Supplemental Information

Structural basis for the identification of the N-terminal domain of coronavirus nucleocapsid protein as an antiviral target

Shing-Yen Lin,[1,2] Chia-Ling Liu,[1,2] Yu-Ming Chang,[3] Jincun Zhao,[4] Stanley Perlman,[4#] and Ming-Hon Hou[1,2,5,*#]

TOC

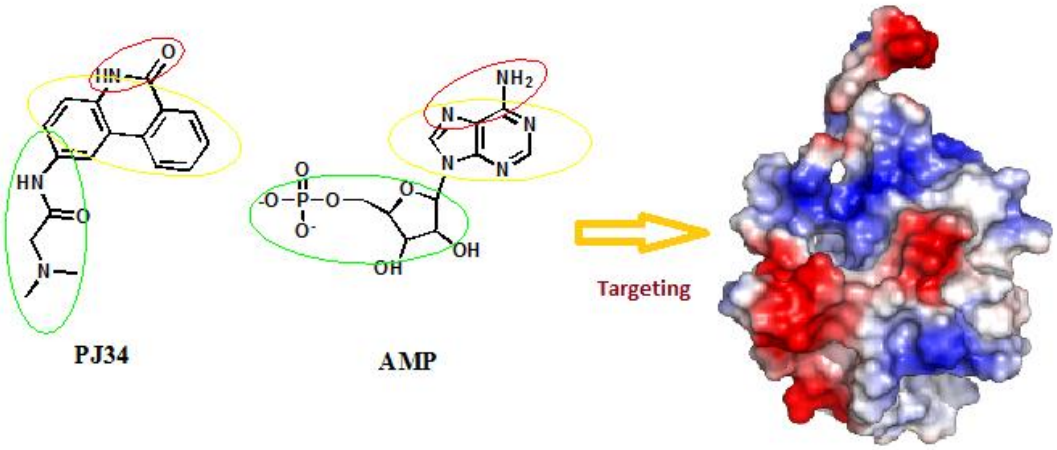


Figure legends

Figure S1. Amino acid sequence alignment (performed using T-coffee) of the N-NTDs from HCoV-OC43 (NP_937954), SARS-CoV (ABI96968), HCoV-229E (AAG48597), IBV (AAB24054), and MHV (ACO72897); these sequences were retrieved from GenBank. The conserved amino acids in the four sequences are shaded in red, while the similar residues are highlighted in blue.

Figure S2. (A) HCoV-OC43 N-NTD sequences aligned with the other N-terminal RNA-binding domains in nucleocapsid proteins from coronaviruses. The conserved amino acids in the four sequences are shaded in black, while the similar residues are highlighted in grey. In addition, the residues marked with an asterisk interact with the RNA using their side chains. **(B)** Left: Superimposition of the CoV N-NTD of HCoV-OC43-AMP complex (green), SARS (purple), IBV (orange), and MHV (blue). Right: Enlargement of the RNA binding site.

Figure S3. (A) The detailed interactions of CMP binding site. The residues in this binding site responsible for interacting with CMP include Ser 64, Phe 66, Arg 122, Tyr 124, Tyr 126 and Arg 164. **(B)** The detailed interactions of GMP binding site. The residues in this binding site responsible for interacting with GMP include Ser 64, Gly 68, Arg 122, Tyr 124, Tyr 126 and Arg 164. **(C)** The detailed interactions of UMP binding site. The residues Ser 64, Gly 68, Phe 66, Arg 122, Tyr 124 and Arg 164. form interaction with UMP. The electron density (*2Fo - Fc*) is contoured at 1.0 sigma level. The dotted green lines indicate hydrogen bonds. The red dashed lines indicate ionic interactions. **(D)** A summary of the amino acids that interact with the N protein of HCoV-OC43 N-NTD.

Figure S4. Virus replication is inhibited by presence of mutant N protein. Cells were with plasmids encoding the WT N protein or the mutant (R122A, Y124A or R164A) N protein and then infected with HCoV-OC43 as described in Materials and Methods. Samples were then analyzed for levels of M protein gene transcript. No transfection and no infection controls are also shown.

Figure S5. Structural superimposition of the HCoV-OC43 N-NTD-AMP (red) with HCoV-OC43 N-NTD-PJ34 (green) at the residues involved in the ligand binding.

Figure S6. Surface charge distribution of N-NTDs from HCoV-OC43(4LI4), SARS-CoV(2OFZ), IBV(2GEC), and MHV(3HD4).

Figure S7. (A) Surface charge distribution of the HCoV-OC43 N-NTD-RNA complex in which blue and red indicate positive and negative charges, respectively. **(B)** A model of the HCoV-OC43 N-NTD-RNA complex. The model was constructed using the crystal structure of an OC43 N-NTD-AMP complex as a template.

Figure S1


HCoV-OC43 NVVPYYSWFSGITQFQKGKEFEFVEGQGVPIAPGVPATEAKGYWYRHNRR
HCoV-HKU1 NTIPHYSWFSGITQFQKGRDFKFSDGQGVPIAFGVPPSEAKGYWYRHSRR
MHV SVVPHYSWFSGITQFQKGKEFQFAEGQGVPIANGIPASEQKGYWYRHNRR
SARS-CoV NTASWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRRATRR
MERS-CoV NTVSWYTGLTQHGK-VPLTFPPGQGVPLNANSTPAQNAGYWRRQDRK
IBV GSSGNASWFQAIKAKKLNTPPPKFEGSGVPDNENIKPSQQHGYWRRQAR-
HCoV-229E YSLYSPLLVDSEQ-PWKVIPRNLVPINKKDK-NKLIGYWNVQKR-
HCoV-NL63 PSFYMPLLVSSDKAPYRVIPRNLVPIGKGNK-DEQIGYWNVQER-
HCoV-OC43 SFKTADGNQRQLLPRWYFYYLGTGPHAKDQYGTDIDGVYWVASNQADVNT
HCoV-HKU1 SFKTADGQQKQLLPRWYFYYLGTGPYANASYGESLEGVFWVANHQADTST
MHV SFKTPDGQQKQLLPRWYFYYLGTGPHAGASYGDSIEGVFWVANSQADTNT
SARS-CoV -VRGGDGKMKELSPRWYFYYLGTGPEASLPYGANKEGIVWVATEGALNTP
MERS-CoV -INTGNG-IKQLAPRWYFYYTGTGPEAALPFRAVKDGIVWVHEDGATDAP
IBV -FKPGKGGRKPVPDAWYFYYTGTGPAADLNWGDTQDGIVWVAAKGADTKS
HCoV-229E -FRTRKGKRVDLSPKLHFYYLGTGPHKDAKFRERVEGVVWVAVDGAKTEP
HCoV-NL63 -WRMRRGQRVDLPPKVHFYYLGTGPHKDLKFRQRSDGVVWVAKEGAKTVN
HCoV-OC43 PADIVDRDPSSDEAIPTRFPPGTVLPQGYYIEGS
HCoV-HKU1 PSDVSSRDPTTQEAIPTRFPPGTILPQGYYVEGS
MHV RSDIVERDPSSHEAIPTRFAPGTVLPQGFYVEGS
SARS-CoV KDHIGTRNPNNNAATVLQLPQGTTLPKGFYAEGS
MERS-CoV ST-FGTRNPNNDSAIVTQFAPGTKLPKNFHIEGT
IBV RSNQGTRDPDKFDQYPLRFSDGG--PDGNFRWDFI
HCoV-229E TG-YGVRRKNSEPEIP---H-NQKLPNGVTVVEE
HCoV-NL63 TS-LGNRKRNQKPLEP---KFSIALPPELSVVEF

Figure S2

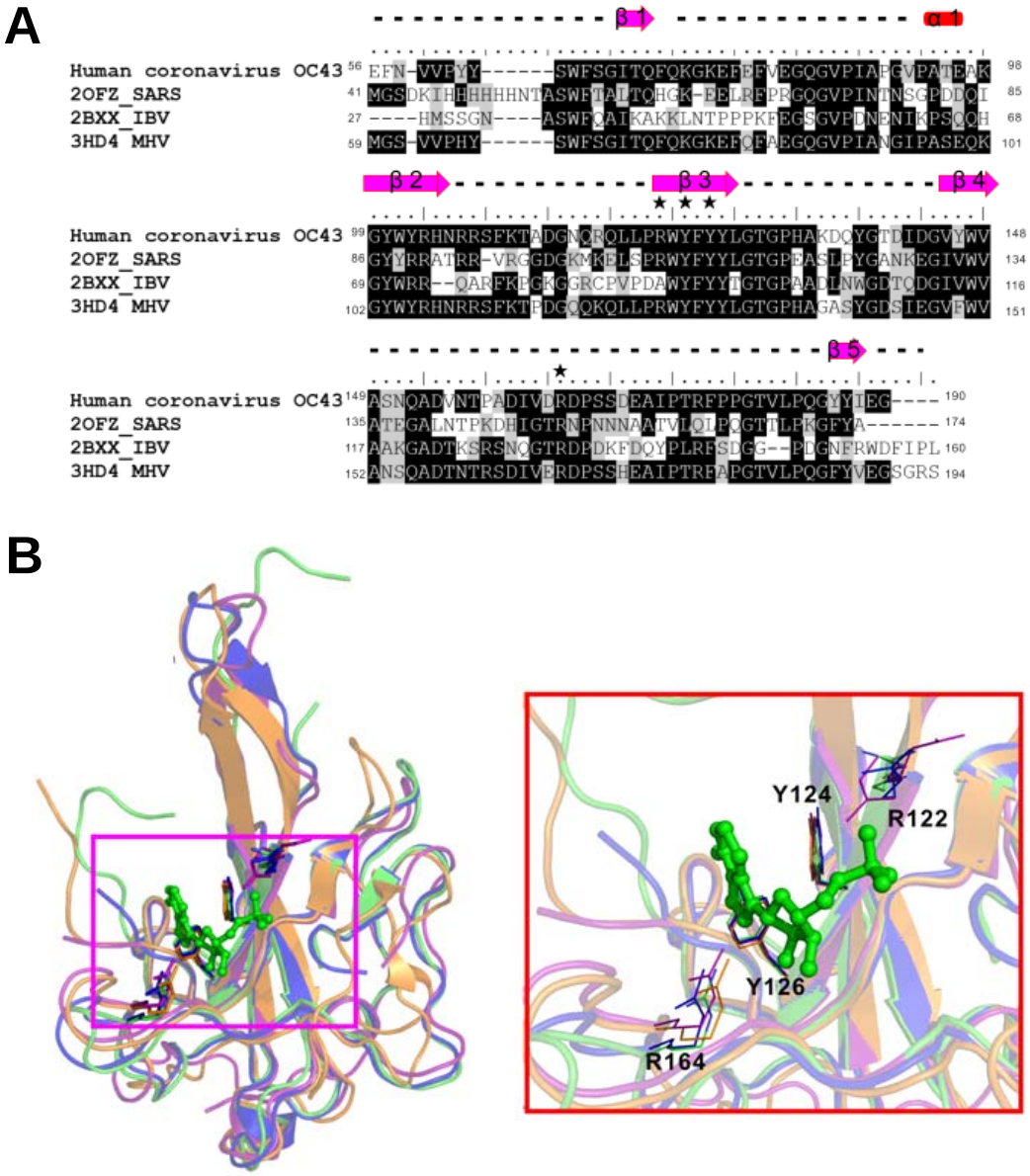

A
β 1
α 1
Human coronavirus OC43
2OFZ_SARS
2BXX_IBV
3HD4_MHV
β 2
β 3
β 4
β 5
B
Y124
R122
Y126
R164

Figure S3

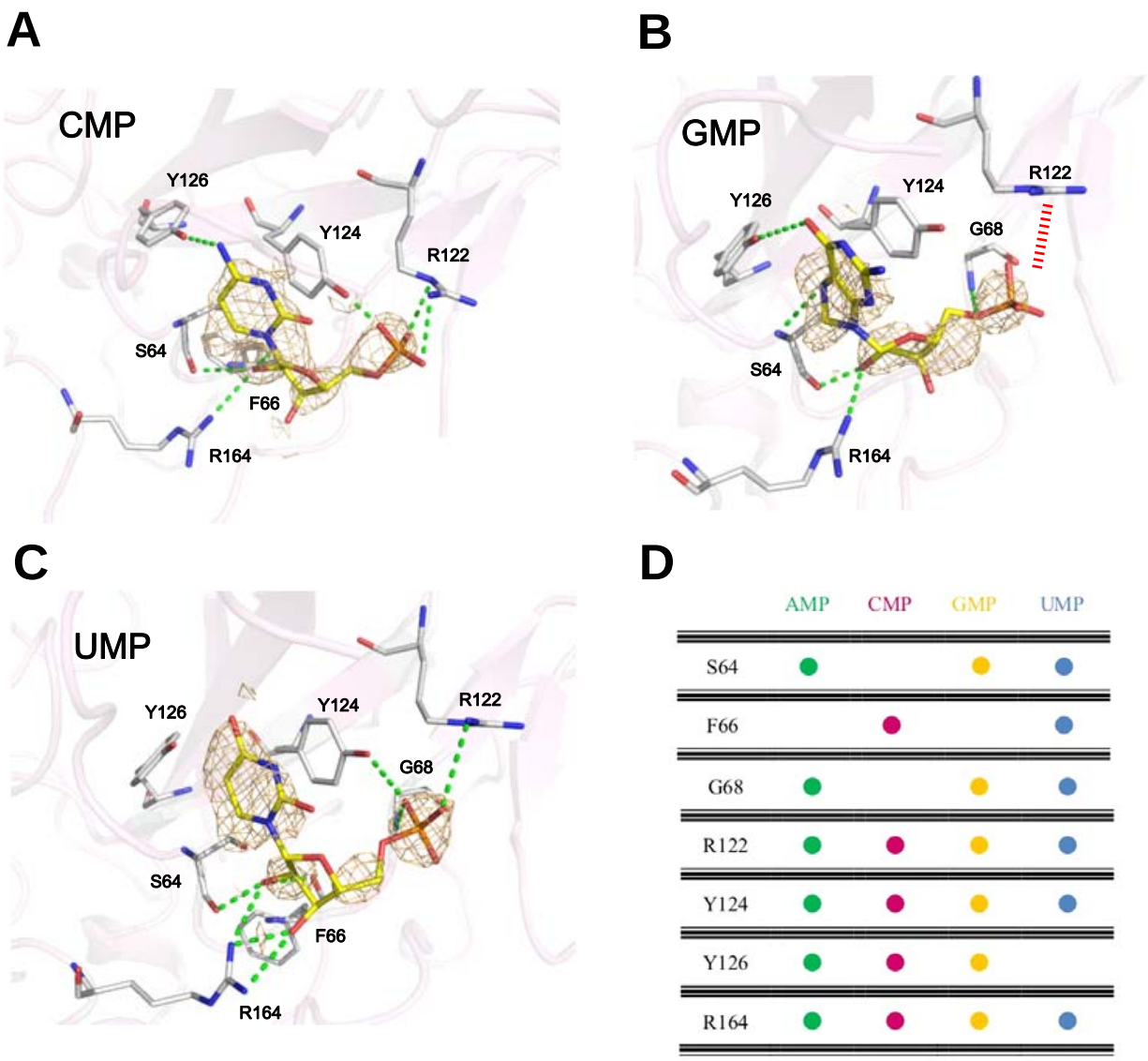

A
CMP
Y126
Y124
R122
S64
F66
R164
B
GMP
Y126
Y124
R122
G68
S64
R164
C
UMP
Y126
Y124
R122
G68
S64
F66
R164
D
AMP
CMP
GMP
UMP
S64
F66
G68
R122
Y124
Y126
R164

Figure S4

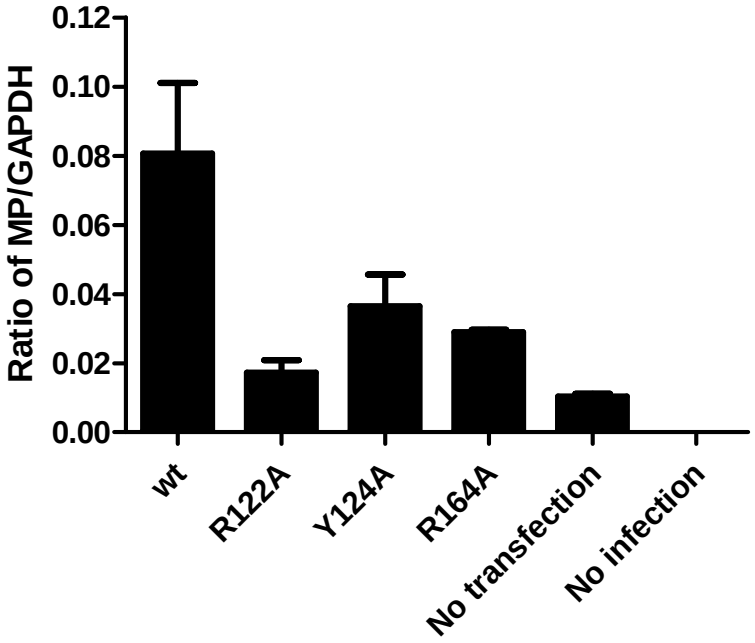

Ratio of MP/GAPDH
0.12
0.10
0.08
0.06
0.04
0.02
0.00
wt
R122A
Y124A
R164A
No transfection
No infection

Figure S5

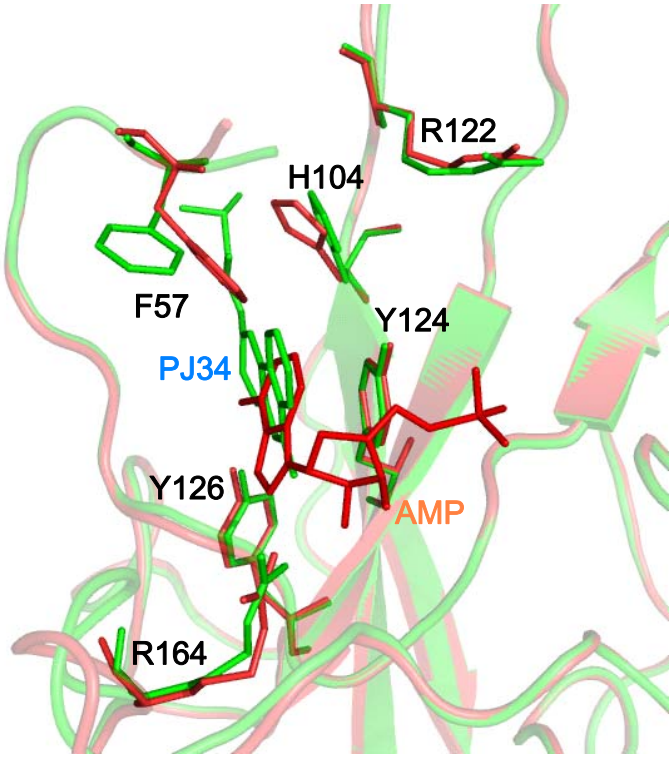


Figure S6

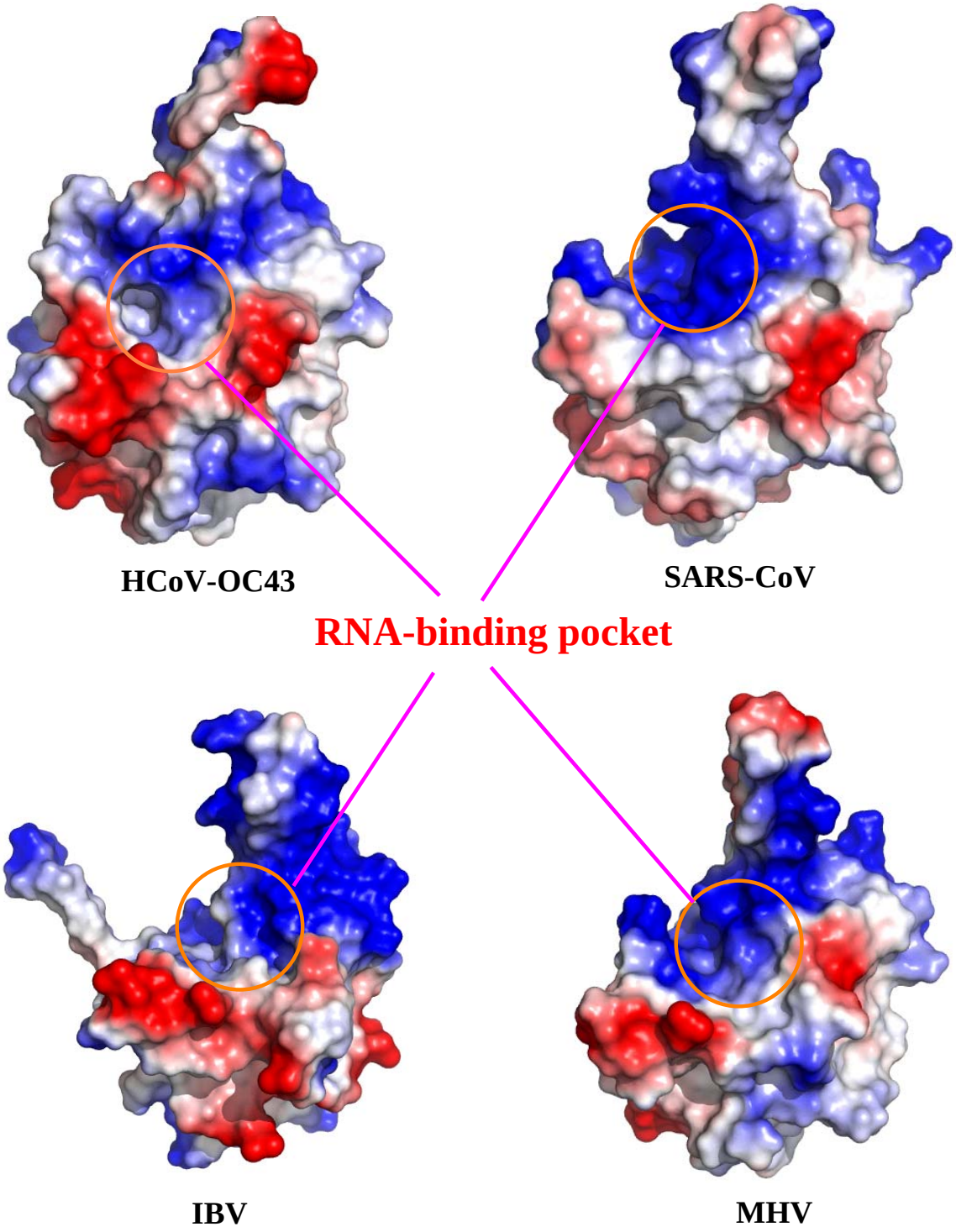


Figure S7

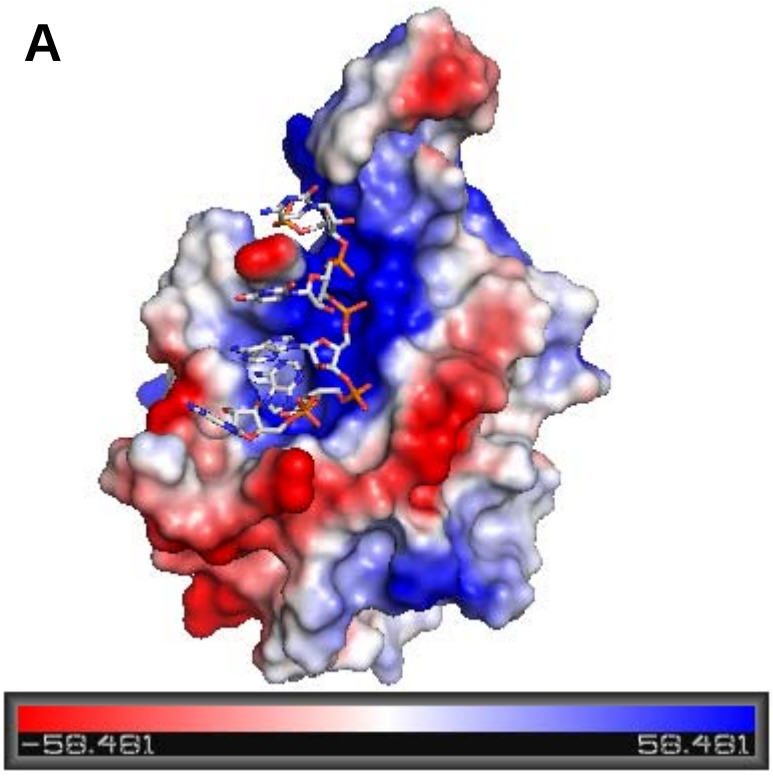
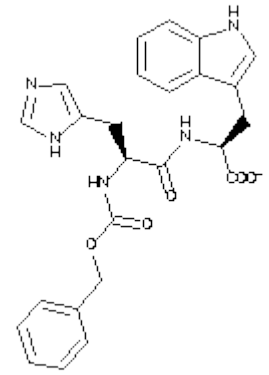
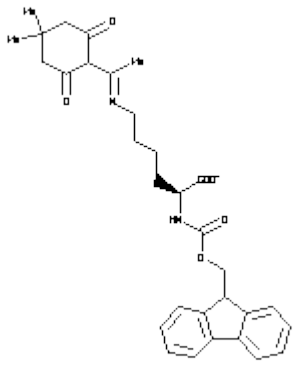
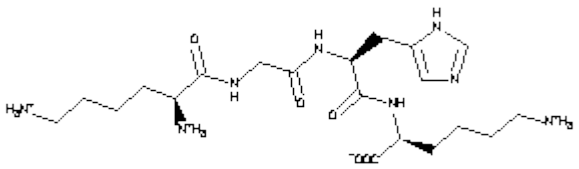
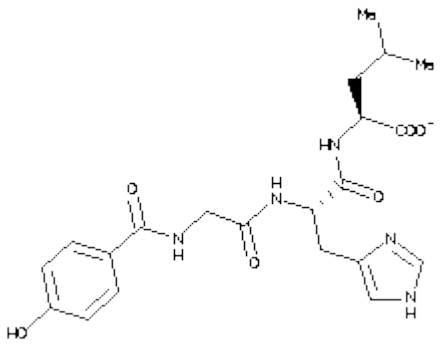


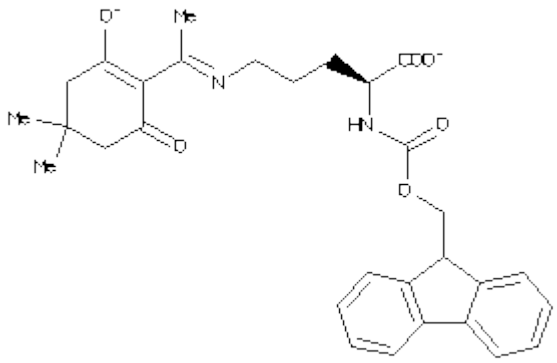
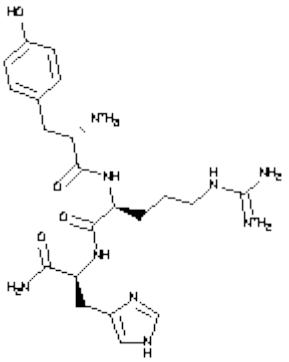
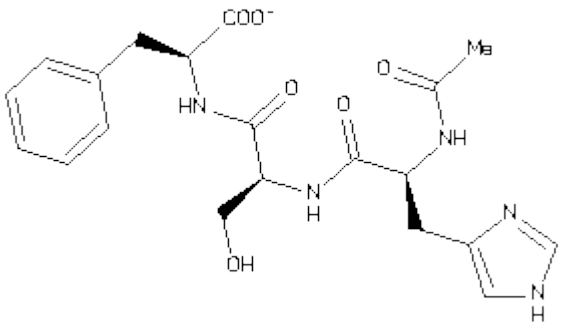
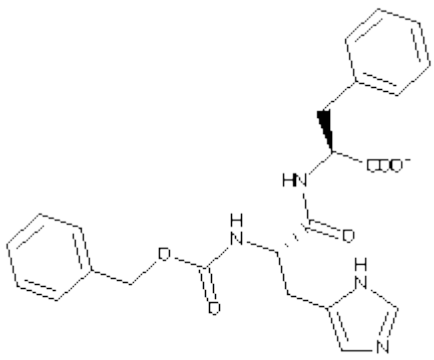
Table S1. Data collection and refinement statistics for HCoV-OC43 N-NTD-CMP, HCoV-OC43 N-NTD-GMP and HCoV-OC43 N-NTD-UMP crystals

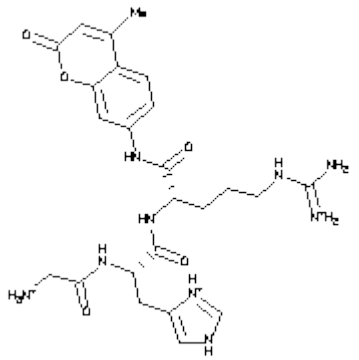
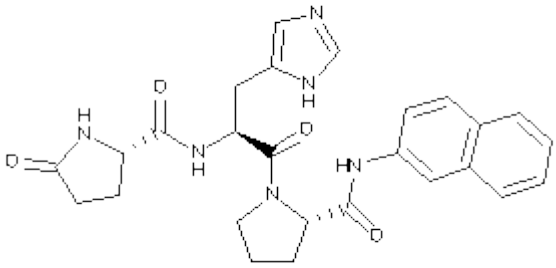
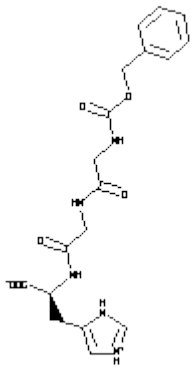
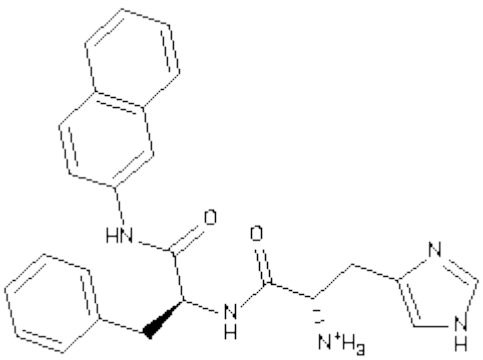
Names	HCoV-OC43 N-NTD-CMP	HCoV-OC43 N-NTD-GMP	HCoV-OC43 N-NTD-UMP
PDB number	4LMC	4LM9	4LM7
Data Collection		NSRRC BL13B1	
Space group	$P6_5$	$P6_5$	$P6_5$
Resolution (Å)[a]	30-1.74 (1.80-1.74)[a]	30-1.60 (1.63-1.60)[a]	30-1.72 (2.59-1.72)[a]
Wavelength (Å)	1.00000	1.00000	1.00000
Unit Cell Dimensions			
$a = b$ (Å)	81.897	81.971	81.938
c (Å)	42.889	42.732	42.891
No. of reflections			
Observed	82635	144518	122776
Unique	17034	21252	17712
Completeness (%)	94.8(94.8)[a]	97.3(98.9)[a]	98.8(100)[a]
R_{merge} (%)	4.0(16.1)[a]	3.1(16.1)[a]	3.9(15.0)[a]
I/ σ(I)	34.06(9.59)[a]	51.97 (2.86)[a]	37.4(13.0)[a]
Refinement			
No. of reflections	16138	20167	17394
R_{work} (95% data)	0.21	0.22	0.20
R_{free} (5% data)	0.25	0.26	0.23
Geometry deviations			
Bond lengths (Å)	0.014	0.013	0.016
Bond angles (°)	1.783	1.565	1.690
No. of all protein atoms			
Mean B-values (Å^2)	34.19	19.49	29.30
No. of ligand atoms			
Mean B-values (Å^2)	50.53	33.54	41.94
No. of water molecules			
Mean B-values (Å^2)	36.81	23.97	44.46
Ramachandran plot (%)			
Most favored	96.0	96.0	95.3
Generously allowed	0.8	1.6	3.1
Others	3.2	2.4	1.6

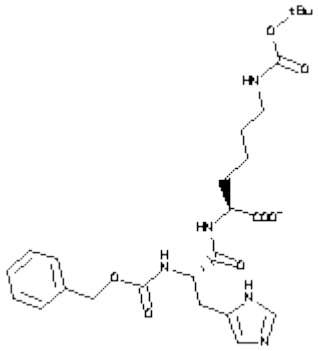
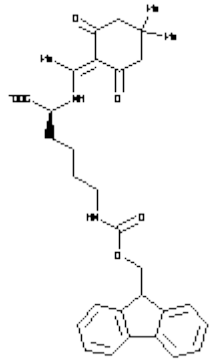
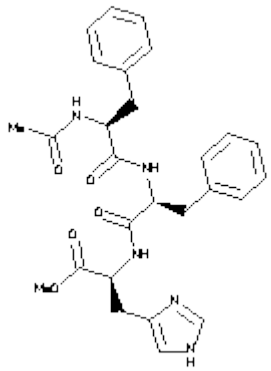
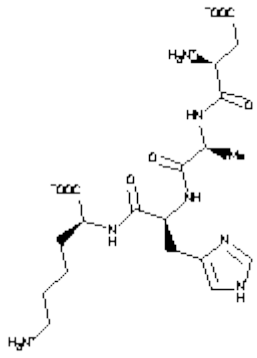
[a] Values in the parenthesis are for the highest resolution

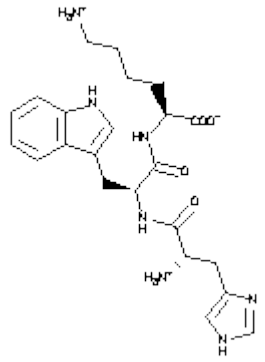
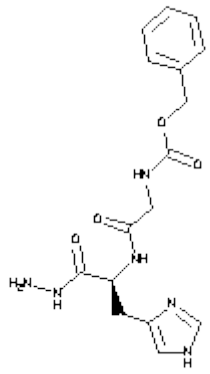
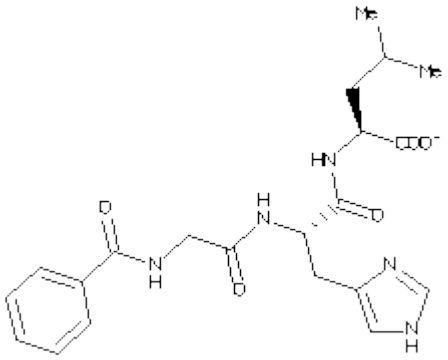
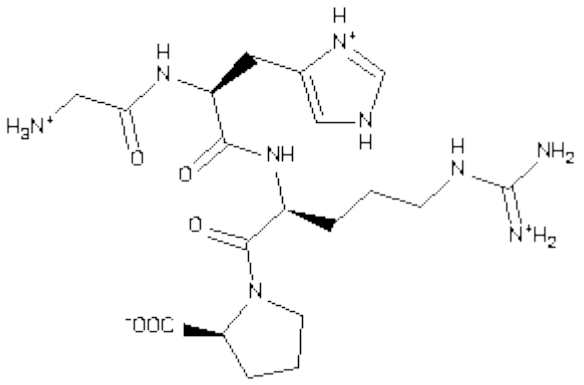
Table S2. The chemical structures and docking scores of 87 potential hits.

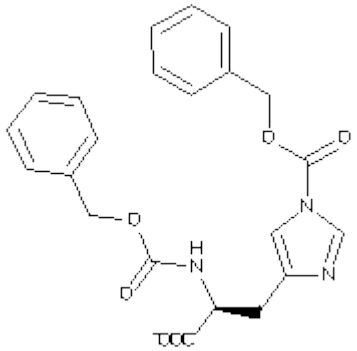
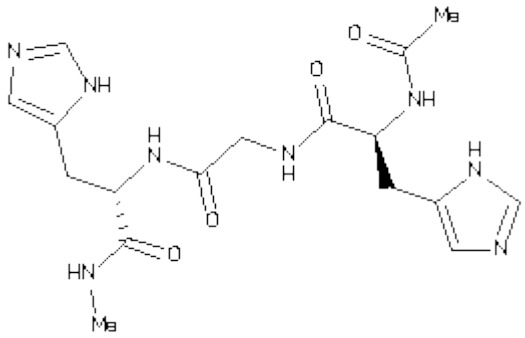
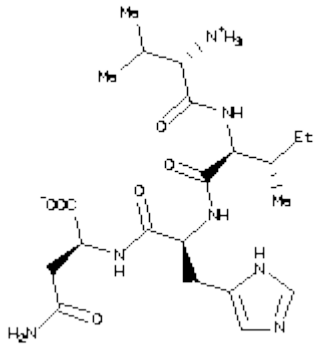
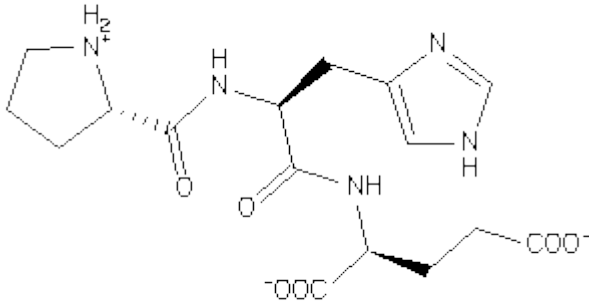
	**ZINC ID**：ZINC15721543 **Score**：135.925
	**ZINC ID**：ZINC08781153 **Score**：129.222
	**ZINC ID**：ZINC15722196 **Score**：127.518
	**ZINC ID**：ZINC15721227 **Score**：123.882

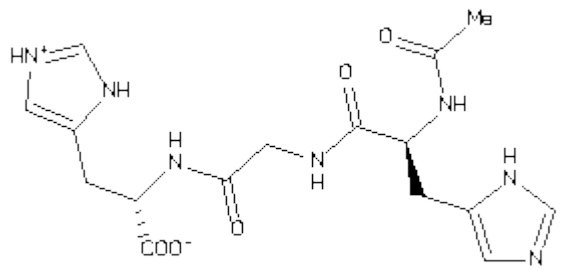
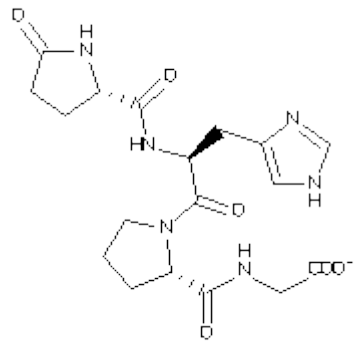
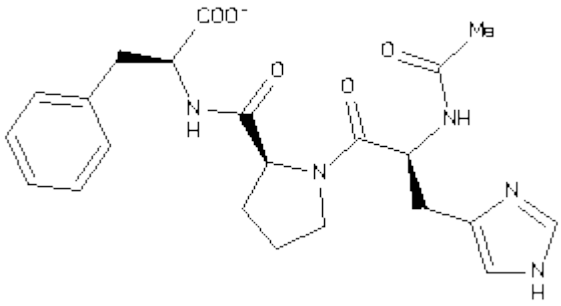
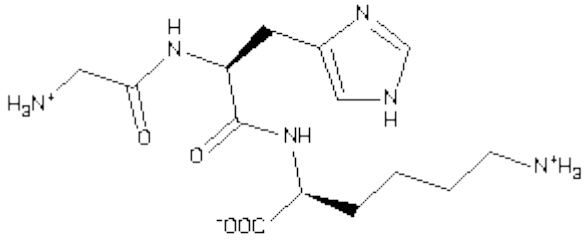
	**ZINC ID**：ZINC15722239 **Score**：122.028
	**ZINC ID**：ZINC15721984 **Score**：120.458
	**ZINC ID**：ZINC15721747 **Score**：119.428
	**ZINC ID**：ZINC13543937 **Score**：119.312

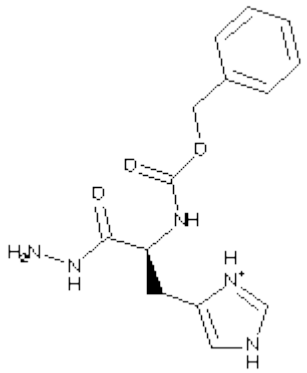
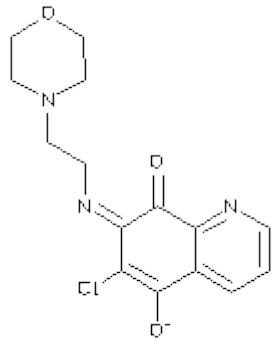
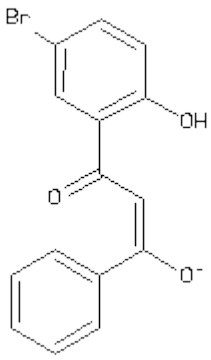
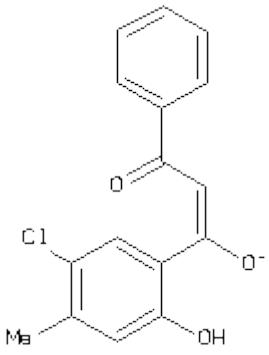
	**ZINC ID**：ZINC15722313 **Score**：118.923
	**ZINC ID**：ZINC15721220 **Score**：117.611
	**ZINC ID**：ZINC15722043 **Score**：117.377
	**ZINC ID**：ZINC15721913 **Score**：117.319

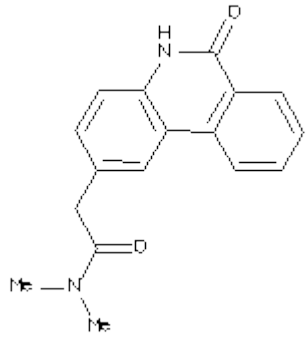
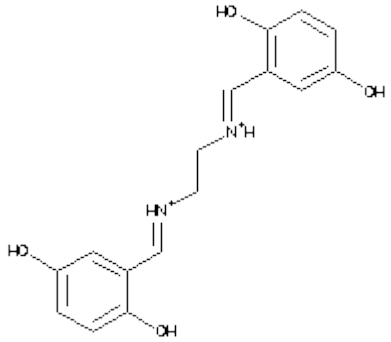
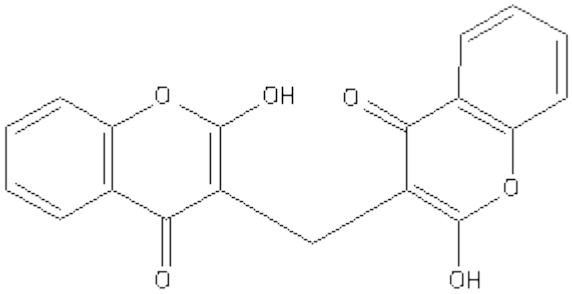
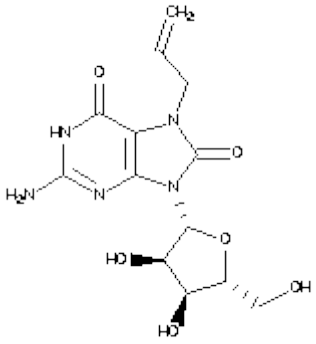
	**ZINC ID**：ZINC15722049 **Score**：115.862
	**ZINC ID**：ZINC15722084 **Score**：115.498
	**ZINC ID**：ZINC15721523 **Score**：115.449
	**ZINC ID**：ZINC15721853 **Score**：113.235

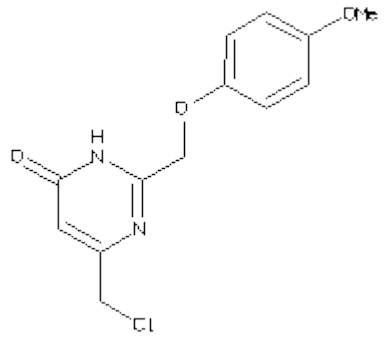
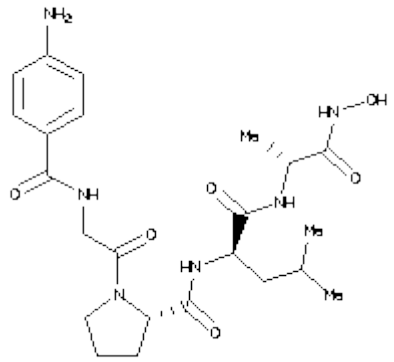
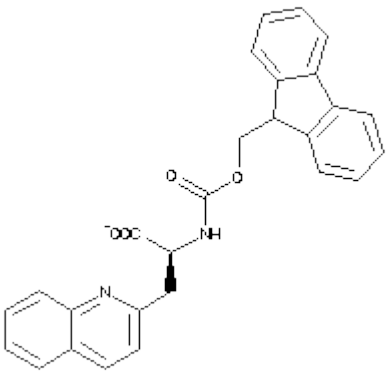
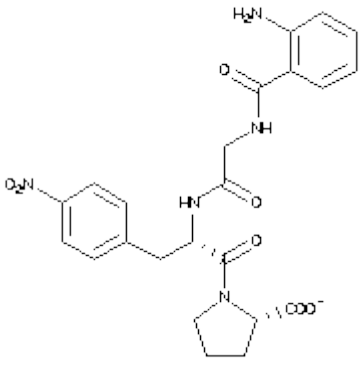
	**ZINC ID**：ZINC15721916 **Score**：112.834
	**ZINC ID**：ZINC13285034 **Score**：112.807
	**ZINC ID**：ZINC06566221 **Score**：112.538
	**ZINC ID**：ZINC15721431 **Score**：111.285

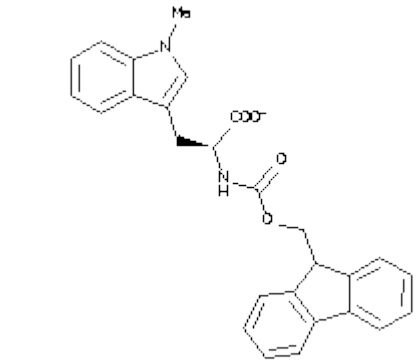
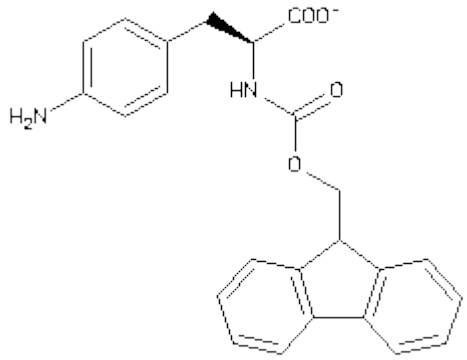
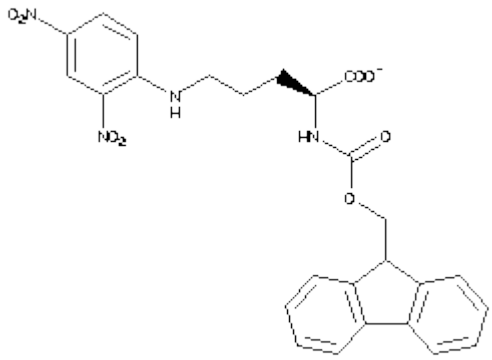
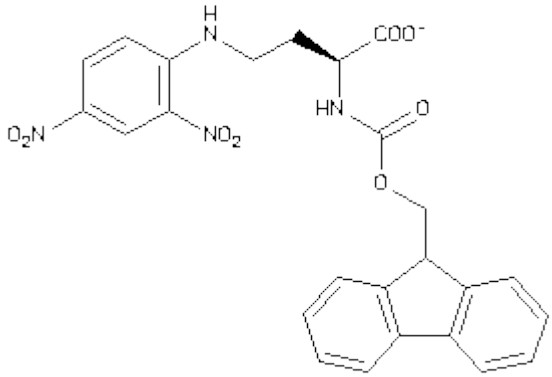
	**ZINC ID**：ZINC15721511 **Score**：110.458
	**ZINC ID**：ZINC15721738 **Score**：109.574
	**ZINC ID**：ZINC15721676 **Score**：106.657
	**ZINC ID**：ZINC15721965 **Score**：105.54

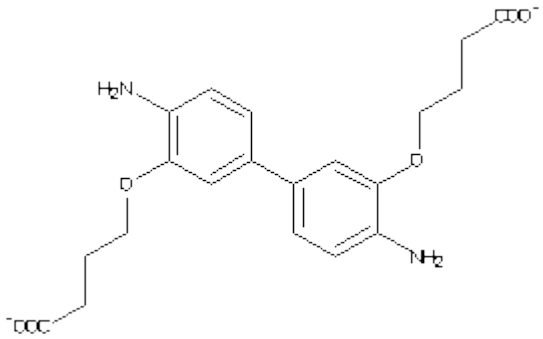
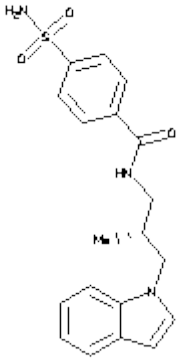
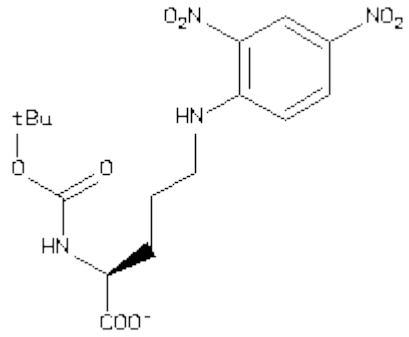
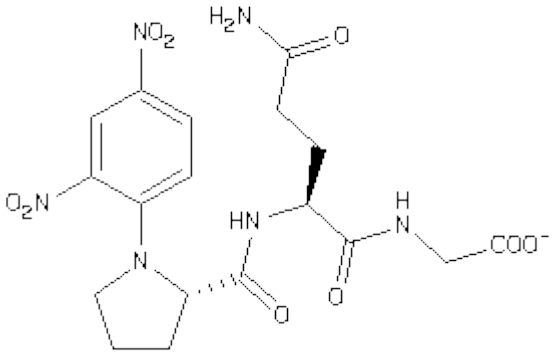
	**ZINC ID**：ZINC15721741 **Score**：104.826
	**ZINC ID**：ZINC15721529 **Score**：103.04
	**ZINC ID**：ZINC15721744 **Score**：101.332
	**ZINC ID**：ZINC13507674 **Score**：100.356

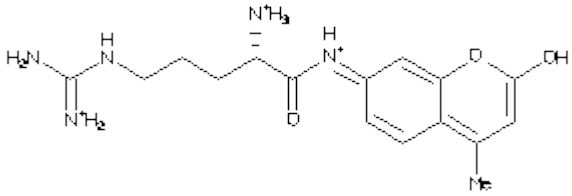
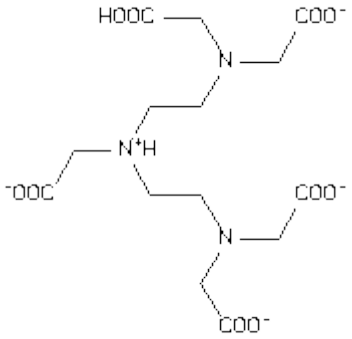
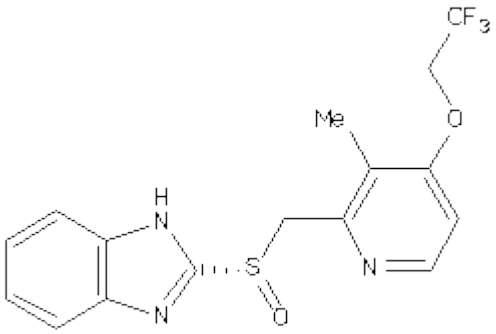
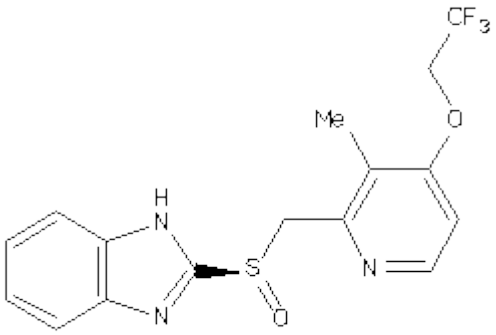
	**ZINC ID**：ZINC01708465 **Score**：99.8826
	**ZINC ID**：ZINC00027263 **Score**：113.528
	**ZINC ID**：ZINC00155947 **Score**：82.894
	**ZINC ID**：ZINC00155949 **Score**：100.623

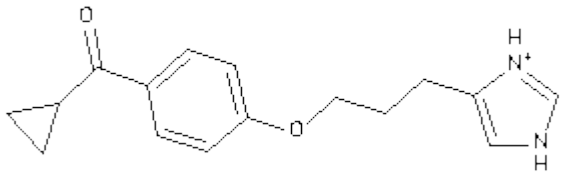
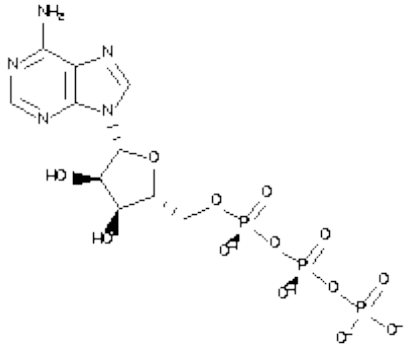
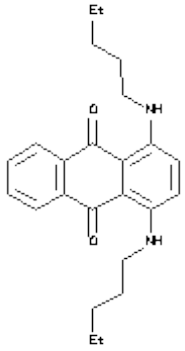
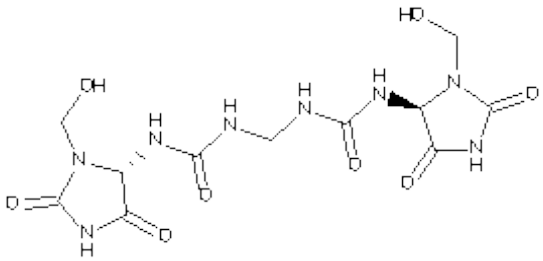
	**ZINC ID**：ZINC02507482 **Score**：102.854
	**ZINC ID**：ZINC02526442 **Score**：128.009
	**ZINC ID**：ZINC03869855 **Score**：109.043
	**ZINC ID**：ZINC04674533 **Score**：113.00

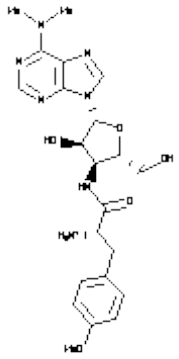
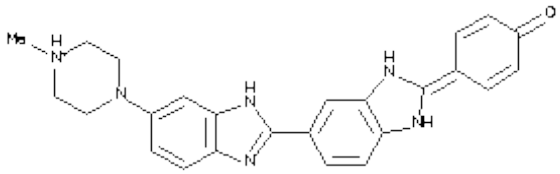
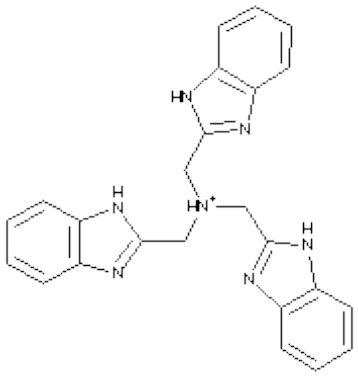
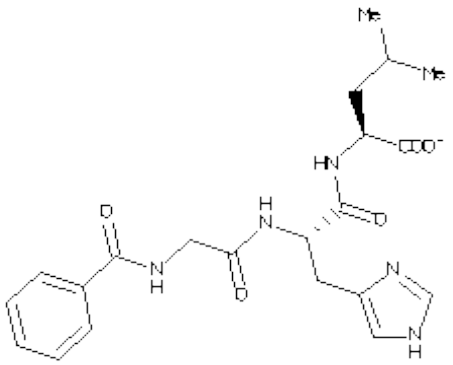
	**ZINC ID**：ZINC14989116 **Score**：116.649
	**ZINC ID**: ZINC03873619 **Score**: 122.625
	**ZINC ID**: ZINC15722209 **Score**: 117.166
	**ZINC ID**: ZINC15657694 **Score**: 115.192

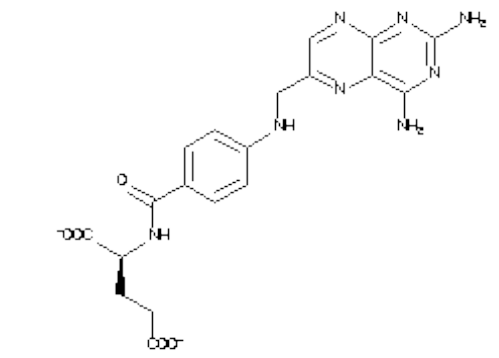
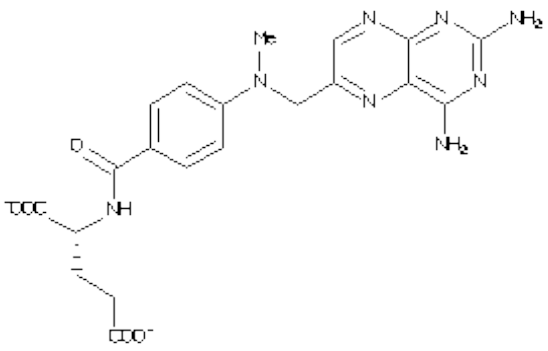
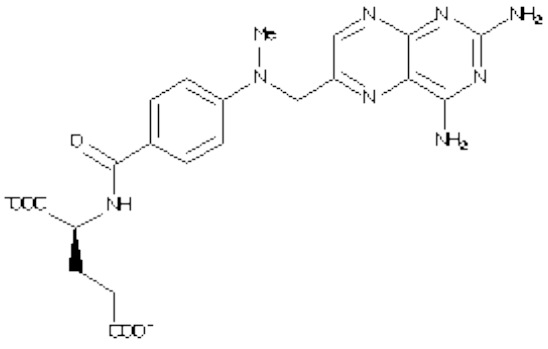
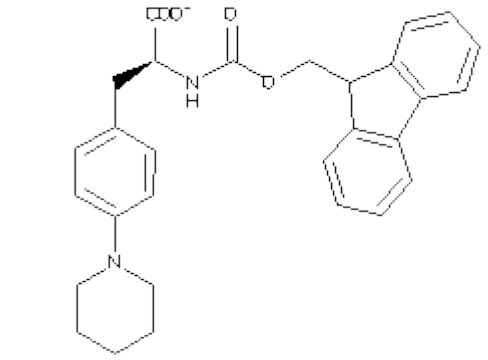
	**ZINC ID**: ZINC02583099 **Score**: 114.743
	**ZINC ID**: ZINC02560905 **Score**: 110.102
	**ZINC ID**: ZINC15722257 **Score**: 107.382
	**ZINC ID**: ZINC15722232 **Score**: 101.696

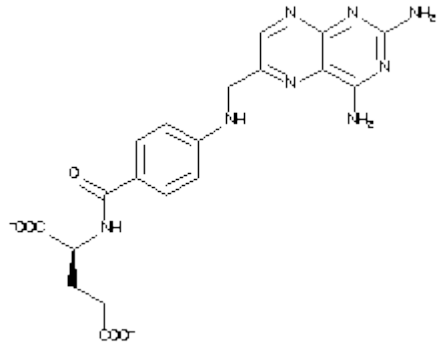
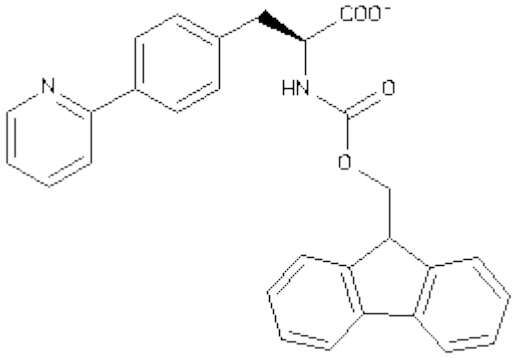
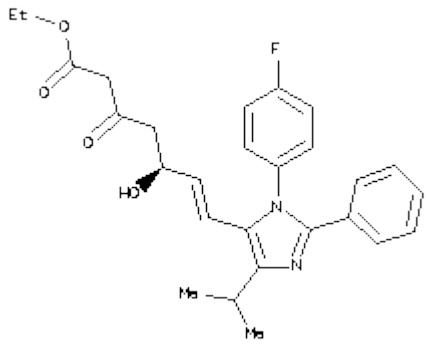
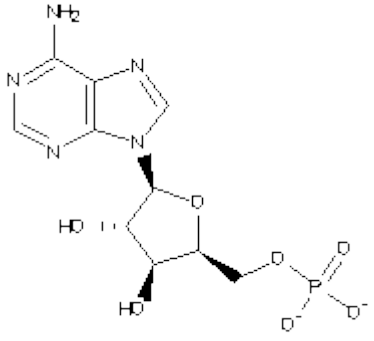
	**ZINC ID**: ZINC02002421 **Score**: 98.3177
	**ZINC ID**: ZINC02547583 **Score**: 95.0994
	**ZINC ID**: ZINC15722251 **Score**: 92.6223
	**ZINC ID**: ZINC15721208 **Score**: 91.1407

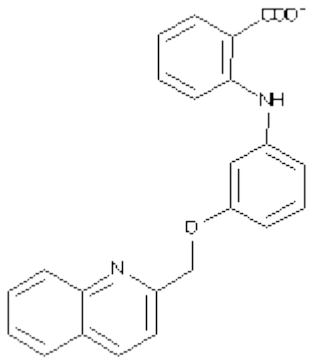
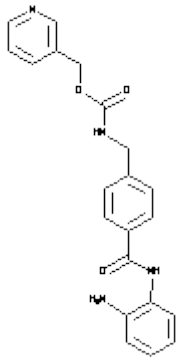
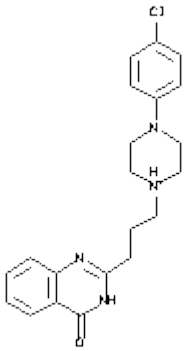
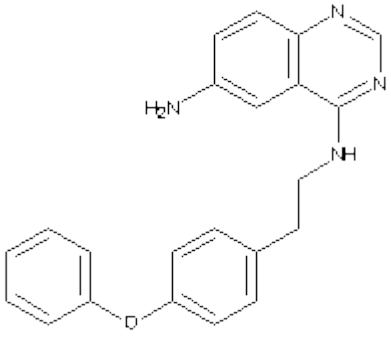
	**ZINC ID**: ZINC25667099 **Score**: 94.488
	**ZINC ID**: ZINC19419017 **Score**: 95.303
	**ZINC ID**: ZINC03830986 **Score**: 91.7811
	**ZINC ID**: ZINC00599734 **Score**: 91.7811

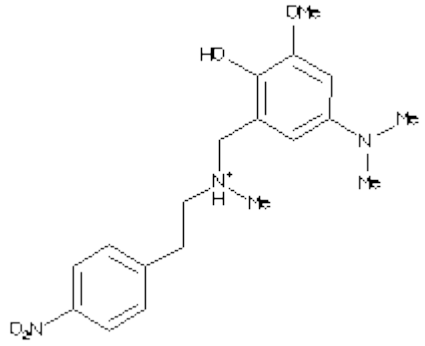
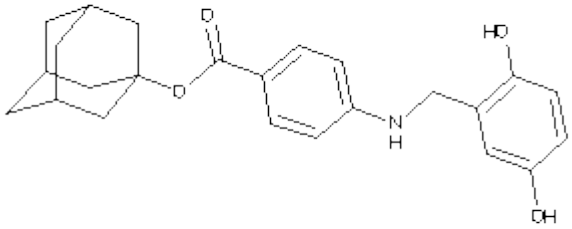
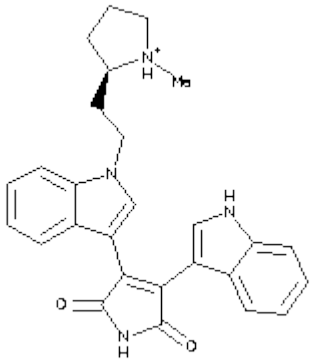
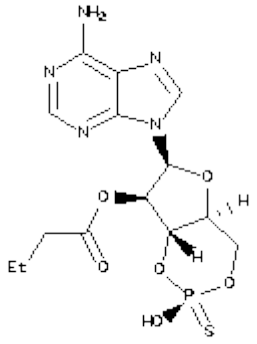
	**ZINC ID**: ZINC01892860 **Score**: 90.4006
	**ZINC ID**:ZINC18456332 **Score**:111.51
	**ZINC ID**: ZINC03861477 **Score**: 101.313
	**ZINC ID**: ZINC04213546 **Score**: 100.98

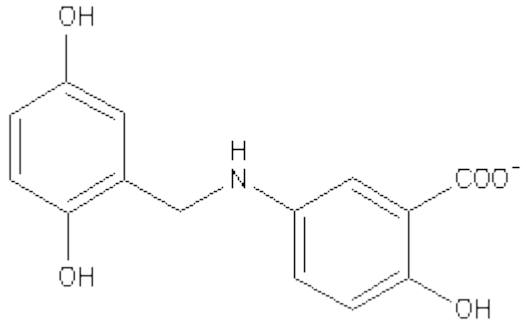
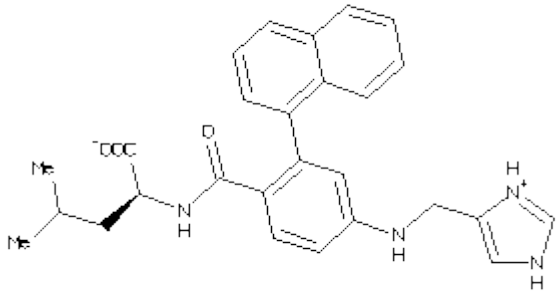
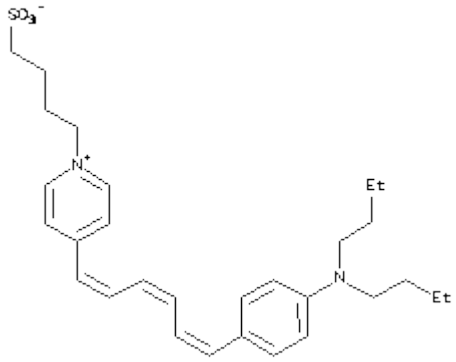
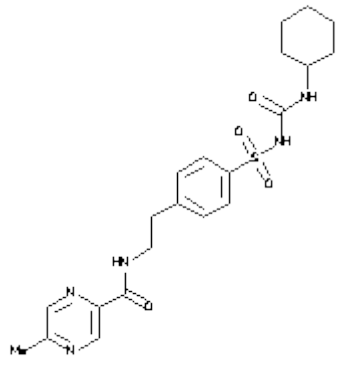
	**ZINC ID**: ZINC03861788 **Score**: 99.7776
	**ZINC ID**: ZINC22912997 **Score**: 99.1555
	**ZINC ID**: ZINC08617147 **Score**: 115.397
	**ZINC ID**: ZINC06566221 **Score**: 112.538

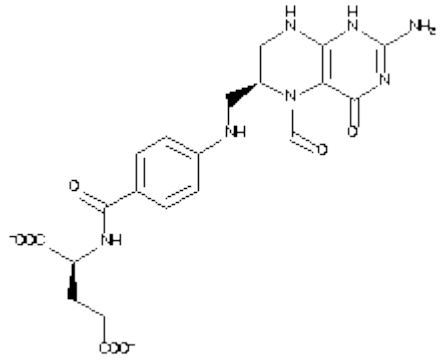
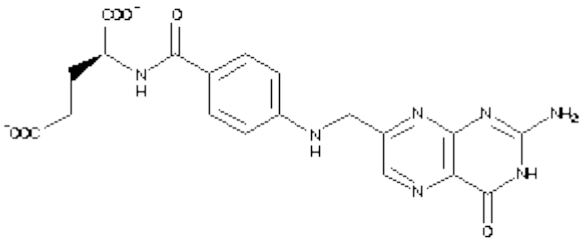
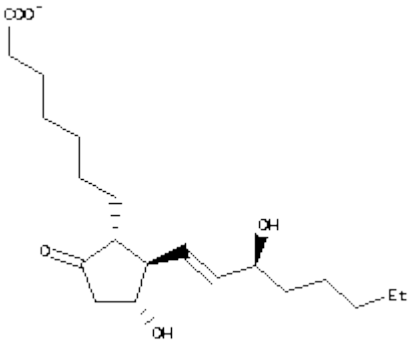
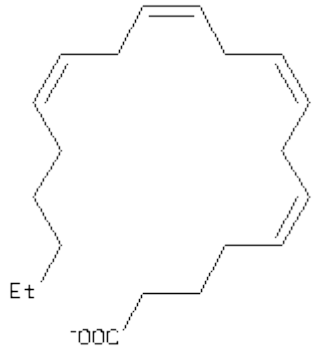
	**ZINC ID**: ZINC02036915 **Score**: 114.859
	**ZINC ID**: ZINC06920406 **Score**: 120.634
	**ZINC ID**: ZINC01529323 **Score**: 122.052
	**ZINC ID**: ZINC12958249 **Score**: 119.59

	**ZINC ID**: ZINC02036915 **Score**: 114.859
	**ZINC ID**: ZINC12958628 **Score**: 113.199
	**ZINC ID**: ZINC12957897 **Score**: 100.462
	**ZINC ID**: ZINC03201893 **Score**: 99.1148

	**ZINC ID**: ZINC00596951 **Score**: 98.7548
	**ZINC ID**: ZINC01488870 **Score**: 105.73
	**ZINC ID**: ZINC03820039 **Score**: 98.2651
	**ZINC ID**: ZINC01489490 **Score**: 97.5368

	**ZINC ID**: ZINC03938697 **Score**: 95.7617
	**ZINC ID**: ZINC03941698 **Score**: 95.3101
	**ZINC ID**: ZINC02560260 **Score**: 92.8234
	**ZINC ID**: ZINC14806350 **Score**: 92.7004

	**ZINC ID**: ZINC00388549 **Score**: 91.4878
	**ZINC ID**: ZINC13796731 **Score**: 91.0907
	**ZINC ID**: ZINC12405163 **Score**: 116.753
	**ZINC ID**: ZINC00537795 **Score**: 114.478

	**ZINC ID**: ZINC18456279 **Score**: 113.363
	**ZINC ID**: ZINC04261891 **Score**: 109.313
	**ZINC ID**: ZINC03813088 **Score**: 105.863
	**ZINC ID**: ZINC04474696 **Score**: 103.915

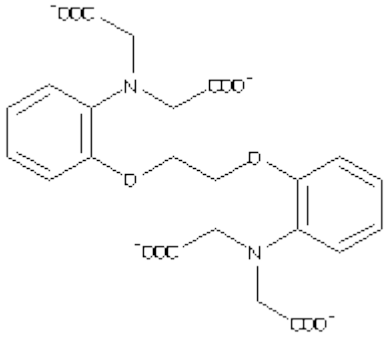
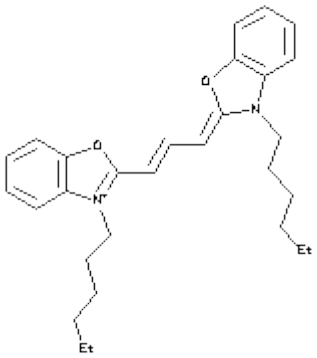
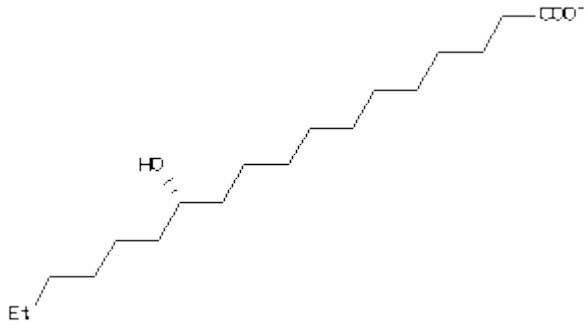
	**ZINC ID**: ZINC03860554 **Score**: 93.4135
	**ZINC ID**: ZINC02167083 **Score**: 92.2656
	**ZINC ID**: ZINC03861566 **Score**: 90.5868

Table S3. The compounds obtained from virtual screening that decreased the RNA-binding capacity of N protein by more than 10%.

No.	Compound	Inhibition of RNA-binding activity
O1	3,3'-Methylenebis(4-hydroxycoumarin)	- [a]
O2	1-(5-Chloro-2-hydroxy-4-methylphenyl)-3-phenyl-1,3-propanedione	-
O3	6-chloro-7-(2-morpholin-4-yl-ethylamino)quinoline-5,8dione	+
O4	1-(5-Bromo-2-hydroxyphenyl)-3-phenyl-1,3-propanedione	-
O5	N-(6-Oxo-5,6-dihydrophenanthridin-2-yl)-(N,N-dimethylamino)acetamide hydrochloride (PJ34)	+
O6	7-Allyl-7,8-dihydro-8-oxoguanosine	-
O7	N,N-Bis(2,5-dihydroxybenzylidene)ethylenediamine	-
O8	6-(chloromethyl)-2-[(4-methoxyphenoxy)methyl]pyrimidin-4-ol	-

[a] No significant inhibition was observed.