Effects of natural polymorphisms in SARS-CoV-2 RNA-dependent RNA polymerase on its activity and sensitivity to inhibitors *in vitro*

Nataliya Miropolskaya, Maxim Kozlov, Ivan Petushkov, Maria Prostova, Danil Pupov, Daria Esyunina, Sergey Kochetkov, Andrey Kulbachinskiy

Supplementary Information

Contains 1 Supplementary Table and 8 Supplementary Figures

Table S1. Frequencies of amino acid substitutions at indicated positions of nsp12. The data were retrived from the GISAID database [21], accessed at the COVID CG site on 25.05.2022, counted for the periods from the start of pandemic until 25.05.2021 (1,893,772 genomic sequences) and 24.05.2022 (11,750,010 genomic sequences). The data for the Omicron lineage for the same period were retrived on 21.09.2022 (2,789,360 Omicron lineage sequences). Substitutions are given in comparison with the reference sequence of the WIV04 SARS-CoV-2 strain (GenBank: MN996528.1). Substitutions analyzed in this study are shown in yellow. The frequencies of substitutions in the Omicron lineage are shown relative to the total number of sequenced Omicron genomes.

24 May 2021, all genomes				25 May 2022, all genomes				25 May 2022, Omicron lineage		
Substitution	# of seq	frequency		Substitution	# of seq	frequency		Substitution	# of seq	frequency
A443S	1699	0.000897		A443S	5580	0.000475		A443S	31	$1.03*10^{-5}$
A443V	911	0.000481		A443V	2190	0.000186		A443V	103	$3.43*10^{-5}$
A443T	3	$1.58*10^{-6}$		A443T	292	$2.49*10^{-5}$		A443T	1	$3.33*10^{-7}$
				A443P	7	$5.96*10^{-7}$				
				A443G	5	$4.26*10^{-7}$				
D445N	513	0.000271		D445N	928	$7.9*10^{-5}$		D445N	61	$2.03*10^{-5}$
D445G	451	0.000238		D445G	1174	$9.99*10^{-5}$		D445G	318	$1.06*10^{-4}$
D445Y	22	$1.16*10^{-5}$		D445Y	112	$9.53*10^{-6}$		D445Y	4	$1.33*10^{-6}$
D445A	3	$1.58*10^{-6}$		D445A	24	$2.04*10^{-6}$		D445V	4	$1.33*10^{-6}$
D445*10	1	$5.28*10^{-7}$		D445C	12	$1.02*10^{-6}$				
				D445*10	12	$1.02*10^{-6}$				
				D445V	9	$7.66*10^{-7}$				
L514F	507	0.000268		L514F	2007	0.000171		L514F	47	$1.56*10^{-5}$
L514I	8	$4.22*10^{-6}$		L514I	147	$1.25*10^{-5}$		L514I	12	$4*10^{-6}$
L514V	3	$1.58*10^{-6}$		L514P	10	$8.51*10^{-7}$		L514P	5	$1.66*10^{-6}$
L514P	2	$1.06*10^{-6}$		L514V	8	$6.81*10^{-7}$				
R583G	1071	0.000566		R583G	1880	0.00016		R583G	80	$2.66*10^{-5}$
R583K	42	$2.22*10^{-5}$		R583K	109	$9.28*10^{-6}$		R583K	36	$1.2*10^{-5}$
R583T	21	$1.11*10^{-5}$		R583S	39	$3.32*10^{-6}$		R583S	24	$8*10^{-6}$
R583S	13	$6.86*10^{-6}$		R583T	31	$2.64*10^{-6}$		R583I	7	$2.33*10^{-6}$
R583I	5	$2.64*10^{-6}$		R583I	25	$2.13*10^{-6}$				
M794V	33	$1.74*10^{-5}$		M794V	88	$7.49*10^{-6}$		M794V	15	$4.99*10^{-6}$
M794I	43	$2.27*10^{-5}$		M794T	28	$2.38*10^{-6}$		M794L	6	$2*10^{-6}$
M794T	4	$2.11*10^{-6}$		M794L	15	$1.28*10^{-6}$		M794I	80	$2.66*10^{-5}$
M794R	1	$5.28*10^{-7}$		M794R	2	$1.7*10^{-7}$				
M794L	1	$5.28*10^{-7}$		M794S	1	$8.51*10^{-8}$				
				M794Q	1	$8.51*10^{-8}$				
				M794K	1	$8.51*10^{-8}$				
S795F	219	0.000116		S795F	2790	0.000237		S795F	90	$3*10^{-5}$
S795Y	10	$5.28*10^{-6}$		S795A	54	$4.6*10^{-6}$		S795A	19	$6.33*10^{-6}$
S795P	9	$4.75*10^{-6}$		S795Y	43	$3.66*10^{-6}$		S795Y	5	$1.66*10^{-6}$
S795C	9	$4.75*10^{-6}$		S795P	38	$3.23*10^{-6}$		S795P	6	$2*10^{-6}$
S795A	2	$1.06*10^{-6}$		S795C	38	$3.23*10^{-6}$		S795C	10	$3.33*10^{-6}$
				S795T	2	$1.7*10^{-7}$				
N911K	38	$2.01*10^{-5}$		N911K	368	$3.13*10^{-5}$		N911K	109	$3.63*10^{-5}$
N911S	76	$4.01*10^{-5}$		N911S	354	$3.01*10^{-5}$		N911S	62	$2.06*10^{-5}$
N911D	23	$1.21*10^{-5}$		N911D	96	$8.17*10^{-6}$		N911D	19	$6.33*10^{-6}$
N911T	5	$2.64*10^{-6}$		N911T	46	$3.91*10^{-6}$				
N911Y	2	$1.06*10^{-6}$		N911Y	12	$1.02*10^{-6}$				
				N911H	7	$5.96*10^{-7}$				

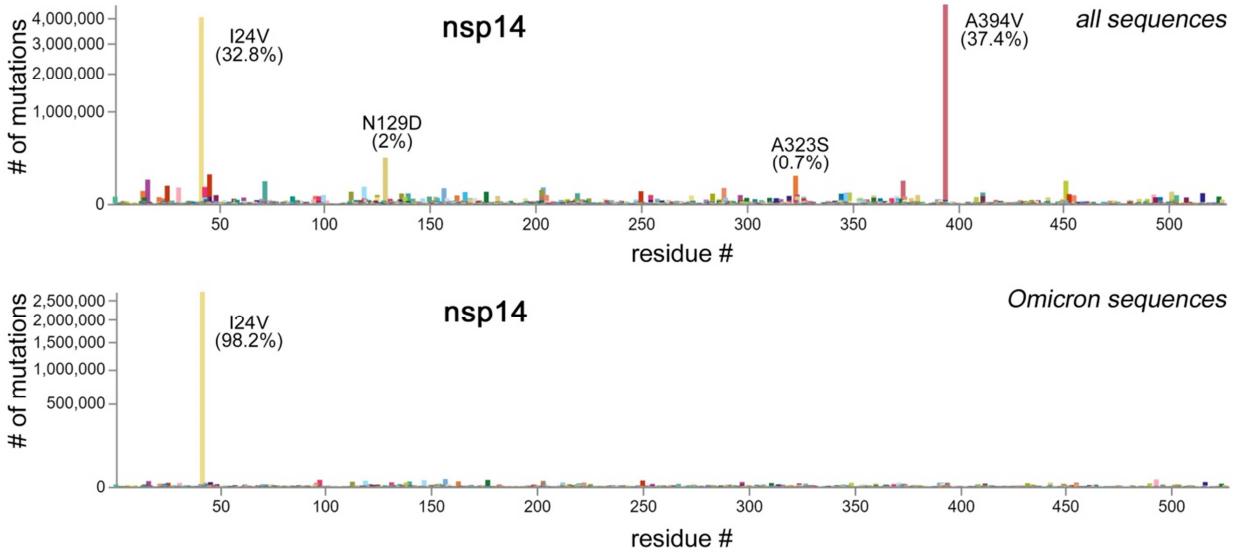


Figure S1. Frequencies of amino acid substitutions found in nsp14 for the period from the start of pandemic until 25.05.2022, for all sequenced SARS-CoV-2 genomes (*top*, 12,365,901 genomes) and for the Omicron lineage (*bottom*, 2,789,360 genomes). The data were retrieved from the GISAID database on 21.09.2022 [21]. The number of sequenced genomes containing substitutions at each nsp14 position is shown on the y-axis (log scale). Substitutions are shown relative to the reference strain WIV04. Most frequent substitutions and their percentage among all SARS-CoV-2 sequences are indicated.

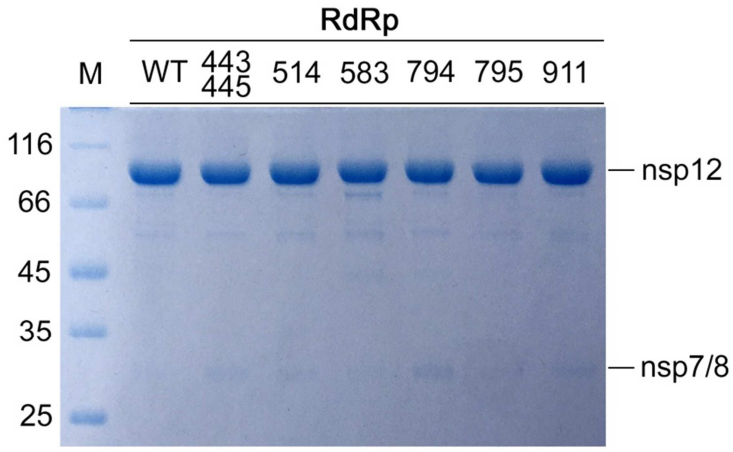


Figure S2. SDS-PAGE analysis of wild-type and mutant RdRp variants analyzed in this study. Each RdRp sample was purified by Ni-affinity (HiTrap TALON crude column) and anion exchange (HiScreenQ HP column) chromatography. Positions of the nsp12 and fused nsp7-nsp8 subunits are indicated.

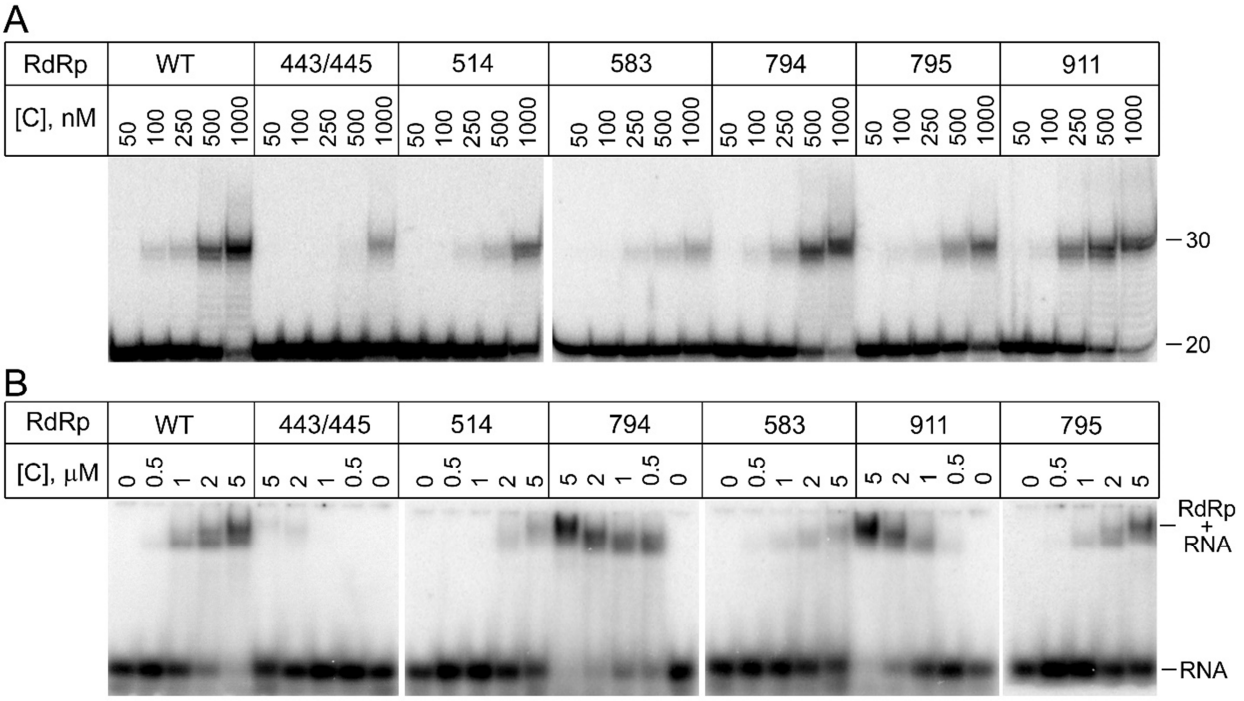


Figure S3. Analysis of the relative activity and RNA binding by the wild-type and mutant RdRp variants. (A) Analysis of RNA extension at different RdRp concentrations. Positions of the starting RNA primer (20 nt) and the extended RNA product (30 nt) are indicated. (B) Analysis of RNA binding by the wild-type and mutant RdRp variants. Increasing amounts of RdRp were incubated with pre-annealed primer-template duplex containing a 5'-P^{32}-labeled RNA primer and the samples were run on native 5% PAGE. Positions of the free RNA substrate and the RdRp-RNA complex are indicated.

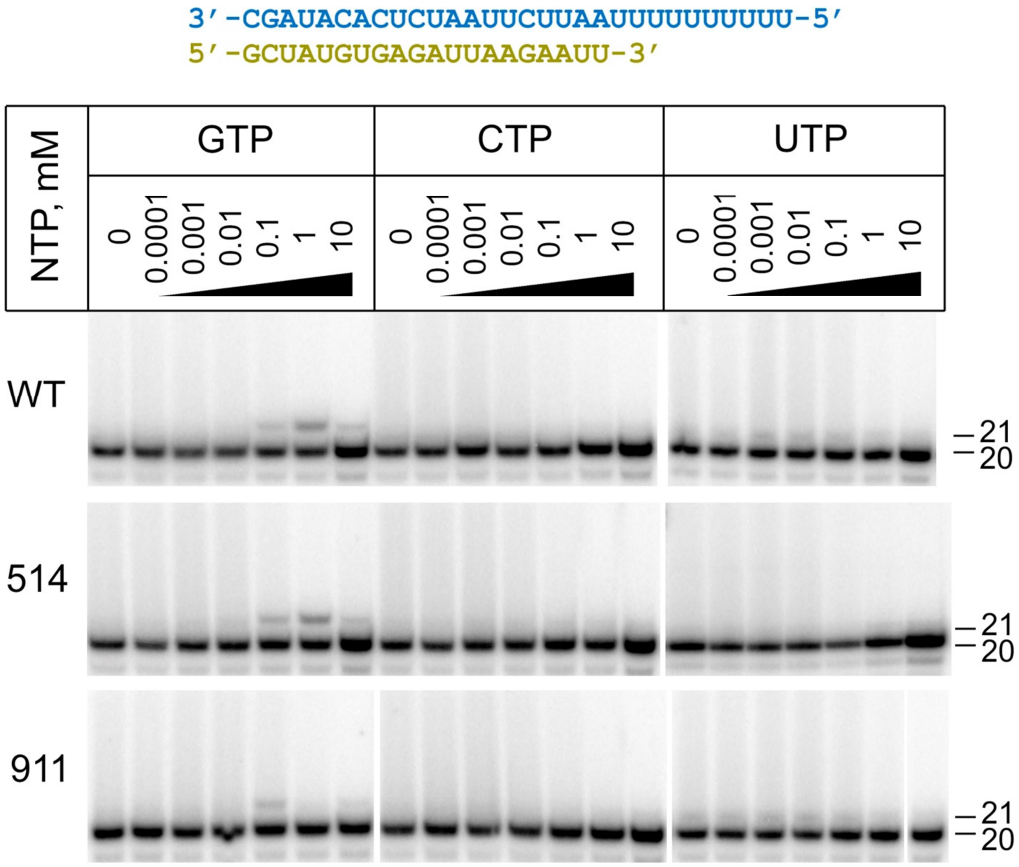


Figure S4. Analysis of nucleotide misincorporation by the wild-type and L514F and N911K RdRp variants. The RNA substrate is shown on the top (primer, yellow; template, blue). RdRp (500 nM) was incubated with primer-template RNA (50 nM) for 5' at 30°C, individual non-complementary NTPs were added to indicated concentrations () and RNA extension was performed for 10 seconds at 30 °C. The results of a single measurement are presented. Positions of the starting RNA primer (20 nt) and the extended RNA product (21 nt) are indicated.

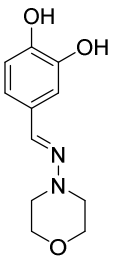
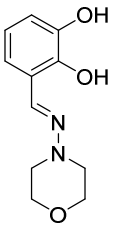
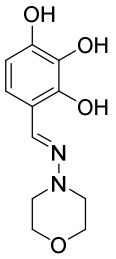
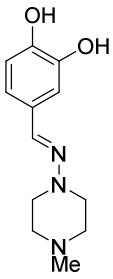
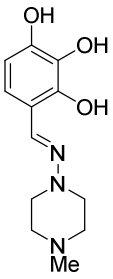
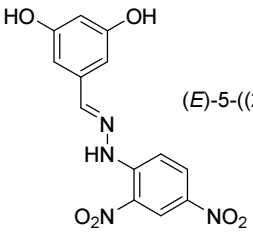
Figure S5. Structures of non-nucleoside compounds tested in this study.
Pages 8-12.

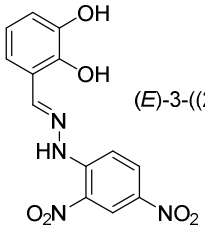
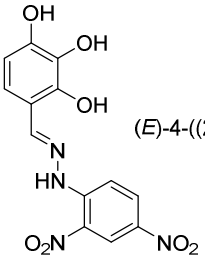
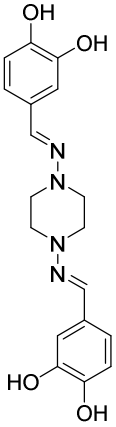
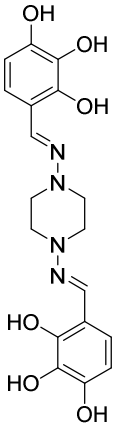
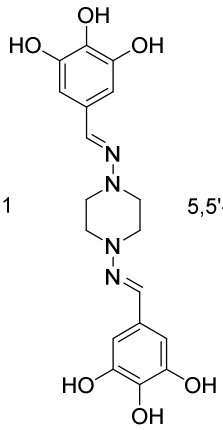
(A) Hydrazons of polyhydroxybenzaldehydes (PP). (B) Diketoacids (DA). (C) Analogs of diketoacids (DAA).
(D) Hydroxyquinolines (HQ).

For each compound, its abbreviation, chemical structure and chemical name are shown. For DAA and HQ, ^{1}H NMR spectra are presented.

A

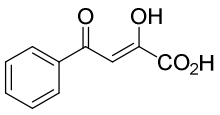
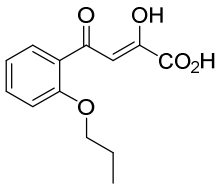
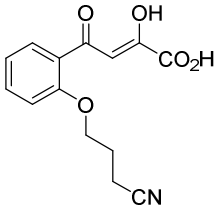
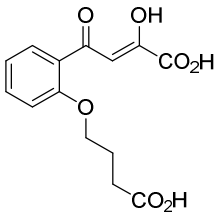
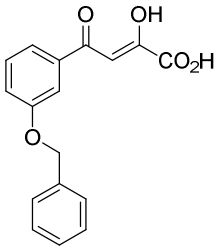
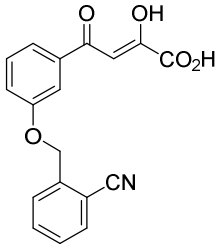
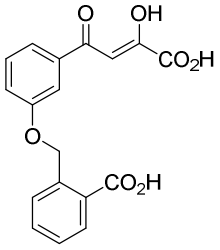
Polyphenols (PPs)

Abr.	*Structure*	*Name*
PP1		(*E*)-4-((morpholinoimino)methyl)benzene-1,2-diol
PP2		(*E*)-3-((morpholinoimino)methyl)benzene-1,2-diol
PP3		(*E*)-4-((morpholinoimino)methyl)benzene-1,2,3-triol
PP4		(*E*)-4-(((4-methylpiperazin-1-yl)imino)methyl)benzene-1,2-diol
PP5		(*E*)-4-(((4-methylpiperazin-1-yl)imino)methyl)benzene-1,2,3-triol
PP6		(*E*)-5-((2-(2,4-dinitrophenyl)hydrazineylidene)methyl)benzene-1,3-diol

Abr.	*Structure*	*Name*
PP7		(*E*)-3-((2-(2,4-dinitrophenyl)hydrazineylidene)methyl)benzene-1,2-diol
PP8		(*E*)-4-((2-(2,4-dinitrophenyl)hydrazineylidene)methyl)benzene-1,2,3-triol
PP9		4,4'-((1*E*,1'*E*)-(piperazine-1,4-diylbis(azaneylylidene))bis(methaneylylidene))bis(benzene-1,2-diol)
PP10		4,4'-((1*E*,1'*E*)-(piperazine-1,4-diylbis(azaneylylidene))bis(methaneylylidene))bis(benzene-1,2,3-triol)
PP11		5,5'-((1*E*,1'*E*)-(piperazine-1,4-diylbis(azaneylylidene))bis(methaneylylidene))bis(benzene-1,2,3-triol)

B

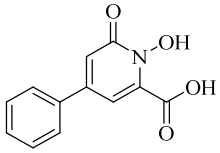
Diketo Acids (DAs)

Abr.	*Structure*	*Name*
DA1		(*Z*)-2-hydroxy-4-oxo-4-phenylbut-2-enoic acid
DA2		(*Z*)-2-hydroxy-4-oxo-4-(2-propoxyphenyl)but-2-enoic acid
DA3		(*Z*)-4-(2-(3-cyanopropoxy)phenyl)-2-hydroxy-4-oxobut-2-enoic acid
DA4		(*Z*)-4-(2-(3-carboxypropoxy)phenyl)-2-hydroxy-4-oxobut-2-enoic acid
DA5		(*Z*)-4-(3-(benzyloxy)phenyl)-2-hydroxy-4-oxobut-2-enoic acid
DA6		(*Z*)-4-(3-((2-cyanobenzyl)oxy)phenyl)-2-hydroxy-4-oxobut-2-enoic acid
DA7		(*Z*)-2-((3-(3-carboxy-3-hydroxyacryloyl)phenoxy)methyl)benzoic acid

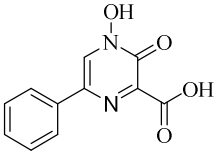
C

Diketo Acid Analogs (DAAs)

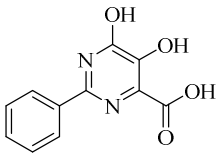
Abr. *Structure* *Name*

DAA1  1-hydroxy-6-oxo-4-phenyl-1,6-dihydropyridine-2-carboxylic acid

^{1}H NMR (400 MHz, DMSO-d_6)
δ 7.79–7.71 (m, 2H), 7.53–7.45 (m, 3H), 7.00 (d, *J* = 2.3 Hz, 1H), 6,97 (d, *J* = 2.4 Hz, 1H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 161.80, 157.14, 147.50, 138.97, 135.90, 129.58, 129.03, 126.71, 115. 95, 104.45.

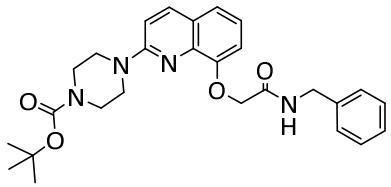
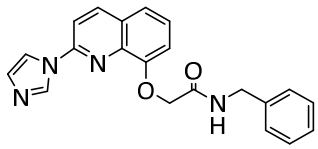
DAA2  4-hydroxy-3-oxo-6-phenyl-3,4-dihydropyrazine-2-carboxylic acid

^{1}H NMR (400 MHz, DMSO-d_6)
δ 8.91 (s, 1H), 7.92 (d, *J* = 7.4 Hz, 2H), 7.45 (t, *J* = 7.6 Hz, 2H), 7.36 (t, *J* = 7.3 Hz, 1H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 164.19, 150.74, 142.86, 134.41, 131.75, 128.70, 128.13, 128.05, 124.99.

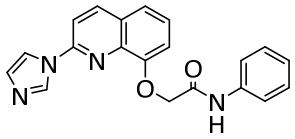
DAA3 (Cmpd3)  5,6-dihydroxy-2-phenylpyrimidine-4-carboxylic acid

D

Hydroxyquinolines (HQs)

Abr.	*Structure*	*Name*
HQ1 (Cmpd13e)		*tert*-butyl 4-(8-(2-(benzylamino)-2-oxoethoxy)quinolin-2-yl) piperazine-1-carboxylate
HQ2		2-((2-(1*H*-imidazol-1-yl)quinolin-8-yl)oxy)-*N*-benzylacetamide

^{1}H NMR (300 MHz, DMSO-d_6)
δ 8.71 (s, 1H), 8.58 (d, *J* = 8.9 Hz, 2H), 8.10 – 8.01 (m, 2H), 7.66 (dd, *J* = 8.2, 1.0 Hz, 1H), 7.53 (t, *J* = 7.9 Hz, 1H), 7.38 – 7.20 (m, 6H), 7.16 (s, 1H), 4.86 (s, 2H), 4.41 (d, *J* = 6.1 Hz, 2H). ^{13}C NMR (75 MHz, DMSO-d_6) δ 168.24, 153.45, 147.09, 140.67, 139.52, 138.45, 136.02, 130.72, 128.76, 128.41, 127.72, 127.36, 126.95, 121.42, 117.17, 113.88, 113.35, 69.42, 42.41.

HQ3		2-((2-(1*H*-imidazol-1-yl)quinolin-8-yl)oxy)-*N*-phenylacetamide

^{1}H NMR (300 MHz, DMSO-d_6)
δ 10.14 (s, 1H), 8.74 (s, 1H), 8.60 (d, *J* = 8.9 Hz, 1H), 8.15 (s, 1H), 8.07 (d, *J* = 8.8 Hz, 1H), 7.65 (t, *J* = 8.3 Hz, 3H), 7.54 (t, *J* = 7.9 Hz, 1H), 7.38 – 7.29 (m, 3H), 7.19 (s, 1H), 7.09 (t, *J* = 7.4 Hz, 1H), 5.02 (s, 2H). ^{13}C NMR (75 MHz, DMSO-d_6) δ 167.11, 153.64, 147.13, 140.72, 138.78, 138.51, 136.03, 130.74, 129.25, 128.49, 127.01, 124.20, 121.41, 120.02, 117.33, 113.91, 113.48, 69.41.

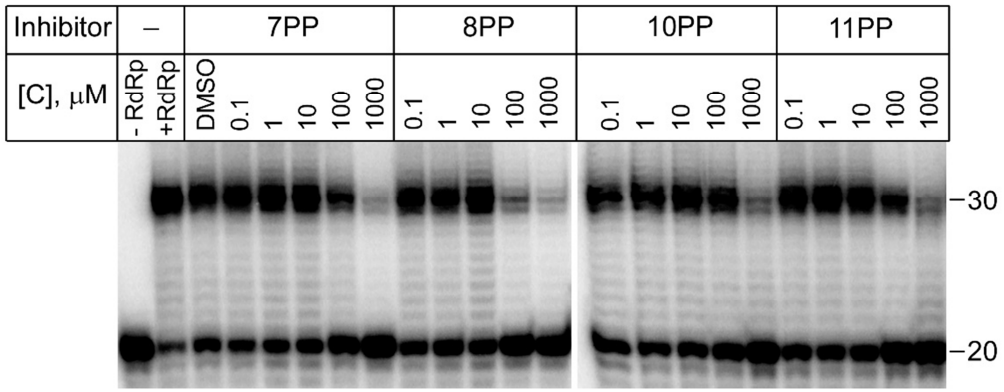


Figure S6. Titration of selected polyphenol compounds (PP). Preformed complex of wild-type RdRp with the RNA substrate was incubated with increasing concentrations of indicated compounds dissolved in DMSO, and RNA extension was analyzed after ATP addition. Control reactions contained either RdRp in the absence of inhibitors (“+RdRp”) or RdRp and DMSO only (“DMSO”). Positions of the starting RNA primer (20 nt) and the extended RNA product (30 nt) are indicated.

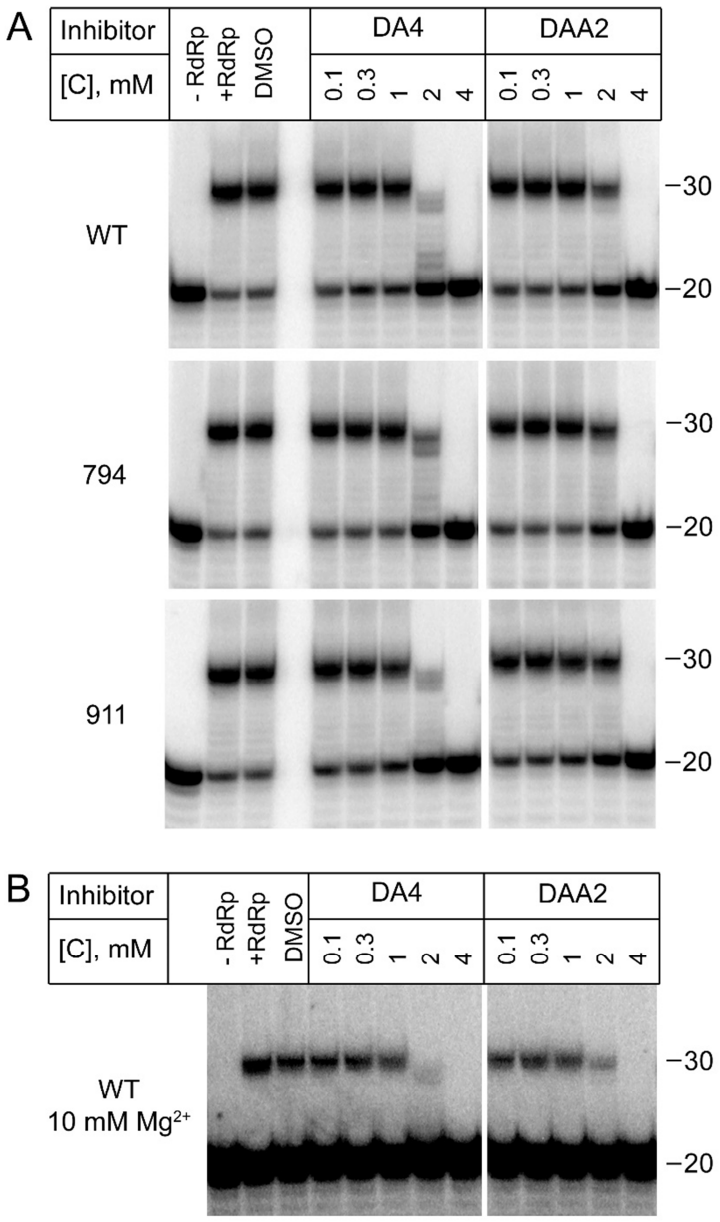


Figure S7. Titration of selected DA and DAA compounds. The reactions were performed in the reaction buffer containing 2 mM $MgCl_2$ (A) or 2 mM $MgCl_2$ (B). Preformed complex of wild-type or mutant (M794V and N911K) RdRp variants with the RNA substrate was incubated with increasing concentrations of indicated compounds dissolved in DMSO, and RNA extension was analyzed after ATP addition. Control reactions contained either RdRp in the absence of inhibitors ("+RdRp") or RdRp and DMSO ("DMSO"). Positions of the starting RNA primer (20 nt) and the extended RNA product (30 nt) are indicated.

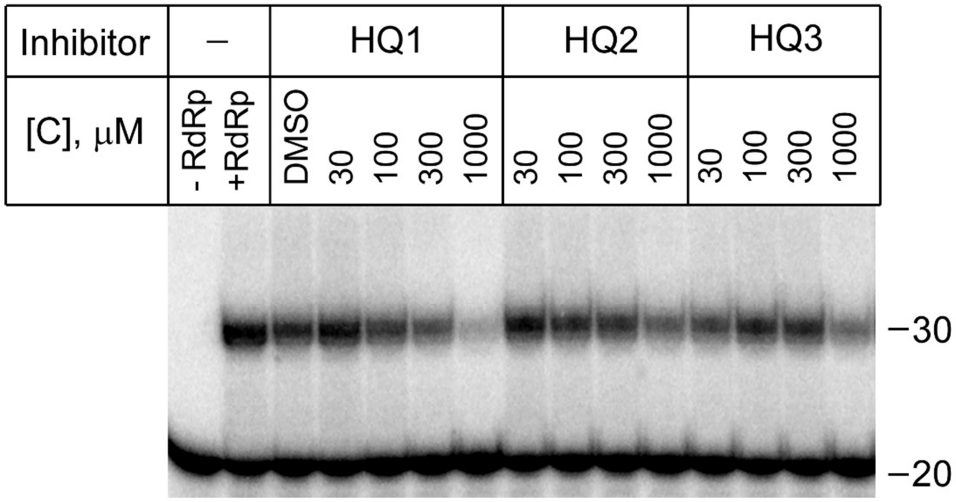


Figure S8. Titration of HQ compounds. Preformed complex of wild-type RdRp with the RNA substrate was incubated with increasing concentrations of indicated compounds dissolved in DMSO, and RNA extension was analyzed after ATP addition. Control reactions contained either RdRp in the absence of inhibitors ("+RdRp") or RdRp and DMSO only ("DMSO"). Positions of the starting RNA primer (20 nt) and the extended RNA product (30 nt) are indicated.