Supporting Information for "Identification of Potential Inhibitors of the SARS-CoV-2 NSP13 Helicase via Structure-Based Ligand Design, Molecular Docking and Nonequilibrium Alchemical Simulations"

Giorgio Di Paco, Marina Macchiagodena , and Piero Procacci*

Dipartimento di Chimica "Ugo Schiff", Università degli Studi di Firenze, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

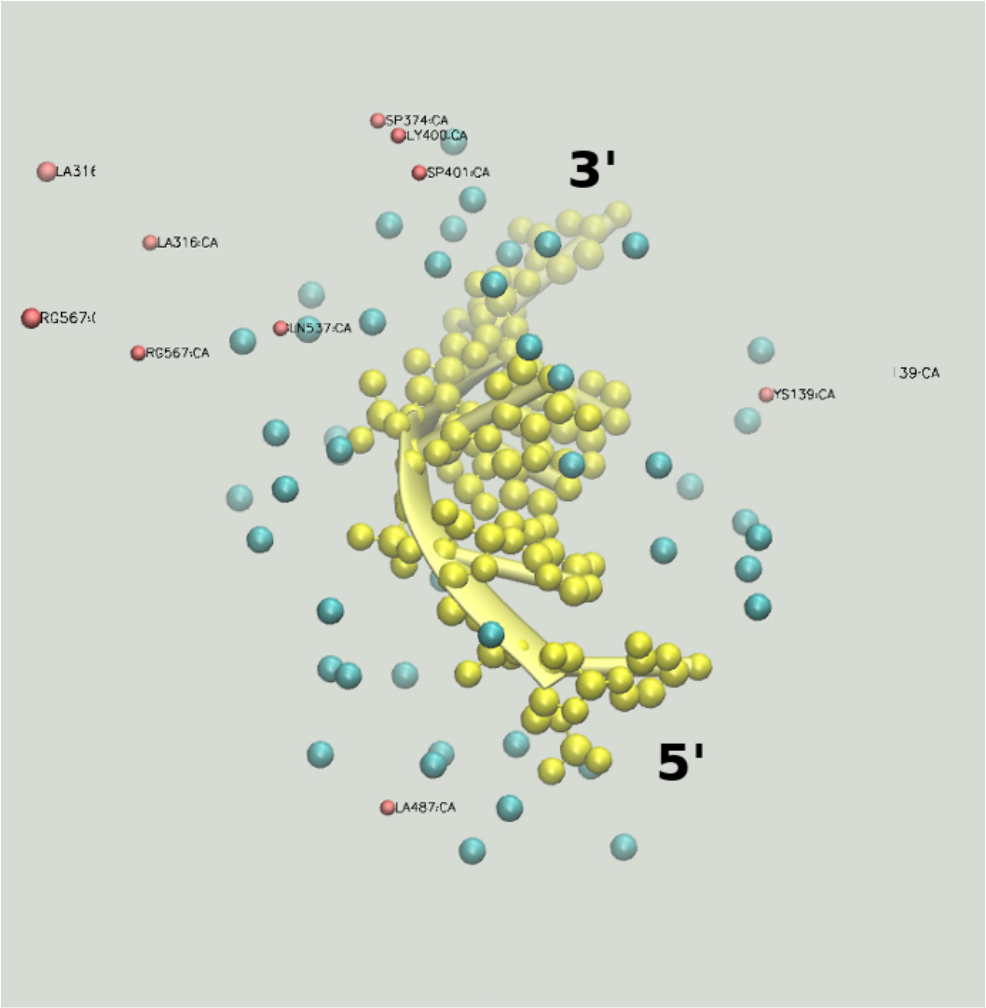
E-mail: piero.procacci@unifi.it

Binding pattern of the compounds 43, 52, 62, 79 in the NSP13-complexes

The channel residues

The channel residues reported in Figure 2 of the main paper are based on a purely geometrical criterion relying on the combination of the the recent 6ZSL X-ray structure of the SARS-CoV-2 helicase with the RNA taken from a earlier structure, 2XZL, referring to an yeast helicase with high sequence similarity.[1] In the figure S1 we have revised the proximal channel residues using a recent cryo-electron microscopy experiment of the transcription replication complex[2] (chain E and L of PDB code 7XCM, for NSP13 monomer and RNA chain respectively). In

the bottom panel, we list all the residues in 7XCM whose C_α is found within 7.5 Å of any atom of the RNA substrate. We note that some of the residues listed in Figure 2 of the main paper are missing from the updated list based on the 7CMX structure. These residues are underlined in red in the bottom panel and the location of their C_α with respect to the substrate can be seen in the upper panel (red spheres).



LYS139 PRO174 PRO175 ASN177 ARG178 ASN179 HIS230 THR231 CYS309 SER310 HIS311
ALA312 ALA313 ALA316 ILE334 ARG337 ALA338 VAL340 THR359 VAL360 ASN361 ALA362
LEU363 ASP374 GLU375 SER377 MET378 ALA379 THR380 GLY400 ASP401 GLN404 LEU405
ALA407 PRO408 THR410 LEU411 THR413 VAL484 SER485 SER486 TYR515 ASN516 THR532
VAL533 ASP534 SER535 GLN537 GLY538 HIS554 ARG567

Figure S1: Alpha carbon (spheres in cyan color) of the channel residues in NSP13. Channel residues are defined as those corresponding to the residues whose distance of the alpha-carbon from any atom of the RNA sequence is less than 7.5 Å as resulting from the chains L (RNA) and E (NSP13 monomer) of PDB structure of Replication Transcription Complex (PDB code 7XCM)

Best Vinardo docking poses

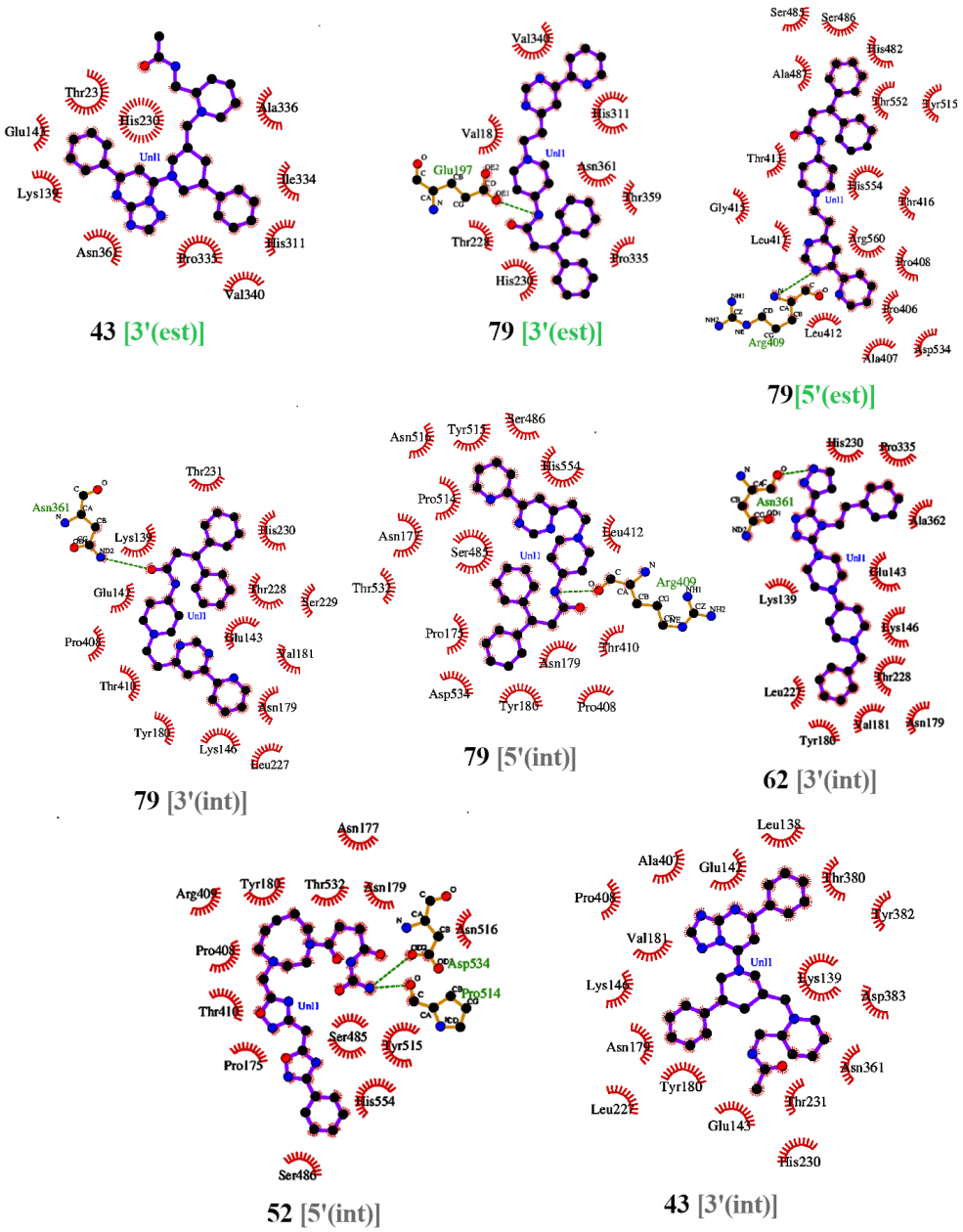


Figure S2: LIGPLOT diagrams obtained for the best Vinardo docking poses on the eight compounds/locations. Runtime parameters for non-bonded contacts are 2.9 and 4.5 Å for the minimum and maximum contact distance, respectively.

As stated in the paper, docking calculations for the eight compounds/locations reported in Table 3 of the main paper, delivered optimal poses with the center of mass of the ligand rarely coincident with the originally selected box centers at the 3' and 5' entrances (see Table

4 of the main paper). Besides, in many instances, docking calculations produced binding poses with disparate RMSD-based distance but with a similar score. In the Figure S2, we show the LIGPLOT[3] diagrams of the pose with the lowest binding free energy for the eight selected compounds/locations obtained using the Vinardo scoring function. We recall that the best Vinardo pose was chosen as the starting configuration in the MD simulation of the complex for the eight compounds/locations (see Section "MD preliminary assessment" in the main paper). The binding pattern found in docking reflects the relative freedom experienced by the ligand in optimally fitting in the channel. For example, when the box is centered on the same LIGANN location 5'(est), compound 79, and compound 43 exhibit best Vinardo poses with only four contact residues in common. The same outcome can be seen for other locations such as the 3'(int) or the 5'(int) centers.

Binding pattern in the equilibrium MD of the bound state

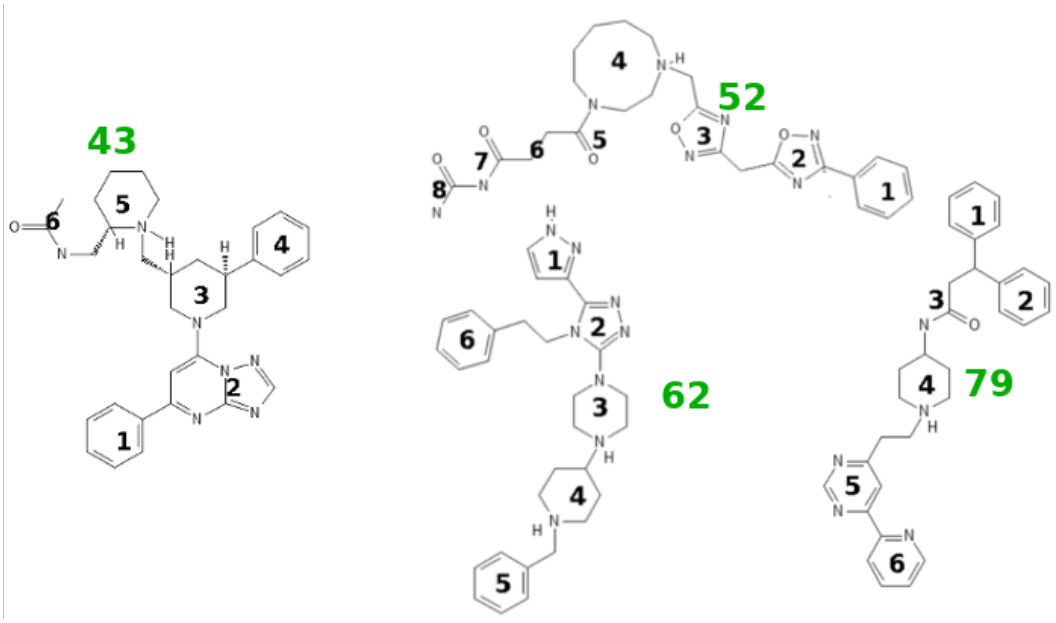


Figure S3: 2D structures of compounds 43, 52, 62, 79 with the labeling of the groups (x-axis in the contact heat-maps of Figures S4-S6).

In the MD simulation of the complexes, the binding modality is strongly affected by the ligand conformational states *and* the protein residue flexibility around the channel. Such dynamical

behavior of the binding pattern can be described by an assessment of the persistence of residue-ligand contacts in the MD sampling of the bound state. This can be achieved by sectioning the ligands into groups as shown in Figure S3. In a given equilibrium configuration of the complex, a ligand group is assumed to be in contact with a given protein residue if any group-residue atom-atom distance is found below a 4.5 Å threshold. For each of the eight compounds/locations, we sampled 400 uncorrelated configurations taken from the equilibrium MD replicates of the bound state spanning a total of 160 ns (see Section "vDSSB: canonical sampling" in the main paper). We then construct the contact heat maps of Figures S4-S6 where on the x-axis we report the group index as defined in Figure S3, and on the y-axis the corresponding contact residue. The heat-bar on the right of the maps refers to the normalized contact probability computed over the 400 configurations, defined as $p_{ij} = N_c(ij)/400$ where $N_c(ij)$ is the number of configurations where group i of the ligand is found to be in contact with residue j of the protein. The darker colors in the xy maps are indicative of stronger group-residue contacts during the simulations. The red dots in the Figures S4-S6 mark the Vinardo docking contacts. We can see, in general, that the binding pattern obtained from MD is far richer in contacts with variable strength compared to the static picture obtained from docking, a clear indication of the relative conformational freedom experienced by the ligand in the channel during the simulation. We also note significant differences in the docking/MD comparison for most of the compounds/locations, indicating that the original starting Vinardo pose is in general not conserved during the MD simulation.

The heat maps S4-S6 are organized as referring to the external site (green tone) and internal sites (black tone). The channel residues are highlighted in red on the y-axis. In general we expectedly observe a larger number of non-zero contact probabilities for the internal than for the external sites. The average number of non-zero contact probabilities for the internal and external locations is 46.7 and and 37.7, respectively. Compound 79 is a remarkable exception to this trend as, at the 5' end, this ligand has 52 and 38 non zero contacts in the external and internal locations, respectively. Since in the COM-COM

distribution of the 20 ns preliminary simulations we did not observe any inversion of the COM-COM residue-ligand average distance for 79 at the 5' end (see Figure 9 of the main paper), the richer contact map at the 5'(est) location must be caused by an enhanced mobility of this ligand due to a weaker stability which ultimately translates in a weaker activity as shown in Table 3 of the main paper. We stress that the predicted activity of the ligand using vDSSB at a specific location is not related to the corresponding overall contact index defined as $I_c = \sum_i^{\text{nres}} \sum_j^{\text{ngroups}} p_{ij}$. Dissociation free energies of Table 3 of the main paper and I_c contact index computed from the heat maps are indeed weakly *anti-correlated* with a Pearson coefficient of -0.37. The binding strength must hence depend on the specificity of the contacts at the four channel locations with a subtle balance of interaction with hydrophobic as well as polar or charged residues with no clear prevalence of one or the other. The heat maps reported in this study, in combination with the calculated dissociation free energies of Table 3 of the main paper, by thoroughly describing the dynamical binding pattern of the ligands in the channel, should be of some help in designing effective drugs for NSP13 inhibition. To this end, the contact map matrices ($N_c(ij) = 400 p_{ij}$), corresponding to the Figures S4-S6, are reported in full for convenience of use and analysis in the last section of the Supporting Information.

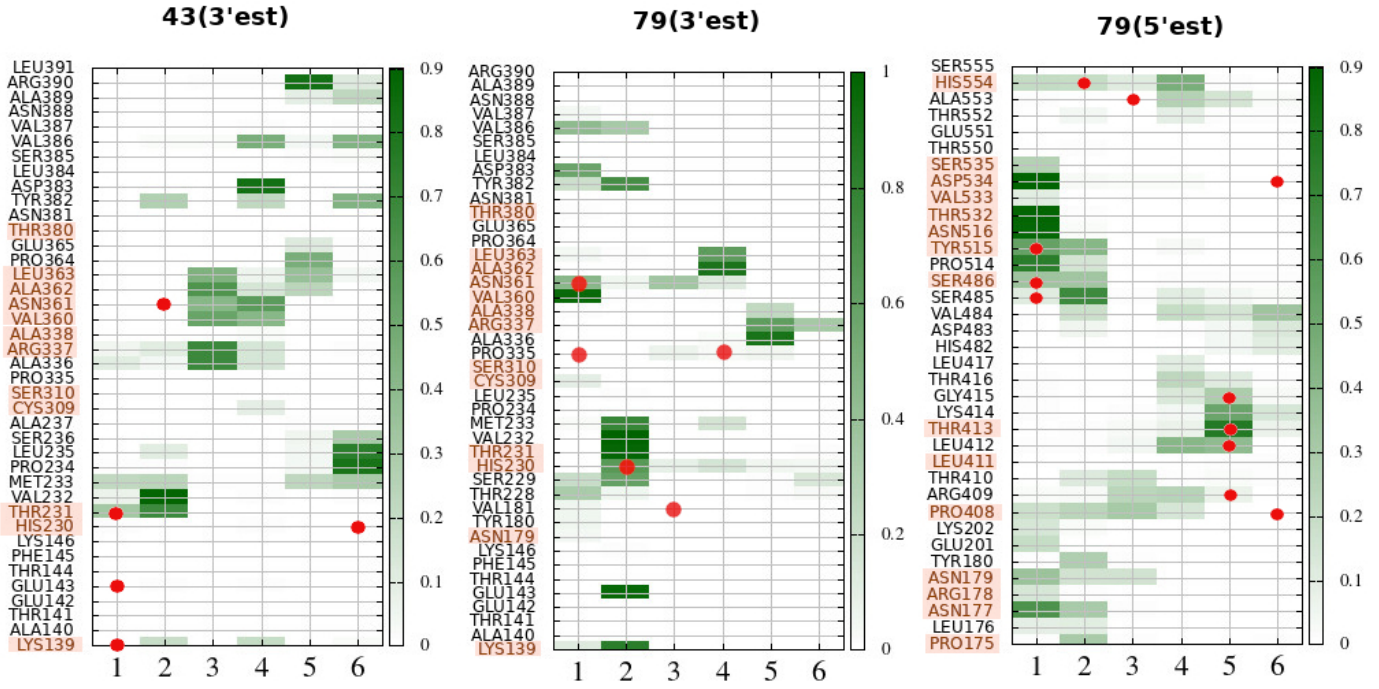


Figure S4: Contact heat map for 43 (3'est), 79 (3'est), and 79 (5'est)

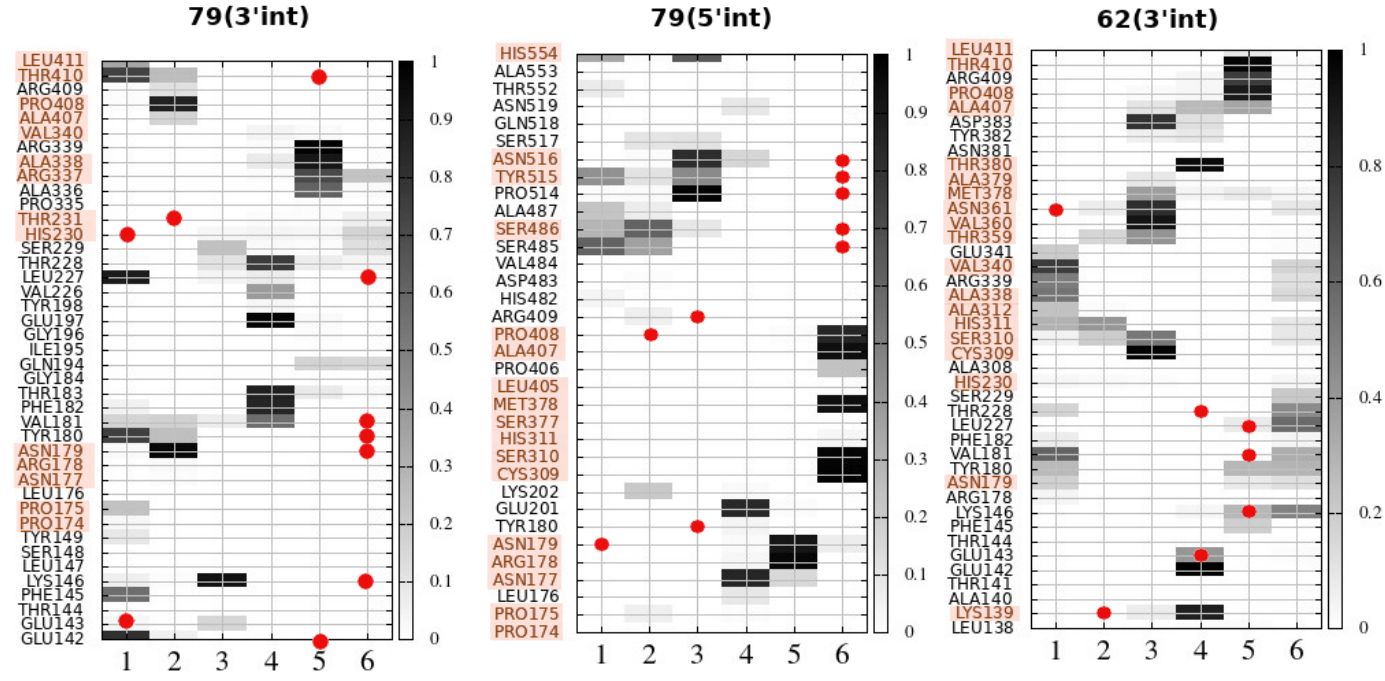


Figure S5: Contact heat map for 79 (3'int), 79 (5'int), and 62 (3'int)

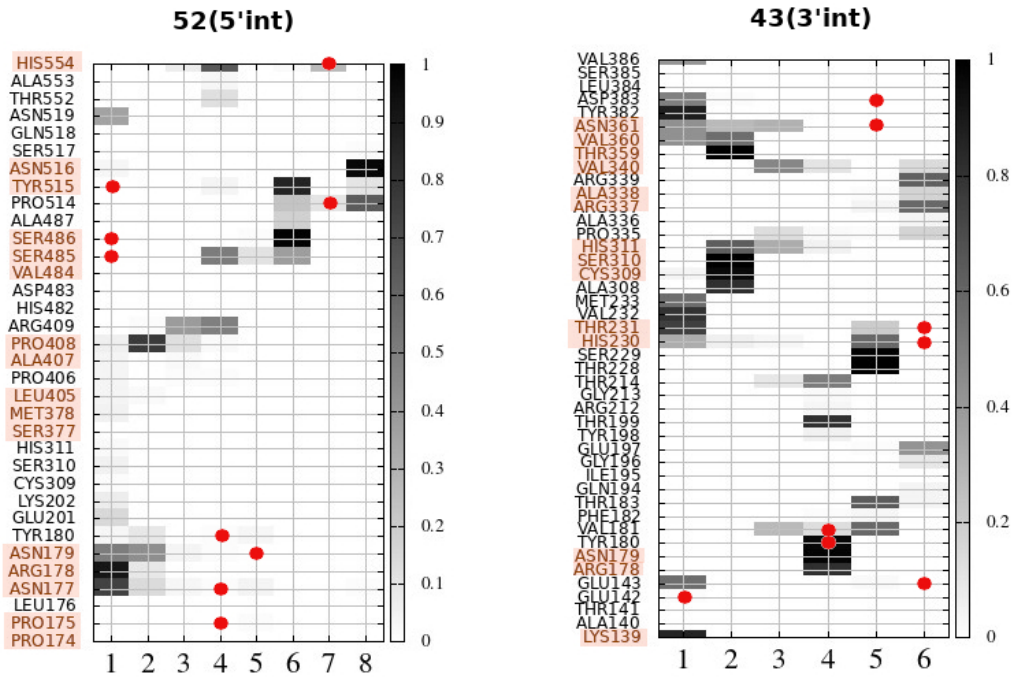


Figure S6: Contact heat map for 52 (5'int), 43 (3'int)

MD-computed contact maps

43(3'est) contact map

nres\ng	1	2	3	4	5	6
136	0	0	0	0	0	6
139	0	75	0	81	0	7
143	0	6	0	4	0	0
230	6	0	0	2	0	0
231	126	270	0	0	0	0
232	29	354	0	6	0	6
233	88	95	2	3	92	120
234	3	4	0	0	18	322
235	0	41	0	0	15	300
236	0	0	0	0	20	119
237	0	0	0	0	4	3
309	0	0	0	37	0	0
336	39	15	276	54	1	0
337	20	42	278	56	0	3
338	0	0	0	7	0	0
360	0	0	212	167	0	0
361	0	0	169	231	0	0
362	0	1	249	54	95	0
363	0	1	179	25	143	18
364	0	0	0	0	186	4
365	0	0	0	0	48	4
380	0	0	0	1	0	0
382	0	109	1	89	0	177
383	0	0	0	332	0	0
385	0	0	0	0	4	9
386	0	9	8	188	18	175
387	0	0	0	2	0	3
389	0	1	0	0	35	91
390	0	1	4	2	333	49

79(3'est) contact map

nres\ng	1	2	3	4	5	6
139	52	334	0	0	0	0
140	0	1	0	0	0	0
143	0	389	0	0	0	0
146	0	5	0	0	0	0
179	18	0	0	0	0	0
180	22	0	0	0	0	0
181	23	0	0	0	0	0
228	132	24	0	0	5	13
229	84	242	0	0	4	59
230	1	259	35	64	21	17
231	0	391	0	0	0	3
232	0	400	0	0	0	0
233	8	312	4	69	0	0
235	1	5	0	0	0	0
309	38	0	0	0	0	0
335	0	0	28	10	19	0
336	0	0	0	9	361	0
337	0	0	0	0	247	140
338	5	0	0	0	96	0
360	384	2	1	3	0	0
361	183	22	144	51	0	0
362	0	0	0	356	0	0
363	16	4	0	253	0	0
365	0	0	0	1	0	0
380	2	0	0	0	0	0
382	81	287	0	0	0	0
383	234	1	0	0	0	0
386	172	137	0	0	0	0
387	18	0	0	0	0	0
390	0	2	0	5	0	0

79(5'est) contact map

nres\ng	1	2	3	4	5	6
115	0	0	0	0	0	1
116	0	0	0	0	1	1
117	0	0	0	0	1	0
145	0	0	0	0	2	0
146	0	0	0	0	0	1
148	0	0	0	0	0	2
149	0	0	0	0	0	11
172	0	0	0	0	0	11
173	0	0	0	0	1	9
174	0	1	0	0	0	7
175	6	121	4	0	0	2
176	36	42	0	0	0	0
177	251	119	6	0	0	0
178	58	2	0	0	0	0
179	139	65	65	0	0	0
180	5	113	5	1	0	0
181	0	1	0	0	0	0
201	78	0	0	0	0	0
202	56	7	0	0	0	1
405	2	0	0	2	0	0
408	71	95	124	58	0	0
409	7	12	95	99	26	0
410	0	48	78	4	7	2
411	0	0	0	0	4	2
412	2	4	15	167	156	0
413	1	3	13	26	309	29
414	4	0	2	18	207	54
415	6	0	1	51	116	13
416	6	0	0	93	43	2
417	0	0	0	40	0	0
482	0	0	0	0	14	38
483	0	21	0	4	15	53
484	0	71	0	78	59	137
485	45	273	5	41	17	8
486	136	131	0	6	2	2
487	0	7	0	1	0	2
514	300	62	1	0	0	0
515	215	167	0	13	0	1
516	352	7	0	0	0	0
517	9	0	0	0	0	0
532	356	5	0	0	0	0

533	41	0	0	0	0	0
534	361	10	9	7	0	0
535	102	0	0	0	0	0
550	0	4	0	1	0	0
552	0	15	0	35	5	6
553	6	0	0	102	59	9
554	84	78	43	190	5	0
555	0	2	0	1	0	0
557	4	0	0	7	0	0
560	1	0	1	1	0	0

79(3'int) contact map

nres\ng	1	2	3	4	5	6
139	4	12	0	0	0	0
142	323	25	0	0	0	0
143	13	1	64	0	0	0
145	230	0	0	0	0	0
146	22	6	372	0	0	0
147	0	0	0	0	0	1
149	36	0	0	0	0	0
174	11	0	0	0	0	0
175	98	0	0	0	0	0
176	1	0	0	0	0	0
177	0	2	0	0	0	0
178	0	7	0	0	0	0
179	8	389	0	0	0	0
180	299	101	0	0	0	0
181	69	73	28	230	0	0
182	21	0	0	340	0	0
183	0	0	0	347	32	7
185	0	0	0	0	1	8
194	0	0	0	0	67	61
196	0	0	0	1	4	0
197	0	0	0	393	7	0
198	0	0	0	13	0	0
199	0	8	0	0	0	0
200	1	0	0	0	0	0
214	0	0	0	4	1	0
226	0	0	0	157	0	0
227	355	0	16	28	0	0
228	6	0	45	307	28	14
229	1	0	97	9	6	46
230	0	0	13	15	13	67
231	0	0	0	3	3	22
232	0	0	0	0	0	3
309	0	11	0	0	0	0
310	0	7	0	0	0	0
311	0	7	0	2	0	0
335	0	0	0	1	4	0
336	0	0	0	0	245	8
337	0	0	0	6	286	90
338	0	0	0	33	365	0
339	0	0	0	2	398	0
340	0	1	0	16	8	0

361	0	8	0	20	0	0
378	0	10	0	0	0	0
407	0	68	0	0	0	0
408	2	348	0	0	0	0
409	5	50	0	0	0	0
410	295	105	0	0	0	0
411	140	0	0	0	0	0

79(5'int) contact map

nres\ng	1	2	3	4	5	6
175	0	23	0	3	0	0
176	0	0	0	34	0	0
177	0	0	0	341	59	0
178	0	0	0	12	388	0
179	0	0	0	7	363	29
180	0	1	0	18	0	0
201	0	0	0	338	2	0
202	0	83	0	9	0	0
309	0	0	0	0	0	396
310	0	0	0	0	1	392
311	0	0	0	0	0	11
312	0	0	0	0	1	0
378	0	0	0	0	0	374
405	0	0	0	0	0	2
406	0	0	0	0	0	95
407	0	0	0	0	0	376
408	0	0	0	0	6	341
409	1	23	0	1	0	0
410	0	1	0	0	0	0
482	18	0	0	0	0	0
483	0	3	0	0	0	0
484	1	0	0	0	0	0
485	245	149	1	0	0	0
486	118	246	34	0	0	0
487	94	26	0	0	0	0
514	2	7	391	0	0	0
515	174	43	183	0	0	0
516	0	6	330	64	0	0
517	0	41	44	0	0	0
519	0	0	0	37	0	0
532	0	0	0	350	50	0
534	0	1	0	24	352	0
535	0	0	0	0	400	0
537	0	0	0	0	16	327
552	33	0	0	0	0	0
554	135	6	259	0	0	0
560	0	1	0	0	0	0

62(3'int) contact map

nres\ng	1	2	3	4	5	6
138	0	0	0	3	4	0
139	0	1	27	352	0	0
141	0	0	0	0	1	0
142	0	0	0	399	1	0
143	0	0	1	166	0	5
145	0	0	0	0	75	0
146	0	0	0	11	111	198
149	0	0	0	0	1	0
175	0	0	0	0	11	2
177	0	0	0	0	11	0
178	14	0	0	0	0	1
179	72	0	0	0	32	46
180	106	0	0	0	111	115
181	244	0	0	0	1	129
182	29	0	0	0	0	14
183	3	0	0	0	0	3
197	8	0	0	0	0	4
199	0	0	0	0	0	8
214	0	0	0	0	0	1
227	0	0	0	0	31	233
228	69	0	0	0	0	178
229	0	0	0	0	0	83
230	11	8	2	0	0	12
308	0	0	1	0	0	0
309	0	1	392	0	0	6
310	21	81	212	0	1	36
311	119	164	1	0	0	37
312	98	2	0	0	0	1
335	1	0	0	0	0	14
337	3	0	0	0	0	0
338	216	1	0	0	0	59
339	203	0	0	0	0	47
340	305	0	0	0	0	69
341	80	0	0	0	0	2
342	1	0	0	0	0	0
359	0	65	167	0	0	3
360	0	0	367	0	0	0
361	0	30	328	0	0	35
375	0	0	0	0	0	2
378	0	0	152	18	31	11
379	0	0	37	4	1	1

380	0	0	0	386	1	0
382	0	0	23	35	0	0
383	0	0	321	48	0	0
407	0	0	41	107	142	0
408	0	0	9	12	359	0
409	0	0	0	10	305	0
410	0	0	0	0	393	2
411	0	0	0	0	37	0
516	0	0	0	0	0	1
534	0	0	0	0	5	4
535	0	0	0	0	0	7
537	0	0	0	0	1	12

52(5'int) contact map

nres\ng	1	2	3	4	5	6	7	8
175	0	0	0	0	5	0	0	0
177	297	58	13	3	15	0	0	2
178	363	36	0	0	0	0	0	0
179	203	177	16	4	0	0	0	0
180	13	35	1	3	10	0	0	0
200	2	0	0	0	0	0	0	0
201	62	0	0	0	0	0	0	0
202	30	0	0	0	0	1	0	0
309	6	0	0	0	0	0	0	0
310	19	0	0	0	0	0	0	0
311	8	0	0	0	0	0	0	0
312	2	0	0	0	0	0	0	0
378	21	0	0	0	0	0	0	0
405	14	7	0	0	0	0	0	0
406	18	0	4	3	0	0	0	0
407	17	0	7	0	0	0	0	0
408	25	307	49	9	0	0	0	0
409	0	2	156	199	0	0	0	0
410	0	15	42	282	0	0	0	0
411	0	0	0	61	0	0	0	0
412	0	0	2	281	0	0	0	0
413	0	0	0	279	0	0	0	0
485	0	0	0	202	45	151	0	0
486	0	0	0	0	3	397	0	0
487	0	0	0	0	0	72	0	0
513	0	0	0	0	0	0	0	3
514	0	0	0	0	0	91	41	257
515	0	0	0	15	0	343	0	42
516	10	0	0	0	0	1	0	389
517	0	0	0	0	0	1	0	4
519	143	0	0	0	0	0	0	0
530	20	0	0	0	0	0	0	0
531	106	0	0	0	0	0	0	0
532	9	0	0	0	0	0	0	391
533	0	0	0	0	0	0	0	166
534	0	8	0	0	0	0	94	0
535	384	14	0	0	0	0	0	0
552	0	0	0	48	0	0	0	0
553	0	0	0	10	0	0	0	0
554	0	0	27	260	0	7	106	0
560	0	0	12	2	0	0	0	0

43(3'int) contact map

nres\ng	1	2	3	4	5	6
139	346	0	0	0	0	0
140	1	0	0	0	0	0
142	2	0	0	0	0	0
143	229	0	0	1	7	0
178	0	0	0	323	0	0
179	0	0	0	391	0	0
180	0	0	0	390	0	0
181	0	0	109	53	238	0
182	0	0	0	4	10	1
183	0	0	0	0	256	15
184	0	0	0	0	0	2
194	0	0	0	0	0	18
196	0	0	0	0	0	34
197	0	0	0	0	6	170
198	0	0	0	19	0	0
199	0	0	0	325	0	0
212	0	0	0	11	0	0
213	0	0	0	8	0	0
214	0	0	34	204	0	0
215	0	0	0	0	0	6
228	0	0	0	0	400	0
229	1	0	0	0	398	0
230	127	24	12	0	237	0
231	300	0	0	0	78	0
232	318	0	0	0	0	0
233	227	0	0	0	0	0
308	0	326	0	0	0	0
309	17	383	0	0	0	0
310	0	399	0	0	0	0
311	0	252	127	21	0	0
312	0	1	0	0	0	0
335	0	0	51	7	2	68
336	0	0	0	0	0	18
337	0	0	0	0	17	236
338	0	0	0	0	3	69
339	0	0	0	0	0	247
340	0	0	187	38	0	58
341	0	0	0	6	0	2
342	0	0	0	3	0	0
359	0	400	0	0	0	0
360	173	227	0	0	0	0

361	171	109	120	0	0	0
362	1	0	0	0	0	0
378	12	0	0	2	0	0
379	7	0	0	0	0	0
380	8	0	0	0	0	0
382	350	0	0	0	0	0
383	201	4	0	0	0	0
386	163	0	0	0	0	0

References

(1) Yazdani, S.; De Maio, N.; Ding, Y.; Shahani, V.; Goldman, N.; Schapira, M. Genetic Variability of the SARS-CoV-2 Pocketome. *J. Proteome Res.* **2021**, *20*, 4212–4215.

(2) Yan, L.; Zhang, Y.; Ge, J.; Zheng, L.; Gao, Y.; Wang, T.; Jia, Z.; Wang, H.; Huang, Y.; Li, M.; Wang, Q.; Rao, Z.; Lou, Z. Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nature Communications* **2020**, *11*, 5874.

(3) Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng. Des. Sel.* **1995**, *8*, 127–134.