AIMS Biophysics, 8(3): 248–263.
DOI: 10.3934/biophy.2021019
Received: 22 April 2021
Accepted: 20 June 2021
Published: 24 June 2021

http://www.aimspress.com/journal/biophysics

Research article

Screening coronavirus and human proteins for sialic acid binding sites using a docking approach

Chia-Wen Wang[1], Oscar K. Lee[2,3,*] and Wolfgang B. Fischer[1,*]

[1] Institute of Biophotonics, School of Biomedical Science and Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan
[2] Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan
[3] Department of Orthopedics, China Medical University Hospital, Taichung, Taiwan

* **Correspondence:** Email: oscarlee9203@gmail.com; wfischer@ym.edu.tw; Tel: +886228267394; Fax: +886228235460.

Appendix

Suppl. Table 1. Averaged estimated binding energies for the sialic acids and sugar molecules. The values are an average over the individual pockets, protein structures derived by the minimization protocols, respective number of three sialic acids and five sugar molecules. Vales are also averaged for over the S-proteins Sp, $Sp2^{h}$ and $Sp2^{or}$, as well as the two versions of the furin protein, furin-a and -h. The results are shown for the oriented sugar molecules (oriented) and the best ranked (rank-1) for $p < 0.05$ (*), for $p < 0.01$ (**) and for $p < 0.001$ (***). $p^{sia\text{-}su}$ denotes the difference of the values between the averaged values for the sialic acid and the sugar molecules.

oriented

	p	**Sp-l**	**Sp, $Sp2^{h}$, $Sp2^{or}$**	**ACE2**	**furin-a and -h**
LeadIT					
Sialic acids		-20.1 ± 1.6	-6.9 ± 1.3	-8.1 ± 0.8	-11.0 ± 0.5
p	**Sp-l**		*	*	-
	Sp, $Sp2^{h}$, $Sp2^{or}$	*		-	**
	ACE2	*	-		**
	furin-a and -h	-	**	**	
Sugar molecules		-16.7 ± 0.7	-11.7 ± 1.5	-10.6 ± 0.5	-13.7 ± 0.3
p	Sp-l		***	***	***
	Sp, $Sp2^{h}$, $Sp2^{or}$	***		-	*
	ACE2	***	-		**
	furin-a and -h	***	*	**	
	Sia-Sug	-	***	*	**

HYDE					
Sialic acids		-16.4 ± 4.0	-4.4 ± 1.3	-4.8 ± 0.7	-3.0 ± 0.7
p	Sp-l		*	*	*
	Sp, $Sp2^{h}$, $Sp2^{or}$	*		-	-
	ACE2	*	-		-
	furin-a and -h	*	-	-	
Sugar molecules		-15.3 ± 2.2	-7.8 ± 0.8	-6.8 ± 0.2	-5.8 ± 0.8
p	Sp-l		-	-	-
	Sp, $Sp2^{h}$, $Sp2^{or}$	-		-	*
	ACE2	-	-		-
	furin-a and -h	-	*	-	
	Sia-Su	-	**	-	*

Suppl. Tab. 1 (cont.)

rank-1

	p	**Sp-l**	**Sp, Sp2^{h}, Sp2^{or}**	**ACE2**	**furin-a and -h**
LeadIT					
Sialic acids		-20.6 ± 1.4	-8.8 ± 1.4	-9.1 ± 0.0	-13.3 ± 0.7
p	Sp-l		*	*	-
	Sp, Sp2^{h}, Sp2^{or}	*		-	**
	ACE2	*			**
	furin-a and -h				
Sugar molecules		-17.3 ± 1.3	-12.3 ± 1.5	-10.9 ± 0.5	-14.4 ± 0.3
p	Sp-l		***	**	*
	Sp, Sp2^{h}, Sp2^{or}	***		-	*
	ACE2	**			***
	furin-a and -h				
	Sia-Su	-	**	**	-

HYDE					
Sialic acids		-19.0 ± 2.3	-4.6 ± 0.9	-5.3 ± 0.7	-1.7 ± 0.7
p	Sp-l		*	*	*
	Sp, Sp2^{h}, Sp2^{or}	*		-	**
	ACE2	*			*
	furin-a and -h				
Sugar molecules		-17.0 ± 2.0	-7.5 ± 0.5	-6.6 ± 0.1	-5.6 ± 0.9
p	Sp-l		***	***	***
	Sp, Sp2^{h}, Sp2^{or}	***		-	*
	ACE2	***			-
	furin-a and -h				
	Sia-Su	-	**	-	***

Suppl. Table 2. Biggest difference, Δ, of the estimated binding energies between sialic acids and sugar molecules. Data are shown for the oriented molecules and the rank-1 molecules from LeadIT and HYDE. Only values for both, sialic acids and sugar molecules smaller than zero are chosen. The difference is evaluated by calculating the *p*-value. If two Δ-values are the same, the one with the smaller *p*-value is listed. The *p*-values are marked for $p < 0.05$ (*), $p < 0.01$ (**) and $p < 0.001$ (***). The difference data are averaged (Ave.) over the S-proteins and the human proteins.

Oriented						rank-1				
	Sia	Gly	Δ	p	Ave.	Sia	Gly	Δ	p	Ave.
Sp-l	-22.8 ± 1.5	-16.8 ± 0.8	6.1 ± 0.8	***		-23.2 ± 1.0	-17.2 ± 1.9	6.0 ± 1.2	**	
Sp	-12.4 ± 3.3	-10.8 ± 1.7	1.7 ± 1.7	-	4.1 ± 2.0	-17.1 ± 4.0	-14.6 ± 1.0	2.5 ± 1.9	-	4.2 ± 1.5
$Sp2^{h}$	-17.2 ± 5.2	-13.8 ± 3.5	3.4 ± 3.1	-		-18.2 ± 4.4	-13.8 ± 3.5	4.4 ± 2.8	-	
$Sp2^{or}$	-15.5 ± 2.6	-10.3 ± 9.5	5.3 ± 5.6	-		-16.5 ± 8.2	-12.6 ± 2.5	3.9 ± 3.9	-	
ACE2	-21.5 ± 3.2	-16.3 ± 2.6	5.2 ± 2.1	*		-22.3 ± 1.4	-16.1 ± 2.4	6.2 ± 1.5	**	
furin-h	-19.8 ± 1.1	-14.0 ± 2.6	5.8 ± 1.6	*	6.0 ± 0.9	-23.3 ± 4.0	-17.9 ± 1.9	5.4 ± 2.1	*	6.0 ± 0.5
furin-a	-21.4 ± 1.8	-14.5 ± 2.3	6.9 ± 1.5	**		-19.7 ± 3.2	13.2 ± 1.0	6.5 ± 1.5	**	

HYDE										
Sp-l	-20.0 ± 6.1	-10.2 ± 13.8	9.8 ± 8.4	-		-22.0 ± 3.0	-17.8 ± 2.8	4.2 ± 2.1	-	
Sp	-16.0 ± 4.6	-6.4 ± 10.1	9.6 ± 6.2	-	9.7 ± 4.0	-13.3 ± 4.9	-4.0 ± 13.8	8.9 ± 8.3	-	8.3 ± 5.4
$Sp2^{h}$	-16.3 ± 8.5	-11.6 ± 4.2	4.7 ± 4.5	-		-16.0 ± 8.5	-11.6 ± 4.2	4.4 ± 4.5	-	
$Sp2^{or}$	-19.7 ± 1.5	-5.2 ± 6.9	14.5 ± 4.1	*		-16.7 ± 5.1	-1.0 ± 7.4	15.7 ± 4.8	*	
ACE2	-16.3 ± 5.9	-7.4 ± 2.6	8.9 ± 3.0	*		-19.0 ± 7.8	6.2 ± 8.8	12.8 ± 6.2		
furin-h	-21.7 ± 6.0	-8.2 ± 8.4	13.5 ± 5.5	-	12.0 ± 2.6	-19.0 ± 10.2	-8.2 ± 8.4	10.8 ± 6.6	-	12.5 ± 1.6
furin-a	-14.3 ± 10.7	-0.8 ± 8.2	13.5 ± 6.7	-		-14.7 ± 10.6	-0.8 ± 8.3	13.9 ± 6.7	-	

Suppl. Table 3. Rank-1 sugar molecules with the lowest estimated binding energies identified for LeadIT and best HYDE values. The values are obtained for S-proteins and the human proteins. '+' marks those sialic acid poses which are the same for both oriented and rank-1. Marked with '++' are those sialic acid molecules which are the best docked ligands (best value) and used for decoy finding. Red: lowest value observed over all protocols (mp-0, mp-1, mp-2) and the respective pockets. $p^{s\text{-}h}$: *p*-value from the difference between the averaged (avg.) estimated binding energies and their standard deviation of S (s) and human (h) proteins (ACE2 and furin).

	LeadIT	Sialic acids		Sugar molecules		*p*
			ΔG (kJ/mol)		ΔG (kJ/mol)	
S-proteins	Sp-l	**+/++Neu5Gc**	-24.5	GalNAc	-21.3	
	Sp	Neu5Gc	-24.4	Gal	-26.9	
	$Sp2^{h}$	Neu5Gc	-24.0	GlcNAc	-26.8	
	$Sp2^{or}$	Neu5Ac	-23.3	GlcNAc	-23.3	
	avg.		-24.1 ± 0.5		-24.6 ± 2.8	-
Human proteins	ACE2	**++Neu5Gc**	-30.5	GlcNAc	-23.1	
	furin-h	Neu5Gc	-26.5	GalNAc	-24.5	
	furin-a	**++Neu5Gc**	-27.6	GalNAc	-21.8	
	avg. ΔG		-28.2 ± 2.1		-23.1 ± 1.4	-
	$p^{s\text{-}h}$		*		-	

	HYDE	Sialic acids		Sugar molecules		*p*
			ΔG (kJ/mol)		ΔG (kJ/mol)	
S-proteins	Sp-l	Neu5Ac	-25.0	Mannose	-25.0	
	Sp	Neu5Gc	-28.0	GlcNAc	-30.0	
				GalNAc	-30.0	
	$Sp2^{h}$	Neu5Ac	-25.0	GlcNAc	-33.0	
	$Sp2^{or}$	**+/++Neu5Gc**	-31.0	GlcNAc	-27.0	
				GlcNAc	-27.0	
				GalNAc	-27.0	
	avg. ΔG		-27.3 ± 2.9		-28.4 ± 2.7	-
Human proteins	ACE2	**++Neu5Gc**	-31.0	GalNAc	-26.0	
	furin-h	**+/++9-*O*-AC-Me-Sia**	-28.0	GlcNAc	-25.0	
		Neu5Gc	-28.0	GlcNAc	-25.0	
	furin-a	9-*O*-AC-Me-Sia	-26.0	GalNAc	-30.0	
	avg. ΔG		-28.3 ± 2.1		-26.5 ± 2.4	-
	$p^{s\text{-}h}$		-		-	

Suppl. Table 4. Number of rotatable bonds in the sialic acids and sugar molecules. The values are derived from Molinspiration property engine v2018.10 (molinspiration.com). Rotatable bond is defined as any single non-ring bond, bounded to nonterminal heavy (i.e., non-hydrogen) atom.

Sugar molecules	Rotatable bonds
9-O-Ac-Me-Sia	8
Neu5Gc*	6
Neu5Ac*	5
Fucose	0
Galactose	1
Mannose	1
GalNAc	2
GlcNAc	2

Suppl. Table 5. Values for the area-under-curve (AUC) for 300 decoys from the ZINK data base. The decoys are docked into the best binding sites for (A) best oriented sites occupied by Neu5Gc at Sp-l, ACE2, and furin-a from LeadIT. (B) Best HYDE sites occupied by Neu5Gc at for $Sp2^{or}$ and Neu5Ac for ACE2 in oriented site and Neu5Gc in rank-1 site, and 9-*O*-Ac-Me-Sia for furin-h.

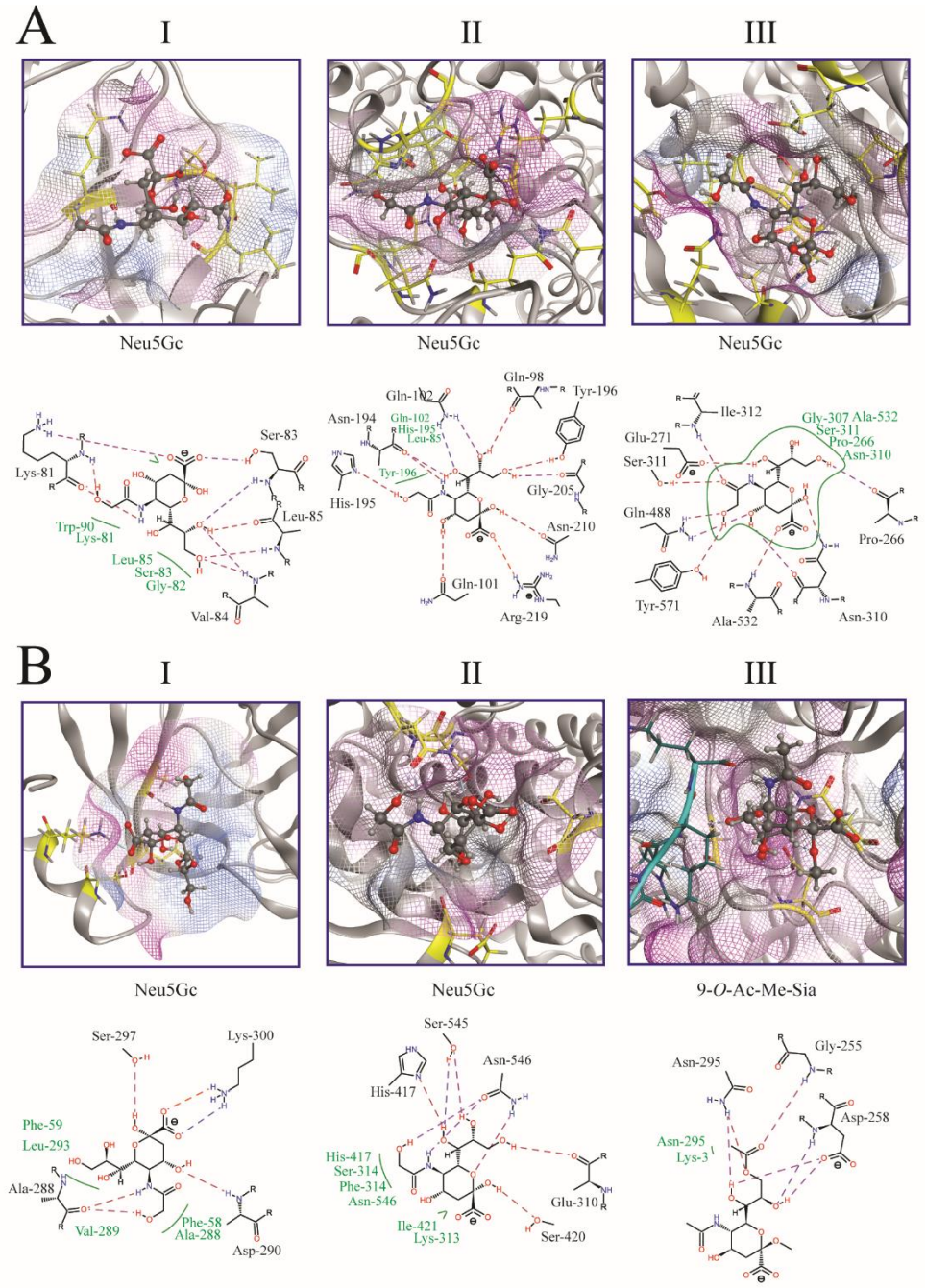
LeadIT	Sp-l	ACE2	furin-a
oriented	0.55	0.29	0.52
rank-1	0.55	0.92	0.36

HYDE	Sp-l	ACE2	furin-h
oriented	0.99	0.99	0.99
rank-1	0.99	0.99	0.99

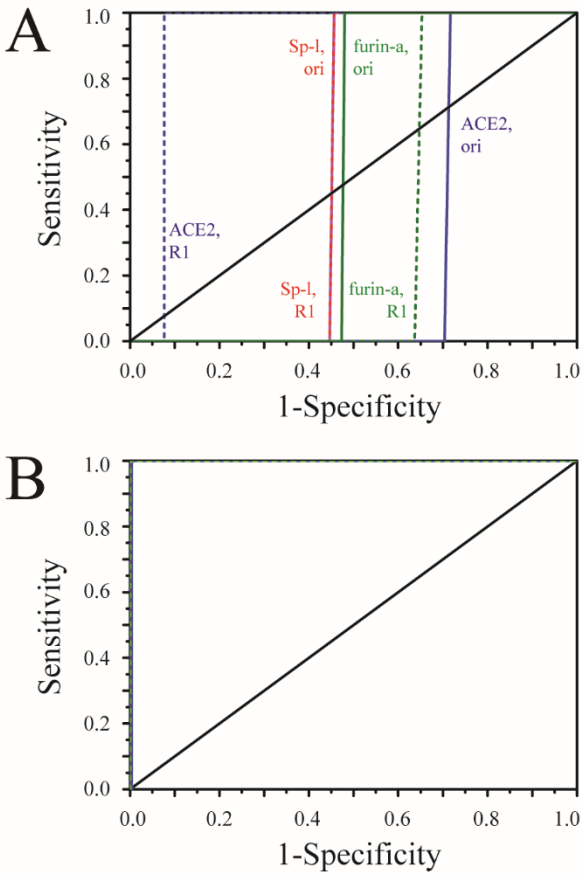
Suppl. Table 6. Values for the area-under-curve (AUC) for 300 decoys from the ZINK data base. The decoys are docked into the best binding sites for (A) best oriented sites occupied by Neu5Gc at Sp-l, ACE2, and furin-a for the respective HYDE values from LeadIT. (B) Respective LeadIT values from the best HYDE sites occupied by Neu5Gc at for $Sp2^{or}$ and Neu5Ac for ACE2 in oriented site and Neu5Gc in rank-1 site, and 9-*O*-Ac-Me-Sia for furin-h.

HYDE	$Sp2^{or}$	ACE2	furin-a
oriented	0.84	0.27	0.73
rank-1	0.84	0.85	0.32

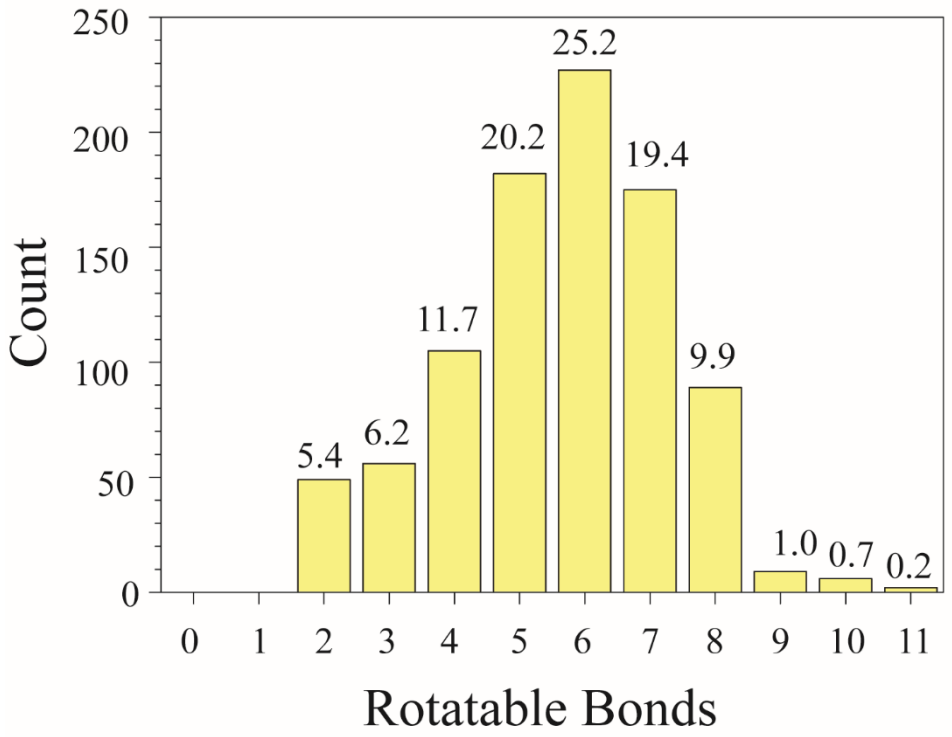
LeadIT	$Sp2^{or}$	ACE2	furin-h
oriented	0.02	0.04	0.07
rank-1	0.02	0.06	0.07



Suppl. Figure 1. Absolute best binding poses of the ranked (rank-1) sialic acids and sugar molecules. (A) from left to right the binding sites based on LeadIT of (I) Neu5Gc to protein Sp-l, (II) ACE2 and (III) furin-a, scoring. (B) in the similar sequence using HYDE from left to write the binding site of Neu5Gc to (I) Sp2^{or} and (II) ACE2 as well as (III) 9-*O*-Ac-Me-Sia to furin-h based on their scoring. The structural features in the sugar in the binding pocket are shown in the upper row and the respective 2D map of the pocket in the lower row. See Suppl. Table 4 for the respective estimated binding energies.



Suppl. Figure 2. Receiver operating characteristic curve (ROC) of the decoys docking to the best binding sites. The data are calculated as sensitivity over 1-Selectivity. (A) Sites identified by LeadIT (L) for the best sialic acid binding to a particular pocket (see Figure 3). The lines represent the oriented sialic acid, the dashed lines the rank-1 sialic acid using the color code as follows, red for the S-proteins, blue for ACE2 and green for furin-proteins. The sites with the best oriented sialic acid are occupied by Neu5Gc at Sp-l, ACE2, and furin-a as well as for the best rank-1, molecules occupied by Neu5Gc at Sp-l, ACE2, and furin-a. (B) the best sites identified by the best HYDE values using the same line and color code. All lines overlap at the left side of the graph. The sites with the best oriented sugar molecules are occupied by Neu5Gc for $Sp2^{or}$, Neu5Ac for ACE2 and 9-*O*-Ac-Me-Sia for furin-h as well as for the best rank-1 sites occupied by Neu5Gc for $Sp2^{or}$ and ACE2 and 9-*O*-Ac-Me-Sia for furin-h.



Suppl. Figure 3. Distribution over the number of rotatable bonds identified in the individual decoys. Numbers on the columns are the percentage contribution on the total number of decoys (300 decoys for each sialic acid, total 900 decoys).