

Supplements to the article:

A tiny viral protein, SARS-CoV-2-ORF7b: Structural features.

Giovanni Colonna

Medical Informatics – AOU Vanvitelli, University of Campania, Naples, Italy.

giovanni.colonna@unicampania.it

Other experimental results.

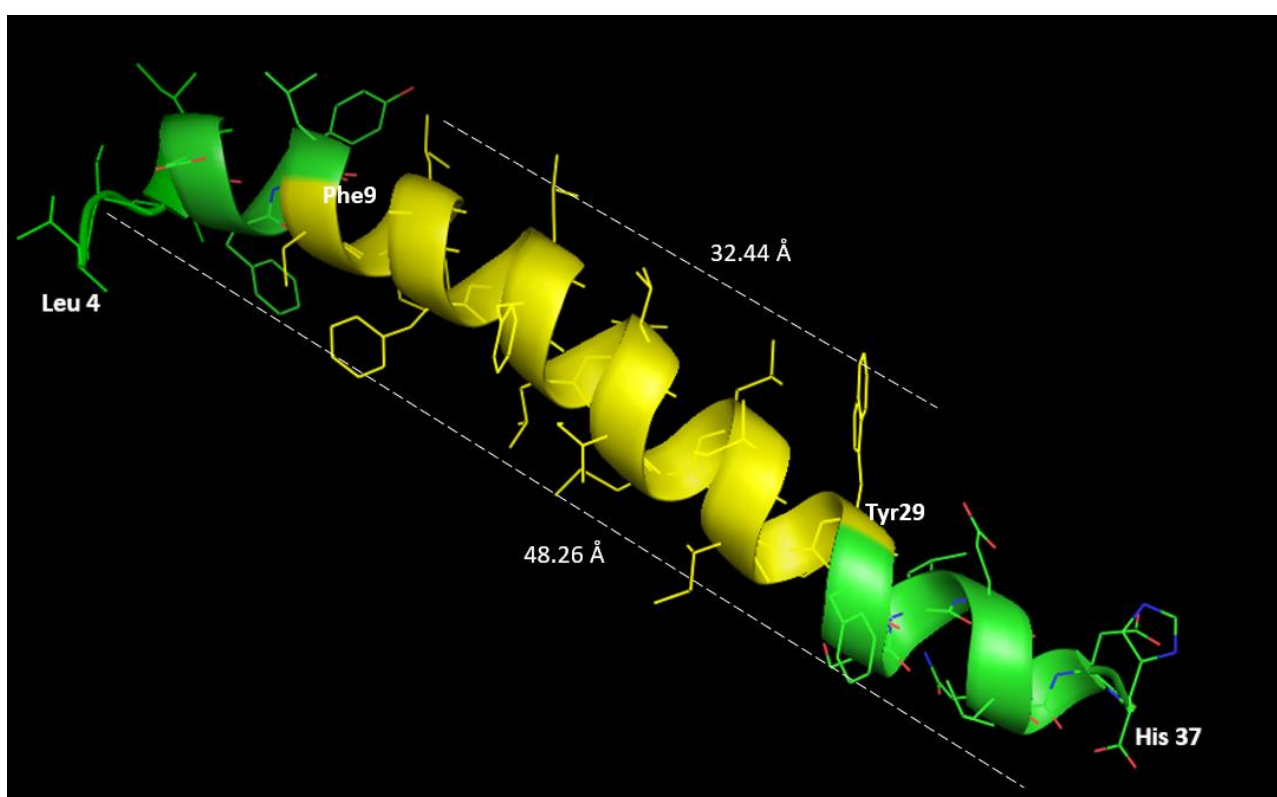


FIGURE 1S - One of the most accredited models, the 3D model of ORF7b-2 from ModBase (University of California San Francisco–UCSF). The model shows only a long single alpha-helix. The prediction does not include terminal residues. Segment 9-29 (in yellow) shows the part of the helix that should be transmembrane. The rigid representation of the helix, without considering its conformational dynamics, its electrostatic properties, and the inter-residue distances, is perhaps the main reason which has led many authors to consider the protein as transmembrane. Template PDB Code: 4uvmA (template region: 478-511 with sequence identity: 29%). Predicted: from residue 4 to 37. Model from PyMol in PNG format.

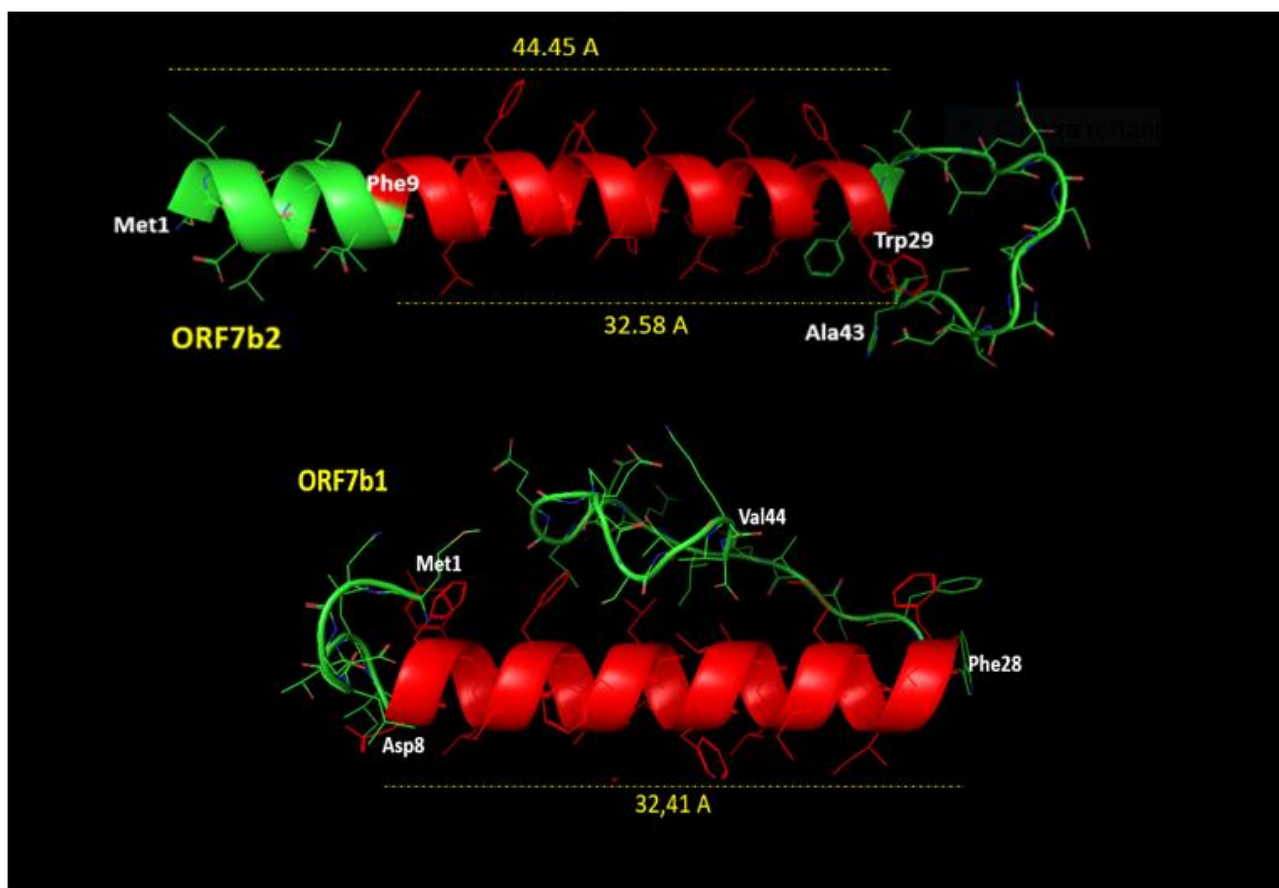


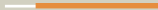







FIGURE 2S – PEP-FOLD3 models of ORF7b-2 and ORF7b-1. The figure shows the dimensions in Å of the two proteins. The segment from 1 to 9 is reported by dynamics as highly mobile around the hinge residue 9. Models from PyMol in PNG format.

TABLE 1S - ORF7b-2 - PHYRE2 - Template analysis

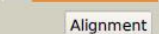

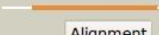





Detailed template information

#	Template	Alignment Coverage	3D Model	Confidence	% i.d.	Template Information
1	c2zxeB_	 <div>Alignment</div>		88.0	21	PDB header: hydrolase/transport protein Chain: B: PDB Molecule: na+,k+-atpase beta subunit; PDBTitle: crystal structure of the sodium - potassium pump in the e2.2k+.pi2 state PDB Entry: PDBe RCSB PDBj
2	c7wyvB_	 <div>Alignment</div>		85.8	21	PDB header: membrane protein Chain: B: PDB Molecule: na+,k+-atpase beta subunit; PDBTitle: cryo-em structure of na+,k+-atpase in the e2p state formed by atp in2 the presence of 40 mm mg2+ PDB Entry: PDBe RCSB PDBj
3	c7wytB_	 <div>Alignment</div>		85.0	14	PDB header: membrane protein Chain: B: PDB Molecule: sodium/potassium-transporting atpase subunit beta-1; PDBTitle: crystal structures of na+,k+-atpase in complex with ouabain PDB Entry: PDBe RCSB PDBj
4	c7e21B_	 <div>Alignment</div>		81.1	15	PDB header: membrane protein Chain: B: PDB Molecule: sodium/potassium-transporting atpase subunit beta-1; PDBTitle: cryo em structure of a na+-bound na+,k+-atpase in the e1 state with2 atp-gamma-s PDB Entry: PDBe RCSB PDBj

Domain analysis	
Rank	Aligned region
1	c2zxeB_
2	c7wyvB_
3	c7wytB_
4	c7e21B_
5	c3kdpD_
6	c2yn9B_
7	c2xzbB_
8	c3b8eB_
9	c7vu5B_
10	c7vu5A_
11	c2kogA_
12	c6lumO_
13	c6lumE_
14	c6lumI_
15	c2n1pA_
16	c7k3gE_
17	c7k3gA_
18	c7k3gC_
19	c7k3gB_
20	c7k3gD_

The domain analysis illustrates where along your sequence matches have been found, color-coded by confidence. The matches are ranked by an alignment that is based on the number of aligned residues and the quality of alignment. This is based on the similarity of residue probability distributions for each position, secondary structure similarity and the presence or absence of insertions and deletions. Each row provides information on the template used for the model and a small graphic showing where along your sequence is the match color-coded by confidence.

TABLE 2S - ORF7b-1 - PHYRE2 - Template analysis

Detailed template information						Cattura rettangolare
#	Template	Alignment Coverage	3D Model	Confidence	% I.d.	Template Information
1	c2zxeB	 Alignment		86.4	22	PDB header: hydrolase/transport protein Chain: B; PDB Molecule: na+,k+-atpase beta subunit; PDBTitle: crystal structure of the sodium - potassium pump in the e2.2k+.pi2 state PDB Entry: PDBe RCSB PDBj
2	c7wyvB	 Alignment		83.9	22	PDB header: membrane protein Chain: B; PDB Molecule: na+,k+-atpase beta subunit; PDBTitle: cryo-em structure of na+,k+-atpase in the e2p state formed by atp in2 the presence of 40 mm mg2+ PDB Entry: PDBe RCSB PDBj
3	c7wytB	 Alignment		83.2	15	PDB header: membrane protein Chain: B; PDB Molecule: sodium/potassium-transporting atpase subunit beta-1; PDBTitle: crystal structures of na+,k+-atpase in complex with ouabain PDB Entry: PDBe RCSB PDBj
4	c7e21B	 Alignment		78.5	15	PDB header: membrane protein Chain: B; PDB Molecule: sodium/potassium-transporting atpase subunit beta-1; PDBTitle: cryo em structure of a na+-bound na+,k+-atpase in the e1 state with2 atp-gamma-s PDB Entry: PDBe RCSB PDBj

Domain analysis	
Rank	Aligned region
1	c2zxeB
2	c7wyvB
3	c7wytB
4	c7e21B
5	c3kdpD
6	c2yn9B
7	c2xzbB
8	c3b8eB
9	c7vu5B
10	c7vu5A
11	c2kogA
12	c7k3gC
13	c7k3gA
14	c7k3gB
15	c7k3gE
16	c7k3gD
17	c6lumO
18	c6lumE
19	c6lumI
20	c2nlpA

TABLE 3S - ORF7b-2 (PHYRE2 model) – Ramachandran statistics

Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	41	100.00%	Goal: >98%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	37	90.24%	Goal: >98%
	Rama distribution Z-score	1.40 ± 1.09		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	4 / 375	1.07%	Goal: 0%
	Bad angles:	2 / 509	0.39%	Goal: <0.1%

TABLE 4S - ORF7b-2 (PEP-FOLD3 model) – Ramachandran statistics

Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	35	97.22%	Goal: >98%
	Ramachandran outliers	1	2.44%	Goal: <0.05%
	Ramachandran favored	36	87.80%	Goal: >98%
	Rama distribution Z-score	0.92 ± 1.15		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	5 / 373	1.34%	Goal: 0%
	Bad angles:	3 / 506	0.59%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 0	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	1	2.6%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/61		

TABLE 5S - ORF7b-1 (PHYRE2 model) – Ramachandran statistics

Protein Geometry	Poor rotamers	0	0.00%	Goal: <0.3%
	Favored rotamers	43	100.00%	Goal: >98%
	Ramachandran outliers	1	2.38%	Goal: <0.05%
	Ramachandran favored	38	90.48%	Goal: >98%
	Rama distribution Z-score	-6.78 ± 0.80		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	12 / 381	3.15%	Goal: 0%
	Bad angles:	21 / 517	4.06%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 1	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	0	0.0%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/64		

TABLE 6S - ORF7b-1 (PEP-FOLD3 model) – Ramachandran statistics

Protein Geometry	Poor rotamers	2	5.26%	Goal: <0.3%
	Favored rotamers	35	92.11%	Goal: >98%
	Ramachandran outliers	3	7.14%	Goal: <0.05%
	Ramachandran favored	34	80.95%	Goal: >98%
	Rama distribution Z-score	-0.52 ± 1.18		Goal: abs(Z score) < 2
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	1 / 380	0.26%	Goal: 0%
	Bad angles:	0 / 516	0.00%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 1	0.00%	Expected: ≤1 per chain, or ≤5%
Low-resolution Criteria	CaBLAM outliers	4	10.0%	Goal: <1.0%
	CA Geometry outliers	0	0.00%	Goal: <0.5%
Additional validations	Chiral volume outliers	0/64		

Tables show the statistical analysis of the best PHYRE2 and PEP-FOLD3 models calculated from the Ramachandran plot,

Details of ORF7b-2 Dynamics

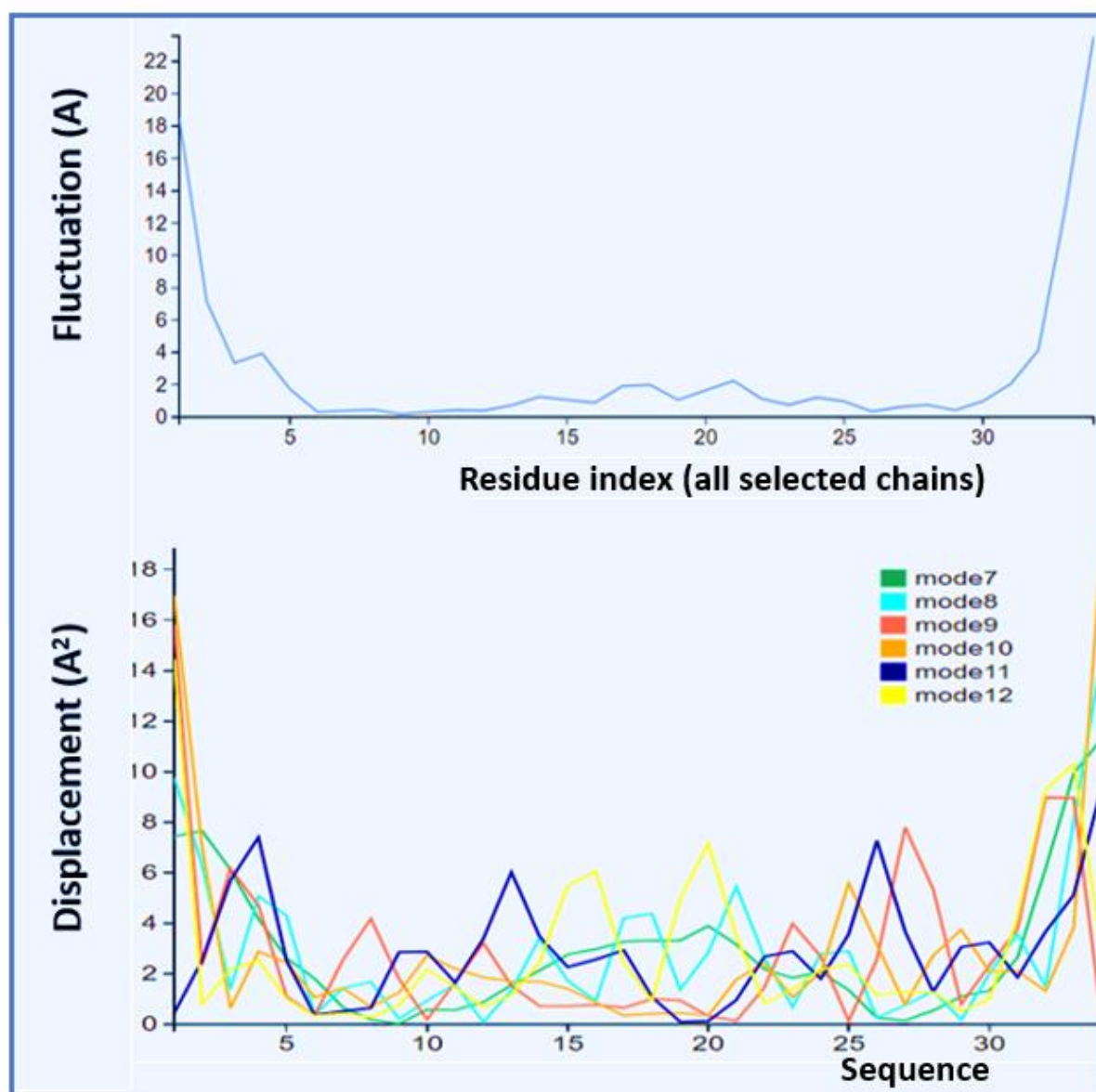


Figure 3S - Normal mode analysis (NMA) of ORF7b-2 - Fluctuations (top) - Mean residue index correlation from the considered modes. Displacements (bottom) for several modes. Profile positions where model estimates are less mobile for each mode are in the center, while the external segments are very mobile and fluctuating.

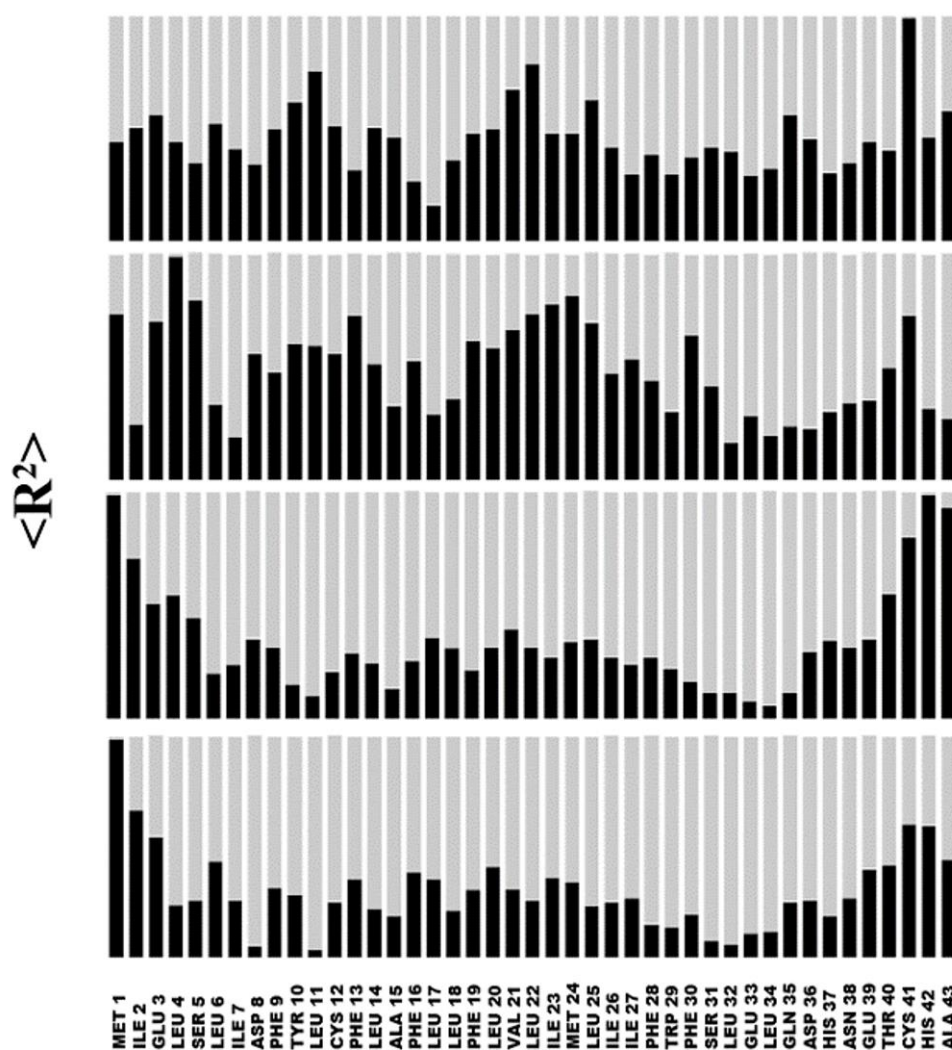


Figure 4S – Normal mode analysis (NMA) of ORF7b-2 - Displacement for single residue according to modes 11, 12, 8, and 7 (from top to bottom). The comparison between the four modes shows a dynamic picture of the residues in which also the central helical residues experience conformational changes because of bending and twisting.

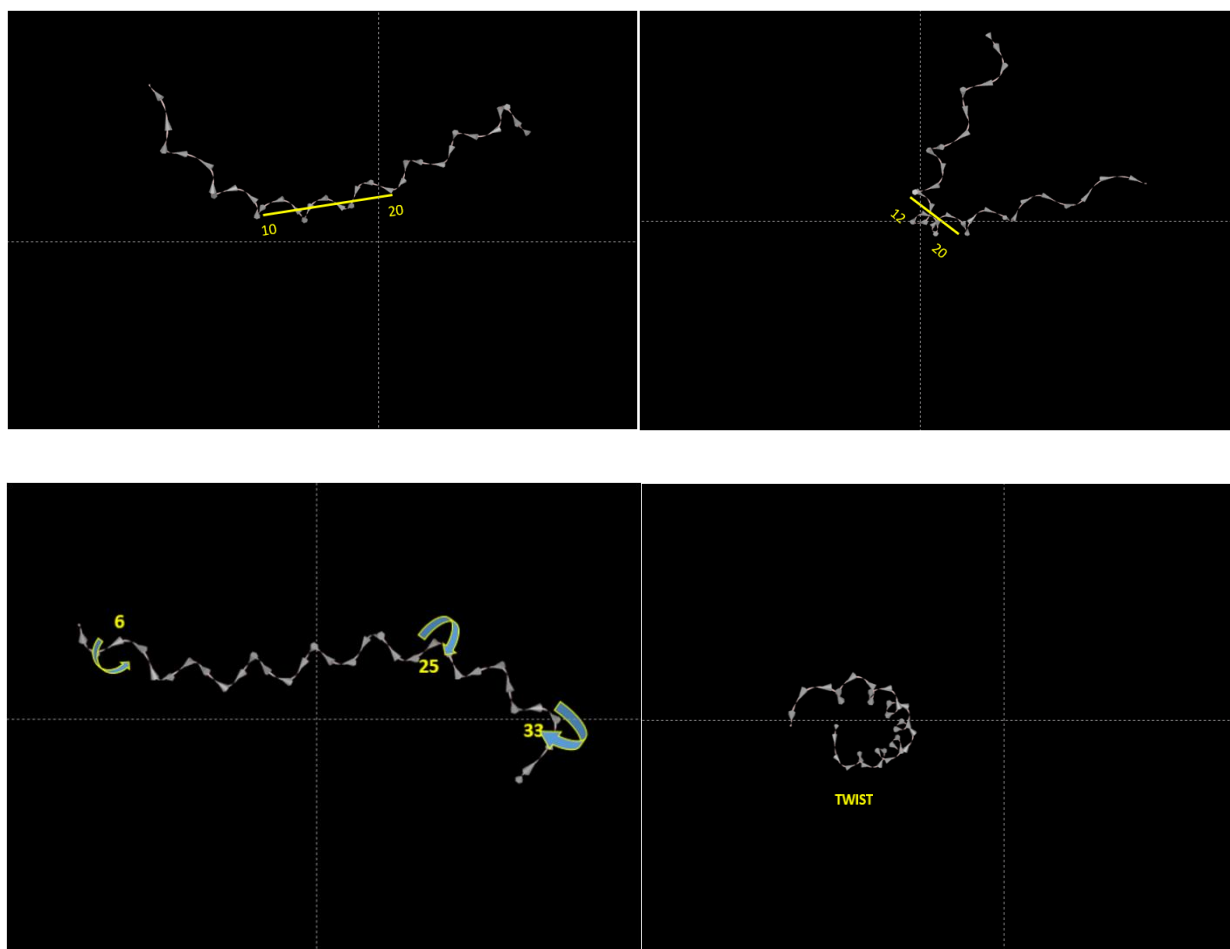


Figure 5S - Normal mode analysis (NMA) of ORF7b-2 - Some snapshots showing the conformational movements of ORF7b-2 calculated by dynamics (bending and twisting). In yellow, the hinge residues.

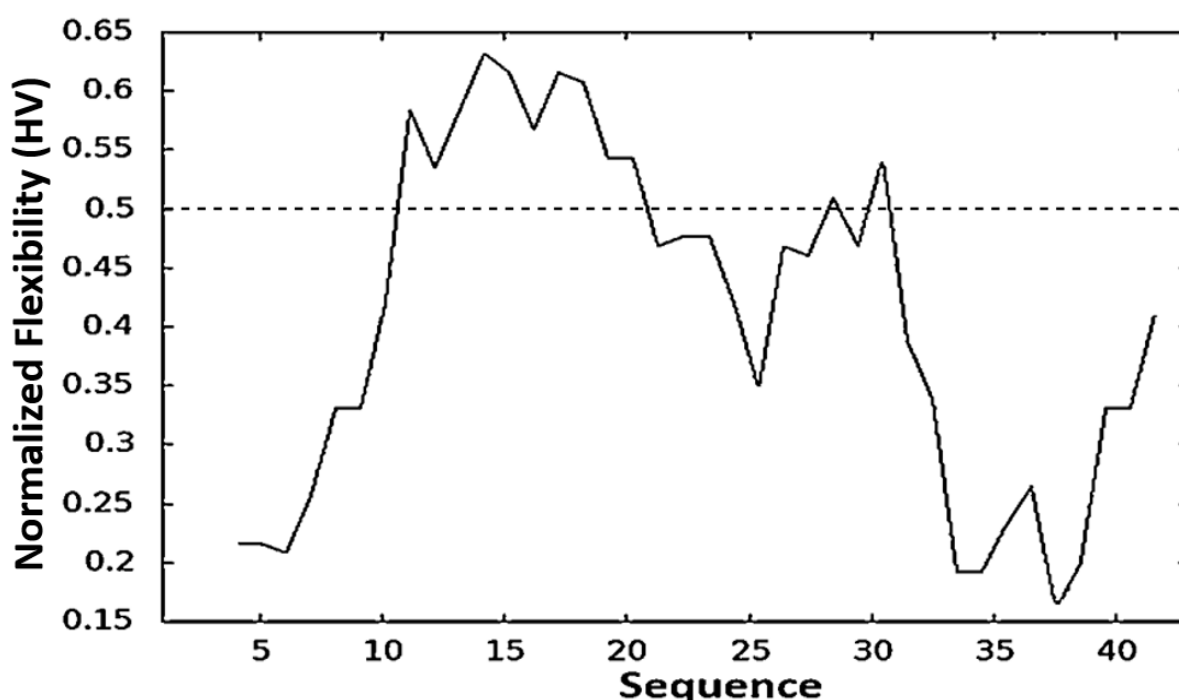


Figure 6S – Flexibility plot of ORF7b-2. The flexibility of a protein depends on the amino acid residues in the highly mobile regions (1). Amino acids possessing the smallest volumes and lowest hydrophobicities are preferred in regions with loops, turns or extended coils because of the flexibility these residues impart. Thus, we can generate a plot by calculating the value of the hydrophobicity-volume product for consecutive quintuplets of amino acid residues. The occurrence of small volumes and low hydrophobicities generate minima representative of flexibility. The low mean values of the HV product show that the protein comprises many residues with small volumes and low hydrophobicity, giving it great flexibility. This result agrees with the results of molecular dynamics. Normalization is on a scale of 0 - 1.

- 1) R. Ragone, F. Facchiano, A. Facchiano, A.M. Facchiano, G. Colonna. Flexibility plot of proteins. *Protein Engineering, Design and Selection*, Volume 2, Issue 7, May 1989, Pages 497–504, <https://doi.org/10.1093/protein/2.7.497>

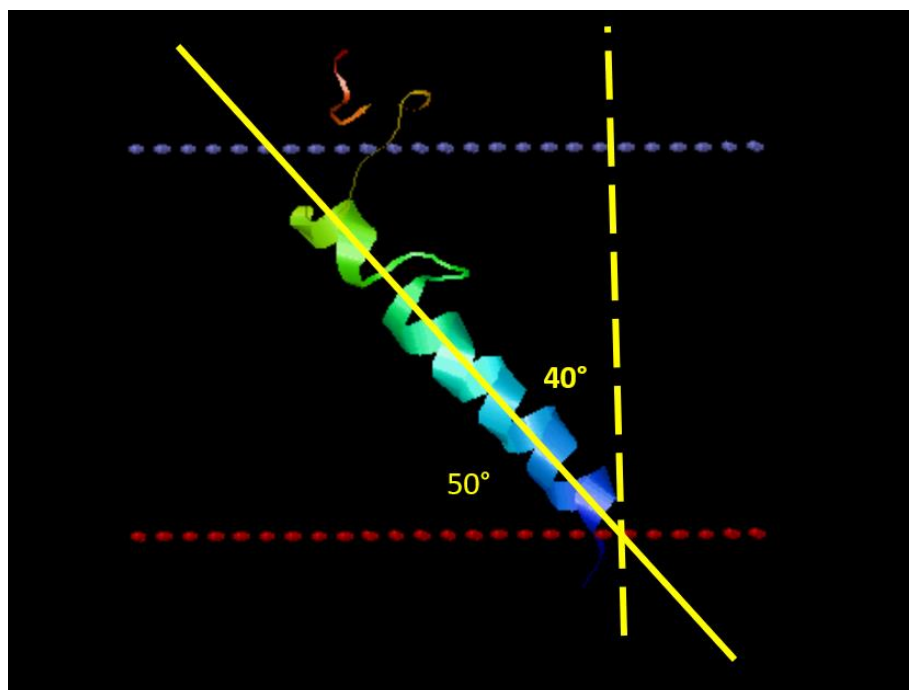




Figure 7S - In the figure, we can observe a model of the ORF7b-2 insertion in the membrane showing a tilt of about 40°. The unfolding of the helix-turns, due to structural stress, is clearly visible.


TABLE 7S

Dipole moment for ORF7b-2

	No. of Chains=1		Prolate							
	No.Atoms	No.Res.	R _M	Pos.Res.	Neg.Res.	Charge	Dipole	Quadrupole	Crg./Nat.	Dip./Nat.
Value	365.	43.	109.01	0.	5.	-5.	488.	1066.	-0.0137	1.3370
No.Dev.Units	-1.19	-1.22	-1.01	-1.47	-1.16	-0.23	-0.13	-0.42	-2.16	3.42

 Dipole vector (in atomic units): -84.37 -16.03 54.30

 Mass Moments vector: 230.47 106.03 206.34

 Open a larger Jmol window.

The table shows the structural and physical parameters involved in the calculation of the dipole vector.

Interaction ORF7b2--ORF7b2

PEPPI results: $\log(\text{LR}) = -1.321$

The chains are not predicted to interact

Top 10 SPRING results

SPRING score ^a	PDB hit
9.726	6FL9
9.726	6FL9
9.681	2XQ3
9.681	2XQ3
9.626	5VKQ
9.626	5VKQ
9.41	6NJL
9.41	6NJL
9.305	6A96
9.305	6A96

SPRING score is a scoring function comprising a weighted sum of the threading Z-score, the model superposition TM-score, and the interface energy.

Top 10 BLAST results

Total SeqID	chain A (seqID) ^a	chain B (seqID) ^b
3.4%	uniprotkb:P02870 (4.0%)	uniprotkb:P02867 (2.9%)
3.4%	uniprotkb:P02867 (2.9%)	uniprotkb:P02870 (4.0%)
3.2%	uniprotkb:P26718 (3.2%)	uniprotkb:P26718 (3.2%)
3.2%	DIP:31100N (3.2%)	DIP:31100N (3.2%)
2.9%	springdb:6FL901A1 (2.9%)	springdb:6FL901B1 (2.9%)
2.9%	springdb:6FL901B1 (2.9%)	springdb:6FL901A1 (2.9%)
2.9%	uniprotkb:O43567 (2.9%)	uniprotkb:O43567 (2.9%)
2.9%	uniprotkb:Q8IEI6 (2.9%)	uniprotkb:Q8IEI6 (2.9%)
2.9%	entrez_gene/locuslink:814054 (2.9%)	entrez_gene/locuslink:814054 (2.9%)
2.8%	uniprotkb:P38310 (2.8%)	uniprotkb:P38310 (2.8%)

(a) id-number of chain A in the sequence database, and the percentage sequence identity between query chain and the database hit.

(b) id-number of chain B in the sequence database, and the percentage sequence identity between query chain and the database hit.

The neural network classifier CT predicted probability of negative result is 1.000.

The Zhang lab (Department of Computational Medicine and Bioinformatics - Department of Biological Chemistry - University of Michigan Medical School) (<https://zhanggroup.org/>) has developed an online platform, called PEPPI (see methods). PEPPI (Pipeline for the Extraction of Predicted Protein-protein Interactions) (<https://seq2fun.dcmdb.med.umich.edu/PEPPI/>) is a computational program for protein-protein interaction prediction. Given a pair of protein amino acid sequences, PEPPI predicts the likelihood of direct, physical interaction for those sequences through several independent prediction methods, including protein structural homology by multimeric threading, protein sequence homology by BLAST search through high-throughput experimental data, functional association from the STRING database, and machine learning-based classification. Scores from each of these approaches are combined through a naive Bayesian consensus model into a final likelihood ratio, expressing the probability of interaction relative to the probability of non-interaction. PEPPI is efficiently implemented for the modeling of proteome-wide interaction networks.

The calculation about the interaction [ORF7b-2--ORF7b-2] does not predict the chains to interact. Two approaches, the Structure similarity method [SPRING] and the Non-interaction similarity method [SPRINGNEG] classify as unfavorable to this interaction; SPRING with a log [LR] of -1.321 and with a low SPRING score of 9.726 for the best model, while SPRINGNEG with a score of 8.997. Where the SPRING score is a scoring function comprising a weighted sum of the threading Z-score, the model superposition TM-score, and the interface energy. This also means that, from a structural and energetic point of view, under the experimental conditions used by PEPPI, the probability of interaction is very low.

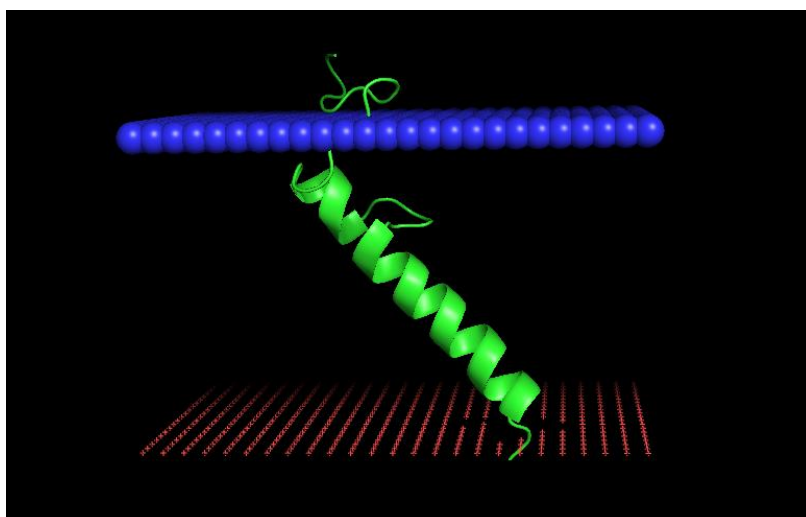
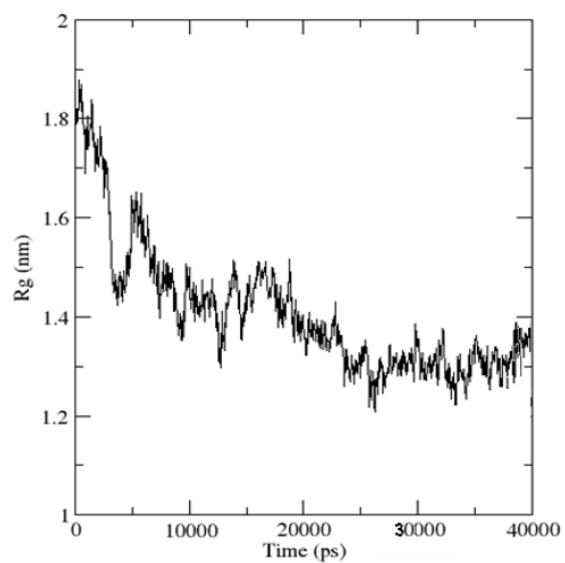
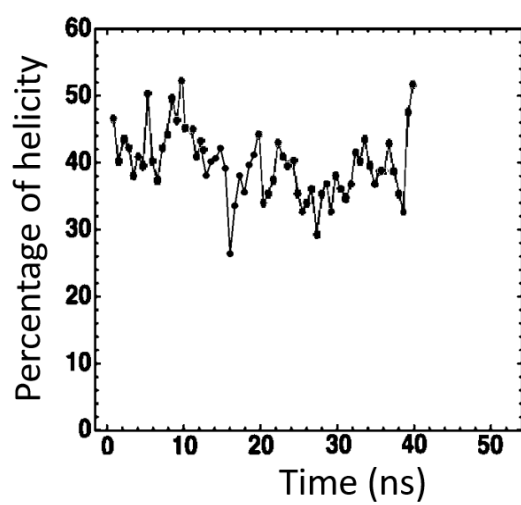
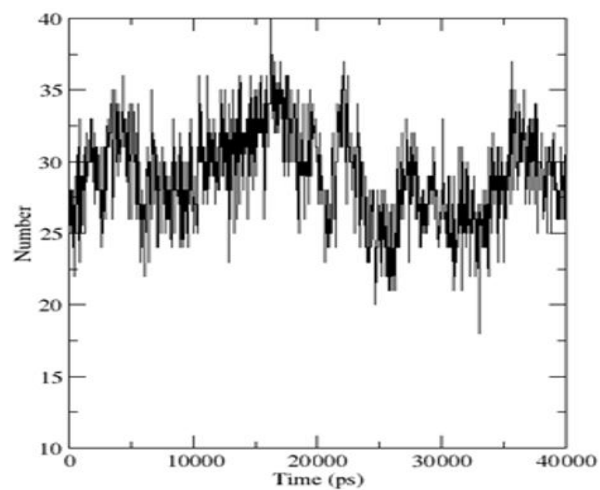


Figure 9S shows another attempt to visualize the insertion of a single molecule of ORF7b-2 into a membrane by using Memembed, an algorithm of the PSIPRED web server [see methods]. The protein shows a tilt angle of 40°, with respect to the axis normal to the surface of the membrane. The terminal segments are free outwards both from the luminal and cytoplasmic sides and the unfolding of some residues inside the membrane appears as a signal of structural stress. We should also consider that both the dipole moment, which is a function of the angle of inclination, and the exposure to the solvent of the terminals, contribute to reducing the dipole of the helix. This could explain the low value found for ORF7b-2

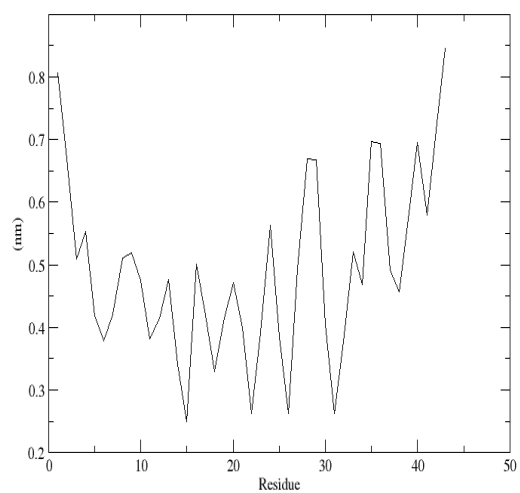
Radius of gyration (total and around axes)



Hydrogen Bonds



RMS fluctuation



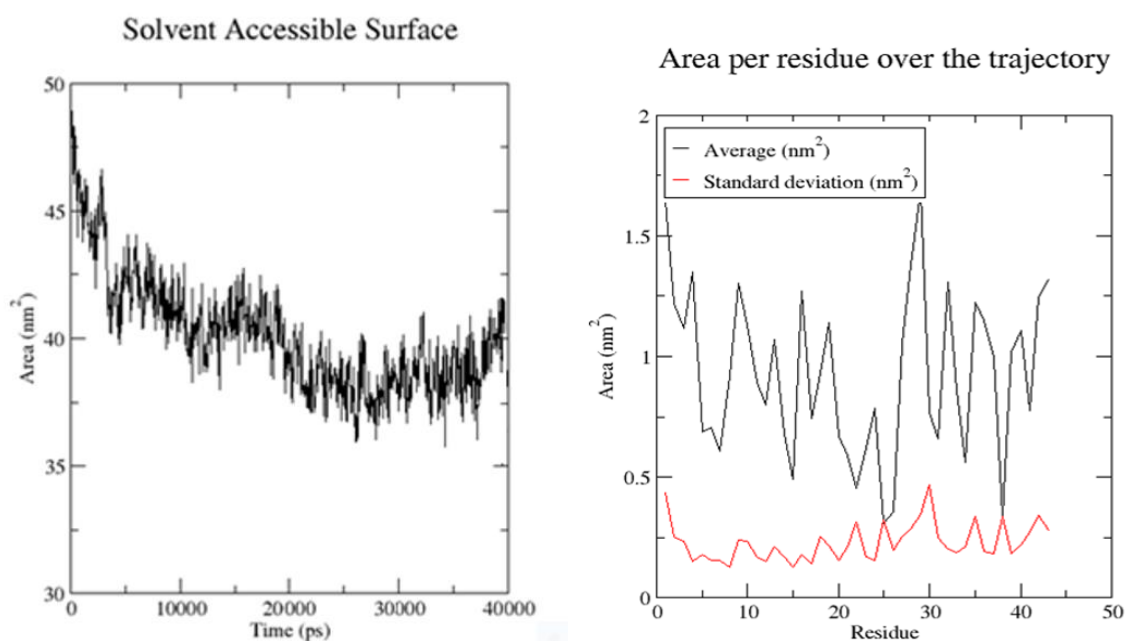
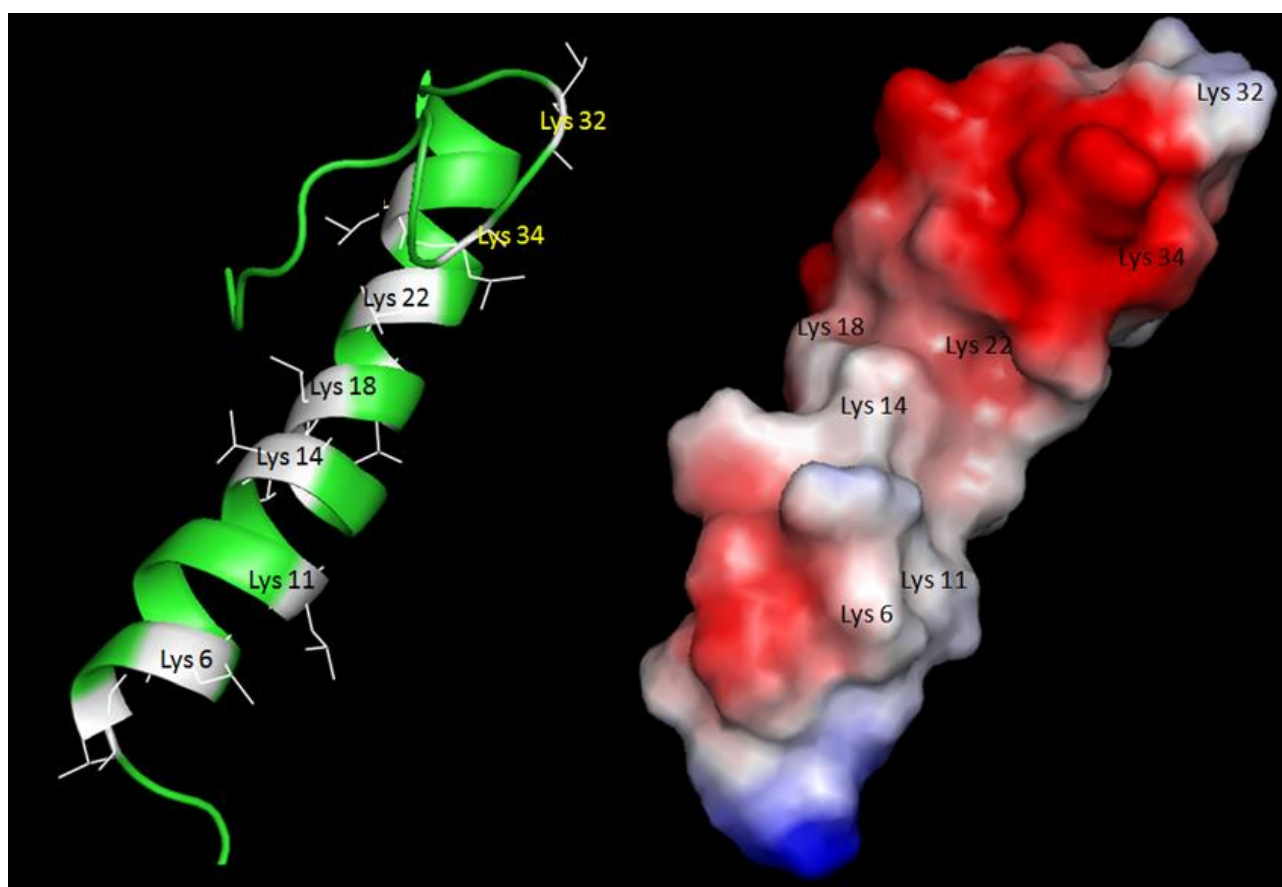


Figure 10S - The figure shows the set of parameters that characterize the trend of the molecular dynamics simulation of ORF7b-2 in water for 40 ns.



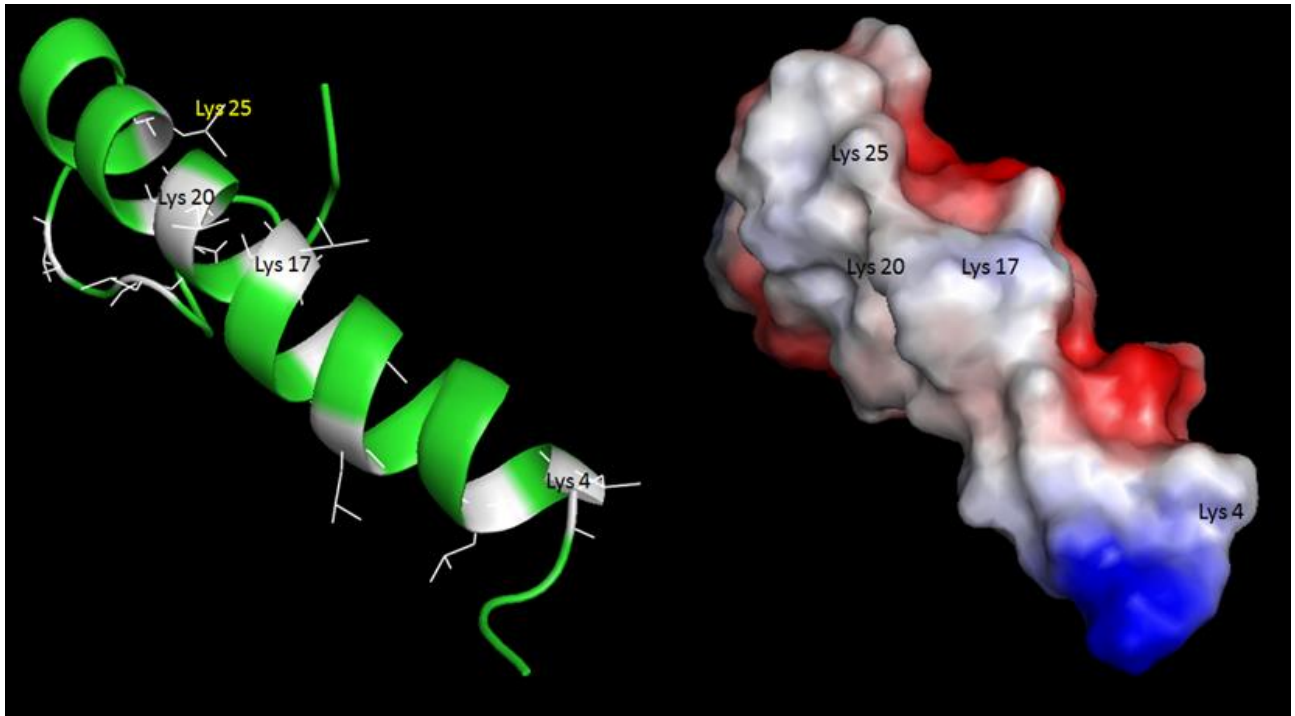


Figure 11S – The two figures show the two sides of the structural model got at 40 ns by molecular dynamics in water. Lysines are distributed on the two opposite sides of the molecule in environments with different electrostatic characteristics. The majority is located on the side with a large negative surface. Thus, a Lys-zip with a linear sequence contained in a single apolar surface is physically missing.

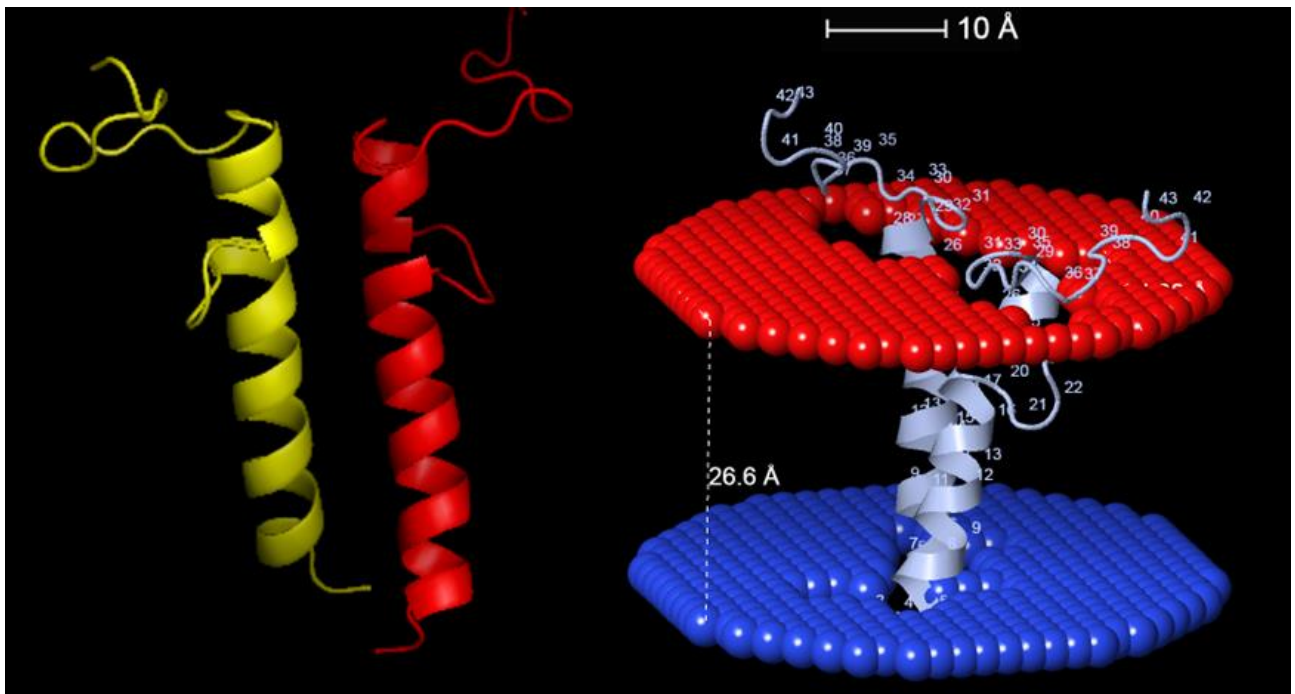


Figure 12S – The figure shows on the left the model got by HDOCK and, on the right, the model simulating the pre-orientation in the Golgi membrane got through the OPM database.

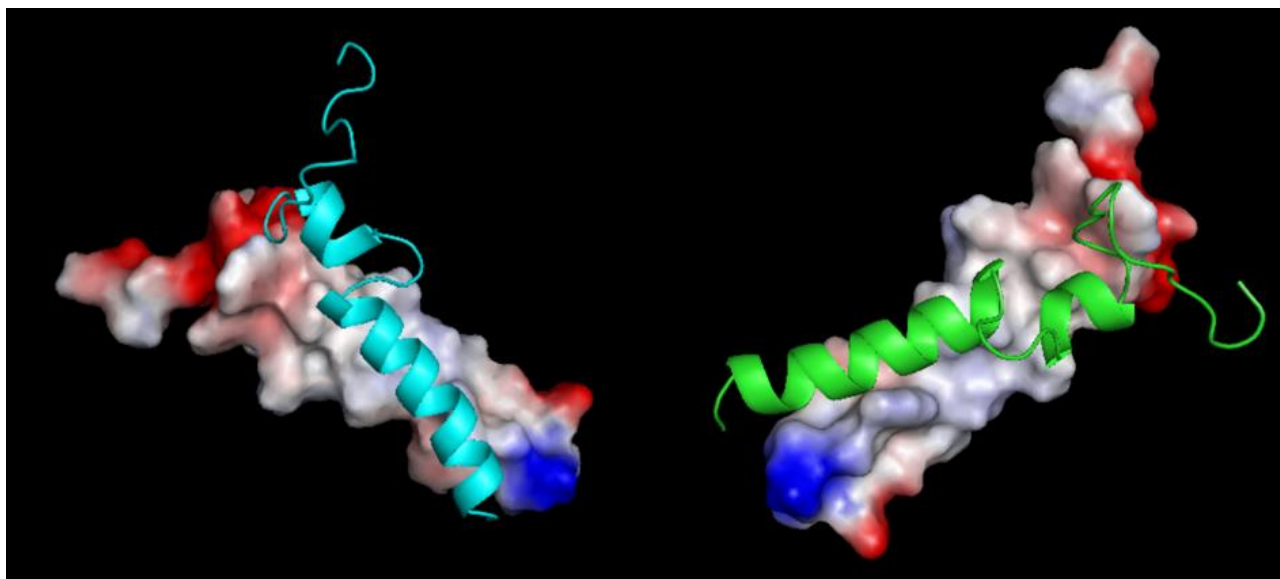
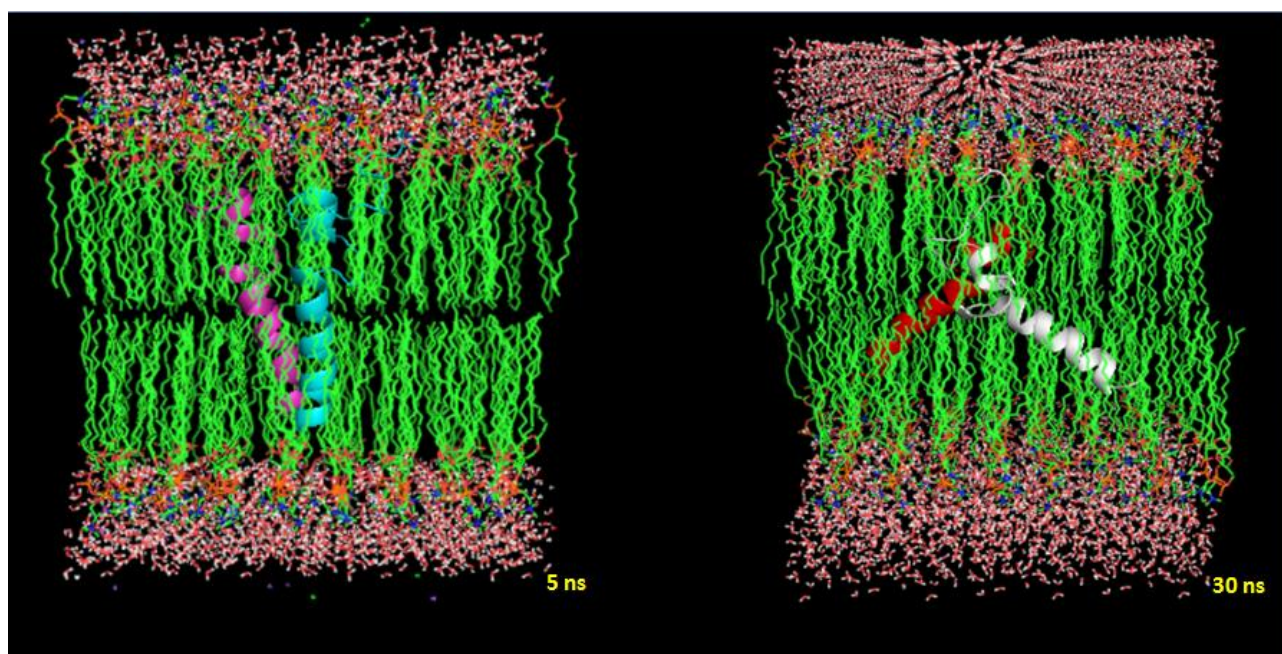
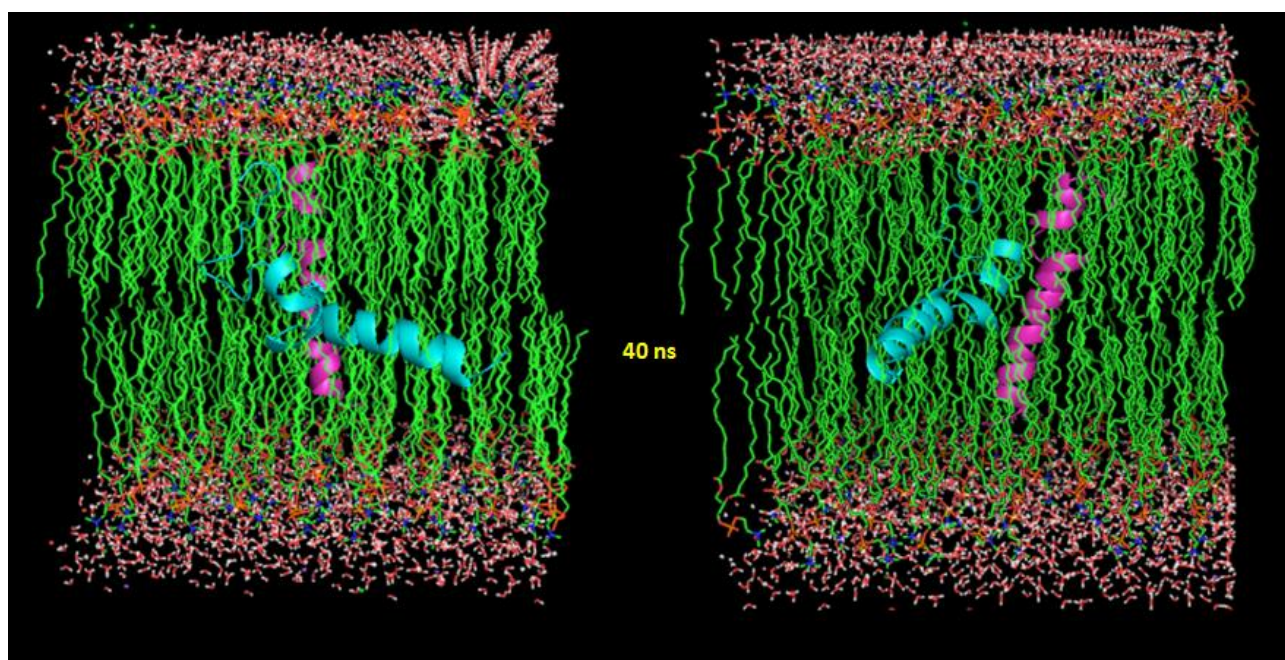
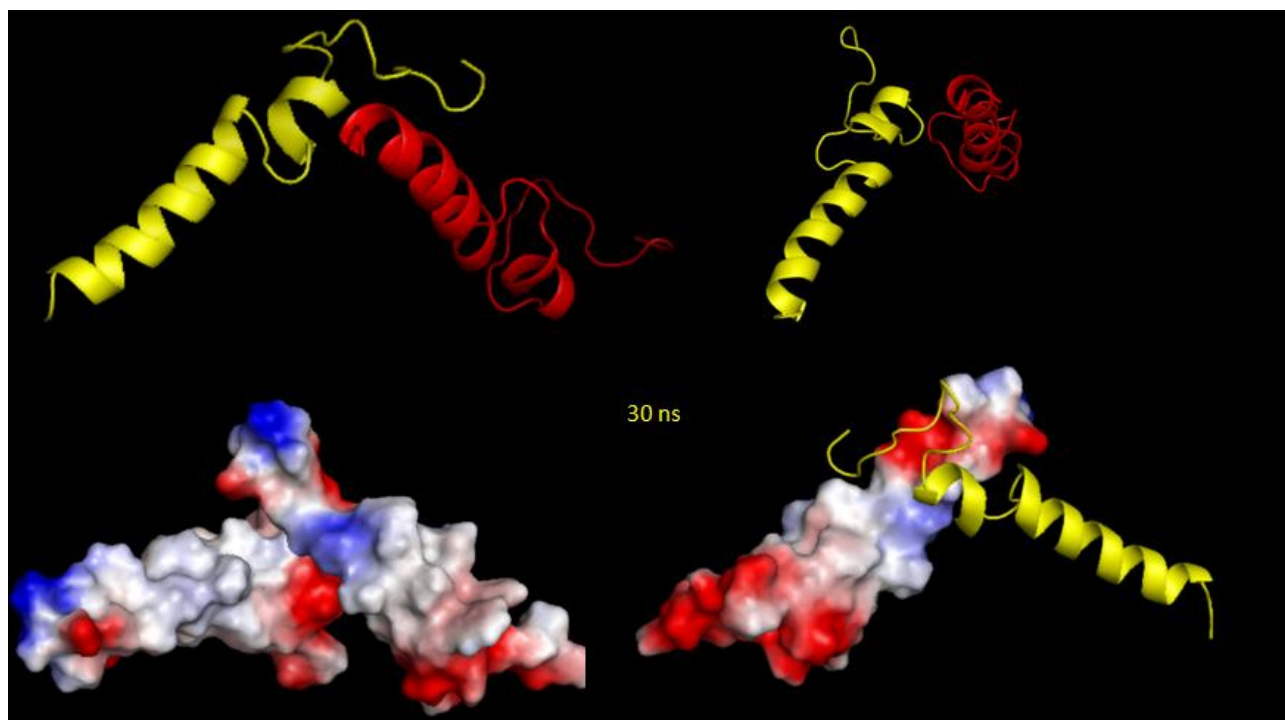


Figure 13S - The figure shows the interaction interface between the two ligands represented by two molecules of ORF7b-2. The model shows that the dominant interaction is implemented through two non-polar patches (see also fig.11), shielded by charged external surfaces. This structural organization exposes a dimeric surface with a widespread negative charge distribution, which does not have a good chance of being structurally favored in apolar environments.





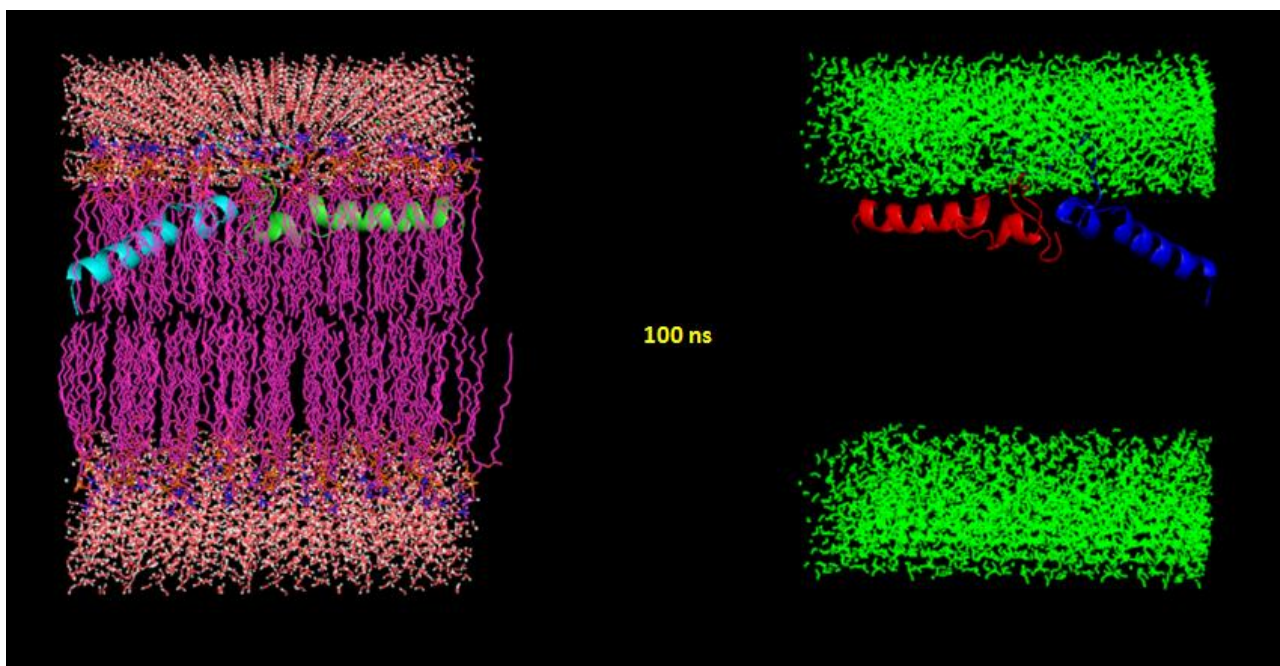


Figure 14S – The figures show some critical steps in the dimer's evolution molecular dynamics simulation down to 100 ns. The different colors or the lack of the lipid sheets (at 100 ns) have the sole purpose of better showing the details of the evolution of the structural organization of the two ORF7b-2 molecules. The organization observed at 100 ns is already present at 70 ns and remains constant. The tendency of ORF7b-2 is not to maintain a dimeric structure in the membrane, but it gradually loses structure and helical organization with a tendency to arrange their polar surfaces towards the lipid layer containing the polar heads. The structural details shown at 30 ns are interesting, because we can see how the distribution of the electrostatic surfaces dynamically generates a head-to-tail interaction of the two proteins (probably because of the interaction of the negative charge of the C-terminus with the positive one of the N-terminus). This organization suggests that it was generated by rotational and translational motions of the two proteins attracted to the polar head layer. Some alpha-helix losses are also clear. At 40 ns, one monomer attaches to the polar layer of the lipids while the other seems to continue its rotation and translation until it reaches a condition of apparent equilibrium where the interaction seems to take place between the C-terminal residues. Here, both structures are arranged in an organization almost parallel to the layer of the lipid heads.

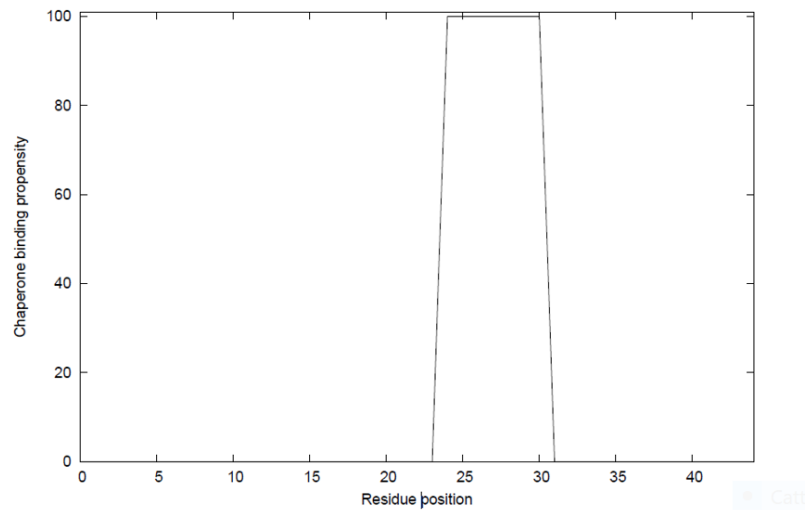


Figure 15S - The graph shows the heptad sequence of ORF7b-2 at position 24-30 (MLIIFWF), as a binding site for Hsp70 (score 23.17) calculated by the specific algorithm at Limbo.switchlab (<https://limbo.switchlab.org>), VIB Switch Laboratory, Vrije Universiteit Brussel, Brussels, Belgium.