

Structure-based pharmacophore modeling and virtual screening using mono- and polyfunctional fragment bases

Maksim V. Sukhorukov, Dmitry A. Shulga, Igor A. Morgunov,
Nikita N. Ivanov and Vladimir A. Palyulin

Contents

Table S1. Set of active and inactive molecules (ValSet) from BindingDB.	S2
Table S2. Coordinates of Grid Boxes for molecular docking using MFB and PFB in Auto Dock Vina.....	S2
Figure S1. The workflow of the experiment.....	S2
Table S3. Parameters used for DBSCAN modes clusterization for MFB and PFB.	S3
Table S4. Probes feature mapping for MFB and PFB.	S3
Figure S2. Pharmacophore features clusters based on results of molecular docking for MFB and PFB located in the binding site of neuraminidase H1N1 complexed with laninamivir.....	S5
Figure S3. SBP models, based on pharmacophore feature clusters, for MFB, MFB-based SBP with aromatic pharmacophore feature and PFB.	S5
Table S5. Metrics based on confusion matrices for MFB- and PFB-based pharmacophores.....	S6
Figure S4. Confusion matrices for SBP models tested on validation set for MFB and PFB.....	S7
Figure S5. Pharmacophore models build in Pharmit.	S8
Figure S6. Model of Ph4M07Ar made in Pharmit using settings described in Table S4.	S9
Figure S7. Distribution of physicochemical properties of compounds with neuraminidase H1N1 bioassays from Binding DB, calculated using RDKit Descriptor Calculation node in KNIME.	S9
Table S6. ChEMBL DB VS settings for SBP Ph4M07Ar used in Pharmit for PDB ID 3TI3, chain A, without ligand (based on Lipinski's rule of five and Figure S7).	S10
Table S7. Plausible influenza A virus neuraminidase inhibitors in ChEMBL DB, identified by SBP model Ph4M07Ar with aromatic pharmacophore feature.	S10
Figure S8. Affinity distribution after re-docking using GNINA for drugs and high potential ligands found by virtual screening using Pharmit and Ph4M07Ar SBP model.	S11
Table S8. Ligands with GNINA CNN energy and ligand efficiency values obtained after re-docking of molecules identified through virtual screening of the ChEMBL database using Ph4M07Ar SBP model in Pharmit.	S12
Figure S9. Alignment of co-crystallized and Pharmit minimized conformers of laninamivir, zanamivir, oseltamivir and oseltamivir carboxylate.....	S17

Table S1. Set of active and inactive molecules (ValSet) from BindingDB (<https://www.bindingdb.org/>, last update on 31.12.2025, accessed on 1 January 2026) made for SBP validation. The set is available online on Pharmit Contributed Libraries (“Neuraminidase_H1N1_10-10000_nM_ValidSet_3D”).

CSV-file “Table_S1_Validation_Set_VSet.csv”:

https://qsar.chem.msu.ru/~shulga/NA_SBP/Table_S1_Validation_Set_VSet.csv

on https://qsar.chem.msu.ru/~shulga/NA_SBP/.

Table S2. Coordinates of Grid Boxes for molecular docking using MFB and PFB in Auto Dock Vina.

PDB ID	Name	Form	Coordinates of binding site's center		
			x	y	z
3TI3	Crystal structure of 2009 pandemic H1N1 neuraminidase complexed with laninamivir	Holo	29.39	14.39	-20.60
3TI5	Crystal structure of 2009 pandemic H1N1 neuraminidase complexed with Zanamivir	Holo	29.45	14.45	-20.78
3TI6	Crystal structure of 2009 pandemic H1N1 neuraminidase complexed with oseltamivir	Holo	-28.95	14.27	20.85
7S0I	Crystal Structure of N1 Neuraminidase from A/Michigan/45/2015(H1N1)	Apo	15.79	5.45	6.01

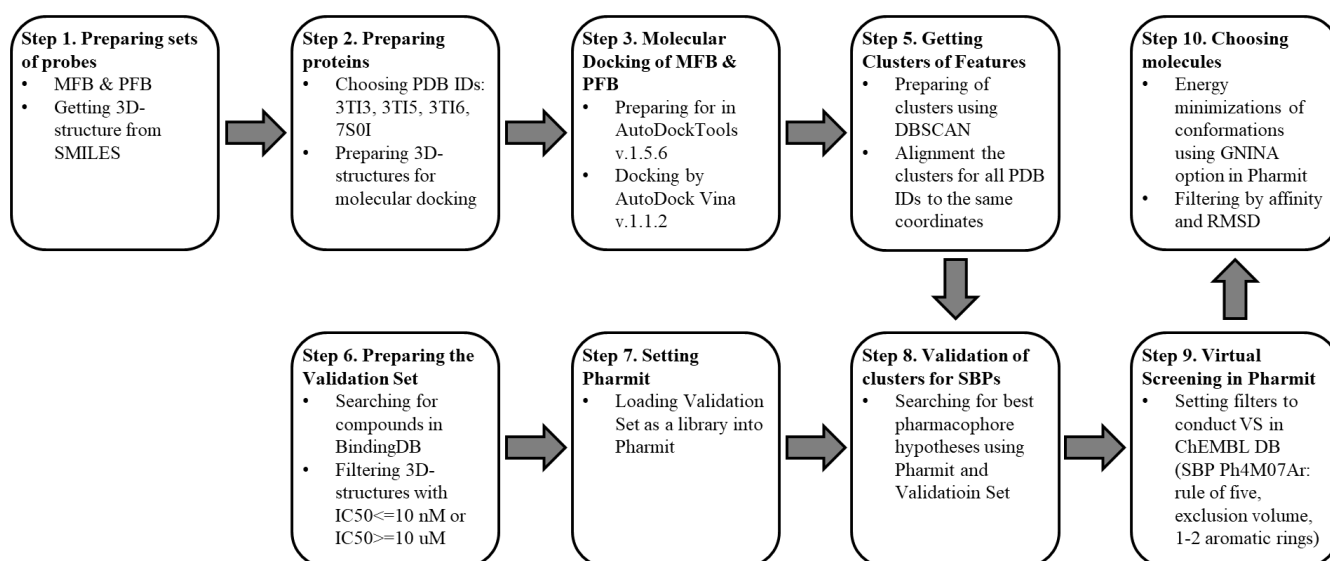


Figure S1. The workflow of the experiment.

Table S3. Parameters used for DBSCAN modes clusterization for MFB and PFB (HYD – hydrophobic, HBA – H-bond acceptor, HBD – H-bond donor, AROM – aromatic feature). According to higher variability of binding modes, just few clusters for PFB were obtained with min_samples = 3, and their performance was low. So, min_samples = 2 was chosen for PFB.

Parameter	Feature	MFB	PFB
min_samples	-	3	2
eps, Å	HYD	1.8	1.8
	HBA	1.2	1.2
	HBD	1.2	1.2
	AROM	1.5	1.5
algorithm	-	auto	auto
noise	-	points without clusters	points without clusters

Table S4. Probes feature mapping for MFB and PFB (HYD – hydrophobic, HBA – H-bond acceptor, HBD – H-bond donor, AROM – aromatic feature).

Name	Probe	Atom SMARTS	Feature	Is Considered?
MFC	CH ₄	C	HYD	Yes
MFF	HCHO	C	HYD	No
		=O	HBA	Yes
MFM	CH ₃ OH	C	HYD	No
		O	HBD	Yes
MFBZ	C ₆ H ₆	c	AROM	Yes, as a feature group
		c		
		c		
		c		
		c		
		c		
PFA	CH ₃ CHO	C	HYD	Yes, as a feature group
		C		
		=O	HBA	Yes
PFBAL	C ₆ H ₅ CHO	c	AROM	Yes, as a feature group
		c		
		c		

Name	Probe	Atom SMARTS	Feature	Is Considered?
		c		
		c		
		c		
		C	HYD	No
		=O	HBA	Yes
PFBSP	C ₆ H ₅ CH ₂ OH	c	AROM	Yes, as a feature group
		c		
		c		
		c		
		c		
		c		
		C	HYD	No
		O	HBD	Yes
PFE	C ₂ H ₅ OH	C	HYD	Yes, as a feature group
		C		
		O	HBD	Yes
PFGA	(CH ₂ OH) ₂	O	HBD	Yes
		C	HYD	Yes, as a feature group
		C		
		O	HBD	Yes
PFT	C ₆ H ₅ CH ₃	c	AROM	Yes, as a feature group
		c		
		c		
		c		
		c		
		c		
		C	HYD	Yes

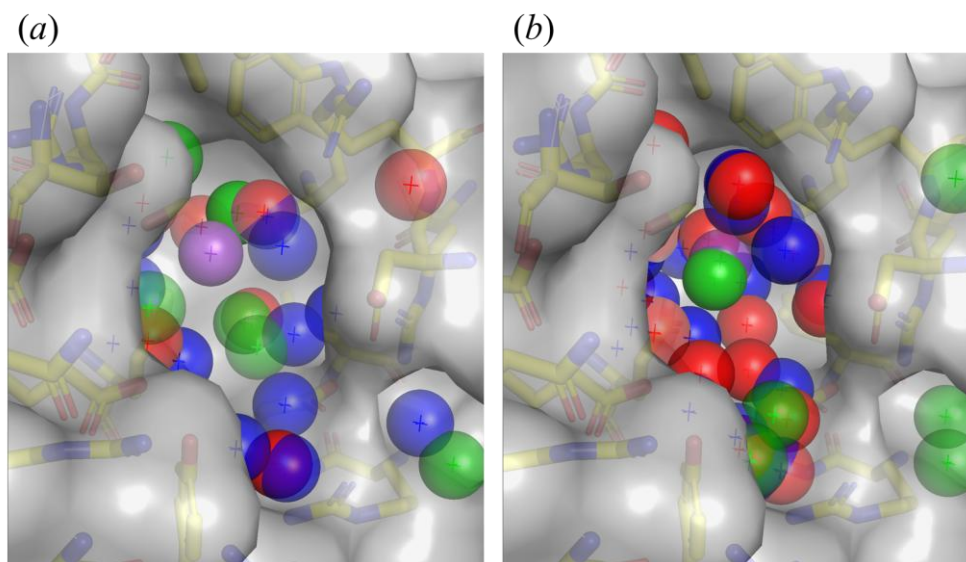


Figure S2. Pharmacophore features clusters based on results of molecular docking for (a) MFB and (b) PFB located in the binding site of neuraminidase H1N1 complexed with laninamivir (PDB ID: 3TI3). Green – hydrophobic, purple – aromatic interactions, blue – HBD, red – HBA.

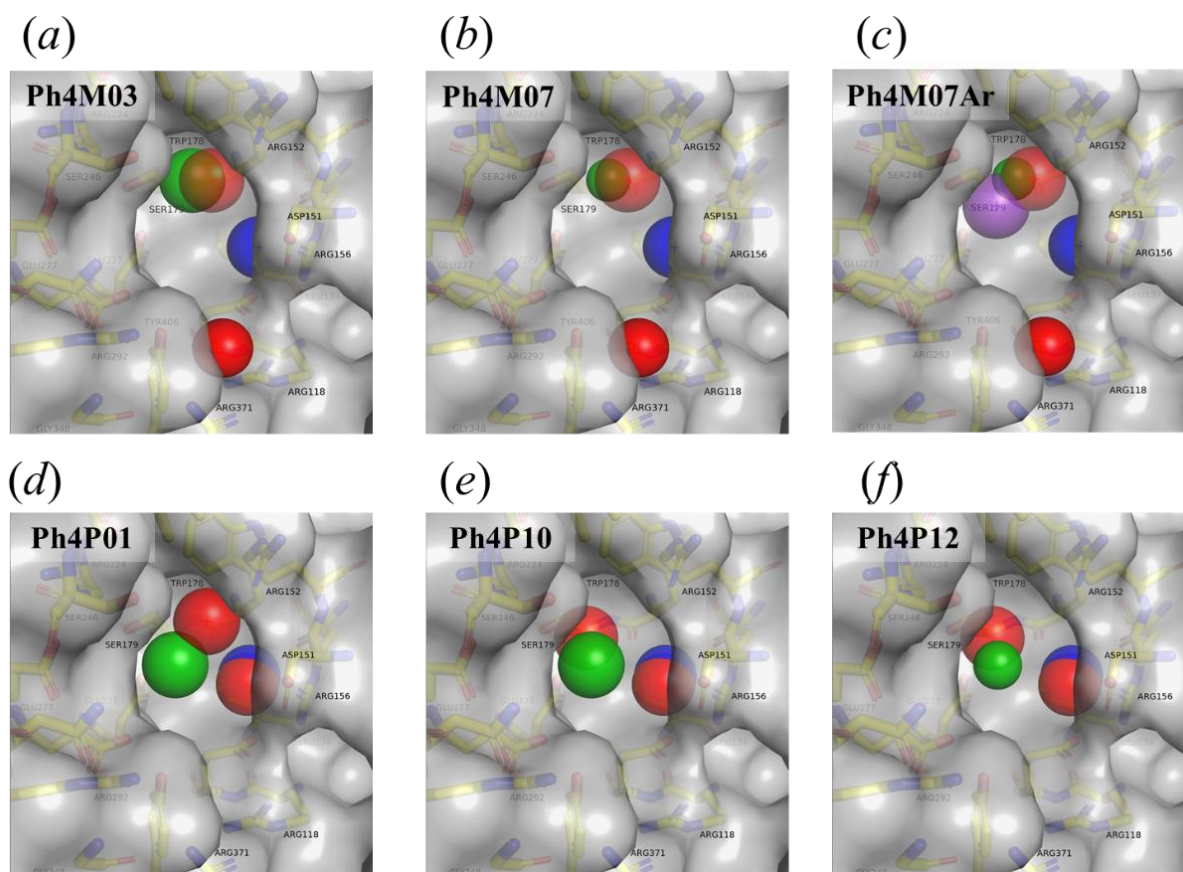


Figure S3. SBP models, based on pharmacophore feature clusters, for MFB (a – Ph4M03, b – Ph4M07), MFB-based SBP with aromatic pharmacophore feature (c – Ph4M07Ar) and PFB (d – Ph4P01, e – Ph4P10, f – Ph4P12). Green – hydrophobic, purple – aromatic interactions, blue – HBD, red – HBA.

Table S5. Metrics based on confusion matrices for MFB- and PFB-based pharmacophores.

Metric	Ph4M03	Ph4M07	Ph4P01	Ph4P10	Ph4P12
N	346	346	346	346	346
Actives	99	99	99	99	99
Decoys	247	247	247	247	247
TP	96	86	91	95	93
FP	103	65	187	190	173
FN	3	13	8	4	6
TN	144	182	60	57	74
Precision	48%	57%	33%	33%	35%
Accuracy	69%	77%	44%	44%	48%
TPR	97%	87%	92%	96%	94%
FPR	42%	26%	76%	77%	70%
AUC	0.83	0.82	0.53	0.56	0.57
EF1%	3.49	3.49	1.75	2.62	2.62
EF5%	2.52	2.91	1.36	1.55	1.55
EF10%	2.40	2.20	0.90	1.40	1.30

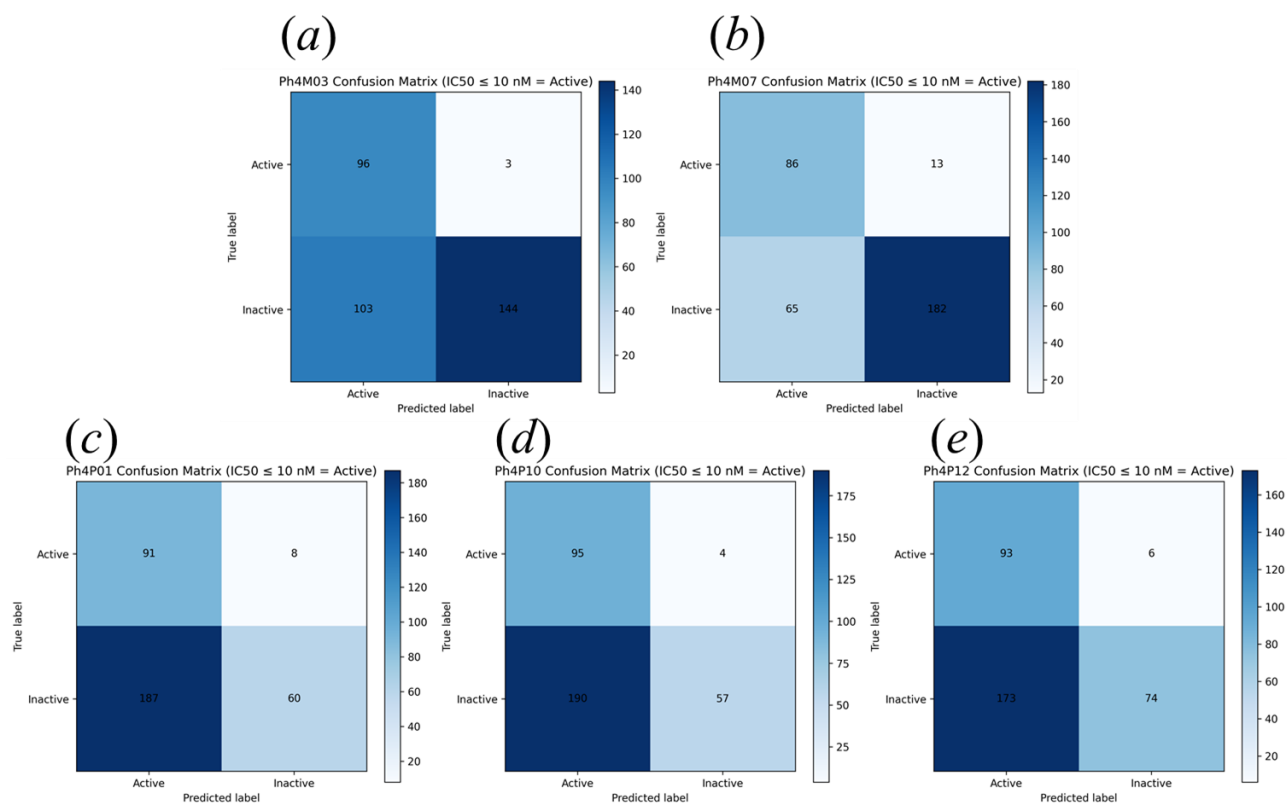


Figure S4. Confusion matrices for SBP models tested on validation set for MFB (a – Ph4M03, b – Ph4M07) and PFB (c – Ph4P01, d – Ph4P10, e – Ph4P12).

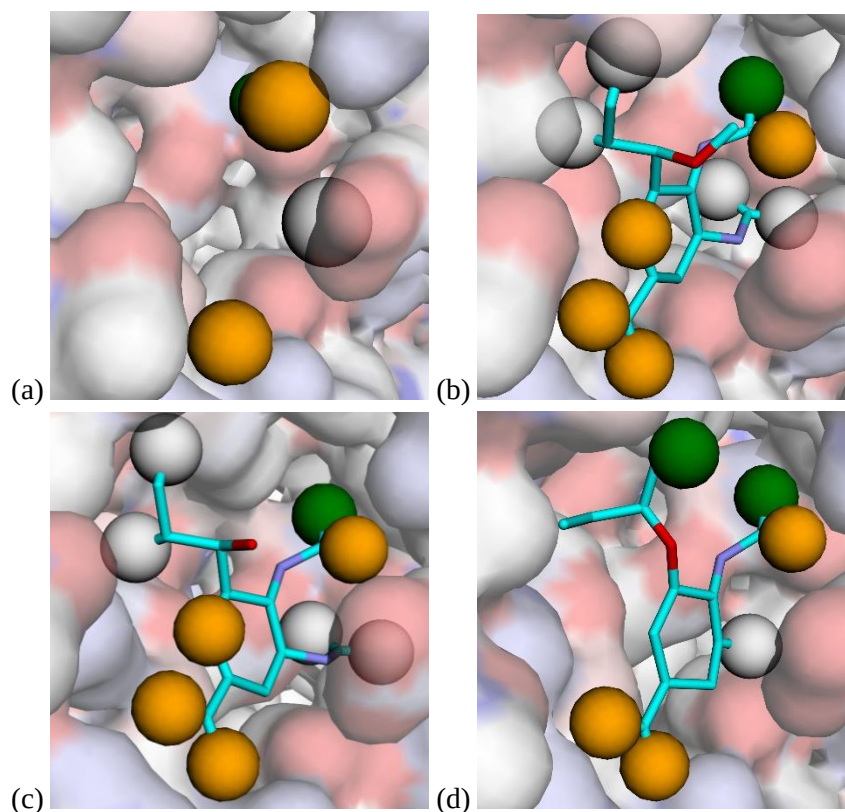


Figure S5. Pharmacophore models build in Pharmit: (a) SBP model Ph4M07 (PDB ID 3TI3 without co-crystallized ligand); (b) LBP model of laninamivir (PDB ID 3TI3); (c) LBP model of zanamivir (PDB ID 3TI5); (d) LBP model of oseltamivir-carboxylate (PDB ID 3TI6). Spheres: green – hydrophobic, white – HBD, orange – HBA.

HBD feature area of Glu-277 and Glu-276 was not used in SBP model Ph4M07, but it was determined by LBP models of laninamivir and zanamivir. It should be mentioned that LBP of oseltamivir also doesn't have that HBD-feature. Other features, determined by SBP are close to LBP models of approved drugs. Moreover, using Pharmit one could find many pharmacophore features for LBP model, but the more features are used in a model the smaller number of potential ligands is found.

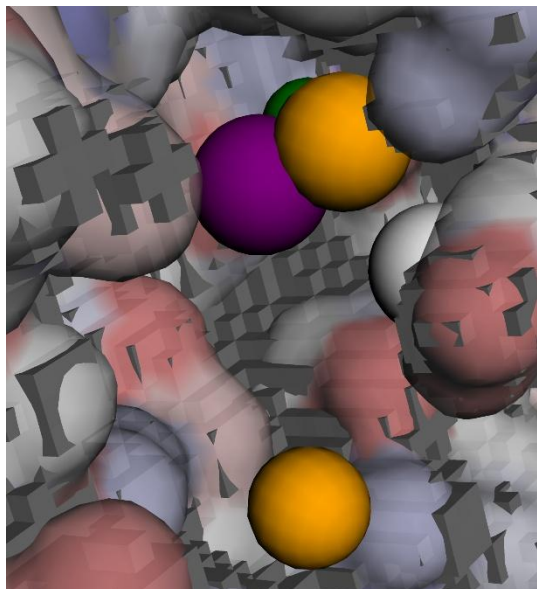


Figure S6. Model of Ph4M07Ar made in Pharmit using settings described in Table S4. The images are made in Pharmit for PDB ID 3TI3 as an example of the neuraminidase. Spheres: green – hydrophobic, purple – aromatic interactions, white – HBD, orange – HBA.

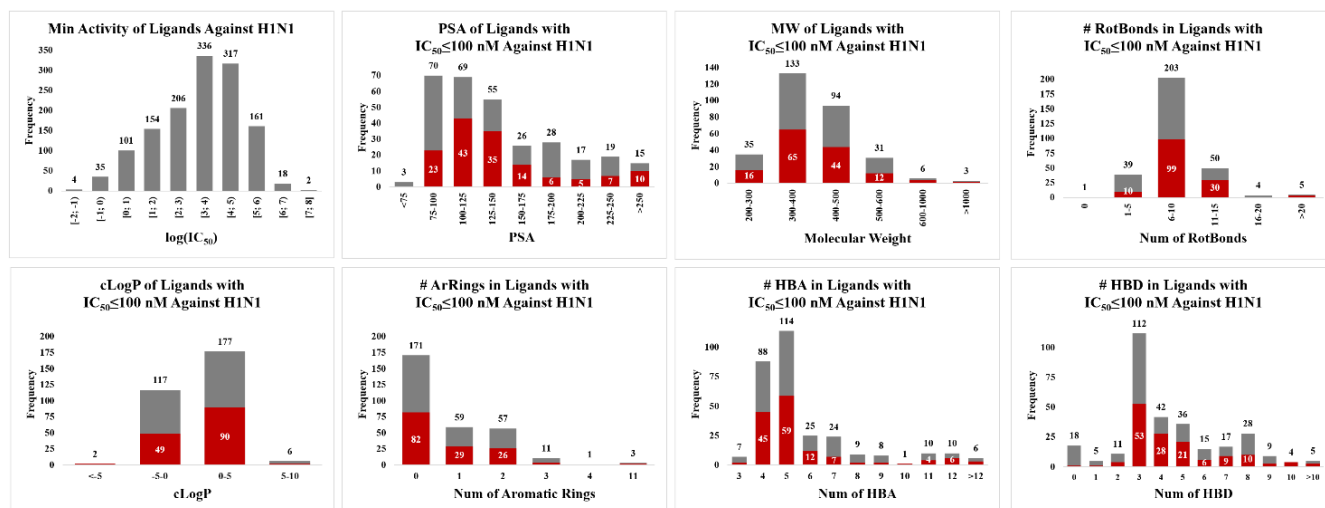


Figure S7. Distribution of physicochemical properties of compounds with neuraminidase H1N1 bioassays from Binding DB, calculated using RDKit Descriptor Calculation node in KNIME. Grey: total number of ligands. Red: number of ligands with $IC_{50} \leq 10nM$.

Table S6. ChEMBL DB VS settings for SBP Ph4M07Ar used in Pharmit for PDB ID 3TI3, chain A, without ligand (based on Lipinski's rule of five and Figure S7).

Settings	Value	Comment
MolWeight	[200;600]	VS is focused on searching for potential structures of hits, not fragments
RotBonds	[0;15]	According to Figure S6
LogP	[-5;5]	LogP value doesn't follow Lipinski's rule of five, but there are active ligands in range [-5;0], i.e. zanamivir (-3.58) (Figure S6)
PSA	[75;175]	According to Figure S6
Aromatics	[1;2]	According to Figure S6. Moreover, the aim of VS is to find ligands using aromatic pharmacophore feature
HBA	[2;10]	There are no active ligands with 2 or less HBA in Binding DB (Figure S6)
HBD	[3;8]	Num. of HBD doesn't follow Lipinski's rule of five, but there some active molecules with 6-8 HBD (Figure S6)
Exclusive Shape	Receptor Tolerance: 0.5Å, Solid	The exclusion volume must be specified to avoid unrealistic superposition of the ligand conformer onto the target structure.

Table S7. Plausible influenza A virus neuraminidase inhibitors in ChEMBL DB, identified by SBP model Ph4M07Ar with aromatic pharmacophore feature.

CSV-file "Table_S7_ChEMBL_ID_IAV_NA_Inhibitors.csv":

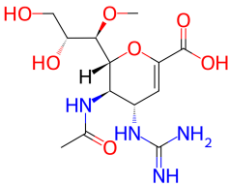
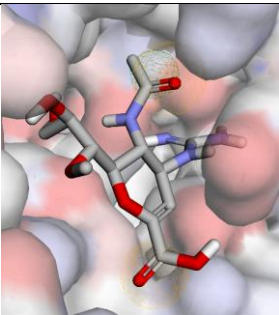
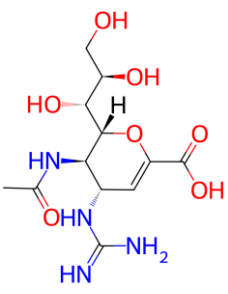
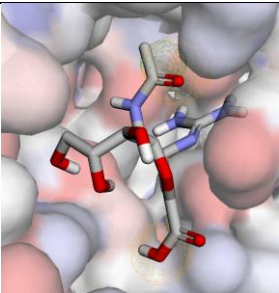
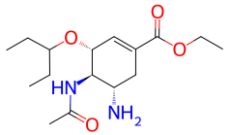
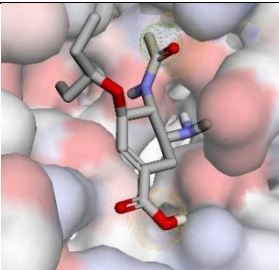
https://qsar.chem.msu.ru/~shulga/NA_SBP/Table_S7_ChEMBL_ID_IAV_NA_Inhibitors.csv
on https://qsar.chem.msu.ru/~shulga/NA_SBP/

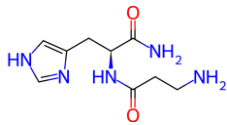
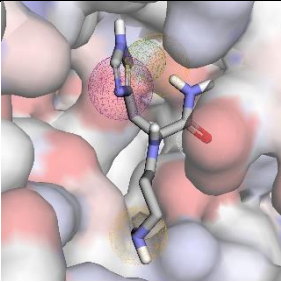
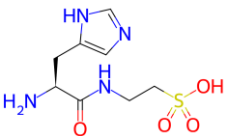
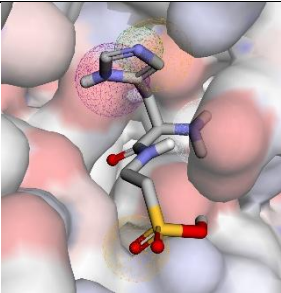
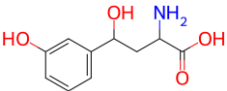
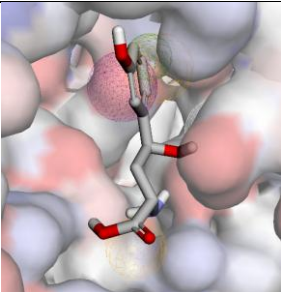
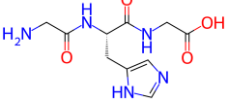
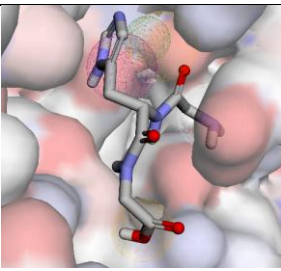
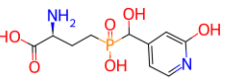
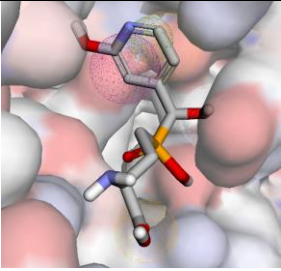


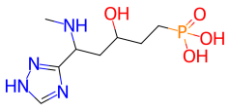
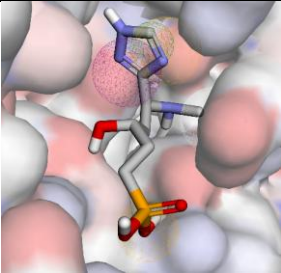
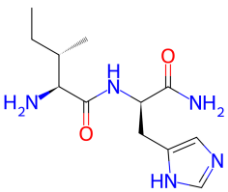
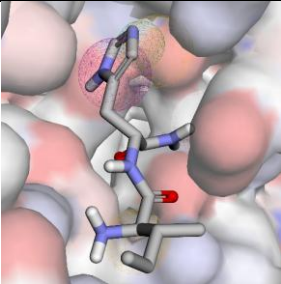
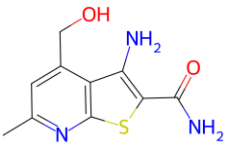
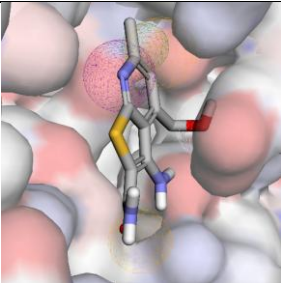
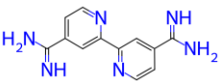
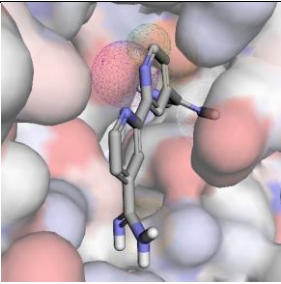
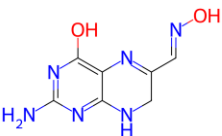
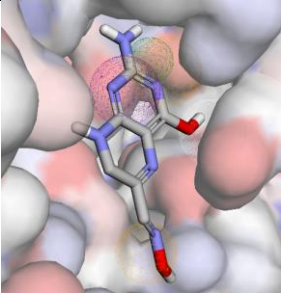
Figure S8. Affinity distribution after re-docking using GNINA for drugs and high potential ligands found by virtual screening using Pharmit and Ph4M07Ar SBP model: (a) PDB ID 3TI3; (b) PDB ID 3TI5; (c) PDB ID 3TI6. Green points: oseltamivir, zanamivir, laninamivir. Red points: top-15 ligands with highest GNINA CNN ligand efficiency (4-18 in Table S8). Yellow points: other top-100 ligands found by virtual screening.

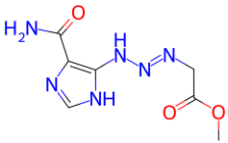
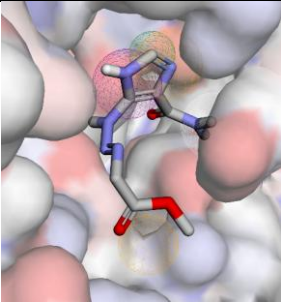
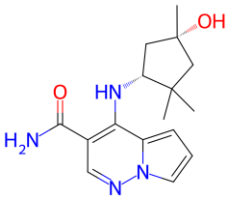
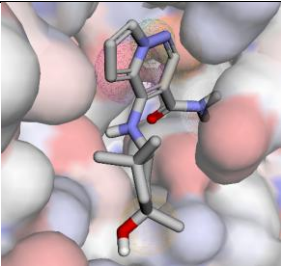
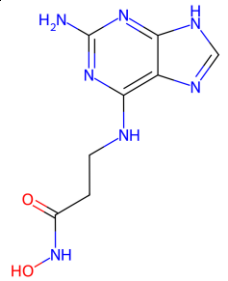
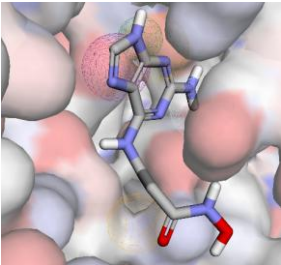
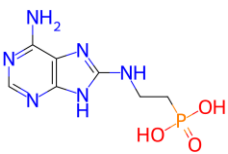
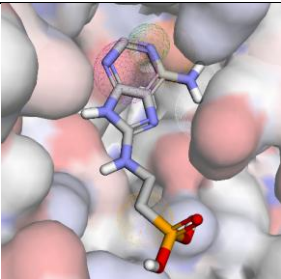
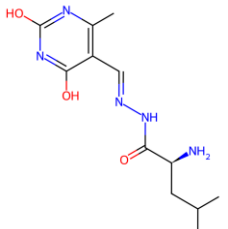
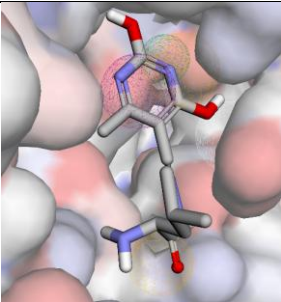
Table S8. Ligands with GNINA CNN energy and ligand efficiency values obtained after re-docking of molecules identified through virtual screening of the ChEMBL database using Ph4M07Ar SBP model in Pharmit. The images are made in Pharmit for PDB ID 3TI3 as an example of the neuraminidase. Spheres: green – hydrophobic, purple – aromatic interactions, white – HBD, orange – HBA. Structures 1-3 – drugs; 4-18 – top-15 ligands by LE; 19-23 – other ligands with highest GNINA Energy in 3TI3, 3TI5 or 3TI6.

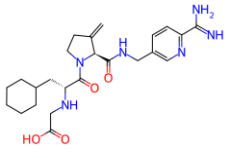
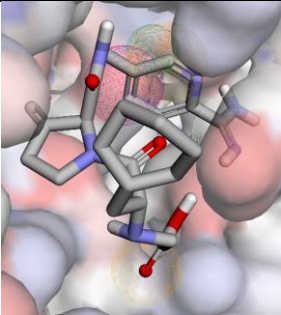
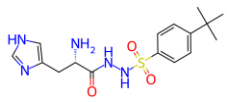
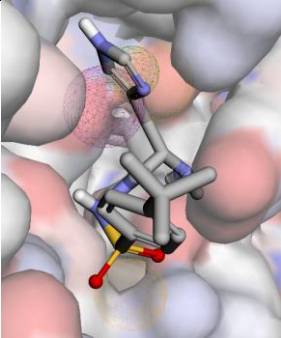
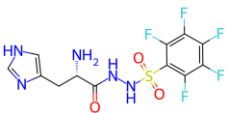
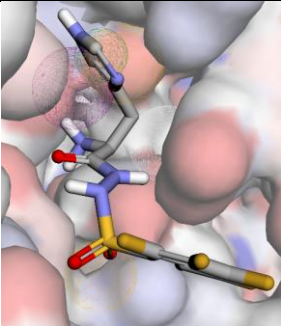
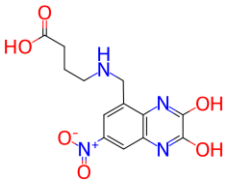
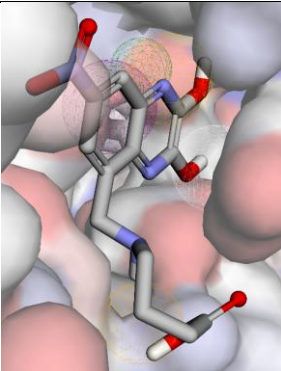
For oseltamivir, laninamivir and zanamivir affinity values for minimized structures are presented. According to conformers alignment, minimized structures with best affinity and lowest RMSD values found by Pharmit are very close to co-crystallized ones; conformers have minimal differences as it presented below in Figure S9. Additional comparison of approved drugs and high potential ligands was performed using GNINA by re-docking procedure after virtual screening. Parameters from Table S1 were chosen. The results of comparison are presented above in Figure S8.

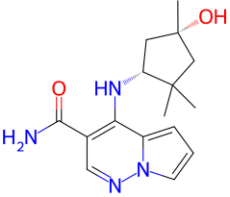
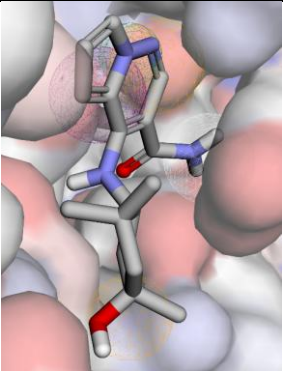
#	ChEMBL ID	Structure	Minimized conformation in Pharmit with SBP	Pharmit Score	Pharmit RMSD	GNINA Energy	GNINA LE
1	Laninamivir (CHEMBL466246) Pharmit Conformer			-8.03	1.09	-8.91	0.37
2	Zanamivir (CHEMBL222813) Pharmit Conformer			-8.35	0.75	-8.94	0.39
3	Oseltamivir (CHEMBL1229) Pharmit Conformer			-7.02	1.11	-8.38	0.42

#	ChEMBL ID	Structure	Minimized conformation in Pharmit with SBP	Pharmit Score	Pharmit RMSD	GNINA Energy	GNINA LE
4	CHEMBL1650080			-6.08	1.00	-7.07	0.44
5	CHEMBL1222396			-6.19	0.67	-7.46	0.44
6	CHEMBL120534			-6.46	1.39	-6.31	0.42
7	CHEMBL1221712			-5.90	0.92	-7.78	0.41
8	CHEMBL4076857			-6.22	1.25	-7.74	0.41

#	ChEMBL ID	Structure	Minimized conformation in Pharmit with SBP	Pharmit Score	Pharmit RMSD	GNINA Energy	GNINA LE
9	CHEMBL356644			-6.20	1.31	-6.83	0.40
10	CHEMBL2408875			-6.21	1.35	-7.50	0.39
11	CHEMBL206961			-6.62	0.65	-6.18	0.39
12	CHEMBL151724			-7.41	1.50	-6.78	0.38
13	CHEMBL3278663			-6.01	1.20	-5.57	0.37

#	ChEMBL ID	Structure	Minimized conformation in Pharmit with SBP	Pharmit Score	Pharmit RMSD	GNINA Energy	GNINA LE
14	CHEMBL4204949			-6.18	1.40	-5.87	0.37
15	CHEMBL3359916			-5.98	1.39	-7.77	0.35
16	CHEMBL3764406			-6.10	1.32	-5.95	0.35
17	CHEMBL483356			-6.18	1.45	-5.93	0.35
18	CHEMBL3195556			-5.88	0.92	-6.58	0.33

#	ChEMBL ID	Structure	Minimized conformation in Pharmit with SBP	Pharmit Score	Pharmit RMSD	GNINA Energy	GNINA LE
19	CHEMBL377056			-6.17	1.42	-8.83	0.26
20	CHEMBL514464			-6.23	0.82	-8.08	0.32
21	CHEMBL519151			-7.40	1.36	-6.95	0.27
22	CHEMBL85211			-6.50	1.36	-6.99	0.30

#	ChEMBL ID	Structure	Minimized conformation in Pharmit with SBP	Pharmit Score	Pharmit RMSD	GNINA Energy	GNINA LE
23	CHEMBL3359916			-5.98	1.39	-7.77	0.35

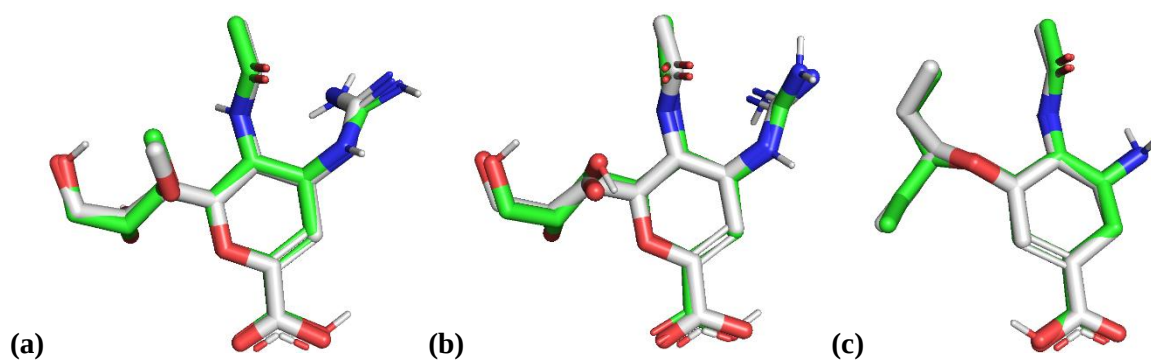


Figure S9. Alignment of co-crystallized (grey) and Pharmit minimized (green) conformers of molecules: (a) laninamivir (PDB ID 3TI3); (b) zanamivir (PDB ID 3TI5); (c) oseltamivir, or oseltamivir carboxylate (PDB ID 3TI6).