Vitamin D3 and its Hydroxyderivatives as Promising Drugs against COVID-19: A Computational Study

Yuwei Song[1], Shariq Qayyum[1], Rory A Greer[2], Radomir M Slominski[1,3], Chander Raman[1,3], Andrzej T Slominski[1,4,5,*], Yuhua Song[2,*]

[1]Department of Dermatology, [2]Department of Biomedical Engineering, [3]Department of Medicine and Microbiology, Division of Clinical Immunology and Rheumatology, [4]Comprehensive Cancer Center, Cancer Chemoprevention Program, University of Alabama at Birmingham, Birmingham, AL 35249. USA. [5]Pathology and Laboratory Medicine Service, VA Medical Center, Birmingham, AL 35249. USA.

Supplemental Materials

*To whom correspondence should be addressed:
Andrzej T Slominski, PhD, MD.
Department of Dermatology
The University of Alabama at Birmingham
VH 476C, Volker Hall
670 University Blvd, Birmingham, AL 35294-0019
Phones: (205)934-5245; Fax: (205)996-0302
E-mail: aslominski@uabmc.edu

Yuhua Song, PhD
Department of Biomedical Engineering
The University of Alabama at Birmingham
803 Shelby Interdisciplinary Biomedical Research Building
1825 University Blvd, Birmingham, AL, 35294
Phone: (205)996-6939; Fax: (205)975-4919
E-mail: yhsong@uab.edu

TMPRSS2 Structure

Since there was not a structure available on the protein data bank for TMPRSS2 to use for molecular docking, the MODELLER[1] was used to generate a structure for TMPRSS2 based on the crystal structure of TMPRSS1 (PDBID: 1Z8G[2]). TMPRSS1 is a homologous protein of TMPRSS2.
T-Coffee[3-6] was used to generate an alignment between TMPRSS2 and TMPRSS1. Figure 1 displays T-Coffee's alignment.

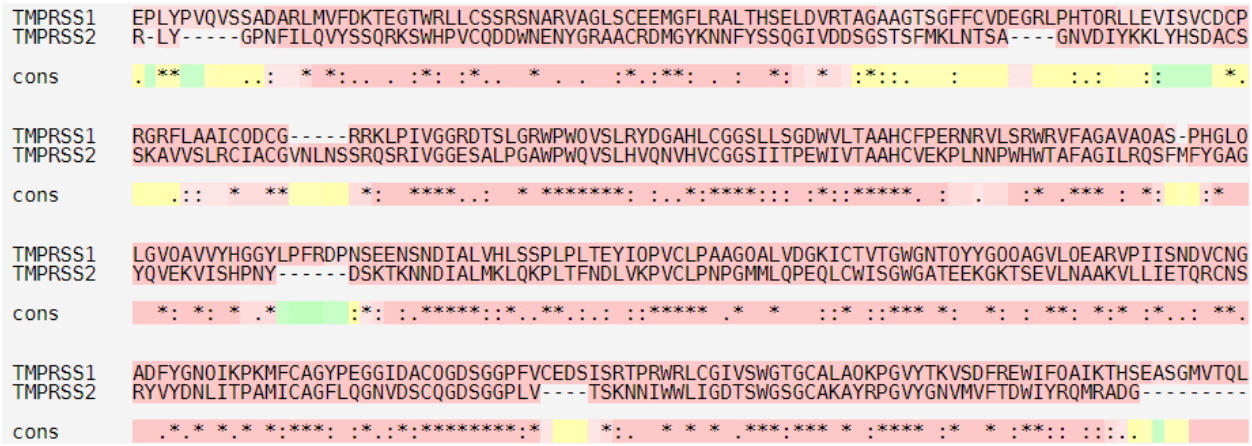

Figure S1. Alignment of TMPRSS2 and TMPRSS1 using T-Coffee.

T-Coffee's alignment resulted in a 31.7% identity match and a 51.3% similarity between TMPRSS1 and TMPRSS2. T-Coffee's alignment closely matched MODELLER's alignment.

After finding the alignment between TMPRSS2 and TMPRSS1, MODELLER[1] was used to predict the structure of TMPRSS2. Using the template TMPRSS1, five predicted TMPRSS2 structures were created. The structure with the highest ERRAT[7] score can be found in **Figure 2** in the main text, which was generated by PyMol[8].

Quality of Predicted Structure

After MODELLER was used to construct the structure for TMPRSS2, ERRAT[7], PROCHECK[9, 10], and WHATCHECK[11] were used to evaluate the quality of constructed structure of TMPRSS2 and compare with those for template structure of TMPRSS1, which were shown as **Table S1**.

Table S1. Evaluation results for constructed TMPRSS2 structure with ERRAT, PROCHECK, and WHATCHECK programs

Sequence ID	ERRAT (Overall Quality Factor)	PROCHECK			
		Favored Residues	Allowed Residues	Generously Allowed Residues	Disallowed Residues
TMPRSS2	59.79%	80.7%	15.9%	2.0%	1.4%
Template (TMPRSS1)	99.60%	86.5%	13.2%	0.3%	0.0%
		WHATCHECK			
		Packing Quality	Ramachandran Appearance	X_1/X_2 Rotamer Normality	Backbone conformation
TMPRSS2		-2.215	-2.401	-1.057	-22.846
Template (TMPRSS1)		-0.411	-1.846	-0.213	-24.189

Data presented in **Table S1** was used to evaluate the quality of the predicted TMPRSS2 structure through three programs: ERRAT, PROCHECK, and WHATCHECK. ERRAT uses statistical information found from real proteins to analyze the statistics of non-bonded interactions between the carbon, oxygen, and nitrogen atoms.[12] For the ERRAT test, the higher the percent value the greater the reliability of the predicted structure. In general, the accepted range for a high quality model is greater than 50%.[13, 14] The PROCHECK test splits all the residues in the predicted structure into four groups, favored residues, allowed residues, generously allowed residues, and disallowed residues. The higher percentage of residues falling in the favored and allowed regions relates to a higher structural quality. The percent for the different groups are based off of a Ramachandran plot, and ideally a structure with a resolution of at least 2 angstroms and an R-factor less than 20% would be expected to have over 90% of the residues in the favored residues section.[15] WHATCHECK was used to analyze the packing quality, Ramachandran appearance, X_1/X_2 rotamer normality, and Backbone conformation. The packing quality is a control test that tests the grade of residues as well as the molecule as a whole. A molecule with a packing quality below -2.7 is considered wrong, a number below -2.0 may be misfolded or poorly refined, and a number below -1.2 does not belong in a reliable structure database.[16] The Ramachandran appearance expresses

well how the backbone conformations of the residues are corresponding to known allowed areas in the Ramachandran plot as a z-score.[16] The X_1/X_2 rotamer normality compares residues X_1 rotamer with an observed rotamer distribution for residues of the same type in a similar local backbone conformation in the database.[16] The backbone conformation checks if there are fragments of 5 alpha carbon coordinates not represented by other structures in the WHAT IF database. This is calculated as a z-score.[16] The results of these tests are shown as comparisons, with zero as average, below zero as below average, and above zero as above average.

The predicted structure of TMPRSS2 had an ERRAT score greater than 50% and PROCHECK displayed over 96% of the residues remained in the favored and allowed region. While TMPRSS2 had a packing quality below -2, it was not below -2.7 and therefore there may be small problems with the structure, but the packing quality as a whole passed. WHATCHECK did recognize two sequences with a questionable packing environment: GLN^{164} to LYS^{166} and MET^{371} to LEU^{373}. Neither of those two sequences are in the active site of TMPRSS2. The Ramachandran appearance and X_1/X_2 rotamer normality both had values within the expected range for TMPRSS2 and TMPRSS1. TMPRSS2 and TMPRSS1 both failed the backbone conformation test due to the backbone conformation comparison with the database proteins appearing unusual. Based on the three quality tests and a comparison between the results of TMPRSS2 and TMPRSS1, the predicted structure of TMPRSS2 is acceptable.

Table S2. Ramachandran plot analyses of the structures of ACE2, RBD for SARS-CoV-2 and TMPRSS2 with and without binding with 1,25(OH)$_2$D3 with PROCHECK.

Sequence ID	PROCHECK (Ramachandran plot)			
	Favored Residues	Allowed Residues	Generously Allowed Residues	Disallowed Residues
ACE2	91.4%	8.0%	0.4%	0.2%
ACE2 binding with 1α,25(OH)2D3	89.2%	10.4%	0.0%	0.4%
SARS-CoV-2 RBD	74.7%	24.7%	0.0%	0.6%
SARS-CoV-2 RBD binding with 1α,25(OH)2D3	75.3%	21.5%	2.5%	0.6%
TMPRSS2	78.7%	18.2%	2.4%	0.7%
TMPRSS2 binding with 1α,25(OH)2D3	78.7%	18.9%	1.7%	0.7%

*Based on the analyses of representative structure of each receptor from the equilibrated MD simulation trajectories.

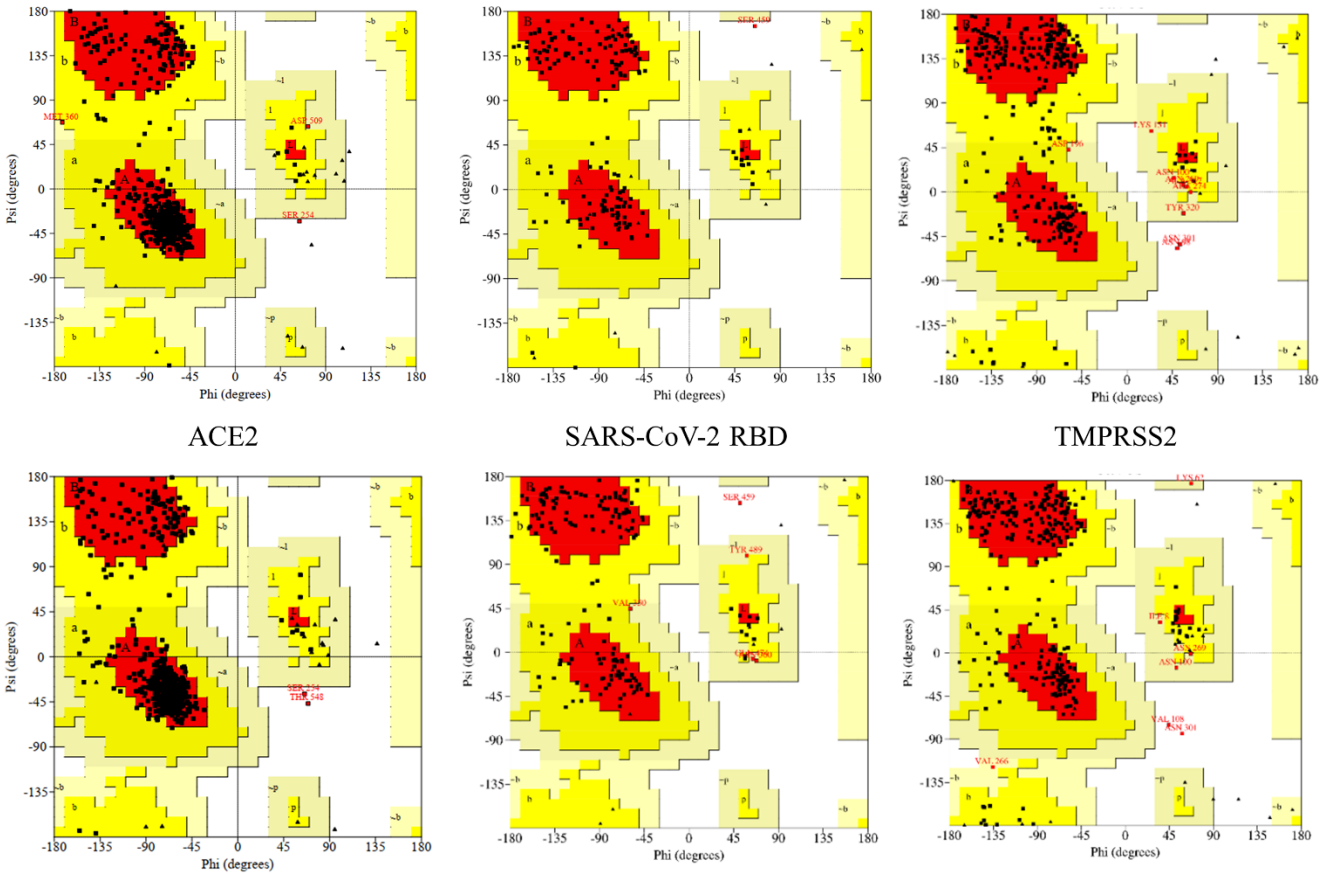

Figure S2 Ramachandran plots of ACE2, RBD for SARS-Cov-2 and TMPRSS2 with and without binding 1,25(OH)$_2$D3 (from the representative structure from equilibrated MD simulation trajectories). Region A& a: Right-handed alpha-helix; Region B& b: Beta-sheet; Region L& l: Left-handed alpha-helix; red, dark yellow and yellow regions represent the favored, allowed, and generously allowed region as defined by PROCHECK. The images were made with UCLA-DOE LAB — SAVES v6.0 web version: https://saves.mbi.ucla.edu/)

Secondary structure analyses of ACE2, RBD for SARS-Cov-2 and TMPRSS2 with and without binding with 1,25(OH)$_2$D3

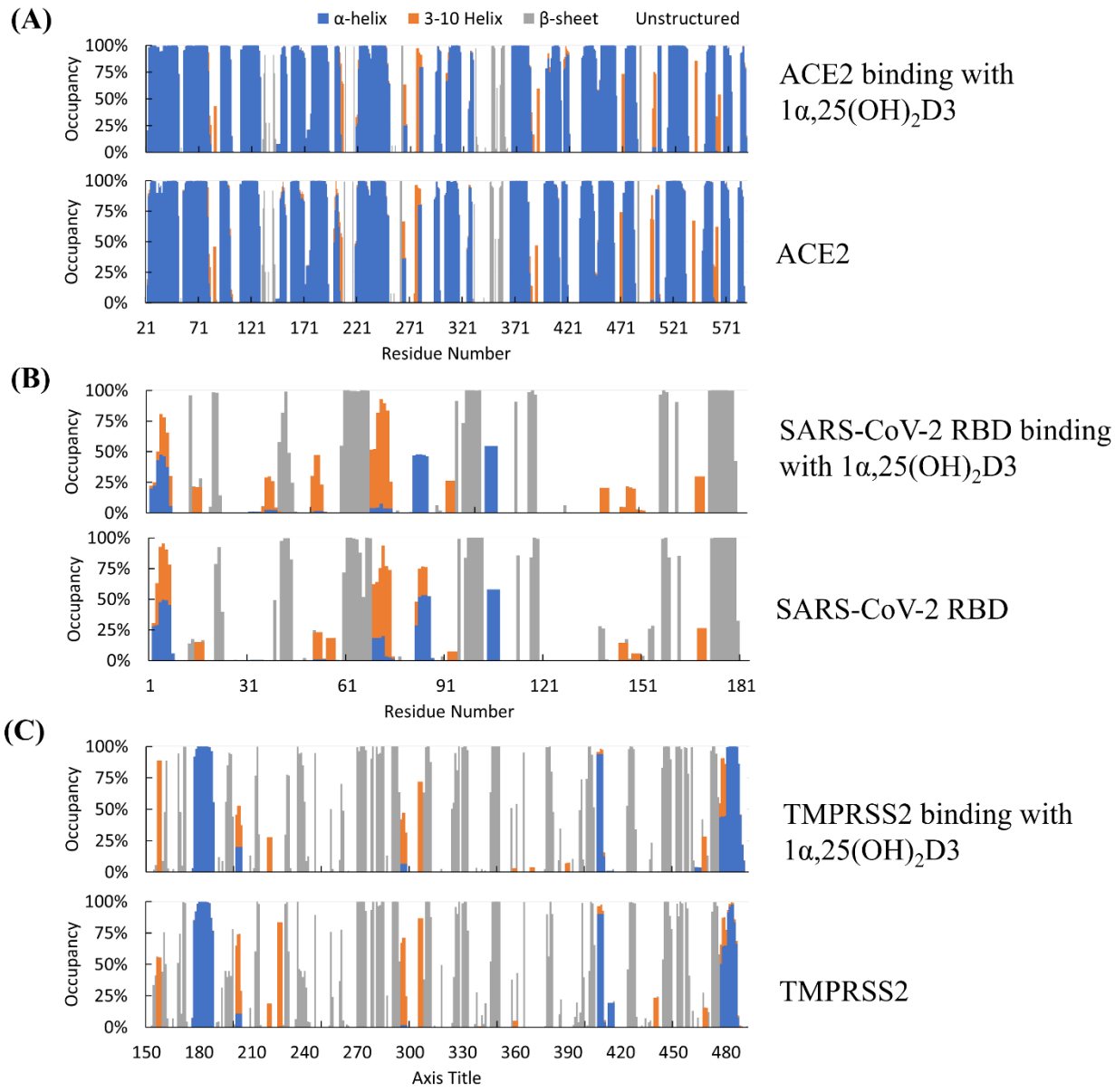

Figure S3. Secondary structure analyses of ACE2 (**A**), RBD for SARS-Cov-2 (**B**) and TMPRSS2 (**C**) with and without binding 1α,25(OH)$_2$D3 (1,25(OH)$_2$D3)

Principal component analysis (PCA) analysis

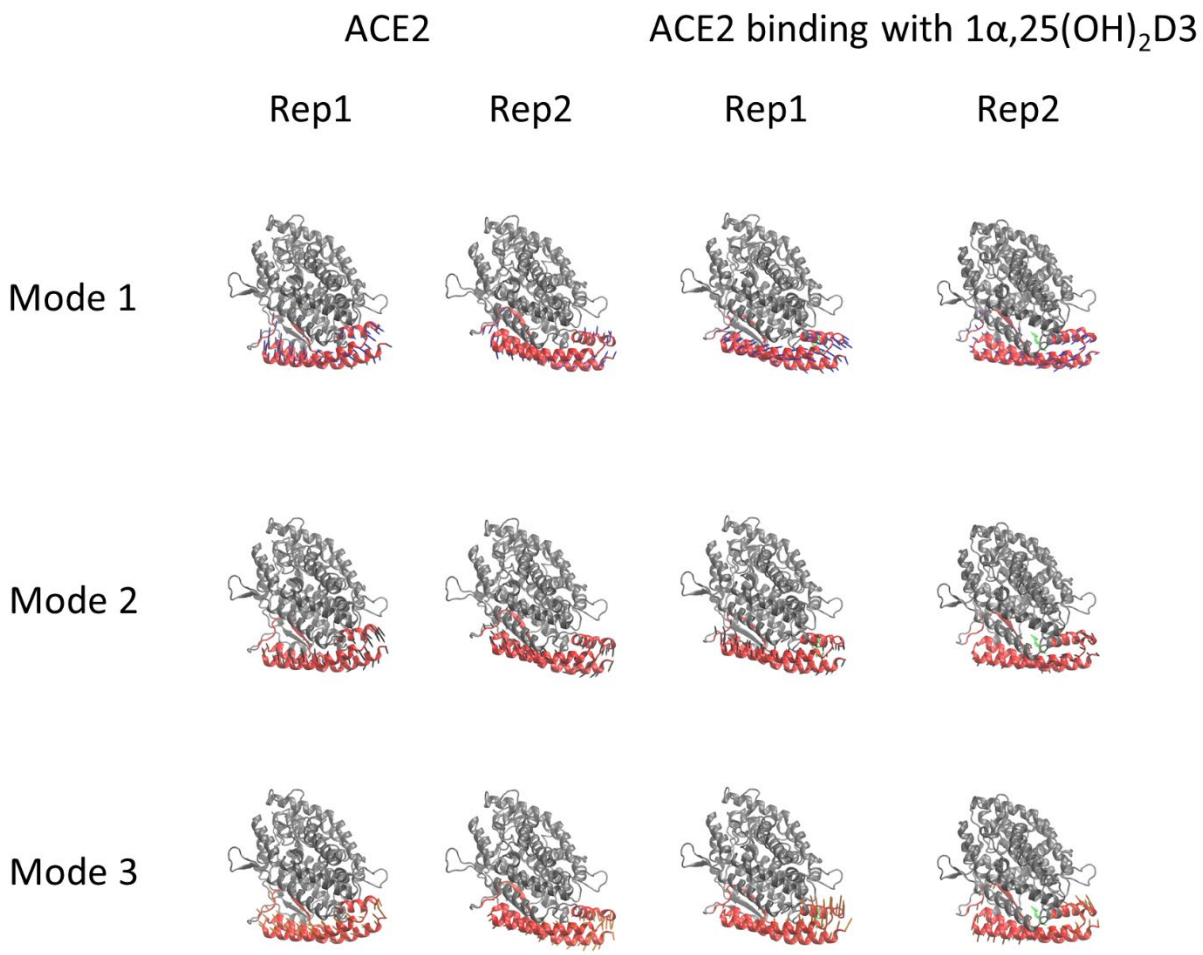

Figure S4. The first, second and third PCA modes of ACE2 binding region for RBD of SARS-CoV-2 with and without binding 1α,25(OH)$_2$D3 (1,25(OH)$_2$D3). ACE2 is shown as new cartoon in grey and its binding region for RBD of SARS-CoV-2 is shown in red. 1,25(OH)$_2$D3 is shown as licorice in green. Arrows of principal dynamic motion are shown as blue, grey and orange separately for the first, second and third PCA modes. The images were made with VMD (v1.9.2, http://www.ks.uiuc.edu/Research/vmd/)[17]

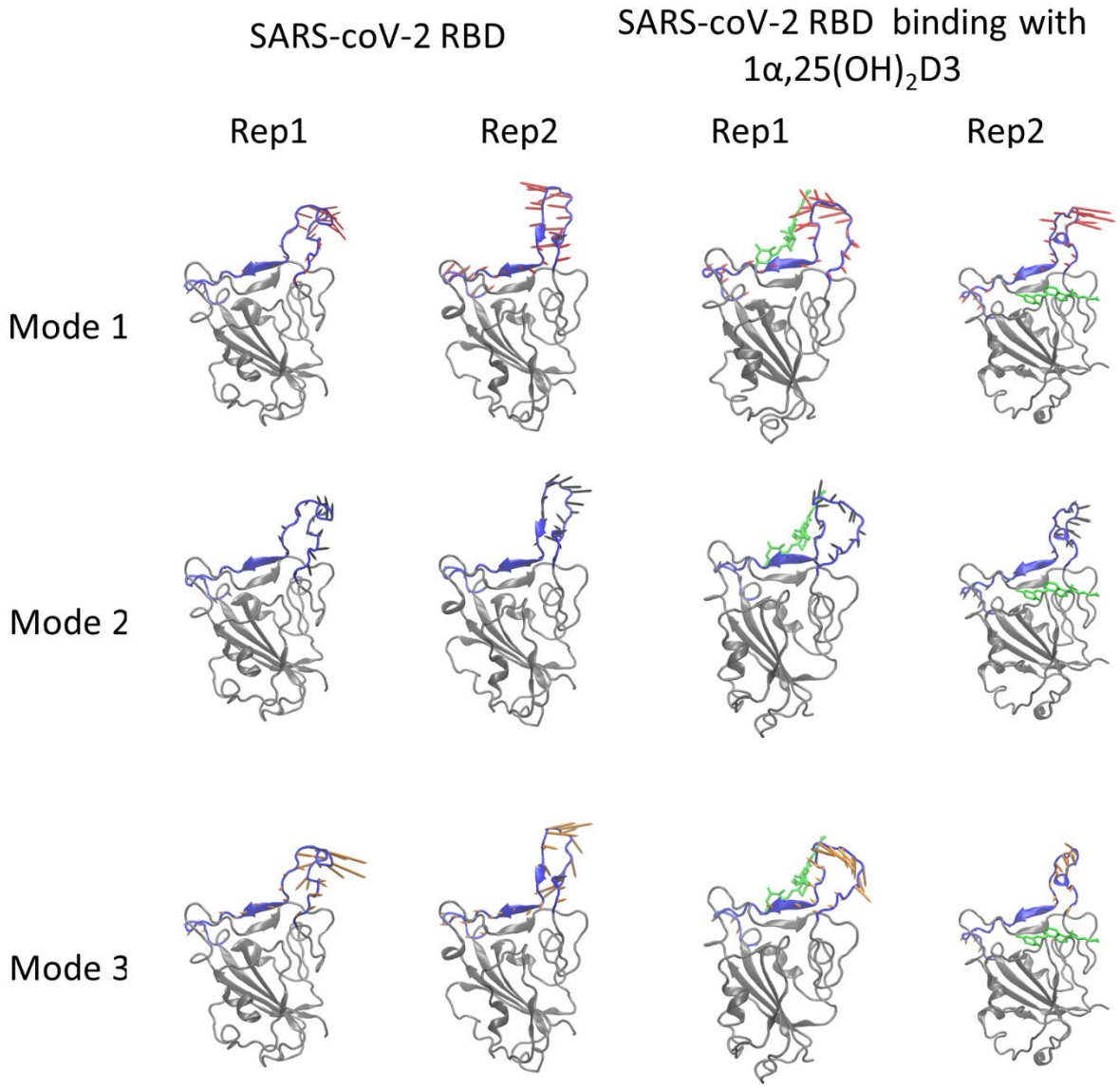

Figure S5. The first, second and third PCA modes of SARS-CoV-2 RBD binding region for ACE2. SARS-CoV-2 RBD are shown as new cartoon in grey and its binding region for ACE2 is shown in blue. 1α,25(OH)2D3 (1,25(OH)2D3) is shown as licorice in green. Arrows of principal dynamic motion shown are as red, grey and orange separately for the first, second and third PCA modes. The images were made with VMD (v1.9.2, http://www.ks.uiuc.edu/Research/vmd/)[17]

REFERENCES

[1] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller, *Curr Protoc Bioinformatics Chapter 5*, Unit-5.6.

[2] Herter, S., Piper, D. E., Aaron, W., Gabriele, T., Cutler, G., Cao, P., Bhatt, A. S., Choe, Y., Craik, C. S., Walker, N., Meininger, D., Hoey, T., and Austin, R. J. (2005) Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers, *Biochem J 390*, 125-136.

[3] Chang, J. M., Di Tommaso, P., and Notredame, C. (2014) TCS: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction, *Mol Biol Evol 31*, 1625-1637.

[4] Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, *Nucleic Acids Res 39*, W13-17.

[5] Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: A novel method for fast and accurate multiple sequence alignment, *J Mol Biol 302*, 205-217.

[6] Chang, J. M., Di Tommaso, P., Lefort, V., Gascuel, O., and Notredame, C. (2015) TCS: a web server for multiple sequence alignment evaluation and phylogenetic reconstruction, *Nucleic Acids Res 43*, W3-6.

[7] Colovos, C., and Yeates, T. O. (1993) Verification of protein structures: patterns of nonbonded atomic interactions, *Protein Sci 2*, 1511-1519.

[8] Pymol. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

[9] Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J Biomol NMR 8*, 477-486.

[10] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK : a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography 26*, 283-291.

[11] Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures, *Nature 381*, 272.

[12] Wallner, B., and Elofsson, A. (2006) Identification of correct regions in protein models using structural, alignment, and consensus information, *Protein Sci 15*, 900-913.

[13] Elengoe, A., Naser, M. A., and Hamdan, S. (2014) Modeling and docking studies on novel mutants (K71L and T204V) of the ATPase domain of human heat shock 70 kDa protein 1, *Int J Mol Sci 15*, 6797-6814.

[14] Messaoudi, A., Belguith, H., and Ben Hamida, J. (2013) Homology modeling and virtual screening approaches to identify potent inhibitors of VEB-1 beta-lactamase, *Theor Biol Med Model 10*, 22.

[15] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *Journal of Applied Crystallography 26*, 283-291.

[16] Rhodes, J. M., Subramanian, S., Laird, E., Griffin, G., and Kenny, R. A. (2020) Perspective: Vitamin D deficiency and COVID-19 severity–plausibly linked by latitude, ethnicity, impacts on cytokines, ACE2, and thrombosis (R1), *Journal of internal medicine*.

[17] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *Journal of molecular graphics 14*, 33-38.