Title: **Spike mimicry of 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: Molecular mimicry between viral antigens and host proteins can produce cross-reacting antibodies leading to autoimmunity. Previous studies have shown that antibodies that cross-react with thrombopoietin, a regulator of platelet production, can induce thrombocytopenia which is characterized by low platelet count. Thrombocytopenia is associated with increased mortality in COVID-19 patients and is a rare side effect following vaccination against SARS-CoV-2. Here we show molecular mimicry between the Spike protein from SARS-CoV-2, the causative agent of COVID-19, and human thrombopoietin. Our computational analyses identify a shared five-amino acid long sequence motif (TQLPP) with similar structure and antibody-binding properties for these proteins, strongly indicative of molecular mimicry with potential for cross-reactivity. Our results are supported by reports of the presence of antibodies against epitopes containing TQLPP in both COVID-19 and pre-pandemic patient samples, indicating antibody cross-reactivity between Spike and a human protein. Since the Spike protein is the antigenic component of the SARS-CoV-2 vaccines, altering the motif may reduce the risk for thrombocytopenia and improve long-term protection against evolving variants. Our findings highlight the importance of considering peptide-level molecular mimicry when developing therapeutic interventions and designing vaccines to avoid potential autoimmune reactions.

Main: Molecular mimicry between pathogen and human proteins can cause antibodies to erroneously attack human proteins sometimes leading to transient or chronic autoimmune disorders (Getts et al. 2013). COVID-19 patients often suffer from thrombocytopenia, characterized by low platelet levels (Yang et al. 2020), which correlates with an almost 5-fold increase in mortality (Shi et al. 2021). Platelet production is regulated by the cytokine thrombopoietin (Varghese et al. 2017) (Fig. 1). 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 (Nazy et al. 2018). Treatments with TPO Receptor Agonists improve thrombocytopenia in both general (Audia and Bonnotte 2021) and COVID-19 (Watts et al. 2021) 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 (Vita et al. 2019)), identified a shared five-amino acid motif, TQLPP, in Spike and hTPO (Fig. 2a). The motif on the Spike protein is located on the surface of the N-terminal Domain (NTD) (Fig. 2b, c), whereas it is found at the interface with a neutralizing antibody in hTPO (Feese et al. 2004) (Fig. 2d). The TQLPP motifs from the two proteins exhibit high structural similarity (Fig. 2e, f). The motif on Spike is adjacent to the NTD supersite targeted by neutralizing antibodies (Cerutti et al. 2021). 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) (Tahara et al. 1998) 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 (Feese et al. 2004) with a published full-length glycosylated model of the Spike protein (Woo et al. 2020). The glycosylated model based on PDB id 6VSB (Wrapp et al. 2020) with the first 26 residues (including the TQLPP motif) reconstructed (Woo et al. 2020) 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. 3a, b) shows that the TQLPP epitope is accessible to the antibody and the adjacent glycan chains do not shield the antibody-binding site (Fig. 3c, Supplementary Fig. 1). 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 (Supplementary Fig. 2). 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. 3f, g). Both the Spike-TN1 and the hTPO-TN1 complexes showed similar contact areas (Supplementary Fig. 3). 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. 3d-g).

To further support our findings, we evaluated the antibody-antigen interface complementarity with MaSIF-search, a recent tool that uses deep learning techniques (Gainza et al. 2020), 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 (Cerutti et al. 2021) 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. 3h, Supplementary Tables 4, 5). Notably, in 7LQV, COVID-19 antibodies bind to Spike at the NTD supersite (Cerutti et al. 2021). These results strongly argue for the possibility of cross-reactivity between Spike and hTPO driven by the molecular mimicry of TQLPP (Fig. 3).

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 (Kanduc 2021). To evaluate structural mimicry between Spike-TQLPP and all human-TQLPP motifs, we utilized AlphaFold predictions (Jumper et al. 2021; Tunyasuvunakool et al. 2021) (Supplementary Fig. 4). The closest structural mimicry region is in 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 (Supplementary Fig. 4). These results indicate that TQLPP motifs have similar conformations (Supplementary Fig. 2), 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. 4). The binding score for NEK10-TN1 (1.44) is comparable to the hTPO-TN1 complex (Fig. 4). NEK10 regulates motile ciliary function responsible for expelling pathogens from the respiratory tract (Chivukula et al. 2020). NEK10 dysfunction can impact mucociliary clearance and lead to respiratory disorders such as bronchiectasis (Chivukula et al. 2020). 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.

Discussion

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(Li et al. 2020), particularly in severe and fatal cases (Voss et al. 2021) 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 (Stoddard et al. 2021). 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 (Stoddard et al. 2021). 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 (Yang et al. 2020; Helms et al. 2021; Greinacher et al. 2021; Schultz et al. 2021) 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 (Hodcroft 2021). 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 evolving variants.

Methods

Identifying epitopes with molecular mimicry

To identify known epitopes with positive assays, we used Epitopedia (Balbin et al. 2021) with a Cryo-EM structure of Spike from SARS-CoV-2 (PDB id 6XR8 (Cai et al. 2020)) as input. Hits containing 5 or more consecutive residues with 100% sequence identity where at least 3 of the input residues are surface accessible are considered sequence-based molecular mimics. For all sequence-based molecular mimics with corresponding structural representation, TM-align (Zhang and Skolnick 2005) was used to generate a structural alignment and Root Mean Square Deviation (RMSD) for all input-hit (sequence mimic) alignment pairs using only the structural regions corresponding to the hit for the source antigenic protein containing the epitope and the input. Epitopes with an RMSD $\leq$ 1Å to Spike were considered structure-based molecular mimics (Supplementary Table 1). 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 (Feese et al. 2004)). 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 (Supplementary Table 2).

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(Hwang et al. 2020)) retrieved from the CHARMM-GUI Archive (Choi et al. 2021). 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 (DeLano) and Visual Molecular Dynamics (VMD 1.9.3 (Humphrey et al. 1996)).

To confirm that the modeled Spike TQLPP region is in agreement with the TQLPP region of solved Spike structures, these regions were extracted. TM-align was used to

superimpose the TQLPP regions from the 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 Supplementary Table 2 and compared in an all-against-all manner (Supplementary Fig 2, Supplementary Table 3).

Molecular dynamics simulation

A simulation system for the modeled Spike-TN1 complex was prepared using CHARMM-GUI (Jo et al. 2008; Brooks et al. 2009; Lee et al. 2016). 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 (Phillips et al. 2005) using CHARMM36m force-field. The production runs were performed under constant pressure of 1 atm, controlled by a Nose−Hoover Langevin piston (Nosé and Klein 1983) 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) (Essmann et al. 1995) was used for long-range electrostatic interactions with periodic boundary conditions and all covalent bonds with hydrogen atoms were constrained by Shake (Ryckaert et al. 1977). 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 (Supplementary Fig. 3).

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 (Gainza et al. 2020). 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 (Sanner et al. 1996). In MaSIF-search, the properties include both geometric and physicochemical properties characterizing the protein surface (Gainza et al. 2020). 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 bound to 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 (Dunbar et al. 2014). 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

(Supplementary Table 4). 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 (Gainza et al. 2020). Columns "Contact AB" and "Contact AG" in Supplementary Table 4 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 (Jumper et al. 2021; Tunyasuvunakool et al. 2021), 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 (Supplementary Table 6).

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*. Following Shapiro-Wilk normality testing, statistical analysis comparing the different distributions was performed using Mann-Whitney U with *Scipy*. 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.

Data availability: All data are available in the supplementary materials.

Code availability: Software packages used are cited.

Acknowledgments: We thank Sathibalan Ponniah, Charles Dimitroff, Sixto Leal, and Kevin Bennett for discussions. We also thank Yuk-Ching Tse-Dinh and Tim Collins for 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).

Contributions:
J.S.-L., J.N.-C., G.N., P.C., K.M., T.C., A.M.M., P.B., and V.S. designed the overall method and approach. J.S.-L., G.N., P.C., A.M.M., K.M. supervised the research. J.N.-C., C.A.B., and J.S.-L. identified molecular mimicry. P.B. and P.C. performed modeling. G.N. and V.S. performed MaSIF-search. J.N.-C., C.A.B., J.S.-L., P.C, G.N., P.B., and V.S. analyzed the data. M.S., J.N.-C., C.A.B., V.S., and P.B. performed visualization. G.N. and J.S.-L. performed project administration. J.S.-L. and J.N.-C. acted as lead authors. C.A.B., G.N., P.C., K.M., T.C., A.M.M., P.B., and V.S. contributed to writing the manuscript. All authors read and commented the manuscript.

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

References

Audia S, Bonnotte B (2021) Emerging Therapies in Immune Thrombocytopenia. J Clin Med 10:1004. https://doi.org/10.3390/jcm10051004

Balbin CA, Nunez-Castilla J, Siltberg-Liberles J (2021) Epitopedia: identifying molecular mimicry of known immune epitopes. BioRxiv. https://doi.org/https://doi.org/10.1101/2021.08.26.457577

Brooks BR, Brooks CL, Mackerell AD, et al (2009) CHARMM: The biomolecular simulation program. J Comput Chem 30:1545–1614. https://doi.org/10.1002/jcc.21287

Cai Y, Zhang J, Xiao T, et al (2020) Distinct conformational states of SARS-CoV-2 spike protein. Science (80-) 369:1586–1592. https://doi.org/10.1126/science.abd4251

Cerutti G, Guo Y, Zhou T, et al (2021) Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe 29:819-833.e7. https://doi.org/10.1016/J.CHOM.2021.03.005

Chivukula RR, Montoro DT, Leung HM, et al (2020) A human ciliopathy reveals essential functions for NEK10 in airway mucociliary clearance. Nat Med 26:244–251. https://doi.org/10.1038/s41591-019-0730-x

Choi YK, Cao Y, Frank M, et al (2021) 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. https://doi.org/10.1021/acs.jctc.0c01144

DeLano WL The PyMOL Molecular Graphics System, Schrödinger, LLC. www.pymol.org

Dunbar J, Krawczyk K, Leem J, et al (2014) SAbDab: the structural antibody database. Nucleic Acids Res 42:D1140–D1146. https://doi.org/10.1093/NAR/GKT1043

Essmann U, Perera L, Berkowitz ML, et al (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593. https://doi.org/10.1063/1.470117

Feese MD, Tamada T, Kato Y, et al (2004) 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. https://doi.org/10.1073/pnas.0308530100

Gainza P, Sverrisson F, Monti F, et al (2020) Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. Nat Methods 17:184–192. https://doi.org/10.1038/s41592-019-0666-6

Getts DR, Chastain EM, Terry RL, Miller SD (2013) Virus infection, antiviral immunity, and autoimmunity. Immunol Rev 255:197–209. https://doi.org/10.1111/IMR.12091

Greinacher A, Thiele T, Warkentin TE, et al (2021) Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. N Engl J Med 384:2092–2101. https://doi.org/10.1056/nejmoa2104840
Helms JM, Ansteatt KT, Roberts JC, et al (2021) Severe, refractory immune thrombocytopenia occurring after sars-cov-2 vaccine. J Blood Med 12:221–224. https://doi.org/10.2147/JBM.S307047
Hodcroft EB (2021) CoVariants: SARS-CoV-2 Mutations and Variants of Interest. https://covariants.org/. Accessed 6 Aug 2021
Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:33–38. https://doi.org/10.1016/0263-7855(96)00018-5
Hwang SS, Lim J, Yu Z, et al (2020) MRNA destabilization by BTG1 and BTG2 maintains T cell quiescence. Science (80-) 367:1255–1260. https://doi.org/10.1126/science.abb2507
Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865. https://doi.org/10.1002/jcc.20945
Jumper J, Evans R, Pritzel A, et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 1–11. https://doi.org/10.1038/s41586-021-03819-2
Kanduc D (2021) Thromboses and Hemostasis Disorders Associated with Coronavirus Disease 2019: The Possible Causal Role of Cross-Reactivity and Immunological Imprinting. Glob Med Genet. https://doi.org/10.1055/s-0041-1731068
Kanehisa M, Furumichi M, Sato Y, et al (2021) KEGG: integrating viruses and cellular organisms. Nucleic Acids Res 49:D545–D551. https://doi.org/10.1093/NAR/GKAA970
Kuter DJ (2013) The biology of thrombopoietin and thrombopoietin receptor agonists. Int J Hematol 98:10–23. https://doi.org/10.1007/s12185-013-1382-0
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. https://doi.org/10.1021/acs.jctc.5b00935
Li Y, Lai D, Zhang H, et al (2020) Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. Cell Mol Immunol 2020 1710 17:1095–1097. https://doi.org/10.1038/s41423-020-00523-5
Nazy I, Kelton JG, Moore JC, et al (2018) Autoantibodies to thrombopoietin and the thrombopoietin receptor in patients with immune thrombocytopenia. Br J Haematol 181:234–241. https://doi.org/10.1111/bjh.15165
Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. Mol Phys 50:1055–1076. https://doi.org/10.1080/00268978300102851
Phillips JC, Braun R, Wang W, et al (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802. https://doi.org/10.1002/jcc.20289
Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341. https://doi.org/10.1016/0021-9991(77)90098-5
Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces - PubMed. Biopolymers 38:305–320
Schultz NH, Sørvoll IH, Michelsen AE, et al (2021) Thrombosis and Thrombocytopenia after ChAdOx1 nCoV-19 Vaccination. N Engl J Med 384:2124–2130. https://doi.org/10.1056/nejmoa2104882
Shi C, Wang L, Ye J, et al (2021) Predictors of mortality in patients with coronavirus disease 2019: a systematic review and meta-analysis. BMC Infect Dis 21:663. https://doi.org/10.1186/s12879-021-06369-0

Stoddard CI, Galloway J, Chu HY, et al (2021) Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs. Cell Rep 35:. https://doi.org/10.1016/j.celrep.2021.109164
Tahara T, Kuwaki T, Matsumoto A, et al (1998) Neutralization of biological activity and inhibition of receptor binding by antibodies against human thrombopoietin. Stem Cells 16:54–60. https://doi.org/10.1002/stem.160054
Taylor WR (1997) Residual colours: a proposal for aminochromography. Protein Eng 10:743–746
Tunyasuvunakool K, Adler J, Wu Z, et al (2021) Highly accurate protein structure prediction for the human proteome. Nature 1–9. https://doi.org/10.1038/s41586-021-03828-1
Varghese LN, Defour JP, Pecquet C, Constantinescu SN (2017) The thrombopoietin receptor: Structural basis of traffic and activation by ligand, mutations, agonists, and mutated calreticulin. Front Endocrinol (Lausanne) 8:1. https://doi.org/10.3389/fendo.2017.00059
Vita R, Mahajan S, Overton JA, et al (2019) The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res 47:D339–D343. https://doi.org/10.1093/nar/gky1006
Voss C, Esmail S, Liu X, et al (2021) Epitope-specific antibody responses differentiate COVID-19 outcomes and variants of concern. JCI Insight 6:. https://doi.org/10.1172/jci.insight.148855
Waterhouse AM, Procter JB, Martin DMA, et al (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033
Watts A, Raj K, Gogia P, et al (2021) Secondary Immune Thrombocytopenic Purpura Triggered by COVID-19. Cureus 13:. https://doi.org/10.7759/cureus.14505
Woo H, Park SJ, Choi YK, et al (2020) Developing a fully glycosylated full-length SARS-COV-2 spike protein model in a viral membrane. J Phys Chem B 124:7128–7137. https://doi.org/10.1021/acs.jpcb.0c04553
Wrapp D, Wang N, Corbett KS, et al (2020) Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 367:1255–1260. https://doi.org/10.1126/SCIENCE.ABB2507
Yang X, Yang Q, Wang Y, et al (2020) Thrombocytopenia and its association with mortality in patients with COVID-19. J Thromb Haemost 18:1469–1472. https://doi.org/10.1111/jth.14848
Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309. https://doi.org/10.1093/NAR/GKI524

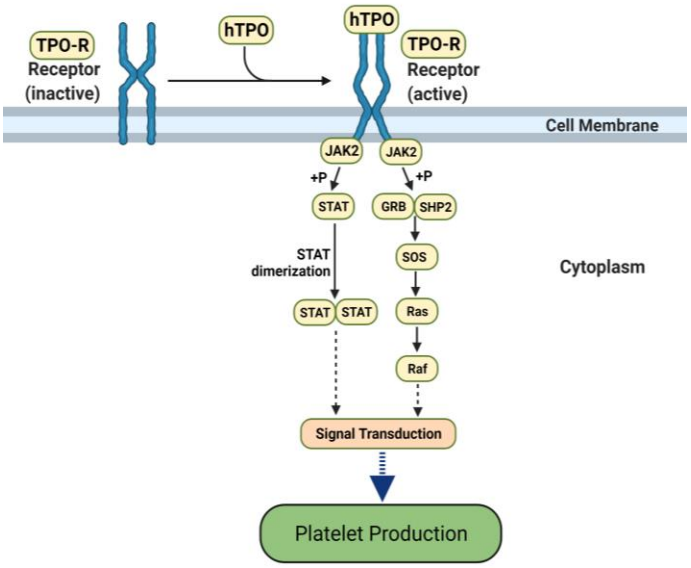


Fig. 1. The hTPO pathway to induce platelet production. Simplified JAK-STAT signaling pathway in megakaryocytes where hTPO activates the TPO receptor and triggers signaling cascades that stimulate platelet production(Kuter 2013; Kanehisa et al. 2021)**.** Created with BioRender.com.

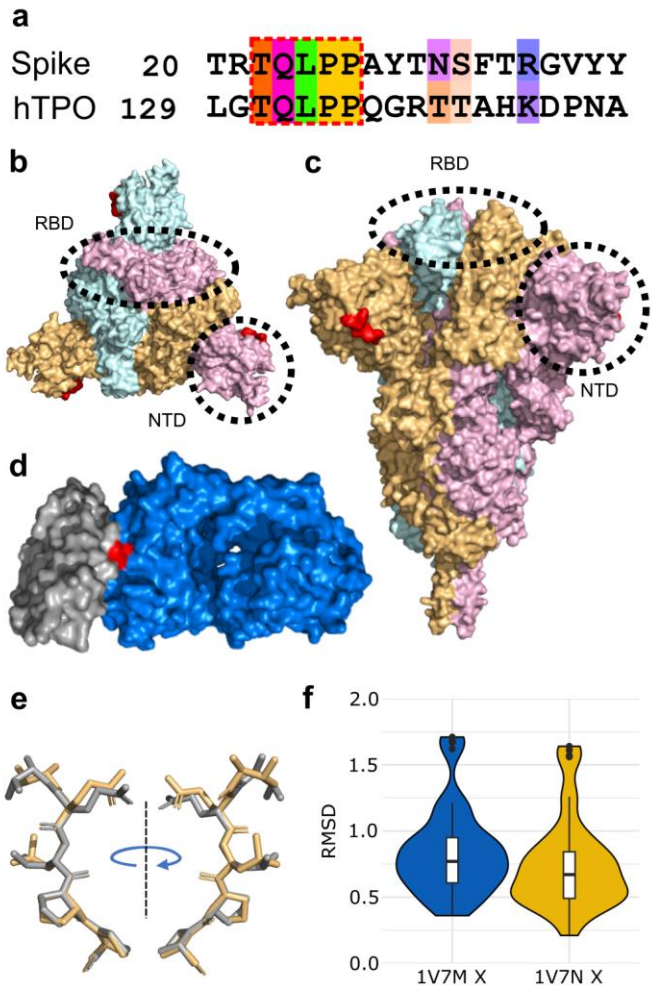


Fig. 2. Structural mimicry between a TQLPP motif in SARS-CoV-2 Spike and an antibody binding epitope in thrombopoietin. a, Pairwise alignment for TQLPP motif in the epitope for human thrombopoietin (hTPO, Epitope ID: 920946) and Spike, colored by Taylor(Taylor 1997) for sites with ≥50% conservation in the amino acid property(Waterhouse et al. 2009). The region of molecular mimicry is highlighted in the red dashed box. **b, c,** Surface representation of Spike from (**b**) above and (**c**) the side, with Spike trimer (PDB id: 6XR8(Cai et al. 2020)) colored by subunit and red indicating the location of the TQLPP epitope fragment, illustrating the surface accessibility of TQLPP. **d**, Surface representation shown for hTPO (gray, PDB id: 1V7M(Feese et al. 2004)) and its TN1 antibody (blue) with the TQLPP motif (red) at the interface. **e**, TM-align generated structural alignment for TQLPP in Spike (beige) and hTPO (gray), with RMSD = 0.61 Å. **f**, Violin plots of RMSD values resulting from the comparison of the TQLPP region in 60 Spike structures vs TQLPP in two hTPO structures (PDB ids: 1V7M and 1V7N, chain X for both(Feese et al. 2004)). Statistical analysis with Mann-Whitney U reveals no statistical significance between the sets. Box plots show median value (horizontal solid bold line) and bounded by 1st and 3rd quartiles, vertical lines (whiskers) represent 1.5 × IQR interquartile range, while outliers are marked as black points. For further details, see supplementary methods.

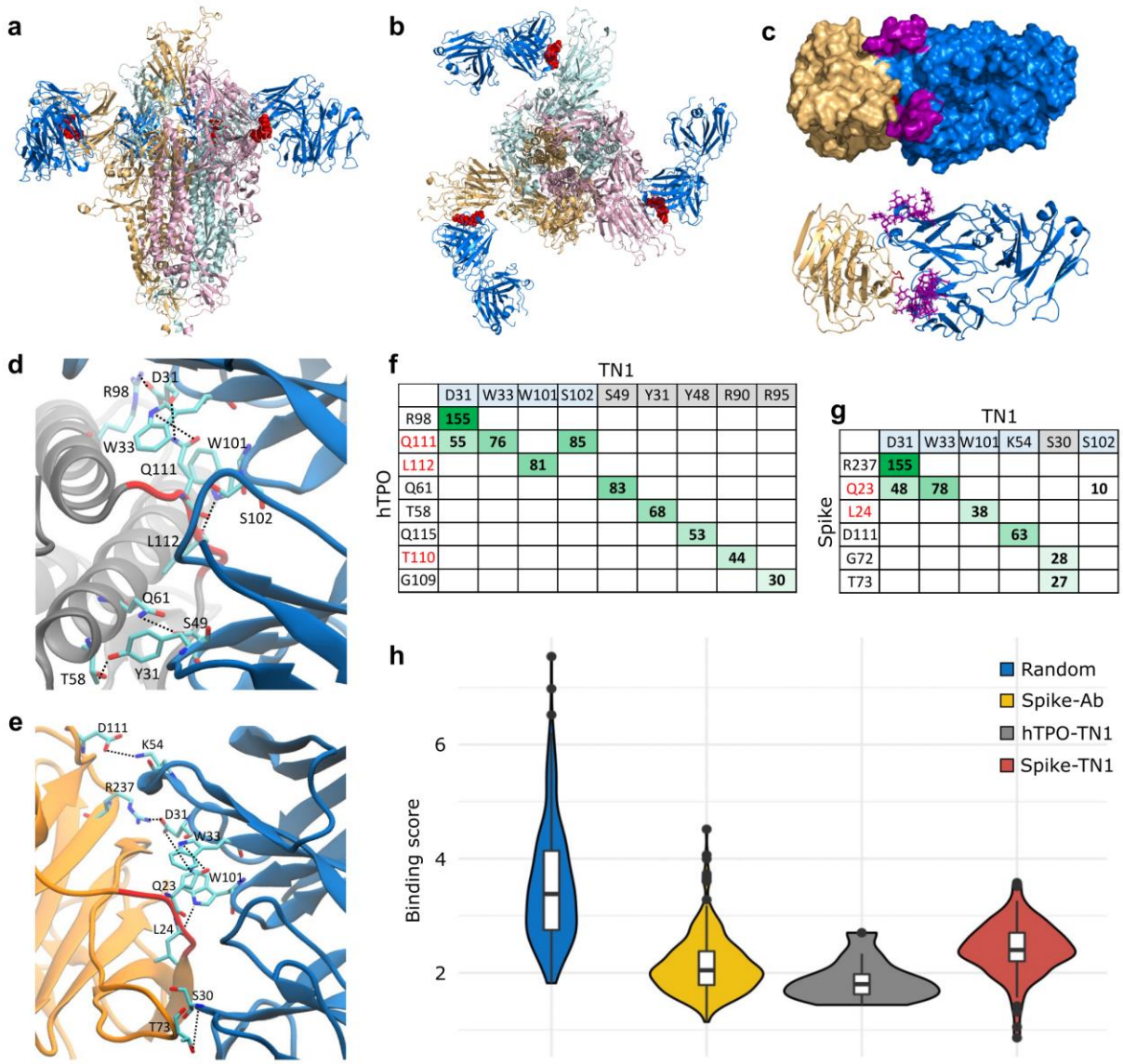


Fig. 3. Binding of SARS-CoV-2 Spike to TN1 Fab antibody. a, **b**, Equilibrated structure (1 ns) of the modeled TN1 Fab antibody (blue, PDB id: 1V7M) complexed with Spike trimer model shown from (**a**) the side and (**b**) the top, with TQLPP shown as red spheres. **c**, The Spike NTD (beige) and TN1 Fab complex were used for MD simulations (200 ns), with adjacent glycans at N17 and N74 highlighted in purple. **d**, **e**, 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. **f**, **g**, 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 (Supplementary Table

5). Box plots show median value (horizontal solid bold line) and bounded by 1st and 3rd quartiles, vertical lines (whiskers) represent 1.5 × IQR interquartile range, while outliers are marked as black points. For further details, see supplementary methods.

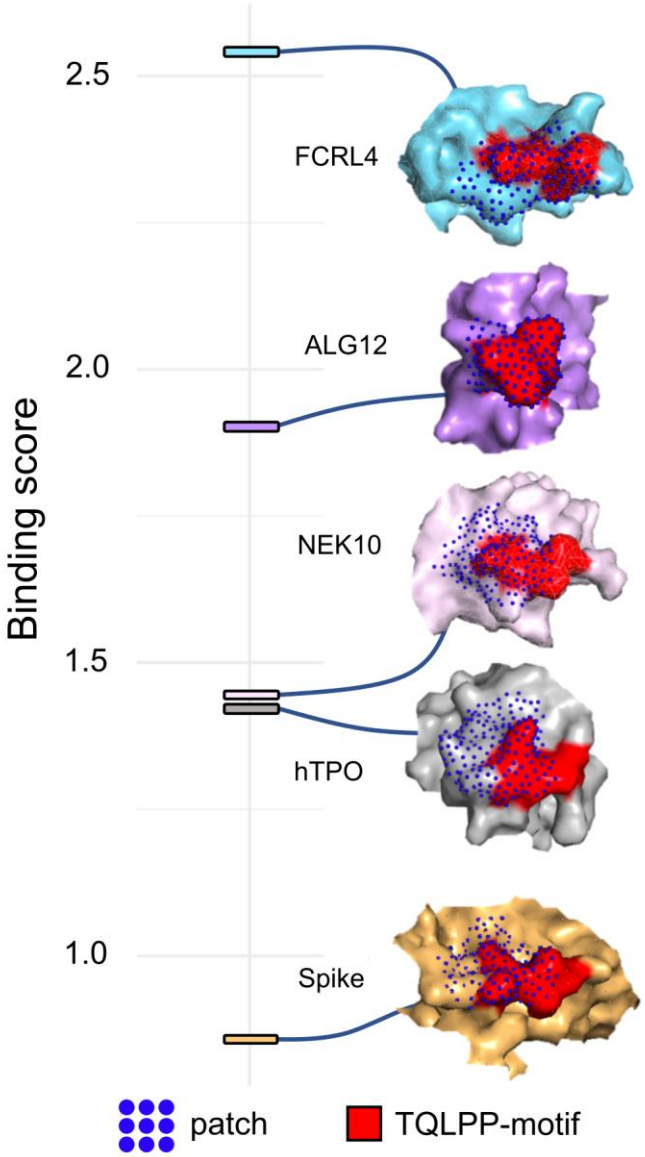
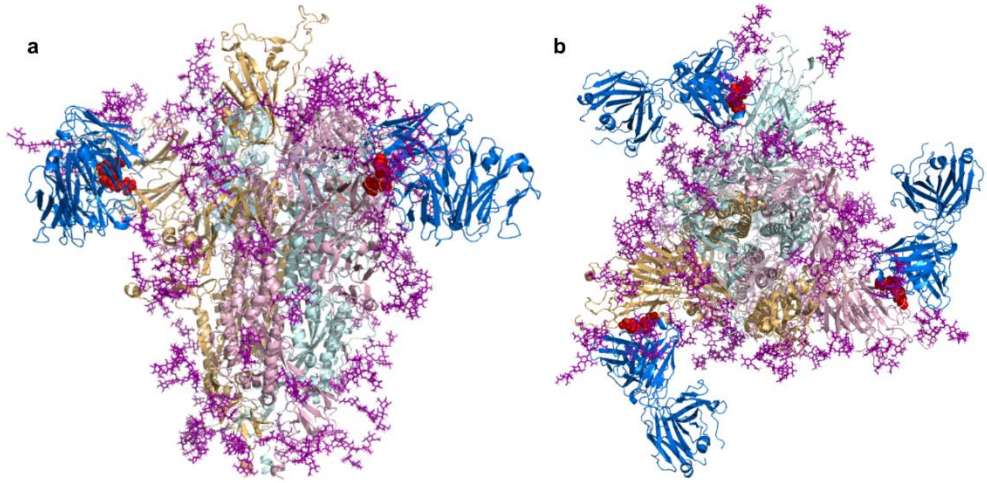
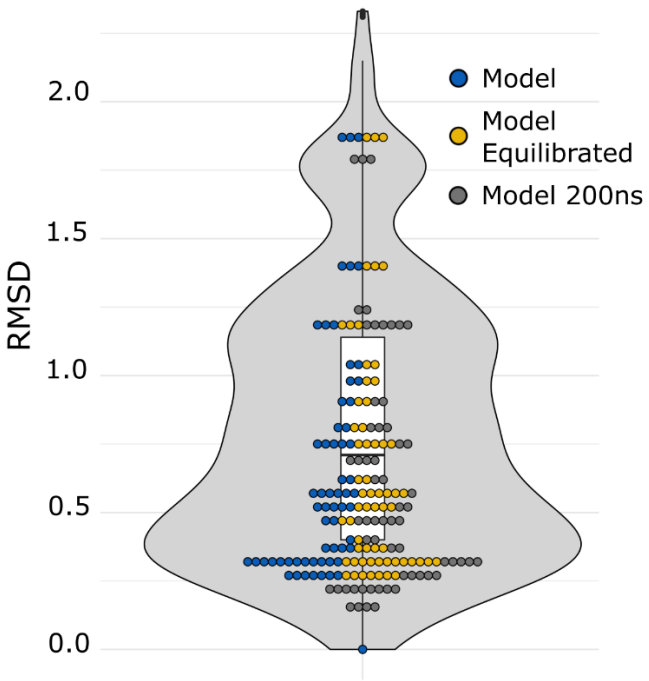


Fig. 4. Predicted interaction patches between TN1 Fab antibody (PDB id: 1V7N) and the TQLPP motif. The best (lowest) binding score is shown for Spike (PDB id: 7LQV, chain A), hTPO (PDB id: 1V7N, chain X), NEK10 (Uniprot: Q6ZWH5), ALG12 (Uniprot: Q9BV10), and FCRL4 (Uniprot: Q96PJ5). For all, red indicates the TQLPP motif and blue dots represent the surface points included in the predicted MaSIF patches.

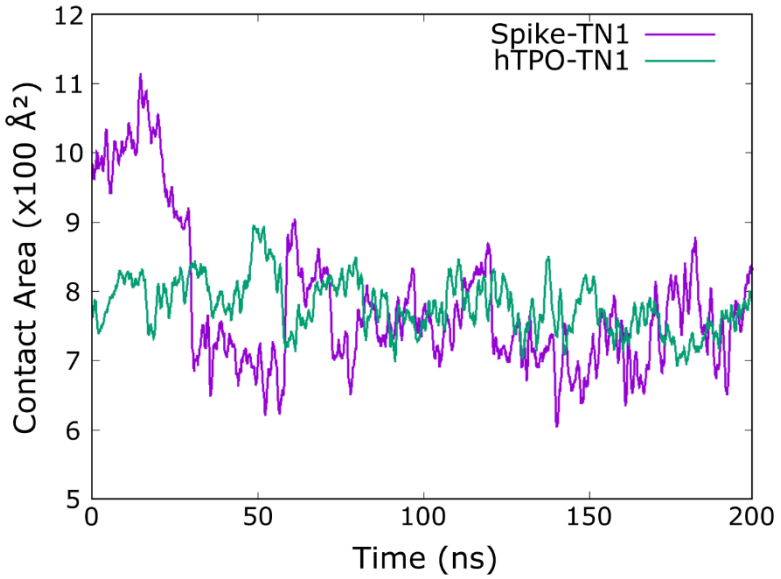
Supplementary materials



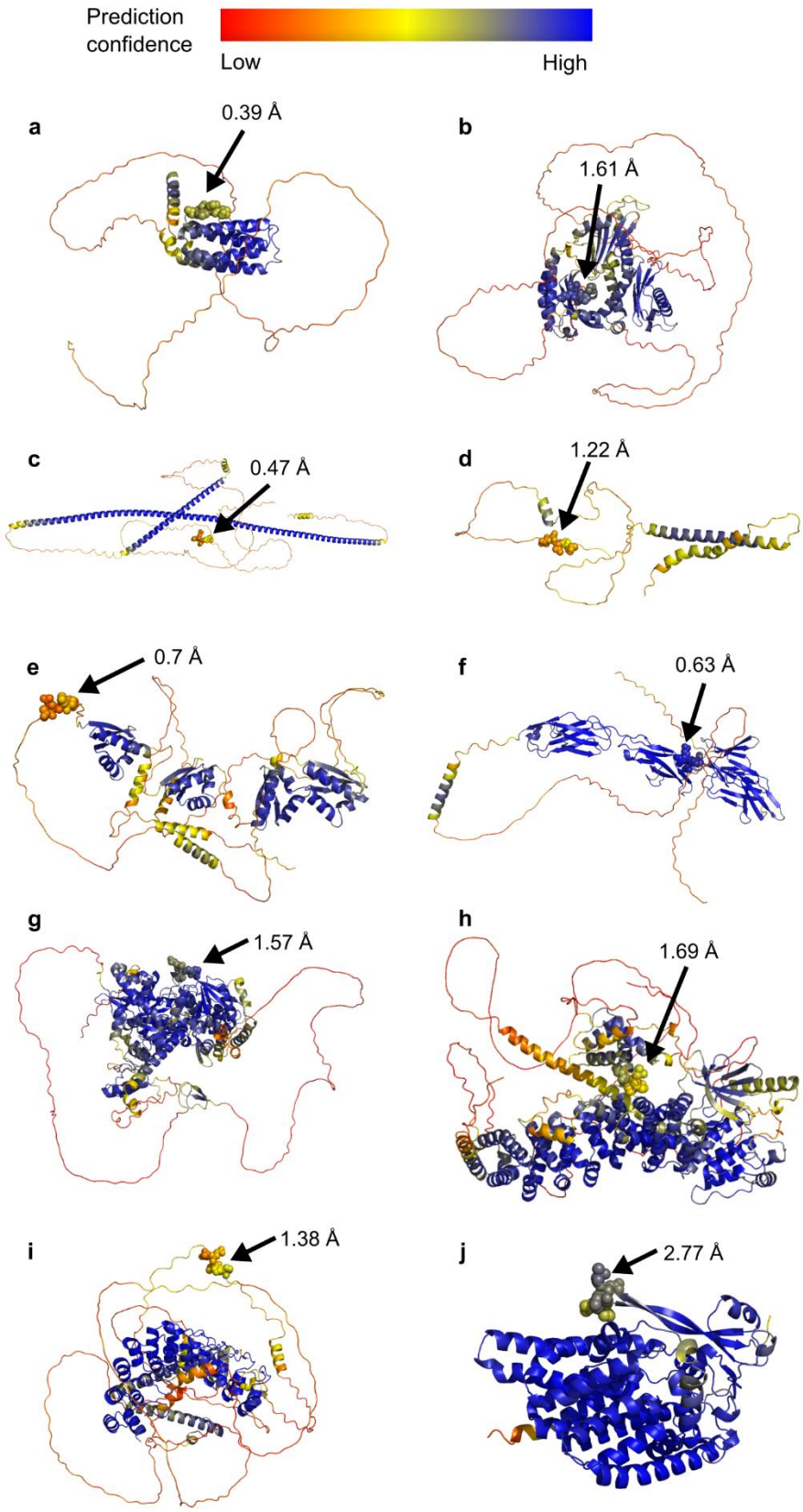
Supplementary Fig. 1. SARS-CoV-2 Spike bound to TN1 Fab antibody. a, **b**, SARS-CoV-2 Spike shown in the trimeric state (PDB id: 6VSB) bound to TN1 Fab antibody (blue, PDB id: 1V7M) as viewed from (**a**) the side and (**b**) the top. The TQLPP motifs are shown as red spheres and glycans are shown in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



Supplementary Fig. 2. 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).



Supplementary Fig. 3. 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.



Supplementary Fig. 4. TQLPP motif for 10 human proteins modeled by AlphaFold. a-j, 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).

Supplementary Table 1. Structure-based molecular mimics between Spike from SARS-CoV-2 and known epitopes with RMSD ≤ 1Å

Motif	Epitope id	Species	Protein	PDB id, chain	RMSD (Å)	Range (Spike)	Range (native)
TQLPP	920946	*Homo sapiens*	Thrombopoietin	1V7N_X	0.46	22-26	131-135
QLPPA	469522	*Homo sapiens*	SMYD3 protein	5XXD_A	0.43	23-27	164-168
KNLRE	1217238	*Homo sapiens*	Toll-like receptor 8	6WML_D	0.87	187-191	125-129
FTVEKG	127202	*Phleum pratense*	Pollen allergen Phl p 2	1WHP_A	0.76	306-311	32-37
GEVFN	926203	*Homo sapiens*	Integrin beta-1	3VI4_D	0.74	339-343	229-233
HAPAT	449143	*Homo sapiens*	Activator of 90 kDa heat shock protein ATPase homolog 1	7DME_A	0.74	519-523	237-241
EHVNN	428030	*Homo sapiens*	Casein kinase 2	2ZJW_A	0.29	654-658	114-118
NLLLQ	1124368	*Homo sapiens*	DNA polymerase subunit gamma-1	5C51_A	0.57	751-755	155-159
LLQYG	1279071	*Homo sapiens*	Ankyrin-1	1N11_A	0.20	753-757	621-625
LPDPS	221943	*Homo sapiens*	BRCA1-A complex subunit BRE	6VGW_C	0.32	806-810	93-97
LPDPS	1108606	*Homo sapiens*	Semaphorin-7A	3NVQ_A	0.84	806-810	309-313
DPSKP	1214157	*Homo sapiens*	60S ribosomal protein L3	6LU8_B	0.10	808-812	36-40
DPSKP	30850	*Mycobacterium tuberculosis*	Alanine and proline-rich secreted protein apa precursor	5ZXA_A	0.21	808-812	230-234
IARRD	531561	*Mus musculus*	Talin	6R9T_A	0.74	844-848	1119-1123
GNCDV	1272132	*Homo sapiens*	Tryptophan-tRNA ligase	1O5T_A	0.91	1124-1128	392-396
ELDKY	96559	Human respiratory syncytial virus	Fusion glycoprotein F0	6EAE_F	0.12	1151-1155	82-86
DKYFK	1230294	*Homo sapiens*	Cytoplasmic FMR1-interacting protein 1	4N78_A	0.14	1153-1157	296-300

Supplementary Table 2. 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

Supplementary Table 3. 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.

Supplementary Table 4. Distribution of MaSIF values.

Supplementary Table 5. 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

Supplementary Table 6. 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