Molecular mimicry between Spike and human thrombopoietin may induce thrombocytopenia in COVID-19

Authors: Janelle Nunez-Castilla[1], Vitalii Stebliankin[2], Prabin Baral[3], Christian A Balbin[1], Masrur Sobhan[4], Trevor Cickovski[2], Ananda Mohan Mondal[2,4,5], Giri Narasimhan[2,5], Prem Chapagain[3,5], Kalai Mathee[5,6], and Jessica Siltberg-Liberles[1,5,*]

Affiliations:

[1]Department of Biological Sciences, College of Arts, Science and Education, Florida International University; 11200 SW 8th St, Miami, 33199, USA

[2]Bioinformatics Research Group (BioRG), Knight Foundation School of Computing and Information Sciences, Florida International University; 11200 SW 8th St, Miami, 33199, USA

[3]Department of Physics, College of Arts, Science and Education, Florida International University; 11200 SW 8th St, Miami, 33199, USA

[4]Machine Learning and Data Analytics Group (MLDAG), Knight Foundation School of Computing and Information Sciences, Florida International University; 11200 SW 8th St, Miami, 33199, USA

[5]Biomolecular Sciences Institute, Florida International University; 11200 SW 8th St, Miami, 33199, USA

[6]Department of Human and Molecular Genetics, Herbert Wertheim College of Medicine, Florida International University; 11200 SW 8th St, Miami, 33199, USA

*Corresponding author. Email: jliberle@fiu.edu

Abstract: Thrombocytopenia, characterized by reduced platelet count, increases mortality in COVID-19 patients. We performed a computational investigation of antibody-induced cross-reactivity due to molecular mimicry between SARS-CoV-2 Spike protein and human thrombopoietin, the regulator of platelet production, as a mechanism for thrombocytopenia in COVID-19 infections. The presence of a common sequence motif with similar structure and antibody-binding properties for these proteins strongly indicate shared molecular mimicry. Recent reports of antibodies in COVID-19 patients and pre-pandemic samples against epitopes containing the motif offer additional support for the cross-reactivity. Altogether, this suggests cross-reactivity between an antibody with affinity for Spike protein and a human protein. Consideration of cross-reactivity for SARS-CoV-2 is important for therapeutic intervention and when designing the next generation of COVID-19 vaccines to avoid potential autoimmune interference.

One-Sentence Summary: SARS-CoV-2 Spike protein may trigger the production of antibodies that can cross-react with human thrombopoietin.

Main Text: Molecular mimicry between pathogen and human proteins can cause antibodies to erroneously attack human proteins sometimes leading to transient or chronic autoimmune disorders (*1*). COVID-19 patients often suffer from thrombocytopenia, characterized by low platelet levels (*2*), which correlates with an almost 5-fold increase in mortality (*3*). Platelet production is regulated by the cytokine thrombopoietin (*4*) (Fig. S1). Thrombocytopenia in COVID-19 patients resembles immune thrombocytopenia, where human thrombopoietin (hTPO) and/or its receptor are mistakenly targeted by autoantibodies leading to reduced platelet count (*5*). Treatments with TPO Receptor Agonists improve thrombocytopenia in both general (*6*) and COVID-19 (*7*) patients, suggesting the mistaken targeting occurs before TPO activates the TPO receptor.

A molecular mimicry search between Spike from SARS-CoV-2, the causative pathogen of COVID-19, and human epitopes from The Immune Epitope Database (IEDB, (*8*)), identified a shared five-amino acid motif, TQLPP, in Spike and hTPO (Fig. 1A). The motif on the Spike protein is located on the surface of the N-terminal Domain (NTD) (Fig. 1B-D), whereas it is found at the interface with a neutralizing antibody in hTPO (*9*) (Fig. 1E, F). The TQLPP motifs from the two proteins exhibit high structural similarity (Fig. 1G, H). The motif on Spike is adjacent to the NTD supersite targeted by neutralizing antibodies (*10*). We hypothesized that COVID-19 infection may trigger the production of TQLPP-specific antibodies against this epitope that can cross-react with hTPO. In the absence of Spike TQLPP antibodies, we investigated the binding of an hTPO neutralizing mouse Fab antibody (TN1) (*11*) to the Spike TQLPP epitope using molecular modeling and machine learning.

We constructed a composite model of Spike and TN1 by coupling a structure of hTPO complexed with TN1 Fab (*9*) with a published full-length glycosylated model of the Spike protein (*12*). The glycosylated model based on PDB id 6VSB (*13*) with the first 26 residues (including the TQLPP motif) reconstructed (*12*), was used. We energy minimized the Spike-TN1 complex and equilibrated it using molecular dynamics (MD) simulation (1 ns). The final model of the Spike trimer complexed with three TN1 Fab antibodies (Fig. 2A, B) shows that the TQLPP epitope is accessible to the antibody and the adjacent glycan chains do not shield the antibody-binding site (Fig. 2C, Fig. S2). To confirm the conformation of TQLPP, we calculated the RMSD for TQLPP regions from 60 Spike proteins from PDB, plus the modeled states (before and after equilibration, and upon 200 ns MD simulation) in an all-vs-all manner (Fig. S3). The reconstructed TQLPP region falls within the conformational ensemble from PDB, suggesting that the modeled representation of TQLPP is valid.

To evaluate the molecular mimicry between the antibody interface areas, we performed MD simulations of hTPO and Spike NTD with TQLPP complexed with the TN1 antibody. The hydrogen bonds were calculated between the TN1 antibody with hTPO and Spike, respectively, from the last 50 ns of both trajectories (Fig. 2F, G). Both the Spike-TN1 and the hTPO-TN1 complexes showed similar contact areas (Fig. S4). Besides the critical hydrogen bonds observed for residues Q and L in the TQLPP motif, interactions with TN1 were observed for R237 in Spike and R98 in hTPO (Fig. 2D-G).

To further support our findings, we evaluated the antibody-antigen interface complementarity with MaSIF-search, a recent tool that uses deep learning techniques (*14*), on a pair of circular surface regions (patches) from an antibody-antigen complex. MaSIF-search produces a score associated with the strength of binding when forming a stable complex. Lower scores are associated with stronger binding. We refer to this score here as the binding score. The results show that Spike-TN1 complexes have a better (lower) binding score than random

complexes and that complexes including Spike from PDB id 7LQV (*10*), have three of the four best binding scores (0.86, 1.05, 1.42) and may bind to TN1 as well as, or better than, hTPO (Fig. 2H, Tables S3, S4). Notably, in 7LQV, COVID-19 antibodies bind to Spike at the NTD supersite (*10*). These results strongly argue for the possibility of cross-reactivity between Spike and hTPO driven by the molecular mimicry of the TQLPP epitope (Fig. 2).

The human proteome contains nine additional occurrences of the TQLPP motif. Two of these motifs, from Hermansky-Pudlak syndrome 4 protein and ALG12 (Mannosyltransferase ALG12 homolog), have been associated with thrombosis and hemostasis disorder (*18*). To evaluate structural mimicry between Spike-TQLPP and all human-TQLPP motifs, we utilized AlphaFold predictions (*15*, *16*) (Fig. S5). The closest structural mimic is hTPO (RMSD = 0.39 Å), followed by coiled-coil domain-containing protein 185, Fc receptor-like protein 4 (FCRL4), and far upstream element-binding protein 1 (Fig. S5). These results indicate that TQLPP motifs have similar conformations (Fig. S3), strengthening the notion of structural mimicry. We investigated the potential cross-reactivity of an antibody targeting TQLPP in these proteins, after discarding six that display the TQLPP motif in low confidence or unstructured regions. The remaining three proteins, NEK10 (ciliated cell-specific kinase), FCRL4, and ALG12 were complexed with TN1 (Fig. 3). The binding score for NEK10-TN1 (1.44) is comparable to the hTPO-TN1 complex (Fig. 3). NEK10 regulates motile ciliary function responsible for expelling pathogens from the respiratory tract (*17*). NEK10 dysfunction can impact mucociliary clearance and lead to respiratory disorders such as bronchiectasis (*17*). Based on our results, it is plausible that the function of NEK10 and thus mucociliary clearance can be affected by cross-reactive Spike antibodies targeting TQLPP.

Our results suggest a high potential for cross-reactivity between Spike and hTPO involving the TQLPP epitope that could possibly affect platelet production and lead to thrombocytopenia. Further, cross-reactivity with other TQLPP-containing proteins such as NEK10 cannot be dismissed. The presence of neutralizing antibodies against peptides with TQLPP in COVID-19 patients' convalescent plasma (*18*), particularly in severe and fatal cases (*19*) adds credence to our hypothesis. It is also interesting to note that antibodies against a TQLPP-containing peptide were found in the serum of pre-pandemic, unexposed individuals (*20*). Prior infection with a different human coronavirus cannot explain the cross-reactivity observed in the unexposed group because TQLPP is situated in a region with low amino acid conservation (*20*). Rather, this suggests the presence of an antibody for an unknown epitope with affinity for the TQLPP region in the Spike protein. The COVID-19 vaccines designed to deliver the Spike protein from SARS-CoV-2, like COVID-19 infection itself, can cause thrombocytopenia (*2*, *21*, *22*) and it is plausible that cross-reactivity can titrate the serum concentration of free hTPO. Consideration of the autoimmune potential of pathogens like COVID-19 is important for therapeutic intervention, and, particularly, when designing the next generation of COVID-19 vaccines to avoid potential predictable autoimmune interference. The evolutionary trends in the TQLPP motif suggest it may not remain in Spike. In the Gamma variant, a P26S mutation has changed TQLPP to TQLPS and two additional mutations are located just before the motif at L18F and T20N in the NTD supersite, while the Delta variant is mutated at T19R (*23*). Neutralizing antibodies targeting the NTD supersite may rapidly lose efficacy against the evolving SARS-CoV-2. Consequently, protein engineering of the TQLPP motif and possibly the NTD supersite for future COVID-19 vaccines may reduce the risk for thrombocytopenia and improve long-term vaccine protection against future variants.

References and Notes

1. D. R. Getts, E. M. Chastain, R. L. Terry, S. D. Miller, Virus infection, antiviral immunity, and autoimmunity. *Immunol. Rev.* **255**, 197–209 (2013).
2. X. Yang, Q. Yang, Y. Wang, Y. Wu, J. Xu, Y. Yu, Y. Shang, D. Lillicrap, Thrombocytopenia and its association with mortality in patients with COVID-19. *J Thromb Haemost*. **18**, 1469–1472 (2020).
3. C. Shi, L. Wang, J. Ye, Z. Gu, S. Wang, J. Xia, Y. Xie, Q. Li, R. Xu, N. Lin, Predictors of mortality in patients with coronavirus disease 2019: a systematic review and meta-analysis. *BMC Infect. Dis.* **21**, 663 (2021).
4. L. N. Varghese, J.-P. Defour, C. Pecquet, S. N. Constantinescu, The Thrombopoietin Receptor: Structural Basis of Traffic and Activation by Ligand, Mutations, Agonists, and Mutated Calreticulin. *Front. Endocrinol. (Lausanne).* **8**, 1 (2017).
5. I. Nazy, J. G. Kelton, J. C. Moore, R. Clare, P. Horsewood, J. W. Smith, N. Ivetic, V. D'Souza, N. Li, D. M. Arnold, Autoantibodies to thrombopoietin and the thrombopoietin receptor in patients with immune thrombocytopenia. *Br. J. Haematol.* **181**, 234–241 (2018).
6. S. Audia, B. Bonnotte, Emerging Therapies in Immune Thrombocytopenia. *J. Clin. Med.* **10**, 1004 (2021).
7. A. Watts, K. Raj, P. Gogia, C. C. T. Gahona, M. Porcelli, Secondary Immune Thrombocytopenic Purpura Triggered by COVID-19. *Cureus*. **13** (2021), doi:10.7759/CUREUS.14505.
8. R. Vita, S. Mahajan, J. A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, D. K. Wheeler, A. Sette, B. Peters, The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2019).
9. M. D. Feese, T. Tamada, Y. Kato, Y. Maeda, M. Hirose, Y. Matsukura, H. Shigematsu, T. Muto, A. Matsumoto, H. Watarai, K. Ogami, T. Tahara, T. Kato, H. Miyazaki, R. Kuroki, Structure of the receptor-binding domain of human thrombopoietin determined by complexation with a neutralizing antibody fragment. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1816–1821 (2004).
10. G. Cerutti, Y. Guo, T. Zhou, J. Gorman, M. Lee, M. Rapp, E. R. Reddem, J. Yu, F. Bahna, J. Bimela, Y. Huang, P. S. Katsamba, L. Liu, M. S. Nair, R. Rawi, A. S. Olia, P. Wang, B. Zhang, G. Y. Chuang, D. D. Ho, Z. Sheng, P. D. Kwong, L. Shapiro, Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe*. **29**, 819-833.e7 (2021).
11. T. Tahara, T. Kuwaki, A. Matsumoto, H. Morita, H. Watarai, Y. Inagaki, H. Ohashi, K. Ogami, H. Miyazaki, T. Kato, Neutralization of biological activity and inhibition of receptor binding by antibodies against human thrombopoietin. *Stem Cells*. **16**, 54–60 (1998).
12. H. Woo, S. J. Park, Y. K. Choi, P. Park, M. Tanveer, Y. Cao, N. R. Kern, J. Lee, M. S. Yeom, T. I. Croll, C. Seok, W. Im, Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. *J. Phys. Chem. B.* **124**, 7128–7137 (2020).
13. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. **367**, 1255–1260 (2020).
14. P. Gainza, F. Sverrisson, F. Monti, E. Rodolà, D. Boscaini, M. M. Bronstein, B. E.

Correia, Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods 2019 172*. **17**, 184–192 (2019).
15. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nat. 2021*, 1–11 (2021).
16. K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, Highly accurate protein structure prediction for the human proteome. *Nat. 2021*, 1–9 (2021).
17. R. R. Chivukula, D. T. Montoro, H. M. Leung, J. Yang, H. E. Shamseldin, M. S. Taylor, G. W. Dougherty, M. A. Zariwala, J. Carson, L. A. Daniels, P. Sears, K. E. Black, L. P. Hariri, I. Almogarri, E. M. Frenkel, V. Vinarsky, H. Omran, M. R. Knowles, G. J. Tearney, F. S. Alkuraya, D. M. Sabatini, A human ciliopathy reveals essential functions for NEK10 in airway mucociliary clearance. *Nat. Med.* **26**, 244 (2020).
18. Y. Li, D. Lai, H. Zhang, H. Jiang, X. Tian, M. Ma, H. Qi, Q. Meng, S. Guo, Y. Wu, W. Wang, X. Yang, D. Shi, J. Dai, T. Ying, J. Zhou, S. Tao, Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell. Mol. Immunol. 2020 1710*. **17**, 1095–1097 (2020).
19. C. Voss, S. Esmail, X. Liu, M. J. Knauer, S. Ackloo, T. Kaneko, L. Lowes, P. Stogios, A. Seitova, A. Hutchinson, F. Yusifov, T. Skarina, E. Evdokimova, P. Loppnau, P. Ghiabi, T. Haijan, S. Zhong, H. Abdoh, B. D. Hedley, V. Bhayana, C. M. Martin, M. Slessarev, B. Chin-Yee, D. D. Fraser, I. Chin-Yee, S. S. Li, Epitope-specific antibody responses differentiate COVID-19 outcomes and variants of concern. *JCI insight*. **6** (2021), doi:10.1172/JCI.INSIGHT.148855.
20. C. Stoddard, J. Galloway, H. Chu, M. Shipley, K. Sung, H. Itell, C. Wolf, J. Logue, A. Magedson, M. Garrett, K. Crawford, U. Laserson, F. Matsen, J. Overbaugh, Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. *Cell Rep.* **35** (2021), doi:10.1016/J.CELREP.2021.109164.
21. J. M. Helms, K. T. Ansteatt, J. C. Roberts, S. Kamatam, K. S. Foong, J. M. S. Labayog, M. D. Tarantino, Severe, refractory immune thrombocytopenia occurring after sars-cov-2 vaccine. *J. Blood Med.* **12**, 221–224 (2021).
22. N. H. Schultz, I. H. Sørvoll, A. E. Michelsen, L. A. Munthe, F. Lund-Johansen, M. T. Ahlen, M. Wiedmann, A.-H. Aamodt, T. H. Skattør, G. E. Tjønnfjord, P. A. Holme, Thrombosis and Thrombocytopenia after ChAdOx1 nCoV-19 Vaccination. *N. Engl. J. Med.*, NEJMoa2104882 (2021).
23. E. B. Hodcroft, CoVariants: SARS-CoV-2 Mutations and Variants of Interest. (2021), (available at https://covariants.org/).
24. W. R. Taylor, Residual colours: a proposal for aminochromography. *Protein Eng.* **10**, 743–746 (1997).
25. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, G. J. Barton, Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*.

25, 1189–1191 (2009).
26. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **245**, 403–410 (1990).
27. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol. 2017 3511*. **35**, 1026–1028 (2017).
28. Y. Cai, J. Zhang, T. Xiao, H. Peng, S. M. Sterling, R. M. Walsh, S. Rawson, S. Rits-Volloch, B. Chen, Distinct conformational states of SARS-CoV-2 spike protein. *Science (80-.).* **369**, 1586–1592 (2020).
29. W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. **22**, 2577–2637 (1983).
30. M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman, C. O. Wilke, Maximum allowed solvent accessibilites of residues in proteins. *PLoS One*. **8** (2013), doi:10.1371/journal.pone.0080635.
31. Y. Zhang, J. Skolnick, TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
32. Y. Choi, Y. Cao, M. Frank, H. Woo, S. Park, M. Yeom, T. Croll, C. Seok, W. Im, Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. *J. Chem. Theory Comput.* **17**, 2479–2487 (2021).
33. W. L. DeLano, The PyMOL Molecular Graphics System, Schrödinger, LLC, (available at www.pymol.org).
34. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
35. S. Jo, T. Kim, V. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
36. B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
37. J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda, W. Im, 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 (2016).
38. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
39. S. Nosé, M. L. Klein, Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **50**, 1055–1076 (1983).
40. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577 (1998).
41. J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
42. M. F. Sanner, A. J. Olson, J. C. Spehner, Reduced surface: an efficient way to compute molecular surfaces - PubMed. *Biopolymers*. **38**, 305–320 (1996).

43. J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, C. M. Deane, SAbDab: the structural antibody database. *Nucleic Acids Res.* **42**, D1140–D1146 (2014).
44. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).
45. D. J. Kuter, The biology of thrombopoietin and thrombopoietin receptor agonists. *Int. J. Hematol. 2013 981*. **98**, 10–23 (2013).

Acknowledgments: We thank Sathibalan Ponniah, Charles Dimitroff, Sixto Leal, and Kevin Bennett for useful discussions. We also thank Yuk-Ching Tse-Dinh and Tim Collins for helpful comments on the manuscript.

Funding: This work was partially supported by the National Science Foundation under Grant No. 2037374 (JNC, PB, VS, KM, GN, PC, AMM, JSL).

Author contributions:

Conceptualization: GN, PC, KM, TC, AMM, JSL
Methodology: JNC, VS, PB, CB, GN, PC, KM, TC, AMM, JSL
Investigation: JNC, VS, PB, CB, MS, GN, PC, KM, TC, AMM, JSL
Visualization: JNC, PB, CB, VS, MS
Funding acquisition: GN, PC, KM, AMM, JSL
Project administration: GN, JSL
Supervision: GN, PC, JSL, AMM
Writing – original draft: JNC, VS, PB, CB, GN, PC, KM, TC, JSL
Writing – review & editing: JNC, VS, PB, CB, GN, PC, KM, TC, AMM, JSL

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or supplementary materials.

Supplementary Materials

Materials and Methods

Figs. S1 to S5

Tables S1 to S5
References (*26–45)*

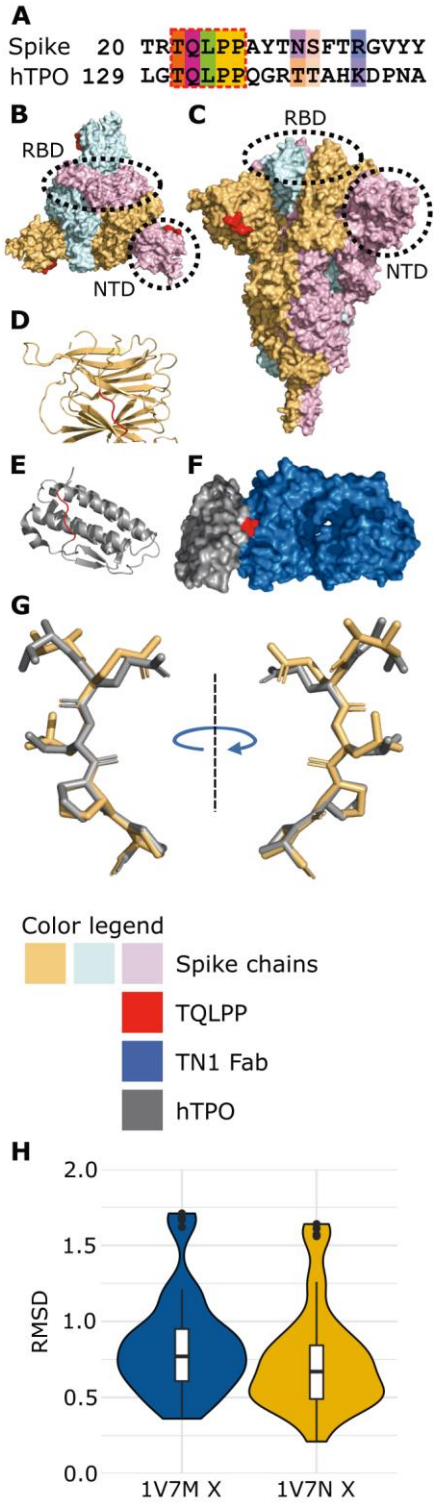

Fig. 1. Structural mimicry between a TQLPP motif in SARS-CoV-2 Spike and an antibody binding epitope in thrombopoietin. (**A**) Pairwise alignment for TQLPP motif in the epitope for human thrombopoietin (hTPO, Epitope ID: 920946) and Spike, colored by Taylor (*24*) for sites with ≥50% conservation in the amino acid property (*25*). The region of molecular mimicry is highlighted in the red dashed box. Surface representation of Spike from (**B**) the side and (**C**) above, with Spike trimer (PDB id: 6XR8) colored by subunit and red indicating the location of

the TQLPP epitope fragment, illustrating the surface accessibility of TQLPP. (**D**) Ribbon representation of Spike shows TQLPP (red) in a loop. (**E**) Ribbon representation of hTPO (gray, PDB id: 1V7M) without antibody shows TQLPP (red) in a loop. (**F**) Surface representation shown for hTPO (gray) and its TN1 antibody (blue). (**G**) TM-align generated structural alignment for TQLPP in Spike (beige) and hTPO (gray), with RMSD = 0.61 Å. (**H**) Violin plots of RMSD values resulting from the comparison of TQLPP region in 60 Spike structures vs TQLPP in two hTPO structures (PDB ids: 1V7M and 1V7N, chain X for both). Statistical analysis with Mann-Whitney U reveals no statistical significance between the sets. For further details, see supplementary methods.

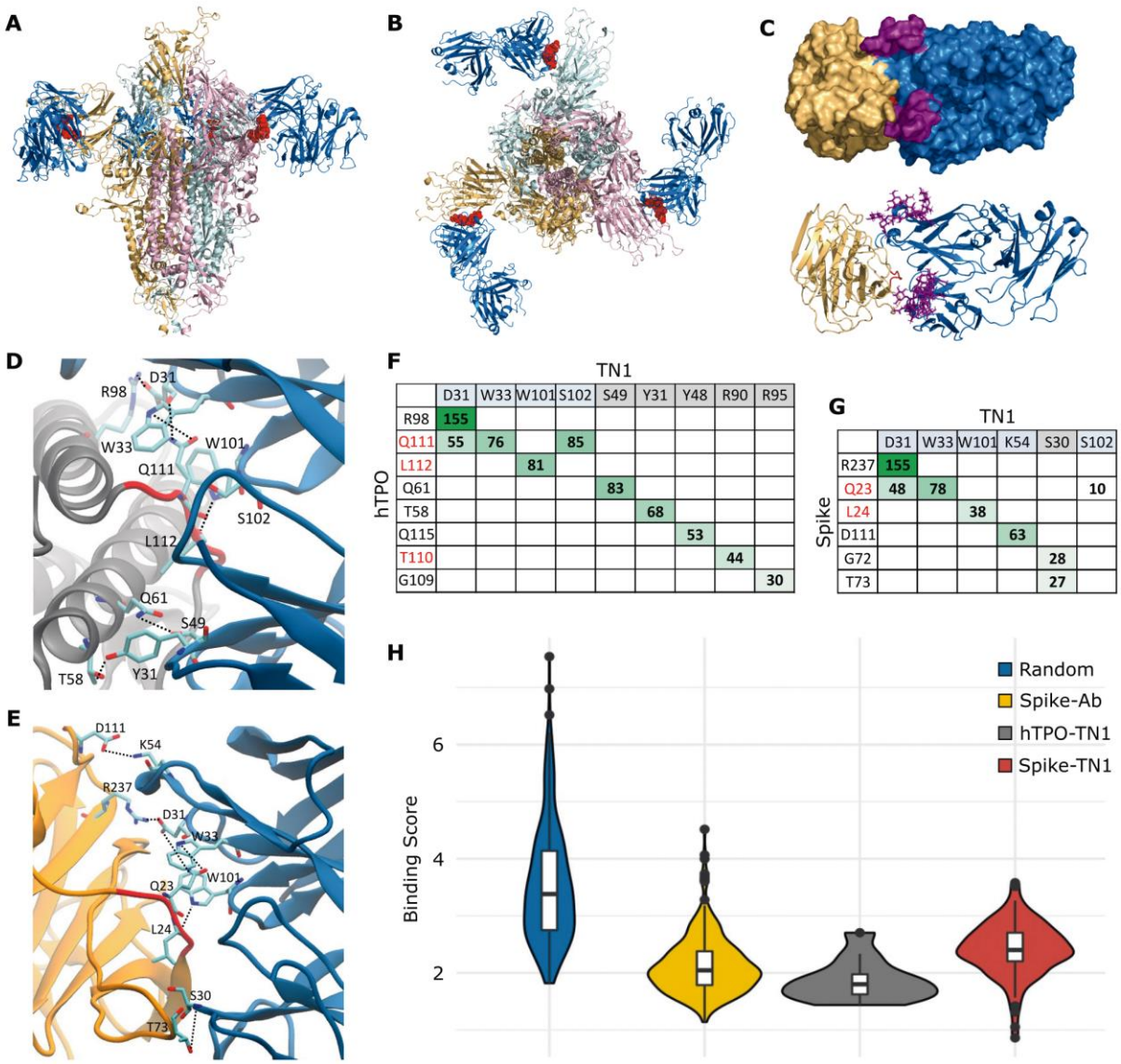

Fig. 2. Binding of SARS-CoV-2 Spike to TN1 Fab antibody. Equilibrated structure (1 ns) of the modeled TN1 Fab antibody (blue, PDB id: 1V7M) complexed with Spike trimer model shown from (**A**) the side and (**B**) the top, with TQLPP shown as red spheres. (**C**) The Spike NTD (beige) and TN1 Fab complex were used for MD simulations (200 ns), with adjacent glycans at N17 and N74 highlighted in purple. The representative amino acids contributing to hydrogen bonds (dashed lines) during the last 50 ns of simulations for the (**D**) hTPO-TN1 and (**E**) Spike-TN1 complexes are highlighted as cyan sticks. Interaction matrices showing the percentage of hydrogen bonds contribution between amino acid residue pairs ordered according to their hydrogen-bond occupancies for the (**F**) hTPO-TN1 and (**G**) Spike-TN1 complexes. Residues belonging to TQLPP are colored in red and positions for hTPO are based on the PDB structure. TN1 Fab residues from heavy and light chains are shaded blue and gray, respectively. (**H**) Violin plot showing the distribution of the MaSIF binding score values for randomly selected patch pairs (blue), the interacting region of Spike-antibody (yellow) and hTPO-TN1 (gray) complexes, and for modeled Spike-TN1 complexes across 40 Spike configurations (red). Statistical analysis with Mann-Whitney U shows that all pairwise comparisons except for Spike-Ab and hTPO-TN1

are significantly different after Bonferroni correction (Table S4). For further details, see supplementary methods.

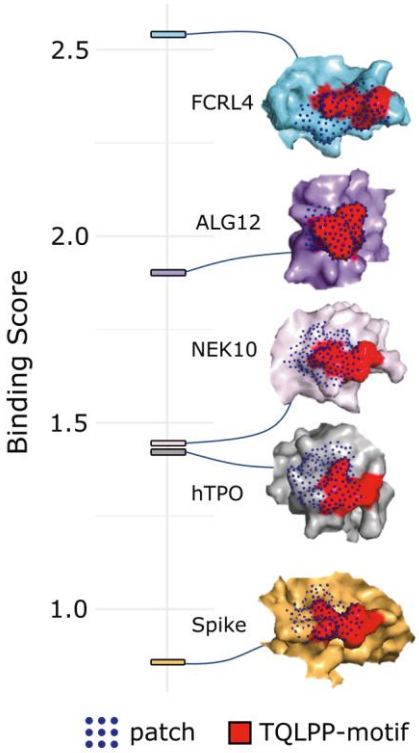

Fig. 3. Predicted interaction patches between TN1 Fab antibody (PDB id: 1V7N) and the TQLPP motif. The best (lowest) binding score is shown for Spike (PDB id: 7LQV, chain A), hTPO (PDB id: 1V7N, chain X), NEK10 (Uniprot: Q6ZWH5), ALG12 (Uniprot: Q9BV10), and FCRL4 (Uniprot: Q96PJ5). For all, red indicates the TQLPP motif and blue dots represent the surface points included in the predicted MaSIF patches.

Supplementary materials including methods

Materials and Methods

Methods

Identifying epitopes with molecular mimicry

Two databases were constructed: (1) EPI-SEQ, a local BLAST (*26*) database for all linear epitopes with positive assays in the Immune Epitope Database (IEDB) (*8*); and EPI-PDB, a library of all PDB structures corresponding to the source sequences of the epitopes in EPI-SEQ, generated using MMseqs2 with sensitivity set to 7.5 (*27*). A Cryo-EM structure of Spike from SARS-CoV-2 (PDB id 6XR8 (*28*)) was used as the representative structure during this step. Accessible surface area (ASA) per residue in this structure was computed using DSSP (*29*). Subsequently, relative accessible surface area (RASA) was determined by the equation *RASA = ASA/MaxASA*, where the Wilke table (*30*) was used to obtain MaxASA values for each residue. Each site was classified as accessible or buried using 20% RASA as a threshold.

To search EPI-SEQ, the amino acid sequence derived from PDB id 6XR8 (chain A) was used as the query in a local BLASTP search against EPI-SEQ using the BLOSUM62 matrix with both an E-value threshold and max target seqs value of 2,000,000, effectively turning off the E-value filter and removing the match limit, respectively. BLAST hits containing 5 or more consecutive residues with 100% sequence identity where at least 3 of the query residues are surface accessible are considered sequence-based molecular mimics.

We then retrieved a structural representation from EPI-PDB for each sequence mimic. All sequence mimics with no corresponding structural representation were discarded. TM-align (*31*) was used to generate a structural alignment for all query-hit (sequence mimic) alignment pairs using only the structural regions corresponding to the hit for the source antigenic protein and the query (SARS-CoV-2 Spike (PDB id: 6XR8))(*31*). In summary, TM-align performed the structural superposition step based on the BLAST alignment, calculated RMSD, and a TM-align score. Mimics with an RMSD $\leq$ 1Å were considered structure-based molecular mimics. One of these hits corresponded to the human thrombopoietin (hTPO) protein (Epitope ID: 920946, UniProt ID: P40225.1, residues 131-135: TQLPP), where the structural representative for hTPO was found bound to the TN1 Fab antibody (PDB id: 1V7M, [11]). In Spike, the TQLPP motif corresponds to residues 22-26 of the NCBI reference sequence (YP_009724390.1).

Conformational ensemble of structural mimicry

To gather all structures of the TQLPP motif in Spike, an NCBI BLASTP search against PDB was performed with the SARS-CoV-2 Spike reference sequence as the query and a SARS-CoV-2 taxa filter. Of 75, close to full-length, hits (>88 % query cover), 20 included a solved structure for the TQLPP motif. The TQLPP region of the PDB structure was extracted for all chains in the 20 structures (all were trimers, as is Spike's biological state) resulting in a TQLPP Spike ensemble of 60 different chains from SARS-CoV-2. Each sequence in the TQLPP Spike ensemble was superimposed with chain X of the two PDB structures of hTPO (PDB ids: 1V7M and 1V7N) to generate an RMSD value distribution for Spike's conformational ensemble vs hTPO for the structural mimicry region (Table S1).

Modeling Spike-TN1 complex

We constructed a composite model of the Spike-TN1 complex using the hTPO-TN1 complex as a template. For this, we first aligned the TQLPP segment of hTPO in the hTPO-TN1 complex with the TQLPP segment of the fully glycosylated model of the Spike protein (PDB id: 6VSB) (*14*) retrieved from the CHARMM-GUI Archive (*32*). We then removed hTPO, leaving

TN1 complexed with Spike. For the Spike-TN1 simulations, only the TN1 interacting N-terminal domain of the Spike protein (residues 1-272) was considered. Geometrical alignments, as well as visualization, were performed with PyMOL version 2.4 (*33*) and Visual Molecular Dynamics (VMD 1.9.3, (*34*)).

To confirm that the modeled Spike TQLPP region is in agreement with the TQLPP region of solved Spike structures, TM-align was used as described above to superimpose the TQLPP regions in different structures, including the modeled TQLPP region from the Spike-TN1 complex, and to calculate the respective RMSD values. Three states of the model were included (before and after equilibration, and after molecular dynamics (described in the following paragraph)) together with the 60 experimentally determined Spike structures in Table S1 and compared in an all-against-all manner (Fig S3, Table S2).

Molecular dynamics simulation

A simulation system for the modeled Spike-TN1 complex was prepared using CHARMM-GUI (*35–37*). The complex was solvated using a TIP3P water model and 0.15 M concentration of KCl and equilibrated for 1 ns at 303 K. All-atom simulations were performed with NAMD2.14 (*38*) using CHARMM36m force-field. The production runs were performed under constant pressure of 1 atm, controlled by a Nose−Hoover Langevin piston (*39*) with a piston period of 50 fs and a decay of 25 fs to control the pressure. The temperature was set to 303 K and controlled by Langevin temperature coupling with a damping coefficient of 1/ps. The Particle Mesh Ewald method (PME) (*40*) was used for long-range electrostatic interactions with periodic boundary conditions and all covalent bonds with hydrogen atoms were constrained by Shake (*41*). The contact area of the interface was calculated as $(S_1+S_2-S_{12})/2$, where S_1 and S_2 represent the solvent accessible surface areas of the antigen and antibody and S_{12} represents that for the complex (Fig. S4).

Antibody interface complementarity

We used the MaSIF-search geometric deep learning tool designed to uncover and learn from complementary patterns on the surfaces of interacting proteins (*14*). Surface properties of proteins are captured using radial patches. A radial patch is a fixed-sized geodesic around a potential contact point on a solvent-excluded protein surface (*42*). In MaSIF-search, the properties include both geometric and physicochemical properties characterizing the protein surface (18). This tool exploits that a pair of patches from the surfaces of interacting proteins exhibit interface complementarity in terms of their geometric shape (e.g., convex regions would match with concave surfaces) and their physicochemical properties. The data structure of the patch is a grid of 80 bins with 5 angular and 16 radial coordinates and ensures that its description is rotation invariant. Each bin is associated with 5 geometric and chemical features: shape index, distance-dependent curvature, electrostatics, hydropathy, and propensity for hydrogen bonding. The model converts patches into 80-dimensional descriptor vectors, such that the Euclidian distance between interacting patches is minimized. Here, we define the binding score as a measure of distance between the descriptor vectors of the two patches. Thus, lower “MaSIF binding scores” represent better complementarity and therefore better matches. The pre-trained MaSIF-search model "sc05" with a patch radius of 12 Å was used.

Using the MaSIF protocol, we evaluated complexes of the TN1 antibody to bind Spike protein in the TQLPP region. To accommodate multiple Spike configurations, we extracted patches from 40 SARS-CoV-2 Spike-antibody complexes from the SAbDab structural antibody database (*43*). Patches centered at Q23 from Spike and W33 from TN1 were selected as representative pairs for the Spike-TN1 interaction type because this potential contact point has

the most hydrogen bonds in the modeled Spike-TN1 TQLPP region. Binding scores of randomly formed complexes (Random), complexes between Spike and its native antibodies (Spike-Ab), and complexes between hTPO and TN1 (hTPO-TN1) were extracted and tabulated (Table S3). The distribution of binding scores from randomly formed complexes was obtained by pairing patches from random locations on Spike with patches from its antibodies. For native antibody-antigen Spike-Ab and hTPO-TN1 complexes, we obtained patch pairs from known interface regions using the MaSIF-search strategy for the selection of interacting patches (*14*). Columns "Contact AB" and "Contact AG" in Table S3 indicate the residue used as the center of the patch from the antibody and the corresponding antigen.

Evaluating further cross-reactivity

The TQLPP sequence motif was used as a query for an NCBI BLASTP search against the RefSeq Select set of proteins from the human proteome. 10 representative isoforms in proteins containing the complete TQLPP motif were found, including hTPO. The other 9 proteins lacked a solved structure containing TQLPP. However, AlphaFold v2 structural predictions were available for all 10 of these RefSeq Select sequences (*15*, *16*), allowing us to extract the region corresponding to TQLPP in these hits and structurally superpose this region with Spike TQLPP (from PDB id 6XR8 with TM-align, as described above).

TN1-protein complexes were generated for three of the remaining 9 proteins (Fc receptor-like protein 4 (residues 190-282), serine/threonine-protein kinase Nek10 (residues 1029-1146), ALG12 (Mannosyltransferase ALG12 homolog (residues 1-488)). The TQLPP segment in hTPO was structurally aligned with each of the TQLPP segments of the modeled proteins, after which, hTPO was removed resulting in the complex of TN1 with the modeled proteins following the methods mentioned for Spike above. The equilibrated structures of these complexes show that TN1 stays firmly with these proteins without any structural clash. Further, to evaluate the shape complementarity of these three proteins and TN1, MaSIF was used to calculate binding scores as described above (Table S5).

It should also be noted that two additional human genes (GeneIDs 8028 and 57110) also have one TQLPP motif, but not in the RefSeq Select isoforms. Since no structure or structural prediction was available for these proteins, they were excluded from further analysis.

Statistical analysis

Distributions were visualized as violin plots with *ggpubr* and *ggplot2*. Statistical analysis comparing the different distributions was performed using Mann-Whitney U with *Scipy*, followed by a simplified Bonferroni correction (alpha/n comparisons) when appropriate.

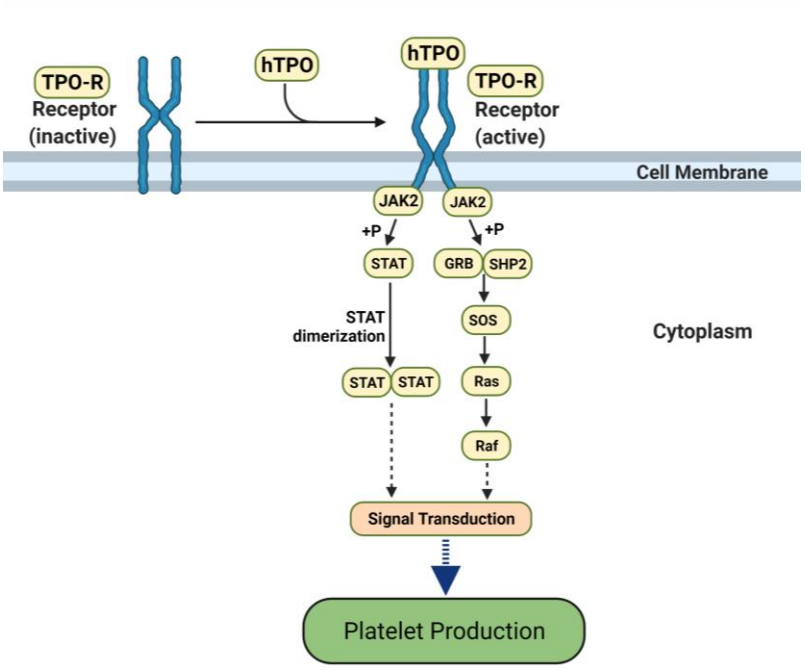

Fig. S1. hTPO pathway to induce platelet production. Simplified JAK-STAT signaling pathway in megakaryocytes where hTPO activates the TPO receptor and triggers signaling cascades that stimulate platelet production (*44*, *45*)**.** Created with BioRender.com.

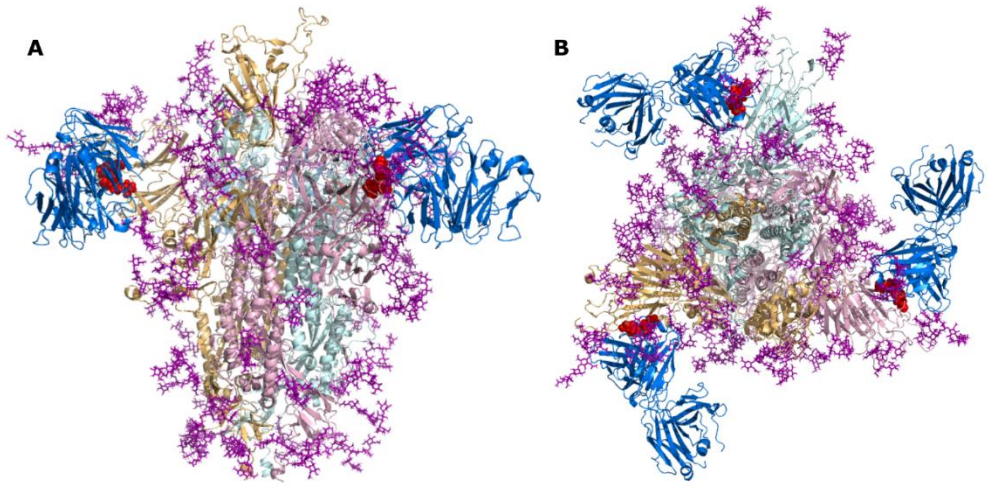

Fig. S2. SARS-CoV-2 Spike bound to TN1 Fab antibody. SARS-CoV-2 Spike shown in the trimeric state (PDB id: 6VSB) bound to TN1 Fab antibody (blue, PDB id: 1V7M) as viewed from (**A**) the side and (**B**) the top. The TQLPP motifs are shown as red spheres and glycans are shown in purple.

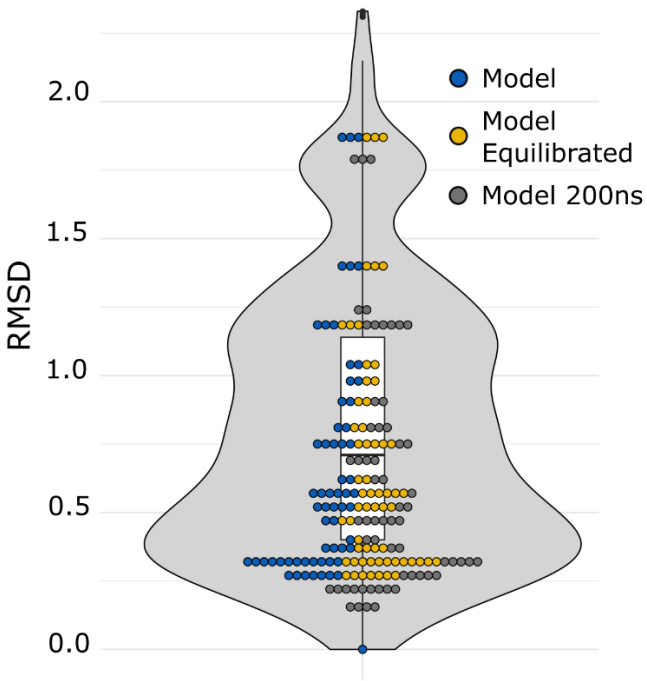

Fig. S3. RMSD value distribution for solved and modeled Spike TQLPP regions. RMSD values resulting from an all-against-all comparison of the Spike TQLPP region of 63 structures, including the model in 3 states (shown as dots).

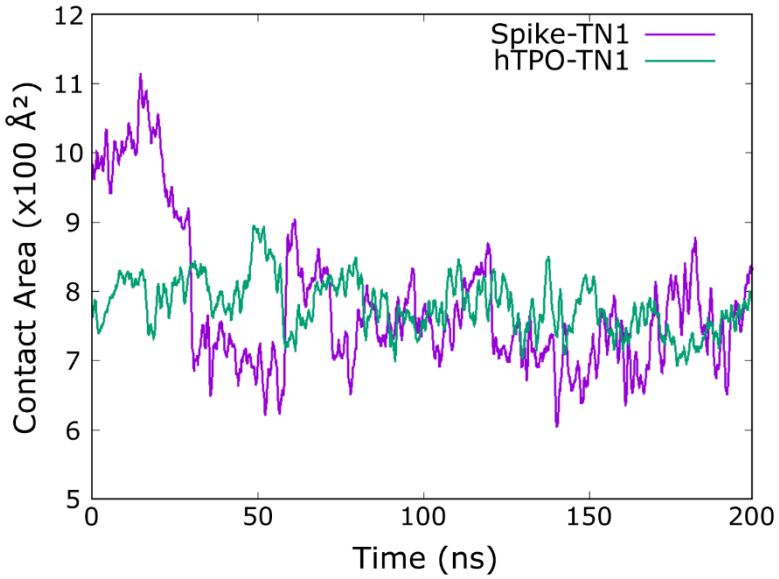

Fig. S4. Contact area of protein-antibody binding interface. Time evolution of the contact areas (100x $Å^2$) for Spike-TN1 (purple) and thrombopoietin-TN1 (green) in the molecular dynamics trajectories.

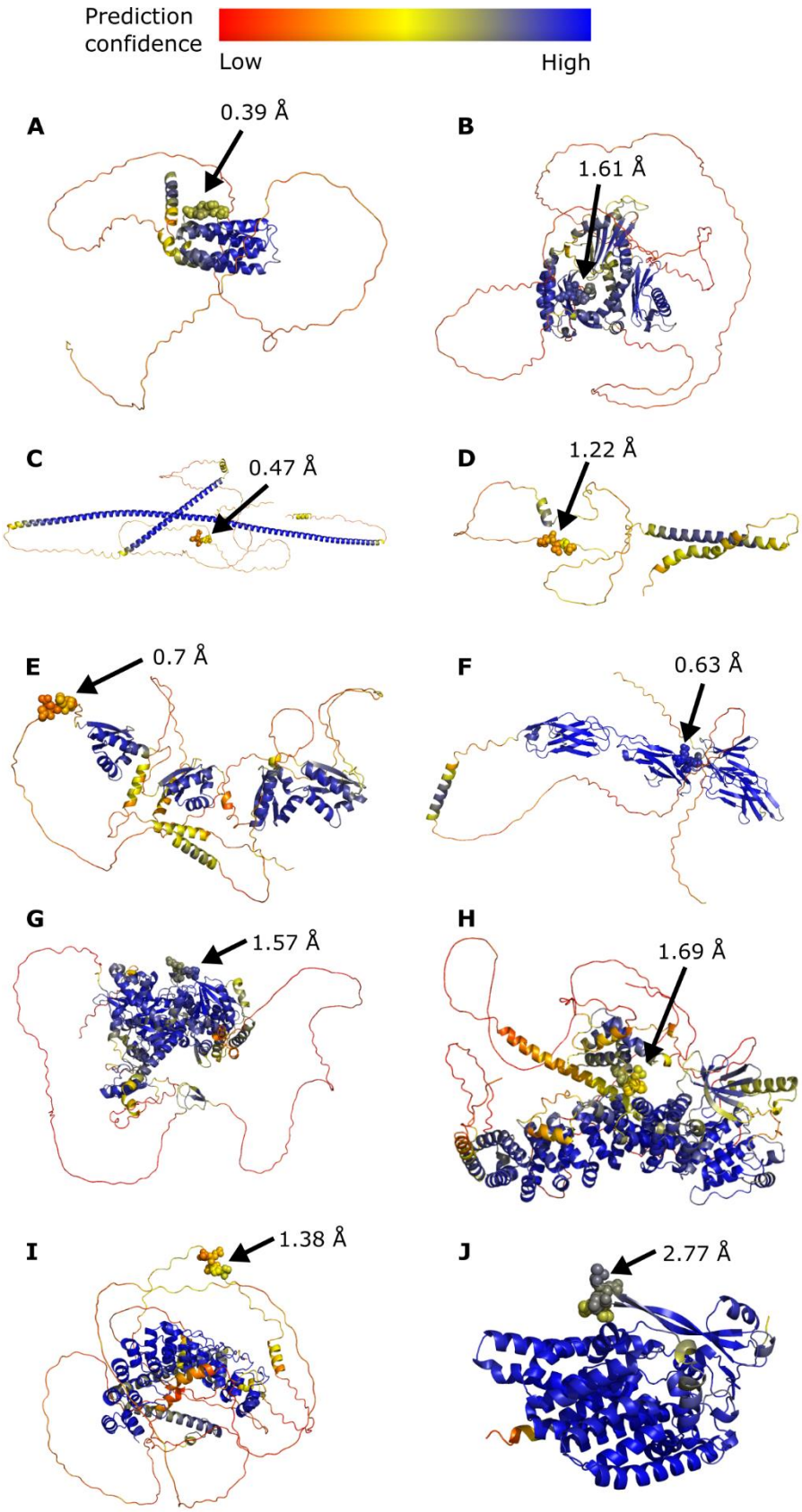

Fig. S5. TQLPP motif for 10 human proteins modeled by AlphaFold. Protein structure models are colored by AlphaFold confidence estimate according to the color bar where red = 25

(low confidence) and blue = 100 (high confidence). TQLPP motif is shown as spheres. RMSD for human TQLPP in the 10 proteins compared to SARS-CoV-2 Spike (PDB id 6XR8, chain A) is shown. The proteins are (**A**) thrombopoietin (Uniprot: P40225), (**B**) Hermansky-Pudlak syndrome 4 protein (Uniprot: Q9NQG7), (**C**) coiled-coil domain containing protein 85 (Uniprot: Q8N715), (**D**) transmembrane protein 52 precursor (Uniprot: Q8NDY8), (**E**) far upstream element-binding protein 1 (Uniprot: Q96AE4), (**F**) Fc receptor-like protein 4 (Uniprot: Q96PJ5), (**G**) DNA annealing helicase and endonuclease ZRANB3 (Uniprot: Q5FWF4), (**H**) serine/threonine-protein kinase Nek10 (Uniprot: Q6ZWH5), (**I**) espin (Uniprot: B1AK53), and (**J**) ALG12 (Mannosyltransferase ALG12 homolog, Uniprot: Q9BV10).

Table S1. RMSD values resulting from the alignment of the TQLPP region of 1V7M chain X and 1V7N chain X against the TQLPP region of 60 Spike structures. Sorted by RMSD.

1V7M X		1V7N X	
Spike Structure	**RMSD**	**Spike Structure**	**RMSD**
6ZGG A	0.36	7DCC E	0.21
6ZGG B	0.40	7DCC I	0.27
7BNN B	0.43	7DCC K	0.28
6ZGG C	0.44	7BNN B	0.42
7DCC E	0.44	7BNM C	0.44
7DCC I	0.48	7BNM B	0.44
7DCC K	0.48	7BNM A	0.44
7BNM A	0.52	7A25 C	0.46
7BNM B	0.52	6XR8 A	0.46
7BNM C	0.52	6ZGE C	0.47
7A25 C	0.59	6ZGE A	0.47
6ZGE A	0.60	6ZGE B	0.48
6ZGE B	0.60	6ZGG A	0.49
6ZGE C	0.60	7KMK B	0.49
7BNN A	0.60	7LRT B	0.49
6XR8 A	0.61	6ZGG C	0.49
7A25 A	0.63	7A25 A	0.50
7LRT B	0.66	6XR8 B	0.51
7LRT C	0.66	7LRT C	0.51
6XR8 B	0.68	6ZGG B	0.53
6XR8 C	0.71	6XR8 C	0.55
7A25 B	0.71	7A25 B	0.57
6ZP2 A	0.72	7LRT A	0.58
6ZP2 B	0.72	7N1U A	0.59
6ZP2 C	0.72	7KRQ A	0.61
7KMK B	0.72	7BNN A	0.61
7LRT A	0.73	7KRQ B	0.62
7KRQ A	0.76	7E8C A	0.64
7KRQ B	0.76	7BNN C	0.64
7N1U A	0.76	7KRQ C	0.66
7BNN C	0.78	7KMK C	0.68
7KRQ C	0.79	7E8C C	0.71
7E8C A	0.82	7E8C B	0.72
7LQV A	0.85	7N1U C	0.73
7KMK C	0.87	7JJI A	0.75
7N1U C	0.87	7JJI B	0.75
7E8C C	0.88	7JJI C	0.75

7LQV C	0.88	7N1U B	0.76
7E8C B	0.90	7KMK A	0.78
7N1U B	0.90	7LQV A	0.78
7LQV B	0.91	7LQV C	0.80
7JJI A	0.92	7LQV B	0.82
7JJI B	0.92	6ZP2 A	0.84
7JJI C	0.92	6ZP2 C	0.84
7CWL B	0.95	6ZP2 B	0.84
7CWS R	0.95	7MJG B	0.85
7KMK A	0.96	7MJG C	0.93
7MJG B	1.02	7MJG A	0.93
7MJG A	1.09	7CWL B	0.94
7MJG C	1.10	7CWS R	0.94
7CWL A	1.17	7CWS O	1.03
7CWS O	1.17	7CWL A	1.03
7C2L A	1.21	7C2L B	1.26
7C2L B	1.21	7C2L C	1.26
7C2L C	1.21	7C2L A	1.26
7N1Q B	1.62	7N1Q B	1.56
7CWL C	1.67	7CWS Q	1.57
7CWS Q	1.67	7CWL C	1.57
7N1Q A	1.68	7N1Q A	1.61
7N1Q C	1.71	7N1Q C	1.64

Table S2. RMSD values resulting from the alignment of the TQLPP region from 60 Spike structures and three modeled states, representing a conformational ensemble of TQLPP in Spike, sorted by RMSD.

Table S3. Distribution of MaSIF values.

Table S4. Statistical comparison of MaSIF binding scores for antibody complexes.

MaSIF binding scores			
Comparison		**p-value**	**Significant** [1]
Random	Spike-Ab	4.83E-75	Yes
Random	hTPO-TN1	3.00E-08	Yes
Random	Spike-TN1	5.24E-26	Yes
Spike-Ab	hTPO-TN1	3.37E-02	No
Spike-Ab	Spike-TN1	7.68E-09	Yes
hTPO-TN1	Spike-TN1	1.92E-04	Yes

[1] Compared to Bonferroni corrected p-value (<8.33E-03) for alpha = 0.05

Table S5. MASIF binding scores of other human proteins in complex with TN1.

UNIPROT ACCESSION[1]	PROTEIN NAME	MASIF BINDING SCORE	CONTACT PROTEIN[2]	CONTACT TN1
Q6ZWH5	**NEK10**	**1.44195044**	**1047 GLN**	**102 SER**
Q9BV10	**ALG12**	**1.897805691**	**466 GLN**	**102 SER**
Q96PJ5	**FCRL4**	**2.539466143**	**215 GLN**	**102 SER**

[1] Reference for AlphaFold prediction
[2] Corresponds to **Q** in T**Q**LPP