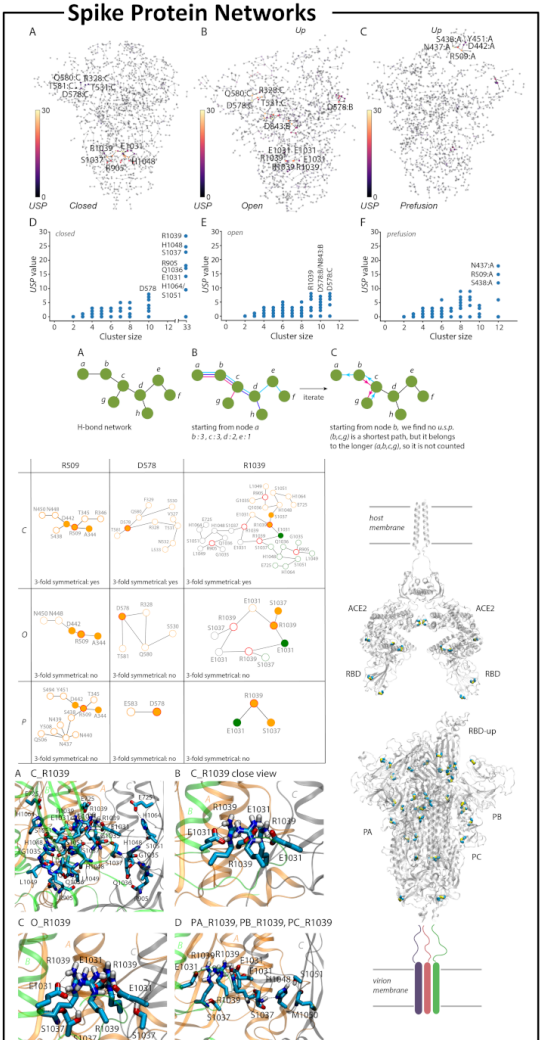
# Identifying long distance communication networks in the SARS-CoV2 S-protein and motifs in a transmembrane protein database.

Michalis Lazaratos, Malte Siemers and Ana-Nicoleta Bondar
Freie Universität Berlin, Department of Physics, Theoretical Molecular Biophysics,
Arnimallee 14, D-14195 Berlin, Germany

## Summary

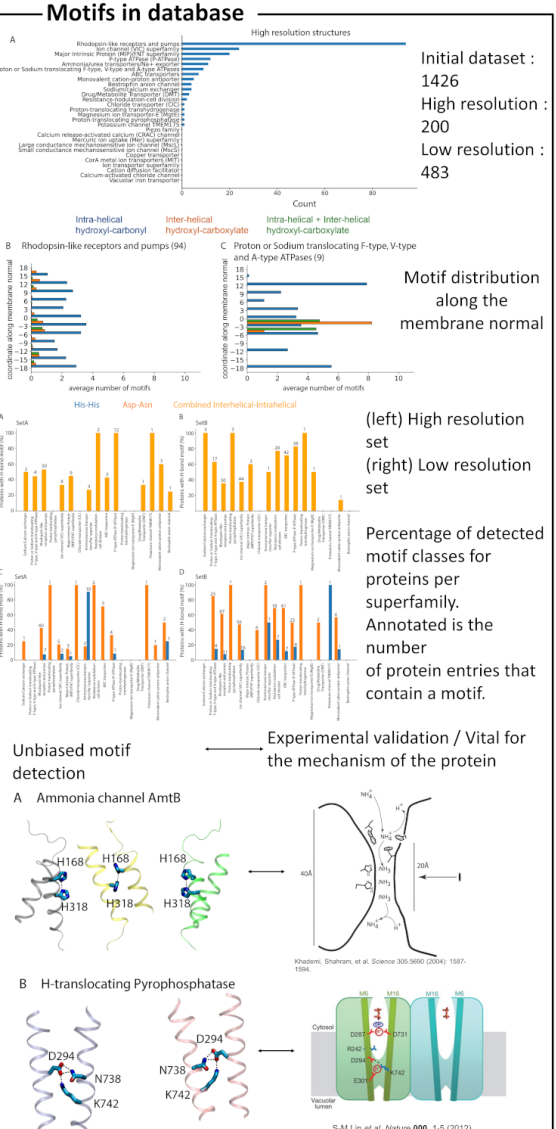
1. Corona virus spike protein S is a large homo-trimeric protein anchored in the membrane of the virion particle. Protein S binds to angiotensin-converting-enzyme 2, ACE2, of the host cell, followed by proteolysis of the spike protein, drastic protein conformational change with exposure of the fusion peptide of the virus, and entry of the virion into the host cell. We have developed a methodology that relies upon graph and centrality analyses, to identify and characterize large H-bond clusters in protein structures.
2. pH sensing by proteins is a fundamental reaction that typically involves chemical binding and unbinding of protons at titratable amino acid residue sidechains. The mechanisms by which proteins sense pH are poorly described, particularly whether groups that titrate at physiological pH, such as carboxylate and imidazole groups, function as pH sensors only when found in specific sequence and structure motifs. Here we address this question by employing data science approaches. We established a large hand-curated dataset of three-dimensional structures of membrane proteins, and a subset of representative protein structures solved at high resolution.

## Spike Protein Networks




We find that, in the closed conformation, the three protomers of protein S bring the same contribution to an extensive central network of H-bonds and contribute symmetrically to a relatively large H-bond cluster at the receptor binding domain, and to a cluster near a protease cleavage site. Markedly different H-bonding at these three clusters in open and pre-fusion conformations suggest dynamic H-bond clusters could facilitate structural plasticity and selection of a protein S protomer for binding to the host receptor, and proteolytic cleavage.

## Motifs in database




Initial dataset : 1426
High resolution : 200
Low resolution : 483

Motif distribution along the membrane normal

(left) High resolution set
(right) Low resolution set

Percentage of detected motif classes for proteins per superfamily. Annotated is the number of protein entries that contain a motif.

Unbiased motif detection ⟷ Experimental validation / Vital for the mechanism of the protein

A Ammonia channel AmtB




Khademi, Shahram, *et al. Science* 305.5690 (2004): 1587-1594.

B H-translocating Pyrophosphatase




S-M Lin *et al. Nature* **000**, 1-5 (2012)
doi:10.1038/nature10963

## References

[1] Siemers, M.; Lazaratos, M.; Karathanou, K.; Guerra, F.; Brown, L. S.; Bondar, A. N., Bridge: A Graph-Based Algorithm to Analyze Dynamic H-Bond Networks in Membrane Proteins. *J Chem Theory Comput* **2019**, 15, 6781-6798.
[2] Karathanou, K., Lazaratos, M., Bertalan, É., Siemers, M., Buzar, K., Schertler, G.F., Del Val, C. and Bondar, A.N., 2020. A graph-based approach identifies dynamic H-bond communication networks in spike protein S of SARS-CoV-2. Journal of structural biology, 212(2), p.107617.
[3] Lazaratos, M.; Brown, L. S.; Bondar, A. N., Conserved H-bond motifs of membrane proteins, in preparation, 2021.
[3] Khademi, S., O'Connell, J., Remis, J., Robles-Colmenares, Y., Miercke, L.J. and Stroud, R.M., 2004. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 Å. Science, 305(5690), pp.1587-1594.
[4] Lin, S.H., Tsai, J.Y., Hsiao, C.D., Huang, Y.T., Chiu, C.L., Liu, M.H., Tung, J.Y., Liu, T.H., Pan, R.L. and Sun, Y.J., 2012. Crystal structure of a membrane-embedded H+-translocating pyrophosphatase. Nature, 484(7394), pp.399-403.

We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support provided to the Sonderforschungsbereich 1078 (SFB 1078) on 'Protonation Dynamics in Protein Function'. An allocation of computing time was provided by the HLRN, the North-German Supercomputing Alliance (bec00193, bec00221).