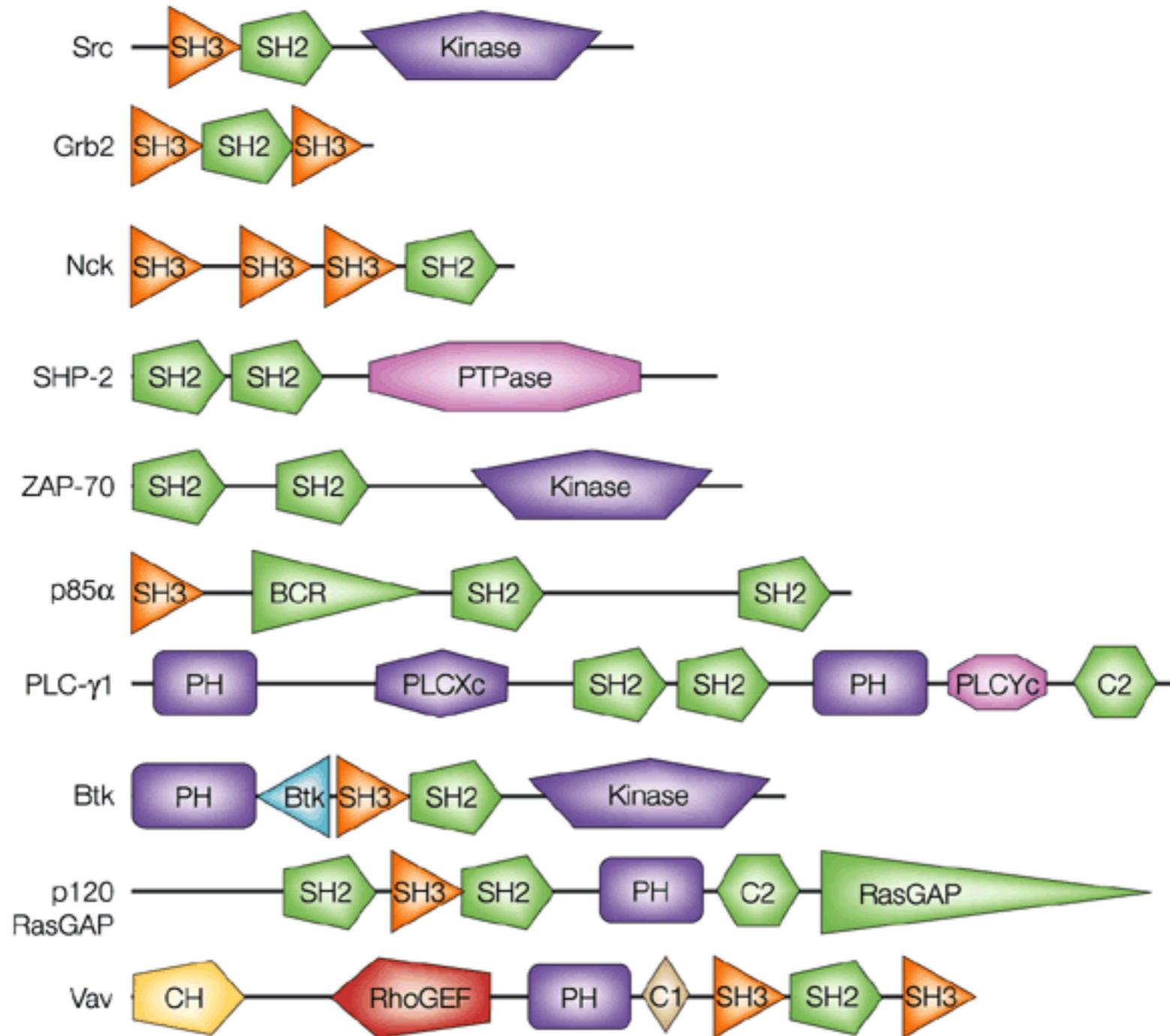
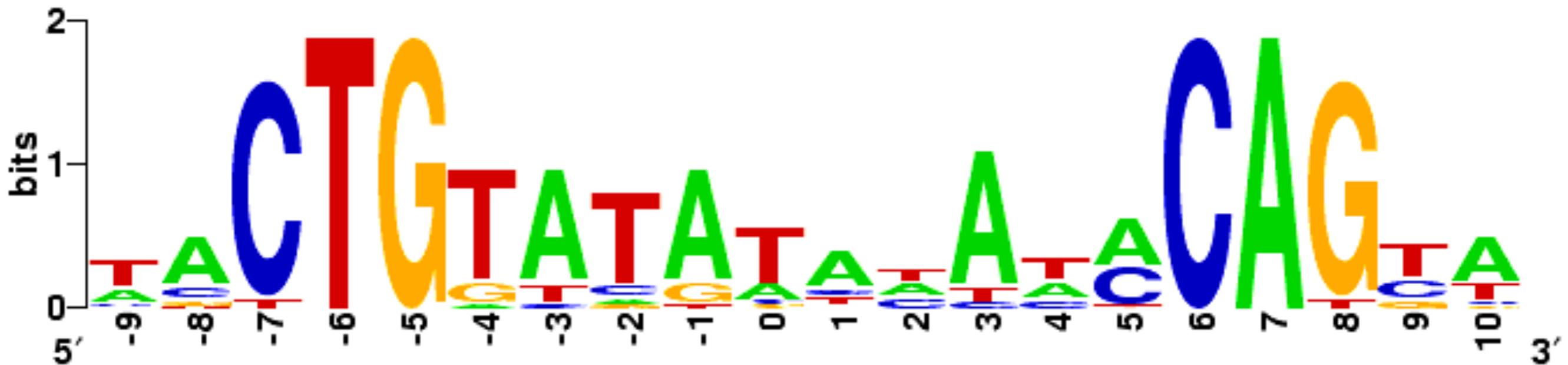


# Domain & Motif Discovery



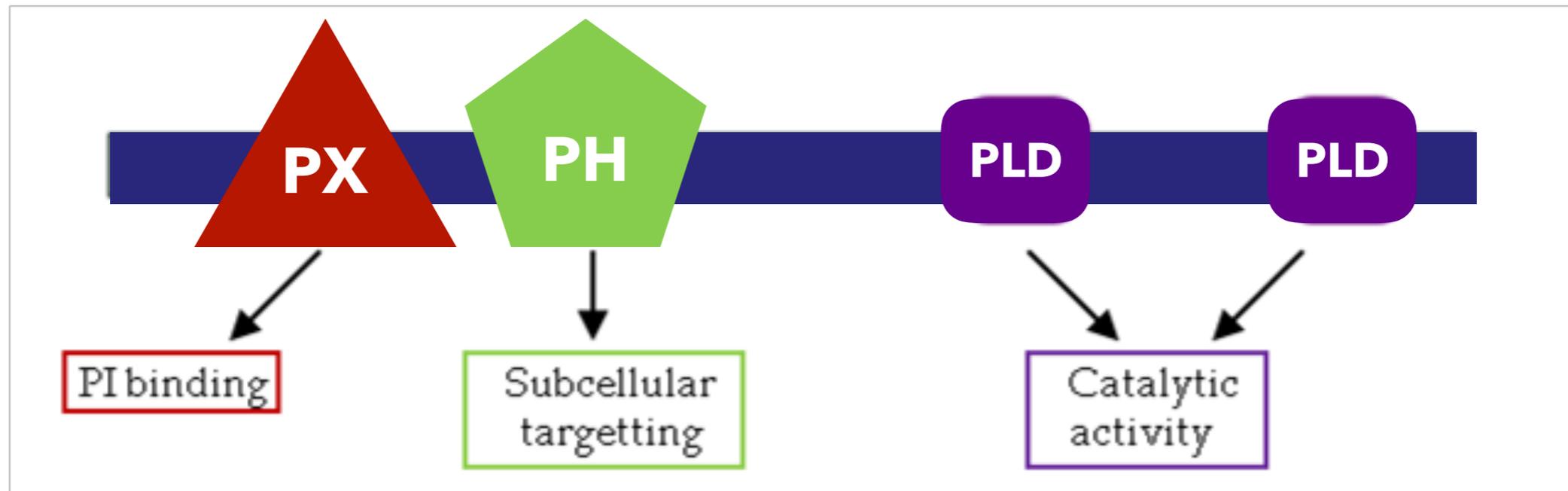
**Laura Williams & Elizabeth Zanetakos**

# What are protein motifs?



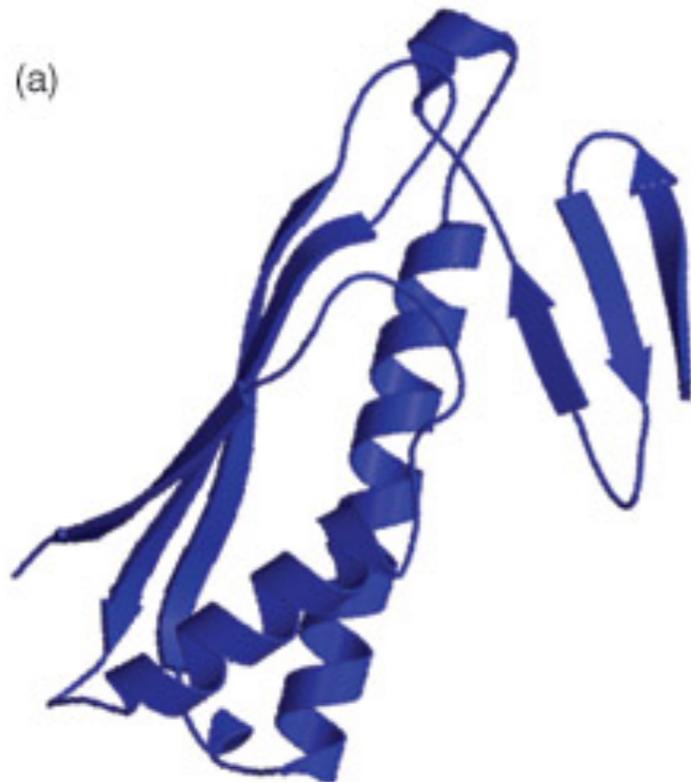
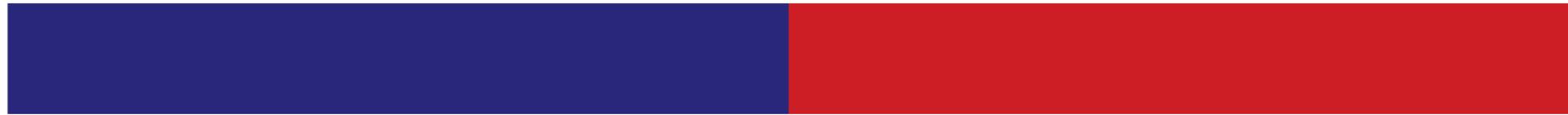
**group of secondary protein structures that contribute to the overall function of the protein**

# What are protein domains?



**a functional or structural unit of a protein that contributes to the overall function**

# Difference between protein domain and motif?

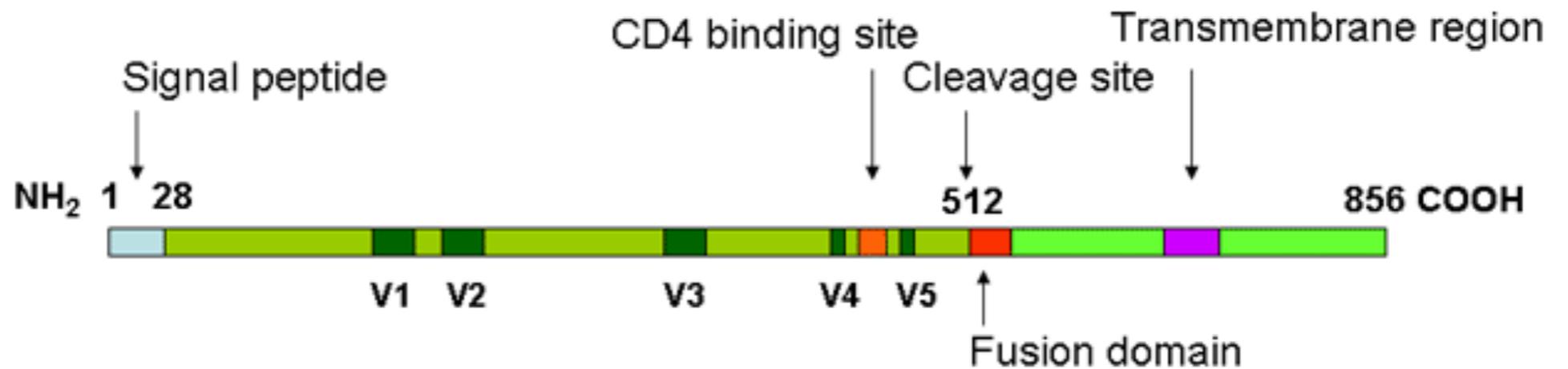


**vs**

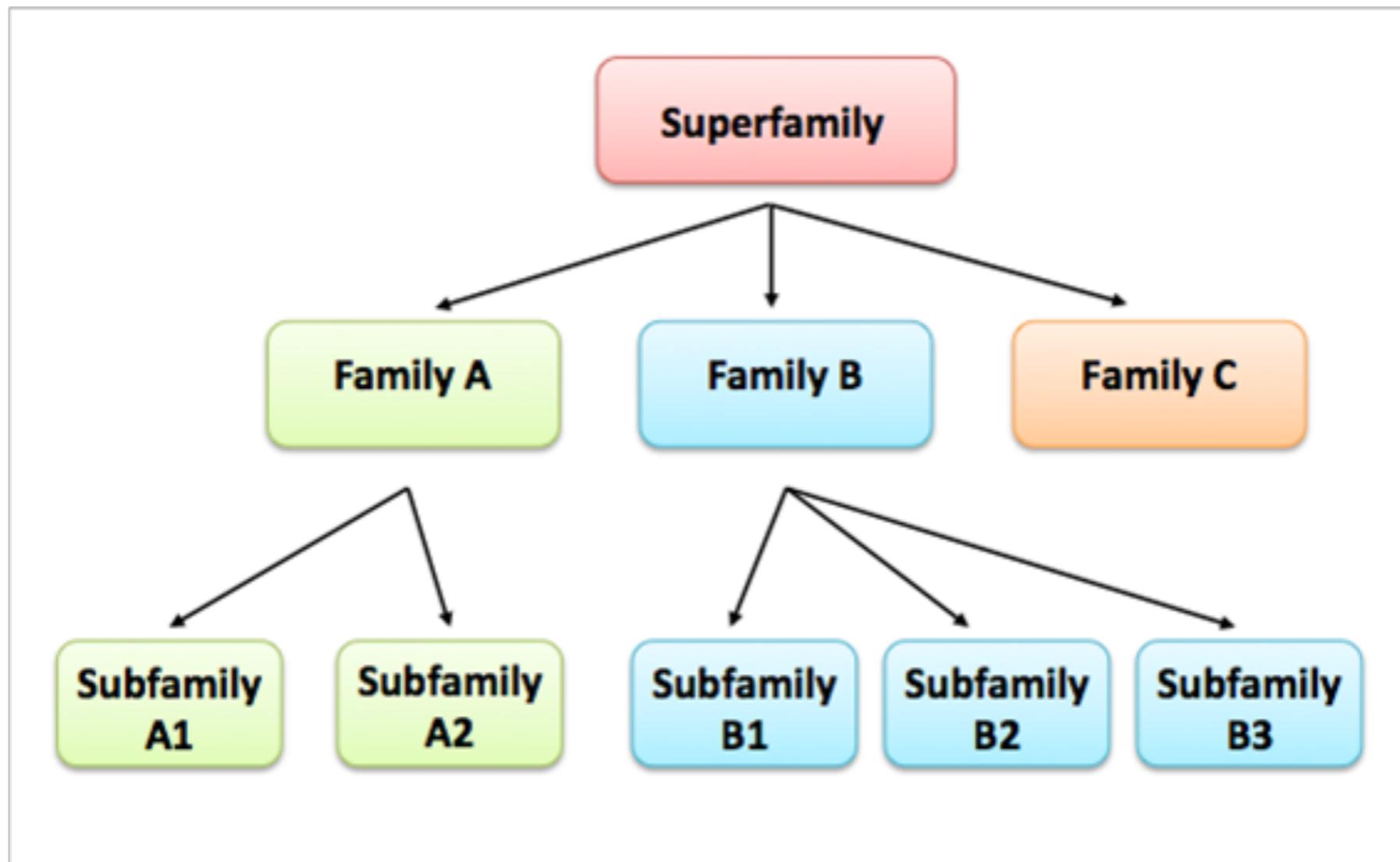


**domains form and function independently of the larger protein**

# What are the function of protein domains?

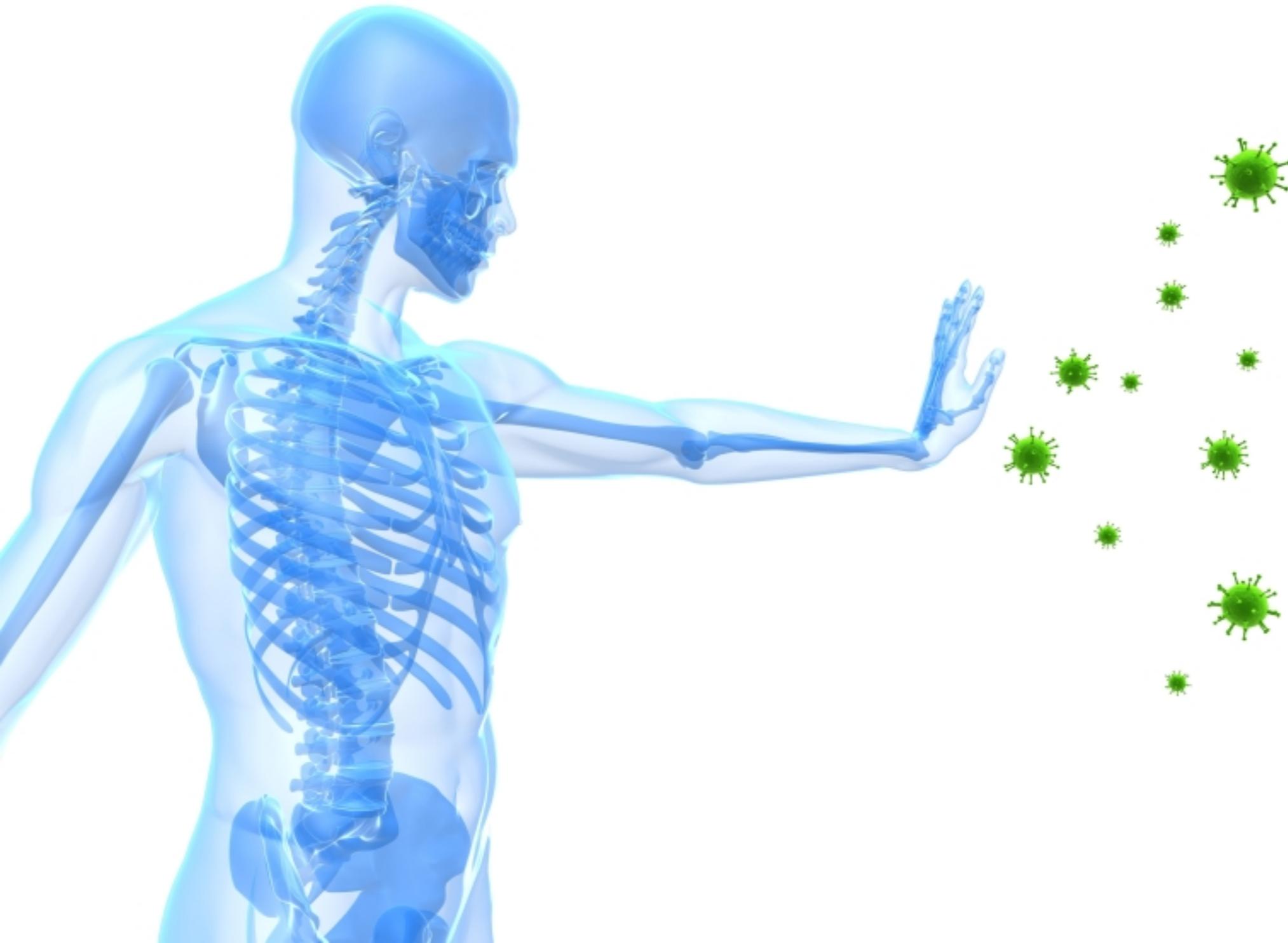


# Why do domains matter?

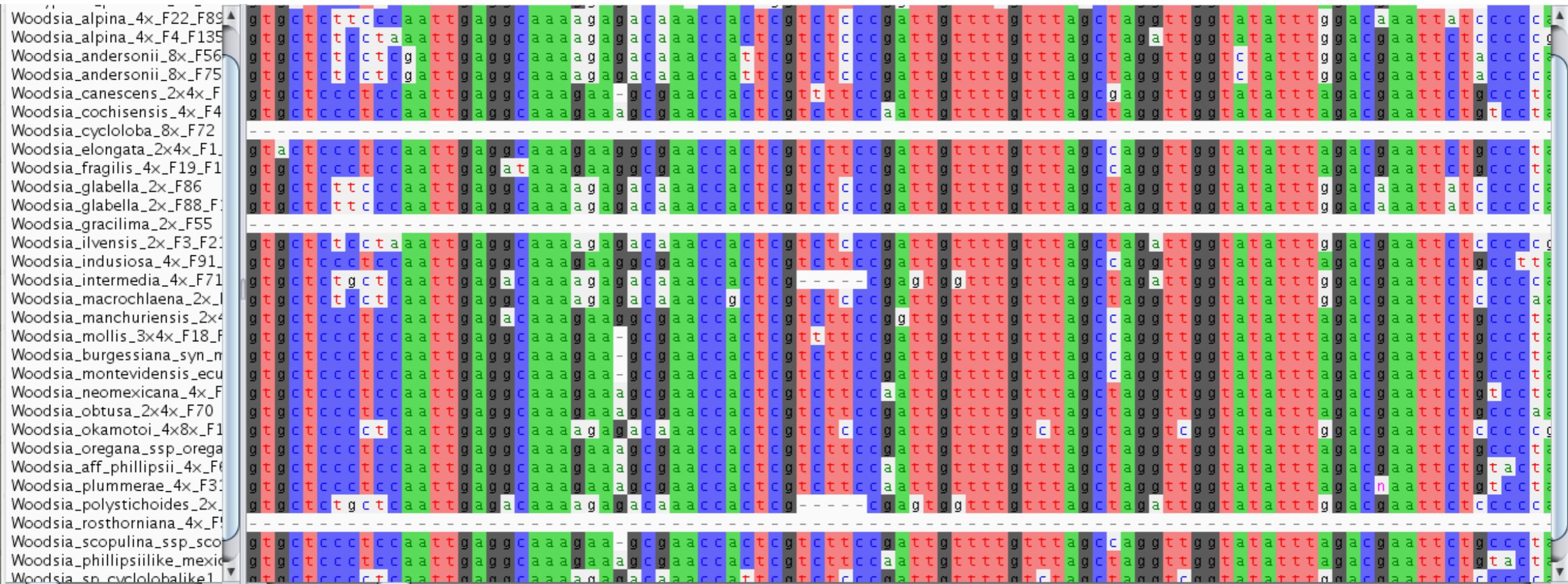


**serve as a unit of protein classification**

# Why is it important to classify proteins?

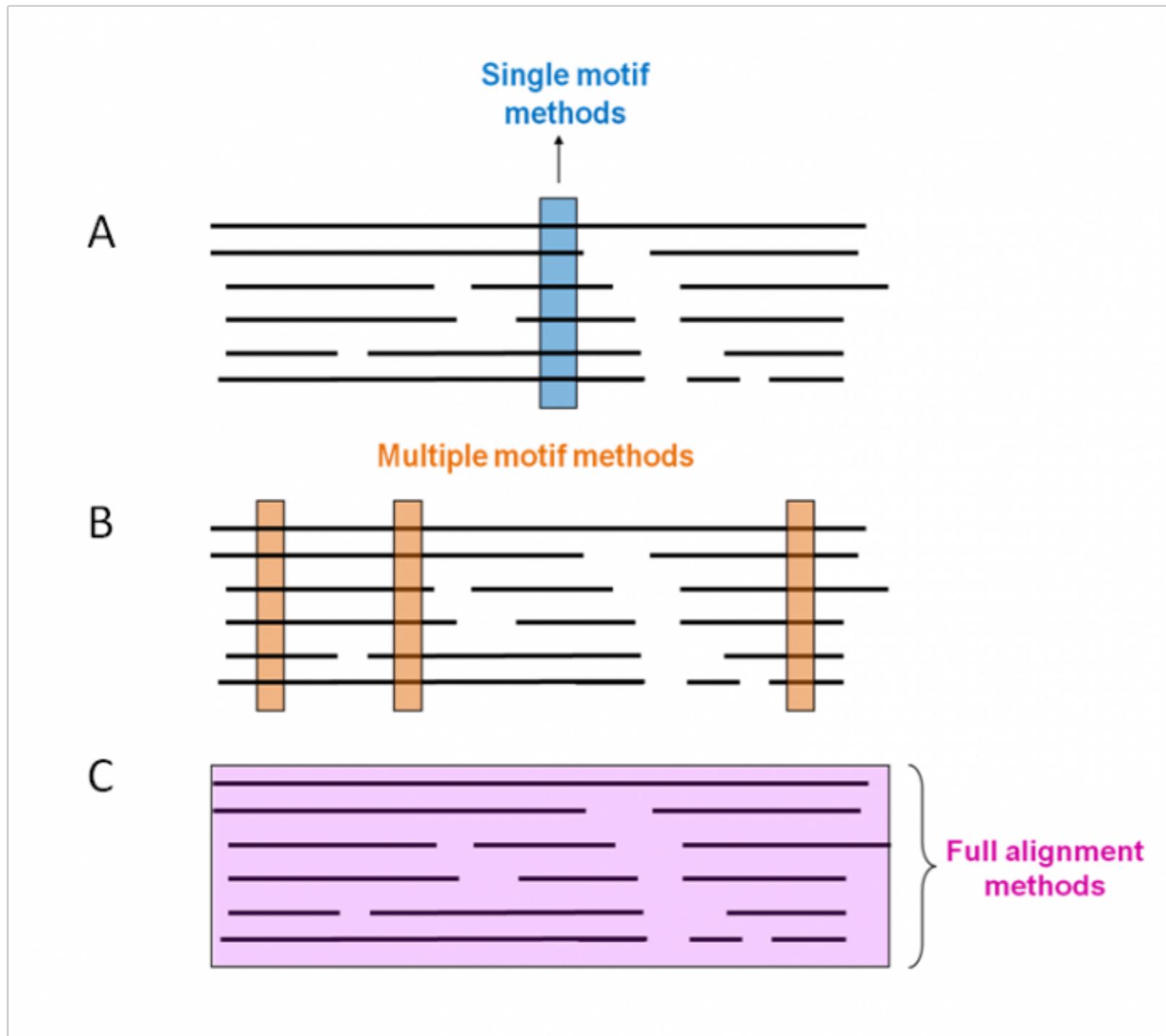


# How do we find protein domains?



**Multiple Sequence Alignment via Multiple Sequence Comparison Log-Expectation (MUSCLE)**

# How can we find protein domains?

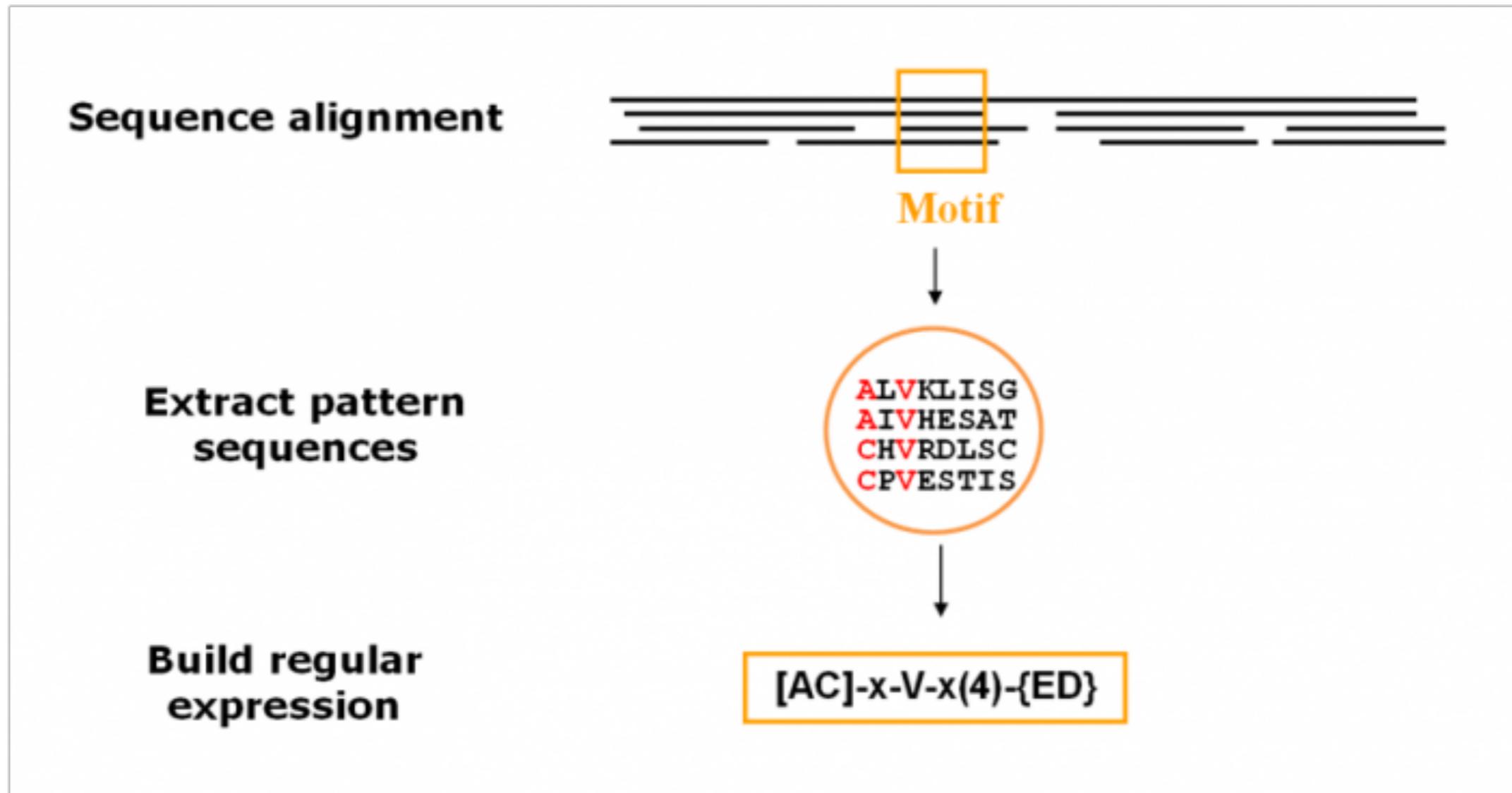


**Patterns**

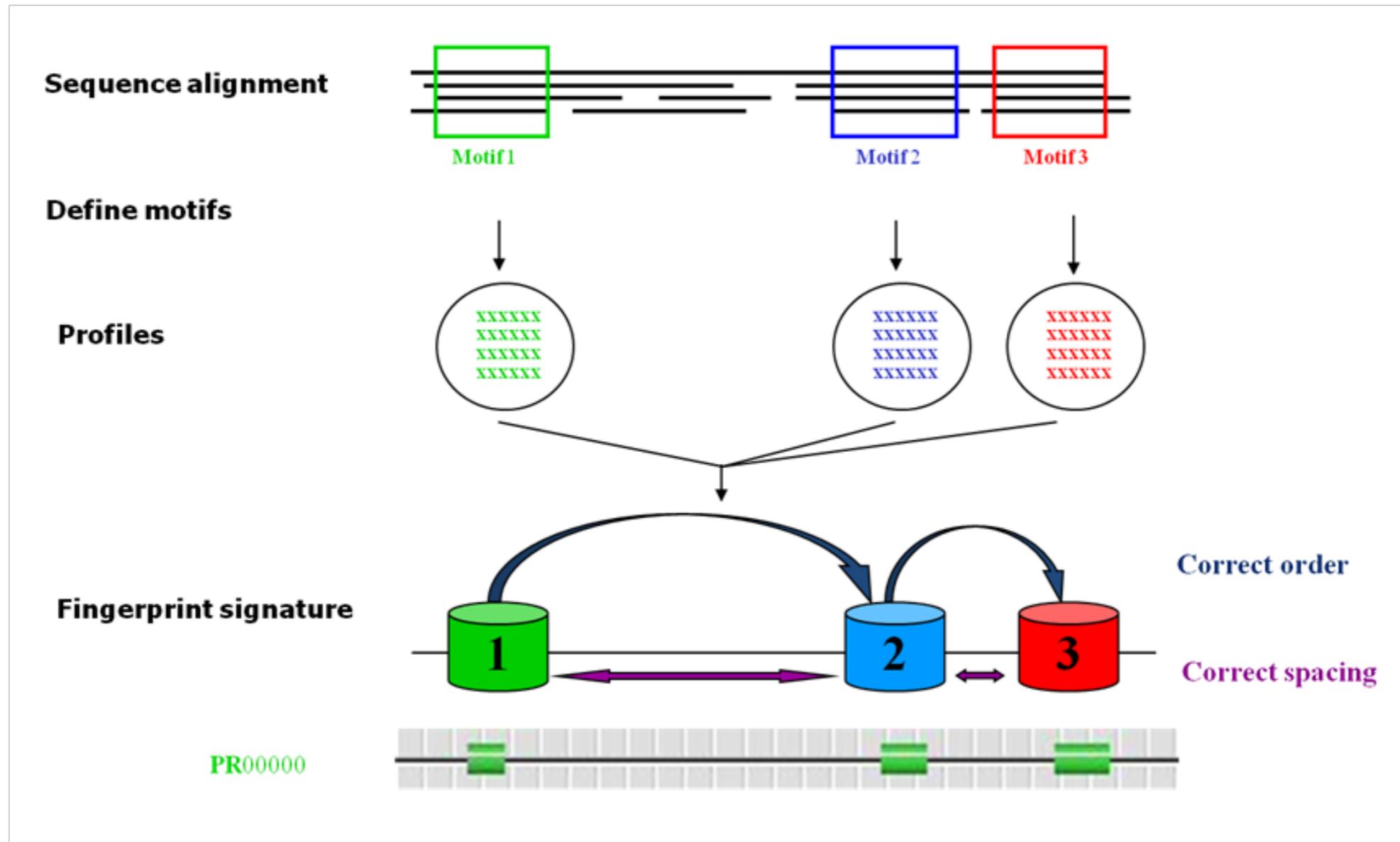
**Fingerprints**

**Profiles and Hidden Markov Model (HMM)**

# Finding domains using patterns



# Finding domains using fingerprints



**PRINTS**

# Fingerprint vs patterns

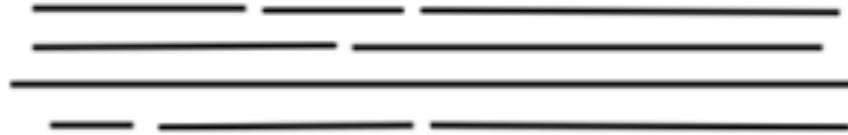
CLCN1_HUMAN	F	P	L	V	L	I	L	F	S	A	L	F	C	H	L	I	S	P	Q	A	V	G	S	G	I	P	E	M	K	T	I	L	R	G	V	V	L	K	E	Y	L	T	M	K	A	F	V	A	K
CLCN1_RAT	F	P	L	I	L	I	L	F	S	A	L	F	C	Q	L	I	S	P	Q	A	V	G	S	G	I	P	E	M	K	T	I	L	R	G	V	V	L	K	E	Y	L	T	L	K	A	F	V	A	K
CLCN2_HUMAN	Y	P	V	V	L	I	T	F	S	A	G	F	T	Q	I	L	A	P	Q	A	V	G	S	G	I	P	E	M	K	T	I	L	R	G	V	V	L	K	E	Y	L	T	L	K	T	F	I	A	K
CLCN2_MOUSE	Y	P	V	V	L	I	T	F	S	A	G	F	T	Q	I	L	A	P	Q	A	V	G	S	G	I	P	E	M	K	T	I	L	R	G	V	V	L	K	E	Y	L	T	L	K	T	F	V	A	K
CLCN3_RAT	W	A	L	S	F	A	F	L	A	V	S	L	V	K	V	F	A	P	Y	A	C	G	S	G	I	P	E	I	K	T	I	L	S	G	F	I	I	R	G	Y	L	G	K	W	T	L	M	I	K
CLCN3_PONAB	W	A	L	S	F	A	F	L	A	V	S	L	V	K	V	F	A	P	Y	A	C	G	S	G	I	P	E	I	K	T	I	L	S	G	F	I	I	R	G	Y	L	G	K	W	T	L	M	I	K
CLCN3_RABIT	W	A	L	S	F	A	F	L	A	V	S	L	V	K	V	F	A	P	Y	A	C	G	S	G	I	P	E	I	K	T	I	L	S	G	F	I	I	R	G	Y	L	G	K	W	T	L	M	I	K

 Amino acids relatively well conserved across all chloride channel protein family members

 Amino acids uniquely conserved in chloride channel protein 3 subfamily members

# Finding domains using profiles

Sequence alignment



Residue frequency at each position

Sequence 1:	F	K	L	L	S	H	C	L	L	V
Sequence 2:	F	K	A	P	G	Q	T	M	P	Q
Sequence 3:	Y	P	I	V	G	Q	E	L	L	G
Sequence 4:	F	P	V	V	K	E	A	I	L	K
Sequence 5:	F	K	V	L	A	A	V	I	A	D
Sequence 6:	L	E	F	I	S	E	C	I	I	Q
Sequence 7:	F	K	L	L	G	N	V	L	V	C



Scoring matrix

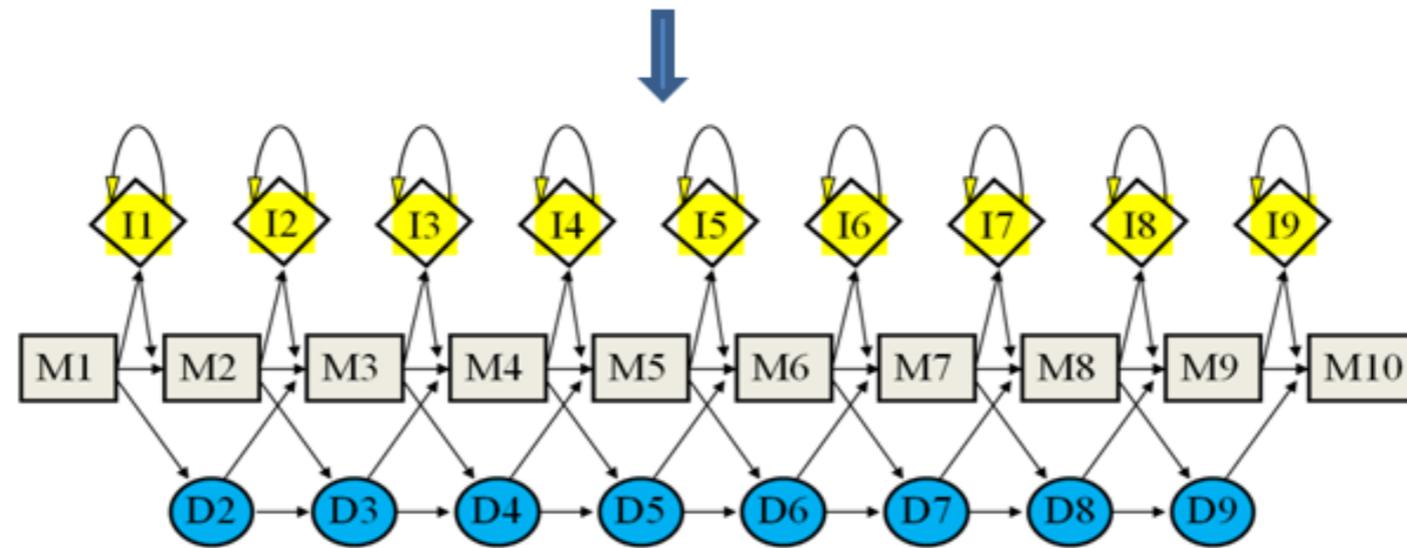
A	-18	-10	-1	-8	8	-3	3	-10	-2	-8
C	-22	-33	-18	-18	-22	-26	22	-24	-19	-7
D	-35	0	-32	-33	-7	6	-17	-34	-31	0
E	-27	15	-25	-26	-9	23	-9	-24	-23	-1
F	60	-30	12	14	-26	-29	-15	4	12	-29
G	-30	-20	-28	-32	28	-14	-23	-33	-27	-5
H	-13	-12	-25	-25	-16	14	-22	-22	-23	-10
I	3	-27	21	25	-29	-23	-8	33	19	-23
K	-26	25	-25	-27	-6	4	-15	-27	-26	0
L	14	-28	19	27	-27	-20	-9	33	26	-21
M	3	-15	10	14	-17	-10	-9	25	12	-11
N	-23	-6	-24	-27	1	8	-15	-24	-24	-4
P	-30	24	-26	-28	-14	-10	-22	-24	-26	-18
Q	-32	5	-25	-26	-9	24	-16	-17	-23	7
R	-18	9	-22	-22	-10	0	-18	-23	-22	-4
S	-22	-8	-16	-21	11	2	-1	-24	-19	-4
T	-10	-10	-6	-7	-5	-8	2	-10	-7	-11
V	0	-25	22	25	-19	-24	6	19	16	-16
W	9	-25	-18	-19	-25	-27	-34	-20	-17	-28
Y	34	-18	-1	1	-23	-12	-19	0	0	-18



# Finding domains using HMM

Multiple sequence alignment

Sequence 1:	F	K	L	L	S	H	C	L	L	V
Sequence 2:	F	K	A	F	G	Q	T	M	F	Q
Sequence 3:	Y	P	I	V	G	Q	E	L	L	G
Sequence 4:	F	P	V	V	K	E	A	I	L	K
Sequence 5:	F	K	V	L	A	A	V	I	A	D
Sequence 6:	L	E	F	I	S	E	C	I	I	Q
Sequence 7:	F	K	L	L	G	N	V	L	V	C



I = insert state

M = match state

D = delete state

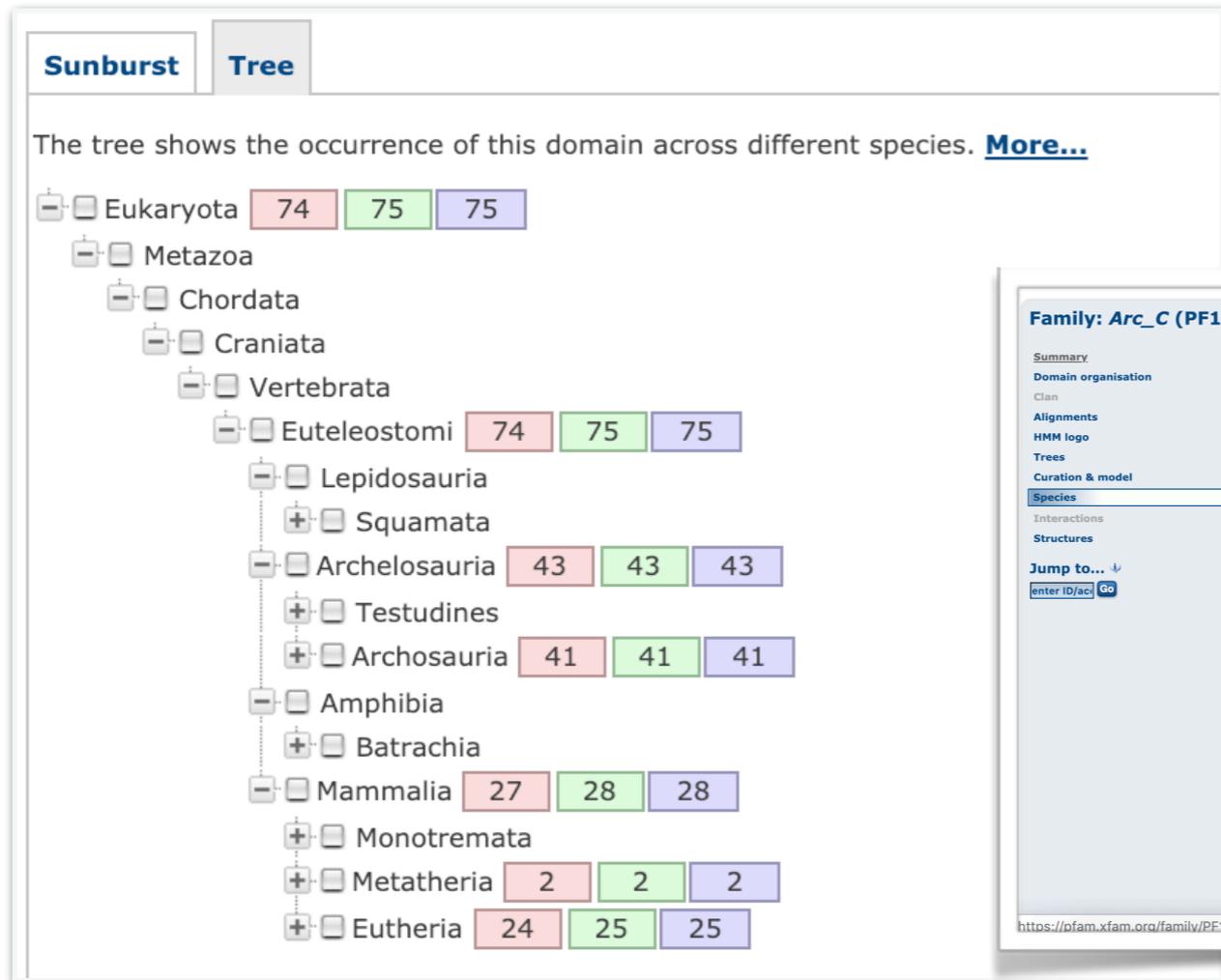
**Pfam**

tigr fams  
tigr protein families

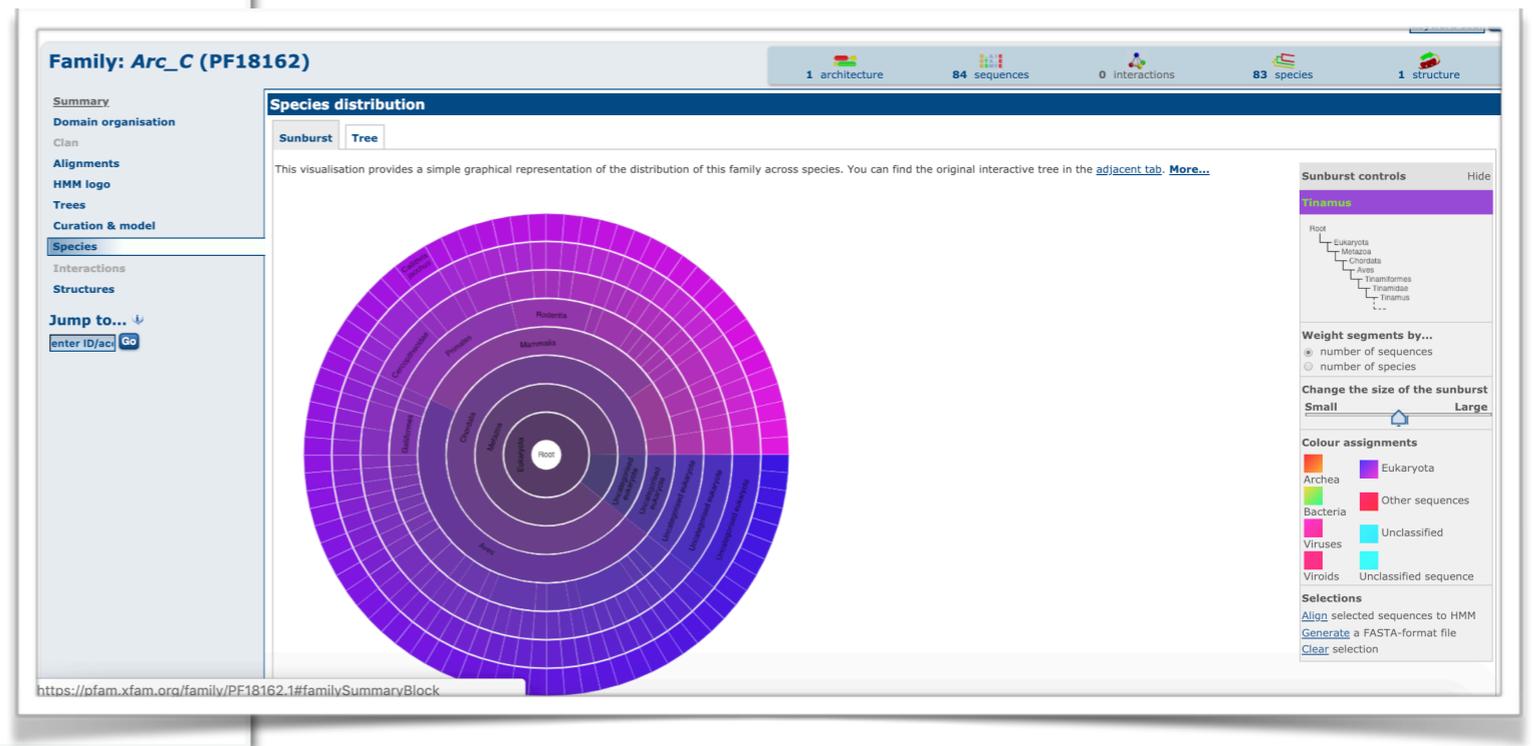
**SMART**

# What results do Pfam provide?

## Phylogenetic Tree



## Sunburst via HMM



Highlighted protein domains

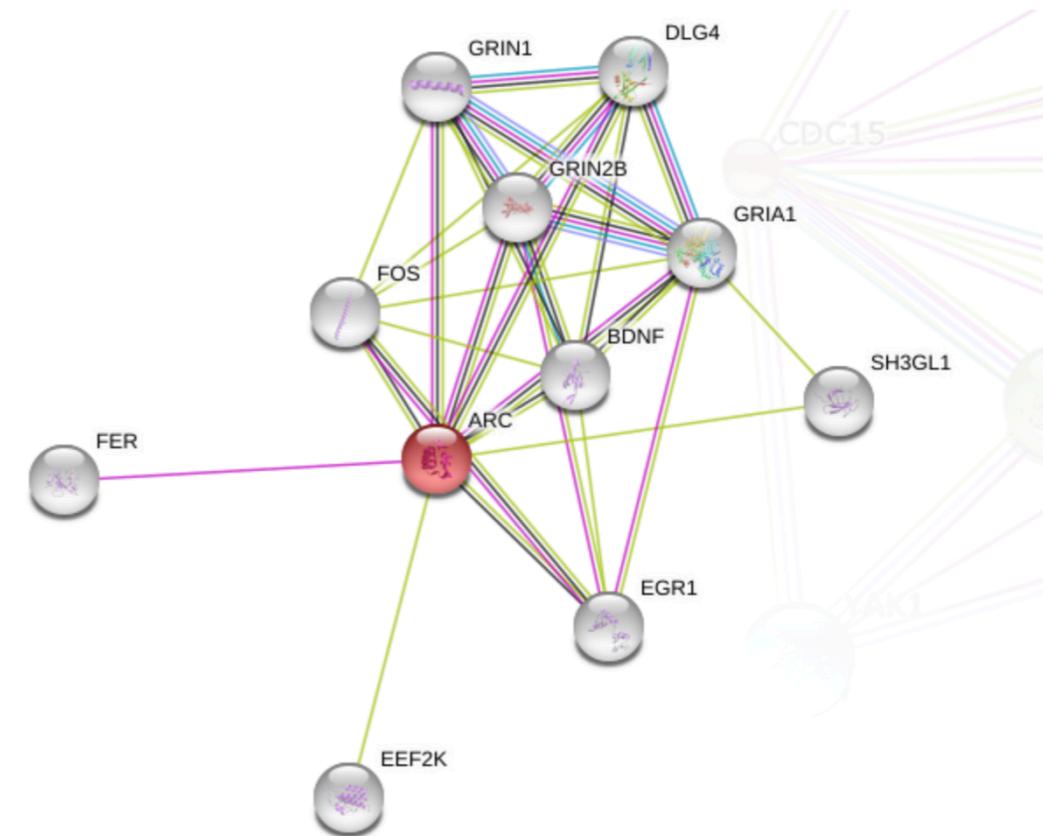
# What results do SMART provide?

Simple Modular Architecture Research Tool

## Protein orthologs

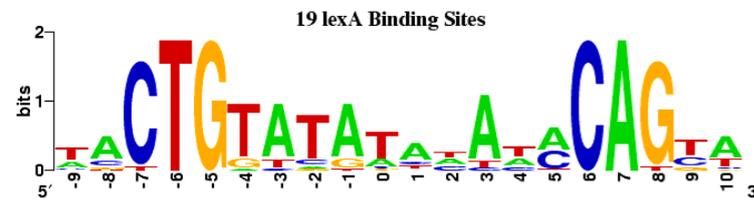
The screenshot shows the 'Orthology' tab of the SMART web interface. It displays a list of orthologous groups for the protein ENSP00000349022. The table below summarizes the data shown in the screenshot.

Orthologous group	Description	Taxonomic class
ENOG410A1GI	activity-regulated cytoskeleton-associated protein	<i>Bilateria</i> (no rank)
ENOG410DP49	activity-regulated cytoskeleton-associated protein	<i>Chordata</i> (phylum)
ENOG410IIAZ	activity-regulated cytoskeleton-associated protein	<i>Eukaryotes</i> (superkingdom)
ENOG410RPU3	activity-regulated cytoskeleton-associated protein	<i>Hominidae</i> (family)
ENOG410UPC0	activity-regulated cytoskeleton-associated protein	<i>Mammals</i> (class)
ENOG410VCC3	activity-regulated cytoskeleton-associated protein	<i>Animals</i> (kingdom)
ENOG410YF34	activity-regulated cytoskeleton-associated protein	<i>All organisms</i> (no rank)
ENOG4113MVY	activity-regulated cytoskeleton-associated protein	<i>Opisthokonts</i> (no rank)
ENOG41167VF	activity-regulated cytoskeleton-associated protein	<i>Primates</i> (order)
ENOG411AUZ9	activity-regulated cytoskeleton-associated protein	<i>Supraprimates</i> (superorder)
ENOG411CY28	activity-regulated cytoskeleton-associated protein	<i>Vertebrates</i> (no rank)

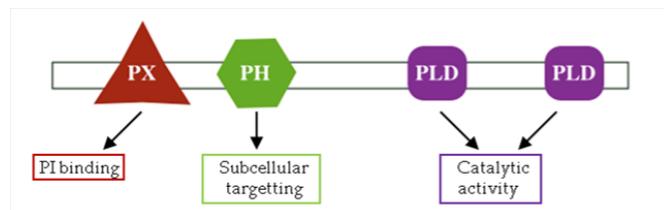


## Protein interactions

# Summary



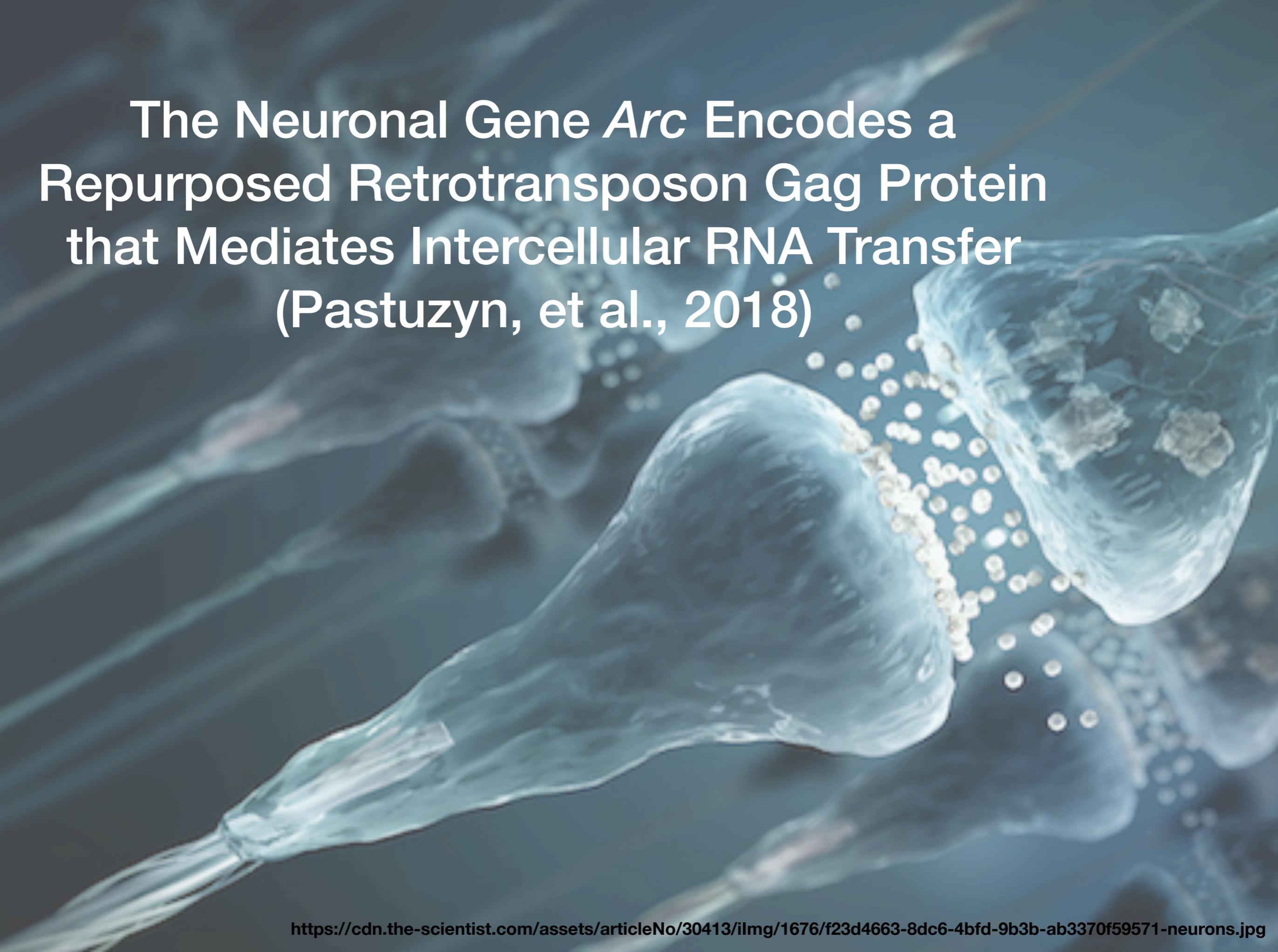
**Motif is a functional unit of a domain**



**Domains are a functional unit of proteins**

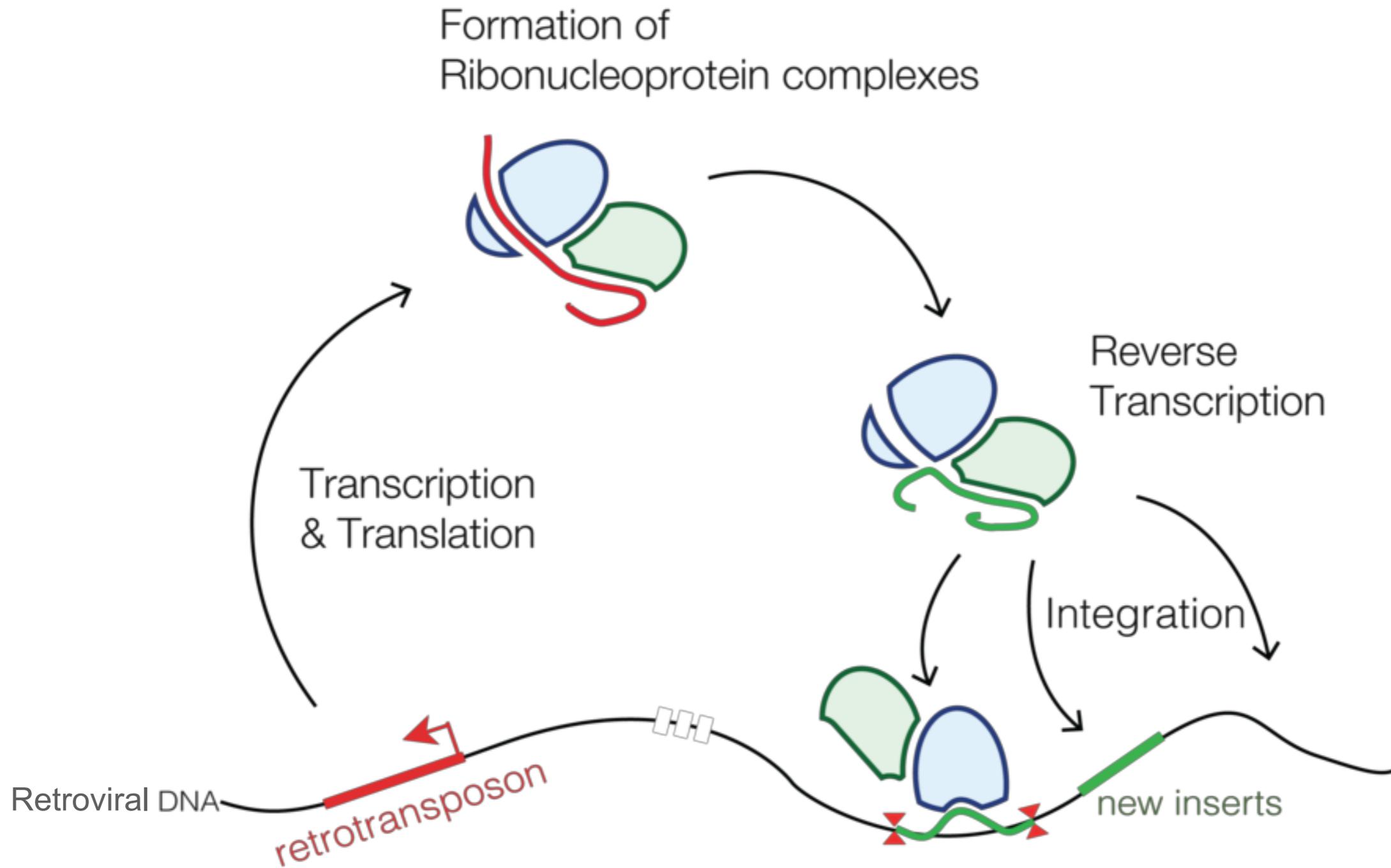


**Classifying and understanding proteins can guide research, preventing diseases/unwanted protein interactions**

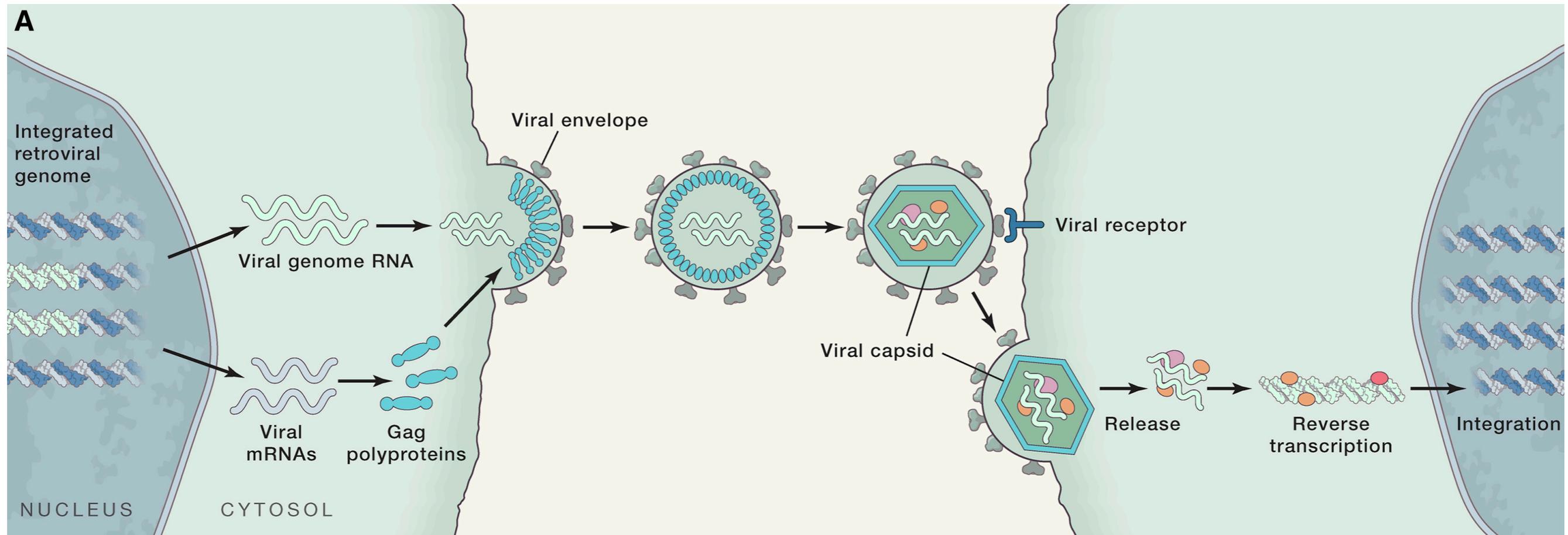


The Neuronal Gene *Arc* Encodes a  
Repurposed Retrotransposon Gag Protein  
that Mediates Intercellular RNA Transfer  
(Pastuzyn, et al., 2018)

# What are retrotransposons?

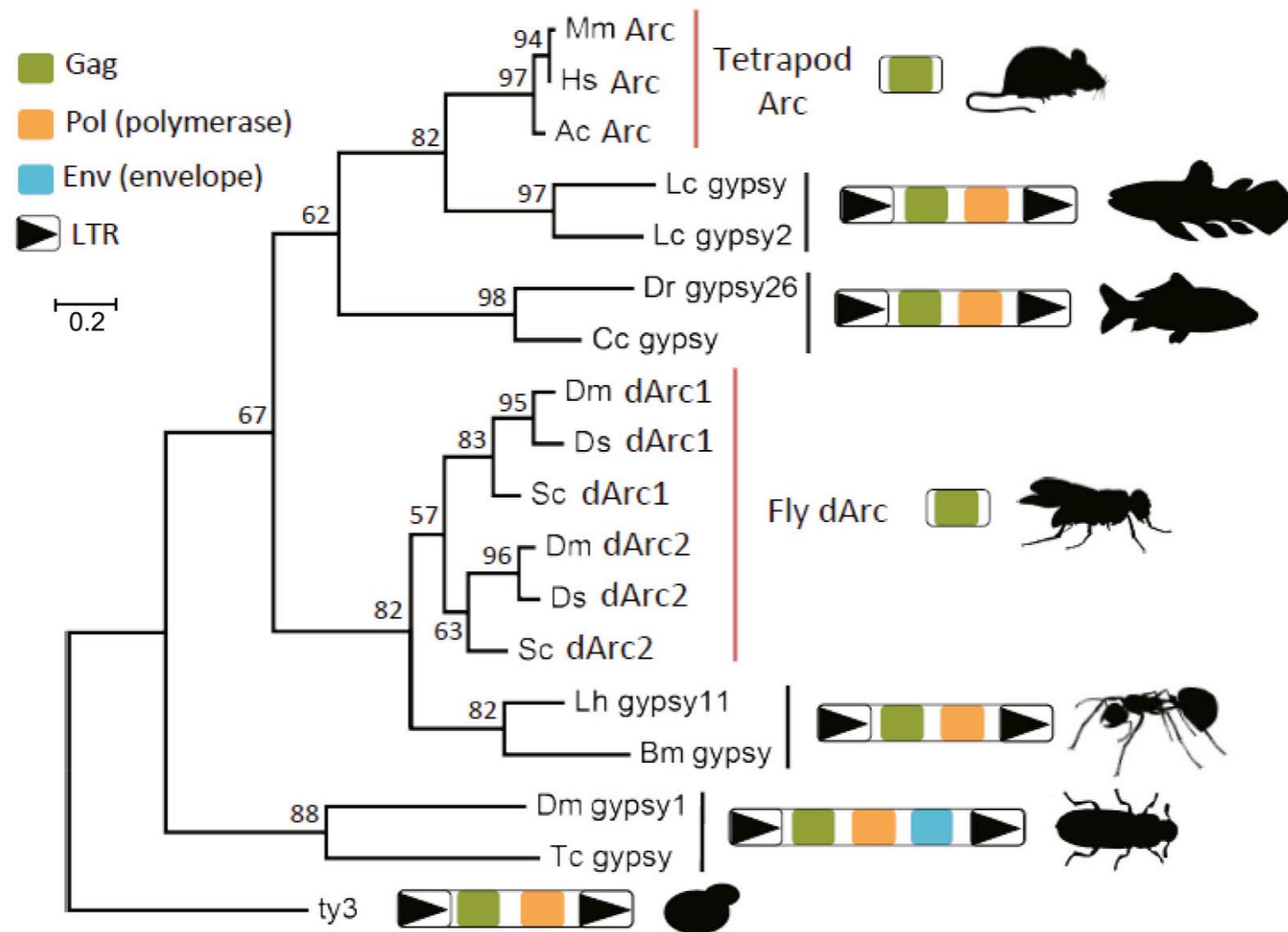


# What do retroviruses do?



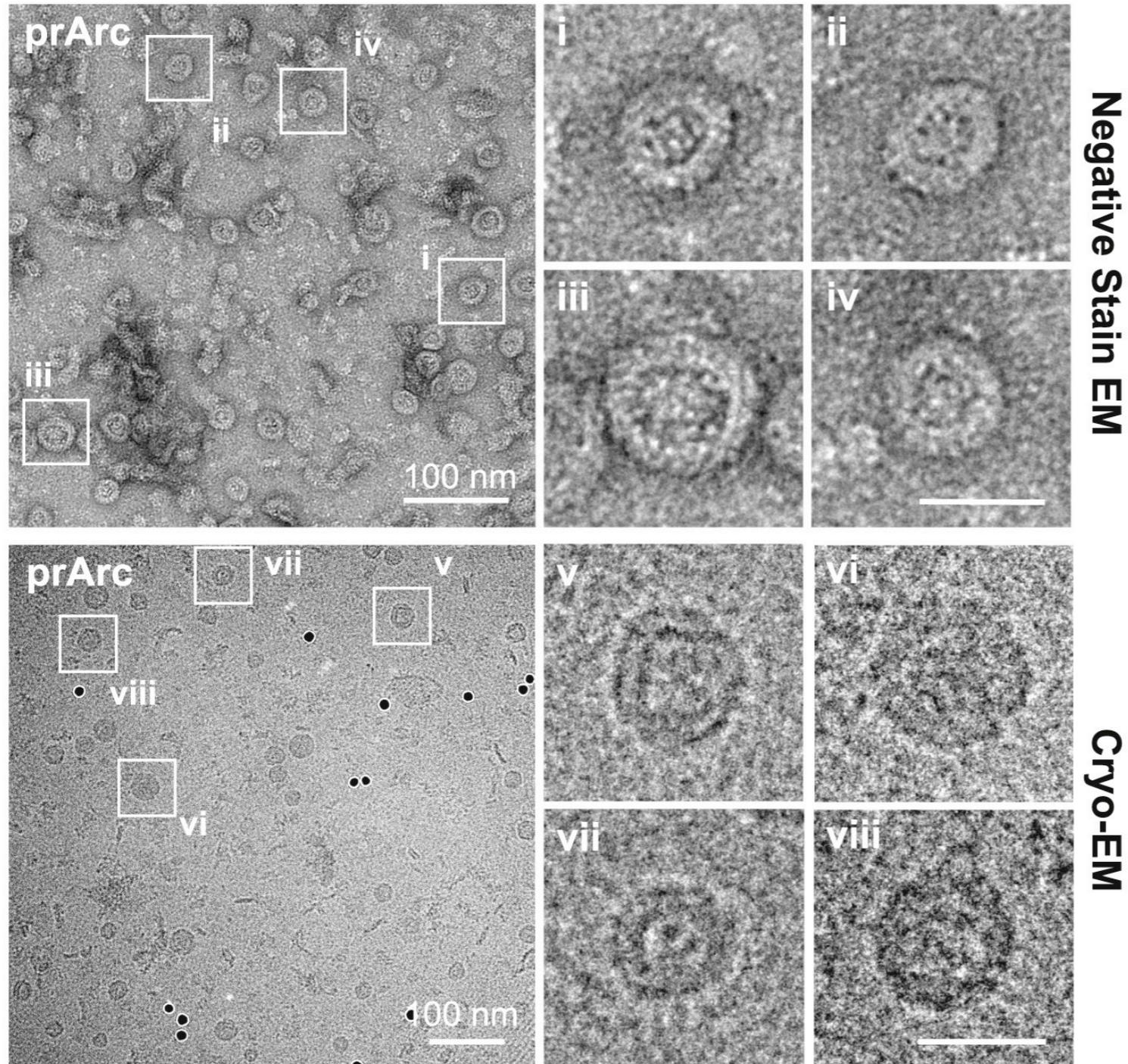
**they insert a DNA copy of their genome into the host cell**

# What is the origin of Arc?



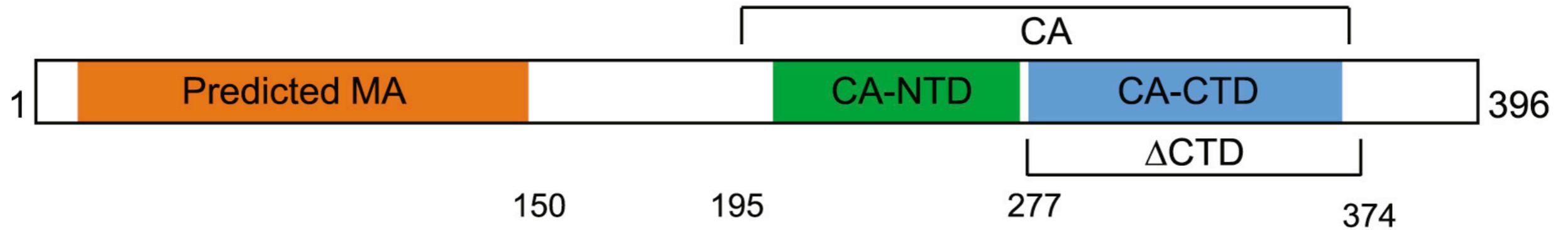
**Arc is closely related to the Ty3/gypsy retrotransposon**

# Can Arc form a virus-like capsid?



**Arc protein forms double-shell capsids**

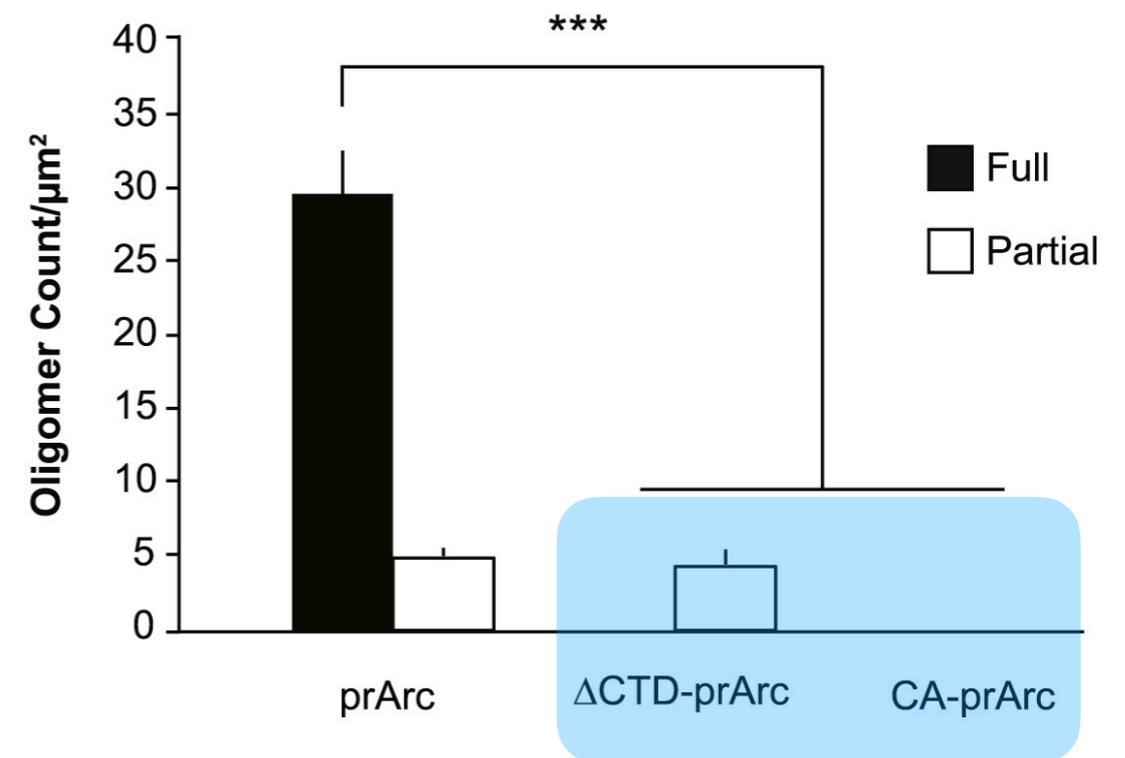
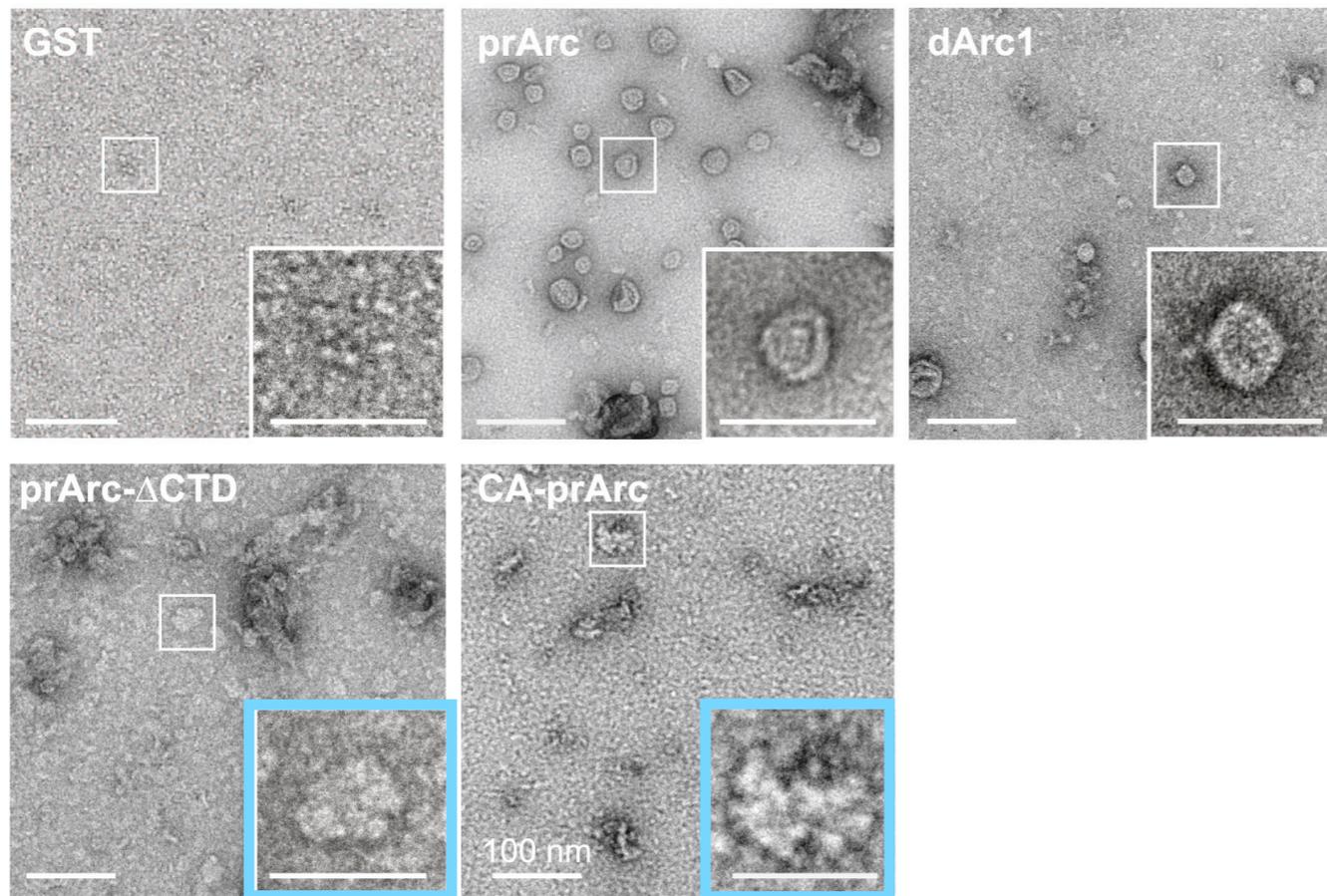
# What does the Arc protein look like?



**MA = Matrix**

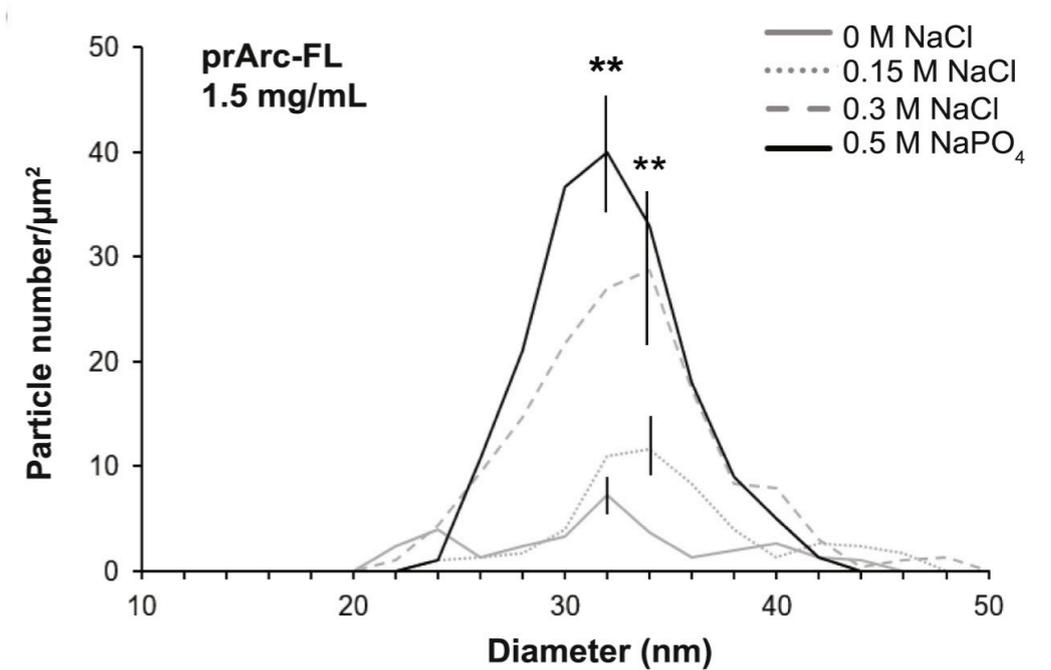
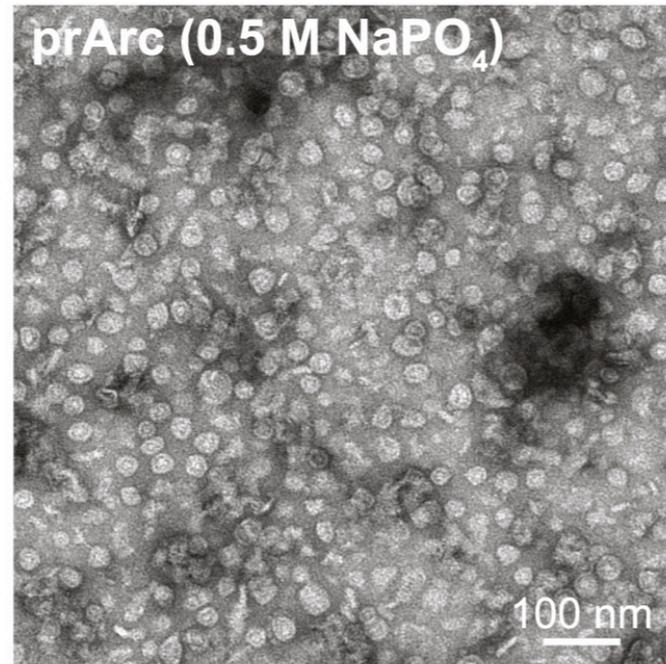
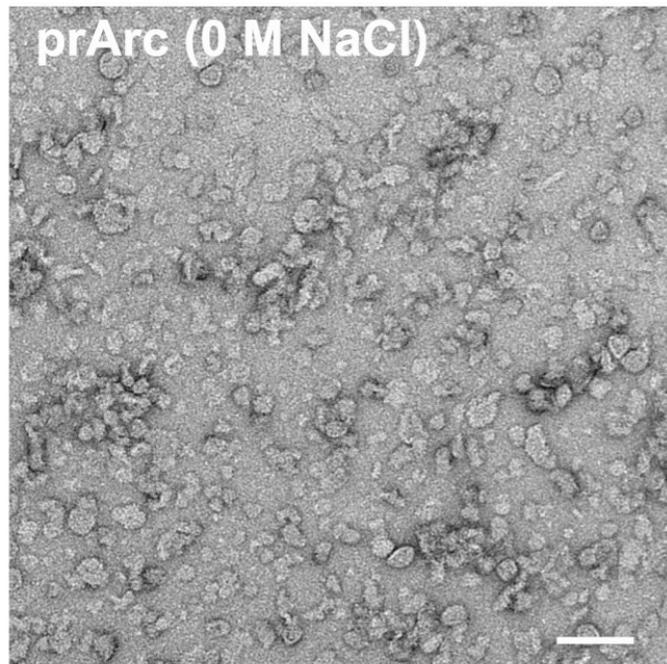
**CA = Capsid**

# Is CTD required for capsid formation?



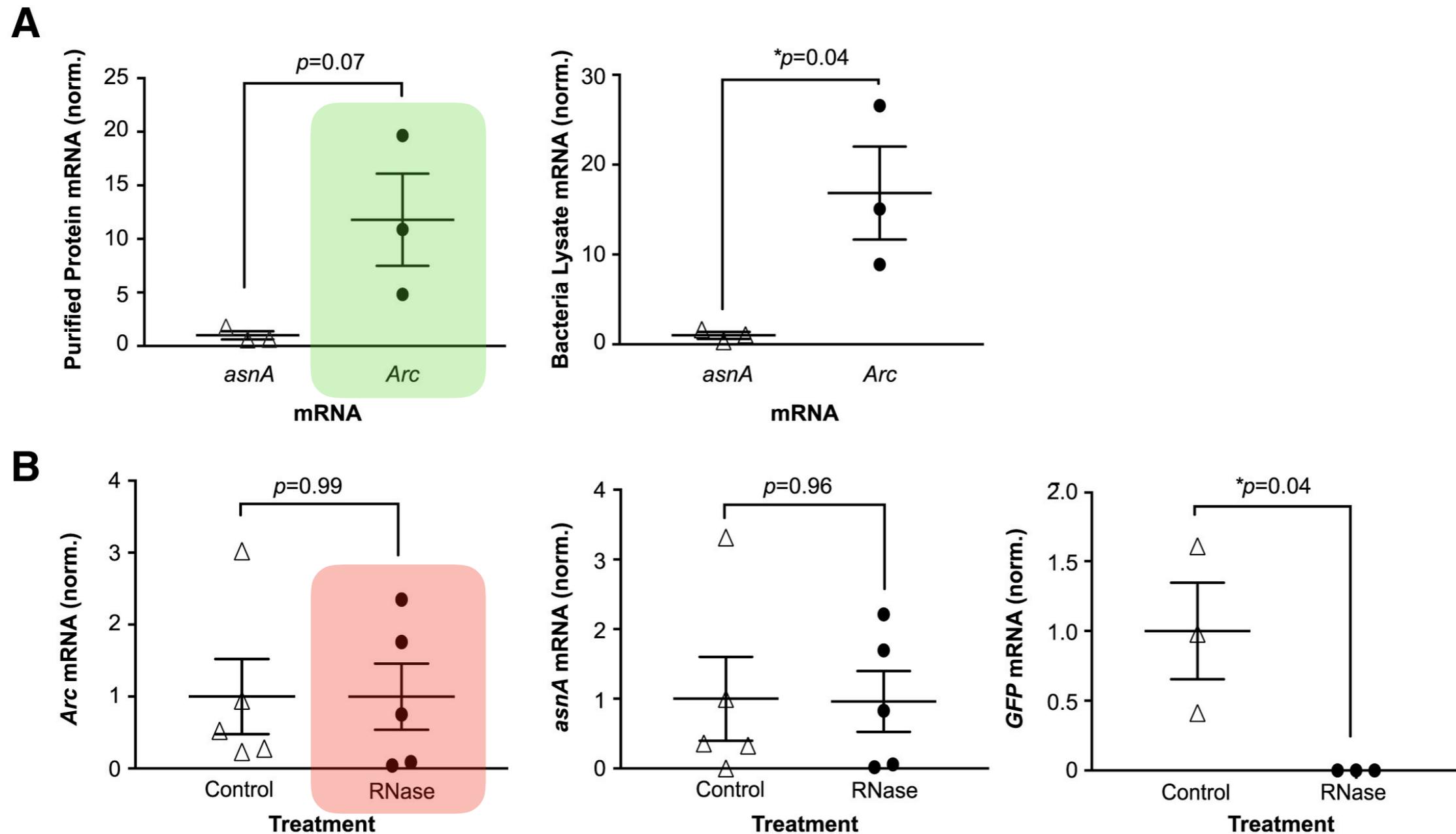
Loss of CA-CTD cannot form capsids

# How stable are these capsids?



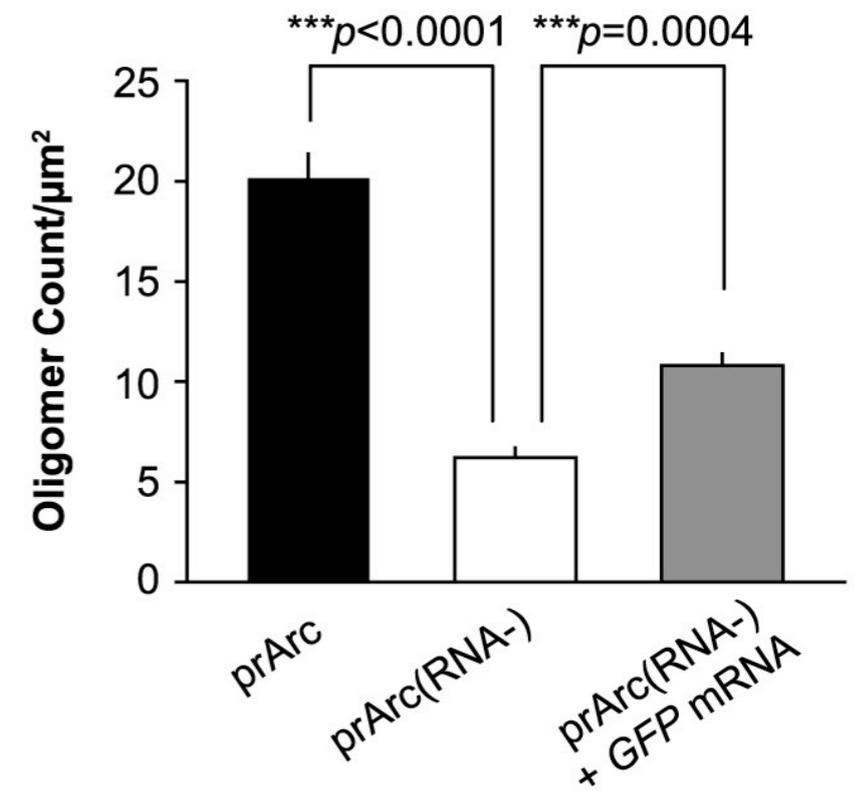
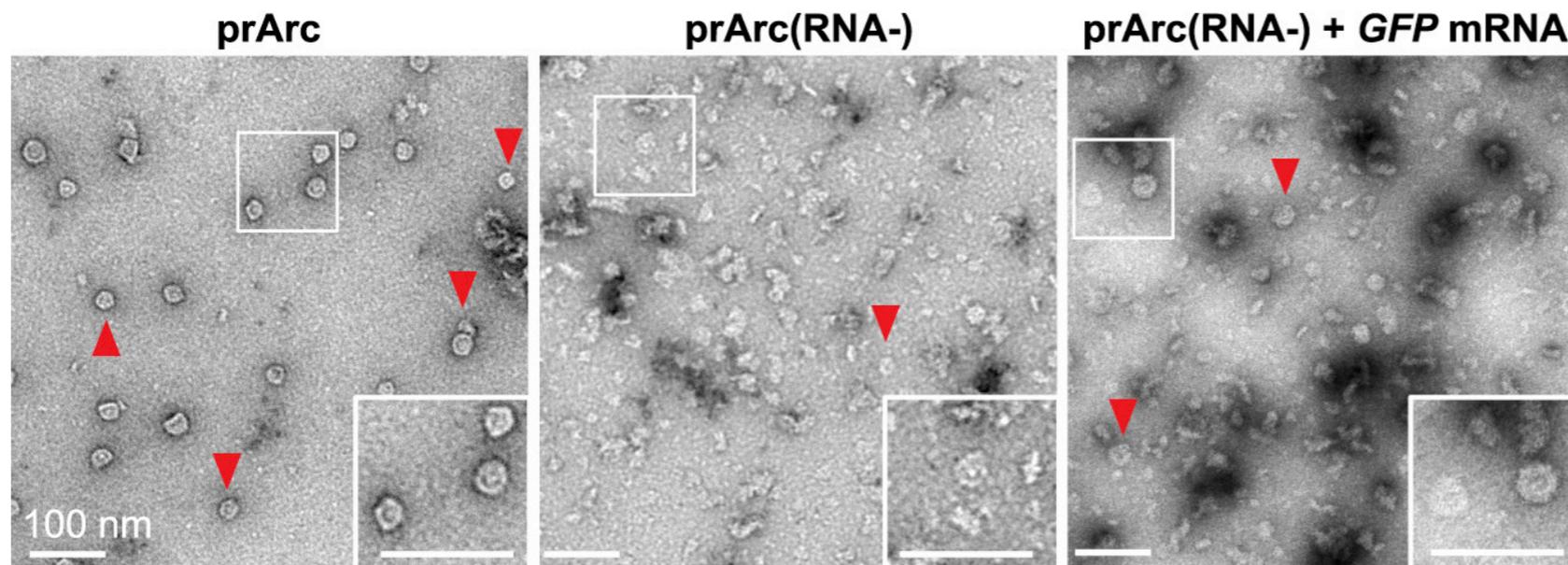
**very stable in high salt and phosphate levels**

# Can Arc bind mRNA?



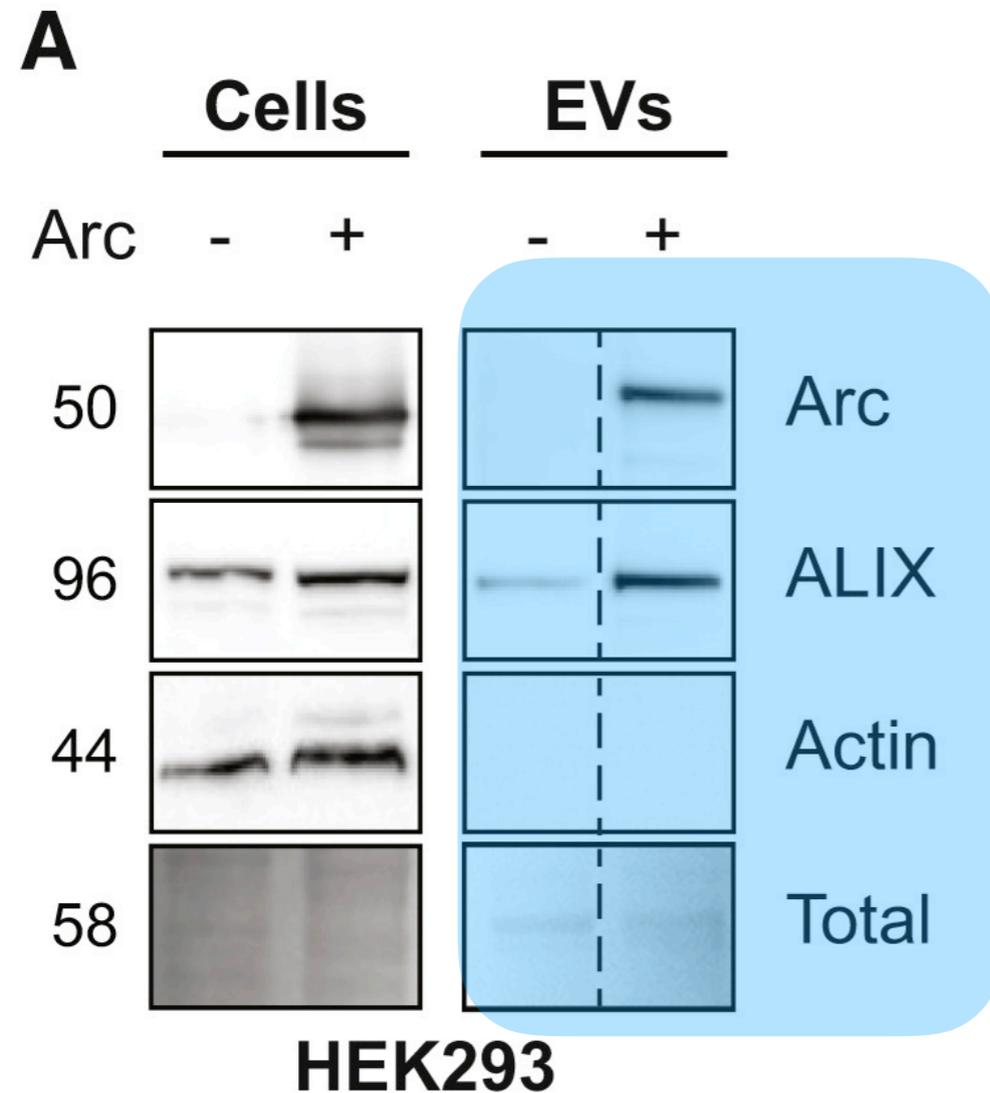
Arc protein binds mRNA and protects it from RNase degradation

# Does Arc capsid formation require RNA?



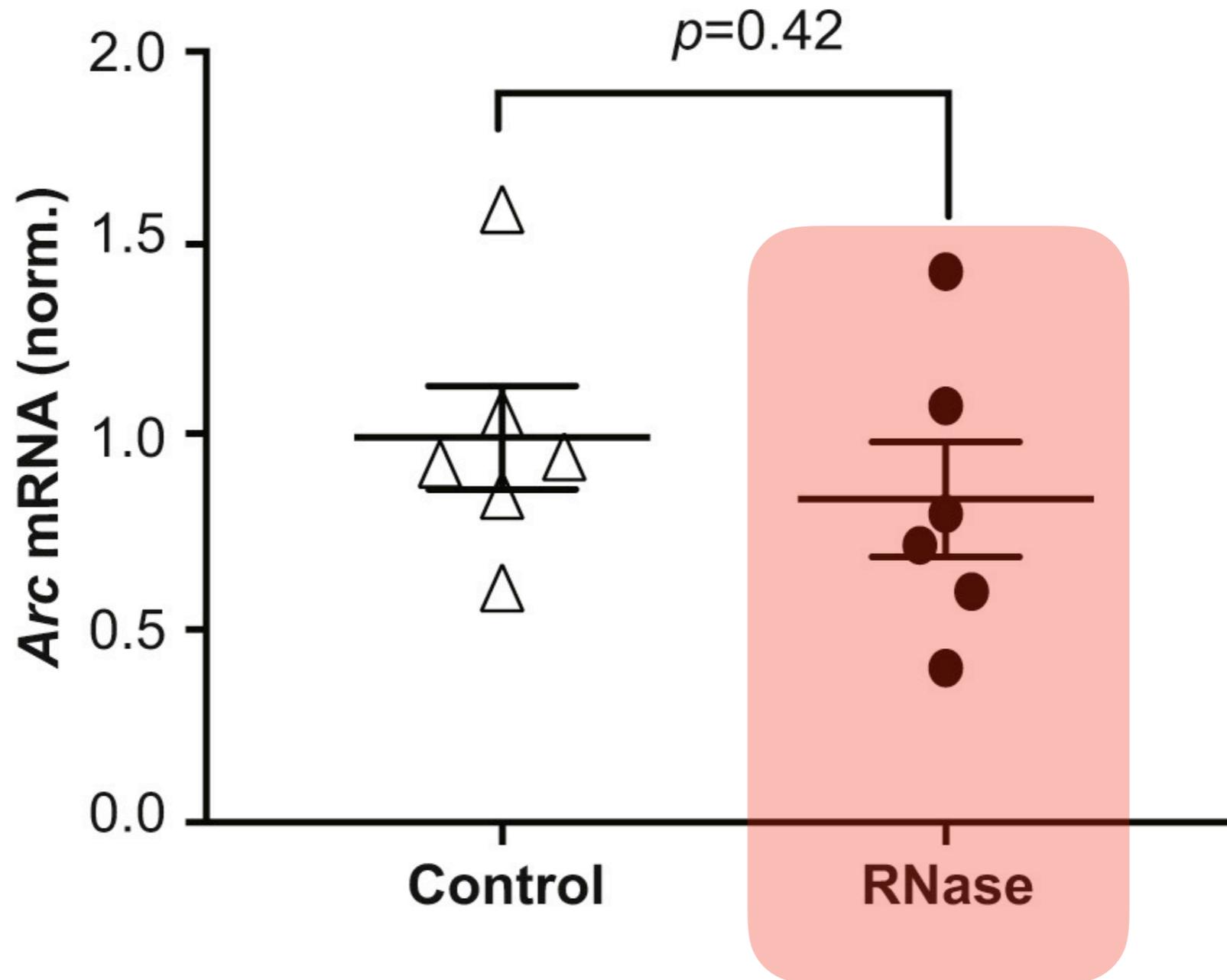
**RNA is necessary for capsid formation**

# Is Arc protein released from cells in extracellular vesicles?



**Arc protein is found in extracellular vesicles**

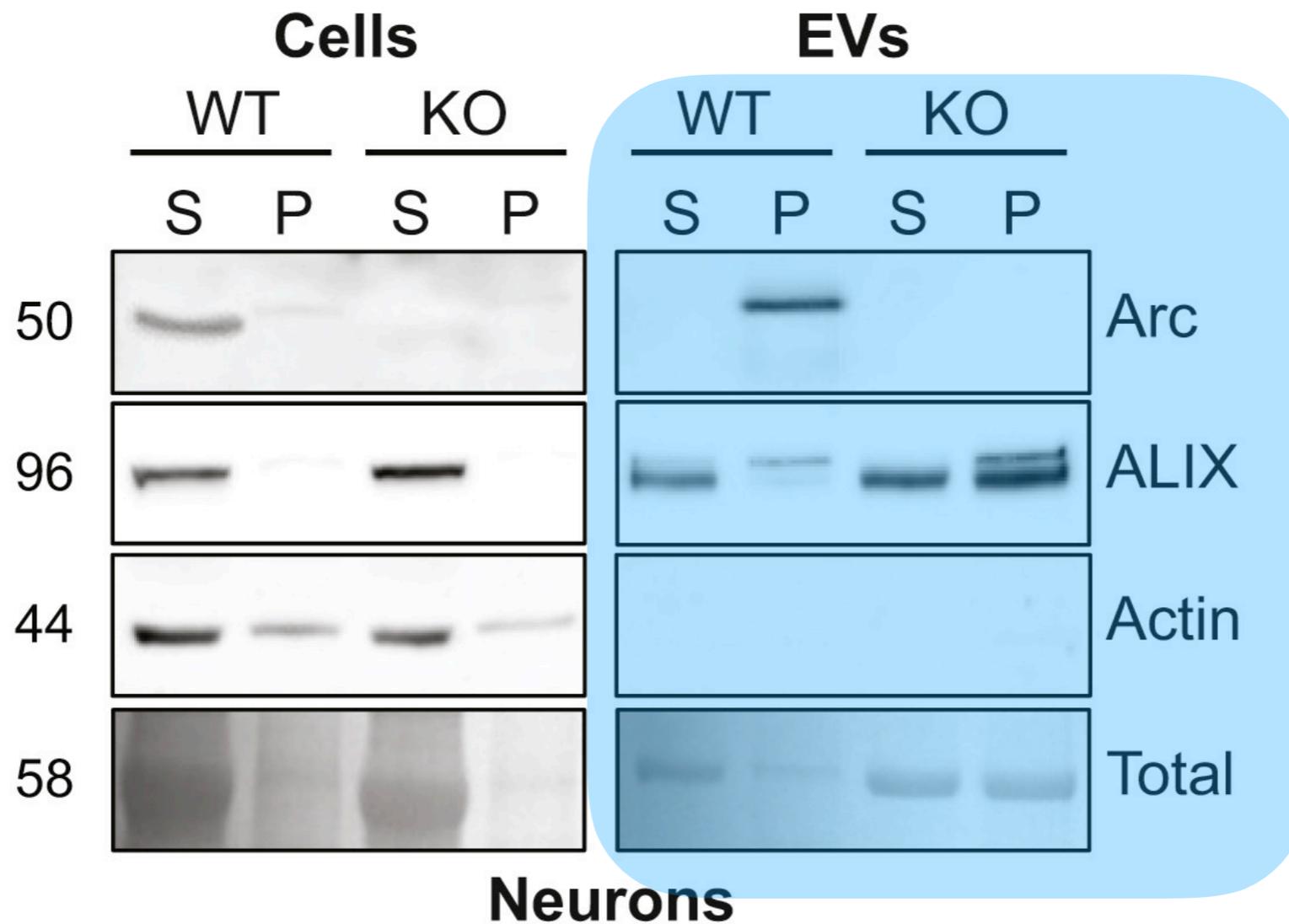
# Is the mRNA **degraded** after release?



**Arc mRNA is protected outside of cells**

# Is Arc released from neurons in extracellular vesicles?

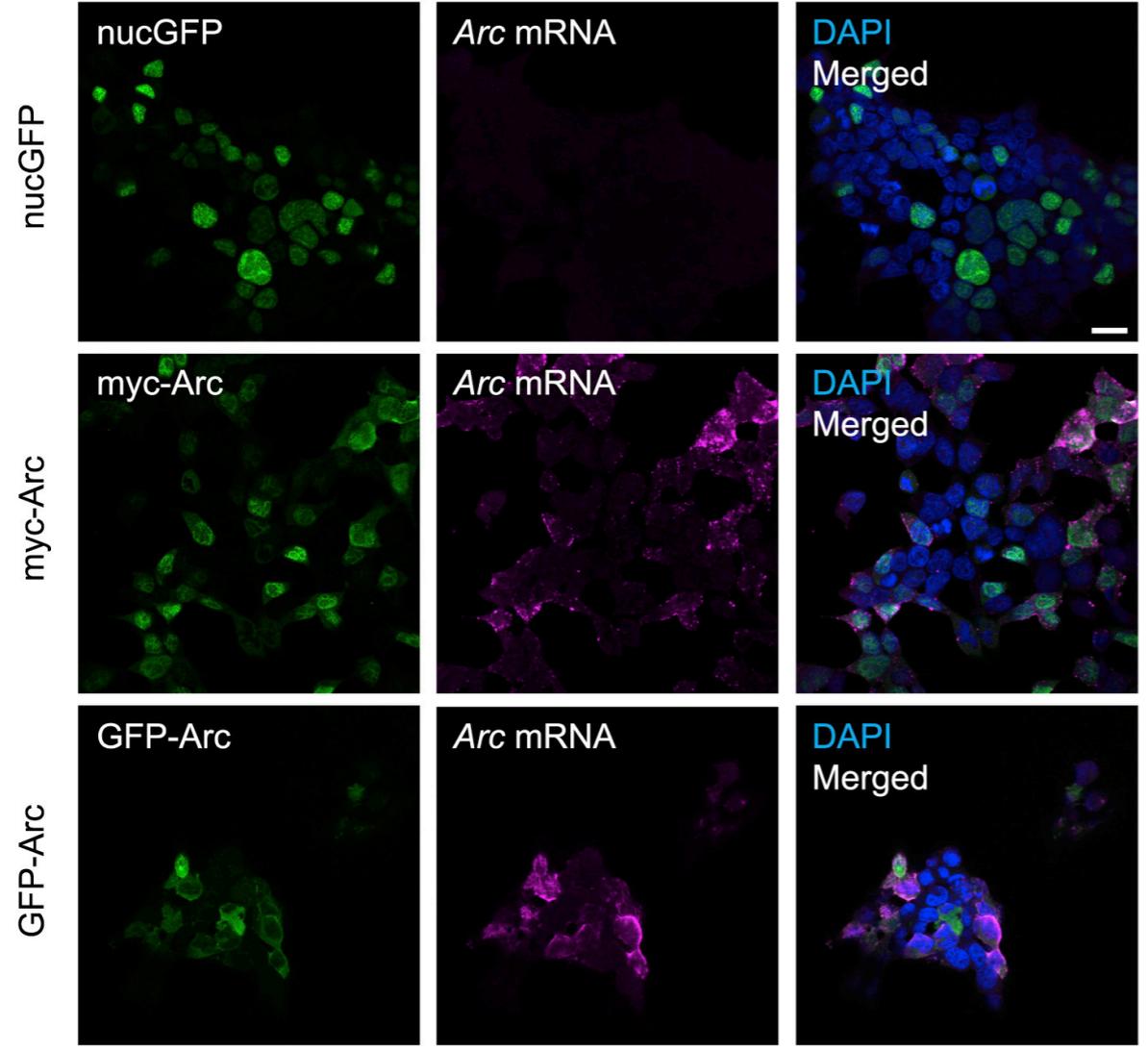
D



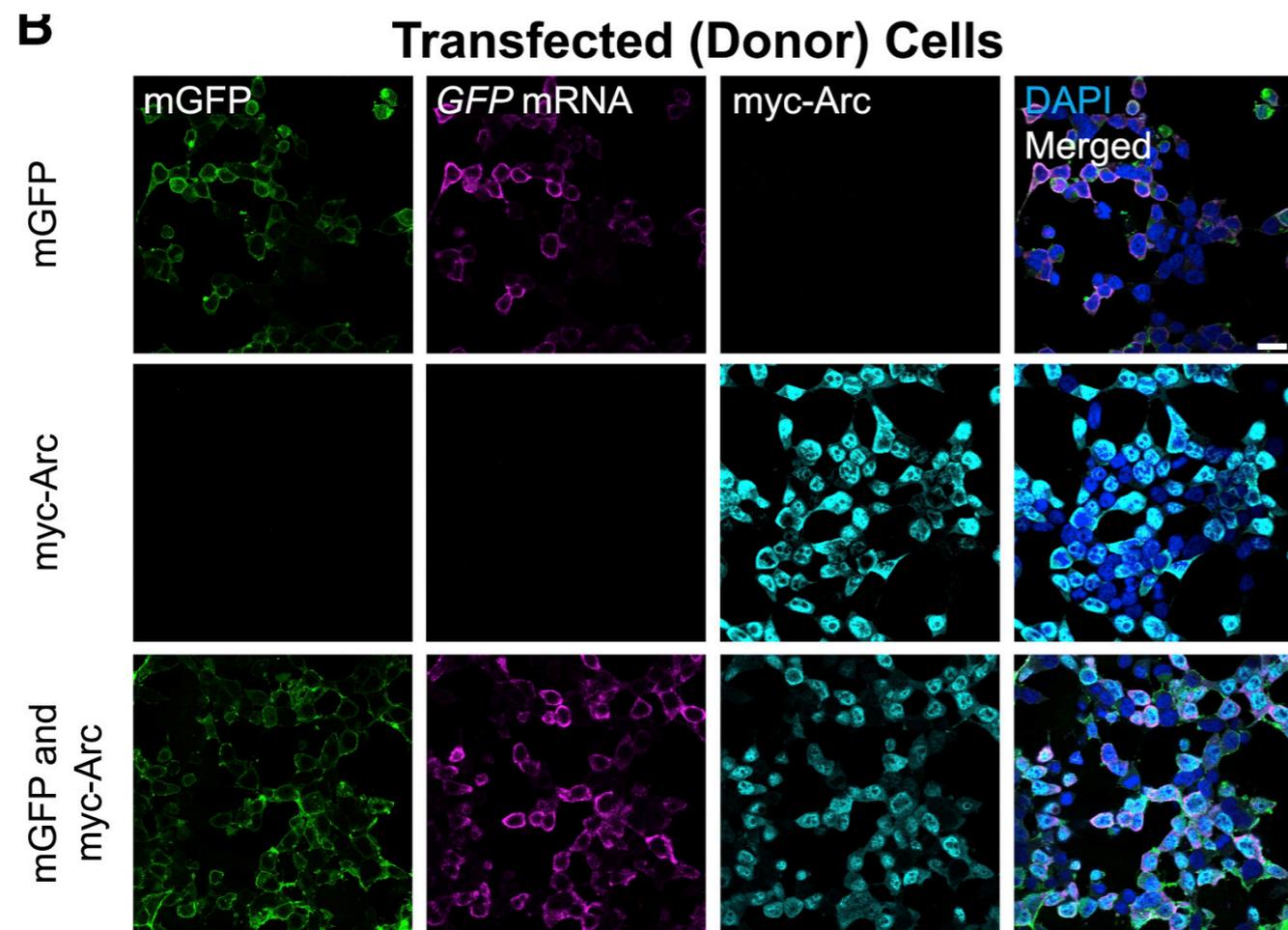
Arc proteins are found in extracellular vesicles outside of neurons

# Can Arc transfer Arc protein and mRNA between cells?

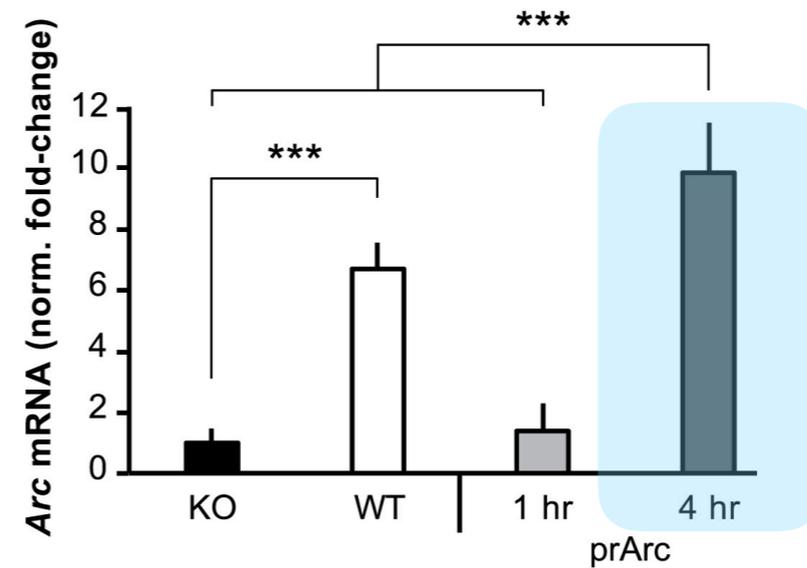
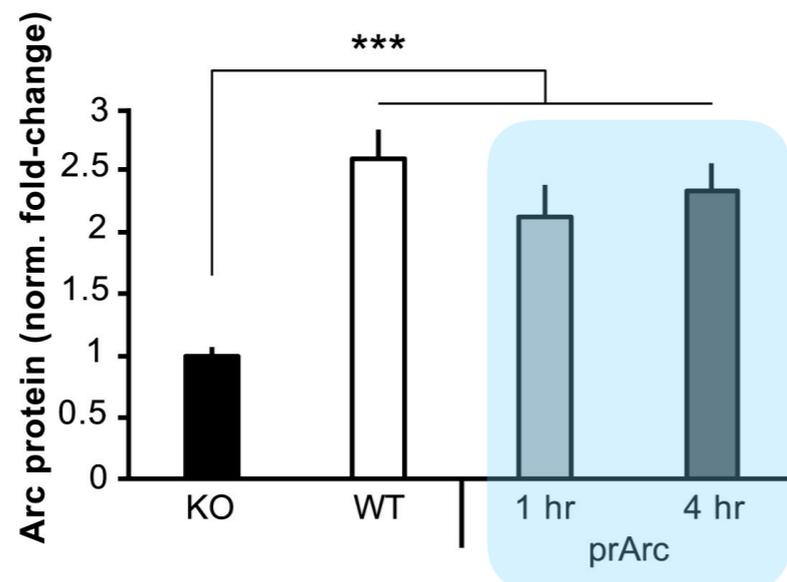
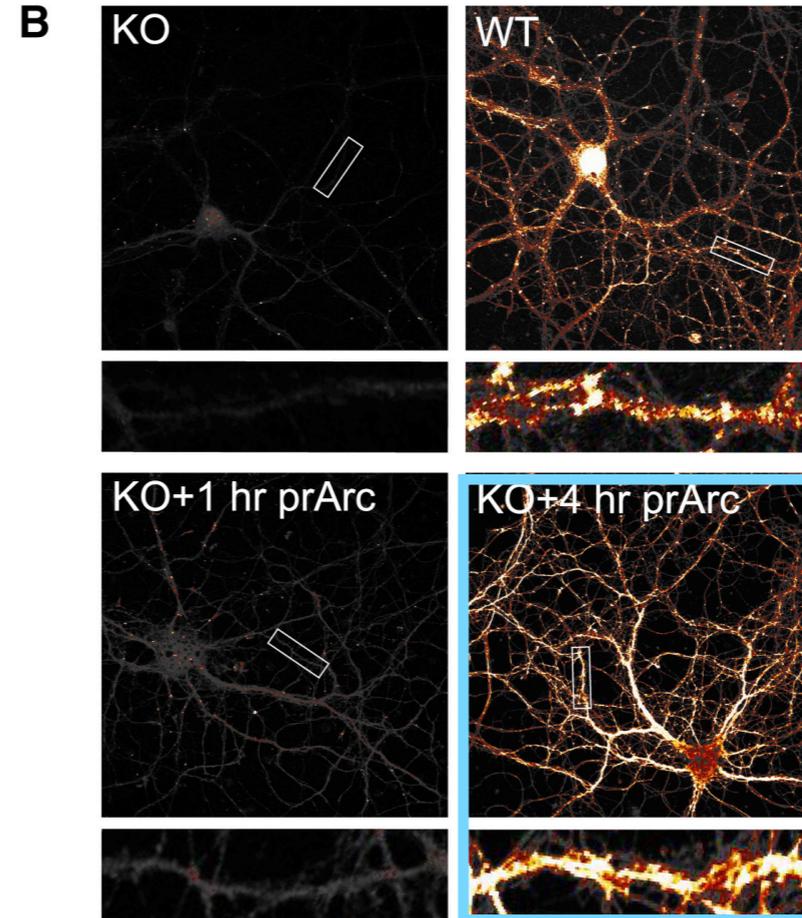
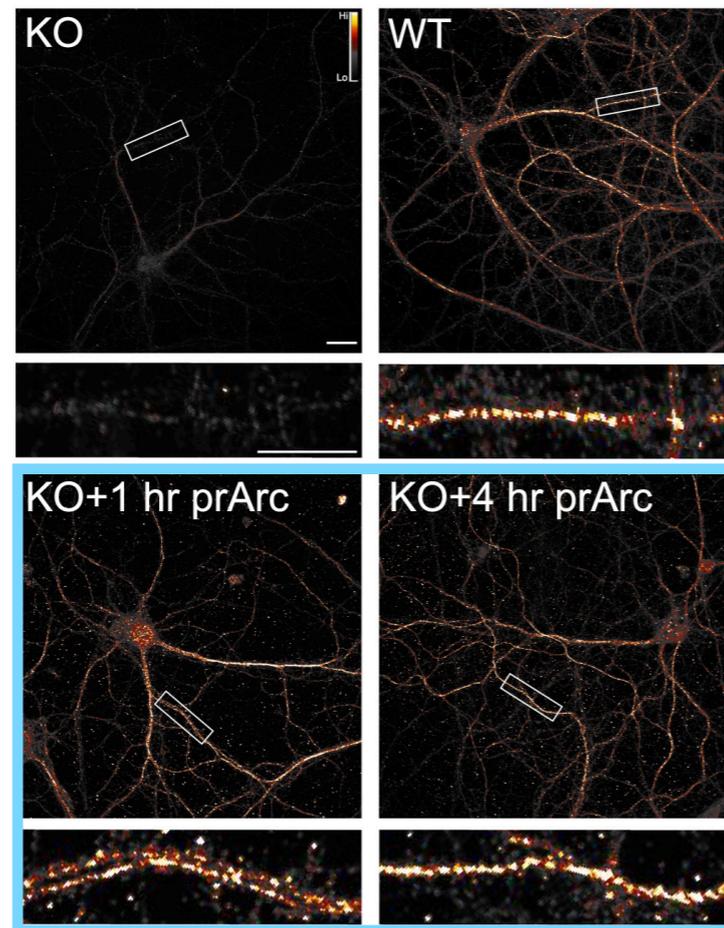
Transfected (Donor) Cells



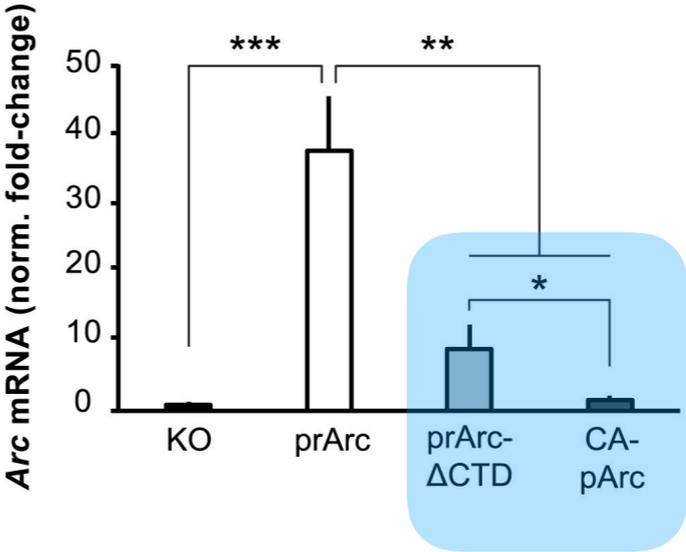
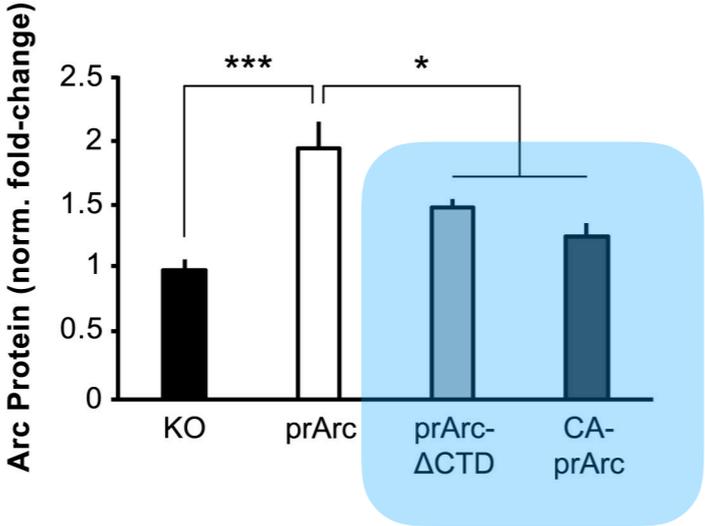
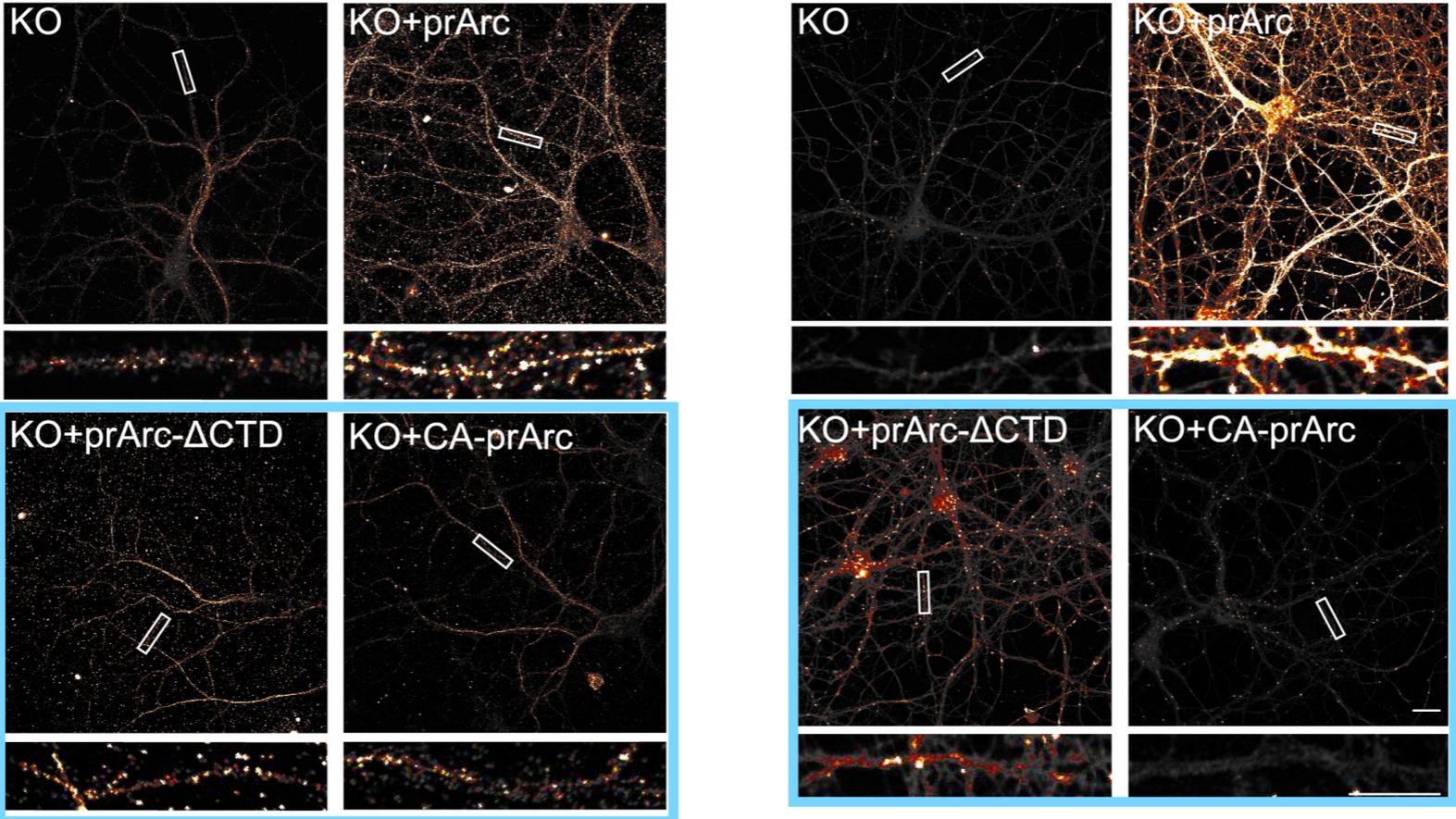
# Can Arc protein transfer mRNA between cells?



# Does this happen in neurons?

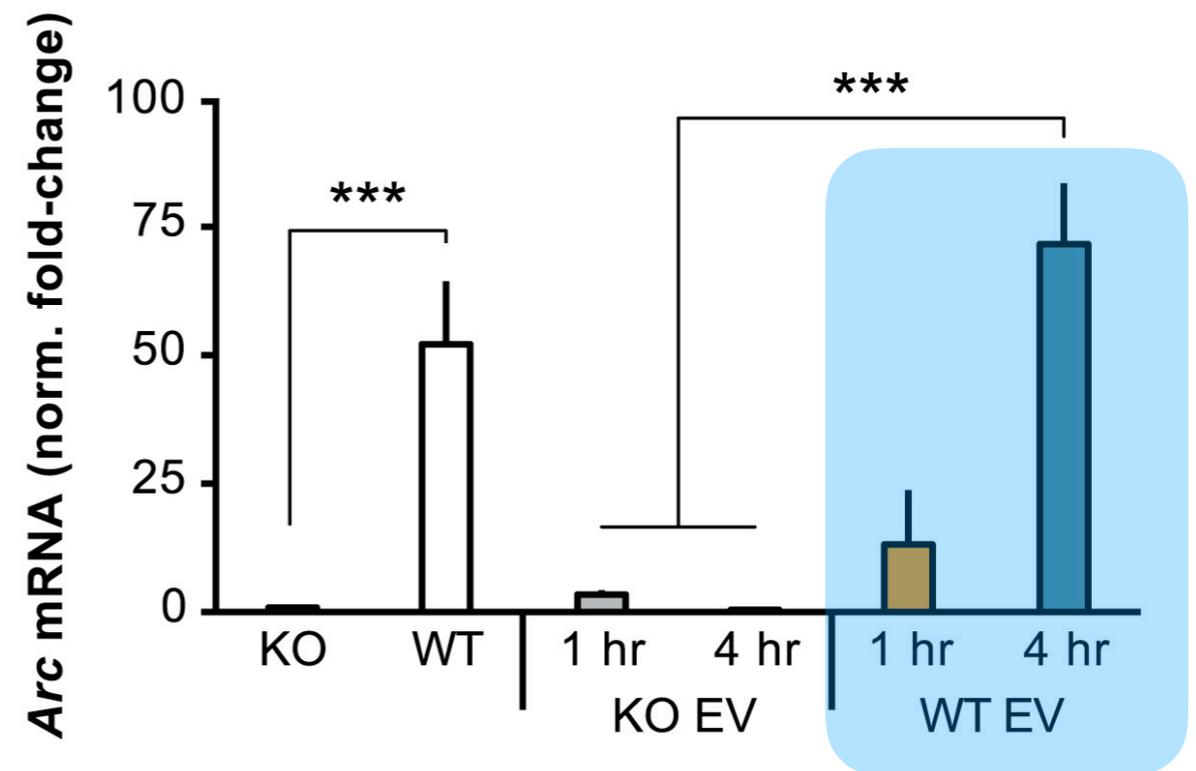
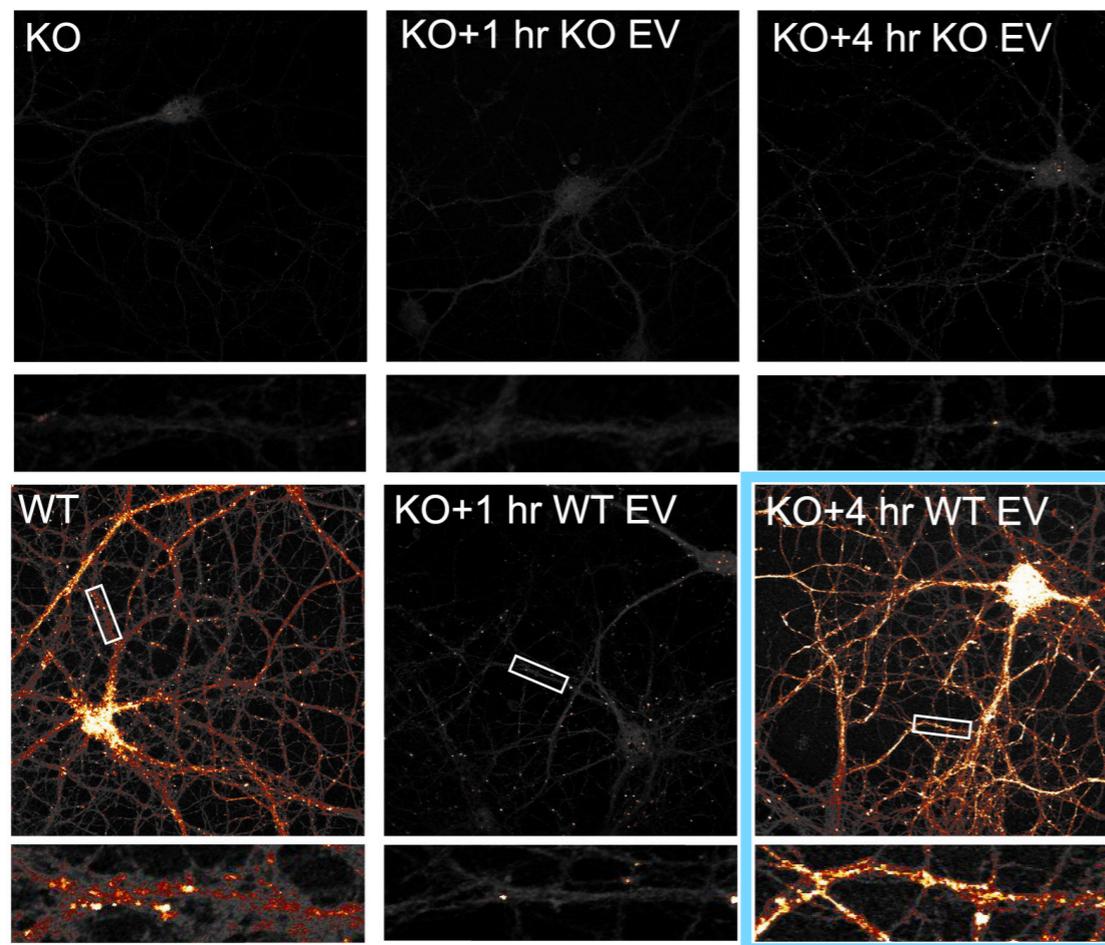


# Is capsid formation required to be taken into neurons?



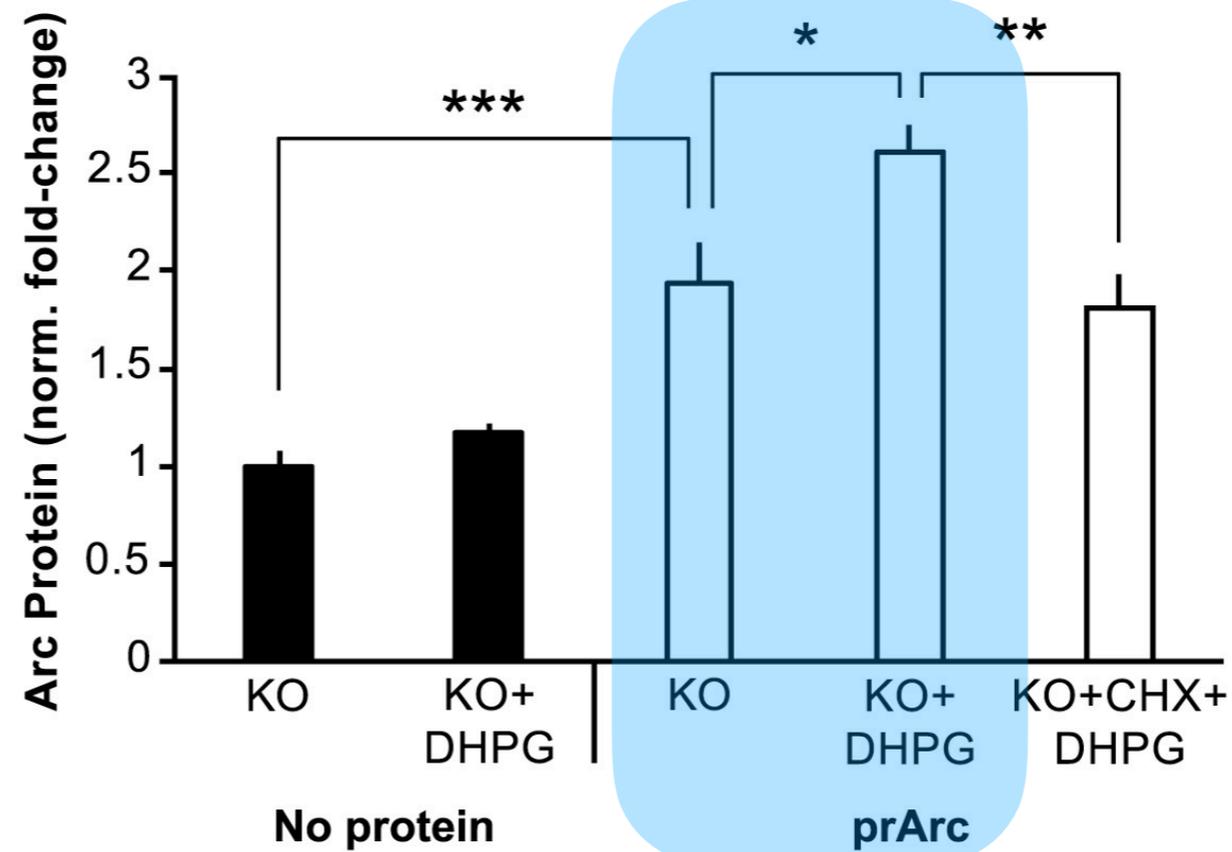
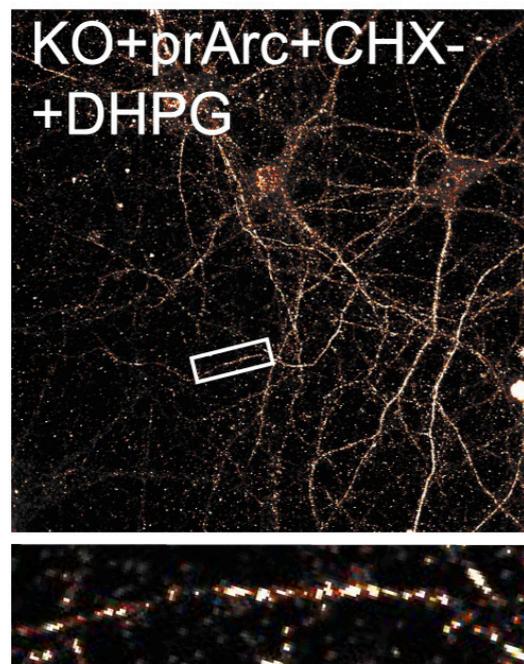
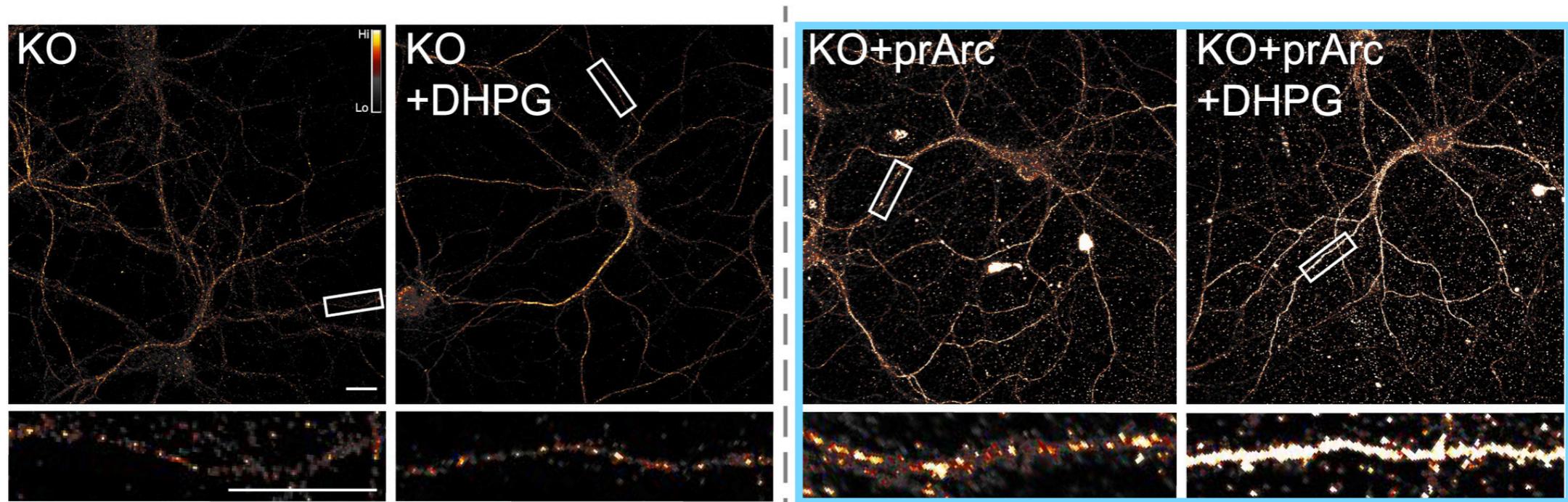
**Capsid formation is required for transfer of mRNA to neurons**

# Can Arc transfer mRNA using **extracellular vesicles**?



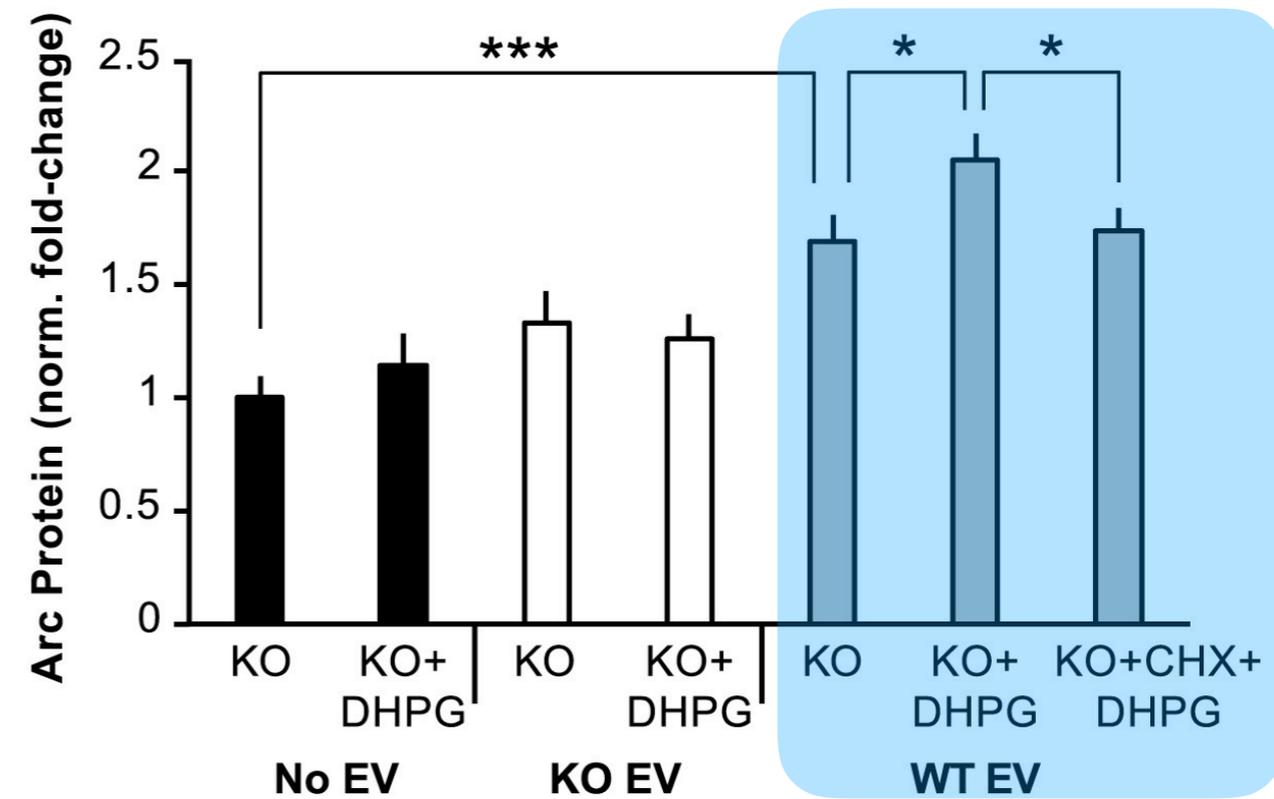
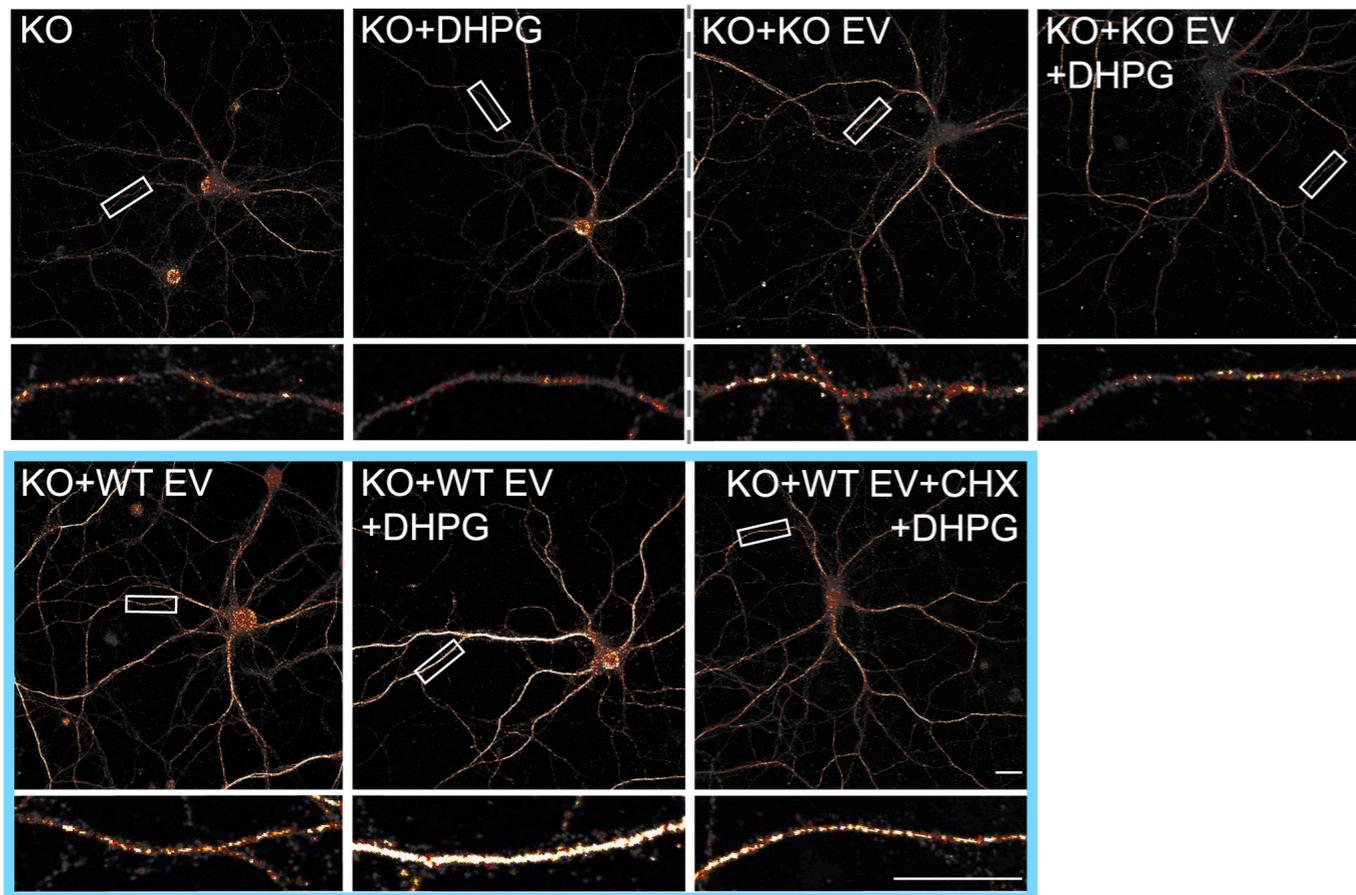
**Arc mRNA transferred to KO neurons**

# Is translation occurring after transfer to a neuron?

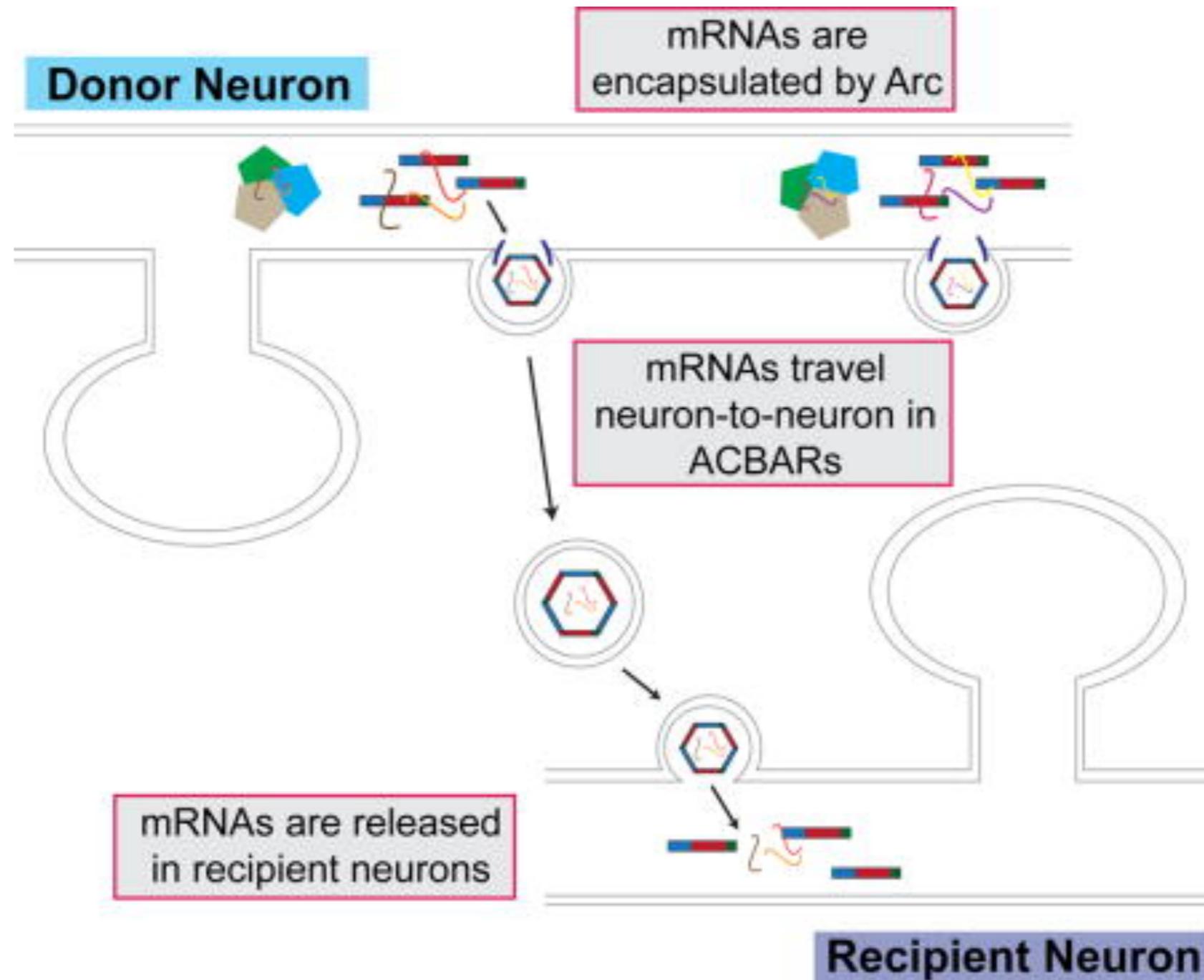


**mRNA is translated once transferred to a neuron**

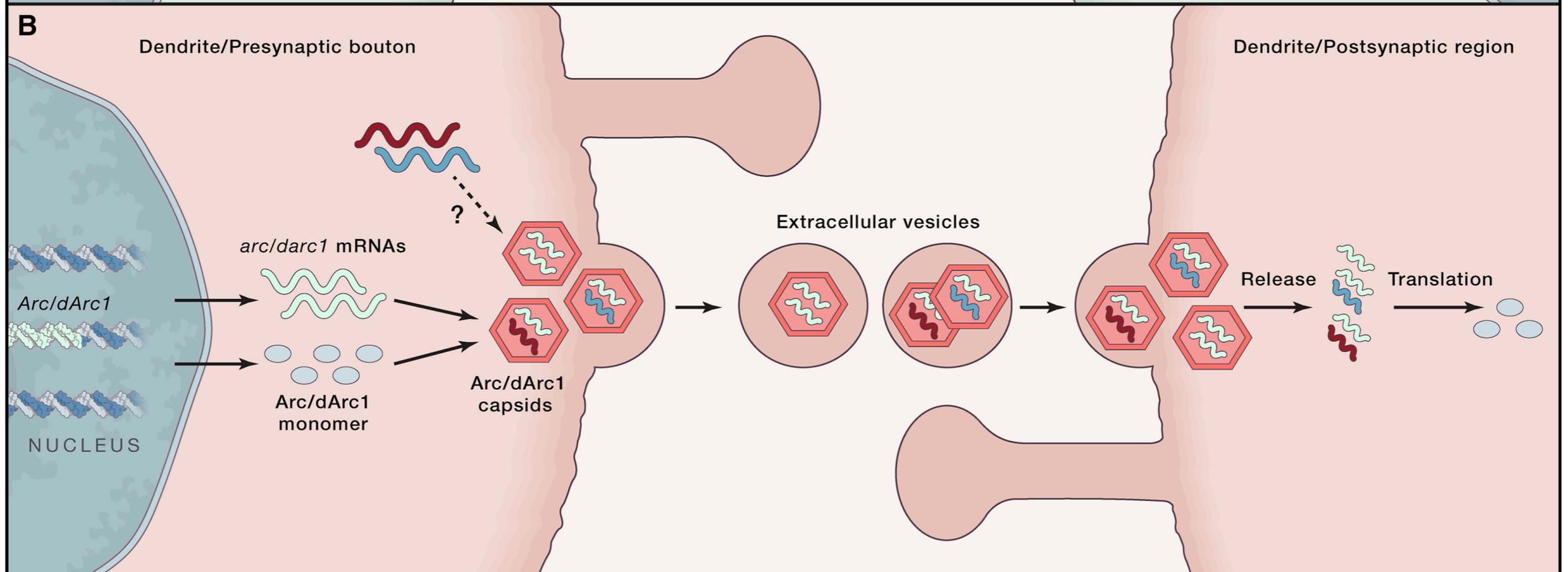
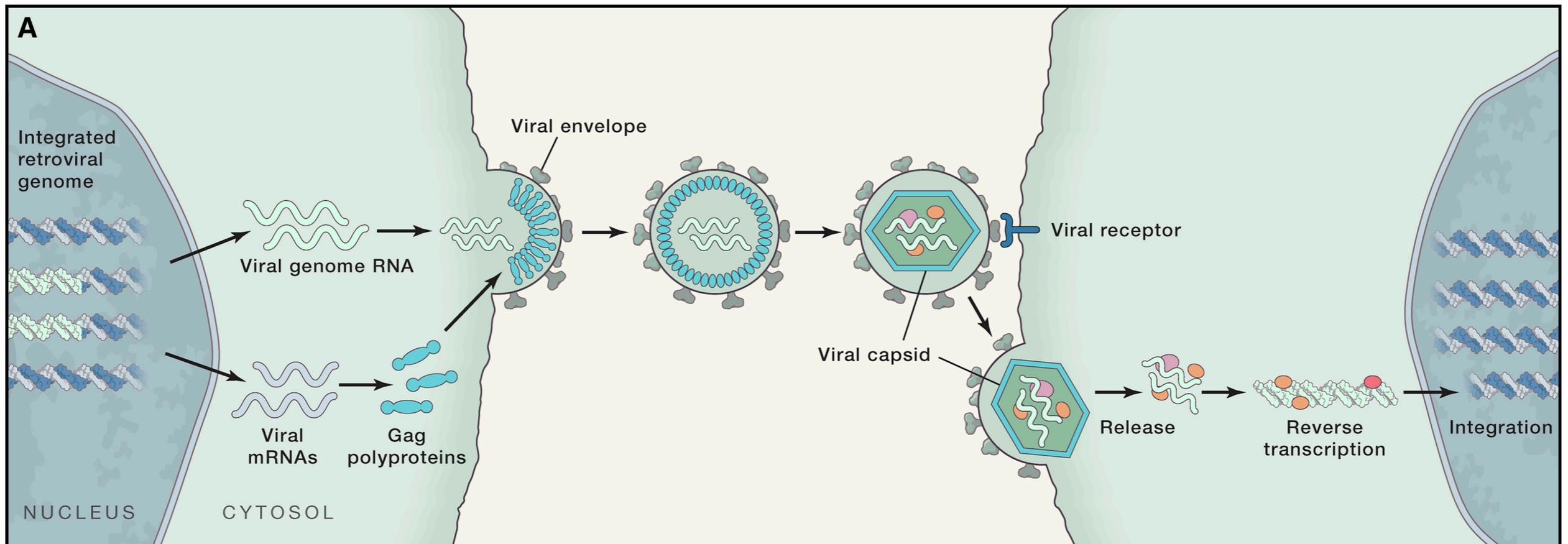
# Is translation occurring after transfer to a neuron?



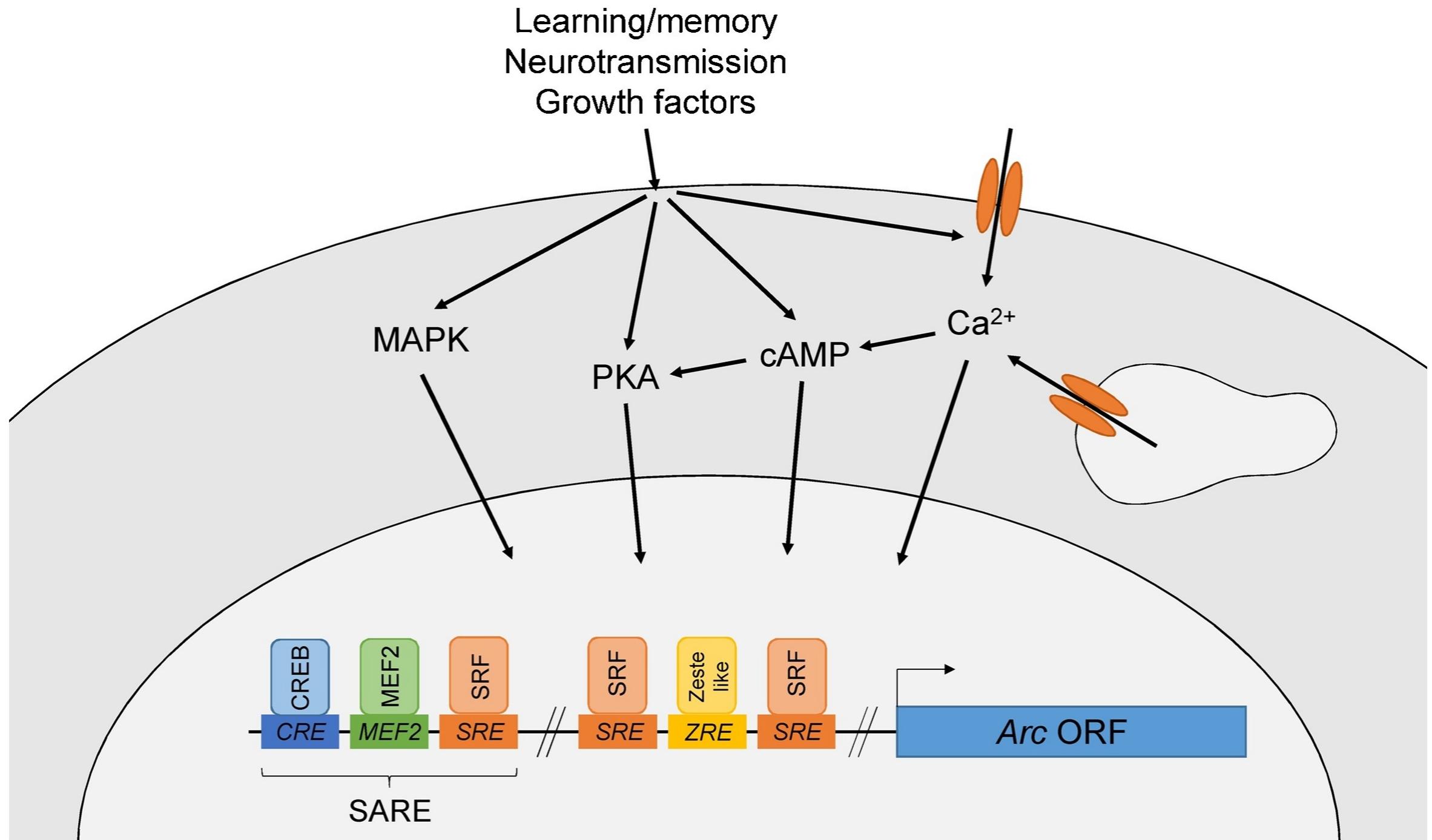
# Transfer of Arc mRNA



# Summary



# Additional Images for Questions



# References

Carmichael, RE, Henley, JM. (2018) Transcriptional and post-translation regulation of Arc in synaptic plasticity. Retrieved from: <https://www.sciencedirect.com/science/article/pii/S1084952117303877?via%3Dihub>

<https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/protein-classification/what-are-protein-families>

<https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/what-are-protein-signatures/signature-types>

<http://smart.embl-heidelberg.de/>

<https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/what-are-protein-signatures/signature-types/what-ar-2>

[https://prosite.expasy.org/prosite\\_details.html](https://prosite.expasy.org/prosite_details.html)

<https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/what-are-protein-signatures/signature-types/what-ar-0>

[http://smart.embl-heidelberg.de/smart/show\\_motifs.pl](http://smart.embl-heidelberg.de/smart/show_motifs.pl)

<https://www.ebi.ac.uk/training/online/course/introduction-protein-classification-ebi/what-are-protein-signatures/signature-types/what-are->

<https://pfam.xfam.org/family/PF18162.1#tabview=tab7>