Team name	CERMN
Team member(s) (firstname lastname; …)	Ronan Bureau ; Mohammed Benabderrahmane ; Jana Sopkova De Oliviera Santos ; Alban Lepailleur.
Affiliation	Universite de Caen Normandie
Contact email	ronan.bureau@unicaen.fr
Contact phone number (optional)	0645972384
Protein targets (for example: 3CLPro/Nsp5, BoAT1, Fc Receptor, Furin, IL6R, M protein, Nspx, OrfXx, N, E, etc…) | 3 required	*3C-like proteinase / ADP ribose phosphatase of NSP3 / Human transmembrane protease serine 2.*

Section 1: methods & metrics

Describe what methods you have used, how they are independent from one another, what your workflow was, how you performed the cross-correlation between your methods. If applicable, please report estimated performance metrics of your methods, such as accuracy, sensitivity, false-discovery rate, etc., and how those metrics were obtained (e.g. cross-validation). Please provide key references if available.

Methods:

1. Virtual Screening (VS) based on docking approaches by two methods : High-Throughput Virtual Screening (HTVS) for big dataset (see below for the databases) and classical VS for a small database (sweetlead).
    a. For HTVS, VIRTUAL FLOW with QVINA2[1,2] for the software. AutoDockTools[3] for the definition of the configuration file associated to QVINA2. The overall process was in four steps :
        i. First run of the overall dataset with a limit for the exhautiveness level (value of 1).
        ii. Ranking (QVINA2 scores) and extraction of the best ligands (around 1 million).
        iii. Second run of the best ligands by considering a flexibility of the active site and the ligands with an exhautiveness level of 8.
        iv. Final extraction of around 50000 ligands in function of a cutoff for the docking score.
    b. For VS, Glide[4] from Schrodinger.
        i. Default parameter for the receptor grid generation (scaling factor of 1.0 / Partial charge cutoff of 0.25, see [4]).
        ii. Standard Precision for the ligand docking (SP, see [5]) / Docking method : rigid.
        iii. Cutoff in function of the docking score (-7 for the three targets and by considering the value for the reference ligands associated to the proteins (around -10)).
2. Clustering of the HTVS / VS results with kmeans (Pipeline Pilot) and selection/ranking of the best ligands. This was done in four steps for the results associated to HTVS and QVINA2.
    a. Clustering by considering these parameters : ECFP4 (1024 bit) / Coefficient of Tanimoto / 3000 clusters.
    b. For each cluster, extraction of the centroid and the 5 best ligands (QVINA2 docking score).
    c. For the selection, determination of a druglikeness with calculation of the Quantitative Estimate of Druglikeness (QED). The ligands should have a value superior or equal to 0.5.

d. Ranking of the final selection in function of the QVINA2 docking score.

3. Definition of constrained pharmacophores with Norns[6] for ligands of proteases described in Chembldb.
    a. Main parameters :
        i. Cutoff between active and inactive ligands : 1000 nM (93364 *vs* 83078 for active *vs* inactive data (see section 3 for more information))
        ii. Number of pharmacophoric features : 4
        iii. Constraint for the pharmacophores : Aromatic and Basic features should be present (in relation with target 3)
        iv. Support : 10
        v. Growth rate : 10
    b. Extraction of 8387 pharmacophores with the following statistics[6].
        i. Precision : 0.85 (6633 compounds predicted as actives and 5629 really actives)/ Recall : 0.060 (5629 predicted active compounds out of 93364 active compounds) / F measure : 0.11
    c. Agreement between these pharmacophores with the best ligands associated to target 3.
4. Definition of closest similarities between two datasets (Pipeline Pilot).
    a. Main parameters :
        i. Fingerprint : ECFP4 (1024 bit)
        ii. Maximum number of closest : 1 / Minimum similarity : 0 / Maximum similarity : 1
        iii. Top N filter : 10000 (best closest similarity values).
        iv. Selection at the end of the process (see results) : closest similarity values >= 0.5.

Section 2: targets

Describe for each protein target: why you chose it, from which source you obtained it (e.g., insidecorona.net / covid.molssi.org / rcsb.org) and why this is the best quality structure, if any pre-processing (e.g., energy minimization, residue correction, alternative folding, …) was performed.

Target 1:

COVID-19 main protease (3C-like proteinase), pdb code : **6W63** (rcsb.org, deposited : 25-03-2020 / resolution of 2.1 Å).

COVID-19 main protease complexed with X77 described as a potent non peptide inhibitor of this protease. Analogs of these compounds are described in a review of 3CLpro non covalent inhibitors (SARS COVID), compounds resulting from a HTS of NIH molecular library collection[7,8]. We did not find a particular IC50 (or Ki) value for X77 but starting with a non covalent non peptide ligand was preferred.

Optimization of the pdb file with Protein Preparation Wizard[9] (Schrödinger) in three steps :

1. Assign bond orders / Add hydrogens / Create disulfide bonds / Delete waters beyond 5 Å from heterotatom groups.
2. H-bond assignment (sample water orientations / optimize the orientation of the H-bond).
3. Restrained minimization with OPLS3e and a RMSD of 0.3 Å for the convergence of heavy atoms.

A definition of the configuration file for QVINA2 with adt (AutoDockTools) was carried out. Water molecules were conserved (water molecules conserved by the previous step : Protein Preparation Wizard).
For the second run (HTVS), the following residues were considered as flexible : His41, Cys145, His163 (two active torsions for the first three), Glu166, Gln189 (3 active torsions for the last one).

Target 2:
ADP ribose phosphatase of NSP3 (Papain-like protease (PLpro)), pdb code : **6W02** (rcsb.org, released on 11-03-2020 / resolution : 1,50 Å)
ADP ribose phosphatase of NSP3 complexed with Adenosine-5-Diphosphoribose. For justifying the potential of this target, a recent reference[10] (but for Chikungunya virus) is a good example for the selection of potential inhibitors by *in silico* approaches on the ADP-ribose site.
Optimization of the pdb file with Protein preparation wizard[9] (Schrodinger) in three steps :

1. Assign bond orders / Add hydrogens / Create disulfide bonds / Delete waters beyond 5 Å from heterotatom groups.
2. H-bond assignment (sample water orientations / optimize the orientation of the H-bond).
3. Restrained minimization with OPLS3e and a RMSD of 0.3 Å for the convergence of heavy atoms.

A definition of the configuration file for QVINA2 with adt (AutoDockTools) was carried out. Water molecules were conserved (water molecules conserved by the previous step : Protein Preparation Wizard).
For the second run (HTVS), the following residues were considered as flexible : Asp22, Asn40 (2 active torsions).

Target 3:
Human transmembrane protease serine 2 (**TMPRSS2**). Source : covid.molssi.org.
Homology model based on the structure of TMPRSS15/enteropeptidase[11](pdb : 4DGJ). We chose a ligand (5-Azaindole inhibitors[12] of factor VIIa described in the pdb file 2FLR) as complexed with TMPRSS2 (definition of the binding site).
Optimization of the pdb file with Protein preparation wizard (Schrodinger) in three steps :

1. Assign bond orders / Add hydrogens / Create disulfide bonds / Delete waters beyond 5 Å from heterotatom groups.
2. H-bond assignment (sample water orientations / optimize the orientation of the H-bond).
3. Restrained minimization with OPLS3e and a RMSD of 0.3 Å for the convergence of heavy atoms.

Definition of the configuration file for QVINA2 with adt (AutoDockTools). Water molecules were conserved (water molecules conserved by Protein Preparation Wizard).
For the second run (HTVS), the following residues were considered as flexible : Glu44 (3 active torsions), Asp180 (2 active torsions).

Section 3: libraries
Describe which libraries you have used, how they were combined, if any compounds were removed / added, why additions are relevant, any unique features of your library, etc. Please provide the sources you obtained the libraries from (if publicly available). Describe the procedure of data preparation (removal of duplicates, standardization, etc). Indicate if different libraries were used for different targets, and why. If possible, provide a download link to your version of the library.

Library 1: **Real database** from HTVS project.

We have chosen to focus on compounds with a MW value superior to 375 g/mol and a large range of log P value in a first step. Afterwards, we have added compounds with a smaller value of MW but with a higher range of value for logP :

- MW between 375 and 500 daltons / logP between 1 and 4,5 : 382 millions compounds.
- MW between 325 to 375 daltons / logP between 2,5 and 4,5 : 240 millions compounds.

At the end : 622 millions of compounds for the three targets were considered (more than **1,8 billions** of compounds) for a computing time around 4-5 millions of hours/cpus (initial HTVS with an average time of 7s by docking / CINES and CRIANN for the facilities).

Library 2 : **Sweetlead library**[13] from SimTK.
Selection of 7134 compounds with MW between 200 and 700 Daltons.
A preparation of the 7134 ligands with ligPrep was carried out. The objective of ligprep is to produce the corresponding low-energy structures for docking programs[14].
The parameters were : OPLS3e for the force field / ionization : do not change / generate tautomers : no / determine chiralities : no / Ignore chiralities : no / number of stereoisomers : 10.

Library 3 : **Merck library**. The library (more than 5 millions compounds) was just used to check the agreement between this library and the final selection resulting from Real database. Indeed, all the compounds in Merck library should be straight available. So we are able (if necessary) to propose a second list in function of the similarities between the hits of real database and the compounds in the Merck library.

Library 4 : **ChEMBLDB for proteases**.

1. Extraction of all data associated to proteases (data extracted on October 2018) :
    a. 264396 biological data associated to 126941 compounds..
    b. 176442 pKi or pIC50 values (PCHEMBL_Value) associated to 105068 compounds (basis for the study with Norn).
    c. 53092 compounds for which at least a pKi/pIC50 value is superior to 6.

Library 5 : **Chembl4303835 (SARS-CoV-2).**

1. Preprint for the publication associated to the data : “Identification of inhibitors of SARS-CoV-2 in-vitro cellular toxicity in human (Caco-2) cells using a large scale drug repurposing collection” (https://www.researchsquare.com/article/rs-23951/v1).
2. 5653 biological data associated to 5653 compounds.
3. Cutoff at 75% for the percentage of inhibition (271 active compounds on this basis).
4. Creation of a sdf file (2D and 3D (same protocols than above) for active and all compounds.

Section 4: results
Briefly describe you key findings, any interesting trends in your data, a description of your top 5 compounds for each target. If possible, provide a link to a code and/or data repository. Please do not submit randomly selected compounds!

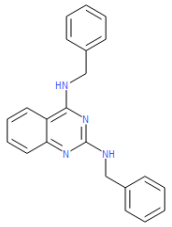
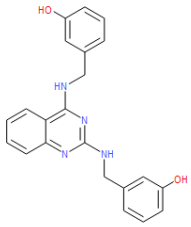
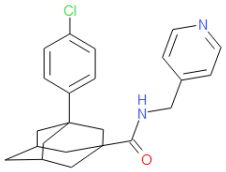
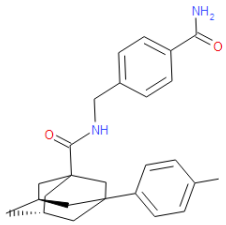
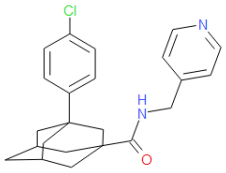
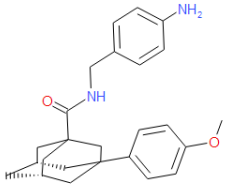
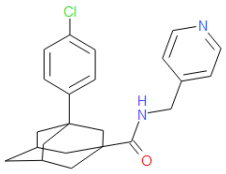
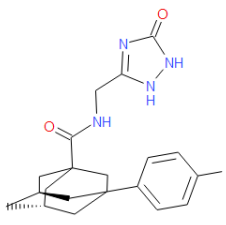
Results:

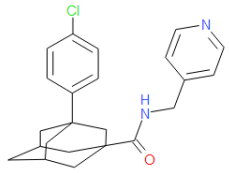
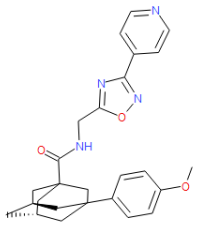
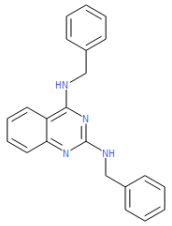
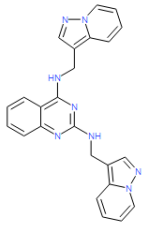
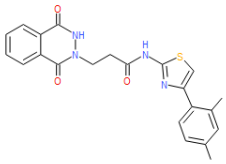
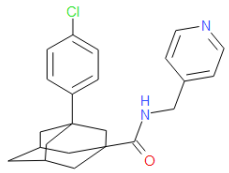
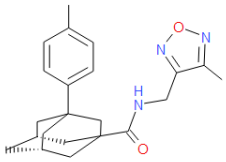
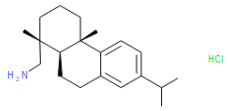
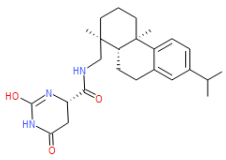
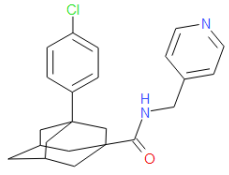
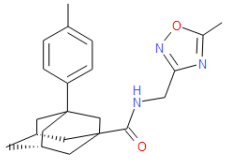
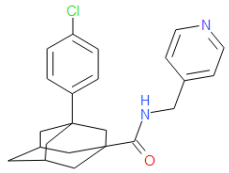
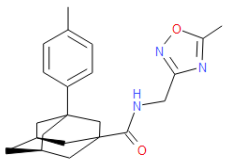
Target 1 :
Studies on Real Database (library 1).

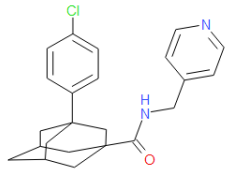
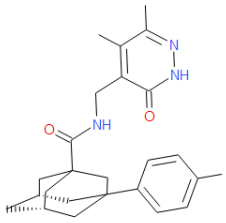
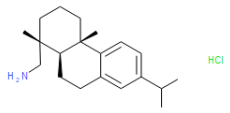
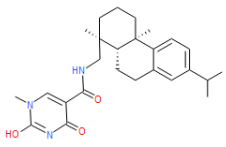
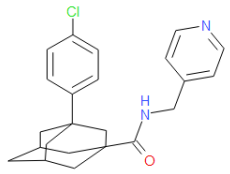
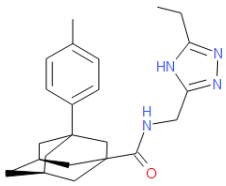
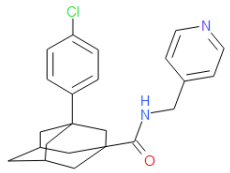
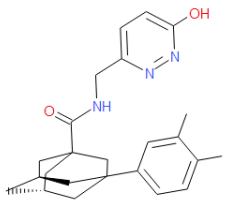
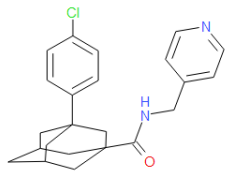
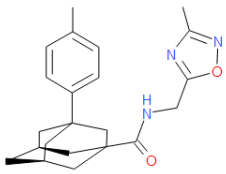
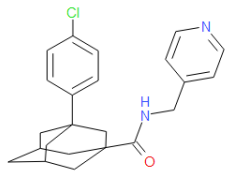
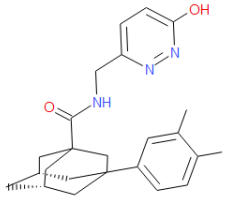
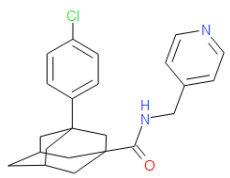
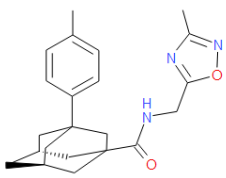
1. First HTVS

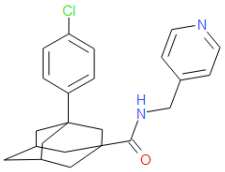
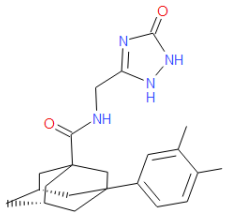
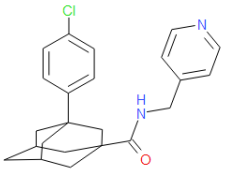
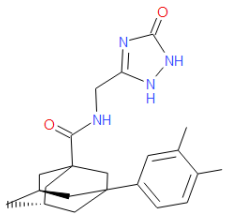
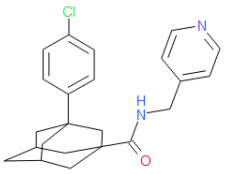
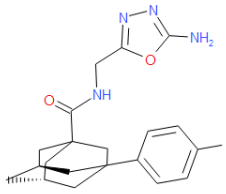
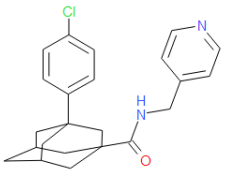
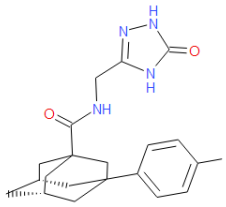
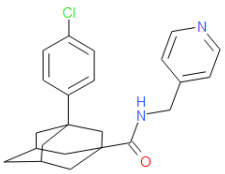
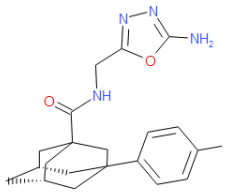
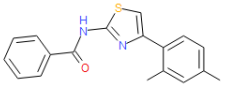
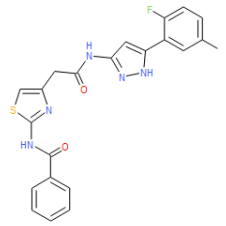
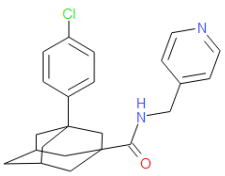
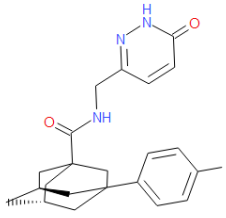
a. Scoring data for 559887710 compounds.
    i. Statistics for the docking scores :
        1. Min : -11.5 / Max : 400.3 / Median : -7.1
    b. Cutoff at -9.0 for the first selection : 930619 compounds (at this step, we wanted to select around 1 million compounds).
2. Second HTVS on the 930619 compounds (selection of around 50000 compounds).
    a. Cutoff at -9.5 for the second selection : 57038 compounds selected.
3. Clustering of the final results (towards 10000 compounds).
    a. 10613 compounds selected (classification in function of the docking score).
4. Similarities between the 57038 compounds (second HTVS) and the 271 compounds described as active towards SARS-COV2 (Chembl4303835).
    a. 32 compounds with similarities >= 0.5 (see Table 1).
        i. Among these compounds, we observed a large series in relation with **Opaganib** (Chembl2158685, line 2 in Table 1). On Clinical Trials web site, two studies[15,16] (one in phase 2) are described for **Opaganib** and COVID-19.
    b. These compounds (columns : HTVS / Name Real) could be considered in the first list.

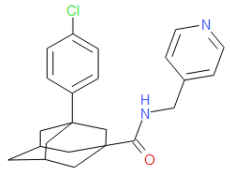
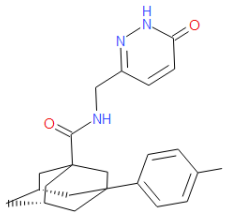
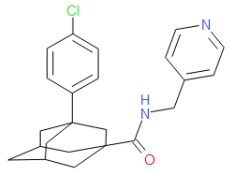
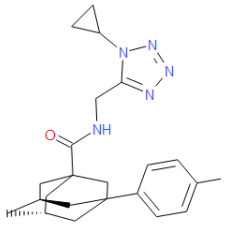
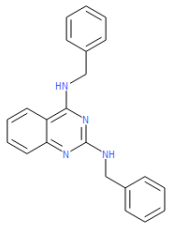
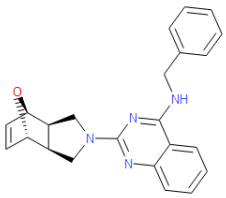
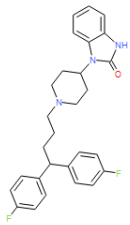
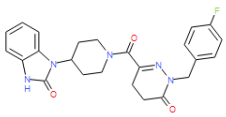
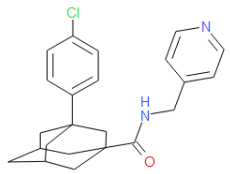
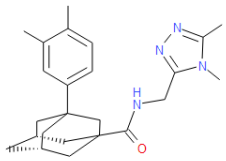
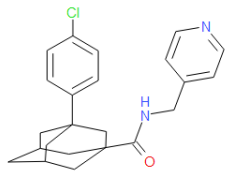
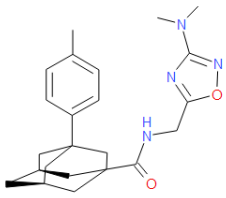
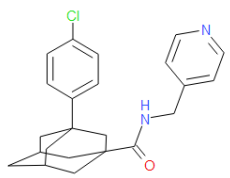
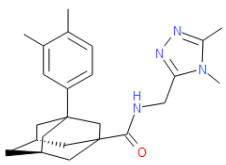
Table 1. Description of 32 compounds (HTVS and Name Real column) similar to active derivatives (Covid_plus and Chembl_ID column) towards SARS-COV2.

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL1624459	CHEMBL4303835	SARS-CoV-2		Z1587196385_1_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z28944530_2_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2301708925_1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1334316909_2_T2_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z240460112_2_T1_replica-1.pdbqt
	CHEMBL1624459	CHEMBL4303835	SARS-CoV-2		Z2094373928_1_T1_replica-1.pdbqt
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		Z88459020_1_T2_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1229914077_3_T1_replica-1.pdbqt
	CHEMBL1552895	CHEMBL4303835	SARS-CoV-2		PV-001831411371_3_T3_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1229818666_3_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1229818666_1_T1_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1174906910_1_T1_replica-1.pdbqt
	CHEMBL1552895	CHEMBL4303835	SARS-CoV-2		PV-001797541819_3_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2282504417_2_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2282739520_2_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1128318094_1_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2282739520_1_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1128318094_2_T1_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1334317467_1_T2_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1334317467_2_T2_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2074207032_2_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1334316909_1_T5_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2074207032_1_T1_replica-1.pdbqt
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		Z1772024632_1_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2282739545_2_T2_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z2282739545_1_T2_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1275585908_2_T1_replica-1.pdbqt
	CHEMBL1624459	CHEMBL4303835	SARS-CoV-2		Z2613463853_6_T1_replica-1.pdbqt
	CHEMBL1423	CHEMBL4303835	SARS-CoV-2		Z743321630_1_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1139171183_3_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1952933636_1_T1_replica-1.pdbqt
	CHEMBL2158685	CHEMBL4303835	SARS-CoV-2		Z1139171183_2_T1_replica-1.pdbqt

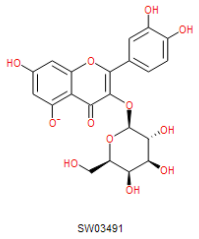
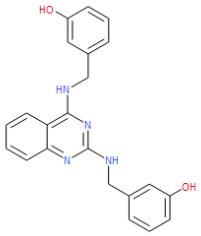
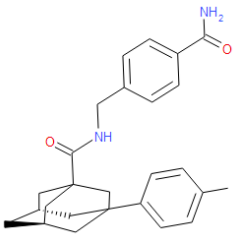
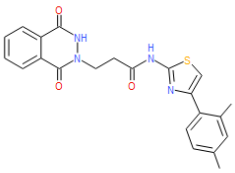
Studies on* Sweetlead library *(library 2).

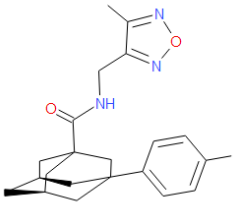
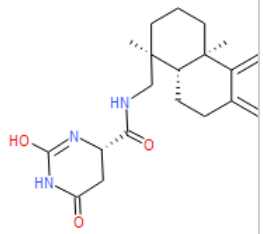
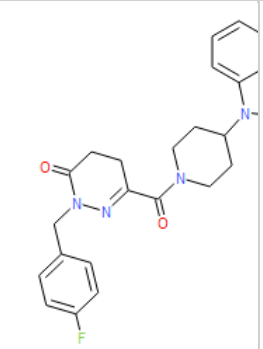
1. 325 compounds with score <= -7.
2. The first ones (best docking score) correspond to **Quercetin** and **Isoquercitrin** which was described as antiviral and recently potentially as therapy for the prevention and treatment of COVID-19[17]. These compounds should be in the first list.

So as suggested, we proposed currently the following top list of compounds beside the 10000 compounds selected.

Top list of compounds (see Table 2) from Table 1 and sweetlead data.

Table 2. Selection of a top list compound (7 compounds)

Molecule	Name	Canonical_smiles	InChI
	SW03491	OC[C@H]1O[C@@H](OC2=C(Oc3cc(O)cc([O-])c3C2=O)c4ccc(O)c(O)c4)[C@H](O)[C@@H](O)[C@H]1O	InChI=1S/C21H20O12/c22-6-13-15(27)17(29)18(30)21(32-13)33-20-16(28)14-11(26)4-8(23)5-12(14)31-19(20)7-1-2-9(24)10(25)3-7/h1-5,13,15,17-18,21-27,29-30H,6H2/p-1/t13-,15+,17+,18-,21+/m1/s1
	Z1587196385_1_T1_replica-1	Oc1cccc(CNc2nc(NCc3cccc(O)c3)c4ccccc4n2)c1	InChI=1S/C22H20N4O2/c27-17-7-3-5-15(11-17)13-23-21-19-9-1-2-10-20(19)25-22(26-21)24-14-16-6-4-8-18(28)12-16/h1-12,27-28H,13-14H2,(H2,23,24,25,26)
	Z28944530_2_T1_replica-1	Cc1ccc(cc1)C23C[C@@H]4C[C@@H](CC(C4)(C2)C(=O)NCc5ccc(cc5)C(=O)N)C3	InChI=1S/C26H30N2O2/c1-17-2-8-22(9-3-17)25-11-19-10-20(12-25)14-26(13-19,16-25)24(30)28-15-18-4-6-21(7-5-18)23(27)29/h2-9,19-20H,10-16H2,1H3,(H2,27,29)(H,28,30)/t19-,20+,25?,26?
	Z88459020_1_T2_replica-1	Cc1ccc(c(C)c1)c2csc(NC(=O)CCN3NC(=O)c4ccccc4C3=O)n2	InChI=1S/C22H20N4O3S/c1-13-7-8-15(14(2)11-13)18-12-30-22(23-18)24-19(27)9-10-26-21(29)17-6-4-3-5-16(17)20(28)25-26/h3-8,11-12H,9-10H2,1-2H3,(H,25,28)(H,23,24,27)

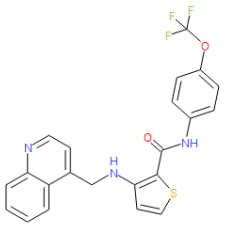
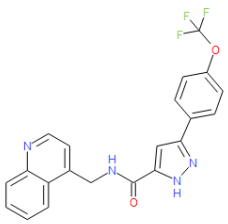
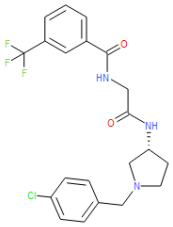
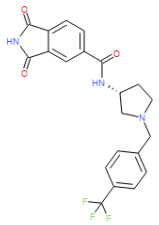
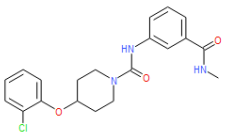
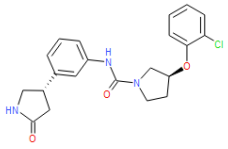
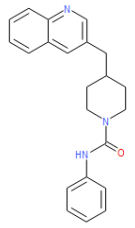
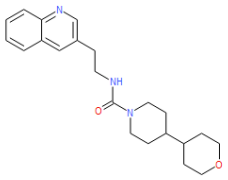
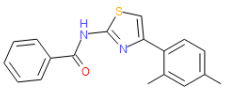
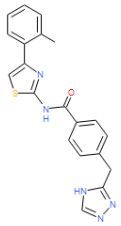
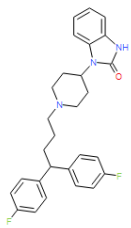
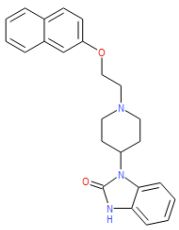
Molecule	Name	Canonical_smiles	InChI
	Z1229914077_3_T1_replica-1	Cc1ccc(cc1)C23C[C@H]4C[C@@H](CC(C4)(C2)C(=O)NCc5nonc5C)C3	InChI=1S/C22H27N3O2/c1-14-3-5-18(6-4-14)21-8-16-7-17(9-21)11-22(10-16,13-21)20(26)23-12-19-15(2)24-27-25-19/h3-6,16-17H,7-13H2,1-2H3,(H,23,26)/t16-,17-,21?,22?/m1/s1
	PV-001831411371_3_T3_replica-1	CC(C)c1ccc2c(CC[C@@H]3[C@@](C)(CNC(=O)[C@@H]4CC(=O)NC(=N4)O)CCC[C@@]23C)c1	InChI=1S/C25H35N3O3/c1-15(2)16-6-8-18-17(12-16)7-9-20-24(3,10-5-11-25(18,20)4)14-26-22(30)19-13-21(29)28-23(31)27-19/h6,8,12,15,19-20H,5,7,9-11,13-14H2,1-4H3,(H,26,30)(H2,27,28,29,31)/t19-,20+,24+,25-/m0/s1
	Z743321630_1_T1_replica-1	Fc1ccc(CN2N=C(CCC2=O)C(=O)N3CCC(CC3)N4C(=O)Nc5ccccc45)cc1	InChI=1S/C24H24FN5O3/c25-17-7-5-16(6-8-17)15-29-22(31)10-9-20(27-29)23(32)28-13-11-18(12-14-28)30-21-4-2-1-3-19(21)26-24(30)33/h1-8,18H,9-15H2,(H,26,33)

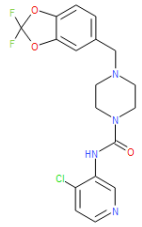
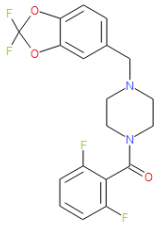
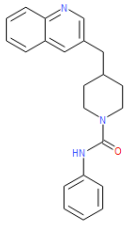
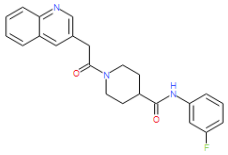
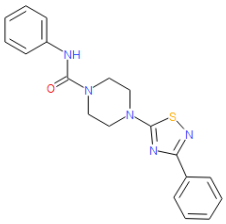
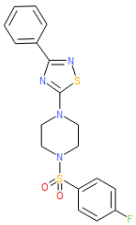
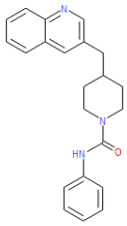
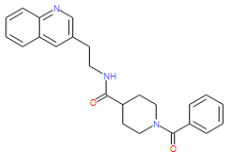
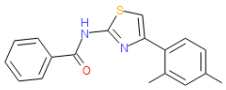
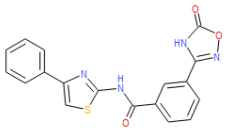
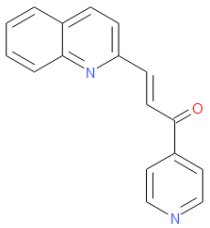
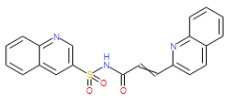
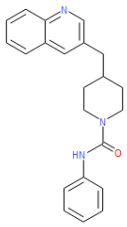
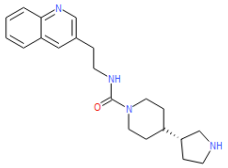
Target 2 :

Studies on Real Database (library 1).

1. First HTVS
    a. Scoring data for 567746375 compounds.
        i. Statistics for the scores :
            1. Min : -13.6 / Max : 355.7 / Median : -4.8
    b. Cutoff at -10.1 for the first selection : 1106522 compounds (at this step, we wanted to select around 1 million compounds).
2. Second HTVS on the 1106522 compounds (selection of around 50000 compounds).
    a. Cutoff at -10.7 for the second selection : 50616 compounds selected.
3. Clustering of the final results (towards 10000 compounds).
    a. 12227 compounds selected (classification in function of the docking score).
4. Similarities between the 50616 compounds and the 271 compounds described as active towards SARS-COV2 (Chembl4303835).
    a. 13 compounds with similarities >= 0.5 (see Table 3).
    b. These compounds (columns : HTVS / Name Real) could be taken in the first list.

Table 3. Description of 12 compounds (HTVS and Name Real column) similar to active derivatives (Covid_plus and Chembl_ID column) towards SARS-COV2

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL1614710	CHEMBL4303835	SARS-CoV-2		PV-001863644225_1_T1_replica-1.pdbqt
	CHEMBL255408	CHEMBL4303835	SARS-CoV-2		Z1205860345_1_T1_replica-1.pdbqt
	CHEMBL469169	CHEMBL4303835	SARS-CoV-2		Z1499837078_2_T1_replica-1.pdbqt
	CHEMBL560590	CHEMBL4303835	SARS-CoV-2		Z2928083636_1_T1_replica-1.pdbqt
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		PV-001830923473_1_T3_replica-1.pdbqt
	CHEMBL1423	CHEMBL4303835	SARS-CoV-2		Z371817720_1_T1_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL4297294	CHEMBL4303835	SARS-CoV-2		PV-001917351551_1_T1_replica-1.pdbqt
	CHEMBL560590	CHEMBL4303835	SARS-CoV-2		Z1786844327_1_T1_replica-1.pdbqt
	CHEMBL460273	CHEMBL4303835	SARS-CoV-2		Z1267751013_1_T1_replica-1.pdbqt
	CHEMBL560590	CHEMBL4303835	SARS-CoV-2		Z2961864866_1_T1_replica-1.pdbqt
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		PV-001821351532_1_T1_replica-1.pdbqt
	CHEMBL3105836	CHEMBL4303835	SARS-CoV-2		Z1843812624_2_T1_replica-1.pdbqt
	CHEMBL560590	CHEMBL4303835	SARS-CoV-2		Z3046389202_2_T1_replica-1.pdbqt

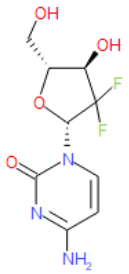
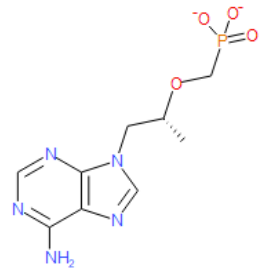
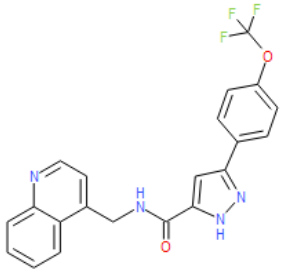
Studies on* Sweetlead library *(library 2).

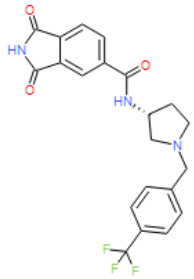
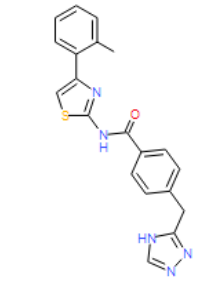
1. 178 compounds with score <= -7.
2. In these 178 compounds, several antiviral compounds appear like cidofovir, **gemcitabine** (SW01334) or **Tenofovir** (SW01591). Two recent publications have shown the interest of **gemcitabine**[18] and **Tenofovir**[19] diphosphate for SARS-COV2. These compounds can be considered in the first list. Clinical studies start currently with the association Emtricitabine / Tenofovir (see clinical trials (COVID / Tenofovir for the search terms (no result for COVID / gemcitabine)).

So as suggested, we proposed currently the following top list of compounds beside the 10000 compounds selected.

Top list of compounds (see* Table 4*) from* Table 3 *and sweetlead data.

Table 4. Selection of a top list compound (5 compounds)

HTVS	Name	smiles	InChI
 SW01334	SW01334	c1cn(c(=O)nc1N)[C@H]2C([C@@H]([C@H](O2)CO)O)(F)F	InChI=1S/C9H11F2N3O4/c10-9(11)6(16)4(3-15)18-7(9)14-2-1-5(12)13-8(14)17/h1-2,4,6-7,15-16H,3H2,(H2,12,13,17)/t4-,6-,7-/m1/s1
 SW01591	SW01591	C[C@H](Cn1cnc2c1ncnc2N)OCP(=O)([O-])[O-]	InChI=1S/C9H14N5O4P/c1-6(18-5-19(15,16)17)2-14-4-13-7-8(10)11-3-12-9(7)14/h3-4,6H,2,5H2,1H3,(H2,10,11,12)(H2,15,16,17)/p-2/t6-/m1/s1
 PV-001863644225_1_T1_replica-1	PV-001863644225_1_T1	FC(F)(F)OC1=CC=C(C=C1)C1=NNC(=C1)C(=O)NCC1=CC=NC2=CC=CC=C12	InChI=1S/C21H15F3N4O2/c22-21(23,24)30-15-7-5-13(6-8-15)18-11-19(28-27-18)20(29)26-12-14-9-10-25-17-4-2-1-3-16(14)17/h1-11H,12H2,(H,26,29)(H,27,28)

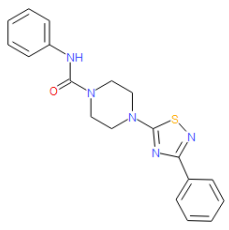
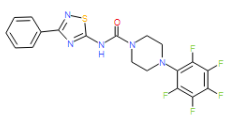
HTVS	Name	smiles	InChI
 Z1205860345_1_T1_replica-1	Z1205860345_1_T1	FC(F)(F)c1ccc(CN2CC[C@H](C2)NC(=O)c2ccc3C(=O)NC(=O)c3c2)cc1	InChI=1S/C21H18F3N3O3/c22-21(23,24)14-4-1-12(2-5-14)10-27-8-7-15(11-27)25-18(28)13-3-6-16-17(9-13)20(30)26-19(16)29/h1-6,9,15H,7-8,10-11H2,(H,25,28)(H,26,29,30)/t15-/m1/s1
 PV-001830923473_1_T3_replica-1	PV-001830923473_1_T3_replica-1	CC1=CC=CC=C1C1=CSC(NC(=O)C2=CC=C(CC3=NN=CN3)C=C2)=N1	InChI=1S/C20H17N5OS/c1-13-4-2-3-5-16(13)17-11-27-20(23-17)24-19(26)15-8-6-14(7-9-15)10-18-21-12-22-25-18/h2-9,11-12H,10H2,1H3,(H,21,22,25)(H,23,24,26)

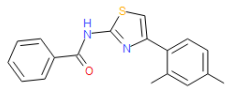
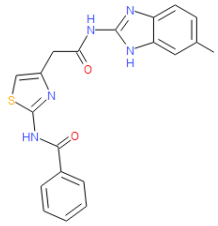
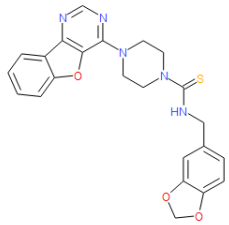
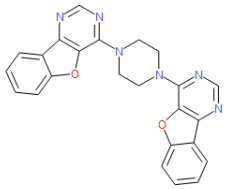
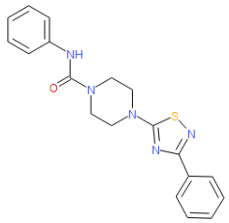
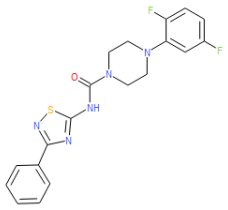
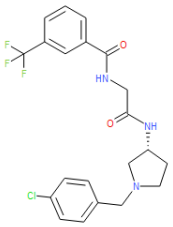
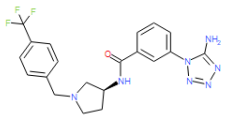
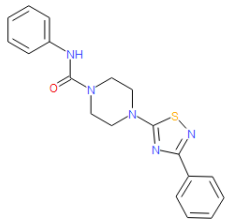
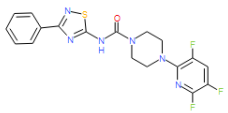
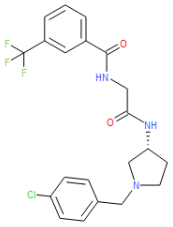
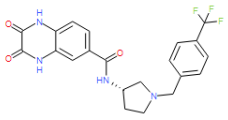
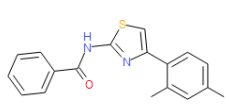
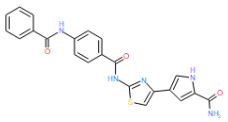
Target 3 :

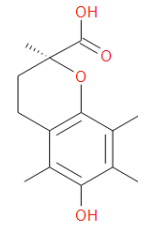
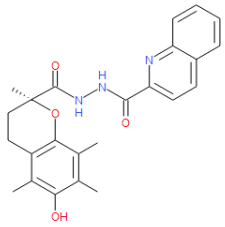
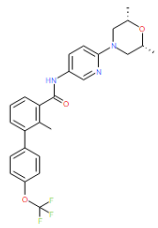
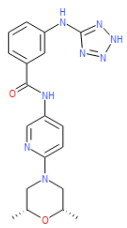
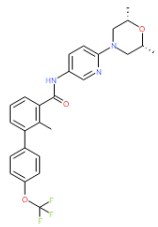
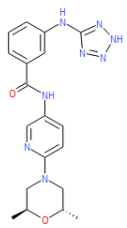
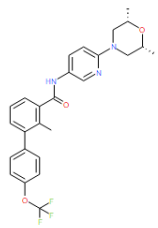
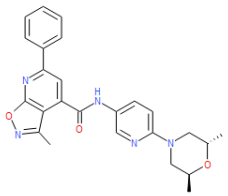
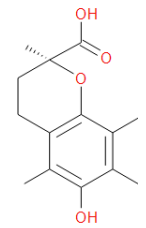
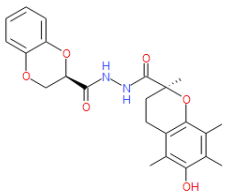
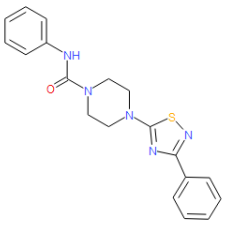
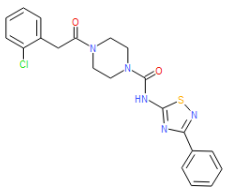
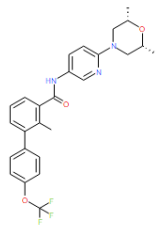
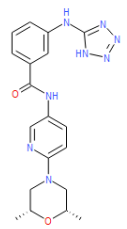
Studies on Real Database (library 1).

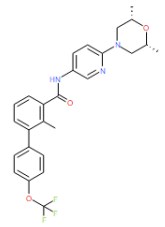
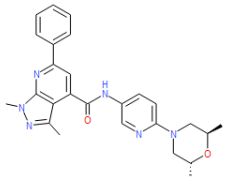
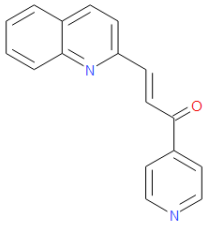
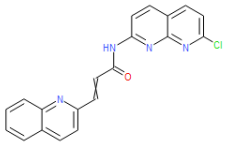
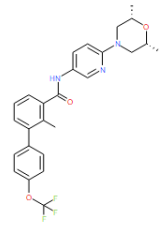
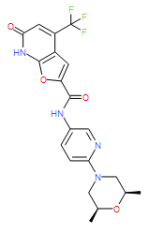
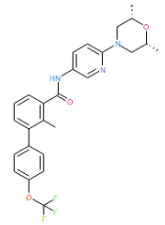
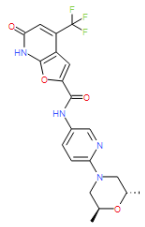
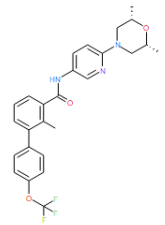
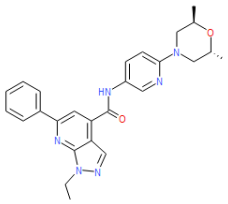
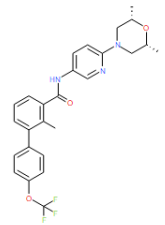
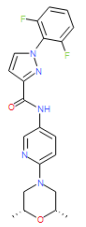
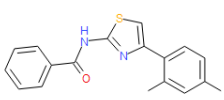
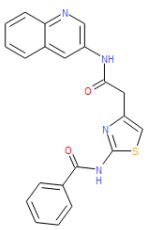
1. First HTVS
    a. Scoring data for 586252945 compounds.
        i. Statistics for the scores :
            1. Min : -11.8 / Max : 127.5 / Median : -6.7
    b. Cutoff at -9.0 for the first selection : 1185453 compound (at this step, we wanted to select around 1 million compounds).
2. Second HTVS on the 1185453 compounds (selection of around 50000 compounds).
    a. Cutoff at -9.6 for the second selection : 58407 compounds selected.
3. Clustering of the final results (towards 10000 compounds).
    a. 9009 compounds selected (classification in function of the docking score).
4. Similarities between the 58407 compounds and the 271 compounds described as active towards SARS-COV2 (Chembl4303835).
    a. 23 compounds with similarities >= 0.5 (see Table 5).
    b. These compounds (columns : HTVS / Name Real) could be taken in the first list.

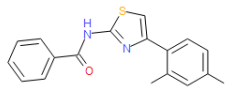
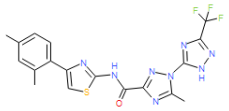
Table 5. Description of 23 compounds (HTVS and Name Real column) similar to active derivatives (Covid_plus and Chembl_ID column) towards SARS-COV2.

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL460273	CHEMBL4303835	SARS-CoV-2		PV-001942186131_1_T1_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		Z2229653598_1_T1_replica-1.pdbqt
	CHEMBL2103851	CHEMBL4303835	SARS-CoV-2		Z276622112_1_T1_replica-1.pdbqt
	CHEMBL460273	CHEMBL4303835	SARS-CoV-2		Z1348188744_1_T1_replica-1.pdbqt
	CHEMBL255408	CHEMBL4303835	SARS-CoV-2		PV-001832741228_2_T1_replica-1.pdbqt
	CHEMBL460273	CHEMBL4303835	SARS-CoV-2		Z1864631732_1_T1_replica-1.pdbqt
	CHEMBL255408	CHEMBL4303835	SARS-CoV-2		Z1205071489_2_replica-1.pdbqt
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		Z284455258_1_T1_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL1319514	CHEMBL4303835	SARS-CoV-2		PV-001820806192_2_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		PV-001846554085_1_T1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		PV-001846554085_2_T1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		Z453672734_3_T1_replica-1.pdbqt
	CHEMBL1319514	CHEMBL4303835	SARS-CoV-2		PV-001820381443_4_T1_replica-1.pdbqt
	CHEMBL460273	CHEMBL4303835	SARS-CoV-2		Z859649658_1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		PV-001846554085_1_T2_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		Z453672616_2_T1_replica-1.pdbqt
	CHEMBL3105836	CHEMBL4303835	SARS-CoV-2		PV-001813614032_2_T1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		Z768485750_1_T1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		Z768485750_2_T1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		Z453675604_2_T1_replica-1.pdbqt
	CHEMBL2105737	CHEMBL4303835	SARS-CoV-2		Z1343595000_1_T1_replica-1.pdbqt
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		Z240714602_1_T1_replica-1.pdbqt

covid_plus	ChEMBL ID	Target ChEMBL ID	Target Name	HTVS	Name Real
	CHEMBL471411	CHEMBL4303835	SARS-CoV-2		Z1325224923_1_T1_replica-1.pdbqt

Studies on ChEMBLDB for proteases (library 4).

1. Agreement between the 58407 compounds (HTVS) and the pharmacophores associated to the 53092 compounds actives on proteases (see point 3 in methods).
    a. 2341 compounds in agreement. These compounds are classified in the first part of the final list and in function of the docking score.
        i. On the 9009 compounds extracted previously (clustering of the 58407 compounds), 587 out of the 2341 compounds are present.
            1. The remaining 8422 compounds are classified in the second part of the final list and in function of the docking score.

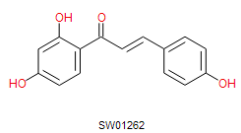
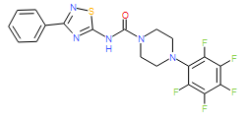
Studies on* Sweetlead library *(library 2).

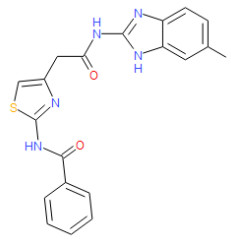
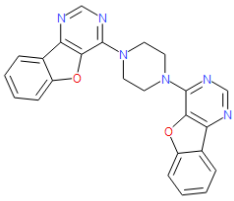
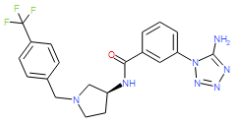
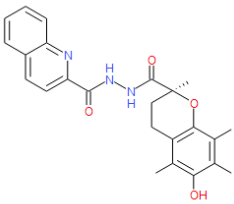
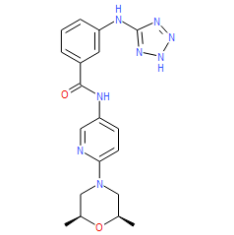
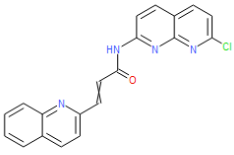
1. 213 compounds with score <= -7.
    a. We have a cluster of 60 derivatives with **1,isoliquiritigenin** as the compound with the best docking score in this cluster (described as a GABA-A benzodiapine receptor positive allosteric modulator). We have one publication with an application of this derivative as antiviral (The Flavonoid Isoliquiritigenin Reduces Lung Inflammation and Mouse Morbidity During Influenza Virus Infection[20]). So, this compound could be in the final list and was classified as first.

So as suggested, we proposed currently the following top list of compounds beside the 10000 compounds selected.

Top list of compounds (see Table 6) from Table 5 and sweetlead data.

Table 6. Selection of a top list compound (8 compounds)

Molecule	Name	Canonical_smiles	InChI
	SW01262	Oc1ccc(\C=C\C(=O)c2ccc(O)cc2O)cc1	InChI=1S/C15H12O4/c16-11-4-1-10(2-5-11)3-8-14(18)13-7-6-12(17)9-15(13)19/h1-9,16-17,19H/b8-3+
	PV-001942186131_1_T1_replica-1	Fc1c(F)c(F)c(N2CCN(CC2)C(=O)Nc3nc(ns3)c4ccccc4)c(F)c1F	InChI=1S/C19H14F5N5OS/c20-11-12(21)14(23)16(15(24)13(11)22)28-6-8-29(9-7-28)19(30)26-18-25-17(27-31-18)10-4-2-1-3-5-10/h1-5H,6-9H2,(H,25,26,27,30)

Molecule	Name	Canonical_smiles	InChI
	Z2229653598_1_T1_replica-1	Cc1ccc2nc(NC(=O)Cc3csc(NC(=O)c4ccccc4)n3)[nH]c2c1	InChI=1S/C20H17N5O2S/c1-12-7-8-15-16(9-12)23-19(22-15)24-17(26)10-14-11-28-20(21-14)25-18(27)13-5-3-2-4-6-13/h2-9,11H,10H2,1H3,(H,21,25,27)(H2,22,23,24,26)
	Z276622112_1_T1_replica-1	C1CN(CCN1c2ncnc3c2oc4ccccc34)c5ncnc6c5oc7ccccc67	InChI=1S/C24H18N6O2/c1-3-7-17-15(5-1)19-21(31-17)23(27-13-25-19)29-9-11-30(12-10-29)24-22-20(26-14-28-24)16-6-2-4-8-18(16)32-22/h1-8,13-14H,9-12H2
	PV-001832741228_2_T1_replica-1	Nc1nnnn1c2cccc(c2)C(=O)N[C@H]3CCN(Cc4ccc(cc4)C(F)(F)F)C3	InChI=1S/C20H20F3N7O/c21-20(22,23)15-6-4-13(5-7-15)11-29-9-8-16(12-29)25-18(31)14-2-1-3-17(10-14)30-19(24)26-27-28-30/h1-7,10,16H,8-9,11-12H2,(H,25,31)(H2,24,26,28)/t16-/m0/s1
	PV-001820806192_2_replica-1	Cc1c(C)c2O[C@@](C)(CCc2c(C)c1O)C(=O)NNC(=O)c3ccc4ccccc4n3	InChI=1S/C24H25N3O4/c1-13-14(2)21-17(15(3)20(13)28)11-12-24(4,31-21)23(30)27-26-22(29)19-10-9-16-7-5-6-8-18(16)25-19/h5-10,28H,11-12H2,1-4H3,(H,26,29)(H,27,30)/t24-/m0/s1
	PV-001846554085_1_T1_replica-1	C[C@@H]1CN(C[C@H](C)O1)c2ccc(NC(=O)c3ccc(Nc4nn[nH]n4)c3)cn2	InChI=1S/C19H22N8O2/c1-12-10-27(11-13(2)29-12)17-7-6-16(9-20-17)21-18(28)14-4-3-5-15(8-14)22-19-23-25-26-24-19/h3-9,12-13H,10-11H2,1-2H3,(H,21,28)(H2,22,23,24,25,26)/t12-,13+
	PV-001813614032_2_T1_replica-1	Clc1ccc2ccc(NC(=O)C=Cc3ccc4ccccc4n3)nc2n1	InChI=1S/C20H13ClN4O/c21-17-10-6-14-7-11-18(25-20(14)23-17)24-19(26)12-9-15-8-5-13-3-1-2-4-16(13)22-15/h1-12H,(H,23,24,25,26)

Other comments:

If you have some problems to get some compounds from Enamine, I can send you data resulting of the comparison between our results from real database and the Merck library (definition of a possible second list).

Concerning some data like the overall docking scores for the three targets, they can be accessed at the following site : https://osf.io/q5fk8/?view_only=36c34803fc9448fba6c45d27cbfc25ee

References :

(1) Gorgulla, C.; Boeszoermenyi, A.; Wang, Z.-F.; Fischer, P. D.; Coote, P. W.; Padmanabha Das, K. M.; Malets, Y. S.; Radchenko, D. S.; Moroz, Y. S.; Scott, D. A.; Fackeldey, K.; Hoffmann, M.; Iavniuk, I.; Wagner, G.; Arthanari, H. An Open-Source Drug Discovery Platform Enables Ultra-Large Virtual Screens. *Nature* **2020**, *580* (7805), 663–668. https://doi.org/10.1038/s41586-020-2117-z.

(2) Alhossary, A.; Handoko, S. D.; Mu, Y.; Kwoh, C.-K. Fast, Accurate, and Reliable Molecular Docking with QuickVina 2. *Bioinforma. Oxf. Engl.* **2015**, *31* (13), 2214–2216. https://doi.org/10.1093/bioinformatics/btv082.

(3) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.

(4) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759. https://doi.org/10.1021/jm030644s.

(5) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749. https://doi.org/10.1021/jm0306430.

(6) Métivier, J.-P.; Cuissart, B.; Bureau, R.; Lepailleur, A. The Pharmacophore Network: A Computational Method for Exploring Structure-Activity Relationships from a Large Chemical Data Set. *J. Med. Chem.* **2018**, *61* (8). https://doi.org/10.1021/acs.jmedchem.7b01890.

(7) Ghosh, A. K.; Brindisi, M.; Shahabi, D.; Chapman, M. E.; Mesecar, A. D. Drug Development and Medicinal Chemistry Efforts toward SARS-Coronavirus and Covid-19 Therapeutics. *ChemMedChem* **2020**, *15* (11), 907–932. https://doi.org/10.1002/cmdc.202000223.

(8) Jacobs, J.; Grum-Tokars, V.; Zhou, Y.; Turlington, M.; Saldanha, S. A.; Chase, P.; Eggler, A.; Dawson, E. S.; Baez-Santos, Y. M.; Tomar, S.; Mielech, A. M.; Baker, S. C.; Lindsley, C. W.; Hodder, P.; Mesecar, A.; Stauffer, S. R. Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(Tert-Butyl)-2-(N-Arylamido)-2-(Pyridin-3-Yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease. *J. Med. Chem.* **2013**, *56* (2), 534–546. https://doi.org/10.1021/jm301580n.

(9) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.

(10) Shimizu, J. F.; Martins, D. O. S.; McPhillie, M. J.; Roberts, G. C.; Zothner, C.; Merits, A.; Harris, M.; Jardim, A. C. G. Is the ADP Ribose Site of the Chikungunya Virus NSP3 Macro Domain a Target for Antiviral Approaches? *Acta Trop.* **2020**, *207*, 105490. https://doi.org/10.1016/j.actatropica.2020.105490.

(11) Huggins, D. Structural Analysis of Experimental Drugs Binding to the COVID-19 Target TMPRSS2. **2020**. https://doi.org/10.26434/chemrxiv.12315449.v1.

(12) Riggs, J. R.; Hu, H.; Kolesnikov, A.; Leahy, E. M.; Wesson, K. E.; Shrader, W. D.; Vijaykumar, D.; Wahl, T. A.; Tong, Z.; Sprengeler, P. A.; Green, M. J.; Yu, C.; Katz, B. A.; Sanford, E.; Nguyen, M.;

Cabuslay, R.; Young, W. B. Novel 5-Azaindole Factor VIIa Inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16* (12), 3197–3200. https://doi.org/10.1016/j.bmcl.2006.03.049.

(13) Novick, P. A.; Ortiz, O. F.; Poelman, J.; Abdulhay, A. Y.; Pande, V. S. SWEETLEAD: An In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PLOS ONE* **2013**, *8* (11), e79568. https://doi.org/10.1371/journal.pone.0079568.

(14) LigPrep | Schrödinger https://www.schrodinger.com/ligprep (accessed Jun 30, 2020).

(15) RedHill Biopharma Limited. *Opaganib, a Sphingosine Kinase-2 (SK2) Inhibitor in COVID-19 Pneumonia: A Randomized, Double-Blind, Placebo-Controlled Phase 2a Study, in Adult Subjects Hospitalized With SARS-CoV-2 Positive Pneumonia*; Clinical trial registration NCT04414618; clinicaltrials.gov, 2020.

(16) Shaare Zedek Medical Center. *Compassionate Use of Opaganib in Patients With Severe COVID-19*; Clinical trial registration NCT04435106; clinicaltrials.gov, 2020.

(17) Colunga Biancatelli, R. M. L.; Berrill, M.; Catravas, J. D.; Marik, P. E. Quercetin and Vitamin C: An Experimental, Synergistic Therapy for the Prevention and Treatment of SARS-CoV-2 Related Disease (COVID-19). *Front. Immunol.* **2020**, *11*, 1451. https://doi.org/10.3389/fimmu.2020.01451.

(18) Zhang, Y.-N.; Zhang, Q.-Y.; Li, X.-D.; Xiong, J.; Xiao, S.-Q.; Wang, Z.; Zhang, Z.-R.; Deng, C.-L.; Yang, X.-L.; Wei, H.-P.; Yuan, Z.-M.; Ye, H.-Q.; Zhang, B. Gemcitabine, Lycorine and Oxysophoridine Inhibit Novel Coronavirus (SARS-CoV-2) in Cell Culture. *Emerg. Microbes Infect.* **2020**, *9* (1), 1170–1173. https://doi.org/10.1080/22221751.2020.1772676.

(19) Ju, J.; Li, X.; Kumar, S.; Jockusch, S.; Chien, M.; Tao, C.; Morozova, I.; Kalachikov, S.; Kirchdoerfer, R. N.; Russo, J. J. Nucleotide Analogues as Inhibitors of SARS-CoV Polymerase. *bioRxiv* **2020**. https://doi.org/10.1101/2020.03.12.989186.

(20) Traboulsi, H.; Cloutier, A.; Boyapelly, K.; Bonin, M.-A.; Marsault, É.; Cantin, A. M.; Richter, M. V. The Flavonoid Isoliquiritigenin Reduces Lung Inflammation and Mouse Morbidity during Influenza Virus Infection. *Antimicrob. Agents Chemother.* **2015**, *59* (10), 6317–6327. https://doi.org/10.1128/AAC.01098-15.