Enhanced binding of SARS-CoV-2 Envelope protein to tight junction-associated PALS1 could play a key role in COVID-19 pathogenesis

Flavio De Maio
Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario “A. Gemelli”, IRCCS, Roma, Italia - Dipartimento di Scienze biotecnologiche di base, cliniche intensivologiche e perioperatorie – Sezione di Microbiologia, Università Cattolica del Sacro Cuore, Roma, Italia

Ettore Lo Cascio
Dipartimento di Scienze biotecnologiche di base, cliniche intensivologiche e perioperatorie – Sezione di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Roma, Italia

Gabriele Babini (gabriele.babini@guest.policlinicogemelli.it)
Dipartimento Scienze della Salute della Donna, del Bambino e di Sanità Pubblica, Fondazione Policlinico Universitario "A. Gemelli", IRCCS, Roma, Italia

Michela Sali
Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario “A. Gemelli”, IRCCS, Roma, Italia - Dipartimento di Scienze biotecnologiche di base, cliniche intensivologiche e perioperatorie – Sezione di Microbiologia, Università Cattolica del Sacro Cuore, Roma, Italia

Stefano Della Longa
Department of Life, Health and Environmental Sciences, University of L'Aquila, 67100 L'Aquila, Italy

Bruno Tilocca
Department of Health Science, University “Magna Græcia” of Catanzaro, Viale Europa, 88100 Catanzaro, Italy

Paola Roncada
Department of Health Science, University “Magna Græcia” of Catanzaro, Viale Europa, 88100 Catanzaro, Italy

Alessandro Arcovito
Dipartimento di Scienze biotecnologiche di base, cliniche intensivologiche e perioperatorie – Sezione di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Roma, Italia - Fondazione Policlinico Universitario “A. Gemelli”, IRCCS, Roma, Italia

Maurizio Sanguinetti
Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario “A. Gemelli”, IRCCS, Roma, Italia - Dipartimento di Scienze biotecnologiche di base, cliniche intensivologiche e perioperatorie – Sezione di Microbiologia, Università Cattolica del Sacro Cuore, Roma, Italia

Giovanni Scambia

Dipartimento Scienze della Salute della Donna, del Bambino e di Sanità Pubblica, Fondazione Policlinico Universitario "A. Gemelli", IRCCS, Roma, Italia - Dipartimento di Scienze della Vita e Sanità pubblica - Sezione di Ginecologia ed Ostetricia, Università Cattolica del Sacro Cuore, Roma, Italia

Andrea Urbani

Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario “A. Gemelli”, IRCCS, Roma, Italia - Dipartimento di Scienze biotecnologiche di base, cliniche intensivologiche e perioperatorie – Sezione di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, Roma, Italia

---

Research Article

Keywords: SARS-CoV-2, COVID-19, Envelope protein, Tight junctions, PALS1

Posted Date: May 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-30903/v1

Version of Record: A version of this preprint was published on September 4th, 2020. See the published version at https://doi.org/10.1016/j.micinf.2020.08.006.

Abstract

Background: The Envelope (E) protein of SARS-CoV-2 is the most enigmatic protein among the four structural ones on the viral genome. Most of the current knowledge on the E protein is based on the direct comparison to the SARS E protein, initially mistakenly undervalued and subsequently proved to be a key factor in the ER-Golgi localization and in tight junction disruption.

Methods: We compared the genomic sequences of E protein of SARS-CoV-2, SARS-CoV and the closely related genomes of bats and pangolins obtained from the GISAID and GenBank databases. Multiple sequence alignments were done with the Geneious software using the MAFFT algorithm. *In silico* modelling analyses of E proteins conformation and docking with PALS1 were performed with the Schrodinger Suite.

Results: When compared to the known SARS E protein, we observed a different amino acidic sequence in the C-terminal of SARS-CoV-2 E protein which might have a key role in the current COVID-19 pathogenesis. *In silico* docking results provide evidence of a strengthened binding of SARS-CoV-2 E protein with the tight junction-associated PALS1 protein.

Conclusions: We suggest that SARS-CoV-2 E protein may interfere with the tight junction stability and formation leading to an enhanced epithelial barrier disruption, amplifying the inflammatory processes, and promoting tissue remodelling. These findings raise a warning on the underestimated role of the E protein in the pathogenic mechanism and could open the route to detailed experimental investigations.

Background

Severe acute respiratory syndrome 2, known as Coronavirus disease 2019 (Covid-19) is caused by a Betacoronavirus named SARS-CoV-2 virus. To date, Covid-19 associated pneumonia accounts (as of 5th May 2020, source: www.who.int, "COVID-19 Situation report - 106") more than 3,5 Million confirmed cases and almost 250000 deaths since its first reported discovery in the Hubei province in China in December 2019 [1].

Bat and Malayan pangolins coronavirus genomes show a high identity percentage to SARS-CoV-2, suggesting these animal species as possible reservoir hosts for SARS-CoV-2 related viruses before adaptation to humans [2]. This hypothesis is well supported by recent findings in genome analysis. In this sense, Spike (S) protein investigation has highlighted a common pattern with the bat and pangolin orthologues, as well as a peculiar motif likely acquired during human adaptation of the SARS-CoV-2 [3].

Coronavirus infection starts with inhalation of droplets containing virus that invades the epithelial cells by using angiotensin converting enzyme 2 (ACE2) [4], or other cell components like integrins, as targets of the SARS-CoV2 S protein [5, 6]. Viral replication in Type II alveolar epithelial cells leads to severe modifications of the innate immune response [7]. Lungs are rapidly compromised following direct damage of the pulmonary tissue mainly through dysregulation of the immune mediators that enhance

the influx of monocytes and neutrophils in the infected tissue [4]. Furthermore, pro-inflammatory cytokine storm positively impairs virus replication and increase its diffusion to nearby cells.

SARS-CoV-2 virus shows a single strand, positive-sense RNA genome of slightly less than 30 kb in length where at 3' end are located the genes codifying for the four structural proteins: spike (S), membrane (M), envelope (E), and nucleocapsid (N) [1]. As an enveloped virus, S, M and E proteins encounter cellular membrane at the initiation of infection and during the replication cycle and are involved in the budding of the mature virions [8]. Whereas Spike protein has been immediately investigated [3, 9–12], few information has been collected on the other surface proteins. Indeed, as occurred with the previously studied *Betacoronavirus*, the role of E protein and its involvement with the host adverse effects was wrongly underestimated. In contrast to S, E protein of the SARS-CoV-2 has not been thoroughly studied and most of the information about its structure and function derive from SARS-CoV experimental studies. While E protein is not represented in the mature virions up to levels of S or M proteins [8, 13], it is abundantly expressed inside the infected cell and actively involved in the pathogenetic viral mechanisms [14].

SARS-CoV E protein is known to account 76 amino acids that are organized in three main domains: a short (approx. 8-10 amino acids) luminal orientated N-terminal domain, a long α-helical transmembrane domain composed of ≈ 22 aminoacidic residues and a cytoplasmically oriented C-terminal domain [15, 16].

Homologous assembling of the E protein contributes to create a pentameric channel with its transmembrane domain that directly impairs virus replication [17].

Conversely, monomeric E protein affects the host's intracellular activities through C-terminal domain, which is predicted to have a β-coil-β structure, leading to its localization in the endoplasmic reticulum, Golgi and the ER-Golgi intermediate compartment[16–18]. Additional targeting information at the N-terminal domain of the E protein would ensure the maintaining of the Golgi complex targeting [19]. Interestingly, the DLLV hydrophobic motif located at the end of the C-terminal has been demonstrated to compete with CRB1 for PALS1 binding [20]. PALS1 is a cellular protein involved in maintaining tight junctions between epithelial cells. Therefore, interactions between the SARS E protein and PALS1 induced relocation of PALS1 to the virus assembly site and disrupted tight junctions promoting virus spread. Little information has been collected yet on SARS-CoV-2 E protein and mainly focused on the sequences conserved from SARS, suggesting its potential interaction with bromodomain proteins [21]. Conversely, aim of this work is to identify the potential implications of sequence dissimilarities between the previous SARS-CoV and SARS-CoV-2.

Methods

Data collection and multiple sequence alignment

Reference coronavirus genomes were obtained either from NCBI Gene or from GISAID (https://www.gisaid.org/) databases as suggested from the recent discoveries on the proximal origins of SARS-CoV-2 [2, 3] (Detailed accession numbers in Supplementary Table 1). Pangolin genomes were pooled in two groups, as shown by Lam et al. phylogenetic analyses [2] and only the consensus genome for each one of the two groups was considered. The Envelope proteins (E) from all genomes were extracted and multiple sequence alignment performed with the MAFFT algorithm (v7.450) in Geneious Prime (version 11.0.4) [22, 23]. Pairwise sequence identities were also calculated using Geneious Prime.

3D Homology modeling of SARS-CoV-2 E

The amino acidic sequence of the SARS-CoV-2 Envelope protein was extracted, and the NMR structure of the homologous Protein of SARS-CoV (PDB code: 2MM4) was used as a template. The starting 3D model was then built using the Homology Modeling protocol Prime of the Schrodinger Suite [24]. According to the Ramachandran plot analysis for 58 residues, 93.1% lie in the most favored regions, 6.9% in the allowed regions, and none in the disallowed regions. The missing C and N terminal residues, not present in the chosen template, were finally added.

Molecular dynamic simulations of E protein

In order to get a more realistic model of the SARS-CoV-2 Envelope protein, the starting 3D homology model was further optimized using molecular dynamics simulations prepared via the CHARMM-GUI [25] server and performed using standard GROMACS tools [26]. At first, the protein was shortly simulated (10ns, 300K) in neutralized solution. It allowed to get conformations in which the transmembrane (TM), the C-term and the N-term domains were well suited for only embedding of the TM domain into a phospholipidic double layer membrane. The chosen conformation of the protein obtained this way was then oriented and inserted into a small POPC double layer (60 molecules per layer) surrounded by water (29472 atoms). A 100ns long MD simulation of the obtained system was performed as proof of stability.

In silico docking of SARS-CoV and SARS-CoV-2 Envelope C-terms to PALS1

The PALS1 structure for the *in silico* docking of the Envelope C-terminus octapeptides, was obtained by the reported PALS1-CRB1 complex (PDB code: 4UU5). It was prepared for docking with the Maestro protein preparation wizard, using default parameters [27]. Neither side-chain atoms nor residues were missing in the protein in the neighborhood of CRB1 peptide at 5 Å of the protein. A search grid was generated with Glide5 by selecting the 8 C-terminal residues of CRB1 (PPAMERLI) to define the binding pocket, thus including the entire binding site of the peptide–protein complex.

Then, the 8 C-terminal residues of each Envelope protein were built, i.e. EGVPDLLV and SRVPDLLV for SARS-CoV and SARS-CoV-2, respectively, whose protonation state was assigned with PROPKA. Using the peptide-protein docking protocol of Glide [28] multiple conformers of the peptide were generated, docked on the protein and post-processed using MM-GBSA.

Results

Identification of genomic differences between SARS-CoV-2 and SARS-CoV E proteins

Multiple sequence alignments showed a quasi-perfect identity between all genomes of bats, pangolin, SARS-CoV and SARS-CoV-2 in the N-terminal and transmembrane regions of the protein E: only few synonymous mutations were identified in these two regions (See Supplementary Table 2). A different outcome is highlighted in the C-terminal region of the E protein sequence (Figure 1A) where two different mutation sites show a clear difference between SARS-CoV-2 and SARS-CoV, confirming what has been found from the whole genome phylogenetic analyses [2].

Specifically, the mutation in the inner part of the C-terminal region consists in 4 nucleotide changes that lead to substitution of Thr^{55}-Val^{56} for SARS-CoV with Ser^{55}-Phe^{56} in SARS-CoV-2. The second mutation regards the deletion of Glu^{69}-Gly^{70} and substitution with an Arg^{69}.

Homology modelling of E proteins

In Figure 1B and 1C, the two predicted monomeric E full length protein structure models have been constructed and show N-terminal (blue), transmembrane (green), C-terminal domains (red) as well as the aminoacidic variants (yellow). As expected, transmembrane domains of both proteins presented the highest accuracy with a total confidence score of more than 90% on ~ 80% of the full-length proteins. The full-length domains of the SARS-CoV variants, have been further characterised, posing them in a membrane bilayer and determining their stability on a short MD simulation of 100 ns (see Supplementary Information Fig S1). In both models, as expected, the trans-membrane helix is perfectly fitted in the hydrophobic core of the membrane bilayer.

The end of the C-terminal, accounting 11 aminoacidic residues, and the begin of the N-terminal did not reach previously indicated accuracy. Moreover, deletion of two aminoacidic residues and arginine substitution at C-terminus could affect the protein structure altering the spatial disposition of the β-coil-β (Figure 1D and E). The structure of this subunit appeared highly mobile but remained substantially unaltered along the short MD simulations for both E variants. Intriguingly, arginine acquisition could alter PDZ peptide-binding to the PALS1 groove, prompting an interaction with the Asn^{315} aminoacidic residue of the PDZ, as predicted for the Arg^{1404} of the natural ligand CRB1 [29].

E protein C-terminals binding to tight junction-associated PALS1

In order to verify the potential implications of the altered amino acidic sequence, the binding pose of the two C-terminus octapeptides belonging to SARS-CoV and SARS-CoV-2 were determined and compared with the crystallographic structure of the complex PALS1-CRB1. The poses with the lowest ΔG, calculated via MM-GBSA by using default parameters, are shown in Figure 2A. Accordingly, the Free Energy of Binding for SARS-CoV and SARS-CoV-2 Envelope C-terminals amounts to -63.62 and -97.10 kcal/mol, respectively. This value must be compared to the value of -92.5 kcal/mol obtained performing the same

analysis on the complex PALS1-CRB1, where the endogenous peptide was shortened to 8 amminoacids in order to compare its *in silico* affinity with the SARS-CoVs variants. Interestingly, the SARS-CoV-2 peptide is able to bind PALS1 with a significantly higher affinity compared SARS-CoV variant, reaching and slightly ameliorating the affinity value of the endogenous ligand, even though the two octapeptides differ for only two out of 8 of the selected amminoacids. In particular, the last four residues of both E C-terminals are the same (Asp, Leu, Leu, Val) and bind PALS1 similarly to what observed for the endogenous CRB1, even if the short sequence of the CRB1 peptide is slightly different (Glu, Arg, Leu, Ile). As shown in the interaction maps described in Figure 2C and 2D, the side-chain of the last residue of the E proteins, which is a valine, interacts with Leu^{267}, Leu^{321} and Phe^{330} of PALS1; its free terminal carboxyl group, instead, makes a salt bridge with Lys^{261} and H-bond interactions with amide hydrogens of Leu^{267}, Gly^{268} and Ala^{269}. The two following leucine residues of the E C-terminals make Van der Waals contacts, in particular the second one makes interaction with Phe^{318} (while CRB1 interacts with this residue via cation-π through its Arg). Aspartate is the last common residue inside the binding pocket and its sidechain residue makes a salt bridge with Arg^{282}. Immediately after this negatively charged amminoacid the two C-terminal residues of SARS-CoV and SARS-CoV-2 differ significantly. The positively charged Arg of the SARS-CoV-2 octapeptide well suits a negatively charged pocket (zoomed region in Figure 2B), being able to contemporarily create a salt bridge with Asp^{313} and several H bonds with Asp^{313} and with a carbonyl oxygen of Met^{275} backbone. In the same position in SARS-CoV, the small size side chain of the Gly residue cannot be involved in any interaction with this pocket thus reducing the interaction strength. Finally, the Ser alcohol moiety of SARS-CoV-2 makes a hydrogen bond with Asp^{299} backbone oxygen, while the Glu residue in SARS-CoV could interact with Arg^{272} with a salt bridge but this condition is never realized among all the poses.

Discussion

In the present work, we compared the genomic sequences of SARS-CoV-2 Envelope protein with the E protein of SARS, and the corresponding bats and pangolin orthologues, in order to identify the implications of dissimilarities in such an enigmatic protein.

Given the very small sequence length of the E protein, full genome multiple sequence alignment might affect the overall precision on this region, likely the reason for erroneously aligned E proteins in previous works, leading to misinterpret these amino acids deletion and substitution in the C-terminal [30, 31]. SARS-CoV matches to orthologue E proteins of bat-CoV Rs3367, while SARS-CoV-2 E protein is identical to bat RaTG13 and, except for synonymous mutations, to bat CoVs and the recently identified pangolin coronaviruses. Interestingly, SARS-CoV and SARS CoV-2 E proteins confirm the two phylogenetic clades observed with the full genome [2].

The subsequent analysis aimed to identify potentially beneficial or detrimental effects of SARS-CoV-2 E protein variant with respect to the previously studied SARS E protein. Docking results of the tight junction complex PALS1-CRB1, used as a reference from the previous study by Ivanova and colleagues [29], were

compared to the docking values of PALS1 with C-terminals of both E proteins. Our findings support the hypothesis that characteristic virulence of SARS-CoV-2 virus could depend on the strengthened interaction between SARS-CoV-2 E protein and PALS1 prompting a strong alteration of the tight junctions. The enhanced binding to PALS1 could represent only the first step of the immunopathogenic process associated to SARS-CoV-2 infection. Moreover, PALS1 – E binding could alter E-cadherin intracellular traffic with change in cell polarization [20] promoting a severe dysregulation of the Th2 mediated response due to enhanced exposition to environmental allergens [32]. All these pathogenetic mechanisms, previously described in the infection with SARS-CoV, could be significantly increased. Epithelial mesenchymal transition could occur with a significant modification of the epithelial tissue and a meaningfully production of chemokines and cytokines in the infection site [33]. Remarkably, changes in epithelial tissue structures, and consequent loss of functions, are age-dependent and associated with defects in tight junctions and cadherin – catenin complex [34]. Although little information is available on lung epithelium ageing-dependent mechanisms, we hypothesize that E protein could significantly increase this dysfunction, especially in elderly people as evidenced by COVID-19 epidemiological data. On the other hand, we cannot exclude that this genomic variant could promote interaction with other viral and host cell components.

Conclusions

The genomic variant highlighted in this work raise concerns on the underestimated role of the Envelope protein of SARS-CoV-2 in the host adverse responses. Indeed, these results shed the lights on the most enigmatic protein among the four structural proteins of coronaviruses, the Envelope protein, laying the foundations for a fundamental detailed experimental investigation of its interaction with both epithelial tight junctions assembly and ER-Golgi intermediate apparatus.

Abbreviations

COVID-19: Coronavirus disease 2019; CoV: Coronavirus; SARS-CoV: Severe acute respiratory syndrome coronavirus; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; S: Spike Protein; M: Membrane protein; E: Envelope protein; N: Nucleocapsid protein; PALS1: Protein Associated with Lin Seven 1; CRB1: Crumbs homolog-1 protein.

Declarations

Acknowledgements

We would like to thank all scientists who have contributed sequences to the GISAID database (https://www.gisaid.org/).

Authors' contributions

This study was designed by FDM, GB, AA, PR, Maurizio Sanguinetti, AU. Data collection was provided by GB, ELC and BT. Data analysis and figure preparation was conducted by FDM, GB, ELC and SDL. FDM, GB, ELC and AA wrote the paper. All authors contributed to literature search ad revision and participate in the discussion and interpretation of the results. All authors contributed to the final revision of the project. All authors read, critically revised and approved the final manuscript.

Funding

No funding was used to conduct this research.

Availability of data and materials

The datasets analysed during the current study are available in the GISAID (https://www.gisaid.org/)

repositories.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Zhu N, Zhang D, Wang W, et al (2020) A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 382:727–733
2. Lam TT-Y, Shum MH-H, Zhu H-C, et al (2020) Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature. doi: 10.1038/s41586-020-2169-0
3. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. Nat Med 2–4
4. Cong Y, Ren X (2014) Coronavirus entry and release in polarized epithelial cells: A review. Rev Med Virol. doi: 10.1002/rmv.1792
5. Sigrist CJ, Bridge A, Le Mercier P (2020) A potential role for integrins in host cell entry by SARS-CoV-2. Antiviral Res 177:104759
6. Letko M, Marzi A, Munster V (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol 5:562–569

7. Sims AC, Burkett SE, Yount B, Pickles RJ (2008) SARS-CoV replication and pathogenesis in an in vitro model of the human conducting airway epithelium. Virus Res 133:33–44
8. J Alsaadi EA, Jones IM (2019) Membrane binding proteins of coronaviruses. Future Virol 14:275–286
9. Pachetti M, Marini B, Benedetti F, et al (2020) Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. J Transl Med 18:1–9
10. Kandeel M, Ibrahim A, Fayez M, Al-Nazawi M (2020) From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes. J Med Virol. doi: 10.1002/jmv.25754
11. Tilocca B, Soggiu A, Musella V, Britti D, Sanguinetti M, Urbani A, Roncada P (2020) Molecular basis of COVID-19 relationships in different species: a one health perspective. Microbes Infect 10–12
12. Watanabe Y, Allen JD, Wrapp D, Mclellan JS, Crispin M (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. 9983:1–9
13. Schoeman D, Fielding BC (2019) Coronavirus envelope protein: Current knowledge. Virol J 16:1–22
14. Venkatagopalan P, Daskalova SM, Lopez LA, Dolezal KA, Hogue BG (2015) Coronavirus envelope (E) protein remains at the site of assembly. Virology 478:75–85
15. Torres J, Parthasarathy K, Lin X, Saravanan R, Kukol A, Ding XL (2006) Model of a putative pore: The pentameric α-helical bundle of SARS coronavirus E protein in lipid bilayers. Biophys J 91:938–947
16. Nieto-Torres JL, DeDiego ML, Álvarez E, Jiménez-Guardeño JM, Regla-Nava JA, Llorente M, Kremer L, Shuo S, Enjuanes L (2011) Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. Virology 415:69–82
17. Li Y, Surya W, Claudine S, Torres J (2014) Structure of a conserved golgi complex-targeting signal in coronavirus envelope proteins. J Biol Chem 289:12535–12549
18. Cohen JR, Lin LD, Machamer CE (2011) Identification of a Golgi Complex-Targeting Signal in the Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Envelope Protein. J Virol 85:5794–5803
19. Cohen JR, Lin LD, Machamer CE (2011) Identification of a Golgi Complex-Targeting Signal in the Cytoplasmic Tail of the Severe Acute Respiratory Syndrome Coronavirus Envelope Protein. J Virol 85:5794–5803
20. Teoh KT, Siu YL, Chan WL, Schlüter MA, Liu CJ, Peiris JSM, Bruzzone R, Margolis B, Nal B (2010) The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. Mol Biol Cell. doi: 10.1091/mbc.E10-04-0338
21. Gordon DE, Jang GM, Bouhaddou M, et al (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 1–13
22. Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066

23. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780
24. Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, Friesner RA (2004) A Hierarchical Approach to All-Atom Protein Loop Prediction. Proteins Struct Funct Genet. doi: 10.1002/prot.10613
25. Lee J, Cheng X, Swails JM, et al (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 12:405–413
26. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindah E (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25
27. Sastry MG, Adzhigirey M, Day T, Annabhimoju R, Sherman W (2013) Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des. doi: 10.1007/s10822-013-9644-8
28. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem. doi: 10.1021/jm051256o
29. Ivanova ME, Fletcher GC, O'Reilly N, Purkiss AG, Thompson BJ, McDonald NQ (2015) Structures of the human Pals1 PDZ domain with and without ligand suggest gated access of Crb to the PDZ peptide-binding groove. Acta Crystallogr Sect D Biol Crystallogr D71:555–564
30. Chatterjee S (2020) Understanding the Nature of Variations in Structural Sequences coding for Coronavirus Spike, Envelope, Membrane and Nucleocapsid Proteins of SARS-CoV-2. PapersSsrnCom 1–12
31. Wu A, Peng Y, Huang B, et al (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 27:325–328
32. Georas SN, Rezaee F (2014) Epithelial barrier function: at the front line of asthma immunology and allergic airway inflammation. J Allergy Clin Immunol 134:509–520
33. Bartis D, Mise N, Mahida RY, Eickelberg O, Thickett DR (2014) Epithelial-mesenchymal transition in lung development and disease: Does it exist and is it important? Thorax 69:760–765
34. Parrish AR (2017) The impact of aging on epithelial barriers. Tissue Barriers. doi: 10.1080/21688370.2017.1343172

Figures

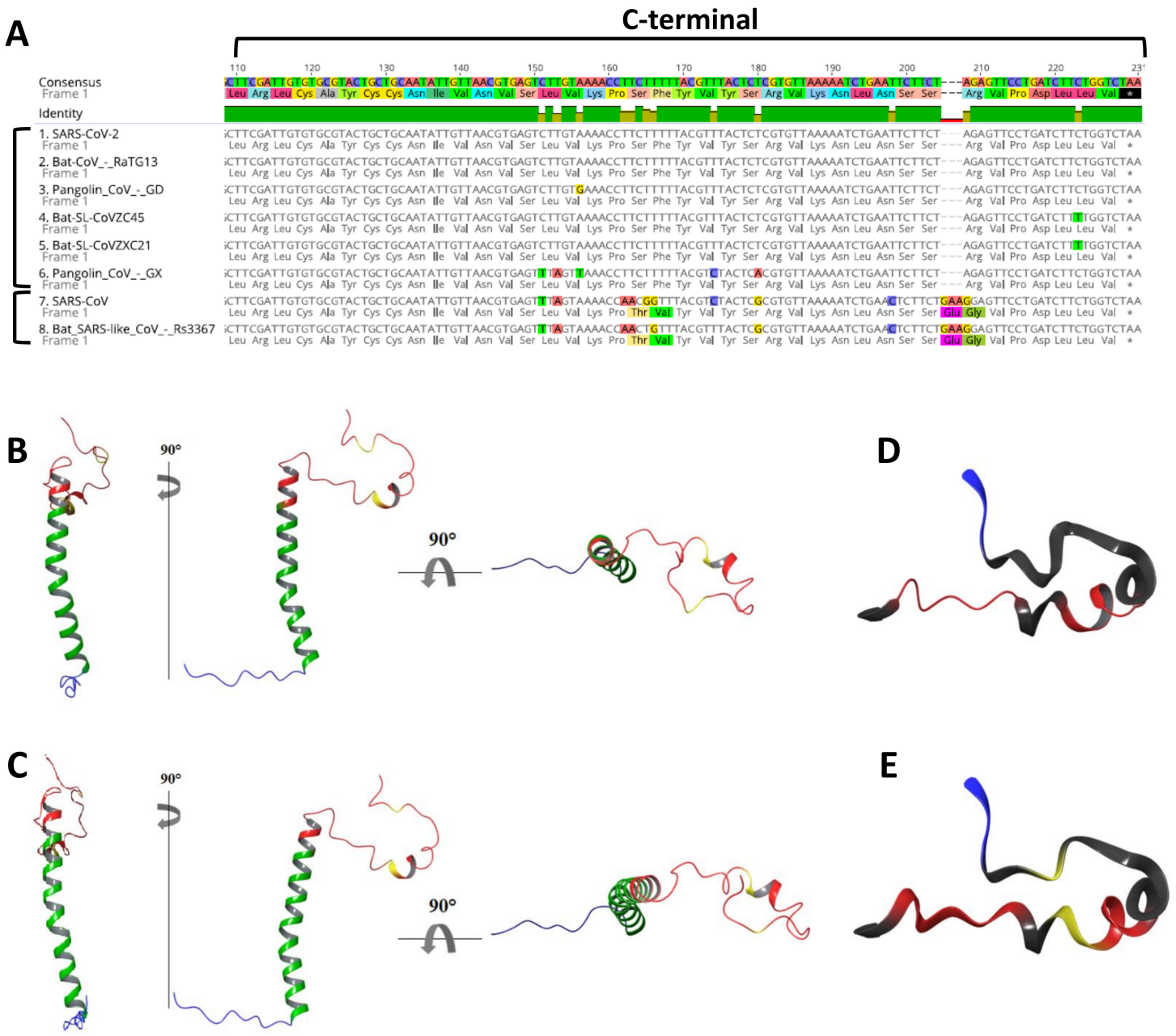

Figure 1

Multiple sequence alignment and homology modelling of E proteins. A) Envelope Protein multiple sequence alignment of C-terminal domain in human, bat and pangolin SARS-like coronaviruses highlights the identity among SARS-CoV-2, 2 bat CoV strains (RaTG13 and CoVZC45/CoVZX21) and pangolins E proteins. The comparison also points out two mutation sites where they differ from SARS-CoV and Bat SARS-like Cov Rs3367. B) E protein structure model for SARS and C) SARS-CoV-2. A closer look at C-terminal domain of E protein in D) SARS and E) SARS-CoV-2 in which are highlighted: in red, the motifs

regulating the transport to Golgi apparatus; in blue, the PDZ-binding domain; in yellow, the two mutation sites identified in SARS-CoV-2 with respect to SARS.

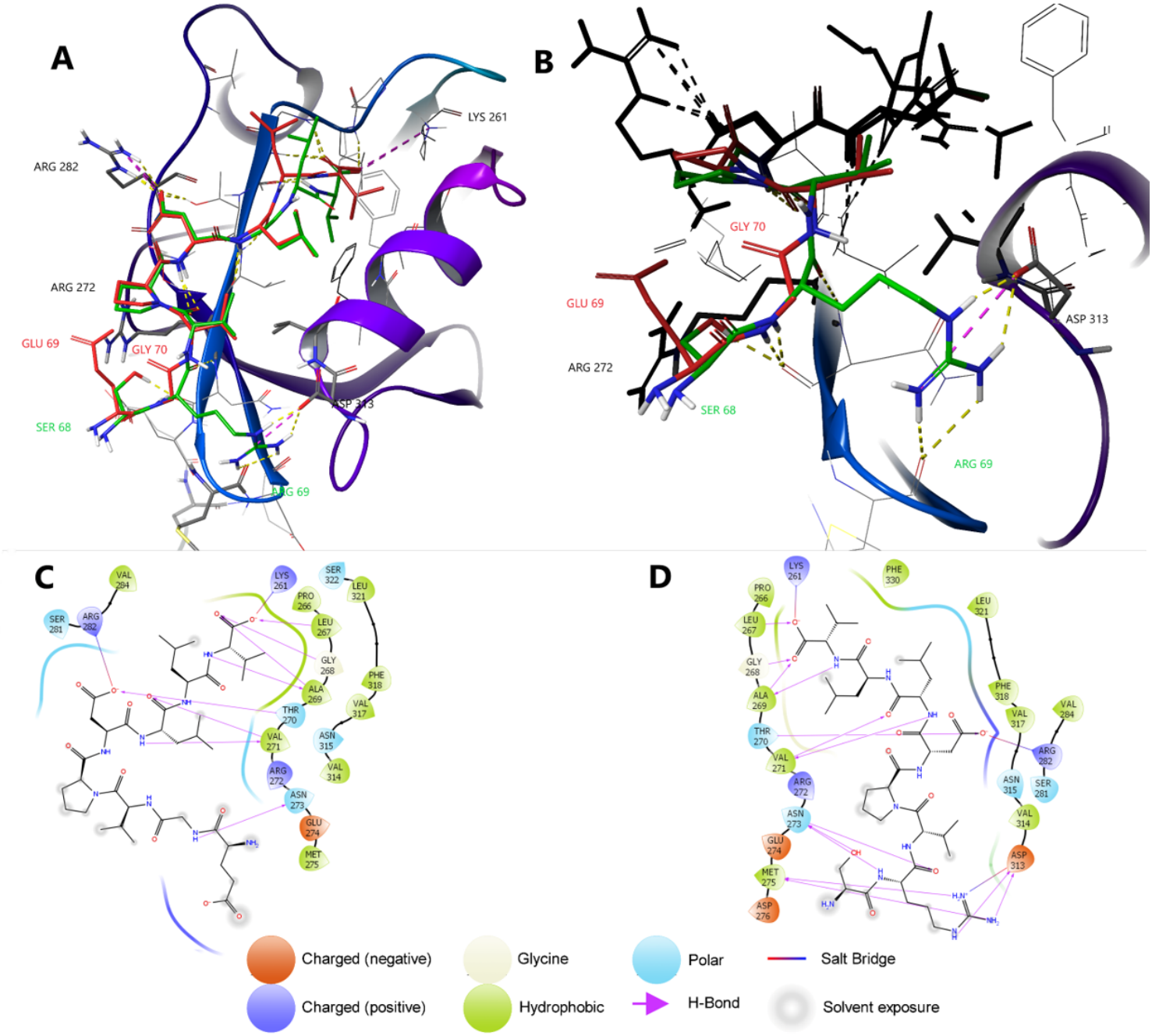

Figure 2

In silico docking of E proteins C-terminals with PALS1. A) SARS-CoV and SARS-CoV-2 octapeptides lowest ΔG poses on PALS1 binding site, representation of H bond in yellow dashed lines, Salt bridge in purple dashed lines. (red structure and label: SARS-CoV; green structure and label: SARS-CoV-2). B) Magnification of SARS-CoV-2 Arg69 inside PALS1 negatively charged pocket showing the interesting salt bridge with ASP313 and H bonds with MET275. C) Ligand Interaction diagram of SARS-CoV octapeptide. D) Ligand Interaction diagram of SARS-CoV-2 octapeptide.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DeMaioetalSupplementaryInformation.docx