNVQQCFGKRGPLQ---NFGNSEMLKLGTNDPQFPILAELAPTPGAF    321
                              *** **  .     .* *.** *.  :   *:*.  ::  *..   :* :*:: *. .*:

sp|Q6Q1R8|NCAP_CVHNL          FFDSEVSTDEVG-----------DNVQITYTYKMLVAKDNKNLPK----FIEQISAFTKP    332
sp|P15130|NCAP_CVH22          LFDSHIVSKESG-----------NTVVLTFTTRVTVPKDHPHLGK----FLEELNAFTRE    347
sp|P59595|NCAP_SARS           FGMSRIGMEVTP-----------SGTWLTYHGAIKLDDKDPQFKDNVILLNKHIDAYKTF    364
tr|A0A6H2L5M8|A0A6H2L5M8_SARS2 FGMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF    363
tr|A0A6G8I2S4|A0A6G8I2S4_SARS2 FGMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF    363
tr|A0A6C0WXA2|A0A6C0WXA2_SARS2 FGMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF    363
tr|A0A6C0N5E8|A0A6C0N5E8_SARS2 FGMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF    363
sp|P0DTC9|NCAP_SARS2          FGMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF    363
tr|A0A7G9U5C5|A0A7G9U5C5_SARS2 FGMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTF    363
tr|U3M826|U3M826_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M804|U3M804_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M816|U3M816_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M6V0|U3M6V0_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M6V7|U3M6V7_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M7Z8|U3M7Z8_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
sp|P33469|NCAP_CVHOC          FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M820|U3M820_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M7K1|U3M7K1_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M7Z6|U3M7Z6_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
tr|U3M822|U3M822_CVHOC        FFGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQ    383
sp|Q5MQC6|NCAP_CVHN1          FFGSKLDLVKRD---SEADSPVKDVFELHYSGSIRFDSTLPGFETIMKVLEENLNAYVNS    378
sp|Q14EA6|NCAP_CVHN2          FFGSKLELFKRD---SDADSPSKDTFELRYSGSIRFDSTLPGFETIMKVLKENLDAYVNS    378
                              :  *.:                .   : :   : . .    :      : :.:.*:

sp|Q6Q1R8|NCAP_CVHNL          SSIK-------EMQSQSSHVAQNTVLNASI----------------PESKPLADDDSAI    368
sp|P15130|NCAP_CVH22          MQQH-------PLLNPSALEFNPSQ--TSP----------------ATAEPVRDEVSIE    381
sp|P59595|NCAP_SARS           PPTEPKKDKKKKTDEAQPLPQRQKKQPTVT----------------LLPAADMDDFSRQL    408
tr|A0A6H2L5M8|A0A6H2L5M8_SARS2 PPTEPKKDKKKKADETQALPQRQKKQQTVT----------------LLPAADLDDFSKQL    407
tr|A0A6G8I2S4|A0A6G8I2S4_SARS2 PPTEPKKDKKKKADETQALPQRQKKQQTVT----------------LLPAADLDDFSKQL    407
tr|A0A6C0WXA2|A0A6C0WXA2_SARS2 PPTEPKKDKKKKADETQALPQRQKKQQTVT----------------LLPAADLDDFSKQL    407
tr|A0A6C0N5E8|A0A6C0N5E8_SARS2 PPTEPKKDKKKKADETQALPQRQKKQQTVT----------------LLPAADLDDFSKQL    407
sp|P0DTC9|NCAP_SARS2          PPTEPKKDKKKKADETQALPQRQKKQQTVT----------------LLPAADLDDFSKQL    407
tr|A0A7G9U5C5|A0A7G9U5C5_SARS2 PPTEPKKDKKKKADETQALPQRQKKQQTVT----------------LLPAADLDDFSKQL    407
tr|U3M826|U3M826_CVHOC        ---D-------GMMNMSPKPQRQRGLKNGQGENDNISVAAPKSRVQQNKIRELTAEDISL    433
tr|U3M804|U3M804_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAAPKSRVQQNKSRELTSEDISL    433
tr|U3M816|U3M816_CVHOC        ---D-------GMMNMSPKPQRQRGLNNGQGENDNISVAAPKSRVQQNKSRELTAEDISL    433
tr|U3M6V0|U3M6V0_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAAPKSRVQQNKSRELTAEDISL    433
tr|U3M6V7|U3M6V7_CVHOC        ---D-------GMMNMSPKPQRQRGLKNGQGENDNISVAAPKSRVQQNKSRELTAEDISL    433
tr|U3M7Z8|U3M7Z8_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAAPKSRVQQNKSRELTAEDISL    433
sp|P33469|NCAP_CVHOC          ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAVPKSRVQQNKSRELTAEDISL    433
tr|U3M820|U3M820_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAVPKSRVQQNKSRELTAEDISL    433
tr|U3M7K1|U3M7K1_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQEENDNISVAVPKSRVQQNKSRELTAEDISL    433
tr|U3M7Z6|U3M7Z6_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAVPKSRVQQNKSRELTAEDISL    433
tr|U3M822|U3M822_CVHOC        ---D-------GMMNMSPKPQRQRGHKNGQGENDNISVAVPKSRVQQNKSRELTAEDISL    433
sp|Q5MQC6|NCAP_CVHN1          NQNT-------DSDSLSSKPQRKRGVKQLPEQFDSLNLSAGTQ-----HISNDFTPEDHSL    427
sp|Q14EA6|NCAP_CVHN2          NQNT-------VSGSLSPKPQRKRGVKQSPESFDSLNLSADTQ----HISNDFTPEDHSL    427
                                                 . .   .                    .
```

Fig 1: Clustal Omega alignment of Nucleocapsid (N) protein sequences. The sequence alignment among SARS, COVID and common cold coronaviruses shows high similarity in chosen epitopes (blue). Except last several residues in last peptide, all SARS2 N-epitopes are highly similar to HuCoV sequences.

A)



B)

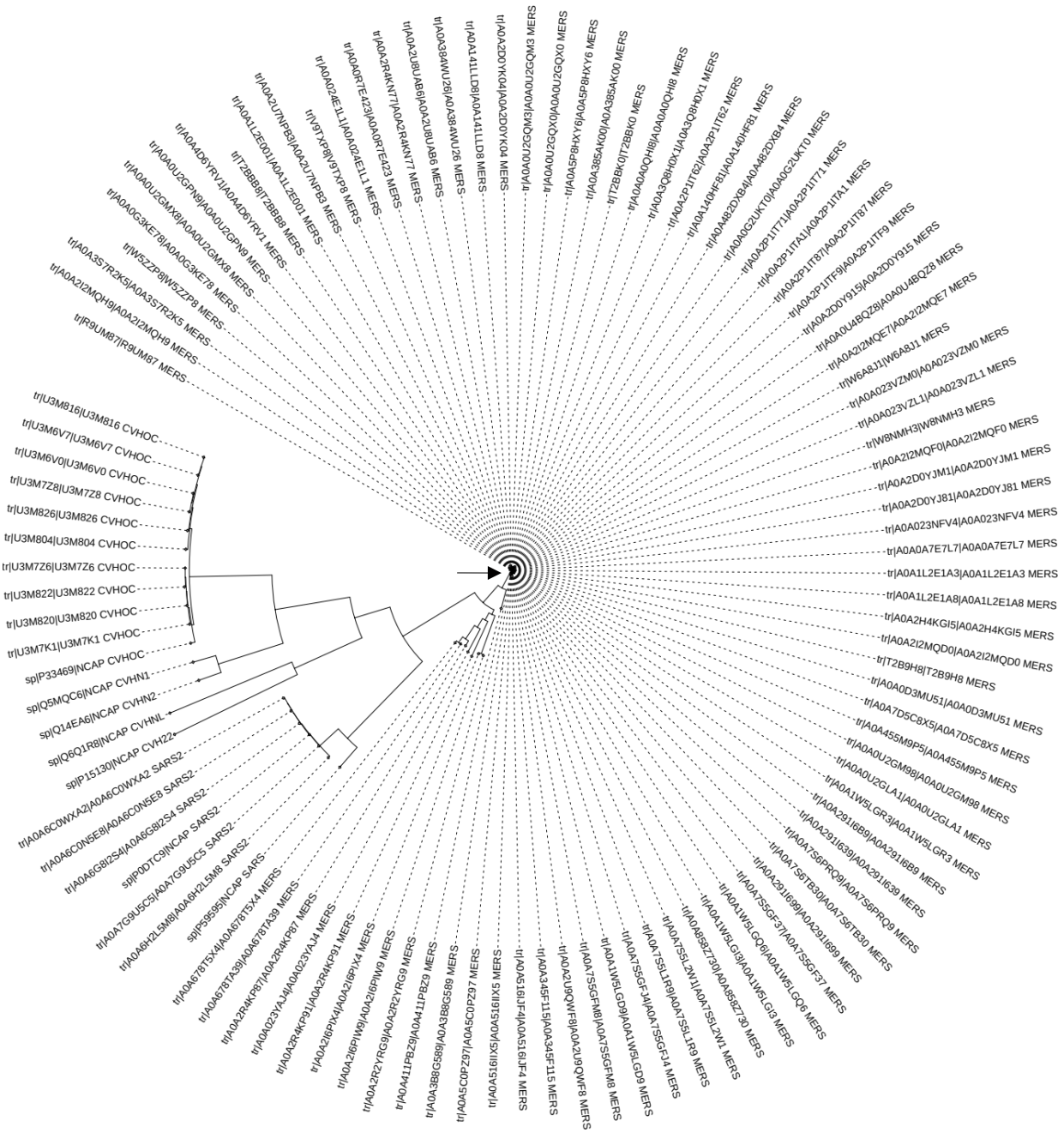


Fig 2: Sequence similarity-based guide tree of Coronavirus nucleocapsid sequences.

A) Preliminary guide tree from Clustal Omega alignment of nucleocapsid sequences. Uniprot-retrieved Corona virus N-protein sequences' alignment demonstrates distinct unique tree formation from later variants of MERS (box enclosed). Here all sequences as available from the database are used. **B)** Sequence similarity-based neighbouring-joining circular tree of coronavirus nucleocapsid proteins sequences. These are chosen from Uniprot after removing fragments and non-human host infecting coronaviruses. The circular tree is plotted in iTOL. Arrow marked are MERS nucleocapsid proteins showing very little changes in their sequences during as compared to others.

A)

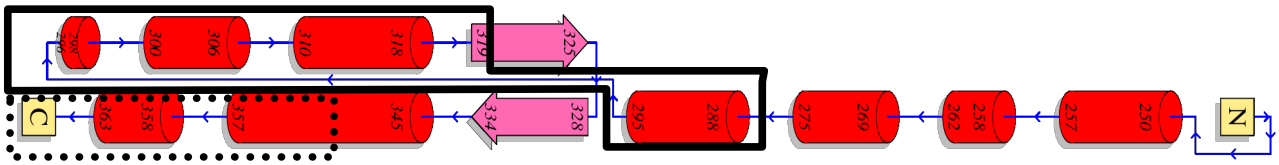
>Nucleocapsid protein

MSDNGPQNQRNAPRITFGGPSDSTGSNQNGERSGARSKQRRPQGLPNNTASWFTALTQHGKE
DLKFPRGQGVPINTNSSPDDQIGYYRRATRRIRGGDGKMK(103)**DLSPRWYFYYLGTGP**(117)
EAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIVLQLPQGTTLPKGFYAEGSRGGS
QASSRSSSRSRNSSRNSTPGSSRGTSPARMAGNGGDAALALLLLDRLNQLESKMSGKGQQQQ
GQTVTKKSAAEASKKPRQKRTATKAYNVTQAFGRRGPEQTQG(285)**NFGDQELIRQGTDYK**
HWPQIAQFAPSASAFFGMSR(319)IGMEVTPSGTWLTYTGAIKLDDKDPNFKD(349)**QVILLN**
KHIDAYKTFPPTEPKKDKKKKADET(379)QALPQRQKKQQTVTLLPAADLDDFSKQLQQSM
SSADSTQA

>Envelope protein
MYSFVSEETGTLIVNSVLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPS(56)**FYVYSRV**
KNLNSSRV(70)PDLLV

B)

I)



II)

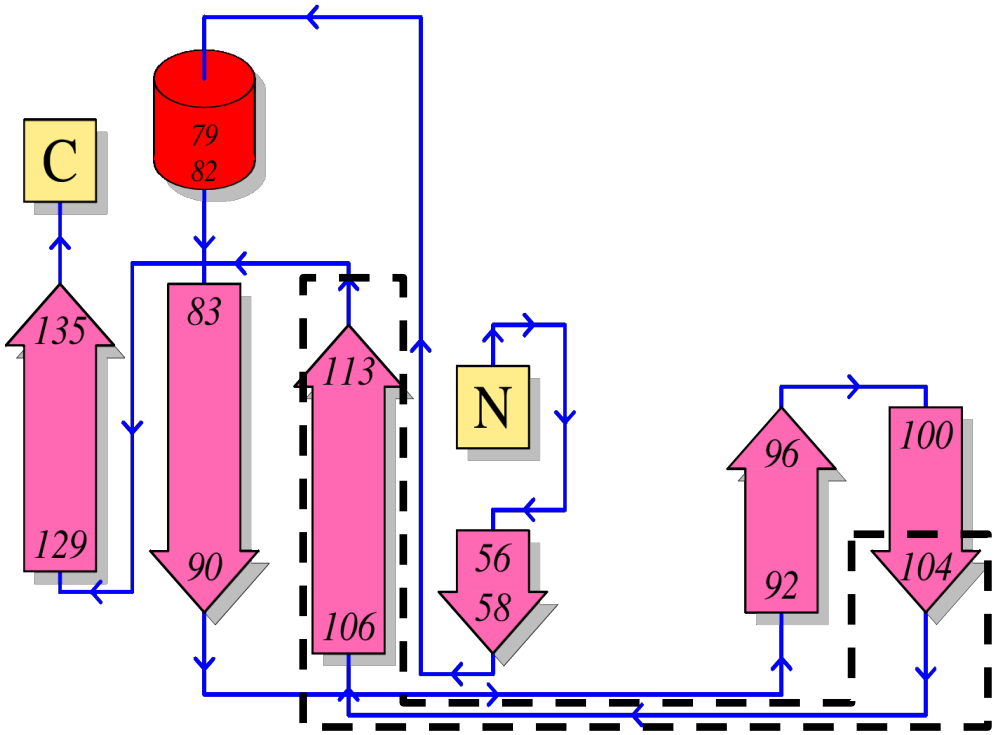


Fig 3: Wuhan COVID Nucleocapsid and Envelope protein epitopes.

A) The T-cell epitopes from N and E protein sequences (in FASTA format), homologous with common cold, chosen for the cytoplasmic domain design of the vaccine construct. These are highlighted yellow, bold-lettered and underlined. The numbers in parenthesis denote amino acid positions.

B) Secondary structures of the chosen nucleocapsid peptides in enclosures from PDBsum: I) **QVILLNKHIDAYKTFPPTEPKKDKKKKADET (6ZCO.pdb)** in dotted rectangle has mostly α-helical structure with adjoining loops; 15 residues missing at C terminus. **NFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSR (6ZCO.pdb)** in continuous lined enclosure has α-helical structures with adjoining loop. II) **DLSPRWYFYYLGTGP (7N0R.pdb)** in dashed enclosure has non-complementary β-strands along with adjoining long loop thereby strong possibility of forming unstructured epitope in construct design. The core sequence in this epitope is **SPRWYFYYL** which induces CD8 response (Supplemental Table 1). CD8+ epitopes are usually linear/unstructured.

```
CLUSTAL O(1.2.4) multiple sequence alignment

                   Signal pep
NP_828851.1        MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL  60
YP_009724390.1     MFVFLVLLPL VSSQCVNLTT --RTQLPPAY --TNSFTRGV YYPDKVFRSS VLHSTQDLFL  56
                   **:**::* * .*.. ::  *        * *   .*  **** ***::***.  .*: ******

NP_828851.1        PFYSNVTGFH TIN------- HTFGNPVIPF KDGIYFAATE KSNVVRGWVF GSTMNNKSQS  113
YP_009724390.1     PFFSNVTWFH AIHVSGTNGT KRFDNPVLPF NDGVYFASTE KSNIIRGWIF GTTLDSKTQS  116
                   **:**** ** :*:        : *.***:** :**:***:** ***::***:* *:*::.*:**

NP_828851.1        VIIINNSTNV VIRACNFELC DNPFFAVSKP MGT----QTH TMIFDNAFNC TFEYISDAFS  169
YP_009724390.1     LLIVNNATNV VIKVCEFQFC NDPFLGVYYH KNNKSWMESE FRVYSSANNC TFEYVSQPFL  176
                   ::*:**:*** **:.*:*::* ::**:.*     ..    ::.   ::..* ** ****:*: *

NP_828851.1        LDVSEKSGNF KHLREFVFKN KDGFLYVYKG YQPIDVVRDL PSGFNTLKPI FKLPLGINIT  229
YP_009724390.1     MDLEGKQGNF KNLREFVFKN IDGYFKIYSK HTPINLVRDL PQGFSALEPL VDLPIGINIT  236
                   :*:. *.*** *:********  **:: :*.  : **::*****.**.:*:*:. .**:*****

NP_828851.1        NFRAILTAFS ------PAQD IWGTSAAAYF VGYLKPTTFM LKYDENGTIT DAVDCSQNPL  283
YP_009724390.1     RFQTLLALHR SYLTPGDSSS GWTAGAAAYY VGYLQPRTFL LKYNENGTIT DAVDCALDPL  296
                   .*:::*: .        :..  * :.****:* ***:* **:* **:******* ****: :**

                                                 RBD Starts
NP_828851.1        AELKCSVKSF EIDKGIYQTS NFRVVPSGDV VRFPNITNLC PFGEVFNATK FPSVYAWERK  343
YP_009724390.1     SETKCTLKSF TVEKGIYQTS NFRVQPTESI VRFPNITNLC PFGEVFNATR FASVYAWNRK  356
                   :* **::***  ::******* **** *: .: ********** *********: * *****:**

NP_828851.1        KISNCVADYS VLYNSTFFST FKCYGVSATK LNDLCFSNVY ADSFVVKGDD VRQIAPGQTG  403
YP_009724390.1     RISNCVADYS VLYNSASFST FKCYGVSPTK LNDLCFTNVY ADSFVIRGDE VRQIAPGQTG  416
                   :********* *****: *** ******* ** ******:*** *****::**: **********

                                                 ACE2 biding region starts
NP_828851.1        VIADYNYKLP DDFMGCVLAW NTRNIDATST GNYNYKYRYL RHGKLRPFER DISNVPFSPD  463
YP_009724390.1     KIADYNYKLP DDFTGCVIAW NSNNLDSKVG GNYNYLYRLF RKSNLKPFER DISTEIYQAG  476
                    ********* *** ***:** *:.*:*:.   ***** ** : *:.:*:**** ***.  :. .

                           ACE2 biding region ends
NP_828851.1        GKPCTP-PAL NCYWPLNDYG FYTTTGIGYQ PYRVVVLSFE LLNAPATVCG PKLSTDLIKN  522
YP_009724390.1     STPCNGVEGF NCYFPLQSYG FQPTNGVGYQ PYRVVVLSFE LLHAPATVCG PKKSTNLVKN  536
                   ..**.   .: ***:**:.** *  *.*:*** ********** **:******* ** **:*:**

                End of RBD                                                   ADCC/ADE
NP_828851.1        QCVNFNFNGL TGTGVLTPSS KRFQPFQQFG RDVSDFTDSV RDPKTSEILD ISPCAFGGVS  582
YP_009724390.1     KCVNFNFNGL TGTGVLTESN KKFLPFQQFG RDIADTTDAV RDPQTLEILD ITPCSFGGVS  596
                   :********** ******* *.*:* ******* *::* **:** **:* ***** :**:*****
```

```
                    Response inducing region
NP_828851.1         VITPGTNASS EVAVLYQDVN CTDVSTAIHA DQLTPAWRIY STGNNVFQTQ AGCLIGAEHV  642
YP_009724390.1      VITPGTNTSN QVAVLYQDVN CTEVPVAIHA DQLTPTWRVY STGSNVFQTR AGCLIGAEHV  656
                    *******:*. :********* **:* .**** *****:**:* ***.*****: **********
                                                Furin
                                  S1 ends                                        Core
NP_828851.1         DTSYECDIPI GAGICASYHT VSL----LRS TSQKSIVAYT MSLGADSSIA YSNNTIAIPT  698
YP_009724390.1      NNSYECDIPI GAGICASYQT QTNSPRRARS VASQSIIAYT MSLGAENSVA YSNNSIAIPT  716
                    :.******** ********:*  :      ** .:.:**:*** *****:.*:* ****:*****
                   β-strand
NP_828851.1         NFSISITTEV MPVSMAKTSV DCNMYICGDS TECANLLLQY GSFCTQLNRA LSGIAAEQDR  758
YP_009724390.1      NFTISVTTEI LPVSMTKTSV DCTMYICGDS TECSNLLLQY GSFCTQLNRA LTGIAVEQDK  776
                    **:**:***: :****:**** **.******* ***:****** ********** *:***.***:

                                                                  Fusion region (FR)
NP_828851.1         NTREVFAQVK QMYKTPTLKY FGGFNFSQIL PDPLKPTKRS FIEDLLFNKV TLADAGFMKQ  818
YP_009724390.1      NTQEVFAQVK QIYKTPPIKD FGGFNFSQIL PDPSKPSKRS FIEDLLFNKV TLADAGFIKQ  836
                    **:******* *:**** :*  ********** *** **:*** ********** *******:**
                           FR ends
NP_828851.1         YGECLGDINA RDLICAQKFN GLTVLPPLLT DDMIAAYTAA LVSGTATAGW TFGAGAALQI  878
YP_009724390.1      YGDCLGDIAA RDLICAQKFN GLTVLPPLLT DEMIAQYTSA LLAGTITSGW TFGAGAALQI  896
                    **:***** * ********** ********** *:*** **:* *::** *:** **********

NP_828851.1         PFAMQMAYRF NGIGVTQNVL YENQKQIANQ FNKAISQIQE SLTTTSTALG KLQDVVNQNA  938
YP_009724390.1      PFAMQMAYRF NGIGVTQNVL YENQKLIANQ FNSAIGKIQD SLSSTASALG KLQDVVNQNA  956
                    ********** ********** ***** **** **.**.:**: **::*::*** **********

NP_828851.1         QALNTLVKQL SSNFGAISSV LNDILSRLDK VEAEVQIDRL ITGRLQSLQT YVTQQLIRAA  998
YP_009724390.1      QALNTLVKQL SSNFGAISSV LNDILSRLDK VEAEVQIDRL ITGRLQSLQT YVTQQLIRAA 1016
                    ********** ********** ********** ********** ********** **********

NP_828851.1         EIRASANLAA TKMSECVLGQ SKRVDFCGKG YHLMSFPQAA PHGVVFLHVT YVPSQERNFT 1058
YP_009724390.1      EIRASANLAA TKMSECVLGQ SKRVDFCGKG YHLMSFPQSA PHGVVFLHVT YVPAQEKNFT 1076
                    ********** ********** ********** ********:* ********** ***:**:***

NP_828851.1         TAPAICHEGK AYFPREGVFV FNGTSWFITQ RNFFSPQIIT TDNTFVSGNC DVVIGIINNT 1118
YP_009724390.1      TAPAICHDGK AHFPREGVFV SNGTHWFVTQ RNFYEPQIIT TDNTFVSGNC DVVIGIVNNT 1136
                    *******:** *:********  *** **:** ***:.***** ********** ******:***

NP_828851.1         VYDPLQPELD SFKEELDKYF KNHTSPDVDL GDISGINASV VNIQKEIDRL NEVAKNLNES 1178
YP_009724390.1      VYDPLQPELD SFKEELDKYF KNHTSPDVDL GDISGINASV VNIQKEIDRL NEVAKNLNES 1196
                    ********** ********** ********** ********** ********** **********
                                                Transmembrane      Cytoplasmic region
NP_828851.1         LIDLQELGKY EQYIKWPWYV WLGFIAGLIA IVMVTILLCC MTSCCSCLKG ACSCGSCCKF 1238
YP_009724390.1      LIDLQELGKY EQYIKWPWYI WLGFIAGLIA IVMVTIMLCC MTSCCSCLKG CCSCGSCCKF 1256
                    ********** *********: ********** ******:*** ********** .*********

NP_828851.1         DEDDSEPVLK GVKLHYT  1255
YP_009724390.1      DEDDSEPVLK GVKLHYT  1273
```

Fig 4: Mapping of B cell epitopes regions of Wuhan COVID spike (YP_009724390.1). Native SARS2 spike RBD, ACE2 binding sequence, furin site, fusion region along with ADCC/ADE inducing regions are aligned with SARS spike (NP_828851.1) in Clustal Omega. The B-cell epitope analysis is from IEDB server. Yellow is B cell epitopes by Kolaskar method, Cyan blue is surface accessibility. Red is having both methods. These B cell epitope regions are present in delta and omicron spike also with ever-changing sequences evading immunity.

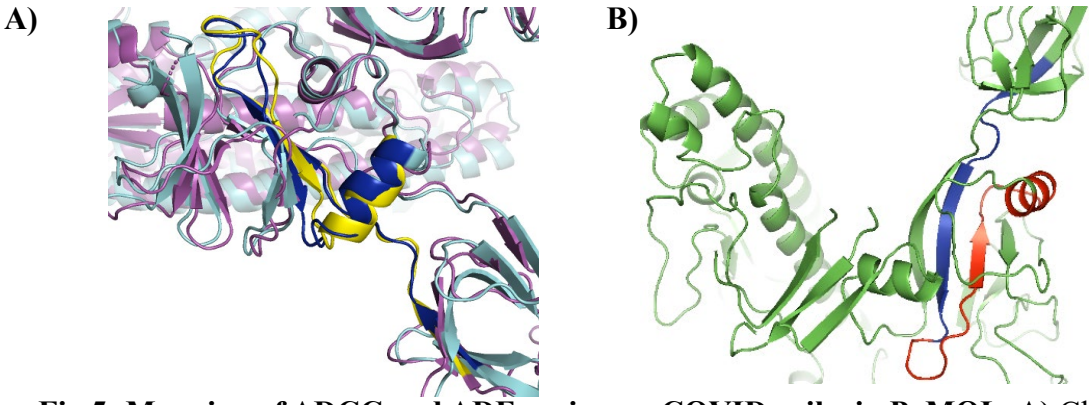


Fig 5: Mapping of ADCC and ADE region on COVID spike in PyMOL. A) Clustal O mapped region (585-625; blue) for COVID spike (cyan) is investigated by structural alignment on ADE inducing region (yellow) in SARS Spike (magenta). **B)** In vaccine construct designed, red segment (Δ600-624) was deleted as that will sufficiently weaken that region (rather a beta cage to encase any possible immunological proteins). Other non-ADE sequences are coloured as green.

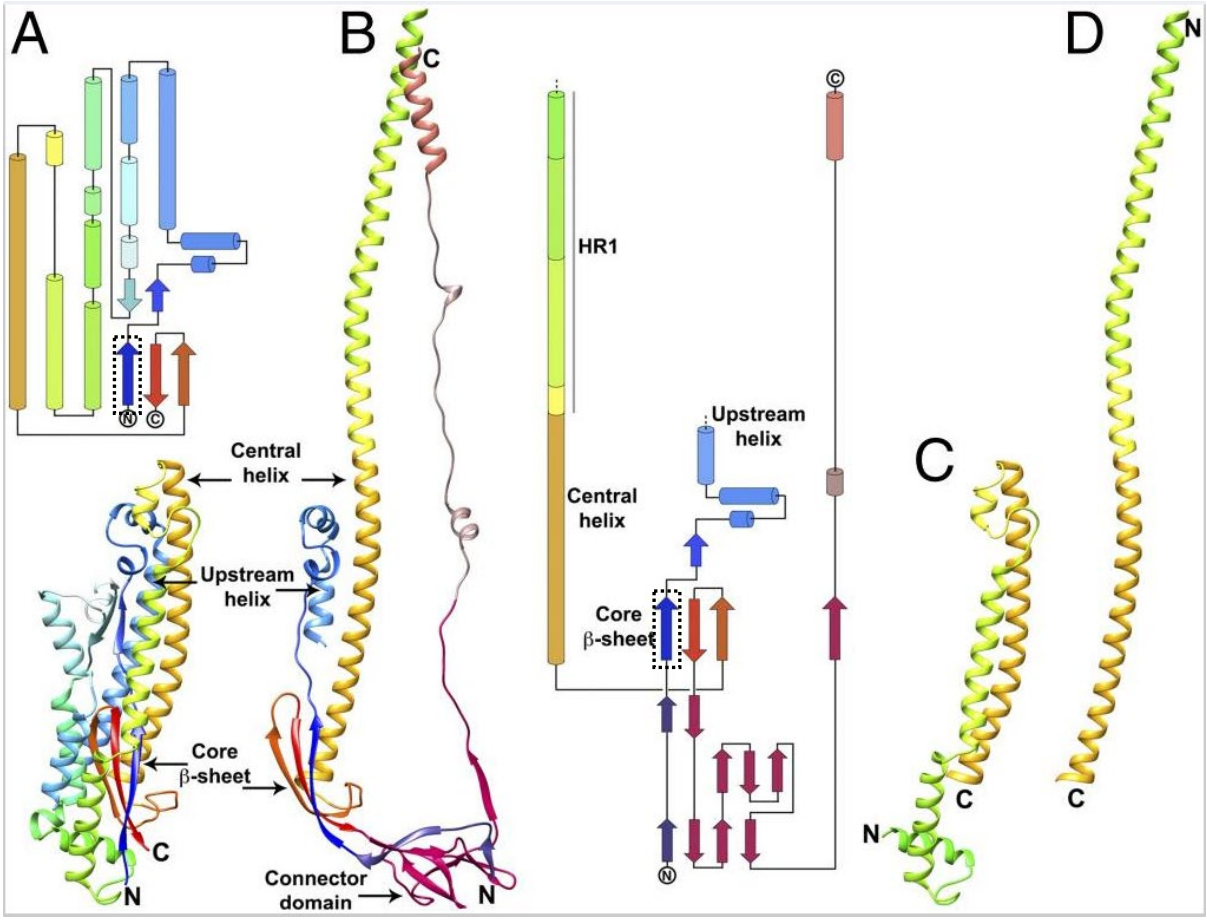


Fig 6: Core β-strand (dark blue, in dotted rectangle, residues: 711-729) is essential for pre- to post-fusion transition of spike. This strand maintains the core β-sheet by exchanging interactions with other two strands from same protomer to homologous strands of different protomer of trimeric spike. A) Pre-fusion spike S2 apparatus comprising central helix, HR region and core β-sheet. B) Post-fusion spike S2 apparatus comprising central

helix fused with HR region; Core β-sheet with exchanged 2nd and 3rd strands between protomers. Core β- strand remains unchanged. C) Central helix in pre and post-fusion conformation. [Reprinted with permission from PNAS (43)].

Final sequence

```
MFVFLVLLPLVSSQCVNLRTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS
NVTWFHVISGTNGTKRFDNPVLPFNDGVYFASIEKSNIIRGWIFGTTLDSKTQSLLIVNN
ATNVVIKVCEFQFCNDPFLDHKNNKSWMESGVYSSANNCTFEYVSQPFLMDLEGKQGNFK
NLREFVFKNIDGYFKIYSKHTPIIVEPERDLPQGFSVLEPLVDLPIGINITRFQTLLALH
RSYLTPGDSSSGLTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKS
FTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNRKRISNCVADY
SVLYNLAPFFTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQTGNIADYNYKL
PDDFTGCVIAWNSNKLDSKVSGNYNYRYRLFRKSNLKPFERDISTEIYQAGNKPCNGVAG
FNCYFPLRSYSFRPTYGVGHQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNG
LKGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITHADQLT
PTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQ
SIIAYTMSLGAENSVAYSNNSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLKRALTG
IAVEQDKNTQEVFAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLA
DAGFIKQYGDCLGDIAARDLICAQKFKGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFG
AGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQ
NVVNHNAQALNTLVKQLSSKFGAISSVLNDIFSRLDKVEAEVQIDRLITGRLQSLQTYVT
QQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVP
AQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVV
IGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEV
AKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLEAAAKDLSPRWYFYY
LGTGPEAAAKNFGDQELIRQGTDYKHWPQIAQFAPSASAFFGMSREAAAKFYVYSRVKNL
NSSRVEAAAKQVILLNKHIDAYKTFPPTEPKKDKKKKADETEAAAKDEDDSE
```

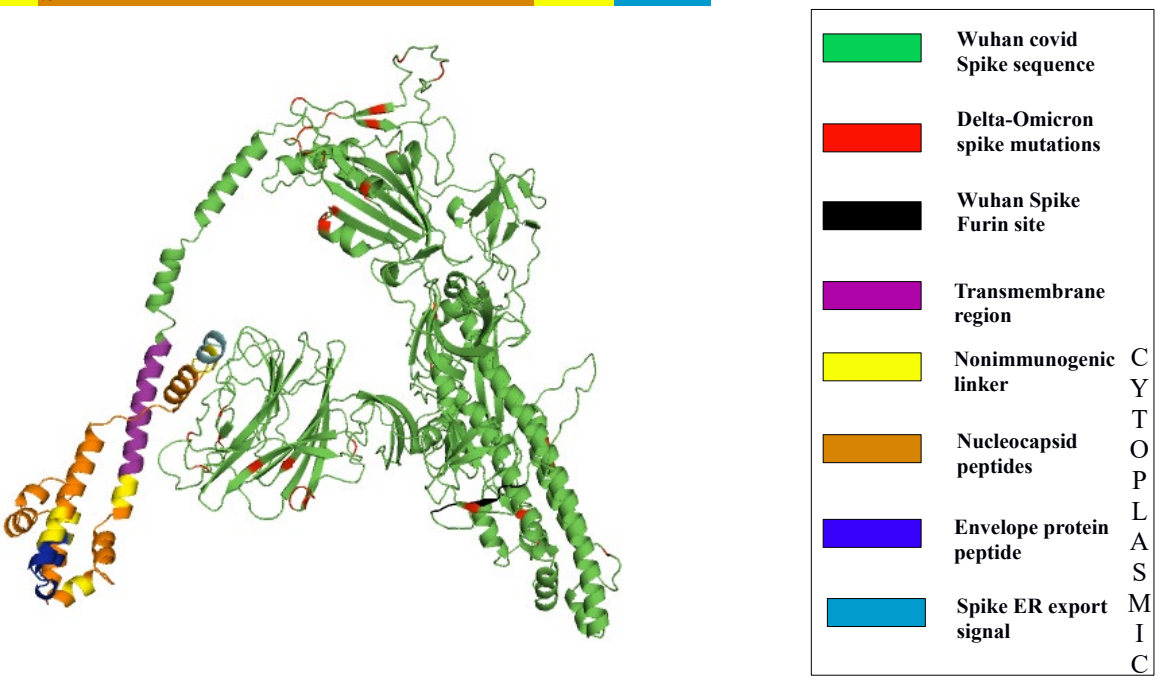


Fig 7: Sequence and 3-D model (RoseTTAFold) of the TM+ vaccine construct. The mutations corresponding to Delta and Omicron variants are coloured red. Transmembrane domain, linkers, cytoplasmic T-epitopes and terminal ER exit signals are coloured differently. All colour codes are explained in the legend on right side. The T-epitopes appear helical in this model. Overall the cytoplasmic domain looks like well folded coiled coil located close to the main protein.

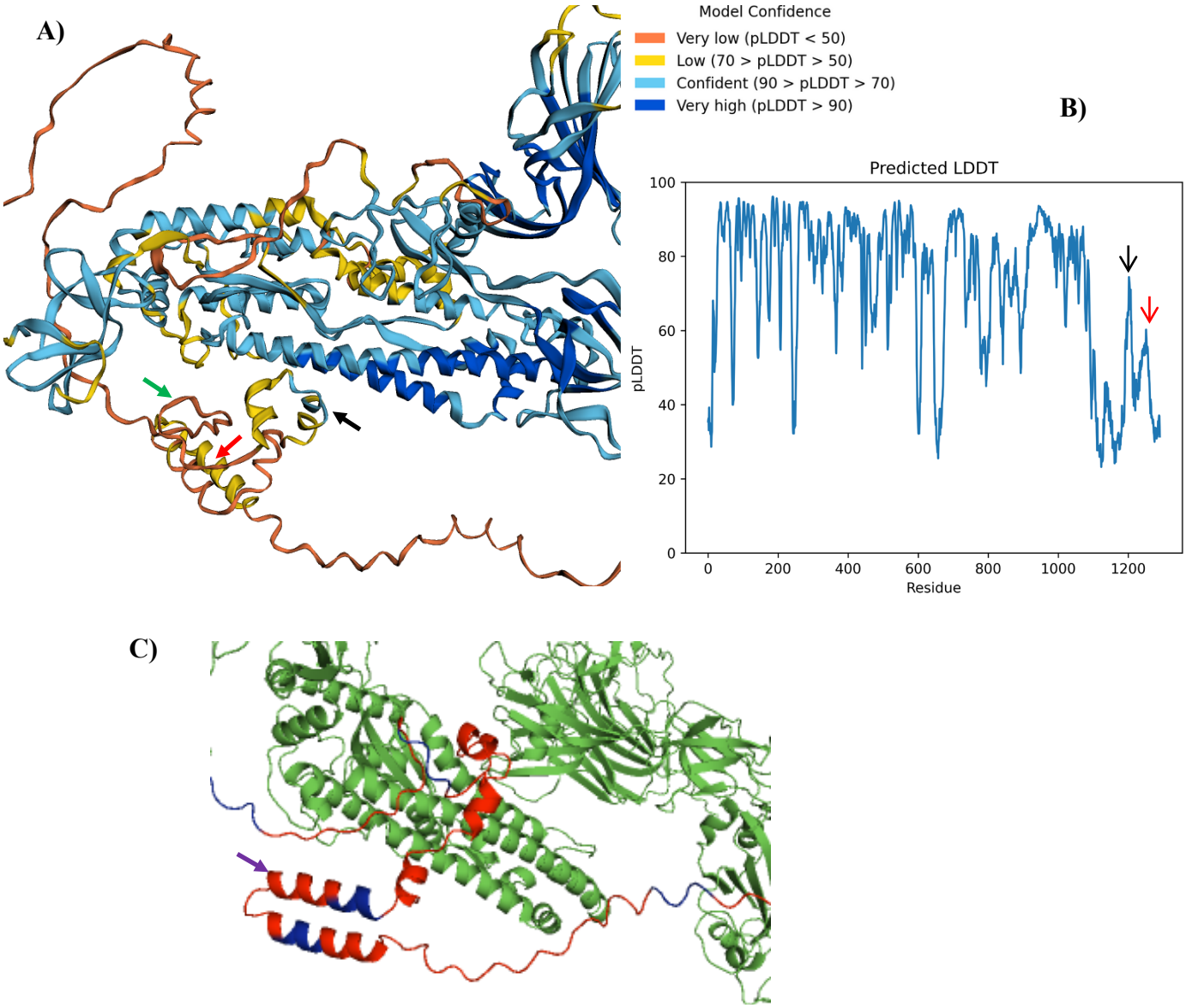


Fig 8: AlphaFold model of the final vaccine construct (without transmembrane domain) appears as well folded. A) Best model for the ΔTM construct and **B)** Predicted local distance difference test (pLDDT) plot. Except cytoplasmic domain, most of the other portions of the molecule have pLDDT values >70. The model is colour coded according to model confidence/pLDDT values in B. In cytoplasmic domain, the central region of folding driver epitope (marked by black arrow) has pLDDT value >70 and >50 for terminal portions. Similarly, part of memory peptide (red arrow marked) was folded with >50 pLDDT value. As expected from noncomplementary β-stranded structure in PDBsum, the first epitope (DLSPRWYFYYLGTGP) (green arrow marked) forms unstructured region (pLDDT<50). **C)** The envelope peptide between folding initiator peptide and memory epitope is folded as helices (from PyMOL; Purple arrow marked). The cytoplasmic epitopes and EAAAK linkers are coloured as red and blue respectively. Rest of the protein is coloured as green.

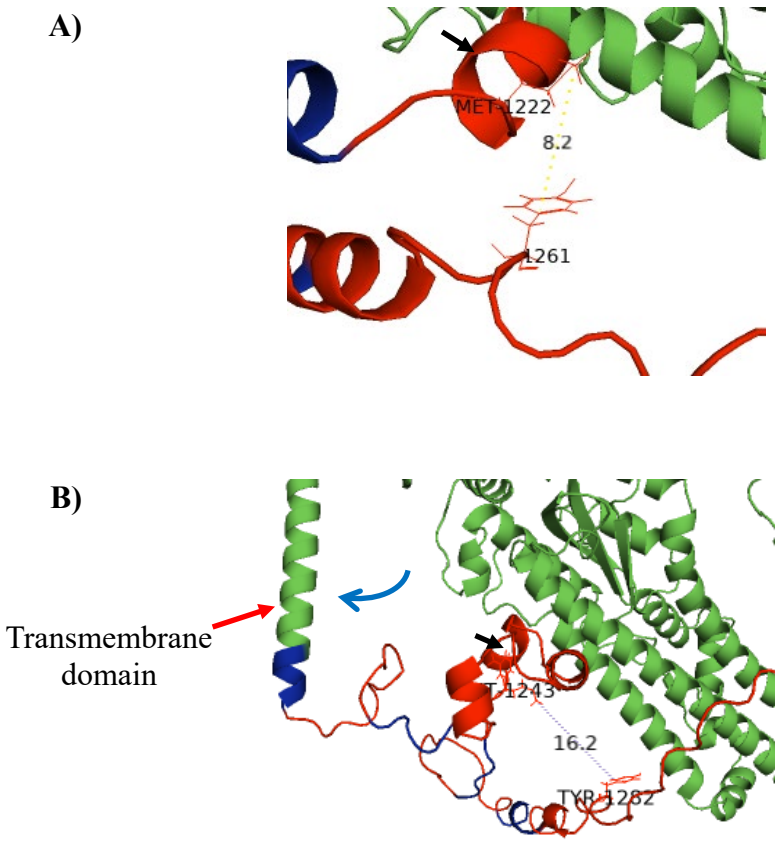


Fig 9: Folding centre of the vaccine constructs with T cell epitopes (red) joined by linkers (blue). A) Without transmembrane domain (ΔTM) and **B)** with transmembrane domain (+TM). The linkers (EAAAK) are coloured as blue while individual T-epitopes in cytoplasmic domain are coloured as red. The closest distance between folding driver epitope (marked by black arrow) and memory peptide is monitored by central methionine of the former and a tyrosine residue of the latter. In ΔTM, this distance is 8.2 Å which almost doubles to 16.2 Å in +TM construct. The transmembrane domain (marked by red arrow) is helical and seems to pull away (blue arrow) the cytoplasmic domain from rest of the protein which is coloured as green.

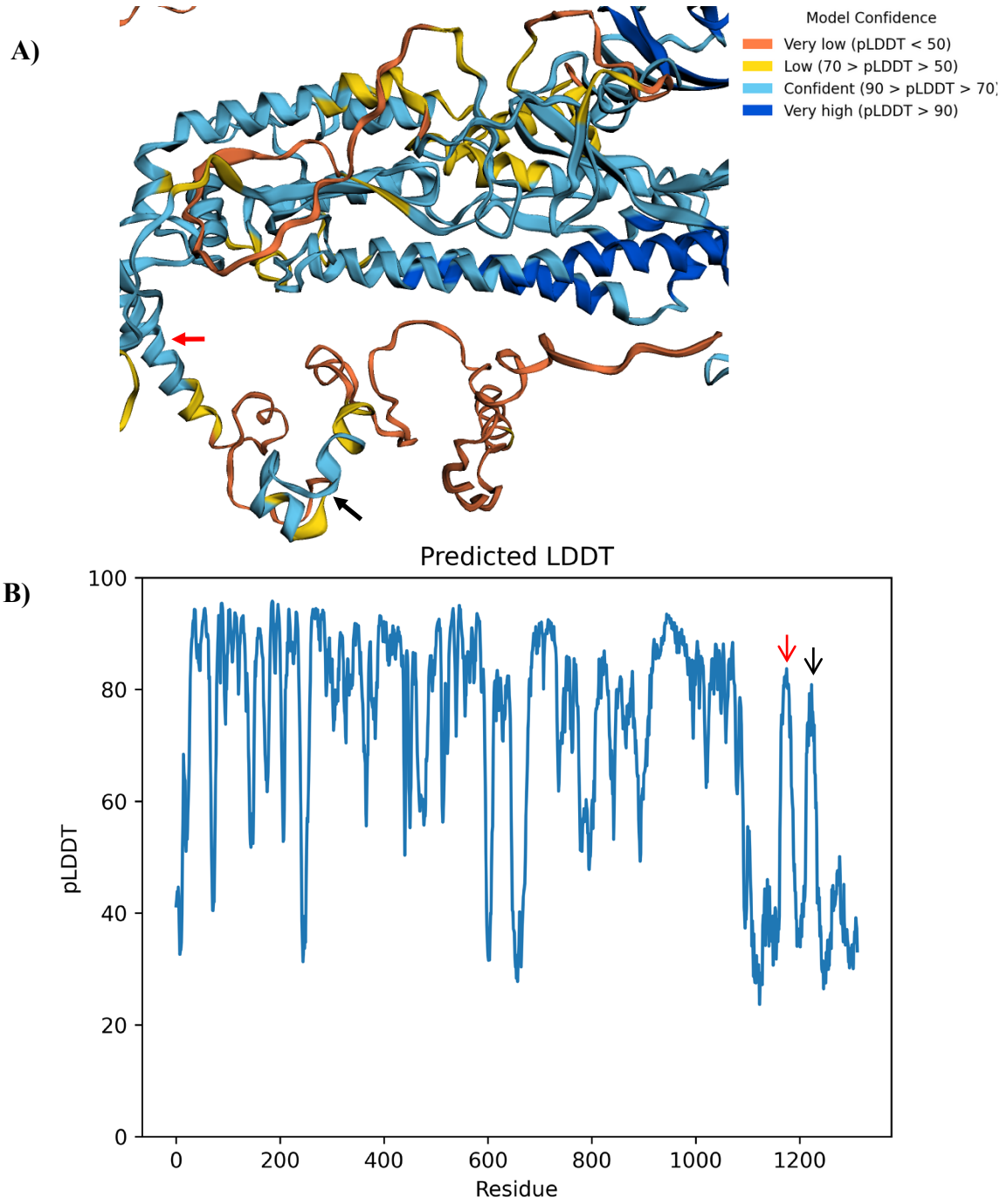


Fig 10: AlphaFold model of the vaccine construct with transmembrane domain (+TM) and predicted LDDT plot. A) The best model is shown. The folding initiator epitope is marked by black arrow. This epitope is well folded. Other epitopes are unstructured in cytoplasmic region. The transmembrane domain is marked by red arrow. **B)** The pLDDT plot shows that the red arrow marked transmembrane region (with pLDDT>80) and most of the other domains are well folded except cytoplasmic region. The model confidence for folding driver peptide (denoted by black arrow) is higher (≈80) than ΔTM indicating no effect from neighbouring unstructured epitopes.

Author Contributions:

AG and MMG designed the study. AG performed the AlphaFold data analysis and other Bioinformatics studies. MKS processed sequences by scripting to get the AlphaFold data from UPPMAX, Uppsala University, Sweden. AG and MMG wrote the paper.

Author Information:

AG: Protein Crystallographer, Bioinformatician and Entrepreneur, India
MKS: Protein and organic Crystallographer, currently post-doctoral researcher in Uppsala University, Sweden
MMG: Scientist G (Retired), ICMR- National Institute of Virology, India

Funding Information: The work is not funded.

Acknowledgements:

The AlphaFold modelling was enabled by resources in project [SNIC 2022/22-709] provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. Special thanks to Rukmini Raju, Centre for Cellular and Molecular Biology, Hyderabad, India for help in editing the manuscript.

References

1. Cascella M, Rajnik M, Aleem A, et al. Features, Evaluation, and Treatment of Coronavirus (COVID-19) [Updated 2022 May 4]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK554776/

2. Wirsching, S., Harder, L., Heymanns, M., Gröndahl, B., Hilbert, K., Kowalzik, F., Meyer, C., & Gehring, S. (2022). Long-Term, CD4+ Memory T Cell Response to SARS-CoV-2. Frontiers in Immunology, 13(April), 1–9. https://doi.org/10.3389/fimmu.2022.800070

3. Mannar, D., Saville, J. W., Zhu, X., Srivastava, S. S., Berezuk, A. M., Tuttle, K. S., Marquez, A. C., Sekirov, I., & Subramaniam, S. (2022). SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex. Science, 375(6582), 760–764. https://doi.org/10.1126/science.abn7760

4. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/myocarditis.html

5. Chen, F., Cao, P., Liu, H., & Cai, D. (2022). The impact of COVID-19 and vaccine on human nervous system. Neuroendocrinology. https://doi.org/10.1159/000524234

6. Escalera, A., Gonzalez-Reiche, A. S., Aslam, S., Mena, I., Laporte, M., Pearl, R. L., Fossati, A., Rathnasinghe, R., Alshammary, H., van de Guchte, A., Farrugia, K., Qin, Y., Bouhaddou, M., Kehrer, T., Zuliani-Alvarez, L., Meekins, D. A., Balaraman, V., McDowell, C., Richt, J. A., … Aydillo, T. (2022). Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission. Cell Host and Microbe, 30(3), 373-387.e7. https://doi.org/10.1016/j.chom.2022.01.006

7. van Zelm, M. C. (2022). Immune memory to SARS-CoV-2 Omicron BA.1 breakthrough infections: To change the vaccine or not? Science Immunology, eabq5901. https://doi.org/10.1126/sciimmunol.abq5901

8. Quiros-fernandez, I., Poorebrahim, M., Fakhr, E., & Cid-arregui, A. (2021). Immunogenic T cell epitopes of SARS-CoV-2 are recognized by circulating memory and naïve CD8 T cells of unexposed individuals. EBioMedicine, 72(103610), 1–14. https://doi.org/10.1016/j.ebiom.2021.103610

9. Grifoni, A., Sidney, J., Vita, R., Peters, B., Crotty, S., Weiskopf, D., & Sette, A. (2021). SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19. Cell Host & Microbe, 29(July 14, 2021), 1076–1092. https://doi.org/https://doi.org/10.1016/j.chom.2021.05.010

10. Talbot, H. K., Shepherd, B. E., Jr, J. E. C., Griffin, Marie, R., Edwards, K. M., Podsiad, A. B., Tollefson, S. J., Wright, P. F., & Williams, J. V. (2009). The Pediatric Burden of Human Coronaviruses Evaluated For 20 Years. 28(8), 682–687. https://doi.org/10.1097/INF.0b013e31819d0d27.

11. Gaunt, E. R., Hardie, A., Claas, E. C. J., Simmonds, P., & Templeton, K. E. (2010). Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected over 3 Years Using a Novel Multiplex Real-Time PCR Method. JOURNAL OF CLINICAL MICROBIOLOGY, 48(8), 2940–2947. https://doi.org/10.1128/JCM.00636-10

12. Dijkman, R., Jebbink, M. F., Gaunt, E., Rossen, J. W. A., Templeton, K. E., Kuijpers, T. W., & Hoek, L. Van Der. (2012). The dominance of human coronavirus OC43 and NL63 infections in infants. Journal of Clinical Virology, 53, 135– 139. https://doi.org/10.1016/j.jcv.2011.11.011

13. Shah, M., Haynes, A. K., Dahl, R. M., Kniss, K., Silk, B., & Killerby, M. E. (2021). Seasonality of Common Human Coronaviruses in the United States, 2014-2021. Open Forum Infectious Diseases, 8(Supplement_1), S753–S754. https://doi.org/10.1093/ofid/ofab466.1523

14. Verhagen, J., Meijden, E. D. Van Der, Lang, V., Kremer, A. E., Völkl, S., Mackensen, A., … Kremer, A. N. (2021). Human CD4 + T cells specific for dominant epitopes of 2 Spike and Nucleocapsid proteins with therapeutic potential. Clin Exp Immunol., 205:363–37(May), 363–378. https://doi.org/10.1111/cei.13627

15. Petra Bacher, Elisa Rosati, Daniela Esser, Gabriela Rios Martini, Carina Saggau, Esther Schiminsky, Justina Dargvainiene, Ina Schröder, Imke Wieters, Yascha Khodamoradi, Fabian Eberhardt, Maria J.G.T. Vehreschild, Holger Neb, Michael Sonntagbauer, Claud, A. F. and A. S. (2020). Low-Avidity CD4 + T Cell Responses to SARS-CoV-2 in Unexposed Individuals and Humans with Severe COVID-19. Immunity, 53(December), 1258–1271. https://doi.org/https://doi.org/10.1016/j.immuni.2020.11.016

16. Cun, Y., Li, C., Shi, L., Sun, M., Dai, S., Sun, L., & Shi, L. (2021). COVID-19 coronavirus vaccine T cell epitope prediction analysis based on distributions of HLA class I loci (HLA-A , -B , -C) across global populations. Human Vaccines & Immunotherapeutics, 17(4), 1097–1108. https://doi.org/10.1080/21645515.2020.1823777

17. Tarke, A., Sidney, J., Methot, N., Crotty, S., Grifoni, A., Tarke, A., Sidney, J., Methot, N., Yu, E. D., Zhang, Y., Dan, J. M., Goodwin, B., Antunes, S., Peters, B., Scheuermann, R. H., Weiskopf, D., & Crotty, S. (2021). Impact of SARS-CoV-2 variants on the total CD4 + and CD8 + T cell reactivity in

infected or vaccinated individuals. Cell Reports Medicine, 2(July), 1–17. https://doi.org/10.1016/j.xcrm.2021.100355

18. Thura, M., Xuan, J., Sng, E., Ang, K. H., Li, J., Gupta, A., Hong, J. M., Hong, C. W., & Zeng, Q. (2021). Targeting intra-viral conserved nucleocapsid (N) proteins as novel vaccines against SARS-CoVs. Bioscience Reports, 41(September), 1–15. https://doi.org/https://doi.org/10.1042/BSR20211491

19. Hajnik, R. L., Plante, J. A., Liang, Y., Alameh, M., Tang, J., Bonam, S. R., Zhong, C., Adam, A., Scharton, D., Rafael, G. H., Liu, Y., Hazell, N. C., Sun, J., Soong, L., Shi, P., Wang, T., Walker, D. H., Sun, J., Weissman, D., … Hu, H. (2022). Dual spike and nucleocapsid mRNA vaccination confer protection against SARS-CoV-2 Omicron and Delta variants in preclinical models. Science Translational Medicine, 14(662), 1–16. https://doi.org/10.1126/scitranslmed.abq1945

20. Bilich, T., Nelde, A., Heitmann, J. S., Maringer, Y., Roerden, M., Bauer, J., Rieth, J., Wacker, M., Peter, A., Horber, S., Rachfalski, D., Marklin, M., Stevanovic, S., Rammensee, H. G., Salih, H. R., & Walz, J. S. (2021). T cell and antibody kinetics delineate SARS-CoV-2 peptides mediating long-Term immune responses in COVID-19 convalescent individuals. Science Translational Medicine, 13(590), 1–15. https://doi.org/10.1126/scitranslmed.abf7517

21. Vo, H. T. M., Maestri, A., Auerswald, H., Sorn, S., Lay, S., Seng, H., Sann, S., Ya, N., Pean, P., Dussart, P., Schwartz, O., Ly, S., Bruel, T., Ly, S., Duong, V., Karlsson, E. A., & Cantaert, T. (2022). Robust and Functional Immune Memory Up to 9 Months After SARS-CoV-2 Infection: A Southeast Asian Longitudinal Cohort. Frontiers in Immunology, 13(February), 1–16. https://doi.org/10.3389/fimmu.2022.817905

22. Pickett, B. E., Greer, D. S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., Gu, Z., Kumar, S., Zaremba, S., Larsen, C. N., Jen, W., Klem, E. B., & Scheuermann, R. H. (2012). Virus Pathogen Database and Analysis Resource (ViPR): A Comprehensive Bioinformatics Database and Analysis Resource for the Coronavirus Research Community. In Viruses (Vol. 4, Issue 11). https://doi.org/10.3390/v4113209

23. Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance, 22(13). https://doi.org/https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

24. Lacek, K., Rambo-Martin, B., Batra, D., Zheng, X., Hassell, N., Sakaguchi, H., Peacock, T., Groves, N., Keller, M., Wilson, M., Sheth, M., Davis, M., Borroughs, M., Gerhart, J., Shepard, S., Cook, P., Lee, J., Wentworth, D., Barnes, J., … Paden, C. (2022). SARS-CoV-2 Delta–Omicron

Recombinant Viruses, United States. Emerging Infectious Disease Journal, 28(7), 1442. https://doi.org/10.3201/eid2807.220526

25. Zhang, Y., Chen, Y., Li, Y., Huang, F., Luo, B., Yuan, Y., Xia, B., Ma, X., Yang, T., Yu, F., Liu, J., Liu, B., Song, Z., Chen, J., Yan, S., Wu, L., Pan, T., Zhang, X., Li, R., … Zhang, H. (2021). The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. Proceedings of the National Academy of Sciences of the United States of America, 118(23), 1–12. https://doi.org/10.1073/pnas.2024202118

26. Yoo, J. S., Sasaki, M., Cho, S. X., Kasuga, Y., Zhu, B., Ouda, R., Orba, Y., de Figueiredo, P., Sawa, H., & Kobayashi, K. S. (2021). SARS-CoV-2 inhibits induction of the MHC class I pathway by targeting the STAT1-IRF1-NLRC5 axis. Nature Communications, 12(1). https://doi.org/10.1038/s41467-021-26910-8

27. Banerjee, A. K., Blanco, M. R., Bruce, E. A., Honson, D. D., Chen, L. M., Chow, A., Bhat, P., Ollikainen, N., Quinodoz, S. A., Loney, C., Thai, J., Miller, Z. D., Lin, A. E., Schmidt, M. M., Stewart, D. G., Goldfarb, D., De Lorenzo, G., Rihn, S. J., Voorhees, R. M., … Guttman, M. (2020). SARS-CoV-2 Disrupts Splicing, Translation, and Protein Trafficking to Suppress Host Defenses. Cell, 183(5), 1325-1339.e21. https://doi.org/10.1016/j.cell.2020.10.004

28. Schmidt, K. G., Nganou-makamdop, K., Tenbusch, M., Kenz, B. El, Maier, C., Lapuente, D., Überla, K., Spriewald, B., Bergmann, S., Harrer, E. G., & Harrer, T. (2021). SARS-CoV-2-Seronegative Subjects Target CTL Epitopes in the SARS- CoV-2 Nucleoprotein Cross-Reactive to Common Cold Coronaviruses. Frontiers in Immunology, 12(April), 1–13. https://doi.org/10.3389/fimmu.2021.627568

29. Katie E. Lineburg, Emma J. Grant, Srividhya Swaminathan, ..., Kirsty R. Short, Corey Smith, S. G. (n.d.). CD8+ T cells specific for an immunodominant SARS- CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses. Immunity, 54(January), 1055–1065. https://doi.org/https://doi.org/10.1016/j.immuni.2021.04.006

30. Kundu, R., Narean, J. S., Wang, L., Fenn, J., Pillay, T., Fernandez, N. D., Conibear, E., Koycheva, A., Davies, M., Tolosa-wright, M., Hakki, S., Varro, R., Mcdermott, E., Hammett, S., Cutajar, J., Thwaites, R. S., Parker, E., Rosadas, C., Mcclure, M., … Lalvani, A. (n.d.). Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. Nature Communications, 13(80), 1–8. https://doi.org/10.1038/s41467-021-27674-x

31. Oliveira, S. C., de Magalhães, M. T. Q., & Homan, E. J. (2020). Immunoinformatic Analysis of SARS-CoV-2 Nucleocapsid Protein and Identification of COVID-19 Vaccine Targets. Frontiers in Immunology, 11. https://doi.org/10.3389/fimmu.2020.587615

32. Nelde, A., Bilich, T., Heitmann, J. S., Maringer, Y., Salih, H. R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., Peter, A., Hörber, S., Traenkle, B., Kaiser, P. D., Rothbauer, U., Becker, M., Junker, D., Krause, G., Strengert, M., Klein, R. (2021). SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. Nature Immunology, 22(January), 74–85. https://doi.org/10.1038/s41590-020-00808-x

33. Agerer, B., Koblischke, M., Gudipati, V., Montaño-Gutierrez, L. F., Smyth, M., Popa, A., Genger, J. W., Endler, L., Florian, D. M., Mühlgrabner, V., Graninger, M., Aberle, S. W., Husa, A. M., Shaw, L. E., Lercher, A., Gattinger, P., Torralba-Gombau, R., Trapin, D., Penz, T., … Bergthaler, A. (2021). SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8+ T cell responses. Science Immunology, 6(57), 17–22. https://doi.org/10.1126/sciimmunol.abg6461

34. Charles, T., Moss, D. L., Bhat, P., Moore, P. W., Kummer, N. A., Bhattacharya, A., Landry, S. J., & Mettu, R. R. (2022). CD4+ T-Cell Epitope Prediction by Combined Analysis of Antigen Conformational Flexibility and Peptide-MHCII Binding Affinity. Biochemistry, 61(15), 1585–1599. https://doi.org/10.1021/acs.biochem.2c00237

35. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7(1), 539. https://doi.org/https://doi.org/10.1038/msb.2011.75

36. Consortium, T. U. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

37. Fleri, W., Paul, S., Dhanda, S. K., Mahajan, S., Xu, X., Peters, B., & Sette, A. (2017). The immune epitope database and analysis resource in epitope discovery and synthetic vaccine design. Frontiers in Immunology, 8(MAR), 1–16. https://doi.org/10.3389/fimmu.2017.00278

38. Zhao, J., Wang, L., Schank, M., Dang, X., Lu, Z., & Cao, D. (2021). SARS-CoV-2 specific memory T cell epitopes identified in COVID-19-recovered subjects. Virus Research, 304(198508), 1–14. https://doi.org/https://doi.org/10.1016/j.virusres.2021.198508

39. Lee, C. H., Pinho, M. P., Buckley, P. R., & Woodhouse, I. B. (2020). Potential CD8 + T Cell Cross-Reactivity Against SARS-CoV-2 Conferred by Other Coronavirus Strains. Frontiers in Immunology, 11(November), 1–10. https://doi.org/10.3389/fimmu.2020.579480

40. Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., & Thornton, J. M. (1997). PDBsum: a web-based database of summaries and analyses of all PDB structures. Trends in Biochemical Sciences, 22(12), 488–490. https://doi.org/10.1016/S0968-0004(97)01140-7

41. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols, 10(6), 845–858. https://doi.org/10.1038/nprot.2015-053

42. Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Research, 49(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

43. Walls, A. C., Tortorici, M. A., Snijder, J., Xiong, X., Bosch, B. J., Rey, F. A., & Veesler, D. (2017). Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proceedings of the National Academy of Sciences of the United States of America, 114(42), 11157–11162. https://doi.org/10.1073/pnas.1708727114

44. Yip, M. S., Leung, N. H. L., Cheung, C. Y., Li, P. H., Lee, H. H. Y., Daëron, M., Peiris, J. S. M., Bruzzone, R., & Jaume, M. (2014). Antibody-dependent infection of human macrophages by severe acute respiratory syndrome coronavirus. Virology Journal, 11, 82. https://doi.org/10.1186/1743-422X-11-82

45. Wang, Q., Zhang, L., Kuwahara, K., Li, L., Liu, Z., Li, T., Zhu, H., Liu, J., Xu, Y., Xie, J., Morioka, H., Sakaguchi, N., Qin, C., & Liu, G. (2016). Immunodominant SARS Coronavirus Epitopes in Humans Elicited both Enhancing and Neutralizing Effects on Infection in Non-human Primates. ACS Infectious Diseases, 2(5), 361–376. https://doi.org/10.1021/acsinfecdis.6b00006

46. Ajmeriya, S., Kumar, A., Karmakar, S., Rana, S., & Singh, H. (2022). Neutralizing Antibodies and Antibody-Dependent Enhancement in COVID-19: A Perspective. Journal of the Indian Institute of Science, 1–17. https://doi.org/10.1007/s41745-021-00268-8

47. Xiaoying Chen, Jennica Zaro, W.-C. S. (2014). Fusion Protein Linkers: Property, Design and Functionality. Adv Drug Deliv Rev, 65(10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039.

48. Tahir, M., Shahid, F., Aslam, S., Ashfaq, U. A., Aslam, S., Fatima, I., … Chen, L. (2020). Reverse vaccinology assisted designing of multiepitope-based subunit vaccine against. Infectious Diseases of Poverty, 9(132), 1–14. https://doi.org/https://doi.org/10.1186/s40249-020-00752-w

49. Safavi, A., Kefayat, A., Mahdevar, E., Abiri, A., & Ghahremani, F. (2020). Exploring the out of sight antigens of SARS-CoV-2 to design a candidate multi-epitope vaccine by utilizing immunoinformatics approaches. Vaccine (2020), 38(October), 7612–7628. https://doi.org/https://doi.org/10.1016/j.vaccine.2020.10.016

50. Minkyung, B., Frank, D., Ivan, A., Justas, D., Sergey, O., Rie, L. G., Jue, W., Qian, C., N., K. L., Dustin, S. R., Claudia, M., Hahnbeom, P., Carson, A., R., G. C., Andy, D., H., P. J., V., R. A., A., van D. A., C., E. A., … David, B. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871–876. https://doi.org/10.1126/science.abj8754

51. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

52. Guo, H. B., Perminov, A., Bekele, S., Kedziora, G., Farajollahi, S., Varaljay, V., Hinkle, K., Molinero, V., Meister, K., Hung, C., Dennis, P., Loughnane, N. K., & Berry, R. (2022). AlphaFold2 models indicate that protein sequence determines both structure and dynamics. Scientific Reports, 12(10696), 1–15. https://doi.org/10.1038/s41598-022-14382-9

53. Dutta, N. K., Mazumdar, K., & Gordy, J. T. (2020). The Nucleocapsid Protein of SARS-CoV-2: a Target for Vaccine Development. In Journal of virology (Vol. 94, Issue 13). https://doi.org/10.1128/JVI.00647-20

54. Smits VAJ, Hernández-Carralero E, Paz-Cabrera MC, Cabrera E, Hernández-Reyes Y, Hernández-Fernaud JR, Gillespie DA, Salido E, Hernández-Porto M, Freire R. The Nucleocapsid protein triggers the main humoral immune response in COVID-19 patients. Biochem Biophys Res Commun. 2021 Mar 5; 543:45-49. https://doi.org/10.1016/j.bbrc.2021.01.073.

55. Bai, Z., Cao, Y., Liu, W., & Li, J. (2021). The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. Viruses, 13(6). https://doi.org/10.3390/v13061115

56. Chai, J., Cai, Y., Pang, C., Wang, L., McSweeney, S., Shanklin, J., & Liu, Q. (2021). Structural basis for SARS-CoV-2 envelope protein recognition of human cell junction protein PALS1. Nature Communications, 12(1), 3433. https://doi.org/10.1038/s41467-021-23533-x

57. Casado, L., Johannes, H., Vizcarra, P., Velasco, H., Velasco, T., Fernandez-escribano, M., & Vallejo, A. (2021). SARS CoV-2 infections in healthcare workers with a pre-existing T-cell response: a prospective cohort study. Clinical Microbiology and Infection Journal, 27(March), 1–5. https://doi.org/https://doi.org/10.1016/j.cmi.2021.02.020

58. Cattin-Ortolá, J., Welch, L. G., Maslen, S. L., Papa, G., James, L. C., & Munro, S. (2021). Sequences in the cytoplasmic tail of SARS-CoV-2 Spike facilitate expression at the cell surface and syncytia formation. Nature Communications, 12(1), 5333. https://doi.org/10.1038/s41467-021-25589-1

59. Dai, L., & Gao, G. F. (2021). Viral targets for vaccines against COVID-19. Nature Reviews Immunology, 21(2), 73–82. https://doi.org/10.1038/s41577-020-00480-0

60. Ebenig, A., Muraleedharan, S., Kazmierski, J., Goffinet, C., Brown, R. J. P., Ebenig, A., Muraleedharan, S., Kazmierski, J., Todt, D., Auste, A., & Anzaghe, M. (2022). Vaccine-associated enhanced respiratory pathology in COVID-19 hamsters after T H 2-biased immunization. Cell Reports, 40(111214), 1–23. https://doi.org/10.1016/j.celrep.2022.111214

61. Thornhill, J. P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L. B., Palich, R., Nori, A., Reeves, I., Habibi, M. S., Apea, V., Boesecke, C., Vandekerckhove, L., Yakubovsky, M., Sendagorta, E., Blanco, J. L., Florence, E., Moschese, D., Maltez, F. M., … Orkin, C. M. (2022). Monkeypox Virus Infection in Humans across 16 Countries — April–June 2022. New England Journal of Medicine, 387(8), 679–691. https://doi.org/10.1056/nejmoa2207323

62. Seifert, S. N., Bai, S., Fawcett, S., Norton, E. B., Zwezdaryk, K. J., Robinson, J., Gunn, B., & Letko, M. C. (2022). An ACE2-dependent Sarbecovirus in Russian bats is resistant to SARS-CoV-2 vaccines. BioRxiv, 2021.12.05.471310. https://doi.org/10.1371/journal.ppat.1010828

63. Bhatt, L., Bhoyar, R. C., Jolly, B., Israni, R., Vignesh, H., & Scaria, V. (2022). The genome sequence of the Lumpy Skin Disease virus from the outbreak in India suggests a distinct lineage of the virus. 0–7, bioRxiv preprint doi: https://doi.org/10.1101/2022.09.15.508131

64. Din, M., Asghar, M., & Ali, M. (2021). COVID-19 and dengue coepidemics: A double trouble for overburdened health systems in developing countries. Journal of Medical Virology, 93(2), 601–602. https://doi.org/10.1002/jmv.26348