

Supplementary Information

Nucleotide Analogues as Inhibitors of SARS-CoV Polymerase

Jingyue Ju^{1,2,3,*}, Xiaoxu Li^{1,2}, Shiv Kumar^{1,2}, Steffen Jockusch^{1,4}, Minchen Chien^{1,2}, Chuanjuan Tao^{1,2}, Irina Morozova^{1,2}, Sergey Kalachikov^{1,2}, Robert N. Kirchdoerfer^{5,6}, James J. Russo^{1,2}

¹Center for Genome Technology and Biomolecular Engineering, Columbia University, New York, NY 10027; Departments of ²Chemical Engineering, ³Molecular Pharmacology and Therapeutics, and ⁴Chemistry, Columbia University, New York, NY 10027; ⁵Department of Biochemistry and ⁶Institute of Molecular Virology, University of Wisconsin-Madison, Madison, WI 53706

*To whom correspondence should be addressed. Email: dj222@columbia.edu

Table of Contents

1. **Figure S1.** Protein sequence alignments for nsp7, nsp8, and nsp12: SARS-CoV vs SARS-CoV-2.
2. **Figure S2.** Incorporation of UTP by SARS-CoV RNA-dependent RNA polymerase.

(A) nsp7

Query: AGT21317.1:3837-3919 replicase polyprotein 1ab [SARS coronavirus wtic-MB] Query ID: lcl|Query_59781 SARS-CoV Length: 83

>QHD43415_7 (L=83) nsp7 SARS-CoV-2
Sequence ID: Query_59783 Length: 83
Range 1: 1 to 83

Score:162 bits(410), Expect:2e-59,
Method: Compositional matrix adjust.,
Identities:82/83(99%), Positives:83/83(100%), Gaps:0/83(0%)

```
Query 1 SKMSDVKCTSVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSVLLSVLL 60
      SKMSDVKCTSVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSVLLSVLL
Sbjct 1 SKMSDVKCTSVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEAFEKMSVLLSVLL 60

Query 61 SMQGAVDINRLCEEMLDNRATLQ 83
      SMQGAVDIN+LCEEMLDNRATLQ
Sbjct 61 SMQGAVDINKLCEEMLDNRATLQ 83
```

(B) nsp8

Query: AGT21317.1:3920-4117 replicase polyprotein 1ab [SARS coronavirus wtic-MB] Query ID: lcl|Query_11547 Length: 198

>QHD43415_8 (L=198) nsp8 SARS-CoV-2 YP_009725304.1
Sequence ID: Query_11549 Length: 198
Range 1: 1 to 198

Score:396 bits(1018), Expect:5e-148,
Method: Compositional matrix adjust.,
Identities:193/198(97%), Positives:196/198(98%), Gaps:0/198(0%)

```
Query 1 AIASEFSSLPSYAA+YATAQEAYEQAVANGDSEVVLKLLKSLNVAKSEFDRDAAMQRKLE 60
      AIASEFSSLPSYAA+ATAQEAYEQAVANGDSEVVLKLLKSLNVAKSEFDRDAAMQRKLE
Sbjct 1 AIASEFSSLPSYAAFATAQEAYEQAVANGDSEVVLKLLKSLNVAKSEFDRDAAMQRKLE 60

Query 61 KMADQAMTQMYQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNII 120
      KMADQAMTQMYQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNII
Sbjct 61 KMADQAMTQMYQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNII 120

Query 121 PLTTAAKLMVV+VPDY+GTYKNTCDGNTFTYASALWEIQVVDADSKIVQLSEI+MDNSPNL 180
      PLTTAAKLMVV+PDY TYKNTCDG TFTYASALWEIQVVDADSKIVQLSEI+MDNSPNL
Sbjct 121 PLTTAAKLMVVIPDYNTYKNTCDGTTFTYASALWEIQVVDADSKIVQLSEISMDNSPNL 180

Query 181 AWPLIVTALRANSVAVKLQ 198
      AWPLIVTALRANSVAVKLQ
Sbjct 181 AWPLIVTALRANSVAVKLQ 198
```

(C) nsp12

Query: AGT21317.1:4370-5301 replicase polyprotein 1ab [SARS coronavirus wtic-MB] SARS-CoV Query ID: lcl|Query_33851 Length: 932

>QHD43415_11 (L=932 aa) nsp12 RNA-directed RNA polymerase (RdRp) SARS-CoV-2
Sequence ID: Query_33853 Length: 932
Range 1: 1 to 932

Score:1894 bits(4907), Expect:0.0,
Method: Compositional matrix adjust.,
Identities:898/932(96%), Positives:916/932(98%), Gaps:0/932(0%)

```
Query 1 SADASTFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNEKVAGFAKFLKTNCCRFQEKD 60
      SADA +FLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYN+KVAGFAKFLKTNCCRFQEKD
Sbjct 1 SADAQSFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNDKVAGFAKFLKTNCCRFQEKD 60

Query 61 EGNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFFKFRVDGDMVPHISRQRLT 120
      E+ NL+DSYFVVKRHT SNYQHEETIYNL+KDCPAVA HDFFKFR+DGDMPHISRQRLT
Sbjct 61 EDDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQRLT 120

Query 121 KYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGE 180
      KYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGE
Sbjct 121 KYTMADLVYALRHFDEGNCDTLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGE 180

Query 181 RVRQSLKKTQVFCAMRDAGIVGVLTLDNQDLNGNWDYFGDFVQVAPGCGVPIVDSYYSL 240
      RVRQ+LLKKTQVFCAMR+AGIVGVLTLDNQDLNGNWDYFGDF+Q PG GVP+VDSYYSL
Sbjct 181 RVRQALLKKTQVFCAMRNAGIVGVLTLDNQDLNGNWDYFGDFIQTTPGSGVPVDSYYSL 240

Query 241 LMPILTLTRALAAESHMDADLAKPLIKWDLKLYDFTEERLCLFDRYFKYWDQTYHPNCIN 300
      LMPILTLTRAL AESH+D DL KP IKWDLKLYDFTEERL LDRYFKYWDQTYHPNC+N
Sbjct 241 LMPILTLTRALTAESHVDTDLTKPYIKWDLKLYDFTEERLKLDRYFKYWDQTYHPNCVN 300

Query 301 CLDDRCILHCANFNVLVSTVFPPTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVN 360
      CLDDRCILHCANFNVLVSTVFPPTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVN
Sbjct 301 CLDDRCILHCANFNVLVSTVFPPTSFGPLVRKIFVDGVPFVSTGYHFRELGVVHNQDVN 360

Query 361 LHSSRSLFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFY 420
      LHSSRSLFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFY
Sbjct 361 LHSSRSLFKELLVYAADPAMHAASGNLLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFY 420

Query 421 DFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYRYNLPMTCDIRQLLFWVEVVDKYF 480
      DFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYRYNLPMTCDIRQLLFWVEVVDKYF
Sbjct 421 DFAVSKGFFKEGSSVELKHFFFAQDGNAAISDYDYRYNLPMTCDIRQLLFWVEVVDKYF 480

Query 481 DCYDGGCINANQVIVNNLDKSAGFPFNKWKARLYDSMSYEDQDALFAYTKRNVIPITIT 540
      DCYDGGCINANQVIVNNLDKSAGFPFNKWKARLYDSMSYEDQDALFAYTKRNVIPITIT
Sbjct 481 DCYDGGCINANQVIVNNLDKSAGFPFNKWKARLYDSMSYEDQDALFAYTKRNVIPITIT 540
```

Motif G

Query	541	QMNLKYAISAKNRARTVAGVSI	CSTMTNRQFHQKLLKSIAATR	GATVWIGTSK	FYGGWHN	600		
		QMNLKYAISAKNRARTVAGVSI	CSTMTNRQFHQKLLKSIAATR	GATVWIGTSK	FYGGWHN	600		
Sbjct	541	QMNLKYAISAKNRARTVAGVSI	CSTMTNRQFHQKLLKSIAATR	GATVWIGTSK	FYGGWHN	600		
							Motif F	
Query	601	MLKTVYSDVE	TPHLMGWDYPKCDRAMPNMLR	IMASLVLARKH	N	TCCNLSHRFYRLANECA	660	
		MLKTVYSDVE	PHLMGWDYPKCDRAMPNMLR	IMASLVLARKH	TCC	+LSHRFYRLANECA		
Sbjct	601	MLKTVYSDVE	NPHLMGWDYPKCDRAMPNMLR	IMASLVLARKH	T	TCCSLSHRFYRLANECA	660	
							Motif A	
Query	661	QVLSEMVMCGGSLYVKPGGTSS	GATTAYANSVFNICQAVTANVN	ALLSTDGNKIADKYV	720			
		QVLSEMVMCGGSLYVKPGGTSS	GATTAYANSVFNICQAVTANVN	ALLSTDGNKIADKYV	720			
Sbjct	661	QVLSEMVMCGGSLYVKPGGTSS	GATTAYANSVFNICQAVTANVN	ALLSTDGNKIADKYV	720			
							Motif B	
Query	721	RNLQHRLYECLYRNRD	VDFHEFVDEFYAYLRKHF	SMMILSDDAVVC	YNS	N	YAAQGLVASIK	
		RNLQHRLYECLYRNRD	+FV+EFYAYLRKHF	SMMILSDDAVVC	+NS	YA	+QGLVASIK	
Sbjct	721	RNLQHRLYECLYRNRD	TD	FVNEFYAYLRKHF	SMMILSDDAVVC	F	N	STYASQGLVASIK
							Motif C	
							Motif D	
Query	781	NFKAVLYYQNNVFMSEAK	WTETDLTKGPHEFCSQHT	MLVKQGDDYVYL	PYPDPSRILGA	840		
		NFK+VLYYQNNVFMSEAK	WTETDLTKGPHEFCSQHT	MLVKQGDDYVYL	PYPDPSRILGA			
Sbjct	781	NFKSVLYYQNNVFMSEAK	WTETDLTKGPHEFCSQHT	MLVKQGDDYVYL	PYPDPSRILGA	840		
							Motif D	
							Motif E	
Query	841	GCFVDDIVKTDGTLMIER	FVSLAIDAYPLTKHPNQEY	ADVFHLYLQYIRKLH	DEL	TGHML	900	
		GCFVDDIVKTDGTLMIER	FVSLAIDAYPLTKHPNQEY	ADVFHLYLQYIRKLH	DEL	TGHML		
Sbjct	841	GCFVDDIVKTDGTLMIER	FVSLAIDAYPLTKHPNQEY	ADVFHLYLQYIRKLH	DEL	TGHML	900	
Query	901	DMYSVMLTNDNTSRYWEPE	FYEAMYPHTVLQ	932				
		DMYSVMLTNDNTSRYWEPE	FYEAMYPHTVLQ	932				
Sbjct	901	DMYSVMLTNDNTSRYWEPE	FYEAMYPHTVLQ	932				

Figure S1. Protein sequence alignments for nsp7, nsp8, and nsp12: SARS-CoV vs SARS-CoV-2. Protein sequences are from NCBI Protein Database, accession ID's as indicated. Sequences were aligned with blastp.^{S1} Consensus is shown in blue between the query and subject sequences; positions of amino acid substitutions are in red; + indicates conservative amino acid substitutions involving amino acids with close physico-chemical properties. (A) nsp7; (B) nsp8; (C) nsp12: functional motifs^{S2} are shown as colored bars underneath the aligned sequences. Comparison of the polymerase complex components (nsp7, nsp8, and nsp12 proteins) shows that these proteins are very similar in SARS-CoV and SARS-CoV-2. There are no indels in any of the three protein pairs. There is only one amino acid substitution in nsp7 (99% sequence identity); nsp8 has 5 amino acid changes (97% sequence identity), out of which 3 are between amino acids with similar properties. Alignment of the nsp12 pair shows that 898 out of 932 amino acids (96%) are identical between SARS-CoV and SARS-CoV-2. Eighteen of the substitutions are between amino acids with similar physico-chemical properties, therefore the level of similarity is higher, at 98%. Most of the amino acid substitutions (24 out of 34) are located within the N-terminal portion of the nsp12 protein. This region corresponds to the NiRAN domain (nidovirus RdRp-associated nucleotidyltransferase; approximately amino acids 1 through 250) which is also less conservative in other coronaviruses.^{S3} Within the next region (the interface domain, aa ~250 through 400), the first 15 amino acid positions have multiple substitutions, but the rest of the interface domain is quite conservative. The region beyond the interface domain, corresponding to the nsp12 C-terminus, contains polymerase functional domains. These domains constitute the canonical *fingers*, *palm*, and *thumb* of the polymerase enzyme and contain several motifs that are conservative among coronaviruses (Motifs A through F). Out of the 34 amino acid substitutions in the nsp12 between SARS-CoV and SARS-CoV-2, only three substitutions are located within these motifs, and all three are between similar amino acids.

RNA Template 5'-UGGUGGACCCUCAGAUUCAACUGGCAGUAACCGAAUGGAGAACGCAGUGG-3'
 RNA Primer (MW = 6892) 3'-GACCGUCAUUGGUCUUACCUCU-5'

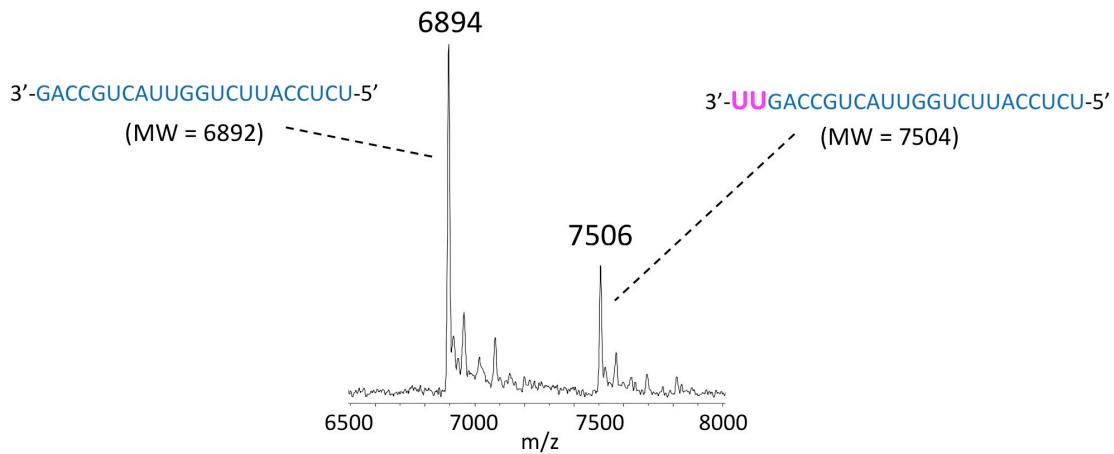


Figure S2. Incorporation of UTP by SARS-CoV RNA-dependent RNA polymerase. The sequence of the primer and template used for this extension reaction is shown at the top of the figure. Polymerase extension reactions were performed by incubating UTP with pre-assembled SARS-CoV polymerase (nsp12, nsp7 and nsp8), the indicated RNA template and primer, and the appropriate reaction buffer, followed by detection of reaction products by MALDI-TOF MS. The detailed procedure is shown in the methods. The accuracy for m/z determination is ± 10 Da.

Reference:

S1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).

S2. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature Commun.* 2019;10:2342. <https://doi.org/10.1038/s41467-019-10280-3>.

S3. Selisko B, Papageorgiou N, Ferron F, Canard B. Structural and functional basis of the fidelity of nucleotide selection by Flavivirus RNA-dependent RNA polymerases. *Viruses.* 2018;10:59. <https://doi.org/10.3390/v10020059>.