Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Molecular Simulations suggest Vitamins, Retinoids and Steroids as Ligands of the Free Fatty Acid Pocket of the SARS-CoV-2 Spike Protein**

Deborah K. Shoemark[+], Charlotte K. Colenso[+], Christine Toelzer, Kapil Gupta, Richard B. Sessions, Andrew D. Davidson, Imre Berger, Christiane Schaffitzel, James Spencer, and Adrian J. Mulholland**

anie_202015639_sm_miscellaneous_information.pdf

Author Contributions

D.S. Conceptualization: Equal; Investigation: Equal; Writing—original draft: Lead; Writing—review & editing: Equal
C.C. Conceptualization: Supporting; Investigation: Equal; Writing—review & editing: Equal
C.T. Writing—review & editing: Supporting
K.G. Writing—review & editing: Supporting
R.S. Software: Supporting; Writing—review & editing: Supporting
A.D. Writing—review & editing: Supporting
I.B. Conceptualization: Supporting; Resources: Supporting; Writing—review & editing: Supporting
C.S. Conceptualization: Supporting; Resources: Supporting; Writing—review & editing: Supporting
J.S. Conceptualization: Equal; Funding acquisition: Equal; Resources: Supporting; Supervision: Equal; Writing—review & editing: Equal
A.M. Conceptualization: Equal; Funding acquisition: Equal; Resources: Equal; Supervision: Equal; Writing—review & editing: Equal.

Supplementary Information

METHODS

Molecular Dynamics Simulations

Proteins were prepared as described in Toelzer *et al*.[1]. Loops for the unstructured regions of the locked (LA^-- bound) and open (apo) cryo-EM structures were built using Chimera (UCSF)[2]. Likely disulphide bonds were reconstructed based on experimentally observed distances (42 for the locked structure and 43 for the open) and each chain sequence was used in an EBI-blast check to verify wild type spike sequence post build. PROCHECK[3] was then used to check the quality of the resulting structure prior to simulation. ACPYPE[4] was used to prepare the topologies for all ligands. Cholesterol and dexamethasone were superimposed onto the equilibrated LA^- positions with most polar ends of the molecules oriented to the carboxylate position of the linoleate. Dexamethasone and cholesterol were fitted into the fatty acid binding pocket in the locked conformations following a brief relaxation with LA^-.

Simulation details: All simulations were performed under the Amber99SB-ildn[5],[6],[7] forcefield in NPT ensembles at 310 K using periodic boundary conditions. Hydrogen atoms, consistent with pH7, were added to the complex. Short-range electrostatic and van der Waals interactions were truncated at 1.4 nm while long-range electrostatics were treated with the particle-mesh Ewald method and a long-range dispersion correction applied. A simulation box extending 2 nm from the protein was filled with TIP3P water molecules and 150 mM Na^+ and Cl^- ions added to attain a neutral charge overall. Pressure was controlled by the Berendsen barostat and temperature by the V-rescale thermostat. The

simulations were integrated with a leapfrog algorithm over a 2 fs time step, constraining bond vibrations with the P-LINCS method. Structures were saved every 0.1 ns for analysis in each of the simulations over 200 ns. Simulations were run on the Bristol supercomputer BlueCrystal 4, the BrisSynBio BlueGem, and the UK supercomputer, ARCHER.
Software: The GROMACS-2019.2[8] suite of software was used to set up and perform the molecular dynamics simulations and analyses for BlueGem and BlueCrystal runs and the ARCHER runs. Molecular graphics manipulations and visualisations were performed using VMD-1.9.1[9] and Chimera-1.10.2[2].

Virtual Screening of Ligands

A database of 2697 FDA-approved drugs in SDF format was downloaded from SelleckChem (https://www.selleckchem.com/). 3D coordinates of each compound (protonated at pH7.5) were generated using The Open Babel Package, version 2.3.2 (http://openbabel.org)[10], and up to 50 conformers per compound were produced using Confort[11] (Tripos Inc.). The receptor used for docking was an equilibrated frame from the molecular dynamics simulation of the LA^{-}- bound SARS-CoV-2 spike glycoprotein cryo-EM structure (PDB code 6ZB5, 2.85 Å resolution). A 13 Å radius around LA^{-} was selected as the docking pocket with C9 being the centroid. LA^{-} was subsequently removed prior to the docking calculations. The virtual screen was run using BUDE1.2.9[12] via a cloud computing platform powered by Oracle Corp. The total number of compounds screened was 2505, accounting for 39 compounds that had undefined atoms, and 153 compounds with atoms that are not parameterised in the BUDE forcefield.

DOCKING RESULTS

Validation of docking method

The presence of LA^{-} in the repurposing library allows for validation of BUDE docking by comparing the predicted mode of LA^{-} binding with that observed in the equilibrated structure of the locked complex (i.e. the cryo-EM_{eq} structure). Experimentally, the K_d of LA^{-} for the isolated SARS-CoV-2 spike RBD has been determined by surface plasmon resonance as ~41 nM[1]. LA^{-} ranked 86th in the list of compounds ordered by BUDE binding energy, with a heavy atom RMSD between the docked (best scoring pose from BUDE) and the cryo-EM_{eq} LA^{-} structure of 1.6 Å. Furthermore, in the cryo-EM_{eq} structure, and as described in the main text, the linoleate carboxylate group forms salt-bridge and hydrogen-bonding interactions with the sidechains of R408 and Q409. In the lowest-energy docked pose generated by BUDE, the linoleate carboxylate group occupies the same position as that observed in the cryo-EM_{eq} structure and also forms salt-bridge and hydrogen-bonding interactions with the R408 and Q409 sidechains (Figure 6E). Thus, the positions of bound LA^{-} in the experimental and docked structures are very similar; their close correspondence and the consistency with the interactions from experimental and MD structures, provides support for the use of BUDE in this virtual screening study.

Vitamins

The list of the 100 highest ranked (‘hit’) compounds ordered by BUDE binding energy can be found in Table S1 along with their predicted ligand efficiencies. Here, we identify from these results a

number of compound classes of particular interest, based upon their demonstrated or proposed relevance to SARS-CoV-2 and/or other respiratory viral infections. Notably, the fat-soluble vitamins D (vitamin D3 metabolite calcitriol), K and A are all present in the FDA library and all rank more highly than LA^- by BUDE binding energy. Indeed, vitamin K2 is the highest ranked compound in the BUDE output list, with vitamin A acetate ranked 17th by predicted binding energy, vitamin D (calcitriol) 24th, vitamin A 38th and vitamin K1 70th. Inspection of the respective lowest energy BUDE docked poses showed that all of these vitamins align very well with the pose adopted by LA^- bound to the locked spike trimer after 200 ns of MD simulation (Figure 7). In the case of calcitriol (Figure 7A, D) and vitamin K2 (Figure 7B, E), the presence of double bonds in their extended sidechains allows them to align closely with the hydrophobic LA^- tail and consequently to fit well into the highly hydrophobic FA binding site.

Retinoids

Numerous studies have sought to identify candidate approved drugs suitable for repurposing as therapies for COVID-19, both computationally and experimentally. Notably, Riva *et al*.[13] screened ~12,000 compounds from the ReFRAME library[14] for inhibition of SARS-CoV-2 replication in Vero E6 cells. Accordingly, we compared the list of highly ranked compounds from the BUDE screen against those identified experimentally by Riva *et al.*[13] as able to inhibit viral replication. Of the 100 agents on their list of most active compounds three (tretinoin, acitretin and tazarotene) are relatively prominent in the ordered list of compounds identified by BUDE, ranking 5th, 8th, and 214th, respectively by BUDE energy. According to Riva *et al*.[13], acitretin at 2.5 μM inhibits viral replication by 40%, while tazarotene and tretinoin at 1 μM inhibit viral replication by 50% and 42%, respectively. Notably, the lowest energy binding poses identified by BUDE for all of these compounds involve salt-bridge and hydrogen-bonding interactions with R408 and Q409, in addition to hydrophobic interactions with the remainder of the binding pocket (Figure 8A, C, D). These data suggest that these compounds may act on viral replication by binding to this free FA site in a manner similar to LA.

Like vitamin A, tretinoin, acitretin and tazarotene are all retinoic acid receptor agonists: tretinoin (Atralin) is also known as all-trans retinoic acid, acitretin is a second-generation retinoid and tazarotene (Avage) is classed as a third-generation retinoid. In their recent study Riva *et al*.[13] highlight retinoic acid receptor agonists as one of 15 target classes that were enriched in their assay, with a total of 13 retinoic acid receptor agonists appearing in their list of 100 top-ranked active compounds. With this in mind, we interrogated the list of BUDE hits to seek additional retinoid compounds, and, in addition to vitamin A, identified a further four retinoids: adapalene (9th, third-generation topical retinoid), fenretinide (10th, synthetic phenylretinamide retinol analogue), etretinate (18th, second-generation retinoid) and isotretinoin (68th, 13-cis retinoic acid) as prominent among the highly ranked hits. Thus, our virtual screening identifies multiple synthetic retinoids in addition to vitamin A as candidate ligands, supporting our contention that the spike protein FA site may also be capable of binding members of this class of compounds.

Steroids

Similarly, reports that the spike protein binds cholesterol[15] as well as the possibility that the LA^- site may be capable of binding dexamethasone, led us to scrutinize the list of BUDE hits for steroid compounds. A variety of steroid agents, including the progestins (hydroxyprogesterone caproate, ranked 29th by BUDE energy, chlormadinone acetate (41st); and medroxyprogesterone acetate (50th)); the androgens (testosterone enanthate, 19th); the corticosteroids (methylprednisolone hemisuccinate, 47th); the bile acids (ursolic acid, glycocholic acid, dehydrocholic acid; 6th, 7th and 49th, respectively) and the synthetic androstane steroid abiraterone acetate (23rd) all feature among candidate ligands that are highly ranked by BUDE. Several other steroid-like natural products (carbenoxolone (2nd),

oleanolic acid (27th), madecassic acid (32nd) and hederagenin (34th)) are also prominent in the BUDE ranking list. Although (and typical of the results of docking studies) the list of hits is subject to some variation with different ranking procedures (by binding energy or ligand efficiency), there is a consistent preponderance of steroids among the most highly ranked compounds.

The anti-inflammatory corticosteroid dexamethasone has attracted considerable attention on the basis of its demonstrated efficacy improving outcomes for COVID-19 patients in clinical trials[16]. Accordingly, we also examined the results of BUDE docking for dexamethasone. Consistent with the MD simulations described in the main text, the dexamethasone binding mode predicted by BUDE does not involve interactions with R408 or Q409 (unlike LA^- and the retinoic acid receptor agonists) and instead is dominated by hydrophobic interactions with residues on the opposite side of the binding site (Figure 8B). Thus, although other steroidal compounds score more highly, our docking results also support the MD simulations described here in identifying a possible interaction of dexamethasone with the SARS-CoV-2 spike protein.

Supplementary Figures

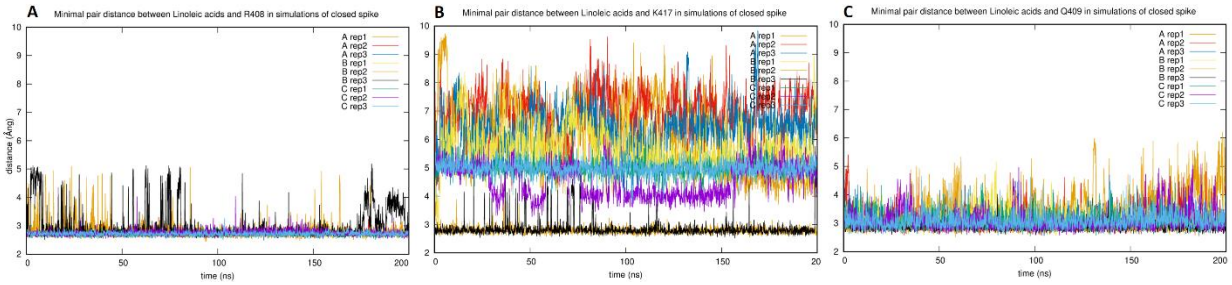

Figure S1. Interaction of linoleate with spike RBD during MD simulations of the locked complex. Plots of minimal pairwise distances between linoleate atoms and: **A** the nitrogen atoms of the R408 guanidinium; **B** the nitrogen atom of the K417 amino group; and **C** the nitrogen and oxygen atoms of the Q409 amide during simulations of locked conformation linoleate complexes. The coloring of individual chains/replicates is consistent across the Figure.

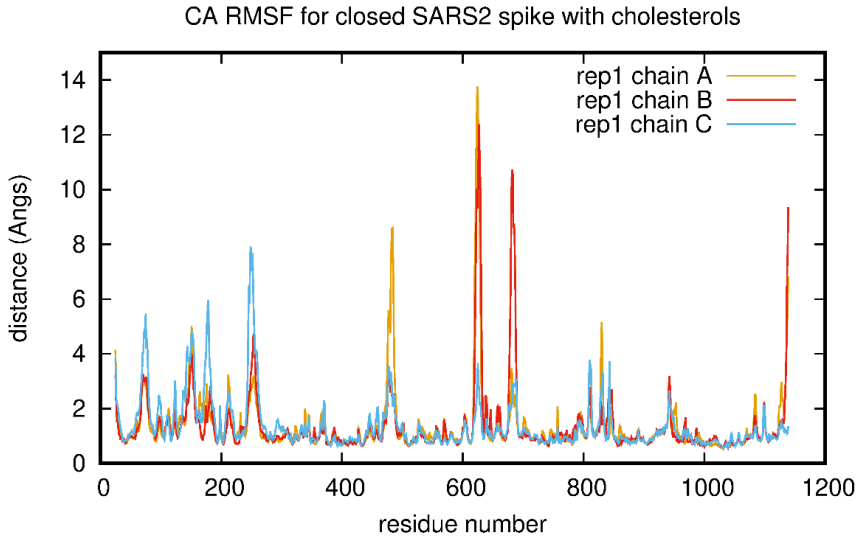

Figure S2. RMSF for the individual chains showing the regions destabilised by the binding of cholesterol in the fatty acid binding site.

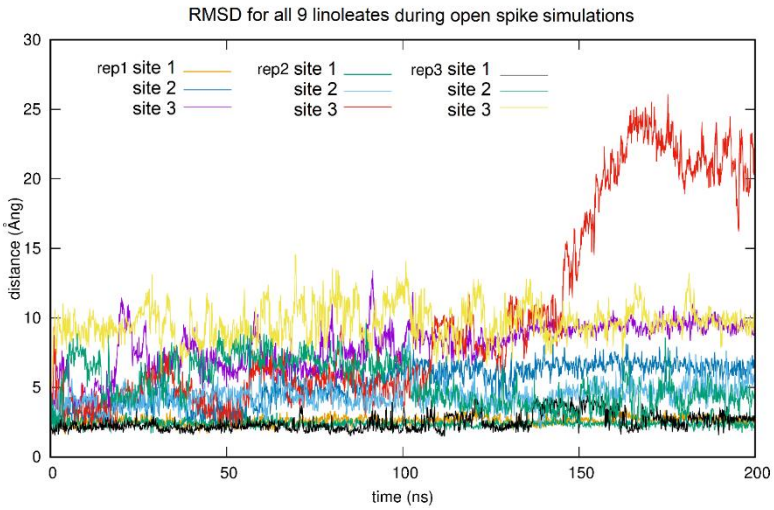

Figure S3. Stability of linoleate binding in simulations of the open spike. Mean distance moved by linoleate molecules from their positions in the starting structures during simulations. The C-alphas for the individual chains were aligned with those in their first frame and the distance calculated for linoleate compared to its coordinate position in the first frame, for each of 3 sites (site 1 = locked, site 2 = closed, site 3 = open as described earlier) during three replicate simulations of the open spike trimer. Note increased distance of LA^- in site 3 (open site) from its starting position (yellow, magenta traces), with dissociation evident in one trajectory (red trace).

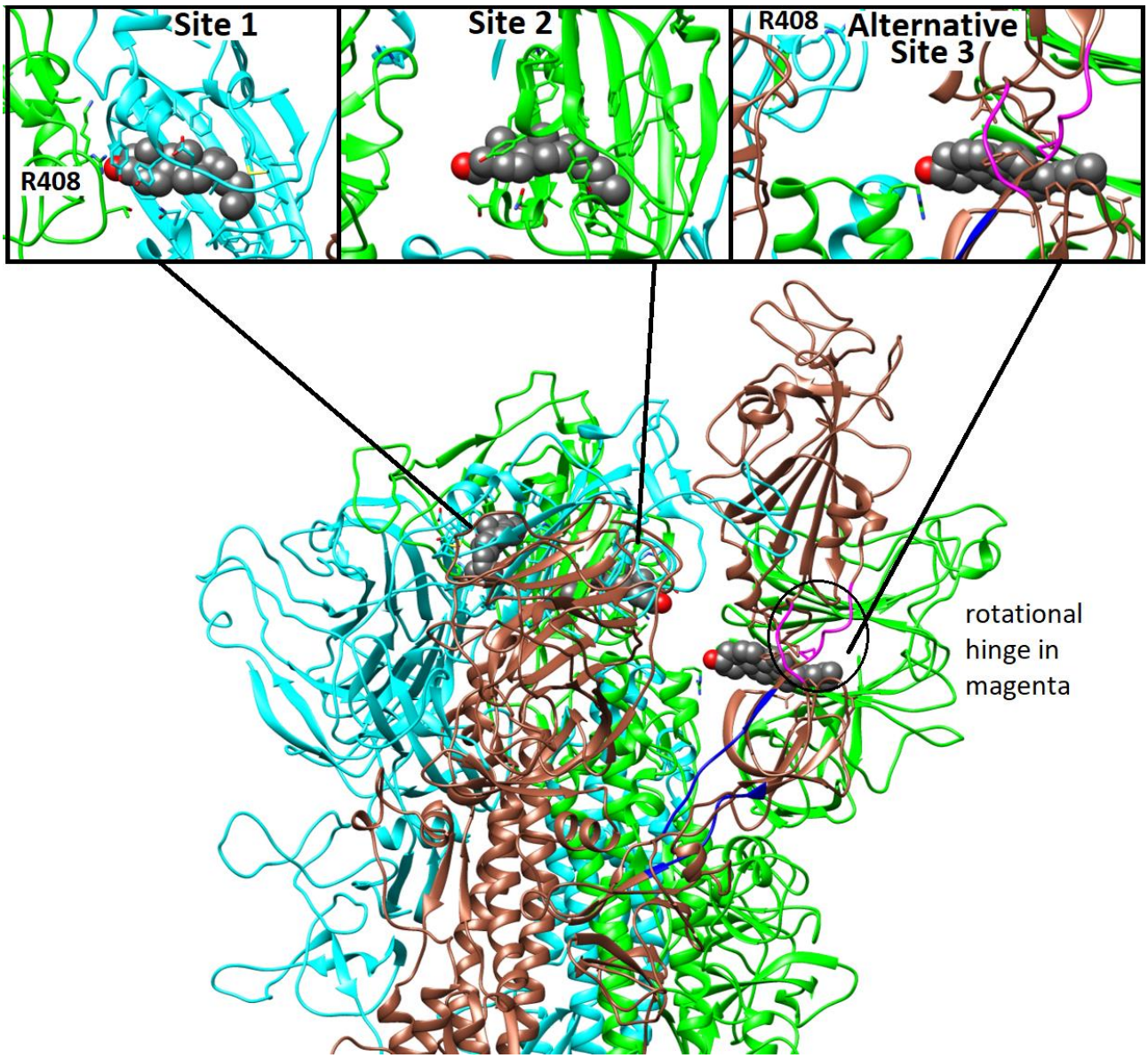

Figure S4. Three distinct cholesterol binding sites in the open spike conformation suggested by MD simulations. SARS-CoV-2 spike chains A (cyan ribbon) B (brown ribbon) and C (green ribbon) with bound cholesterols (carbon atoms grey spheres) from the last frame of a 200 ns MD trajectory of the open SARS-CoV-2 spike trimer. Top row inserts illustrate cholesterol binding sites 1 and 2 and the alternative site 3. Main image shows relative positions of the three sites. Alternative site 3 is shown in the context of residues required for the rigid body rotational (circled and magenta ribbon) and vertical motion (dark blue ribbon) necessary to raise the RBD[17] for interaction with the ACE2 receptor.

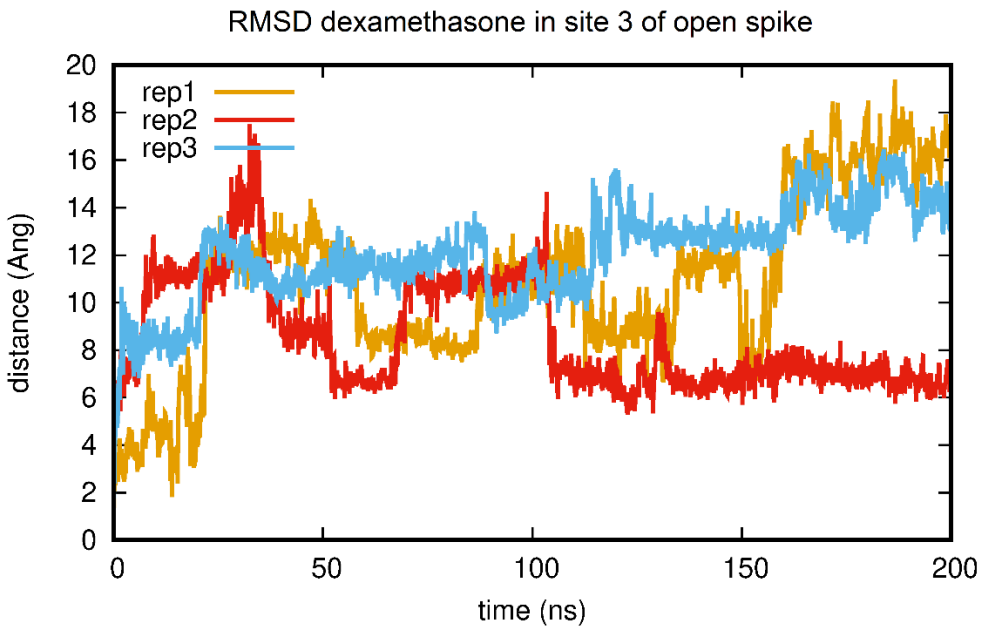

Figure S5. RMSD for dexamethasone bound to site 3 during MD simulations of the open spike complex. Plot shows RMSD of dexamethasone with respect to its position in site 3 of the starting structure (associated with the underside of the raised RBD).

FDA drugs with 15-45 heavy atoms	BUDE binding energy kJ/mol	Ligand efficiency	FDA drugs with 15-45 heavy atoms	BUDE binding energy kJ/mol	Ligand efficiency
Vitamin K2	-155.4074	-4.71	Droperidol	-119.3248	-4.26
Carbenoxolone disodium	-144.0183	-3.51	Alectinib (CH5424802)	-119.23	-3.31
Doxercalciferol	-143.4565	-4.78	Sitagliptin phosphate monohydrate	-119.1617	-4.26
Tedizolid	-143.404	-5.31	Norethisterone Enanthate	-118.9326	-3.96
ATRALIN (tretinoin)	-137.6396	-6.26	Obeticholic acid	-118.7365	-3.96
Ursolic acid	-136.9199	-4.15	Cilostazol	-118.7236	-4.4
Glycocholic acid	-135.8867	-4.12	Fluconazole	-118.4734	-5.39
Acitretin	-135.0294	-5.63	VX661	-118.1015	-3.19
Adapalene	-134.9415	-4.35	Ceftezole sodium	-117.8625	-4.21
Fenretinide	-134.8654	-4.65	Fusidate Sodium	-117.1164	-3.17
Pemirolast potassium	-134.1502	-7.89	Fusidine	-117.1164	-3.17
Alizapride hydrochloride	-134.0645	-5.83	Radotinib (IY-5511)	-117.0468	-3
Maraviroc	-133.9345	-3.62	Clobetasol propionate	-117.0436	-3.66
Calcipotriene	-133.9261	-4.46	halobetasol propionate	-116.9628	-3.54
Pranlukast	-133.7867	-3.72	17-Beta-Estradiol-3,17-Dipropionate	-116.8183	-4.17
Alfacalcidol	-133.5167	-4.6	Miconazole	-116.4486	-4.66
Vitamin A acetate	-133.3426	-5.56	DESOWEN (desonide)	-116.2427	-3.87
TEGISON (etretinate)	-133.035	-5.12	Isotretinoin	-116.2182	-5.28
Testosterone Enanthate	-132.6409	-4.57	SOLATENE (beta carotene)	-116.0116	-2.9
Apramycin Sulfate	-132.4513	-3.58	Vitamin K1	-115.978	-3.51
Cefamandole (nafate)	-131.8277	-3.99	Doxazosin	-115.8407	-3.51
Forsythin	-131.6235	-3.46	Cyproterone acetate	-115.5778	-3.99
Abiraterone Acetate	-131.1496	-4.52	Altrenogest	-115.4188	-5.02
Calcitriol	-129.4745	-4.32	Eicosapentaenoic Acid	-115.04	-5.23
ODM-201	-128.0165	-4.57	Ethynodiol diacetate	-114.4823	-4.09
Sofalcone	-127.091	-3.85	Finasteride	-114.3894	-4.24
Oleanolic Acid	-126.7326	-3.84	Donepezil HCl	-114.0946	-4.07
Fumagillin	-125.3788	-3.8	Andrographolide	-113.9383	-4.56
Hydroxyprogesterone caproate	-124.7928	-4.03	Octenidine Dihydrochloride	-113.8641	-2.85
Ceftezole	-124.4288	-4.44	Deoxycholic acid	-113.8574	-4.07
Peimine	-124.1502	-4	RISPERIDONE (risperidone)	-113.795	-3.79
Madecassic acid	-122.9159	-3.41	Donepezil	-113.7222	-4.06
Fipexide hydrochloride	-122.8516	-4.55	Ruscogenin	-113.6938	-3.67
Hederagenin	-122.6472	-3.61	Cortisone acetate	-113.6429	-3.92
Trazodone hydrochloride	-122.436	-4.71	Nefazodone hydrochloride	-113.6076	-3.44
Teprenone	-122.406	-5.1	Linoleic acid	-113.5997	-5.68

INVEGA (paliperidone)	-122.029	-3.94	Squalene	-113.4198	-3.78
Vitamin A	-121.9105	-5.81	canrenoic acid, potassium salt	-113.3962	-4.36
IRBESARTAN (irbesartan)	-121.1952	-3.79	Cariprazine HCl	-113.2954	-4.05
irinotecan	-121.0838	-2.82	atovaquone (atavaquone, Mepron)	-113.2767	-4.36
Chlormadinone acetate	-120.6775	-4.31	Misoprostol	-113.1274	-4.19
Buspirone hydrochloride	-120.263	-4.3	pentoxifylline	-112.9423	-5.65
JNJ-42756493 (Erdafitinib)	-120.1199	-3.64	Ouabain Octahydrate	-112.8282	-2.75
Sitagliptin	-119.9671	-4.28	Hyodeoxycholic acid	-112.5414	-4.02
Diflucortolone valerate	-119.8251	-3.52	Estradiol Valerate	-112.2776	-4.32
Betulin	-119.7406	-3.74	Quillaic acid	-112.2117	-3.21
Methylprednisolone hemisuccinate	-119.6263	-3.52	CETACORT (hydrocortisone)	-112.0461	-4.31
Alectinib hydrochloride	-119.6058	-3.32	ABILIFY (aripiprazole)	-111.8436	-3.73
Dehydrocholic acid	-119.4235	-4.12	CHENIX (chenodiol)	-111.6111	-3.99
Medroxyprogesterone acetate	-119.3268	-4.26	Mosapride	-111.4791	-3.84

Table S1. The top 100 drugs from the database of FDA-approved compounds (with between 15 and 45 heavy atoms) from docking (with BUDE) into the fatty acid binding site of the locked SARS-CoV-2 spike protein, ranked according to predicted binding energy, also showing their predicted ligand efficiency (kJ/mol/number of heavy atoms).

References

[1] C. Toelzer, K. Gupta, S. K. N. Yadav, U. Borucu, A. D. Davidson, M. Kavanagh Williamson, D. K. Shoemark, F. Garzoni, O. Staufer, R. Milligan, J. Capin, A. J. Mulholland, J. Spatz, D. Fitzgerald, I. Berger, C. Schaffitzel, *Science* **2020**, *370*, 725-730.

[2] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J Comput Chem* **2004**, *25*, 1605-1612.

[3] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, J. M. Thornton, *J Biomol NMR* **1996**, *8*, 477-486.

[4] A. W. Sousa da Silva, W. F. Vranken, *BMC Res Notes* **2012**, *5*, 367.

[5] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J Comput Chem* **2004**, *25*, 1157-1174.

[6] D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J Comput Chem* **2005**, *26*, 1668-1688.

[7] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins* **2010**, *78*, 1950-1958.

[8] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J Comput Chem* **2005**, *26*, 1701-1718.

[9] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph* **1996**, *14*, 33-38, 27-38.

[10] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J Cheminform* **2011**, *3*, 33.

[11] R. S. Pearlman, R. Balducci, *National Meeting of the American Chemical Society, New Orleans* **1998**.

[12] S. McIntosh-Smith, J. Price, R. B. Sessions, A. A. Ibarra, *Int J High Perform Comput Appl* **2015**, *29*, 119-134.

[13] L. Riva, S. Yuan, X. Yin, L. Martin-Sancho, N. Matsunaga, L. Pache, S. Burgstaller-Muehlbacher, P. D. De Jesus, P. Teriete, M. V. Hull, M. W. Chang, J. F. Chan, J. Cao, V. K. Poon, K. M. Herbert, K. Cheng, T. H. Nguyen, A. Rubanov, Y. Pu, C. Nguyen, A. Choi, R. Rathnasinghe, M. Schotsaert, L. Miorin, M. Dejosez, T. P. Zwaka, K.

Y. Sit, L. Martinez-Sobrido, W. C. Liu, K. M. White, M. E. Chapman, E. K. Lendy, R. J. Glynne, R. Albrecht, E. Ruppin, A. D. Mesecar, J. R. Johnson, C. Benner, R. Sun, P. G. Schultz, A. I. Su, A. Garcia-Sastre, A. K. Chatterjee, K. Y. Yuen, S. K. Chanda, *Nature* **2020**, *586*, 113-119.

[14] J. Janes, M. E. Young, E. Chen, N. H. Rogers, S. Burgstaller-Muehlbacher, L. D. Hughes, M. S. Love, M. V. Hull, K. L. Kuhen, A. K. Woods, S. B. Joseph, H. M. Petrassi, C. W. McNamara, M. S. Tremblay, A. I. Su, P. G. Schultz, A. K. Chatterjee, *Proc Natl Acad Sci U S A* **2018**, *115*, 10750-10755.

[15] Y. Peng, L. Wan, C. Fan, P. Zhang, X. Wang, J. Sun, Y. Zhang, Q. Yan, J. Gong, H. Yang, X. Yang, H. Li, Y. Wang, Y. Zong, F. Yin, X. Yang, H. Zhong, Y. Cao, C. Wei, *medRxiv* **2020**, DOI:2020.2004.2016.20068528.

[16] R. C. Group, P. Horby, W. S. Lim, J. R. Emberson, M. Mafham, J. L. Bell, L. Linsell, N. Staplin, C. Brightling, A. Ustianowski, E. Elmahi, B. Prudon, C. Green, T. Felton, D. Chadwick, K. Rege, C. Fegan, L. C. Chappell, S. N. Faust, T. Jaki, K. Jeffery, A. Montgomery, K. Rowan, E. Juszczak, J. K. Baillie, R. Haynes, M. J. Landray, *N Engl J Med* **2020**, DOI:10.1056/NEJMoa2021436.

[17] R. Melero, C. O. S. Sorzano, B. Foster, J-L. Vilas, M. Martínez, R. Marabini, E. Ramírez-Aportela, R. Sanchez-Garcia, D. Herreros, L. del Caño, P. Losana, Y.C. Fonseca-Reyna, P. Conesa, D. Wrapp, P. Chacon, J. S. McLellan, H. D. Tagareb, J-M. Carazoa, 2020 bioRxiv preprint doi: https://doi.org/10.1101/2020.07.08.191072