AIMS Biophysics

AIMS Biophysics, 8(2): 165–181.
DOI: 10.3934/biophy.2021013
Received: 25 February 2021
Accepted: 13 April 2021
Published: 20 April 2021

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

Research article

Sequence–function correlation of the transmembrane domains in NS4B of HCV using a computational approach

Ta-Chou Huang and Wolfgang B. Fischer*

Institute of Biophotonics, School of Biomedical Science and Engineering, National Yang Ming Chiao Tung University, Taipei, Taiwan

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

Appendix

Suppl. Table 1. Best scored target proteins derived from multiple sequence alignment with NS4B using BLAST. The length of the NS4B sequence used varies from (i) a stretch which includes the r-TMDs (L70 to I190), to (ii) the individual sequences of the four r-TMDs, as well as (iii) specifically r-TMD2 and 4 only. Information is provided about their function, the percentage identity (I) and similarity (S), the availability of a crystal structure or the uniport accession number (PDB ID / Uniprot), the oligomeric state (Oligo. state) and whether the alignment is with the TMDs of the target protein (TMDs at target protein). cov (covered): the scoring is reported as the percentage number of identical residues identified in the alignment derived from the percentage of the TMD stretch of NS4B used in BLAST.

	Segment	Target protein	Function	I (cov) [%]	PDB ID / Uniprot[a]	Oligo. state	TMDs at target protein
i	r-TMDs + loops	Succinate-CoA ligase subunit β	Enzyme	28 (57.9)	2FP4 /P53590		
ii	r-TMD1	Ccc1-like protein	Transport protein	64 (100)	-[b] /A0A2P5CTQ4	-	
	r-TMD2	Signal transduction histidine kinase	Enzyme	80 (100)	-/A0A1N6G2L8		
	r-TMD3	Chromate transport protein	Transport protein	68 (100)	-[c]/Q0BU73	2	
	r-TMD4	inositol transport system permease protein	Transport protein	62 (100)	-[d] /A0A1G4QXH9	1	
iii	r-TMD2	CIC11C00000002157	Transport protein	56 (90)	-[e] /A0A1L0BG09	1	
	r-TMD4	Ca^{2+}/Na^{+}:H^{+} antiporter	Transport protein	52 (94)	-[e] /A0A0T9KYC6	2	

[a]: The crystal structure information is not found by BLAST result and comes from the same/similar functional protein but different species/organisms. [b]: Cation-Chloride Cotransporter (CCC1) protein is suggested to be a novel type of Ca^{2+} and/or Mn^{2+} transporter. Taking the crystalized Mn^{2+} transporter (PDB ID: 5M87) as a mapping bundle (DOI: 10.1038/ncomms14033). [c]: TMD3 target protein lacks the crystal structure information but has sequence information in Uniprot database with topology information. [d]: TMD4 target protein lacks the crystal structure information but belongs to the Major Facilitator Superfamily (MFS). Taking proton-coupled sugar transporter XylE (PDB: 4GBY), another MFS member, as a mapping bundle (DOI: 10.1038/ncomms5521). [e]: Although TMD2 and TMD4 have different target proteins from BLAST, the target proteins covered the same Na^{+}/Ca^{2+} exchanger (NCX) domain. Using crystalized NCX protein (PDB ID: 3V5S) as a mapping bundle (DOI: 10.1126/science.1215759).

Suppl. Table 2. Sequence alignment of r-TMDs. Calculation of r-TMD identity (I) and similarity (S) using Sequence Manipulation Suite of Clustal Omega. Each of the r-TMDs of NS4B are compared with either pore lining or Ca-binding domains of selected target membrane proteins: Orai, $Ca_v1.1$, $Ca_v1.2$, RyR1, IP_3R, MCU, SERCA, PLN, STIM1. Standard deviation is given if the target protein consists of multiple domains or more than one TMD which is identified as pore lining (PL) or identified as Ca-binding site (CaBS). Notation: e.g. 1(4) x 1 (TMD1) = the protein is a homo tetramer and only one of the subunits is consequently used (1(4)), and has only on TMD (x 1), and the TMD is TMD1 (1). For Ca_v channels the PL domains TMD5 and 6 (5/6) are used, for RyR1 also TMD5 and 6 as well as the pore region (marked as (2+1) because of two TMDs plus one pore helix). SERCA consists of three domains, from which one of them having 10 TMDs with 4 of them involved in Ca-binding (1x4). For PLN the name of the transmembrane domain is given as domain II (DOI: 10.1073/pnas.1016535108). Dark shaded boxes indicate TMDs of NS4B which show highest values of I/S with PL domains of target protein, light shaded boxes indicate the corresponding data for CaBS of target protein. mv = maximum value. For SERCA two TMDs of NS4B are shaded for I and S due to their almost similar values.

	TMDs of target protein	I/S	r-TMD1	r-TMD2	r-TMD3	r-TMD4
Orai	1(4) x 1 (1)	I	8.9	9.8	5.3	13.5
		S	8.9	24.4	23.7	27.0
$Ca_v1.1$	4 x 2 (5/6)	I	14.3 ± 6.6	12.5 ± 4.6	10.7 ± 5.0	13.4 ± 6.2
		S	25.3 ± 8.7	31.1 ± 9.5	24.3 ± 7.4	29.1 ± 11.0
$Ca_v1.2$	4 x 2 (5/6)	I	10.9 ± 4.4	9.8 ± 3.8	10.7 ± 5.8	12.8 ± 5.7
		S	22.1 ± 5.6	27.7 ± 11.1	23.5 ± 8.2	37.8 ± 7.3
RyR1	1(4) x (2+1) (5/6)	I	5.7 ± 3.3	12.1 ± 4.3	14.2 ± 4.7	12.9 ± 7.7
		S	14.8 ± 3.1	31.8 ± 6.1	32.6 ± 5.5	36.0 ± 12.5
IP_3R	1(4) x 2 (5/6)	I	8.3 ± 1.4	12.4 ± 5.0	11.7 ± 5.0	18.4 ± 2.4
		S	16.6 ± 2.8	44.0 ± 0.5	39.6 ± 6.3	36.8 ± 0.8
MCU	1(5) x 1 (2)	I	9.5	20.0	16.7	9.5
		S	14.3	35.0	27.8	28.6
SERCA	1 x 4	I	13.3 ± 4.3	17.5 ± 1.5	13.9 ± 5.8	17.6 ± 7.6
		S	31.6 ± 5.3	40.5 ± 1.7	28.8 ± 6.4	39.8 ± 13.5
PL/mv		-	1	4	1	6
CaBS/mv		-	-	2	-	2
Total		-	1	6	1	8
PLN	Domain II	I	10.0	11.1	4.0	7.4
		S	16.7	37.0	40.0	33.3
STIM1	1(2) x 1	I	3.0	11.1	20.0	20.8
		S	18.2	33.3	32.0	37.5

Suppl. Table 3. Sequence alignment on the basis of the individual p-TMDs with specifically the TMD of Vpu of HIV-1. Calculation of identity (I) and similarity (S) using Sequence Manipulation Suite of Clustal Omega on the Vpu sequence UniProt A0A076V5C1.

	TMDs of target protein	I/S	p-TMD1	p-TMD2	p-TMD3	p-TMD4
Vpu	1	I	22.2	17.4	16.7	25.9
		S	33.3	56.5	50.0	51.9

Suppl. Table 4. Predicted sequences of the TMDs of Nsp3 proteins. The proteins are from murine coronavirus especially mouse hepatitis virus (MHV), MERS-CoV (MERS), SARS-CoV, and SARS-CoV-2. Italics: amino acids belonging to the ECTO domain [1].

TMD1

MHV: SRGFFLVATV FLLWFNFLYA NV[22]

MERS: LRLLLMLCTT MVLLSSVYHL YV[22]

SARS-CoV: FTIAMWLLL LSICLGSLIC VTAAFGVLL[28]

SARS-CoV-2: LINIIIWFLL LSVCLGSLIY STAALG[26]

aligned

MHV: SRGFF--LV---ATVF----LLWFNFLYANV[22]

MERS: LRL-----LLMLCTTMVLLSSVYHLYV[22]

SARS-CoV: FTIAMWLLLLSICLGSL----IC--------------V------TAAFGVLL[28]

SARS-CoV-2:

LINIIIWFLLLSVCLGSL----IY--------------S------TAALG[26]

TMD2 no ECTO

MHV: CFY PLFVLIGMQL LTT[16]

MERS: AFNWLLL AGTLHYFFAQ TS[19]

SARS-CoV: FFYL LGLSAIMQVF FGYFASH[21]

SARS-CoV-2: FFYVL GLAAIMQLFF SYFAVHF[22]

TMD2 with ECTO (TMD2)*

MHV: *LFKLVVELVI GYSLYTV*CFY PLFVLIGMQL LTT[33]

MERS: *AFETGLAYML YTS*AFNWLLL AGTLHYFFAQ TS[32]

SARS-CoV: *LGLAAEWVLA YMLFTK*FFYL LGLSAIMQVF FGYFASH[37]

SARS-CoV-2: *GLVAEWFLAY ILFTR*FFYVL GLAAIMQLFF SYFAVHF[37]

aligned

MHV: LFKLVVELVIGYSLYTVCFYPLFVLIGMQLLTT[33]

MERS: AFETGLAYMLYTSAFNWLLLAGTLHYFFAQTS[32]

SARS-CoV: LGLAAEWVLAYMLFTKFFYLLGLSAIMQVFFGYFASH[37]

SARS-CoV-2: GLVAEWFLAYILFTRFFYVLGLAAIMQLFFSYFAVHF[37]

1. Lei J, Kusov Y, Hilgenfeld R (2018). Antiviral Res. 149:58-74

Suppl. Table 5. Sequence alignment of p-TMDs. Calculation of p-TMD identity (I) and similarity (S) using Sequence Manipulation Suite of Clustal Omega. I / S values obtained from using the sequences (I.) without the stretch of TMD2 (-E) which overlaps with the ECDO domain of Nsp3 and (II.) the stretch for TMD2* which includes the overlapping sequence (+E). Each of the p-TMDs of NS4B is compared to the two TMDs of Nsp3 of coronaviruses species murine coronavirus especially mouse hepatitis virus (MHV), middle east respiratory syndrome coronavirus (MERS), SARS-CoV (SARS), and SARS-CoV2 (SARS2). The highest values amongst all the TMDs are highlighted in light and dark orange for identity (I) and similarity (S), respectively. All I and S values are separately averaged (avg.).

I.

p (-E)	TMDs of target protein	I/S	p-TMD1	p-TMD2	p-TMD3	p-TMD4	
MHV	1	I	7.41	3.13	7.41	10.71	
		S	14.81	9.38	22.22	28.57	
	2	I	7.41	14.29	9.09	9.09	
		S	11.11	33.33	31.82	27.27	
MERS	1	I	16.67	7.41	9.09	3.23	
		S	33.33	22.22	36.36	22.58	
	2	I	3.70	10.00	13.64	15.38	
		S	14.81	35.00	31.82	19.23	
SARS1	1	I	15.63	10.34	5.71	18.18	
		S	28.13	31.03	22.86	30.30	
	2	I	15.38	6.25	17.39	13.04	
		S	19.23	12.50	43.48	21.74	
SARS2	1	I	12.90	14.81	9.68	13.79	
		S	29.03	37.04	25.81	41.38	
	2	I	3.85	18.18	12.50	10.00	
		S	23.08	40.91	37.50	23.33	
avg.	avg.	I	10.37 ± 5.39	10.55 ± 4.99	10.56 ± 3.75	11.68 ± 4.54	
		S	21.69 ± 7.98	27.68 ± 11.68	31.48 ± 7.53	26.80 ± 6.98	
	1	I	13.15 ± 4.15	8.92 ± 4.92	7.97 ± 1.79	11.48 ± 6.30	10.38 ± 4.59
		S	26.33 ± 8.01	24.92 ± 12.01	26.81 ± 6.55	30.71 ± 7.85	27.19 ± 8.85
	2	I	7.59 ± 5.47	12.18 ± 5.18	13.16 ± 3.42	11.88 ± 2.88	11.20 ± 4.38
		S	17.06 ± 5.21	30.46 ± 12.39	36.16 ± 5.57	22.89 ± 3.37	26.64 ± 7.47

II.

p (+E)	TMDs of target protein	I/S	p-TMD1	p-TMD2	p-TMD3	p-TMD4	
MHV	1	I	7.41	3.13	7.41	10.71	
		S	14.81	9.38	22.22	28.57	
	2	I	12.12	15.15	9.09	11.76	
		S	24.24	30.30	24.24	29.41	
MERS	1	I	16.67	7.41	9.09	3.23	
		S	33.33	22.22	36.36	22.58	
	2	I	21.88	6.06	11.43	10.26	
		S	28.13	18.18	20	12.82	
SARS1	1	I	15.63	10.34	5.71	18.18	
		S	28.13	31.03	22.86	30.30	
	2	I	13.51	10.81	10.81	7.69	
		S	16.22	21.62	27.03	12.82	
SARS2	1	I	12.90	14.81	9.68	13.79	
		S	29.03	37.04	25.81	41.38	
	2	I	7.5	11.9	8.11	16.22	
		S	12.5	16.67	29.73	27.03	
	avg.	I	13.45 ± 4.78	9.95 ± 4.20	8.92 ± 1.84	11.48 ± 4.74	
		S	23.30 ± 7.74	23.31 ± 8.99	26.03 ± 5.15	25.61 ± 9.50	
avg.	1	I	13.15 ± 4.15	8.92 ± 4.92	7.97 ± 1.79	11.48 ± 6.30	10.38 ± 4.59
		S	26.33 ± 8.01	24.92 ± 12.01	26.81 ± 6.55	30.71 ± 7.85	27.19 ± 8.85
	2	I	13.75 ± 6.00	10.98 ± 3.76	9.86 ± 1.53	11.48 ± 3.58	11.52 ± 4.04
		S	20.27 ± 7.17	21.69 ± 6.10	25.25 ± 4.16	20.52 ± 8.94	21.93 ± 6.82

Suppl. Table 6. Sequence alignment of p-TMDs. Number count of how frequent the individual TMDs score the highest with their identity (I) and similarity (S) values amongst the four TMDs. The summed values of highest I (light orange) and S (dark orange) are calculated for the individual four p-TMDS of NS4B aligned with two TMDs of Nsp3. Each of the p-TMDs of NS4B is compared to the TMDs of Nsp3 of coronaviruses species murine coronavirus especially mouse hepatitis virus (MHV), middle east respiratory syndrome coronavirus (MERS), SARS-CoV (SARS), and SARS-CoV2 (SARS2) (see also Suppl. Table 4). Two different sequences are used for TMD2, one in which the sequence includes part of the ECTO domain of Nsp3 marked by '*' as e.g. TMD2* and one in which this sequence is excluded (TMD2). Formulas used for data for TMD2* are also marked with a star.

Nsp3	NS4B				
TMDs of target protein	I/S	p-TMD1	p-TMD2	p-TMD3	p-TMD4
TMD1	I	1	1		2
	S		1	1	2
TMD2	I		2	1	1
	S		3	1	
∑I	I	1	3	1	3
∑S	S		4	2	2
∑(∑I+∑S)		1	7	3	5
TMD1	I	1	1		2
	S		1	1	2
TMD2*	I	2	1		1
	S	1	1	2	
∑I	I	3	2	-	3
∑S	S	1	2	3	2
∑*(∑I+∑S)		4	4	3	5
∑*(∑I+∑S) - ∑(∑I+∑S)		+3	-3	0	0

Suppl. Table 7. Sequence alignment of r-TMDs. Number count of how frequent the individual TMDs score the highest with their identity (I) and similarity (S) values amongst the four TMDs. The summed values of highest I (light orange) and S (dark orange) are calculated for the individual four r-TMDS of NS4B aligned with two TMDs of Nsp3. Each of the r-TMDs of NS4B is compared to the TMDs of Nsp3 of coronaviruses species murine coronavirus especially mouse hepatitis virus (MHV), middle east respiratory syndrome coronavirus (MERS), SARS-CoV (SARS), and SARS-CoV2 (SARS2) (see also Suppl. Table 7). Two different sequences are used for TMD2, one in which the sequence includes part of the ECTO domain of Nsp3 marked by '*' as e.g. TMD2* and one in which this sequence is excluded (TMD2). Formulas used for data for TMD2* are also marked with a star.

Nsp3	NS4B				
TMDs of target protein	I/S	r-TMD1	r-TMD2	r-TMD3	r-TMD4
TMD1	I	3	1		
	S	1	1	1	1
TMD2	I	2	2		
	S	1	2	2	1
∑I	I	5	3	-	-
∑S	S	2	3	3	2
∑(∑I+∑S)		7	6	3	2
TMD1	I	3	1		
	S	1	1	1	1
TMD2*	I	1		2	1
	S	1		2	1
∑I	I	4	1	2	1
∑S	S	2	1	3	2
∑*(∑I+∑S)		6	2	5	3
∑*(∑I+∑S) - ∑(∑I+∑S)		-1	-4	+2	+1

Suppl. Table 8. Sequence alignment of r-TMDs. Calculation of r-TMD identity (I) and similarity (S) using Sequence Manipulation Suite of Clustal Omega. I / S values obtained from using the sequences (I.) without the stretch of TMD2 (-E) which overlaps with the ECDO domain of Nsp3 and (II.) the stretch for TMD2 including the overlapping sequence (+E). Each of the r-TMDs of NS4B is compared to the two TMDs of Nsp3 of coronaviruses species murine coronavirus especially mouse hepatitis virus (MHV), middle east respiratory syndrome coronavirus (MERS), SARS-CoV (SARS), and SARS-CoV2 (SARS2). The highest values amongst all the TMDs are highlighted in light and dark orange for identity (I) and similarity (S), respectively. All I and S values are separately averaged (avg.).

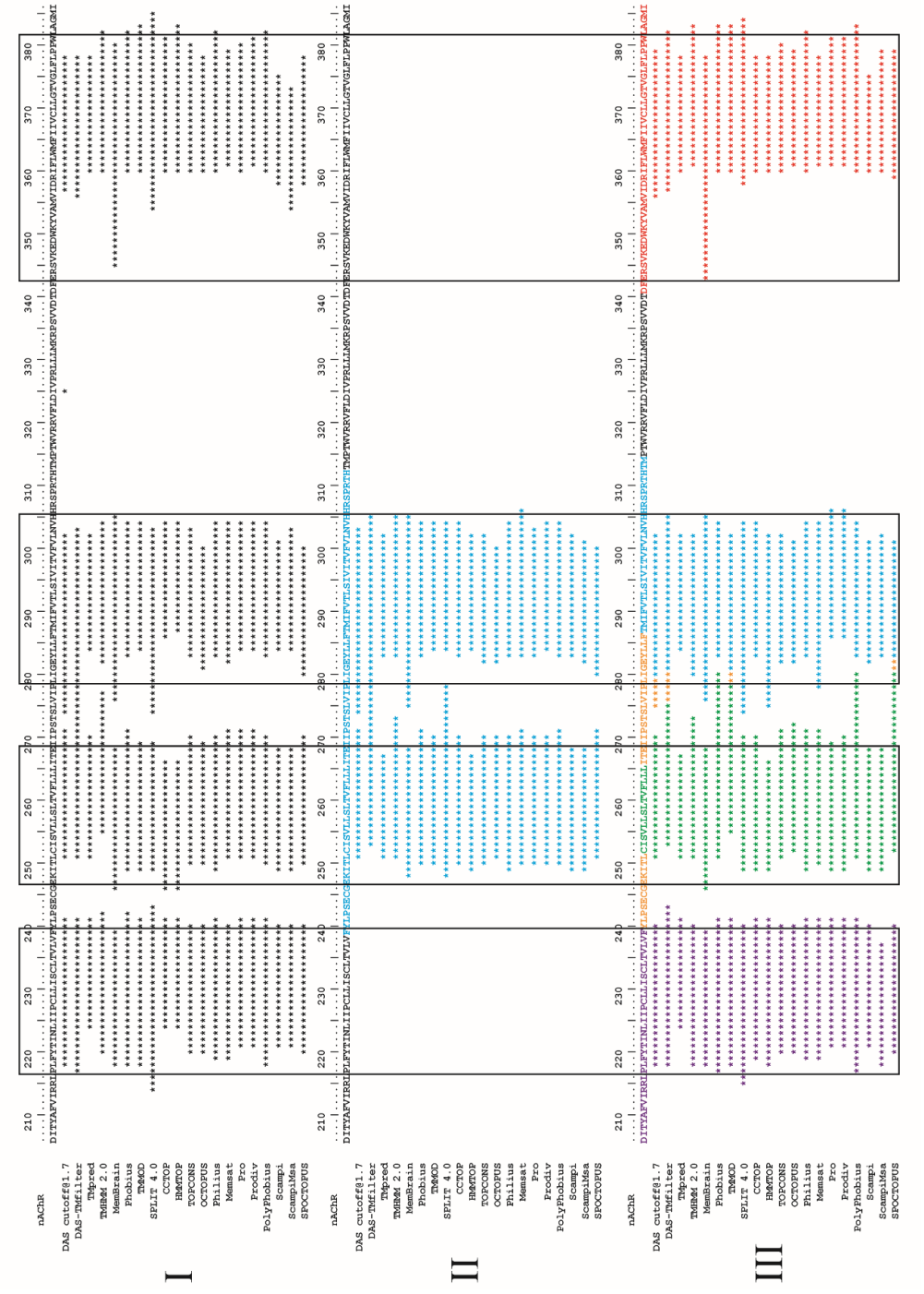
I.

r (-E)	TMDs of target protein	I/S	r-TMD1	r-TMD2	r-TMD3	r-TMD4	
MHV	1	I	18.18	12.00	3.03	11.11	
		S	*40.91*	*32.00*	*9.09*	*29.63*	
	2	I	7.14	15.00	5.26	10.53	
		S	17.86	40.00	26.32	36.84	
MERS	1	I	11.54	12.00	6.45	3.83	
		S	30.77	32.00	16.13	20.00	
	2	I	18.52	7.69	16.67	15.00	
		S	22.22	11.54	25.00	25.00	
SARS1	1	I	17.86	11.76	8.82	10.17	
		S	32.14	26.47	29.41	39.29	
	2	I	20.83	15.38	11.11	13.64	
		S	33.33	26.92	22.22	22.73	
SARS2	1	I	17.24	10.34	16.67	9.38	
		S	31.03	34.48	36.67	18.75	
	2	I	8.33	14.81	7.41	9.09	
		S	16.67	29.63	29.63	27.27	
	avg.	I	14.96 ± 5.18	12.37 ± 2.63	9.43 ± 5.06	10.28 ± 3.48	
		S	28.12 ± 8.40	29.13 ± 8.32	24.31 ± 8.57	27.44 ± 7.49	
avg.	1	I	16.21 ± 3.13	11.53 ± 0.80	8.74 ± 5.79	8.62 ± 3.27	11.27 ± 3.70
		S	33.71 ± 4.83	31.24 ± 3.39	22.83 ± 12.50	26.92 ± 9.57	28.67 ± 8.41
	2	I	13.71 ± 6.97	13.22 ± 3.69	10.11 ± 4.99	12.07 ± 2.73	12.28 ± 4.87
		S	22.52 ± 7.59	27.02 ± 11.76	25.79 ± 3.08	27.96 ± 6.20	25.82 ± 7.81

II.

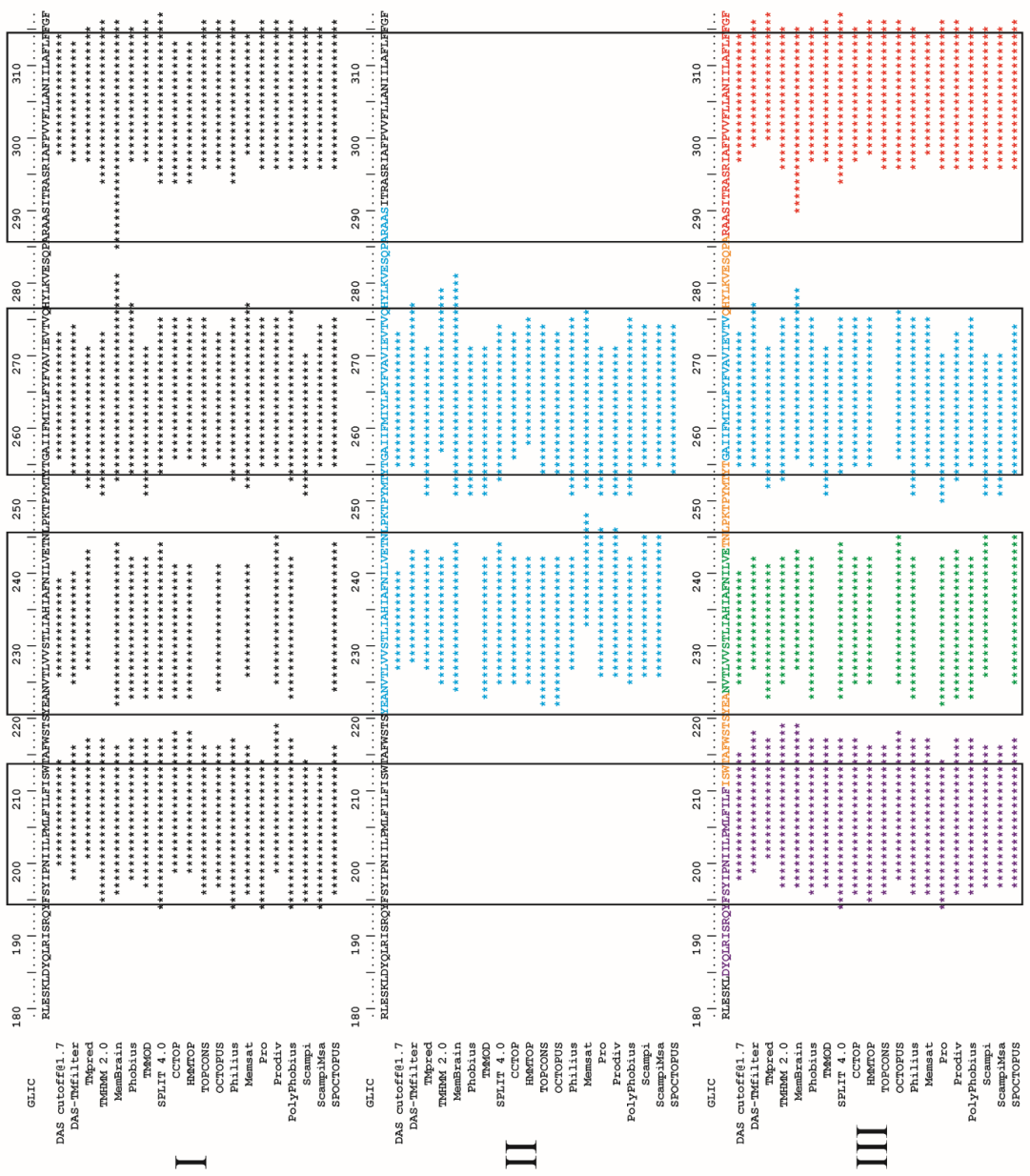
r (+E)	TMDs of target protein	I/S	r-TMD1	r-TMD2	r-TMD3	r-TMD4	
MHV	1	I	18.18	12.00	3.03	11.11	
		S	40.91	32.00	9.09	29.63	
	2	I	4.44	8.33	9.09	5.26	
		S	11.11	27.78	30.30	15.79	
MERS	1	I	11.54	12.00	6.45	3.83	
		S	30.77	32.00	16.13	20.00	
	2	I	9.30	5.13	5.26	9.38	
		S	11.63	7.69	13.16	25.00	
SARS1	1	I	17.86	11.76	8.82	10.17	
		S	32.14	26.47	29.41	39.29	
	2	I	12.50	4.08	6.98	7.89	
		S	20.00	12.24	13.95	13.16	
SARS2	1	I	17.24	10.34	16.67	9.38	
		S	31.03	34.48	36.67	18.75	
	2	I	7.69	6.38	13.51	7.69	
		S	12.82	10.64	32.43	23.08	
	avg.	I	12.34 ± 5.11	8.75 ± 3.24	8.73 ± 4.45	8.09 ± 2.48	
		S	23.80 ± 11.38	22.91 ± 10.90	22.64 ± 10.61	23.09 ± 8.36	
avg.	1	I	16.21 ± 3.13	11.53 ± 0.80	8.74 ± 5.79	8.62 ± 3.27	11.27 ± 3.70
		S	33.71 ± 4.83	31.24 ± 3.39	22.83 ± 12.50	26.92 ± 9.57	28.67 ± 8.41
	2	I	8.48 ± 3.36	5.98 ± 1.83	8.71 ± 3.56	7.56 ± 1.71	7.68 ± 2.75
		S	13.89 ± 4.14	14.59 ± 9.00	22.46 ± 10.32	19.26 ± 5.68	17.55 ± 7.70

A



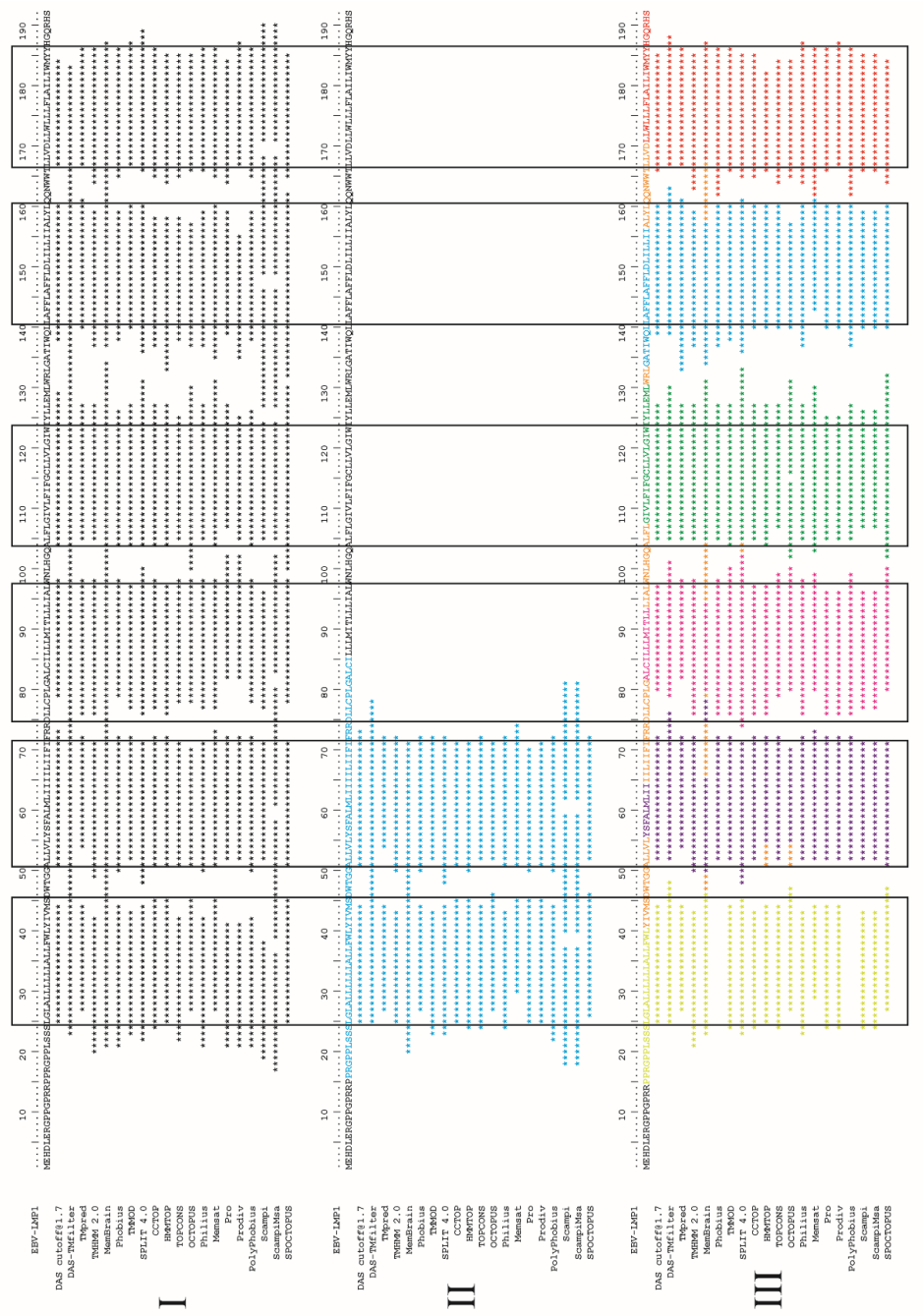
Suppl. Figure 1.

B



Suppl. Figure 1 (cont.)

C

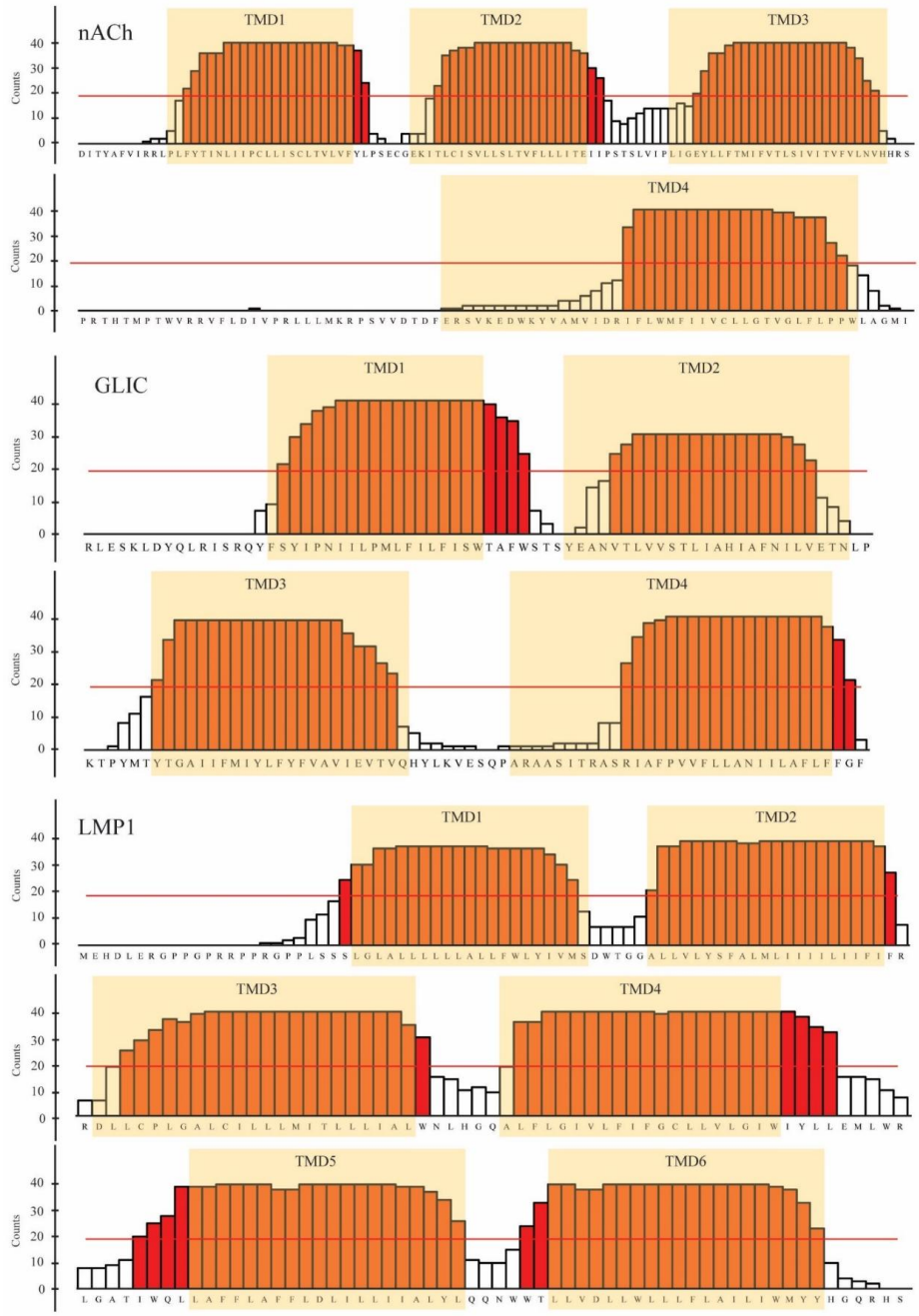


Suppl. Figure 1 (cont.). Identification of the TMDs using secondary structure prediction programs: (A) nAChR, (B) GLIC and (C) LMP1 of EBV. Prediction are made for using (I) the entire length of the NS4B sequence, (II) the region containing the anticipated TMDs only, and (III) the individual proposed stretches which are supposed to contain the respective TMDs. Black boxes are taken from the literature (nAChR: DOI: 10.1038/nature19785, GLIC: DOI: 10.1038/nature07462 and LMP1: DOI: 10.1007/bf02256110) and colour coding: colours refer to the individual TMDs. In case of an overlap the amino acids are highlighted in orange).

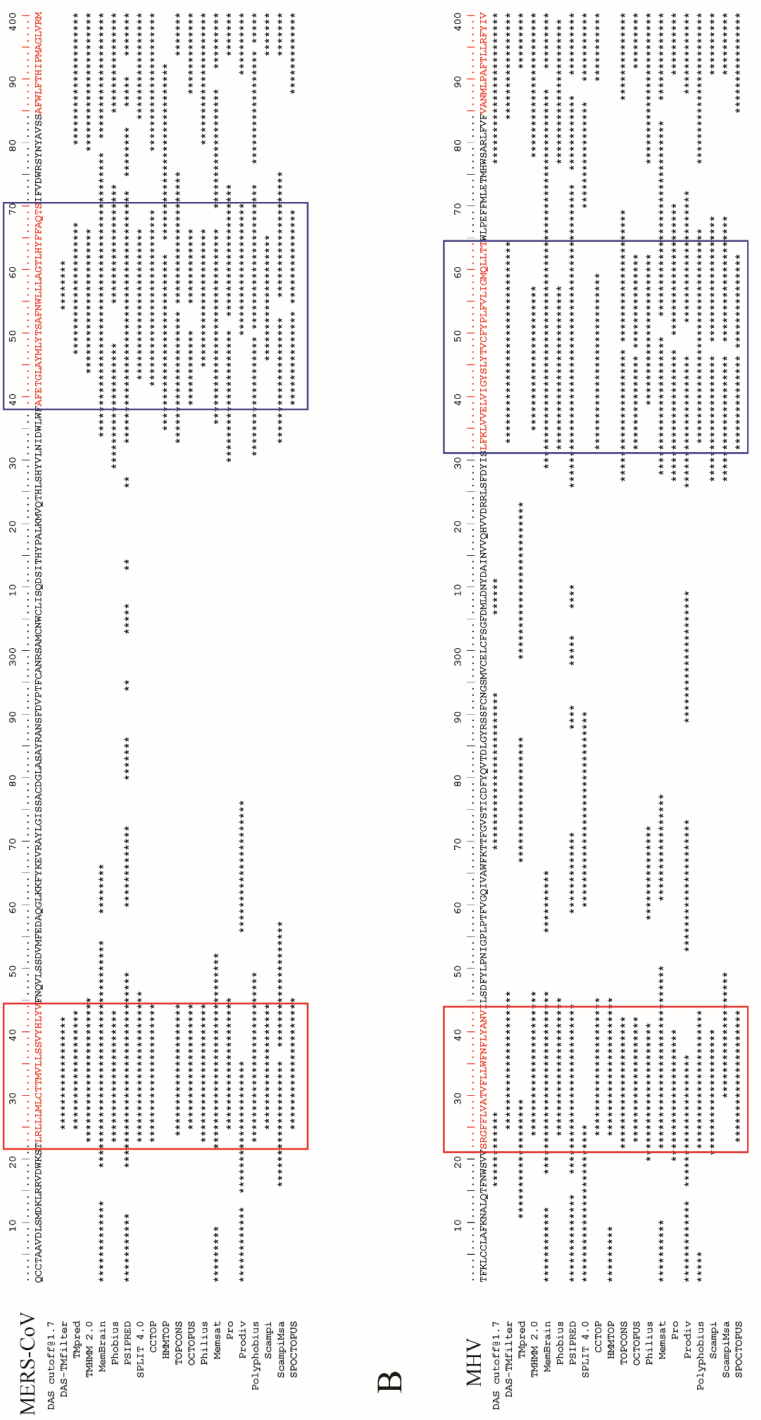
A

B

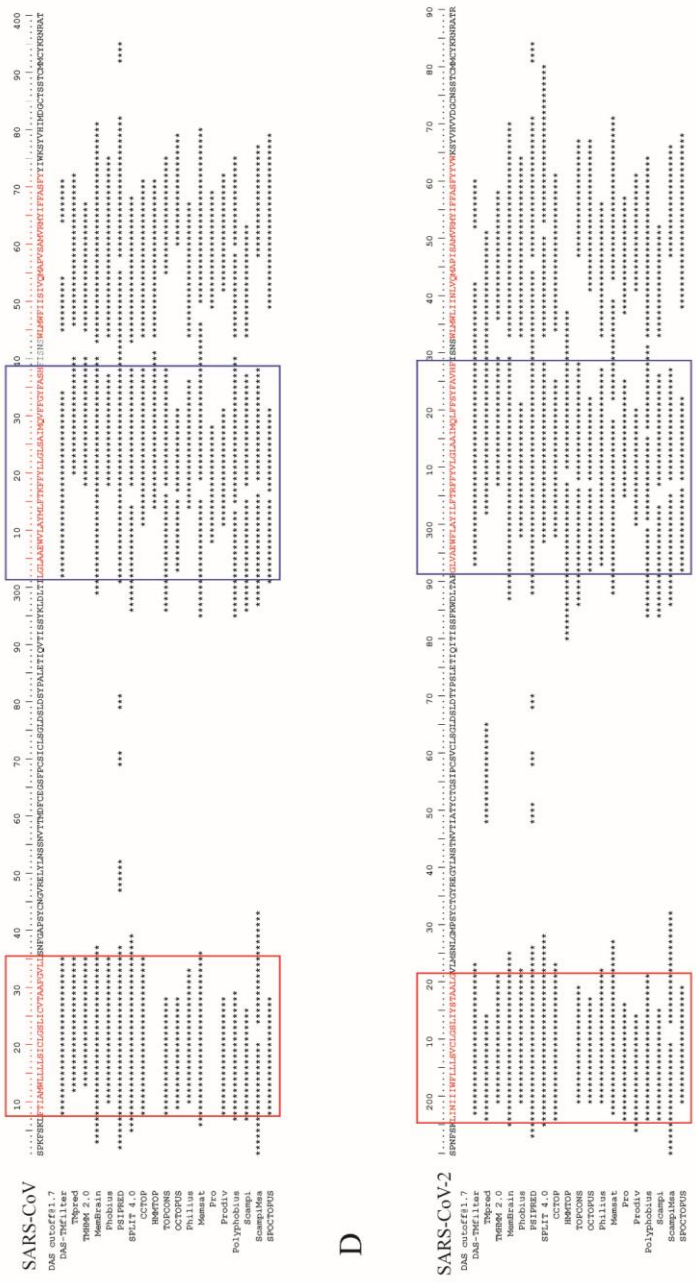
C



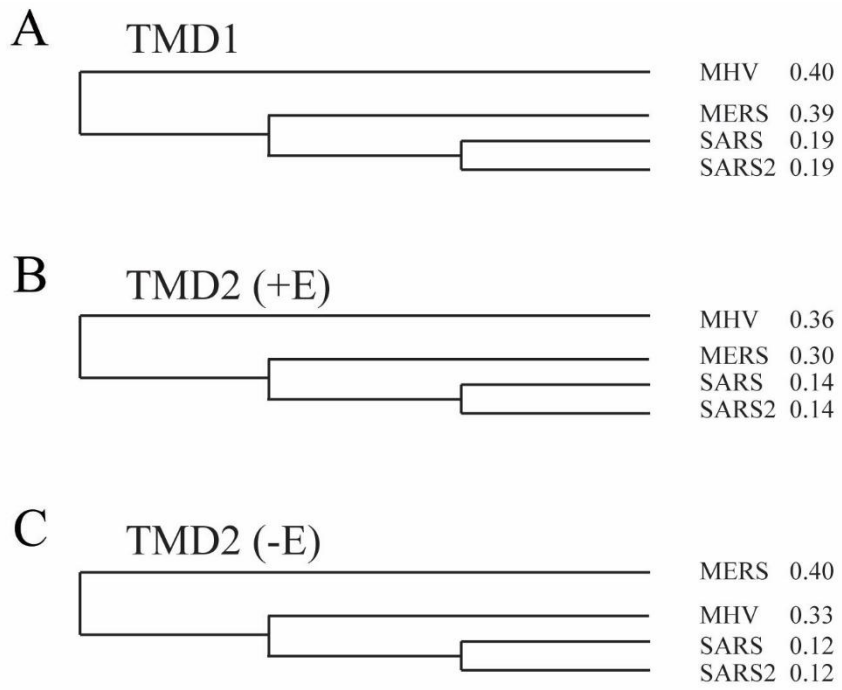
Suppl. Figure 2. Frequency of the amino acids identified to be in a helical conformation by the secondary structure prediction programs (SSPPs): (A) nAChR, (B) GLIC, and LMP1. The numbers are derived from the data presented in Suppl. Figure 1, I and III. The orange/red columns depict the results when using a threshold of equal and more than 19 programs predicting a helical motif for the amino acid. The yellow shade marks the stretches proposed in the literature (nAChR: DOI: 10.1038/nature19785, GLIC: DOI: 10.1038/nature07462 and LMP1: DOI: 10.1007/bf02256110). The match of the predicted results with those from experiments are for nAChR: TMD1 84.0 % TMD2 79.2 %, TMD3 85.2 %, TMD4 53.9%; GLIC TMD1 78.3 %, TMD2 72.0 %, TMD3 95.7 %, TMD4 61.3 %; LMP1: TMD1 90.9 %, TMD2 95.5 %, TMD3 87.5 %, TMD4 79.2 %, TMD5 83.3 %, TMD6 90.9 %.



Suppl. Figure 3.



Suppl. Figure 3 (cont.). Identification of the TMDs of Nsp3 using 20 secondary structure prediction programs: (A) MERS-CoV, (B) MHV (C) SARS-CoV, and (D) SARS-CoV-2. Prediction are made for using the region containing the anticipated TMDs as shown. Red and blue boxes denote the chosen sequence for TMD1 and TMD2, respectively.



Suppl. Figure 4. Phylogram of the individual TMDs of Nsp3 proteins of MERS, HMV, SARS-CoV and SARS-CoV-2. TMD2 includes the sequence with (+E) and without (-E) of the ECDO domain (E) of the proteins. The difference in similarity is given in per centum.