Team name	Belarus
Team member(s) (firstnamelastname; …)	Alexander Tuzikov , Alexander Andrianov, Yuri Kornoushenko, Hanna Karpenka, Ivan Bosko, Nikita Shuldov, Artsemi Yushkevich, Konstantin Furs
Affiliation	The National Academy of Sciences of Belarus
Contact email	rfe.karpenko@gmail.com
Contact phone number (optional)	+375296665458
Protein targets (for example: 3CLPro/Nsp5, BoAT1, Fc Receptor, Furin, IL6R, M protein, Nspx, OrfXx, N, E, etc…) | 3 required	3CLpro (binded active site), 3CLpro (unbinded active site), spike

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:

Screening		
Pharmit provides an online (http://pharmit.csb.pitt.edu/), interactive environment for the virtual screening of large compound databases using pharmacophores, molecular shape and energy minimization.		
3CLpro (binded active site)	3CLpro (unbinded active site)	*Spike*
The pharmacophore model was built on the basis of the interaction of the X77 inhibitor with the main protease SARS-CoV2. Multiple pharmacophore models were created.	*The pharmacophore model was built on the basis of the interaction of the X77 inhibitor with the main protease SARS-CoV2. Several pharmacophore models were created. The first model includes three aromatic and one hydrophobic pharmacophore.*	*The pharmacophore model was built on the basis of the interaction of Lys353 ACE2 receptor residue with S1 protein residues. The model includes one hydrogen donor, one hydrogen acceptor and three hydrophobic interactions.*
The selected databases were searched for compounds that match the specified pharmacophore using the Pharmer search technology. The best screening compound was used to create a new pharmacophore model. A repeated search was performed on the selected databases. Energy minimization of results was used to optimize both the pose and conformation of identified hits with respect to the provided receptor using the AutoDock Vina scoring function and smina, a fork of AutoDock Vina with enhanced minimization functionality. Minimized results were sorted by affinity.		
The complete set of 870000 minimized compound structures was saved, including scoring annotations, as a compressed SDF structure file.	*The complete set of 870000 minimized compound structures was saved, including scoring annotations, as a compressed SDF structure file.*	*The complete set of 17000 minimized compound structures was saved, including scoring annotations, as a compressed SDF structure file.*
Docking		

<table>
<tr><td colspan="3">Before docking, hydrogen atoms were added to the ligand and receptor structures and their optimization was performed in the UFF force field. For this purpose, the Open Babel program (http://openbabel.org/wiki/Main_Page) was used.</td></tr>
<tr><td>3CLpro (binded active site)</td><td>3CLpro (unbinded active site)</td><td>Spike</td></tr>
<tr><td>Molecular docking of 870000 compounds was performed using the Quick Vina 2 program (https://omictools.com/quickvina-tool) with a conformational coverage parameter of 10.</td><td>Molecular docking of 870000 compounds was performed using the Quick Vina 2 program (https://omictools.com/quickvina-tool) with a conformational coverage parameter of 10.
The dissociation constant for the top 10,000 compounds was converted from the binding energies after docking.
The pharmacophore model was built on the basis of the interaction of the X77 inhibitor with the main protease SARS-CoV2</td><td>Molecular docking of 17,000 compounds was performed using the Quick Vina 2 program (https://omictools.com/quickvina-tool) with a conformational coverage parameter of 10.</td></tr>
<tr><td colspan="3">Dynamics</td></tr>
<tr><td colspan="2">3CLpro (binded active site)/ 3CLpro (unbinded active site)</td><td>Spike</td></tr>
<tr><td colspan="2">In the case of dynamic ligand MPro models, the values of binding energy were calculated with Amber18 using the MM/GBSA method . The calculations were made for 200 snapshots extracted from the final 40 ns of the MD trajectories, by keeping the snapshots every 0.2 ns. The polar solvation energies were computed in continuum solvent using Poisson-Boltzmann continuum-solvation model with ionic strength of 0.10. The non-polar terms were estimated using solvent accessible surface areas.</td><td>In the case of dynamic ligand Spike S1 models, the values of binding energy were calculated with Amber18 using the MM/GBSA method . The calculations were made for 200 snapshots extracted from the final 40 ns of the MD trajectories (overall 50 ns), by keeping the snapshots every 0.2 ns. The polar solvation energies were computed in continuum solvent using Poisson-Boltzmann continuum-solvation model with ionic strength of 0.10. The non-polar terms were estimated using solvent accessible surface areas.</td></tr>
<tr><td colspan="3">ML</td></tr>
<tr><td colspan="3">MACCS molecular fingerprints were obtained from SDF structure files and free binding energy was parsed using docking results. Fingerprints for each compound were matched with corresponding binding energy to form a dataset used for training of an adversarial generative autoencoder. The training was conducted in a semi-supervised manner, where fingerprints were fed to the encoder, while binding energy was fed to the special corresponding neuron on the latent layer, and decoder used both compressed representation obtained by encoder AND binding energy to restore original fingerprints. Latent layer was additionally discriminated with a normal distribution, to enforce</td></tr>
</table>

encoder to make meaningful compressed representation. Later, the model was used to generate new molecular descriptors (fingerprints) with a preset property of binding energy. Namely, in order to generate new fingerprints, numbers from normal distribution were sampled to the latent layer, while the neuron responsible for binding energy was given a threshold value to generate with (e.g. -10 kkal/mol).
For the generated fingerprints a similarity search was conducted among ZINC library compounds, using L1 distance as a metric. Closest compounds were subjected to docking procedures.
Rdkit python package was used for preprocessing, MACCS fingerprints generation. Tensorflow 2.1 for python was used as a deep learning framework.

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: **3CLpro (binded active site).**

The main protease of the SARS-CoV2 virus (3CLpro) controls the maturation of viral proteins. Therefore, the ability to inhibit the activity of 3CLpro in order to block the stage of maturation of viral proteins is highly relevant.
The crystal structure of the 3CLpro protein in complex with the X77 inhibitor, determined by X-ray diffraction analysis, was borrowed from the protein data bank ((PDB) code 3W63; https://www.rcsb.org/structure/6W63).
The resolution of the complex is 2.10 Å and is the best for the complex 3CLpro / X77.

Target 2: **3CLpro (unbinded active site)**

The main protease of the SARS-CoV2 virus (3CLpro) controls the maturation of viral proteins. Therefore, the ability to inhibit the activity of 3CLpro in order to block the stage of maturation of viral proteins is highly relevant.
The crystal structure of the free 3CLpro protease form determined by x-ray diffraction analysis was taken from the protein database ((PDB) code 6Y84; https://www.rcsb.org/structure/6Y84).
The resolution of the complex is 1.39 Å and is best for the unbound form of 3CLpro.

Target 3: **spike**

Protein S1 is required for SARS-CoV2 penetration into sensitive cells. Therefore, its inhibition in order to block the stage of virus penetration is highly relevant.
The crystal structure of the S1 protein in complex with the ACE2 protein, determined by X-ray diffraction analysis, was taken from the protein database ((PDB) code 6M0J; https://www.rcsb.org/structure/6M0J).
For a complex of S1 protein with ACE2 protein, a resolution of 2.45 Å is one of the best found in PDB.

Section 3: libraries

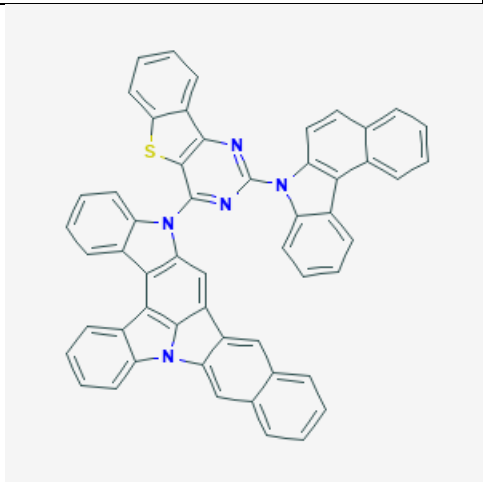
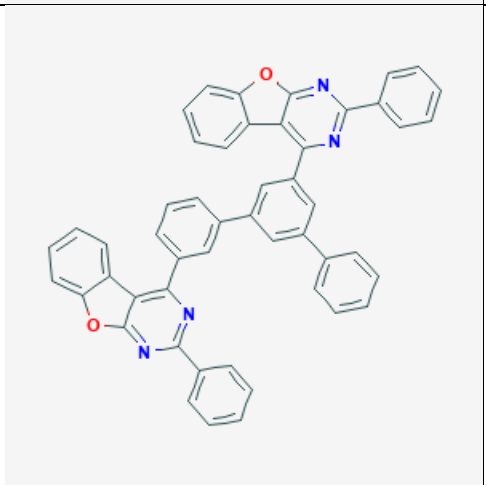
Describe which libraries you have used, how they were combined, if any compounds were removed / added, why additions are relevant, any unique features of your library, etc. Please provide the sources you obtained the libraries from (if publicly available).Describe the procedure of data preparation (removal 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.

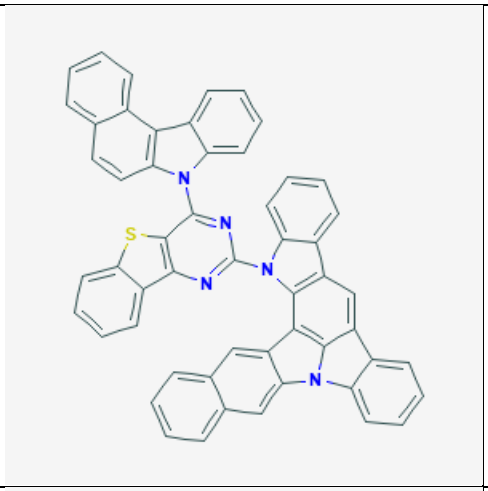
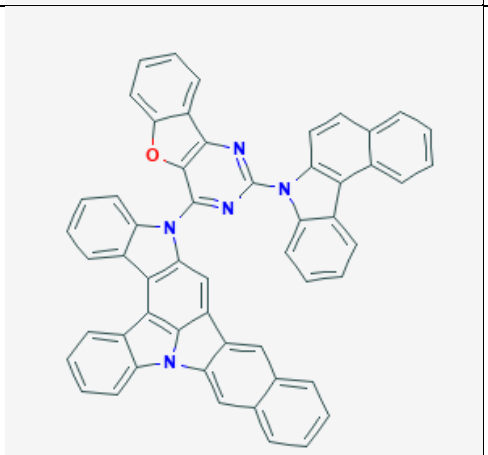
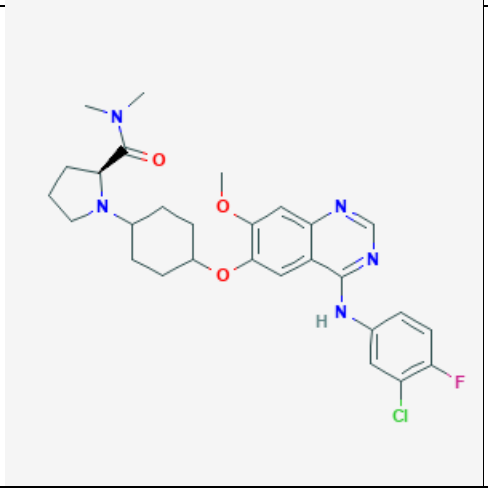
All the libraries (Pubchem, Molport, ZINC, ChemDiv etc.), integrated into the Pharmit web server, were used for virtual screening.

Section 4: results

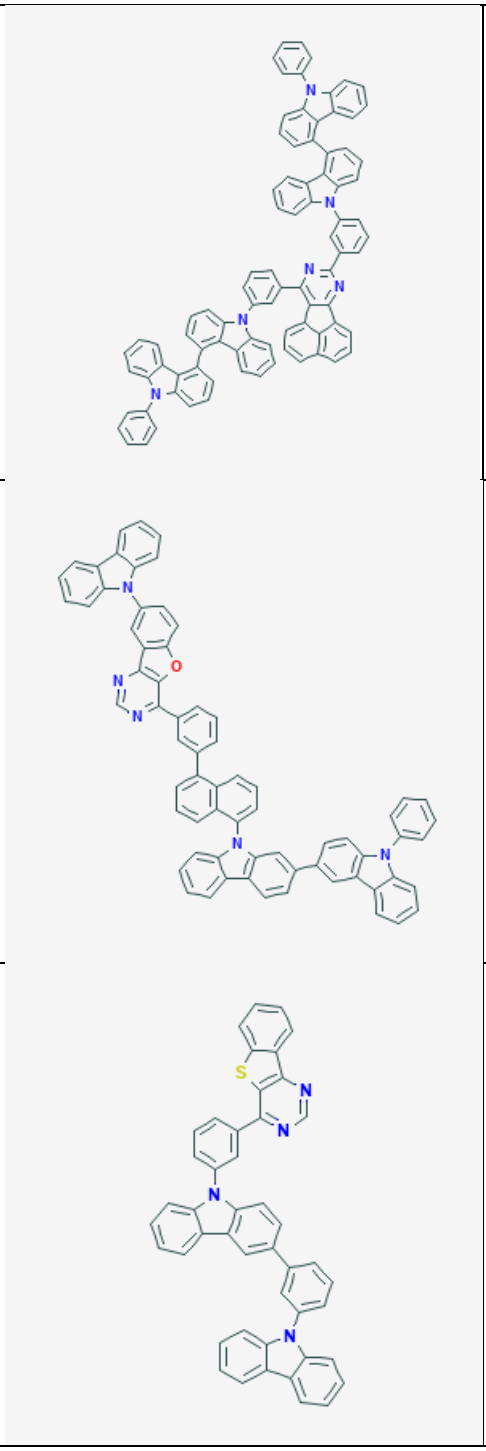
Briefly describe you key findings, any interesting trends in your data, a description of your top 5 compounds for each target.If possible, provide a link to a code and/or data repository. Please do not submit randomly selected compounds!

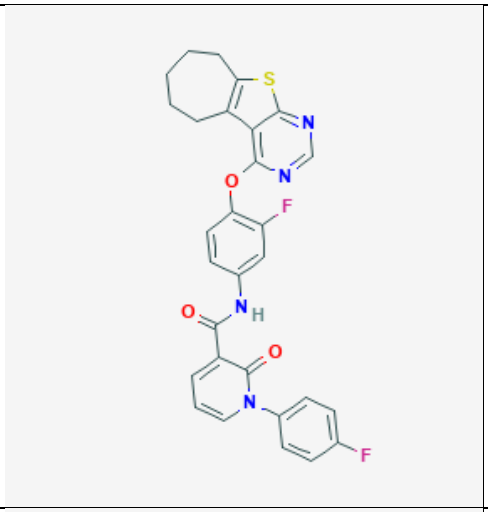
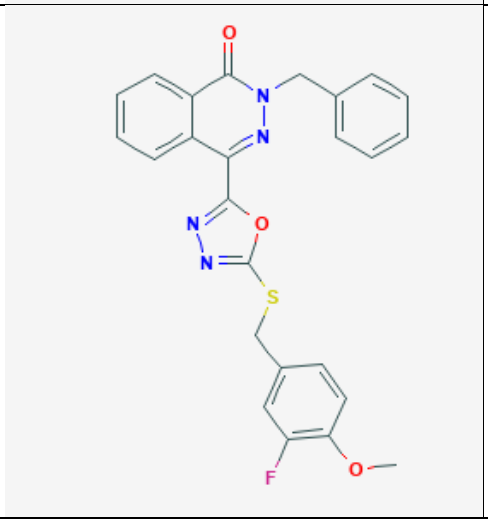
Results:

TOP 5 3CLpro (binded active site)					
Data base ID	Ligand	Vina (kcal/mol)	NNScore2.0 (nm)	Amber ΔH (kcal/mol)	Link
PubChem – 130258290		-13.4	12.7	-56.8±5.1	https://pubchem.ncbi.nlm.nih.gov/compound/130258290
PubChem – 122489503		-13.4	3.9	-62.8±5.1	https://pubchem.ncbi.nlm.nih.gov/compound/122489503

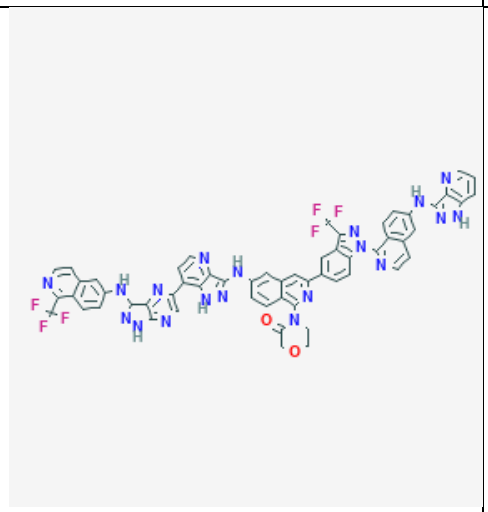
PubChem – 130258575		-13.9	326.5	-62.4±4.2	https://pubchem.ncbi.nlm.nih.gov/compound/130258575
PubChem – 130258293		-13.6	4.2	-59.3±4.4	https://pubchem.ncbi.nlm.nih.gov/compound/130258293
PubChem – 591581887		-10.0	0.034	-56.2 ± 3.1	https://pubchem.ncbi.nlm.nih.gov/compound/59158187

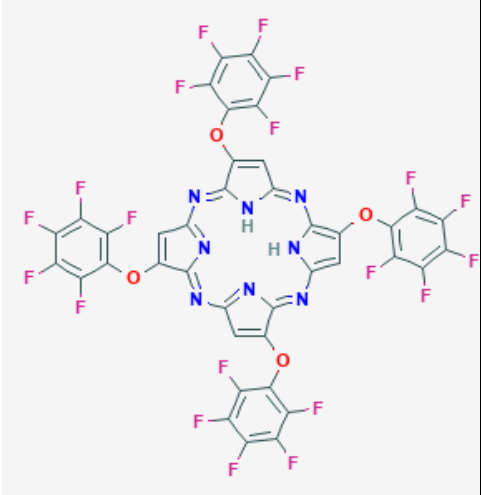
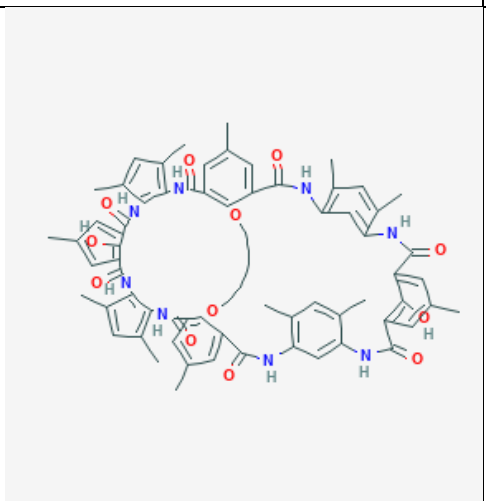
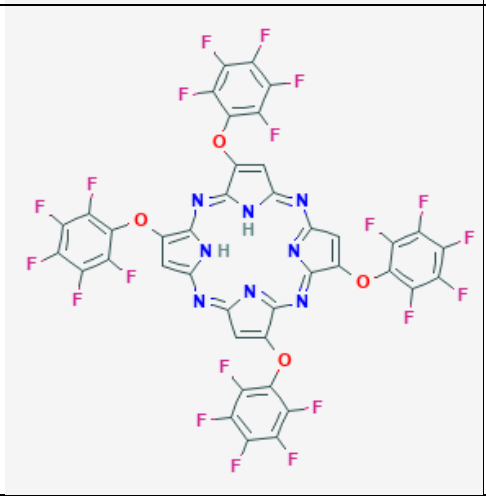
TOP 5 3CLpro (unbinded active site)					
Data base ID	Ligand	Vina (kcal/mol)	NNScore2.0 (nm)	Amber ΔH (kcal/mol)	Link

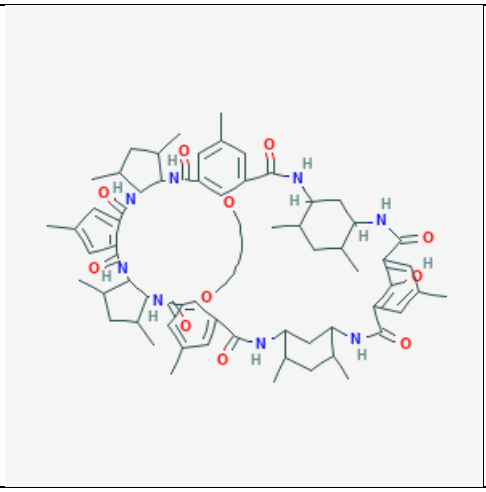
PubChem - 122466316		-11.5	2.0	-53.6 ± 6.0	https://pubchem.ncbi.nlm.nih.gov/compound/122466316
PubChem - 135163806		-11.7	22.5	-58.7 ± 8.6	https://pubchem.ncbi.nlm.nih.gov/compound/135163806
PubChem - 118328152		-11.4	3.8	-59.4 ± 4.3	https://pubchem.ncbi.nlm.nih.gov/compound/118328152

CHEMBL1813644		-8.9	0.014	-58.3 ± 7.2	https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL1813644/
PubChem-24329117		-8.2	0.059	-57.1 ± 4.4	https://pubchem.ncbi.nlm.nih.gov/compound/24329117

TOP 5 Spike

Data base ID	Ligand	Vina (kcal/mol)	NNScore2.0 (nm)	Amber ΔH (kcal/mol)	Link
PubChem-137551823		-12.6	479410.00	-45.1 ± 5.7	https://pubchem.ncbi.nlm.nih.gov/compound/137551823

PubChem - 11051446		-10.2	1.23	-37.9 ± 4.2	https://pubchem.ncbi.nlm.nih.gov/compound/11051446
PubChem - 89269379		-9.6	32000	-48.1 ± 5.3	https://pubchem.ncbi.nlm.nih.gov/compound/59847533
PubChem - 10996796		-10.0	0.978	-37.9 ± 3.5	https://pubchem.ncbi.nlm.nih.gov/compound/10996796

PubChem - 59847533		-9.5	**3210**	-44.0 ± 3.6	https://pubchem.ncbi.nlm.nih.gov/compound/89269379

Other comments: