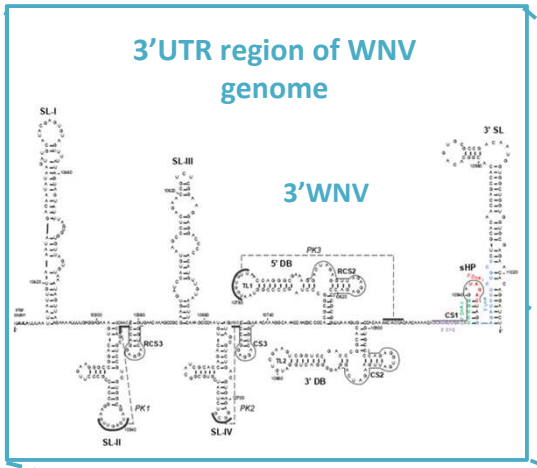
Selection of RNA aptamers targeting the 3' untranslated region of the West Nile Virus genome

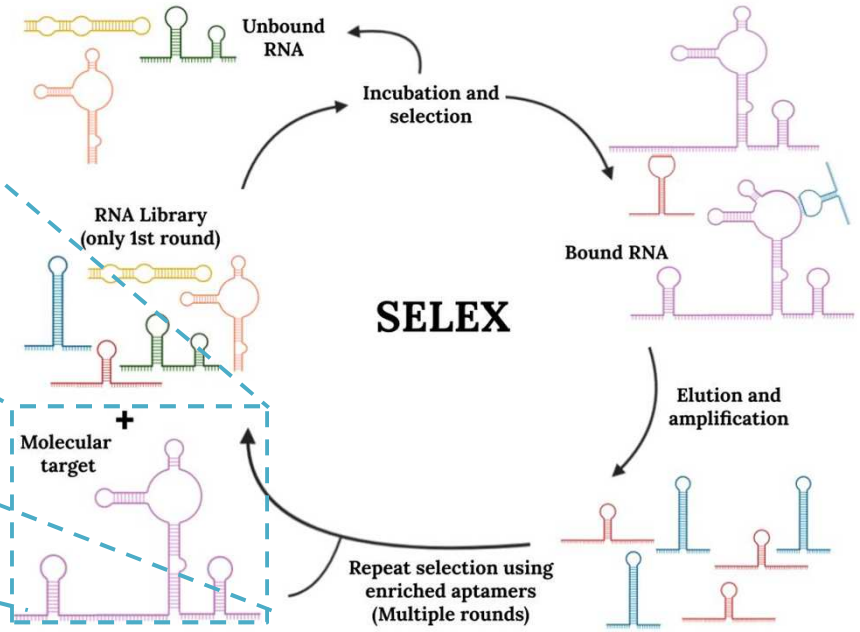
Ana Hinckley Boned, Cristina Romero-López, and Alfredo Berzal-Herranz*

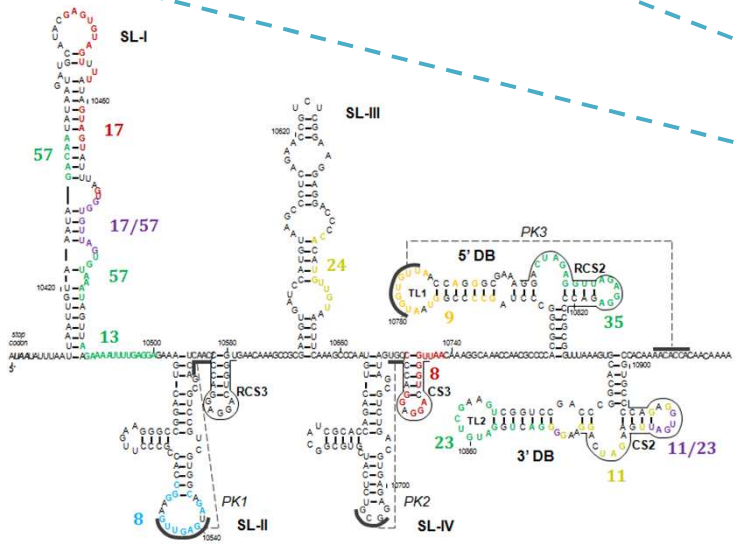
[1] Instituto de Parasitología y Biomedicina "López-Neyra", IPBLN-CSIC PTS Granada. Av del Conocimiento 17, 18016 Granada, Spain.

* Corresponding author: aberzalh@ipb.csic.es

Selection of RNA aptamers targeting the 3' untranslated region of the West Nile Virus genome







In silico **analysis of the aptamers from rounds 6 and 9 and their putative target sites**

Abstract: West Nile Virus (WNV) is a positive polarity, single-stranded RNA virus that causes West Nile fever, for which no cure has been found to date. WNV, like other RNA viruses, needs to compact all the information to complete the viral cycle into a very small genome. Beyond the information that is stored in the primary structure, the genome of RNA viruses bear functional structural domains that perform multiple essential functions for the viral cycle. In WNV, several of these functional domains are found in the 3'UTR region. Based on the importance of these functional domains, in this work, RNA aptamers have been studied as a possible therapeutic agent. Aptamers are oligonucleotides with the ability to efficiently bind to a molecule, not taking into account only the sequence of the target but also its structural motifs. In this work, various aptamers directed against the 3'UTR region of WNV, which could potentially inhibit processes of the WNV viral cycle, have been analysed and selected by *in silico* analysis. We have also studied certain characteristics of the SL-I structural element of the WNV 3'UTR, which shows a high chance of interacting with host molecules. This work will lead further studies towards the generation of antiviral aptamers against WNV and a deeper understanding of WNV interaction with the host cell.

Keywords: aptamer, functional RNA domains, RNA genome, RNA structure, RNA-RNA interactions, West Nile Virus.

Introduction- WNV

Positive polarity single-stranded RNA virus

Disease

West Nile Fever

-Outbreaks in 6 out of 7 continents.
-Symptoms: fever, neuroinvasive disease, death.
-No cure or vaccination.

Vector

***Culex* mosquitoes**

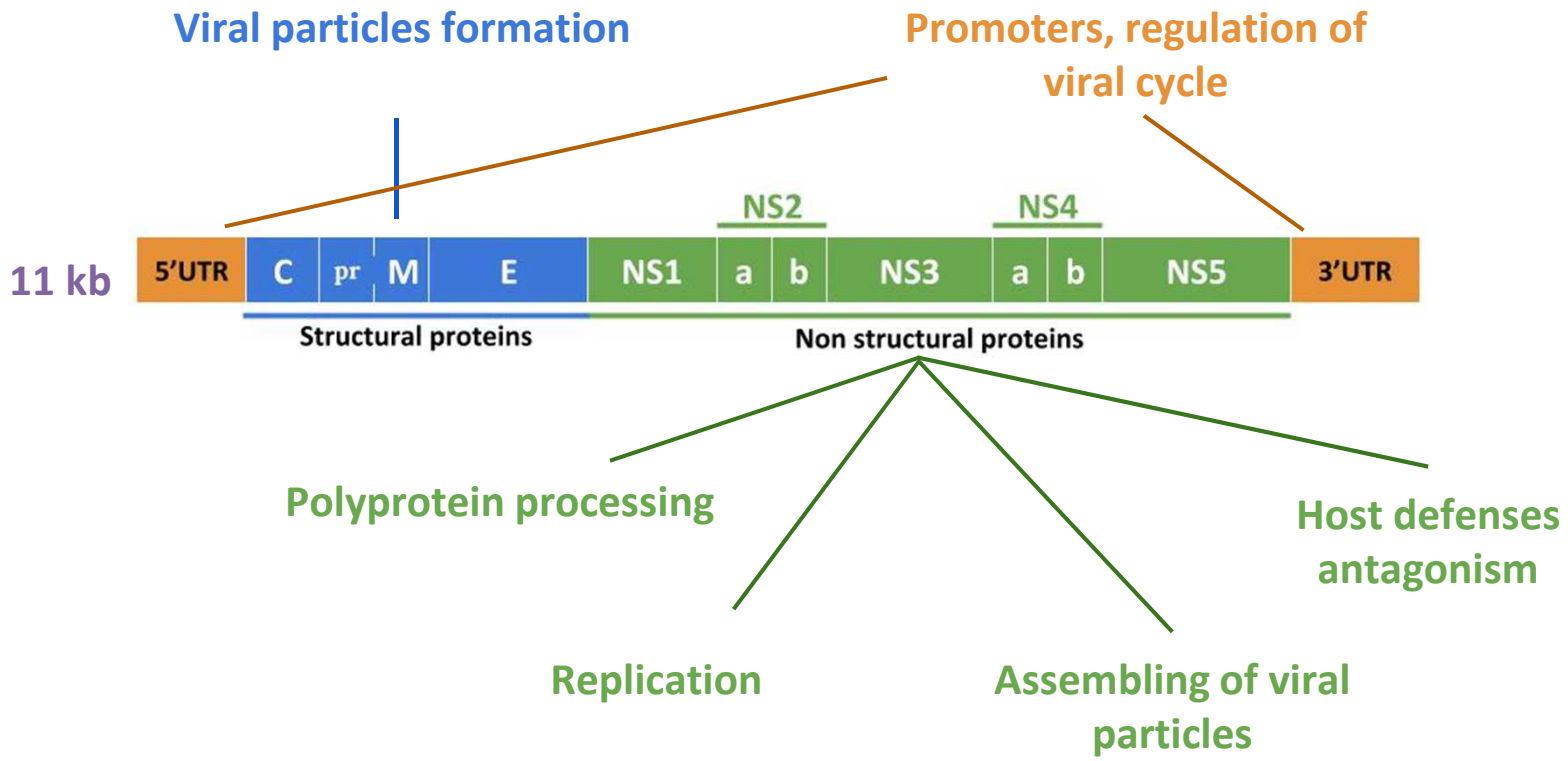
Principal transmission cycle: bird-mosquito-bird. Humans are accidental hosts.

Genus

Flavivirus

Viruses causing: dengue, Zika, yellow fever, Japanese encephalitis ...

Introduction- Genome organization of WNV



Introduction- Functional structural RNA elements of WNV genome

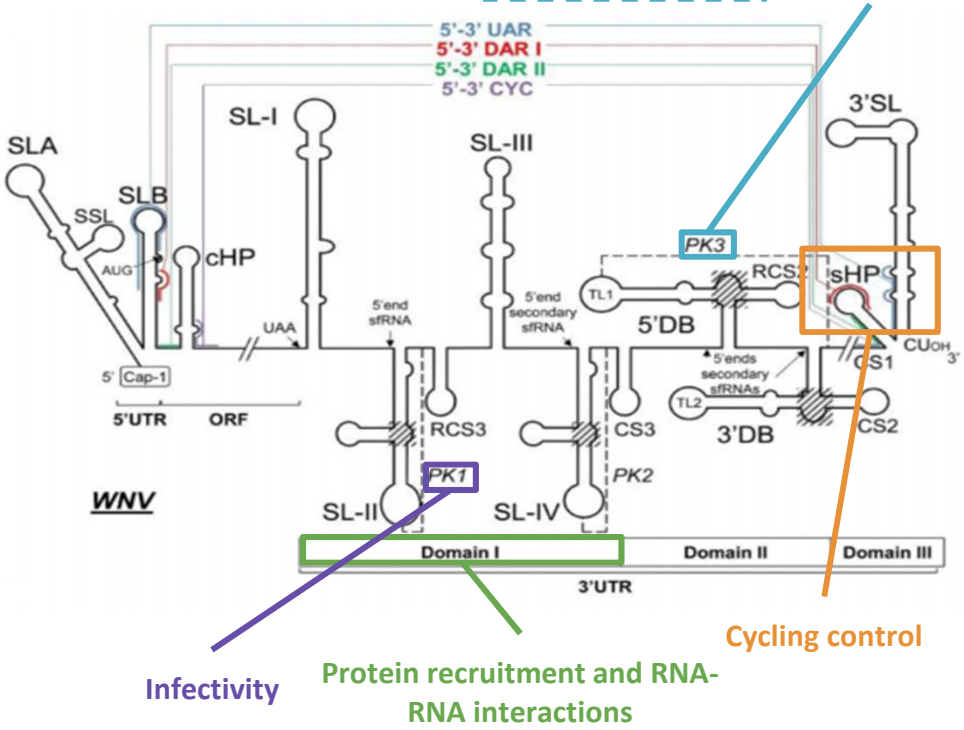
WNV → Small genome
+
Complex viral cycle
+
Need for coordination between replication and translation

Necessary compaction of information in elements superimposed on the sequence

Structural elements with fundamental functions for viral cycle regulation

Conserved on the *Flavivirus* genus

Potential therapeutic targets

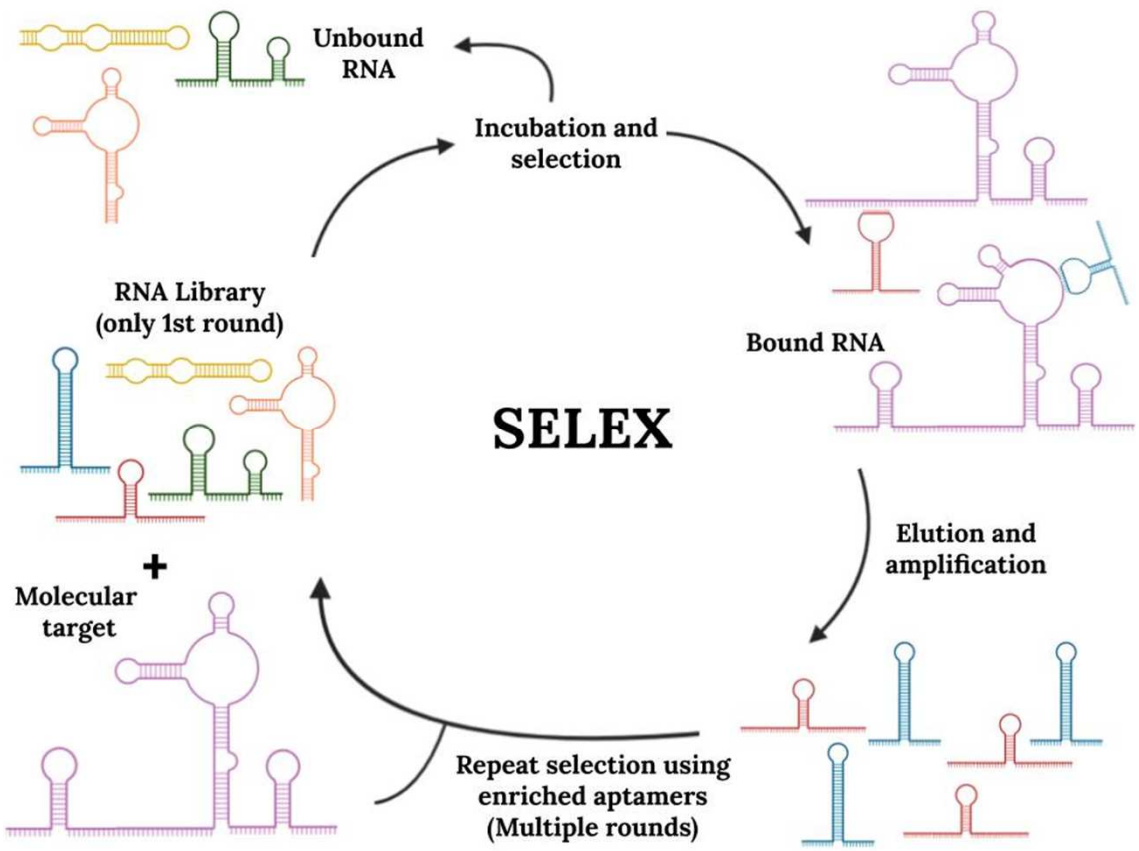


Schematic representation of the secondary structure of the 5' and 3' UTRs of the WNV RNA genome. The main structural motifs and proposed functions are depicted.

Introduction- Selection of Aptamers

Starting from a large synthetic single-stranded population of variable sequence oligonucleotides (DNA or RNA), typically ranging from 10^{12} to 10^{16} variants, molecules able to efficiently bind to a target molecule with high specificity can be selected by a SELEX procedure. They are named Aptamers

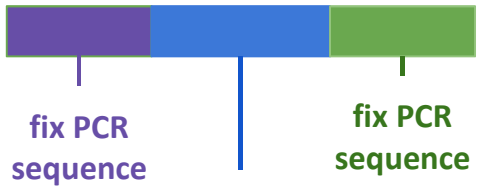
They recognize both the primary structure and the three-dimensional conformation of the target



Schematic representation of the standard procedure for the selection of aptamers (SELEX). It consists of iterative cycles of binding, selection and amplification

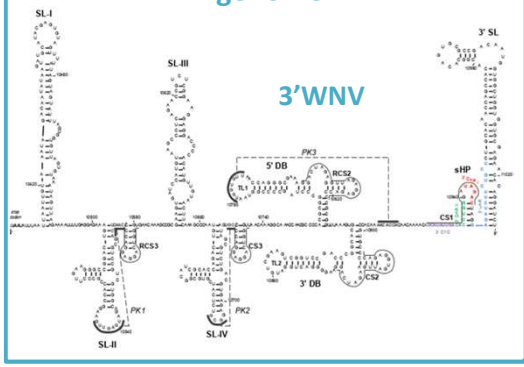
Introduction- Aptamers against WNV genomic 3'UTR

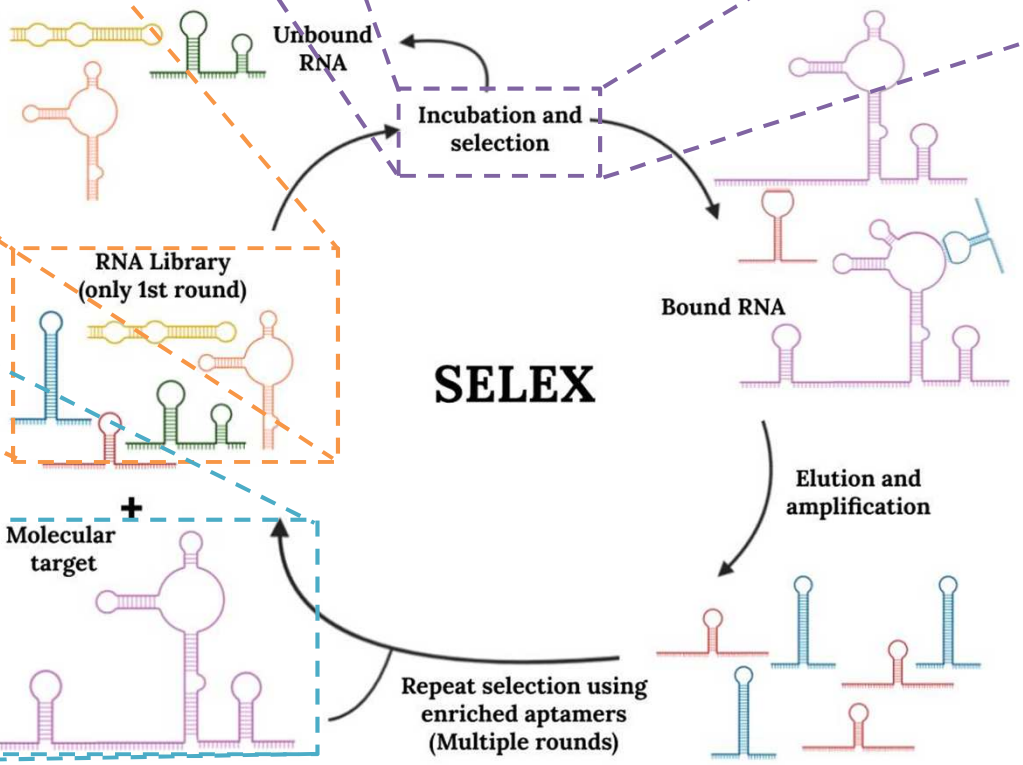
Initial population of 66 nucleotides-long RNA molecule variants



Selection conditions	[RNA Pool] : [RNA Target]	Tº	Time	[MgCl2]	[NaCl]
Cycle 5-6	10:1	37ºC	10'	1mM	50mM
Cycle 7-9	100:1	37ºC	10'	1mM	50mM

3'UTR region of WNV genome





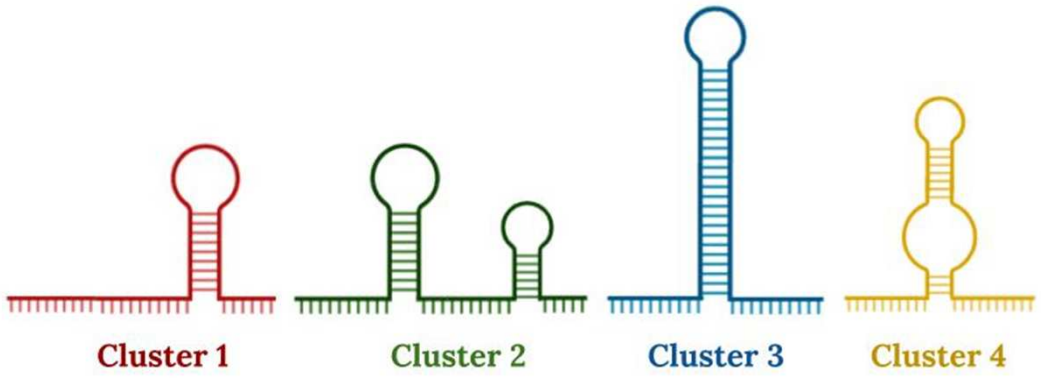
Results- Aptamers are structurally selected in for clusters

RNAstructure

Theoretical RNA aptamer secondary structure prediction

Aptamers secondary structure prediction by RNAStructure

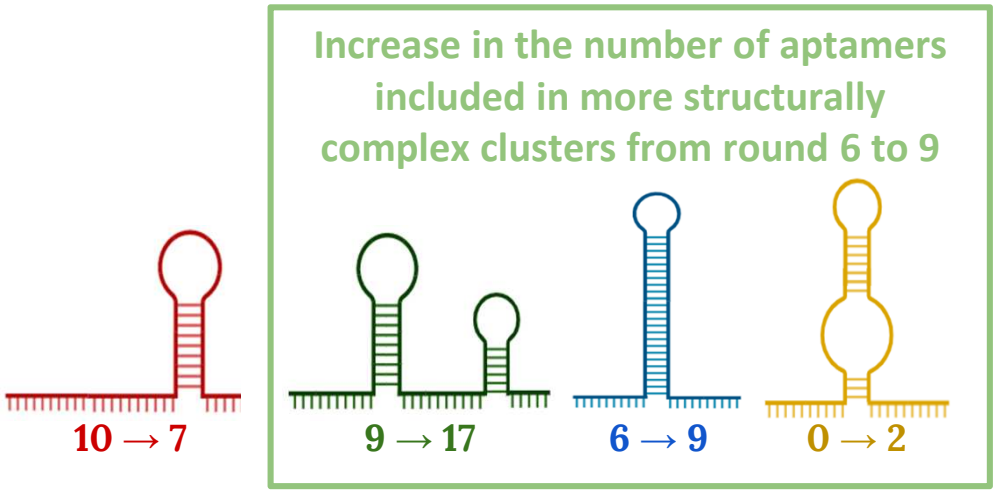
Structural clustering by visual inspection



Cluster	P6-1	P6-2	P6-3	Excluded	P9-1	P9-2	P9-3	P9-4	Excluded
Aptamers	10	9	6	11	7	17	9	2	9

Table summarizes the number of aptamers classified in each cluster. Those representatives showing more than one theoretical structure prediction were excluded of the analysis.

Results- Increase in the interaction affinity leads towards structurally complex aptamers



Number of non excluded aptamers in round 6 and (→) 9 for each cluster

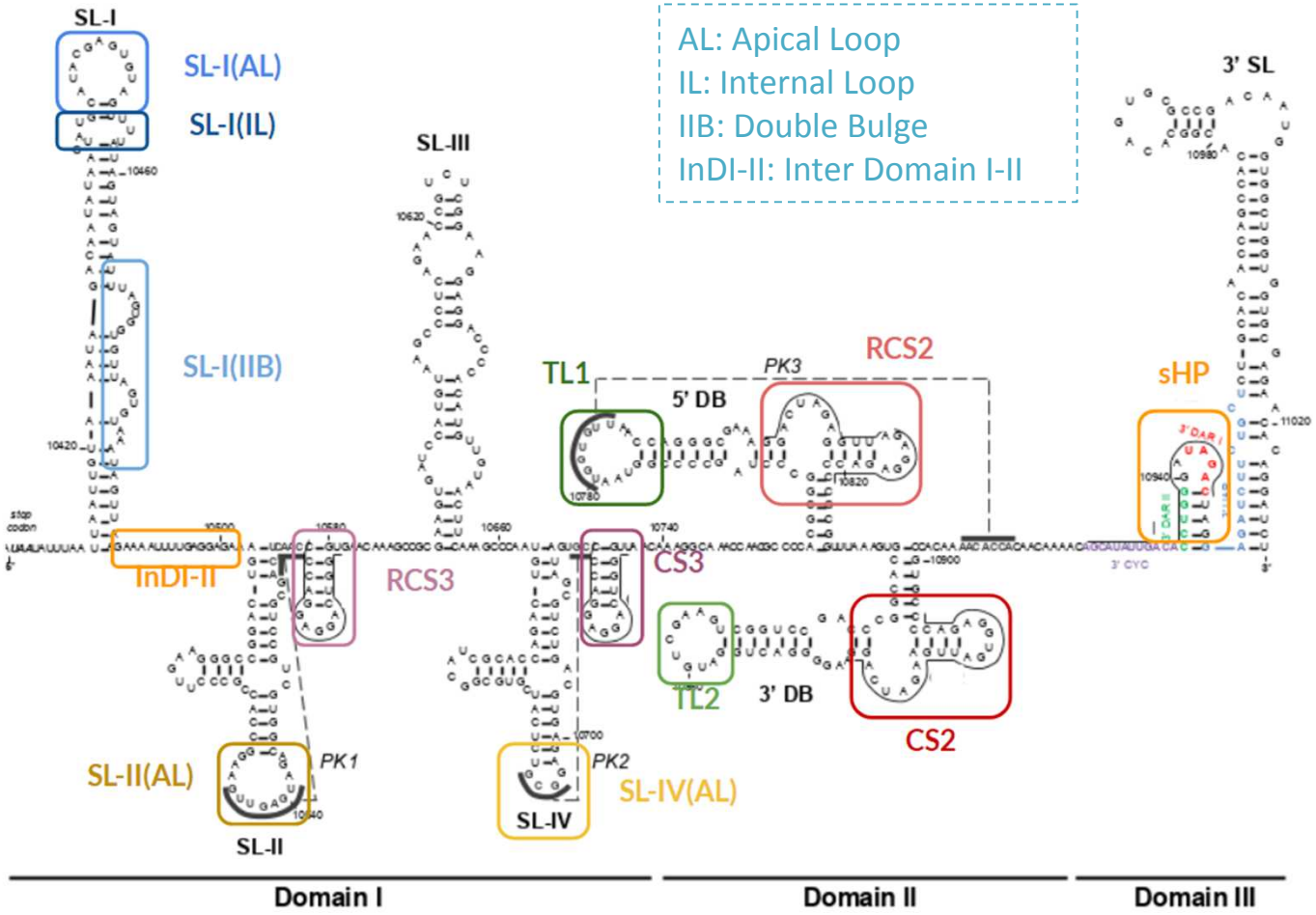
Decrease of excluded aptamers
P6: 11/36 → P9: 9/44

Fixation of secondary structures with the increase of affinity → higher structural complexity = higher affinity

Results- Sequence motifs and structural elements of interest in the WNV 3’ UTR

Structural elements and sequence motifs in the WNV 3’UTR with functional and biological interest presented as putative target sites for the selected aptamers

AL: Apical Loop
IL: Internal Loop
IIB: Double Bulge
InDI-II: Inter Domain I-II



Results- *In silico* strategies based on sequence homology predict 10 different targets

Clustal Omega

Two strategies aimed at searching for the targets of homologous sequences in the selected aptamers

Analysis through apical loop sequence homology

Multiple alignment of the apical loop sequences of each subgroup

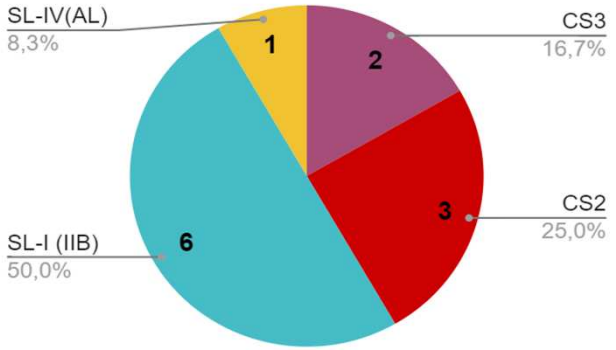
Analysis through complete aptamer sequence homology

Multiple alignment of the aptamers of each subgroup

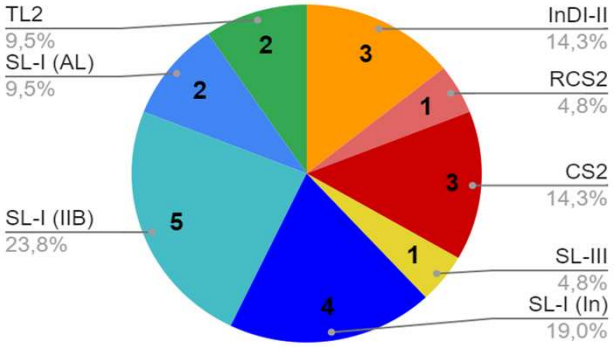
Homologous sequences

Multiple alignment of the complementary sequence from the homologous sequences with WNV 3‘ UTR

P6



P9



Number of aptamers that interact with each structural motif in each round

Long interactions and high number of targets

It does not take into account the target folding

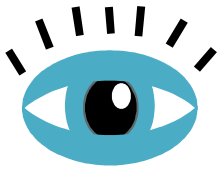
Results- RNAcofold predicts five different targets considering secondary structure

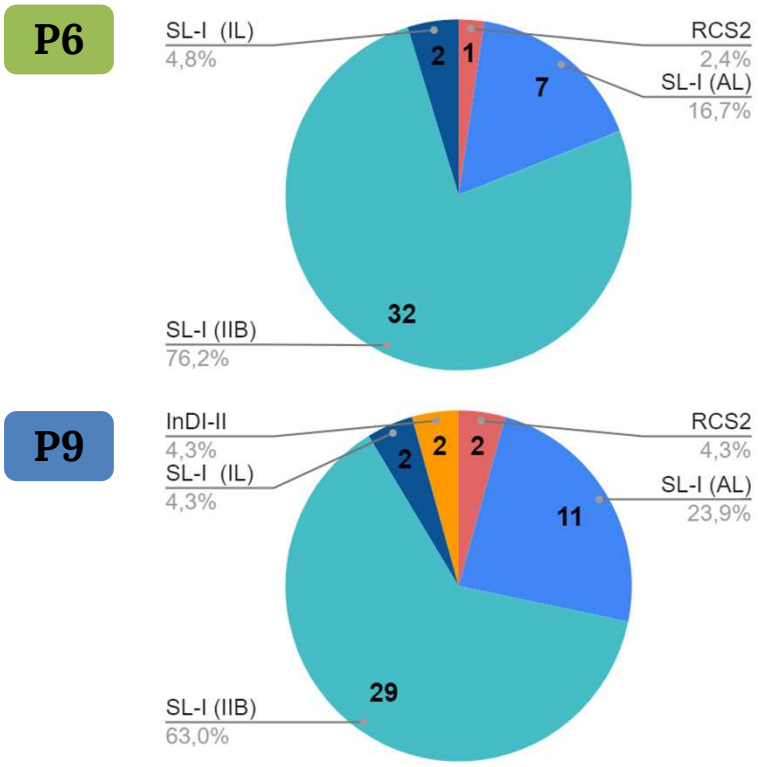
RNAcofold

Target prediction considering the complete sequence and secondary structure of the interacting molecules

Generation of partition matrixes

Visual inspection of the matrixes and obtainment of the interaction sequences in the aptamer and in WNV 3'UTR



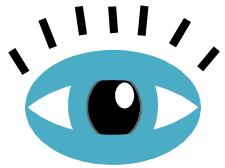


Number of aptamers that interact with each structural motif in each round

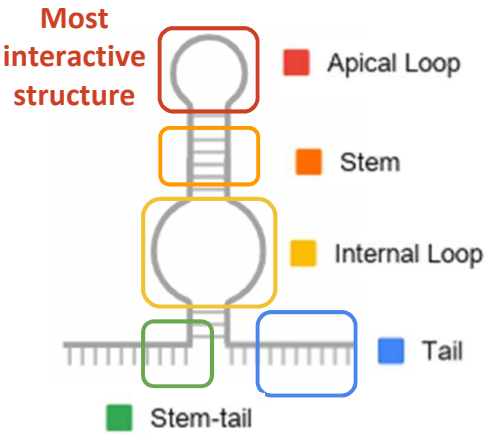
Target prediction for most of the selected aptamers

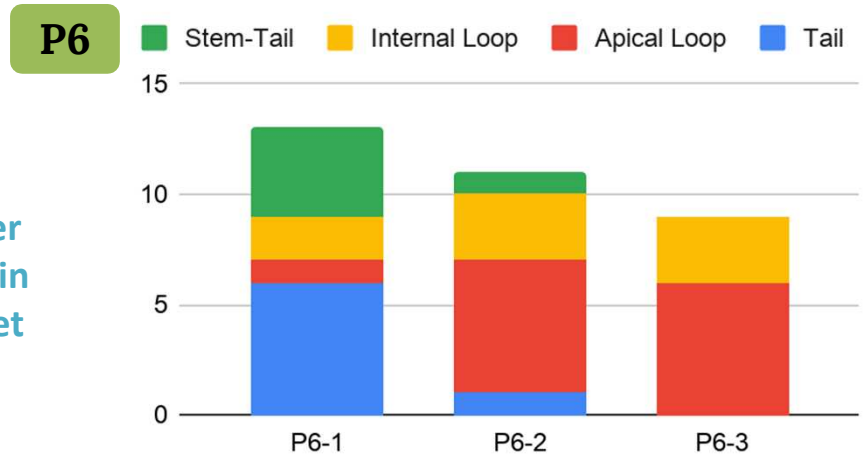
Smaller target diversity

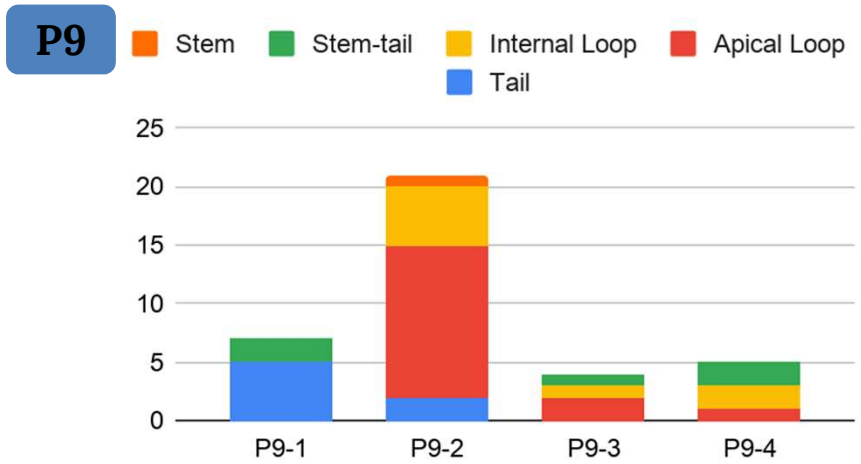
Results- The aptamer population presents a selection towards the interaction through apical and internal loops



Visual analysis of the aptamer structural elements involved in the interaction with the target





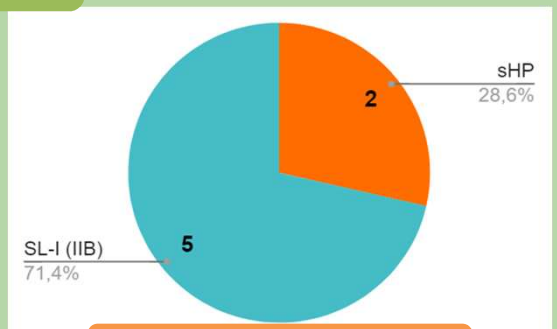


Cluster 1, which interacts mostly through sequences at the tail of the structure, is reduced from round 6 to 9; while cluster 2, which interacts through the loops, increases significantly

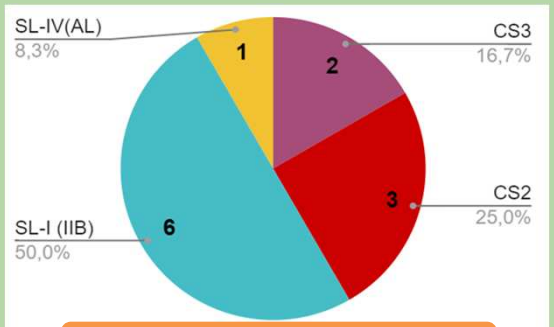
Selection towards more interactive clusters

Results- SL-I(IIB) target dominance and round 9 target diversity

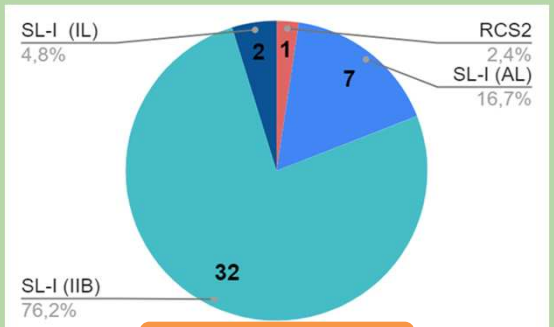
P6



Apical loop analysis

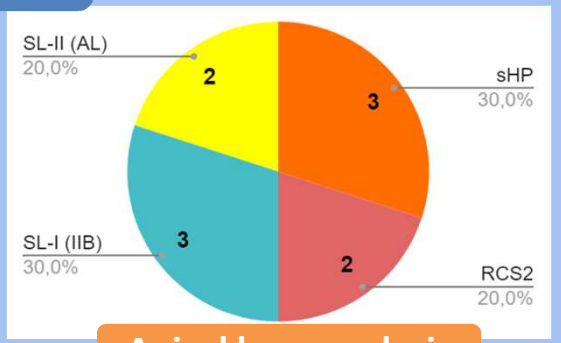


Sequence homology

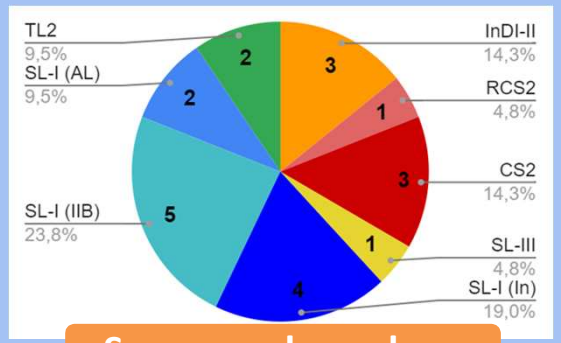


RNAcofold

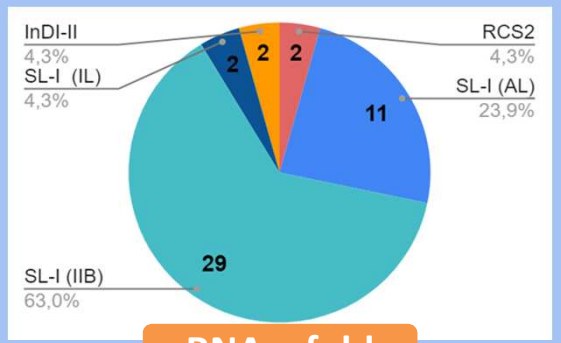
P9



Apical loop analysis



Sequence homology



RNAcofold

Dominance of the interaction with SL-I, specially SL-I (IIB) compared to the other targets

Higher prevalence of SL-I (IIB) with RNAcofold predictions, meaning that not only sequences but also structural motifs have been selected

Target diversification between rounds 6 and 9 ¿SL-I saturation?

Results- There is an increase in sequence conservation from round 6 to 9

Compseq

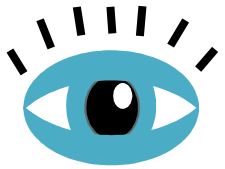
It renders the number of repetitions of a given sequence motif in an aptamer

Increase in sequence conservation in round 9, represented as repeated sequences

	P6	P9	P9 without repseq*
Hexanucleotides repeated 6 times or more in the aptamer population	10	29	15
Most repeated hexanucleotide	ACACUA (18 repeats)	CACUAA (16 repeats)	CACUAA (15 repeats)

*In round 9 there is one whole aptamer sequence repeated 5 times and 3 whole aptamer sequences repeated twice.

Results- Sequences are conserved because of their interaction with WNV 3'UTR

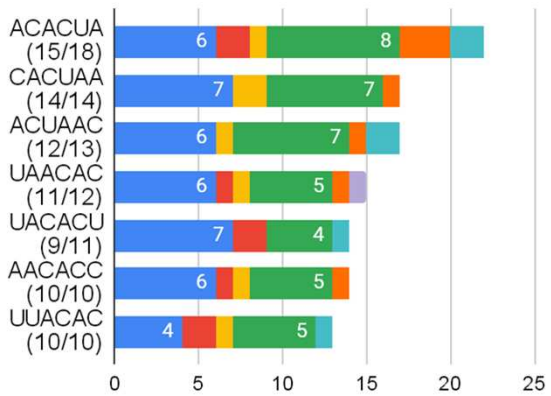


Visual analysis of the location of hexanucleotide motifs within the aptamer

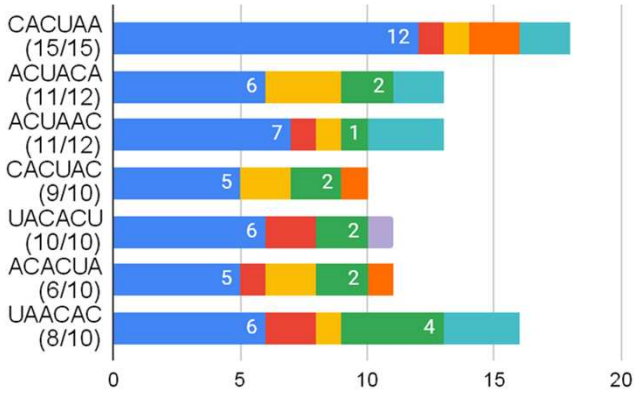
Structural element

3'WNV interaction

P6



P9



Most repeated hexanucleotides and their hexanucleotide motif location. Between brackets, the amount of hexanucleotides from the total of repeats that are part of the sequence that interacts with WNV 3'UTR.

The conserved sequences interact, at least in part, with WNV 3'UTR

Results- The putative target of the conserved sequences is preferently SL-I(IIB)

Clustal Omega

Multiple alignment software

Analysis of the putative targets in WNV 3'UTR of the conserved hexanucleotide sequences

Analysis of the number of aptamers in each round that don't have any conserved hexanucleotide

Round	Target interactions	Apt without hexants
6	SL-I(IIB) (7/7), CS2 (2), SL-I (LA) (1)	7
9	SL-I(IIB) (5/6), CS2(2), SL-I (LA) (2), no target (1)	13

Supports the hypothesis of a general evolution of aptamers towards SL-I

Supports the hypothesis of target diversification between rounds 6 and 9

Results- The putative target of the conserved sequences is preferently SL-I(IIB)

Search for longer conserved sequences and their putative targets through analysis of the adjoining nucleotides of the conserved hexanucleotides

Clustal Omega

Repeat	Round 6	Round 9	Target
CACUAACACC	6 repeats	2 repeats	SL-I
UUACACUA	7 repeats	0 repeats	SL-I
CACUACAC	3 repeats	5 repeats	SL-I
ACUACACUCG	1 repeats	4 repeats	SL-I

Supports the hypothesis of a general evolution of aptamers towards SL-I

Through this process a new target was found: TL-1

Results- Most represented targets present high U% and low C%

Compseq

%nts

Infoseq

%GC

Sequence	%GC	A	C	G	U	Sequence	%GC	A	C	G	U
3'UTR	28,57	29,1	23,3	28,3	19,4	Apt P9 (re)	52,36	28	37,1	15,3	19,6
Apt P6	46,59	**30,9**	**34,4**	**12,2**	22,5	**Apt P9**	51,48	**27,4**	**39,7**	**12,3**	20,6
Apt P6(-E)	45,96	32	34,8	11,1	22,1	Apt P9 (-E)	52,22	27,3	39,9	12,3	20,5
SL-I (IIB)	28,57	23,8	**0**	28,6	**47,6**	**SL-III**	51,61	27,4	**27,4**	24,2	21
SL-I (LA)	46,15	30,8	15,4	30,8	23,1	**SL-IV**	58,82	19,6	**27,4**	31,4	21,6
SL-I (LI)	11,1	22,2	**0**	11,1	**67**	**TL1**	35,71	28,6	**7,1**	28,6	**35,7**
SL-I (total)	25,97	35,8	**3,7**	21	**39,5**	RCS2	56,52	30,4	17,4	39,1	13,1
InDI-II	31,25	43,8	**0**	31,2	**25**	**TL2**	45,45	27,3	**9,1**	36,3	**27,3**
SL-II (LA)	46,67	33,3	6,7	40	20	sHP	56,25	25	18,7	37,5	18,7
RCS3	73,33	18,7	25	43,7	12,5	**3'SL**	58,97	24,3	**29,5**	29,5	16,7

Most interesting data presented in bold letters.

The targets with higher number of predictions present high U% and low C%, which concordates with the percentages in the aptamers if we consider the G-U pairs

Results- Higher GC content in the sequences that interact with SL-I compared to the adjoining nucleotides

GC pairs make structures more stable because of their triple hydrogen bond and their stacking → Analysis of the GC content in the sequences that interact with WNV 3'UTR putative targets

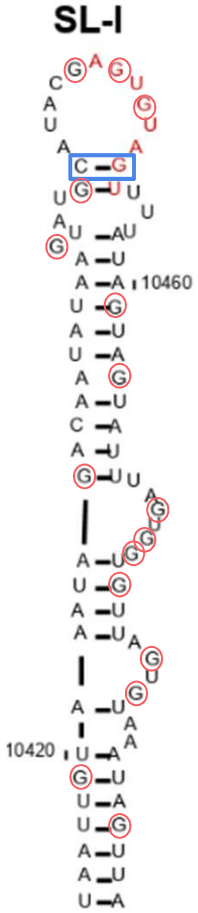
Comparison of the %GC of the aptamer interacting (int) sequence with the %GC of the whole aptamer

Infoseq

Comparison of the %GC of the target site with the %GC of the 10 adjoining nucleotides (a.n.)

Software	Target	Media % int	Media % a.n.	Higher %GC in int
RNAcofold	SL-I	**38,35**	19,77	63/67
RNAcofold	Other targets	47	**47,24**	2/5
RNAcofold	Aptamers	34,79	**49,03**	-
Clustal Omega	SL-I	**36,56**	28,43	13/16
Clustal Omega	Other targets	40,42	**55,5**	0/10
Clustal Omega	Aptamers	35,18	**49,03**	-

Results- G residues make SL-I domain the preferred target site for the selected aptamers



High G% and really low C% in SL-I → Analysis of the Gs in the SL-I structure reveal that they are not forming GC pairs because they are located in loops, bulges or G-U pairs

Guanines size

Guanine in loops and bulges → Exclusion of bases towards the solvent

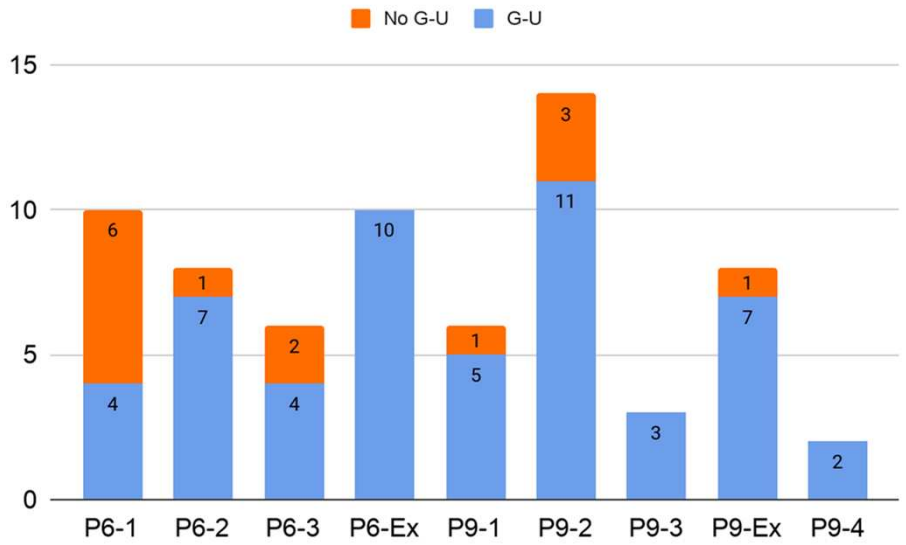
Guanine stacking → Minor groove size variation

Improved molecular intercalation

Missing C and high frequency of G in loops, bulges or G-U pairs→ unique characteristics of SL-I = preferred putative target of the selected aptamers

Results- G-U pairs are highly present in the interactions between aptamers and WNV 3'UTR

After observing several G-U pairs in the SL-I structure, we analysed the G-U pairs formed in the interactions between aptamers and WNV 3'UTR

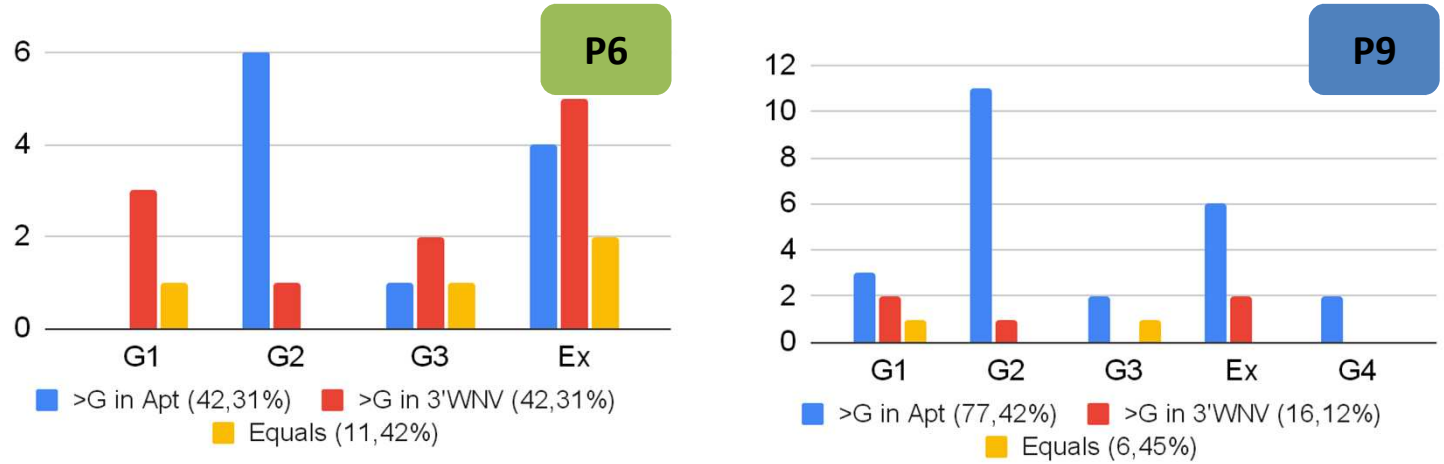


High frequency of G-U pairs in interactions, especially for those aptamers derived from round 9

Results- G-U pair directionality facilitates target-aptamer interaction

G-U pairs are non isosteric to U-G pairs, they generate different torsions of the helix, yielding differences in the minor groove size

Analysis of the location of the G from the G-U pair in either the aptamer or the target



G nucleotide location from the G-U pair in aptamer or in WNV 3'UTR

The selection of a specific directionality of the G-U pair and consequent minor groove size variation seems to be facilitating target-aptamer interaction

Results- Criteria for the selection of aptamers for biochemical analysis

Selection criteria for the aptamers that would be further studied *in vitro*:

1-To obtain a group of aptamers that together, have the maximum number of WNV 3'UTR targets with biological interest predicted.

2-If there are several predictions for the same target, select those aptamers, which yields interactions with lower ΔG and / or greater number of interacting nucleotides.

3-If the data were similar between rounds, aptamers from round 9 were selected because of their theoretical higher affinity.

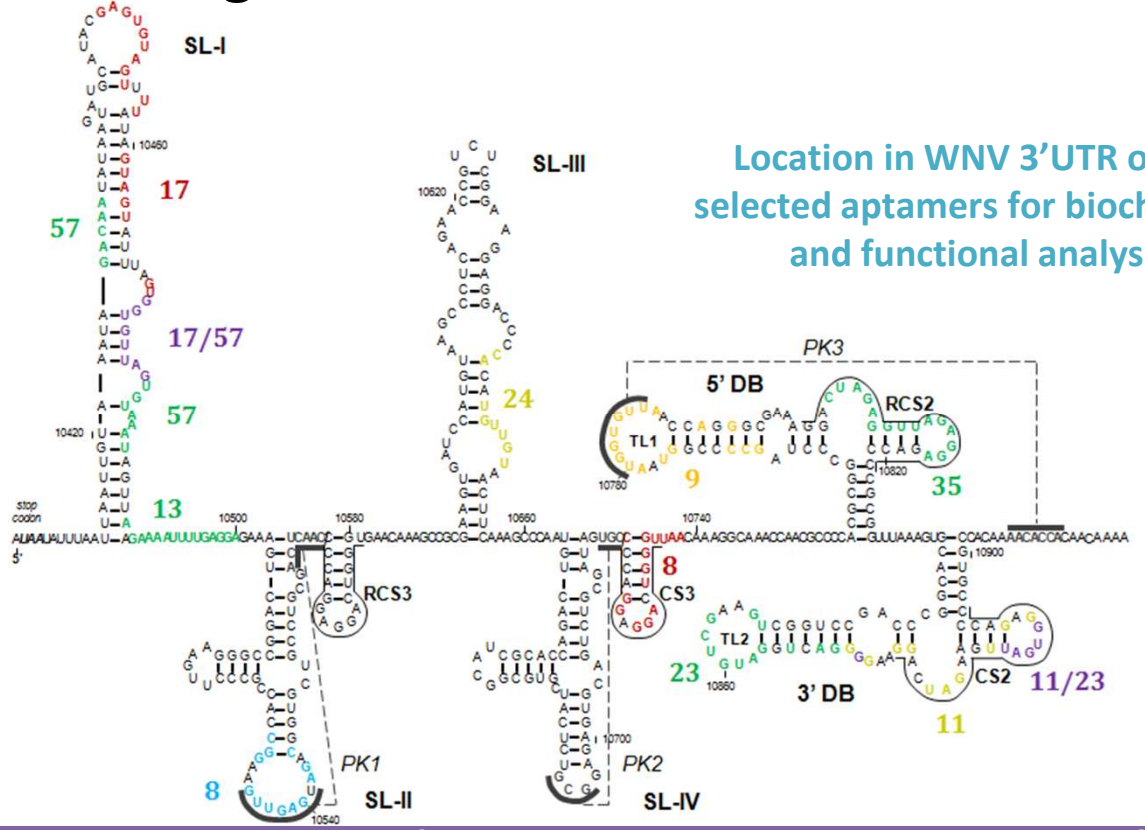
4- Conserved hexanucleotides have SL-I and CS2 as targets. Search for aptamers lacking these hexanucleotide sequence motifs.

Results- Selected aptamers for biochemical analysis

Round	Group	Apt	Apt struct	ΔG	Rep hexant	3'WNV putative targets
6	1	17	Tail	-17,87	2	**[SL-I (LA,IIB)(10)]**
	1	8	Tail	-	0	**CS3 (8)** [SL-I (IIB)(12)]
	2	57	Apical loop	-19,05	7	**[SL-I (IIB)(15)]**
9	2	35	Stem, AL	-13,37	0	**[RCS2(15)]**
	2	23	Apical loop	-	2	**TL2-CS2 (8)** [SL-I (LA, IIB)(11)]
	2	13	Apical loop	-9,8	0	**[InDI-II (14)]**
	3	8[5]	Apical loop	-	0	**SL-II(LA) (6)**
	4	11	IL, AL, Stem	-	2	**CS2 (9)** [SL-I (IIB)(14)]
	4	24	Internal loop	-	0	**SL-III (6)** [SL-I (LA, IIB)(11)]
	Ex	9	Stem-AL	-	3	**TL-1 (11)** [SL-1 (IIB)(12)]

Apt struct, structure of the aptamer that interacts with WNV 3'UTR; rep hexant, number of most repeated hexanucleotides in each aptamer. [] Structures predicted by RNAcofold, () number of nucleotides that interact.

Results- The selected aptamers cover most of the biologically interesting structural motifs of WNV 3'UTR



Location in WNV 3'UTR of the selected aptamers for biochemical and functional analysis

The selected aptamers interact with most of the biologically interesting structural motifs of WNV 3'UTR and may permit the generation of antiviral agents. They also have potential as molecular tools for studying the functions of different structural motifs for a deeper understanding of *Flavivirus* replication and infectious cycles

Conclusions

-The aptamer population shows an evolution throughout the SELEX process, both at the structural complexity level and at the chosen target. Thus, the application of restrictive conditions has promoted the isolation of aptamers against structural elements of the WNV 3'UTR distinct from the SL-I domain.

-The structural element SL-I presents some structural characteristics that make it a highly disposable domain to interact. This supports the interest of a further study of such structural features and their contribution to the completion of the viral cycle.

- 10 aptamers, which theoretically recognize 9 of the biologically relevant structural elements in WNV 3'UTR, have been selected for biochemical analysis.

-*In silico* analysis of the RNA structure and interaction has provided useful information that would reduce the cost of analysis of the whole aptamer population *in vitro* and also given interesting data about structural preferences for RNA-RNA interactions.

Acknowledgments

This work has been funded by the Spanish *Agencia Estatal de Investigación*:
PID2019-104018RB-100/ AEI /10.13039/501100011033