



UNIVERSITY OF TRENTO - Italy

**International Doctoral School in Biomolecular  
Sciences**

**XXV Cycle**

**“Understanding gene expression with a  
pore forming toxin”**

**Tutor**

Alessandro Quattrone

*Laboratory of Translational Genomics, Center for Integrative Biology, University of Trento-Italy*

**Advisor**

Gabriella Viero

*Institute of Biophysics- CNR Trento Italy & Laboratory of Translational Genomics, Center for Integrative Biology, University of Trento-Italy*

**Ph.D. Thesis of**

Massimiliano Clamer

Academic Year 2012-2013



## Declaration

The work described in this thesis was carried out at the Centre for Integrative Biology (CIBIO), University of Trento (first chapter) in the laboratory of Prof. Alessandro Quattrone and in the Chemistry Research Laboratory of Prof. Hagan Bayley, University of Oxford (second and third chapters) between September 2009 and September 2013. I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged. This work has not been submitted previously for any other degree at the University of Trento or Oxford or any other universities.

Massimiliano Clamer

September 2013



## Acknowledgments

Massive thanks go to my supervisor Dr. Gabriella Viero for mentoring and teaching me during all these four years, for the many valuable discussions, for the great time and results achieved together and for cheering me up during hard times.

I would like to thank Prof. Alessandro Quattrone (University of Trento) and Prof. Hagan Bayley (University of Oxford) for the opportunity to work in their labs. I would like to thank Prof. Hagan Bayley, for giving me the opportunity to discover and work on interesting and challenging projects. I appreciate his interest in my work and his understating toward my “entrepreneurial feeling”. I feel privileged to have worked in such dynamic lab.

Thanks must also go to Dr. Toma Tebaldi (University of Trento) and Dr. Lajos Höfler (University of Oxford) for their help with the system analysis on microarray data and with the script for the analysis of the electrophysiological data respectively. I am very grateful for their guidance and help. I’m honored to have worked with such talented scientists/friends.

I would like to thank past and present members of the Quattrone’s and Bayley’s groups and all the PhD students of XXV cycle of the *Doctoral school in bimolecular sciences* for having made these years thoroughly enjoyable.

Thanks to the CARITRO fellowship for the first two years financial support in Trento. I am grateful to the DEMATTE’ foundation for financial support during my first year in Oxford (Sep 2011 – Sep 2012), through my DEMATTE’ fellowship for international training, in partnership with the financial support from prof. Hagan Bayley for my second year (Sep 2012 – Sep 2013).

Finally, thanks to my friends and family for all their support over the years.



Dedication

To my beautiful wife Valentina.



## TABLE OF CONTENTS

<b>Compendium.....</b>	<b>1</b>
------------------------	----------

<b>Chapter 1: Treatment with staphylococcal <math>\alpha</math>-hemolysin results in host genome-wide translome variations and recruitment of RNA binding proteins.....</b>	<b>5</b>
---	----------

<b>1.1 Introduction.....</b>	<b>6</b>
------------------------------	----------

1.1.0 Overview.....	6
---------------------	---

1.1.1 Host-pathogen interactions.....	8
---------------------------------------	---

1.1.2 Staphylococcus aureus alpha hemolysin ( $\alpha$ HL).....	10
---	----

1.1.3 Post-transcriptional mechanisms involved in the cap-dependent translation initiation .....	12
--	----

1.1.4 The mTOR pathway.....	13
-----------------------------	----

<b>1.2 Results and discussion.....</b>	<b>15</b>
--	-----------

1.2.1 Definition of the Sublytic doses of AHL.....	15
--	----

1.2.2 Sublytic AHL redirects factors controlling translation.....	18
---	----

1.2.3 The host cell translation is hijacked in opposite directions.....	23
---	----

1.2.4 An almost exclusive translational response dictates an overall reshaping of host gene expression.....	25
---	----

1.2.5 Biological themes of uncoupling: ontological enrichments reveal a translational-specific response to sublytic AHL toward the plasma membrane.....	28
---	----

1.2.6 Post-transcriptional trans-acting factors involved in AHL translational specific response.....	30
--	----

1.2.7 The recruitment of ELAVL1 is pore-independent.....	34
--	----

<b>1.3 Conclusion.....</b>	<b>37</b>
<b>1.4 Methods.....</b>	<b>45</b>
1.4.1 Cell culture.....	45
1.4.2 Adenilate Kinase assay.....	45
1.4.3 Determination of hemolytic activity.....	45
1.4.4 Protein extraction and immunoblotting.....	46
1.4.5 Methyl- $\beta$ CD treatment .....	47
1.4.6 Recombinant protein expression.....	47
1.4.7 Fluorescence microscopy.....	47
1.4.8 De-novo protein synthesis assay.....	48
1.4.9 Polysomal profiling.....	48
1.4.10 RNA extraction .....	49
1.4.11 Quantitative Real time-PCR (q-PCR).....	50
1.4.12 Microarray.....	50
1.4.13 Gene Ontology Enrichment analysis.....	51
1.4.14 AURA enrichment analysis.....	52
1.4.15 FACS.....	52
1.4.16 Migration assay.....	52
1.4.17 Lipid extraction .....	53
1.4.18 Planar lipid bilayer recording .....	53
1.4.19 List of TaqMan probes and primers used.....	54
<b>1.5 References.....</b>	<b>56</b>

<b>Chapter 2: Detection of 3'-end RNA uridylation with a protein nanopore.....</b>	<b>64</b>
<b>2.1 Introduction .....</b>	<b>65</b>
2.1.0 Overview.....	65
2.1.1 RNA 3'-uridylation.....	66
2.1.2 The <i>state-of-art</i> on single molecule RNA detection.....	67
2.1.3 Biological nanopores for nucleic acid detection.....	68
2.1.4 Planar (PLM) and Dropled (DIB) interface bilayers.....	70
<b>2.2 Result and discussion.....</b>	<b>73</b>
2.2.1 Selective RNA sequence detection.....	73
2.2.2 Effect of the ionic strength on the binding.....	83
2.2.3 The RNA binding site.....	84
2.2.4 Binding modularity.....	86
2.2.5 Proof of translocation.....	88
2.2.6 The $\alpha$ HL pore as a detector of 3'-end uridylation length and post-transcriptional modifications.....	92
2.2.7 Purification of defined RNA signature fragments.....	100
<b>2.3 Conclusion.....</b>	<b>104</b>
<b>2.4 Methods.....</b>	<b>105</b>
2.4.1 Single channel recordings.....	105
2.4.2 Data analysis.....	105
2.4.3 <i>In vitro</i> protein expression .....	107

2.4.4 Mutagenesis .....	107
2.4.5 Thermodynamic analysis.....	108
2.4.6 Periodate oxidation.....	108
2.4.7 HPLC.....	109
2.4.8 MALDI-MS.....	109
2.4.9 RNA sequences.....	109
2.4.10 BindN analysis.....	110
<b>2.5 References.....</b>	<b>112</b>

<b>Chapter 3: A step forwards the sequential recognition of RNA by custom RNA-binding nanopores.</b> .....	<b>120</b>
<b>3.1 Introduction</b> .....	<b>121</b>
3.1.0 Overview.....	122
3.1.1 A new challenge: RNA detection by modular recognition.....	122
3.1.2 A doughnut shape for RNA capture: Lsm proteins.....	124
3.1.3 The archaea <i>Haloferax volcanii</i> Lsm.....	127
<b>3.2 Results and discussion</b> .....	<b>128</b>
3.2.1 Design and expression .....	128
3.2.2 Oligomerization and activity .....	131
3.2.3 Electrophysiological characteristics of LadyA and LadyB.....	133
3.2.4 RNA binding: biochemical evidence.....	138
3.2.5 RNA binding: electrophysiological characterization.....	140
3.2.6 Discrimination of the Lsm binding signal.....	150
3.2.7 Detection of RNAs with complex secondary structure.....	152
<b>3.3 Conclusion</b> .....	<b>156</b>
<b>Appendix of chapter 3: The PUF case</b> .....	<b>158</b>
<b>A.1 Introduction</b> .....	<b>159</b>
A.1.1 PUF repeats.....	159
<b>A.2 Results and discussion</b> .....	<b>162</b>

<b>A.3 Conclusion.....</b>	<b>163</b>
<b>3.4 Methods.....</b>	<b>165</b>
3.4.1 Gene sequences.....	165
3.4.2 <i>In vivo</i> protein expression and purification.....	167
3.4.3 Enterkinase digestion .....	168
3.4.4 Hemolytic assay.....	168
3.4.5 <i>In vitro</i> protein expression.....	169
3.4.6 Chimera oligomers preparation.....	169
3.4.7 Radioactive labeling of ssRNAs.....	170
3.4.8 EMSA assay.....	170
3.4.9 Single channel recordings.....	171
3.4.10 Data analysis.....	171
3.4.11 Gel trypsin digestion.....	172
3.4.12 MALDI-MS.....	173
<b>3.5 References.....</b>	<b>174</b>
<b>Concluding Remarks.....</b>	<b>178</b>

## List of Figures and Illustrations

### Chapter 1.

<b>Figure 1.1</b> Cellular responses to PFTs.....	9
<b>Figure 1.2</b> The $\alpha$ -hemolysin. ....	11
<b>Figure 1.3</b> $\alpha$ HL insertion in a lipid bilayer. ....	11
<b>Figure 1.4</b> Cap-dependent translation initiation.....	14
<b>Figure 1.5</b> Definition of AHL sublytic doses in human cell lines.....	17
<b>Figure 1.6</b> Effect of sublytic doses of AHL on the cell cycle of HeLa, CHP-134 and SH-SY5Y.....	18
<b>Figure 1.7</b> Sublytic doses of AHL induce mTor pathway activation...	19
<b>Figure 2.8</b> Time course of Akt (A) and eIF4E (B) phosphorylation after treating SH-SY5Y with 3 nM, 6 nM and 12 nM of native AHL..	20
<b>Figure 2.9</b> Time course of Akt (A) and eIF4E (B) phosphorylation after treating CHP-134 with 3 nM, 6 nM and 12 nM of native AHL...	20
<b>Figure 1.10</b> NMR spectra of the total lipid extracts for SH-SY5Y (A) and CHP-134 (B) .....	22
<b>Figure 1.11</b> Effect of sublytic doses of AHL on translation.....	24
<b>Figure 1.12</b> Genome-wide gene expression analysis reveals a huge reshaping of the translational level of gene expression. ....	27
<b>Figure 1.13</b> qRt-PCR of selected gene for microarray validation.....	28
<b>Figure 1.14</b> The uncoupling between transcriptome and translome reflects enrichments of specific biological themes.....	29
<b>Figure 1.15</b> Sublytic doses of AHL recruit post-translational trans-acting factors.....	32
<b>Figure 1.16</b> Monomers of AHL are responsible for Akt activation.....	34
<b>Figure 1.17</b> Oligomerization ability: rWT-AHL vs H <sub>35</sub> N.....	36

## Chapter 2.

<b>Figure 2.1</b> Biological nanopores for RNA/DNA sequencing.....	70
<b>Figure 2.2</b> Selective RNA sequence detection. ....	76
<b>Figure 2.3</b> C <sub>3</sub> U <sub>7</sub> does not affect the current of the WT-αHL pore from the <i>trans</i> .....	77
<b>Figure 2.4</b> Ion current traces showing C <sub>5</sub> U <sub>5</sub> interacting with the WT-αHL pore at voltages. ....	77
<b>Figure 2.5</b> Effect of the 3' ribose structure on the binding.....	79
<b>Figure 1.6</b> ssRNAs binding in the WT-αHL pore and effect of the ionic strength of the buffer.....	81
<b>Figure 2.7</b> Periodate oxidation of C <sub>3</sub> U <sub>7</sub> assayed by MALDI mass spectrometry and RP-HPLC.....	82
<b>Figure 2.8</b> The RNA binding pocket in the WT-αHL nanopore.....	86
<b>Figure 2.9</b> Binding modularity.....	88
<b>Figure 2.10</b> RNA current blockades arise from the translocation of the RNA through WT-αHL pore.....	90
<b>Figure 2.11</b> Examples of blocking events observed with folded RNAs translocating through the αHL pore.....	92
<b>Figure 2.12</b> The αHL pore as a sensor for different RNA-uridylylations.....	95
<b>Figure 2.13</b> Progressive additions of RNAs with different U-tails length.....	96
<b>Figure 2.14</b> All data points scatter plot I <sub>RES%</sub> vs dwell time for a typical experiment with C <sub>5</sub> U <sub>5</sub> , C <sub>4</sub> U <sub>6</sub> , C <sub>3</sub> U <sub>7</sub> , U <sub>10</sub> , U <sub>20</sub> in the <i>cis</i> chamber.....	97
<b>Figure 2.15</b> Detection of A <sub>15</sub> U <sub>x</sub> (x < 5) with the D128N-αHL pore	

and comparison with the WT- $\alpha$ HL pore.....	99
<b>Figure 2.16</b> Protocol for RNA purification and analysis.....	101
<b>Figure 2.17</b> Effects of neomycin and incubation times on the final yield.....	103

## Chapter 3.

<b>Figure 3.1</b>	Examples of fused RNA-binding proteins or domains....	123
<b>Figure 3.2</b>	Views of an archea Lsm.....	125
<b>Figure 3.3</b>	Examples of resolved Lsm structures.....	126
<b>Figure 3.4</b>	LadyB design.....	129
<b>Figure 3.5</b>	<i>In vivo</i> protein expression.....	130
<b>Figure 3.6</b>	MADLDI-MS of LadyB monomers.....	131
<b>Figure 3.7</b>	SDS-PAGE of in vitro expressed chimeric proteins.....	132
<b>Figure 3.8</b>	Hemolytic assay.....	132
<b>Figure 3.9</b>	Insertion of LadyA oligomers and permanent block of the channel after insertion.....	133
<b>Figure 3.10</b>	Pore insertion in PLM and conductance distribution of LadyB and WT- $\alpha$ HL.....	134
<b>Figure 3.11</b>	Mean conductance of a single-channel il PLM at different pH. G: conductance .....	134
<b>Figure 3.12</b>	Electrophysiological characterization of LadyB.....	135
<b>Figure 3.13</b>	LadyA vs ladyB: effect of linker length and the mutations T62C, A75C, N76A on the Vrev.....	136
<b>Figure 3.14</b>	LadyB modeling.....	137
<b>Figure 3.15</b>	Determination of the $K_D$ of binding in bulk.....	139
<b>Figure 3.16</b>	EMSA assay.....	140
<b>Figure 3.17</b>	Detection of a polyU <sub>20</sub> ssRNA.....	142
<b>Figure 3.18</b>	Voltage dependence of the blockades for U <sub>20</sub> translocation.....	143
<b>Figure 3.19</b>	Behavior of the blocking events.....	144
<b>Figure 3.20</b>	A zoom in on the signals blockades.....	145

<b>Figure 3.21</b> Effect of the U on ssRNA detection .....	147
<b>Figure 3.22</b> RNA A10C10 titration.....	148
<b>Figure 3.23</b> LadyB <sup>-</sup> RNA sensing in single channel recoding.....	150
<b>Figure 3.24</b> Single channel recording: <i>in vitro</i> vs <i>in vivo</i> protein expression.....	151
<b>Figure 3.25</b> Model of RNA binding for LadyB.....	152
<b>Figure 3.26</b> Mfold structure prediction.....	153
<b>Figure 3.27</b> Circular dichroism analysis of the MGA-U RNA.....	153
<b>Figure 3.28</b> Effect of 2M urea on the recognition of ssRNAs.....	155
<b>Figure 3.29</b> Nucleic acid sequences LadyB.....	165
<b>Figure 3.30</b> Nucleic acid sequence of $\alpha$ HL-(PUF) <sub>2</sub> chimeric protein.....	166
<b>Figure 3.31</b> Plasmids.....	167

### Chapter 3 APPENDIX

<b>Figure A.1</b> PUF repeats. ....	160
<b>Figure A.1.2</b> Design of LadyB and PUF-RBnP: an overview.....	161
<b>Figure A.2.3</b> Oligomerization and insertions in a planar lipid bilayer.....	163

## List of Tables

### Chapter 1

<b>Table 1.1</b> List of qRT-PCR primers. ....	54
--	----

### Chapter 2

<b>Table 2.1</b> Effect of the salt concentration on the oligo U binding.....	83
---	----

<b>Table 2.2</b> Mean dwell time and residual current (%) of different RNA oligos detect by WT- $\alpha$ HL.....	94
--	----

<b>Table 2.3</b> Mean dwell time and residual current (%) of different RNA oligos detect by D128N- $\alpha$ HL.....	97
---	----

<b>Table 2.4</b> Thermodynamic analysis.....	108
--	-----

<b>Table 2.5</b> RNA sequences.....	110
-------------------------------------	-----

### Chapter 3

<b>Table 3.1</b> Proteins that bind ssDNA or ssRNA in a modular fashion.....	125
--	-----

<b>Table 3.2</b> Dwell time and residual current for different RNA interacting with the LadyB.....	146
--	-----

<b>Table 3.3</b> Characterization of the different behaviors: LadyB pore vs WT- $\alpha$ HL pore. ....	149
--	-----

## Compendium

This thesis aimed to explore eukaryotic cellular processes upon the virulent attack of low doses of a well-known pore forming toxin (staphylococcal  $\alpha$ -hemolysin ( $\alpha$ HL)) and to develop a new biotech application using the same protein.

As an important crossroad in gene expression, RNA is involved in modulation, transport and temporary storage of biological information. Despite the similarities with DNA, RNA has different and unique mechanical, chemical and biological properties. At the biological level, RNA can perform catalytic functions (e.g. ribosomes and spliceosome) and regulate the expression of the encoded information at different levels (e.g. mRNA, microRNA, siRNA). During the flow of gene expression, the final choice of a messenger RNA (mRNA) to be translated, after transcription, is a dynamic process that involves balanced forces between transcription and translation<sup>1-3</sup> and ends in ribosome recruitment. Nevertheless, there are still open questions concerning the ability of the cell to reshape, independently from transcription, the process of mRNA loading onto polysomes, the complex machineries in charge for protein synthesis<sup>4</sup>.

The first chapter of the thesis involved a cell biology and omic analysis of the non-immune response of mammalian cells to sublytic doses of a bacterial pore forming toxin,  $\alpha$ -hemolysin ( $\alpha$ HL). We studied the reshaping of gene expression at the transcriptional and translational level after toxin stimulation. We characterized the changes of gene expression combining a genome-wide analysis and a biochemical characterization of the main protein factors involved in translation initiation for the ribosome recruitment and consequent polysomes formation. We observed a strong mRNA hijacking (i.e. selection of specific mRNAs) at the translational level and an extensive uncoupling between transcriptome and translome, together with the involvement of the RNA binding protein ELAVL1, a trans-acting factor, known to regulate post transcriptional events. Gene Ontology analysis of differentially expressed genes proved that the cellular response is compartment-specific and characterized by a boost for translation of membrane proteins. In particular we focused our attention on the membrane mechanism of protein action, finding that the modulation of both the translational effectors and the RNA binding protein was pore-independent and lipid composition-

dependent. In this chapter we tried to shed light into a completely new host-pathogen interaction from multiple points of views, providing new understandings of the biological relationship between host and virulence factors.

A conceptually different but semantically connected task to the first chapter topic is the relevance of RNA signatures and coordinate networks of RNA binding protein in the regulation of gene expression. A specific RNA, or a group of RNAs under a similar regulatory system, can be detected by peculiar sequence signatures. These RNA signatures are secondary structure and/or simple sequence strands that are believed to dictate specific combination with trans-acting elements (miRNA, RNA binding proteins), rationalizing the cell biology of gene expression. The development of new technologies for the identification and monitoring of RNAs signatures without needs of labeling, amplification and at low cost is still an unmet need. Few technologies are available for single molecule detection of nucleic acids, but little work has been done on the selective detection of specific RNA sequences<sup>5-7</sup>. In the second chapter the  $\alpha$ HL pore was used as a sequence-specific RNA sensor. We used the toxin as a single molecule sensor for 3'-end RNA uridylation, a recently characterized non-templated post-transcriptional modification<sup>8</sup>. We performed an extensive electrophysiological characterization of the RNA-nanopore interaction using single channel recordings and mutagenesis approaches. The addition of a variable numbers of uridines in 3' of RNAs affects their turnover, biogenesis *in vivo*. The quantification of this modification in term of overall abundance in the cell and length distribution on single RNAs is still unclear<sup>9</sup>. Moreover, currently available sequencing technologies are not optimal for a rapid and cheap detection of this RNA feature. Here, we demonstrated the selective, transient and modular binding of 3' poly-uridylated ssRNAs inside the  $\beta$ -barrel of the  $\alpha$ HL. Our results illustrated the selective detection of a biologically relevant sequence signature using short and unstructured RNAs as model. Our findings are of particular interest, because stochastic nanopore sensing is a convenient method for rapid, label-free, PCR-amplification free and low cost detection of RNA modification. We reported a number of "proof of principle" experiments for a prospective application of this RNA sensor in the detection of the 3'-end uridylation of short single stranded RNAs.

In the third chapter of the thesis we then tried to implement the RNA-binding property of the  $\alpha$ HL pore using the nanopore as a framework to engineer chimeric RNA-binding nanopores (RBnPs) able to sense more complex (in sequence and structure) RNAs. In this chapter different approaches of possible custom RNA-binding nanopores were presented and characterized using the same experimental setup used in the previous part and, finally, possible new directions for nanopore RNA sensing were suggested.

In conclusion, in this work we explored eukaryotic cellular processes during host-virulence factor interaction, we discovered a hitherto undisclosed RNA-binding property of a well-studied pore forming toxin and we engineer new custom RNA binding nanopores developing a new biotech application with a unique common biological tool: the staphylococcal  $\alpha$ -hemolysin.

1. Sharp, P. a. The centrality of RNA. *Cell* **136**, 577–80 (2009).
2. Shoemaker, C. J. & Green, R. Translation drives mRNA quality control. *Nature structural & molecular biology* **19**, 594–601 (2012).
3. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
4. Tebaldi, T. *et al.* Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. *BMC genomics* **13**, 220 (2012).
5. Zhang, J. *et al.* Rapid and label-free nanomechanical detection of biomarker transcripts in human {RNA.}. *Nature nanotechnology* **1**, 214–20 (2006).
6. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nature nanotechnology* **5**, 807–14 (2010).
7. Wang, Y., Zheng, D., Tan, Q., Wang, M. X. & Gu, L.-Q. Nanopore-based detection of circulating {microRNAs} in lung cancer patients. **6**, 668–74 (2011).
8. Scott, D. D. & Norbury, C. J. RNA decay via 3' uridylation. *Biochimica et biophysica acta* 1–12 (2013). doi:10.1016/j.bbagr.2013.01.009
9. Choi, Y. S. U. N. S., Patena, W., Leavitt, A. D. & McManus, M. T. Widespread RNA 3'-end oligouridylation in mammals. *RNA (New York, N.Y.)* **18**, 394–401 (2012).



## CHAPTER 1

Treatment with staphylococcal alpha-hemolysin results in host genome-wide translome variations and recruitment of RNA binding proteins

## 1.1 Introduction

### 1.1.0 Overview

The host-pathogen interplay represents a complex and dynamic biological system that depends on the pathogen activity and on the host ability to reply. Bacterial pathogens can activate a wide variety of responses in host cells by the secretion of groups of interfering biological macromolecules<sup>1,2</sup>. Although bacterial pathogens have evolved a plethora of strategies to subvert host defense functions, several general and common mechanisms can be identified<sup>1,2</sup>.

One of the most ancient form of attack exerted by bacterial virulence factors is the formation of proteinaceous pores that cross the plasma membrane<sup>3-5</sup>. These proteins, called Pore Forming Toxins (PFTs), are widely used by bacteria, plants, sea anemones or earthworms as powerful attack and defense weapons. All PFTs share a common multi-step mechanism of action: (i) release of the soluble monomers, (ii) binding of monomers to the target membrane requiring or not a protein receptor, (iii) oligomerization in a non-lytic pre-pore, (iv) insertion of the pore forming protein portion into the lipid bilayer<sup>3</sup>. The final pore formation leads, at certain concentrations, to osmotic imbalance and, finally, to cell death. Despite this mechanism of action, at lower doses of PFTs cellular responses have been observed to be in some cases extremely diverse<sup>6</sup>. Therefore, when cell death doesn't occur and the toxin concentration is low or the exposure time is short, target cells may recover and respond to the damage with several cellular responses. For a number of PFTs it has been shown that after pore formation, target cells are able to restore the plasma membrane integrity. Cultured cells and tissues respond to a membrane damage with mechanical survival mechanisms by using different calcium dependent strategies<sup>7,8</sup>. At the molecular level, effects on metabolic pathways, intracellular signalling, proteasome activity, transcription<sup>6,9</sup> and translation<sup>9-11</sup> have been demonstrated. Additionally, regulation of cell proliferation<sup>12,13</sup> and activation of several pathways related to stress response such as p38 and JNK MAPK<sup>14-17</sup> have been reported. Nevertheless, very

few studies have portrayed at the genome-wide level the global picture of the changes in translation and transcription occurring as a host response to virulent attacks<sup>16</sup>.

Translational regulation mediated by RNA binding proteins (RBPs) and non coding RNAs (ncRNAs), is increasingly recognized as a crucial layer of control, capable of reshaping and diversifying changes in mRNA production rates introduced by transcriptional regulation. Recent experiments have shown that variance in protein levels can be explained more by translation rates than by transcription rates<sup>18,19</sup>. The comparison between transcriptome (i.e. all the mRNAs in the cells) and translome (i.e. the mRNAs in active translation) has proved to be an useful way to reveal translation controls in different conditions<sup>20</sup>. It is therefore of utmost importance to investigate the genome-wide translational response to PFTs to unveil how host cells react and eventually resolve the “crisis” triggered by these toxins employing translational control mechanisms. This is particularly true when considering the faster dynamics associated to translational regulation in comparison to transcriptional regulation. Indeed a sudden stress such as the one generated by PTFs can be more effectively addressed by rapid changes in protein translation, resulting in an immediate outcome. This approach was previously performed on macrophages and dendritic cells exposed to LPS<sup>21,22</sup>, and on cells infected with hepatitis C virus<sup>23</sup>, resulting in the discovery of multiple novel translationally-regulated genes involved in response processes.

To date, any knowledge on the role played by translational regulation of gene expression in a host cell responding to both lytic and sublytic doses of PFTs is available. The most important paradigm of all PFTs is the staphylococcal exotoxin  $\alpha$ -hemolysin ( $\alpha$ HL), a PFT secreted by *Staphylococcus aureus*. *S. aureus* is a major cause of community-associated methicillin resistant infections, accounting for 20,000 deaths/year in the United States<sup>24,25</sup>. Being  $\alpha$ HL the primary causative agent for an extraordinary multitude of host-cell responses<sup>26-30</sup>, it represents a virulent-factor of particular interest.  $\alpha$ HL is released by *S. aureus* as a 33.2 kDa monomers that oligomerizes into heptameric<sup>31</sup> and/or hexameric<sup>32</sup> transmembrane pores on the host-cell plasma membrane. It causes swelling and cellular lysis in a wide range of concentrations, depending on the target cell<sup>33,34</sup>. Be-

side these effects due to high protein doses, several signalling pathways have been found to be activated in response to pore-formation<sup>6</sup>. A precise definition of low doses and/or sublytic doses of  $\alpha$ HL and the underlying genome-wide impairment at the transcriptional and translational level remains completely elusive and dependent on the cell type. Few studies approached the changes occurring at the transcriptional level by using single genes<sup>35,36</sup> or low-throughput assays<sup>12</sup> which may underestimate and ultimately hide many massive effects of the global cellular response.

Here, we used  $\alpha$ HL as a model of PFTs and a combination of classical cellular biology analysis, genome wide bioinformatics approaches and translational assays to identify and characterize the host cellular response to virulence attacks. This study represents the first attempt to reduce the lack of specific information on the global impact of sublytic concentration of PFTs on the translational control of the host cell. Our results shed light into a completely new host-pathogen interaction level and opens new ways for possible approaches to attenuate  $\alpha$ HL-induced disease.

### **1.1.1 Host-pathogen interactions**

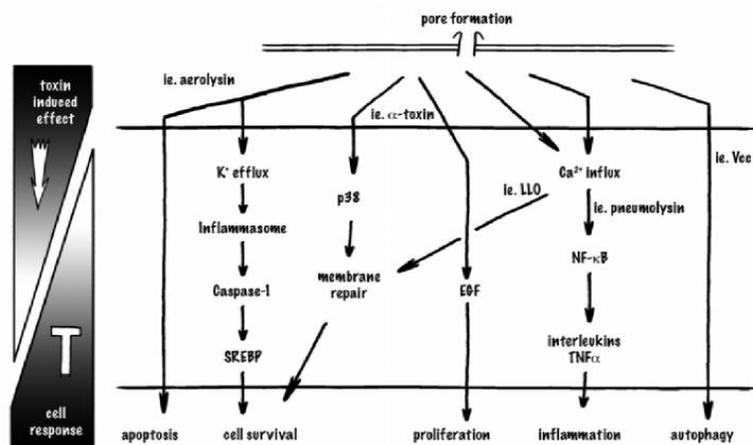
The host-pathogen biology represents a complex interplay and a dynamic system balanced by the pathogen properties as well as the host ability to counteract to infections. During a bacterial or viral infection, pathogenic weapons can be secreted in the surrounding micro-environment. Passing in biological fluids (e.g. blood as well as lymphoid tissues) they can affect tissues far away from the infection site.

Host cells regulatory pathways can be corrupted by secreted pathogen factors. Mechanisms range from constitutive activation of a pathway, to irreversible inactivation of a critical signaling molecule as well as subversion of a whole signaling system to favor the invading pathogen<sup>37</sup>. Examples are (i) the effect of *Bacteroides fragilis* toxin (BFT) on human T cells<sup>38-41</sup>, (ii) the *Pastorella multocita* toxin stimulation of cell growth and proliferation acting on G protein-coupled receptors<sup>42,43</sup>, (iii) the *E.Coli* cytotoxic necrotizing

factor 1 (CNF1) action on Rho family proteins<sup>44</sup> and (iv) the *Vibrio cholerae* cytolysin pro-inflammatory stimulation on human neutrophil granulocytes<sup>45</sup>.

*Staphylococcus aureus* (*S. aureus*) is involved in different diseases such as pneumonia, infective endocarditis, skin infections, arthritis, and septic shock<sup>46</sup>. Methicillin resistant *S. aureus* strains are leading cause of hospital and community infection worldwide<sup>47,48</sup>. Moreover, *S. aureus* seems to be related with the squamous cell carcinoma of the skin<sup>49</sup>.

At the cellular level the plasma membrane is the first barrier between host cells and pathogens. It also represents the source of signal inputs and outputs (clustering of proteins as well as endo- or exocytosis) surrounding the complex adaptive system of gene expression<sup>16</sup>. Bacterial PFTs have been investigated for their ability to interact with the host cell signaling. As example, the *Escherichia coli*  $\alpha$ -toxin activates the Akt/protein kinase B signaling<sup>13</sup>. It has been recently reported that the  $\alpha$ HL can activates the p38 MAPK protein and pro-autophagyc signals<sup>15,50</sup>. While pore formation and cellular lysis are a prominent consequence of  $\alpha$ HL action, recent studies<sup>51</sup> have defined cellular responses to lower intoxication, notably the alteration of cell signaling pathways that govern cell proliferation, inflammatory responses, cytokine secretion, and cell-cell interactions (Figure 1.3) and, as a common mechanism, the reshape of gene expression<sup>5</sup>.

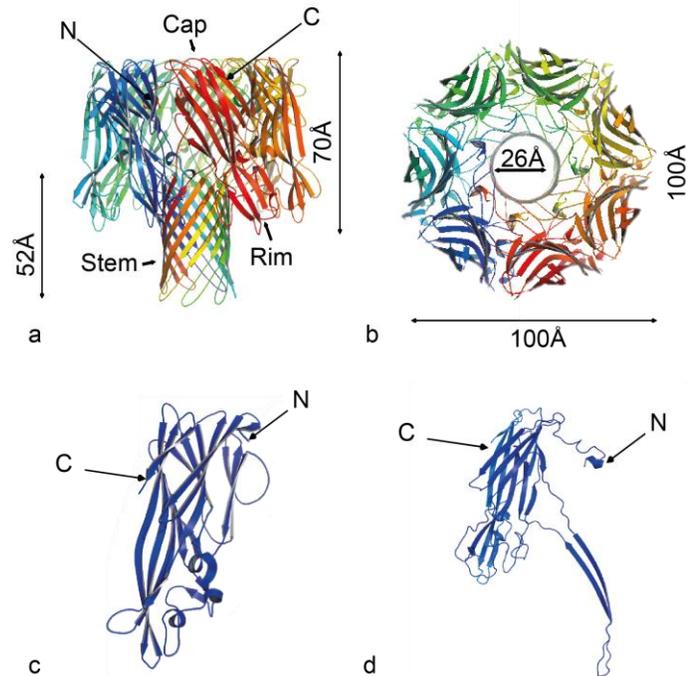


**Figure 1.1 Cellular responses to PFTs.** Different pathways modulated in the host cell by PFTs. EGF, epithelial growth factor; SREBP, sterol responsive element binding protein; LLO, listeriolysin O; Vcc: *Vibrio cholerae* toxin. (reported from Gonzalez M.R. et al., 2008)<sup>5</sup>

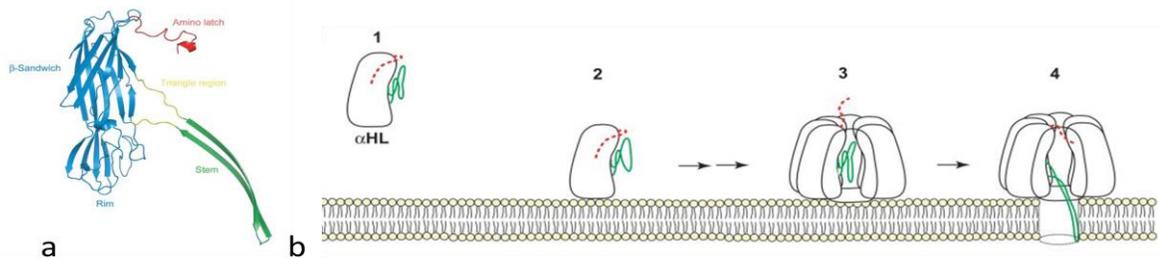
### 1.1.2 *Staphylococcus aureus* alpha-hemolysin (AHL/ $\alpha$ HL)

*Staphylococcus aureus* is a gram positive bacterium, commensal on human skin and involved in several pathological conditions. *S. aureus* produce a broad spectrum of different toxins; one of them is the  $\alpha$ HL, a protein that belongs to the  $\beta$ -pore-forming toxin family ( $\beta$ -PFT). PFTs are proteinaceous virulence factors representing 25–30% of all bacterial protein toxins<sup>52</sup>. The  $\alpha$ HL protein is synthesized and secreted as 33.2-kDa monomer that self-assembles into lipid bilayers forming a transmembrane mushroom shaped pore of nanometer scale<sup>53</sup>. The protein, initially named for its properties as a red blood cell lytic toxin, has a complex action on nucleated cells<sup>51</sup>. The toxin itself does not require any protein to punch a lipid bilayer. A preferential interaction of  $\alpha$ HL in lipids microdomains enriched in cholesterol and sphingomyelin<sup>54,55</sup> has been also suggested. The recent identification of the zinc-dependent metalloprotease ADAM10<sup>56</sup> as a cellular cofactor for  $\alpha$ HL binding has provided keen insight on the biology of the toxin action. This discovery answers to the difference in susceptibility to  $\alpha$ HL-mediated lysis between rabbit erythrocytes (with lysis occurring in the low nanomolar range) and human erythrocytes (with lysis occurring in the high nanomolar to low micromolar range). In fact, the species specificity exhibited by  $\alpha$ -toxin was demonstrated to correlate with ADAM10 expression on rabbit erythrocytes, in contrast to its absence on the surface of the human red cell (Wilke, G. A. & Bubeck-Wardenburg, J. Role of a disintegrin and metalloprotease 10 in *Staphylococcus aureus* alpha-hemolysin-mediated cellular injury. Proc. Natl. Acad. Sci. U.S.A. 107, 13473–13478 (2010)). The heptameric nanopore consists of three domains; (i) a globular cap with an internal vestibule, (ii) a fourteen-stranded, antiparallel  $\beta$  barrel transmembrane (stem) domain (~ 5 nm long and ~ 1.4 nm wide at the point where the vestibule meets the  $\beta$ -barrel) and (iii) a rim domain rich in aromatic residues in contact with the plasmatic membrane (Fig. 1.2). The narrowest part of the pore is in the  $\beta$ -barrel formed by a ring of the residues Glu-111 to Lys-14757. The crystal structure of the oligomer shows a crevice at the base of the cap domain (around Arg-200, on the rim domain) described as a potential binding pocket for a charge head group of a sphingomyelin<sup>54</sup>.

The structure of the  $\alpha$ HL monomer has not yet been solved but it can be deduced from the solved structure of the related staphylococcal LukF leukocidin<sub>58</sub>. The molecular basis of membrane insertion has become clearer after the determination of the oligomeric crystal structure of the  $\alpha$ HL<sub>57</sub>.



**Figure 1.2 The  $\alpha$ -hemolysin.** Representation of the  $\alpha$ HL heptamer (top) and monomer (bottom). Upper: Each subunit of the heptameric protein is shown in a different colour. **a**, Perpendicular view to the sevenfold axis of the nanopore. **b**, Bottom view. **c**, Crystal structures of LukE monomer in solution. **d**, A single protomer of  $\alpha$ HL taken from the heptamer PDB file. In both monomer and oligomer the position of the N- and C-terminal is indicated. (PDB files: 3ROH for LukE and 7 $\alpha$ HL for WT- $\alpha$ HL).



**Figure 1.3  $\alpha$ HL insertion in a lipid bilayer.** **a**,  $\alpha$ HL monomer. The N-terminal (red),  $\beta$ -sandwich and rim (blue), stem (green), and triangle domains (yellow) are labeled and color-coded across the next panel. **b**, The assembly pathway of the  $\alpha$ HL pore (modified from Miles G. et al., 2006.). The assembly pathway on

membrane has been proposed to proceed via four distinguishable stages (i) the secretion of the water soluble monomer, (ii) the monomer binds to membranes, (iii) the formation of a not lytic prepore and at the end (iv) the final channel formation when two antiparallel  $\beta$ -strands for each monomer are inserted in the bilayer<sup>59,60</sup>.

This toxin doesn't act always at lytic concentration. Sub-lytic concentrations  $\alpha$ HL stimulates cell proliferation<sup>12</sup> and activates several stress response pathway such as p38 MAPK, and JNK MAPK<sup>16,50,61</sup>. Moreover,  $\alpha$ HL can inactivate specific pro-survival proteins such as Akt<sup>13</sup> at concentration higher than 100 nM. Finally, Kao and coworker<sup>16</sup> demonstrated a coordinate network of cellular gene expression signals involved in PFTs response, using *Caenorhabditis elegans* as a biological model.

**1.1.3 Post-transcriptional mechanisms involved in the cap-dependent translation initiation.** Protein synthesis is a process highly consuming in term of cell energy. Therefore, regulation of translation plays a profound effect during the changes of the cell phenotype by shaping protein synthesis on a specific behavioral commitment. Rapid changes in protein synthesis are therefore extremely important to support the cellular demands.

Recently, studies have published quantitative information on how genomic informations are processed to obtain a specific cellular proteome at a specific time. It has been demonstrated that only 40% of the variance in protein levels can be explained by transcription<sup>19</sup>. Hence, protein abundance seems to be regulated mainly at the polysomal (i.e. the protein synthesis machinery) and protein level, highlighting the importance of a selective translational control. Translational control has been renowned as the most important emerging player during many cellular processes when compared to the transcriptional control<sup>19</sup>.

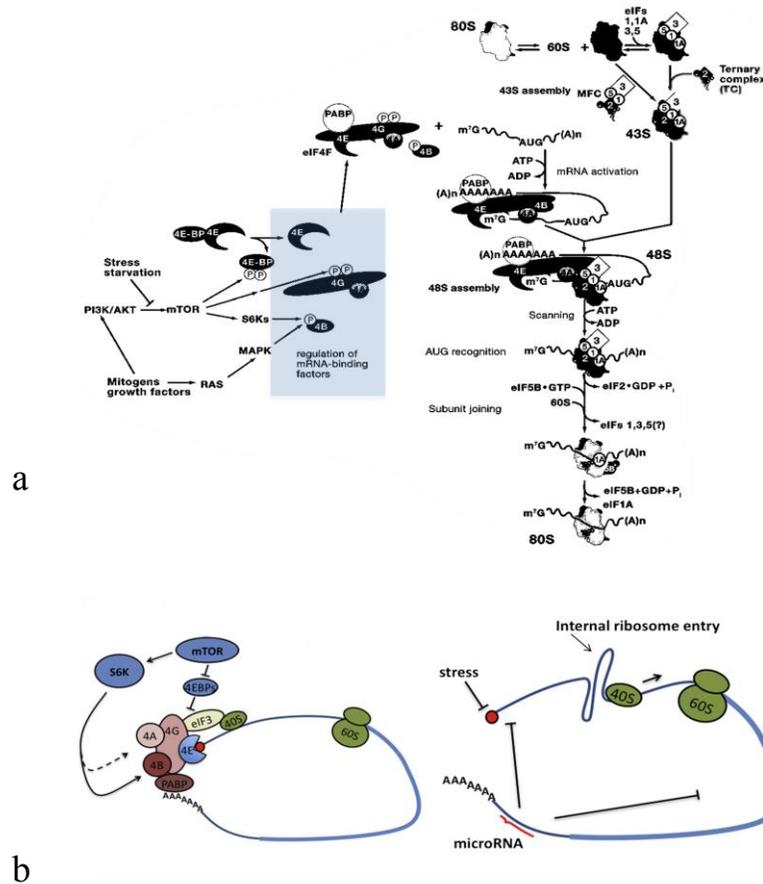
In the canonical model of eukaryotic cap-dependent translation initiation, the small ribosomal subunit is recruited at the 5' end of the mRNA. The cap-dependent scanning of the 40S subunit, before reaching the start codon, requires the binding of the trimetric complex eIF4F (formed by the initiator factors eIF4E, eIF4G and eIF4A) to the 7-methyl

G cap structure in 5'. The eIF4F complex allows the small subunit 40S to form the pre-initiation 48S complex and aids scanning to the start codon with the help of the helicases eIF4B and eIF4A. These last two initiator factors are responsible for the simultaneous unwinding of inhibitory RNA secondary structures. The interaction between eIF4G and eIF4B on the 5'-UTR (untranslated region of the mRNA) with the poly (A)-binding protein (PABP) on the 3'-UTR, allows the circularization of the mRNA and the stabilization the RNA-protein loop complex (Sonenberg 2009 for a review<sup>62</sup>). These protein-protein interaction drive the circularization of the mRNA, which is considered a way to enhance translational efficiency. Other factors are then recruited after circularization, such as eIF2, methionyl tRNA<sup>i</sup>, and GTP<sup>62</sup>. The mRNA loaded small ribosomal subunit scans to the first AUG codon during 5'-UTR's unwinding. Afterwards, the large ribosomal subunit 60S is recruited and translation elongation starts. The ability of the ribosome to be recruited to and scan along the RNA is influenced by sequence elements within both the 5' and 3' untranslated regions of the mRNA (Fig 1.4a). Most RNAs fold to create complex structures with intrinsic regulatory functions, encoding information both at the sequence and the structural level. Functional *cis*-acting elements (e.g. internal ribosome entry segments (IRES), upstream open reading frames (ORF), 3'-terminal uridylation and terminal oligopyrimidine tracts (TOP)) are able to modulate the protein synthesis and to regulate cytoplasmic mRNA fate, interacting with defined *trans*-acting factors such as RNA binding proteins (RBPs) and/or non coding RNAs<sup>63</sup>.

#### **1.1.4 The mTOR pathway**

A key pathway that integrates environmental cues from the plasma membrane involves the Akt-mTORC1 axis. The main downstream targets of mTORC1 include the eukaryotic initiation factor 4E-binding protein-1 (4E-BP1) and the ribosomal protein S6. The phosphorylation of the 4E-BP1 leads to the release of eIF4E the recruitment on the mRNA of other initiation factor (as eIF4G, eIF4A and eIF3). This allows the formation of the complex eukaryotic initiator factor eIF4F. mTORC1 can activate the kinase S6K1, a protein acting

both on the ribosomal protein S6 and on the co-helicase initiator factor 4B (Figure 1.4b). S6K1 is a regulator of ribosome biogenesis and it plays an important role in the unwinding of secondary structures on the 5' UTR<sup>64-66</sup>. Moreover, in plant, mTORC1 can under certain conditions, directly stimulate translation reinitiating<sup>67</sup>.



**Figure 1.4 Cap-dependent translation initiation.** **a**, mTOR promotes eIF4G and eIF4B phosphorylation either directly or via S6Ks. Mitogens and growth factors promote all of these phosphorylation events by activating mTOR via PI3K/Akt signaling or RAS/MAPK signaling. 4E-BP dissociates from eIF4E after phosphorylation. The mRNA translation is activated by binding of eIF4F (eIF4E, eIF4G, eIF4A) in 5' and PABP to the poly(A) tail in 3'. The small ribosomal subunit 43S binds near the 5' cap, facilitated by eIF3/eIF5 interactions with eIF4G/eIF4B, and scans to the AUG codon in an ATP-dependent reaction, with partial hydrolysis of the eIF2-bound GTP in the ternary complex to eIF2-GDP-Pi. Start codon recognition allows the release of Pi and eIF2-GDP. The joining of the ribosomal 60S subunit is catalyzed by eIF5B-GTP. GTP hydrolysis triggers release of eIF5B-GDP and eIF1A to yield the final 80S initiation complex. **b**, Right: A zoom on the proteins involved in the cap-dependent translation initiation and in the formation of the mRNA loops. Left: examples of RNA sequence and structural elements that can affect translation are shown. IRESs in the 5' UTRs (that allow the protein synthesis when cap-dependent translation is compromised) and miRNA binding sites in the 3' UTRs (left) are schematically reported<sup>62,63</sup>.

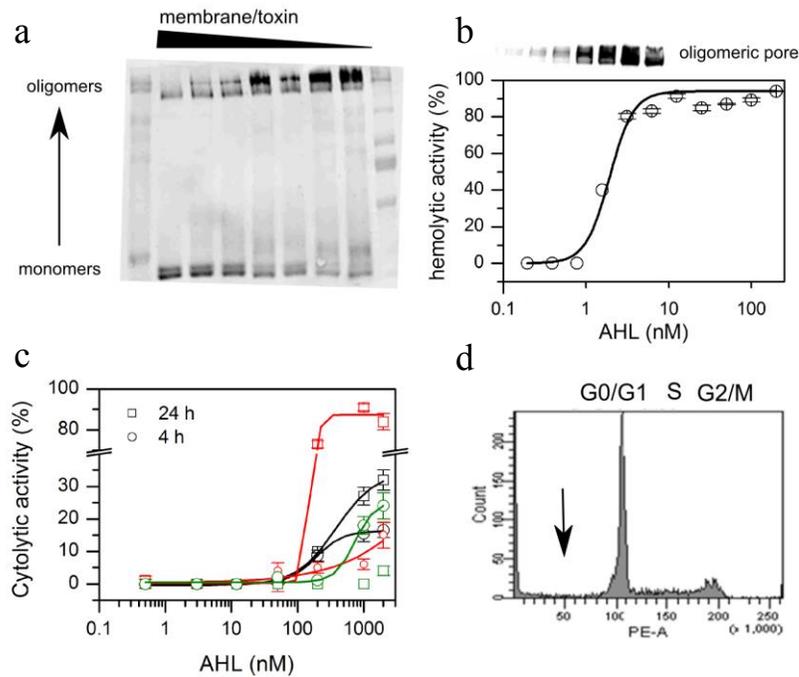
## 1.2 Results and discussion

Genome-wide analysis of translation has the potential to provide major contributions in understanding the pathophysiology of infection processes, given the complex interplay between pathogens and host cells. Informations about the translational state of mRNAs or the activity of RNA binding proteins and ncRNAs after treatment with sublytic doses of pore forming toxins are completely missing. This study uncovers the reshaping undergoing in the translational control system of the host in response to sublytic doses of  $\alpha$ HL. By comparing for the first time variations in the transcriptome and in the translome, our results give evidence that the host gene expression is chiefly translationally rewired after the virulent attack of  $\alpha$ HL and that these effects occur together with the recruitment of RNA binding proteins.

**1.2.1 Definition of the Sublytic doses of  $\alpha$ HL .** With this study we intend to obtain specific information concerning the global impact on translational control of gene expression at sublytic PFT doses. Thus it is of primary importance to define, as precisely as possible, these concentrations in the specific host cell line under study. Here, we used three common human cell lines (SH-SY5Y, CHP-134 and HeLa) as host and the pore forming toxin  $\alpha$ HL as staphylococcal virulence factor. Being the pore formation a strictly sequential and dynamic process, at any time point after toxin addition, all intermediates (i.e. monomers, non-lytic pre-pores and pores) may be present on the plasma membrane<sup>68,69</sup>. In a similar way, depending on the toxin concentration in the extracellular environment, the number of pore/target cell can be tuned. We addressed the ability of native  $\alpha$ HL to cause cytolysis at different lipid to toxin ratio. We employed Rabbit Red Blood Cell (RRBC) a widely used cell target model for  $\alpha$ HL. First, we incubated different amount of plasma membranes obtained from RRBC with 70 nM  $\alpha$ HL (Figure 1.5a). As expected we found that high lipid to toxin ratio drove to a decrease of the stable oligomeric forms of  $\alpha$ HL in favor of the monomeric form.

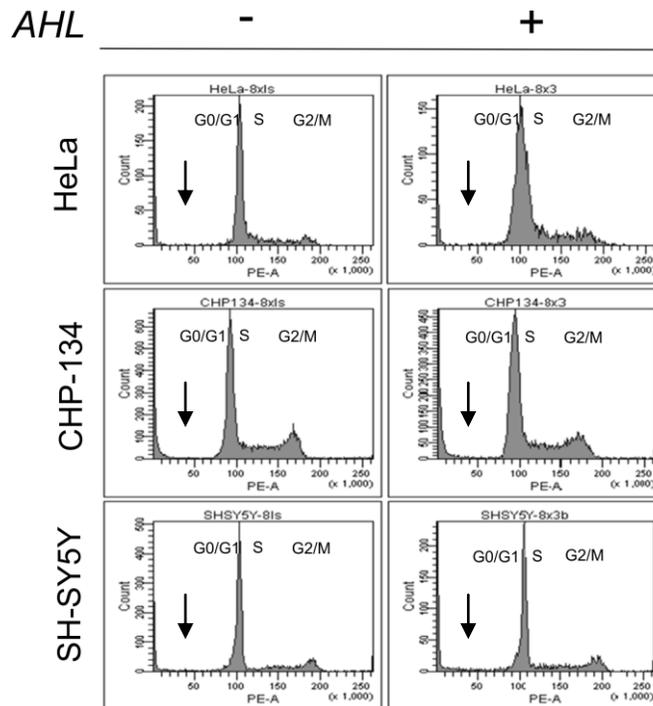
The hemolysis induced by  $\alpha$ HL is complete at 4 nM and correlates with the presence of stable oligomers on the membrane (Figure 1.5B). Nevertheless, oligomers were still detectable on the membranes even in the absence of hemolysis, suggesting that at certain sublytic doses few pores may be inserted into the membrane even in the absence of detectable hemolysis. The first concentration at which no hemolysis occurred but still some oligomers were present can be considered the upper concentration threshold of the so called sublytic doses. By comparing these evidences with the results obtained in Figure 1.5a, we reasoned that 4-6 fold lower concentration than the sublytic threshold concentration is needed to have most of the toxin in the monomeric and membrane bound state, with very few pores inserted in the lipid bilayer. We called this concentration window “testing sublytic concentration”. In this condition we would be able to discriminate between massive ion imbalance effects from the other possible effects occurring at very low doses of native  $\alpha$ HL.

Then we measured the cytolytic effect of different concentration of the native  $\alpha$ HL on two human neuroblastoma cell lines (SH-SY5Y and CHP-134) and on the cervical cancer derived Hela cell line (Figure 1.5c). By using the bioluminescent cytotoxicity assay based on Adenilate Kinase (AK) release, we determined the sublytic concentration threshold after 4 h and 24 h of incubation with  $\alpha$ HL. Each cell line showed its specific susceptibility to the pore formation and its own kinetic. The sublytic threshold for SH-SY5Y (Figure 1.5C, black lines) is 50 nM, which set our testing sublytic concentration in the range of 3-12 nM. The same sublytic concentration window was chosen for CHP-134 which displayed very similar sublytic concentration threshold, even if the cytolysis reached a maximum value after 24 h only in this cell line. The less sensitive cell line to  $\alpha$ HL pore formation was the Hela cell line, with a sublytic concentration threshold of 300 nM. Therefore, in this case the testing sublytic concentration was set between 12 and 50 nM. The observed differences of the cytolytic kinetic among the cell lines tested and between the two time points may be due to a different membrane binding ability of the toxin and/or interaction with ADAM10 and/or to cell wound healing recovery ability<sup>15</sup>.



**Figure 1.5 Definition of  $\alpha$ HL sublytic doses in human cell lines.** (A)  $\alpha$ HL oligomers and monomers distribution at different membrane to toxin ratio. After treatment of a serial dilution of rabbit red blood cells (RRBCs, starting from a suspension with 0.1 optical density) with 70 nM  $\alpha$ -hemolysin. (B) Hemolytic activity of  $\alpha$ HL and detection of oligomers on rabbit plasma membrane. The concentrations for each point are specified in the internal panel. It is possible to observe that at sublytic concentrations (on the left of the dot line) oligomers are still detected. The hemolysis trend follows a logistic curve. The mean value  $\pm$  sd of 3 independent experiments is reported. In the upper panel the SDS-PAGE analysis of RRBC plasma membrane after toxin addition shows a correspondence between oligomeric pore formation and hemolytic activity. (C) Cytolytic activity of  $\alpha$ HL on three human cell lines SH-SY5Y (in black), CHP-134 (in red) and HeLa (in green) after 4 and 24 hours of treatment (open square and open circle, respectively). The first sublytic concentration is 50 nM. The ranges of sublytic concentrations chose in the present work were between 3 nM and 12 nM. The mean value  $\pm$  sd of 3 independent experiments is reported. (D) Effect of sublytic doses of  $\alpha$ HL on the cell cycle of SH-SY5Y. The cell cycle by propidium iodide (PI) staining and FACS CANTO (AB) analysis after cell synchronization. Eight hours treatment with 3nM  $\alpha$ HL was performed. The black arrow points to regions corresponding to damage and/or apoptotic cells.

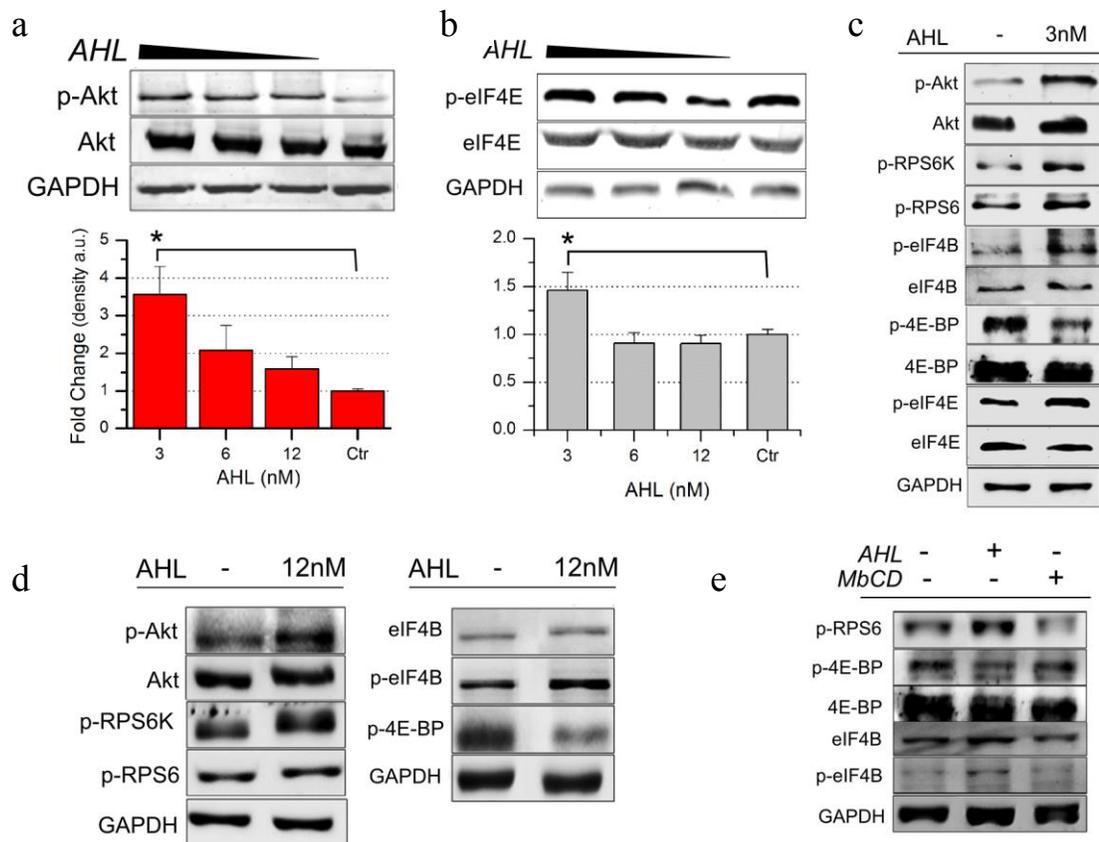
Given the fact that apoptosis or necrosis may still occur at these concentration, we analyzed by flow cytometry the cells after PI staining (Figure 1.5 D and Figure 1.6). After 8 h of  $\alpha$ HL treatment with the previously chosen concentrations, the cells were found to be unaffected by the toxin exposure and neither cell death nor cell cycle changes have been observed. Summarizing, we set two cell specific windows of sublytic concentrations for three cell lines to be used for our purposes, i.e. 3-12 nM for SH-SY5Y and CHP-134 and 12-50 nM for HeLa.



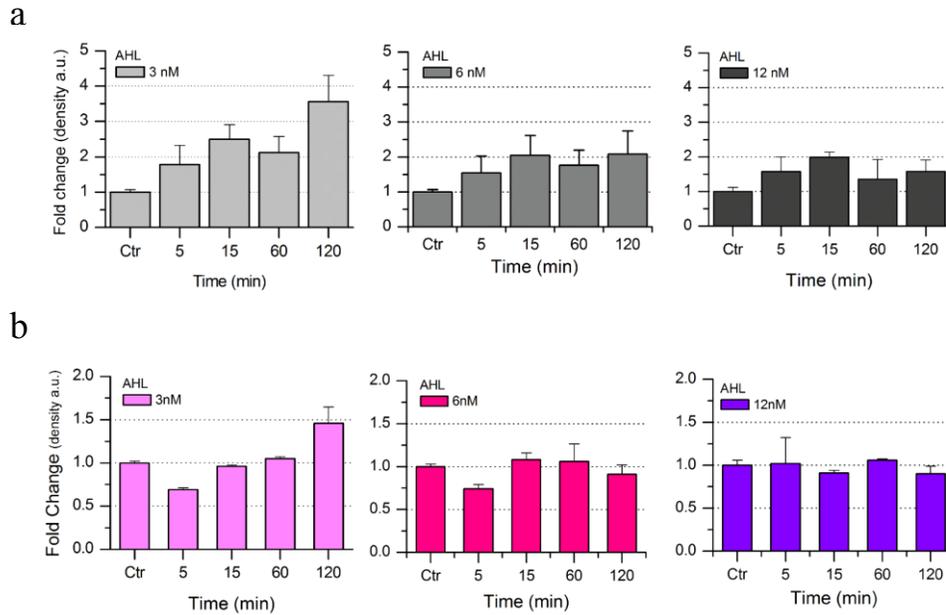
**Figure 1.6** Effect of sublytic doses of  $\alpha$ HL on the cell cycle of HeLa, CHP-134 and SH-SY5Y. The cell cycle by propidium iodide (PI) staining and FACS CANTO (AB) analysis after cell synchronization. Eight hours treatment with 3nM  $\alpha$ HL was performed. The black arrow points to regions corresponding to damaged and/or apoptotic cells.

**1.2.2 Sublytic  $\alpha$ HL redirects factors controlling translation.** To address any changes at the translational level, we first considered the translational effectors known to directly regulate translation initiation and to integrate translational activity with cell environmental cues. We evaluate the phosphorylation state of Akt, of the main mTORC downstream targets (4E-BP, RPS6K, RPS6, eIF4B) and of eIF4E as cellular markers for translational activation or inhibition<sup>62,66</sup> due to exposure to the testing sublytic concentration of  $\alpha$ HL obtained above. The effect of sublytic doses has been tested in three cell lines for 3 sublytic concentrations (i.e. 3, 6, 12 nM for SH-SY5Y and CHP-134 and 12, 25, 50 nM for HeLa), at four times of exposure (i.e. 5, 15, 60, 120 minutes), see Figures 1.7, 1.8 and 1.9. SH-SY5Y, showed a significant activation of Akt and eIF4E at the lower sublytic concentration (3 nM) and at longer incubation time (2 h) (Figure 1.7a and b and Figure

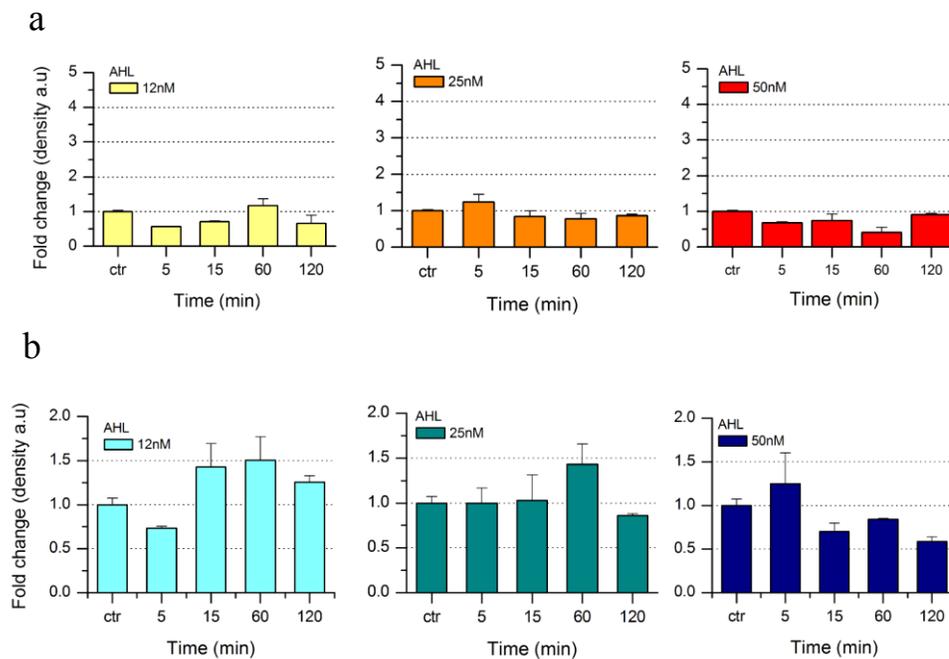
1.8). This result suggests that an activation of key translational proteins occurred uniquely at sublytic concentrations. Higher concentrations are probably suggesting a different effect because neither Akt nor mTOR were activated at 6 nM and 12 nM. These higher concentration are close to previously reported data (> 50 nM) on the  $\alpha$ HL activity human cells.<sup>16,50,61</sup>



**Figure 1.7 Sublytic doses of  $\alpha$ HL induce mTor pathway activation** (a, b) Western blotting and quantification of time course Akt and eIF4E phosphorylation state for three different concentrations of AHL 12nM, 6nM and 3nM in SH-SY5Y. Mean values of at 3-4 independent experiments are shown. t-test  $p$ -val < 0.05. Western blot on downstream target of Akt-mTORC1 axis on SH-SY5Y cells (c) and on HeLa cell after 2 hours of treatment with 3 nM native AHL. (d) Effect of Methyl-beta-cyclodextrin (MBCD) treatment on SH-SY5Y before and after exposure to 3nM AHL for 2h. The effect of cholesterol depletion on the phosphorylation of key proteins involved in translational control are shown. Cells were pre-treated with 10 mM of MBCD before toxin addition.



**Figure 1.8** Time course of Akt (a) and eIF4E (b) phosphorylation after treating SH-SY5Y with 3 nM, 6 nM and 12 nM of native AHL. Densitometric quantification of immunoblotting p-Akt bands after treatment was obtained from three to four independent experiment. Mean value +- s.e.m. are shown.



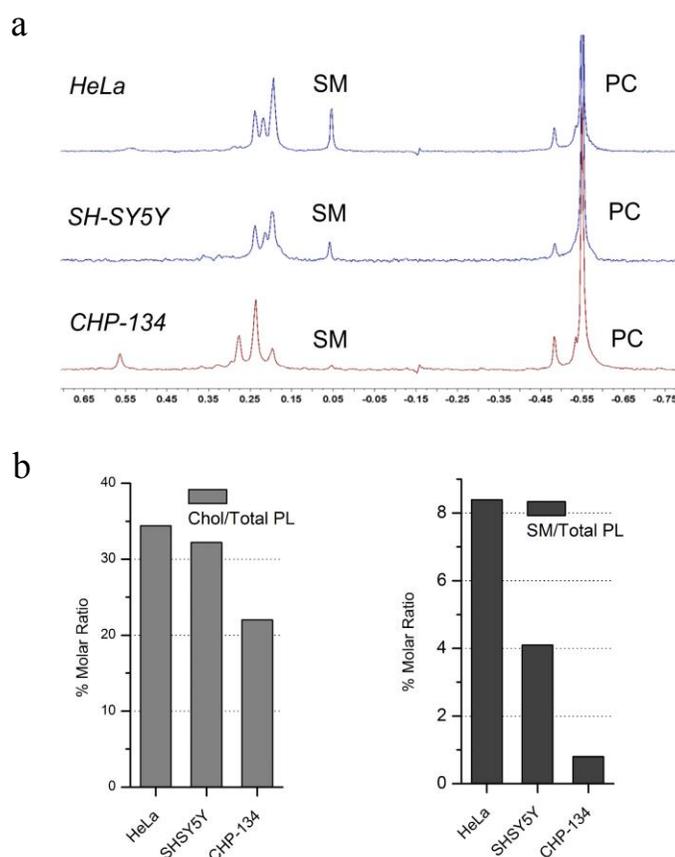
**Figure 1.9.** Time course of Akt (a) and eIF4E (b) phosphorylation after treating CHP-134 with 12 nM, 25 nM and 50 nM of native  $\alpha$ HLL. Densitometric quantification of immunoblotting p-Akt bands after treatment was obtained from three to four independent experiments. Mean value +- s.e.m. are shown.

The possible translational modulation suggested by the phosphorylation of eIF4E, was confirmed by the concomitant increase in the phosphorylation state of RPS6K, which in turns phosphorylated RPS6 and eIF4B (Figure 1.7C). Together this set of phosphorylation gave evidences of a general stimulation of translational initiation. Therefore, we found quite surprisingly the decrease of 4E-BP phosphorylation (Figure 1.7C), which is generally considered a signal depressing translation initiation. Noteworthy, the very same activation of Akt, RPS6K, RPS6, eIF4B, eIF4E and the inactivation of 4E-BP occurred also on HeLa cells at the corresponding lower sublytic concentration (Figure 1.7D), confirming the opposite signals on translation. On the contrary, CHP-134 were almost insensitive to  $\alpha$ HL-dependent activation of Akt (Figure 1.9), a possible reason owing this behavior will be given below.

To confirm that the observed effects were indeed caused by the binding of monomers and by the few pores inserted into the membrane, we interfered with the toxin binding to the membrane by using cholesterol depletion. It is known that the membrane lipid composition is crucial for the proper interaction of  $\alpha$ HL with plasma membranes, and membrane cholesterol or sphingomyelin depletion are able to prevent  $\alpha$ HL binding to the plasma membrane<sup>54</sup>. Most probably this depends on the localization into lipid rafts of ADAM10, the most probable receptor of  $\alpha$ HL<sup>30,70,71</sup>. After pretreating the cells with beta-methyl-cyclodextrin<sup>72</sup>, we incubated SH-SY5Y cells with 3nM native  $\alpha$ HL and observed the complete abrogation of the previously detected phosphorylation of Akt and eIF4B. In accordance, the dephosphorylation of 4E-BP did not occur in the absence of cholesterol, confirming the overall specificity of the effects provoked by low doses of native  $\alpha$ HL. Given these results, we suspected that CHP-134 insensitivity to  $\alpha$ HL could be caused by the plasma membrane composition. In fact, lower amount of cholesterol would impede the toxin binding or the interaction with ADAM10 preventing the induction of the cellular responses observed for SH-SY5Y and HeLa. Therefore, we compared the proportion of cholesterol vs total lipid amount by NMR in HeLa, SH-SY5Y and CHP-134 lipid extracts (Figure 1.10). Consistently with the cholesterol depletion experiments, we found a relative reduction in the cholesterol and sphingomyelin content in CHP-134,

meaning that the absence of cholesterol may prevent and/or delay the effects of low doses of  $\alpha$ HL.

These results demonstrate that i) the host cells stroke back the presence of  $\alpha$ HL monomers and of the few pores (i.e. sublytic concentrations) in two apparently diametrically opposite messages on translational effectors and ii) the host cell membrane composition may affect the sensitivity to such a cellular response.



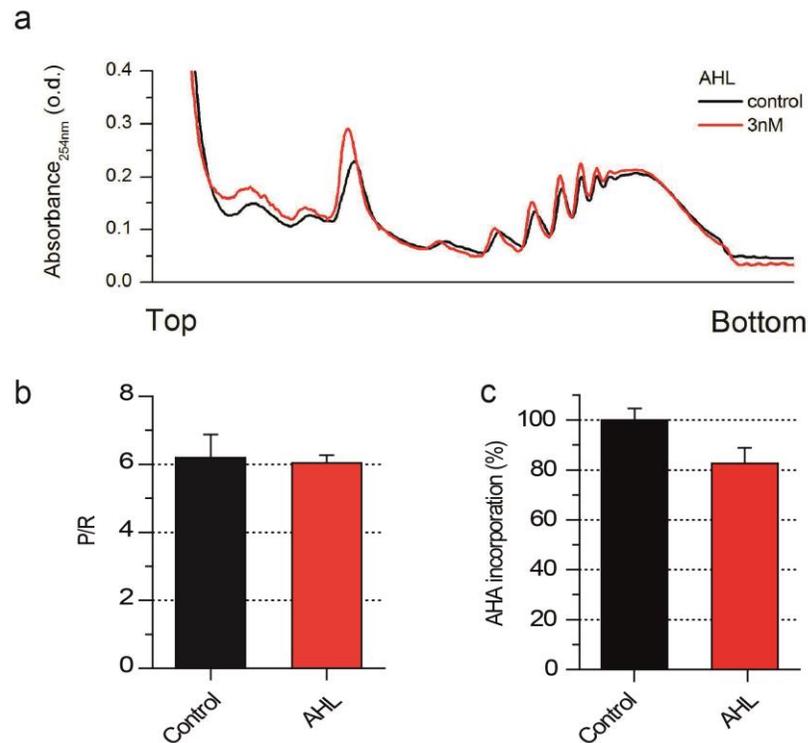
**Figure 1.10** H-NMR and  $^{31}\text{P}$ -NMR spectra of lipid extracts obtained from HeLa, SH-SY5Y and CHP-134 cell lines were recorded in  $\text{CD}_3\text{OD}$  at 298 K on a Bruker-Avance 400 NMR spectrometer. (a)  $^{31}\text{P}$ -NMR spectra (161.9 MHz, PC as calibration signal at -0.55 ppm) showing the main PL classes. Relative area integration shows that SM lipids are more abundant in HeLa (8%) and SH-SY5Y (4%) than in CHP-134 (0.8%) cell line. (b) relative molar ratio of cholesterol (Chol Total), SM and plasmenyl lipids with respect total PL as obtained from the corresponding  $^1\text{H}$ -NMR spectra (400.12 MHz, solvent residual signals as calibration signal at 3.31 ppm) showing that HeLa (34%) and SH-SY5Y (32%) are richer in CHL than CHP-134 (22%) cell line.  $^1\text{H}$ -NMR measurements are in full agreement with  $^{31}\text{P}$ -NMR analysis concerning the relative amount of SM. (NMR analysis has been done in collaboration with prof. Graziano Guella, Department of Physics, University of Trento).

**1.2.3 The host cell translation is hijacked in opposite directions.** As a first step toward delineating the precise role of translation and as a general way to get further insights about the observed opposite translational messages after  $\alpha$ HL treatment, we studied the functionality of protein synthesis by using two different approaches: the quantification of ribosomes engaged in active translation and the incorporation of the methionine analog, homo-L-azido alanine (AHA)<sup>73</sup>.

Template mRNAs, that are going to be used for protein synthesis, are bound to increasing number of ribosomes, forming the so called polysome<sup>74</sup>. Therefore, the higher the synthesis level, the higher the number of ribosomes engaged in translation. Being the polysome-bound mRNAs significantly heavier than the translationally silent mRNAs, they can be purified by sucrose gradient fractionation and the amount of polysomes can be used as a hint of translational activation. The unbound and translationally inactive mRNAs are sequestered into lighter messenger ribonucleoprotein (mRNP) particles or are associated to single ribosomes (80S). When these latter are not engaged in translation, their level increase and this increase represents a symptom of translation initiation defects. In this condition, the polysome to ribosome-free ratio decreases and it can be used as a way to measure initiation defects with respect to elongation<sup>75,76</sup>.

We performed a parallel polysomal profiling analysis of SH-SY5Y cells before and after native  $\alpha$ HL sublytic treatment (Figure 1.11A). Then, we measured the corresponding overall polysome (P) vs ribosome (R) content (Figure 1.11B). As clearly stated by the absorption profiles in Figure 1.11A and by the quantitative comparison in Figure 1.11B,  $\alpha$ HL did not induce a clear impairment in the P/R ratio, even if a slight increase in the ribosomes content could be noted. Next, we quantified the incorporation of the methionine analog AHA, which can be used as convenient alternative to radiolabelled methionine for studying *de novo* protein synthesis<sup>73</sup>. After 2 h of sublytic  $\alpha$ HL treatment, a decrease of around 20% in the protein synthesis was observed (Figure 1.11C), even if it did not correspond to any consistent impairment in the overall ribosome loading on polysomes

Figure 1.11B. The decrease in newly synthesized proteins suggests a partial block of *de novo* protein synthesis and is in agreement with the decreased phosphorylation of 4E-BP (Figure 1.7C), which is known to trap eIF4E and inhibit cap-dependent translation<sup>62</sup>. These findings did not get light onto the previously observed activation of RPS6 nor onto translation initiation factors activation, that should both induce a burst in the rate of protein synthesis.

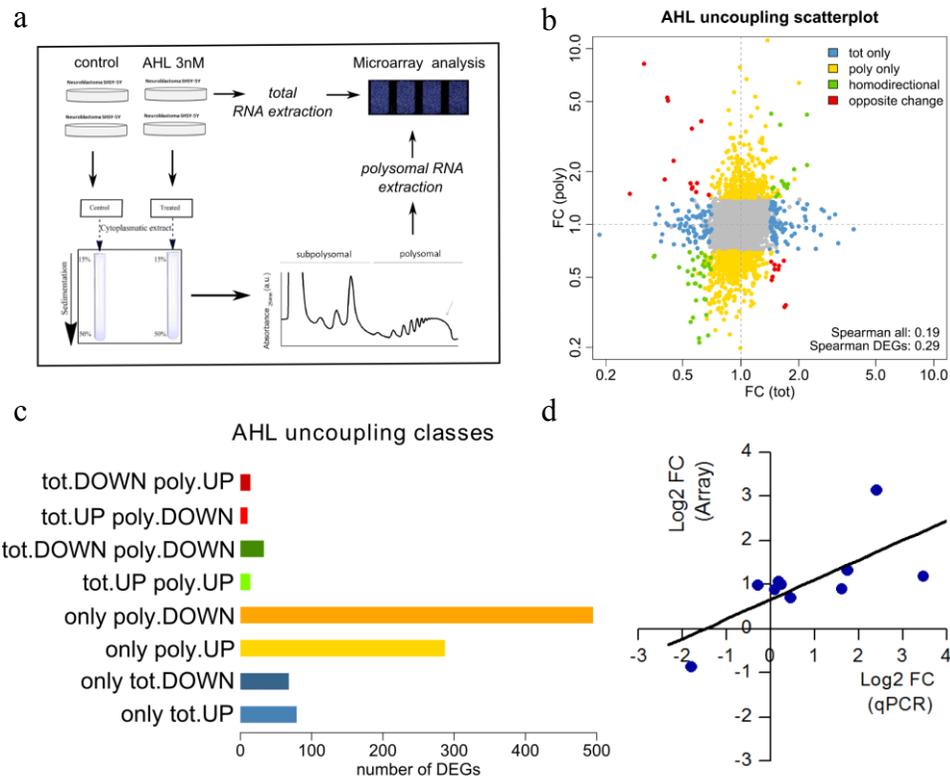


**Figure 1.11** (a) Sedimentation profiles obtained in a concave 15–50% sucrose gradient of cytoplasmic lysates from control (untreated) and treated (3nM AHL for 2h) SH-SY5Y cells. The purification allows the separation of RiboNucleoParticles (RNPs) from ribosomal subunits (40S and 60S), ribosomes (80S) and polysome-bound RNA measuring the UV absorption profile at 254nm along the sucrose gradient. (b) The absorbance profiles were used to derive the relative change in polysomes amount after 3nM treatment with AHL respect to the control by dividing the area under the curve corresponding to the polysomal fractions. (c) The detection of *de-novo* protein synthesis was measured by means of AHA incorporation using Click-iT AHA Alexa Fluor 488 Protein Synthesis HCS Assay (Life Technologies) following the manufacturer's protocol. The methionine analog AHA in methionine-free DMEM was added to cells for 30 min. Cells were fixed and incubated with fluorescent alkyne to label the AHA incorporated into nascent proteins. The relative AHA incorporation was assessed by using Operetta HCS. Experiments were run in biological triplicate and 300-500 cells were considered for each sample.

Therefore, we hypothesized the occurrence of two opposite messages on translation given the opposite activating/inactivating events on translational factors and the absence of polysomal decrease. The opposite activation of translation factors reflects a selective choice of mRNAs that the cell needs to actively translate or to translationally repress rather than a rough non-specific impairment of the overall protein synthesis rate. We stress the point that a non-specific impairment of the overall protein synthesis rate seems not to be involved. In other words, we suspected that the host/virulence-factor interaction could reveal a more sophisticated, general and still unknown remodelling of host gene expression. To investigate this possibility we employed a high-throughput analysis of the host transcriptome and translome before and after sublytic treatment with native  $\alpha$ HL.

**1.2.4 An almost exclusive translational response dictates an overall reshaping of host gene expression.** To get light into the opposite messages on translation and to systematically monitor the gene expression response of the host cells to sublytic doses of PFTs, we performed a genome-wide multi-level analysis based on polysomal profiling<sup>77,78</sup>. The polysomal RNA (portraying the mRNAs in active translation - the translome) and the total RNA (portraying the total amount of transcribed mRNAs, the transcriptome) from SH-SY5Y cells treated or untreated with  $\alpha$ HL 3 nM for 2 hours (Figure 1.12A) were extracted. Polysomal and total mRNAs levels were quantified with gene expression microarrays, profiling the cell translome and the transcriptome, respectively. A total of 16086 genes were quantified after data preprocessing, normalization and quality check with Bioconductor packages (see Methods). A scatter plot containing the whole set of transcriptome and translome variations is displayed in Figure 1.12B. The translational response to  $\alpha$ HL is globally independent from the transcriptional response, as resumed by the low spearman correlation value (0.19). Differentially expressed genes (DEGs) induced by  $\alpha$ HL treatments were determined adopting a double threshold based on 1) the magnitude of the change, 2) the statistical significance of the change. Using this approach, 218 DEGs were detected in the transcriptome, 853 in the translome. Comparing the two lists of genes, 78 % of DEGs (782 genes, in yellow in Figure 1.2.8B)

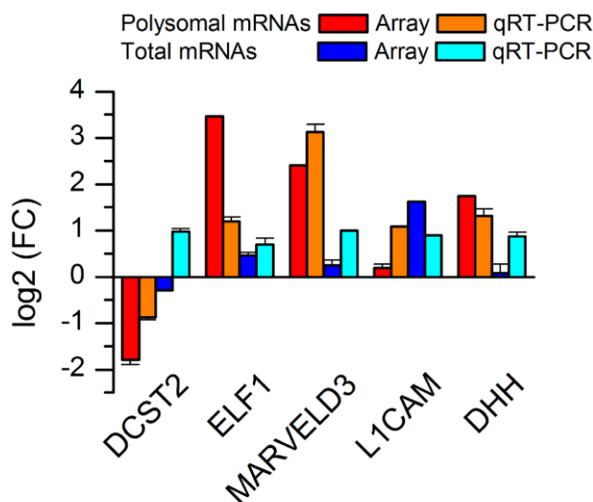
had significant changes only at the polysomal level, 15 % of DEGs (147 genes, in blue in Figure 1.12B) changed significantly only at the transcriptome level. According to these results, cells replied consistently to the  $\alpha$ HL treatment by changing the mRNAs uploaded onto polysomes rather than inducing *de novo* mRNAs transcription. Focusing on genes with significant variations both in the transcriptome and the translome, 24 genes (2% of DEGs, in red in Figure 1.12B) were characterized by anti-directional variations, i.e. the higher degree of discrepancy between translational and transcriptional changes. Finally, only 47 genes (5% of DEGs, in green in Figure 1.12B) showed significative homodirectional changes, reflecting a concordance between transcriptional and translational movements. These classes of genes, further divided according to the verse of the change, are summarized in Figure 1.12C. While transcriptome DEGs were equally balanced between upregulation and down-regulation (47% upregulation, 53% downregulation), translome DEGs were remarkably shifted towards down-regulation (36% upregulation, 64% downregulation). In agreement with the low global correlation mentioned before, for 95% of DEGs the translational effect of  $\alpha$ HL was different from the transcriptional effect, indicating that post-transcriptional regulation possibly results in a strong uncoupling between the two gene expression levels.



**Figure 1.12 Genome-wide gene expression analysis reveals a huge reshaping of the translational level of gene expression.** (a) Diagram outlining the experimental design followed for transcriptome and translome microarray profiling on SH-SY5Y after treatment with  $\alpha$ HL 3nM. Experiments were performed in biological triplicate. (b) Scatterplot displaying, for each gene, the transcriptome and translome gene fold changes (Differentially Expressed Genes, DEGs) after  $\alpha$ HL treatment. Differentially expressed genes (DEGs) induced by  $\alpha$ HL treatment were determined adopting a double threshold based on 1) the magnitude of the change ( $\log_2$  fold change  $> 0.5$  and  $< -0.5$  for induced and repressed genes, respectively); 2) the statistical significance of the change, measured with a rank product test ( $p$ -value  $< 0.01$ ). Genes are coloured according to how they react to the treatment: grey for genes without significant changes, cyan for DEGs with significant variations only in the transcriptome, yellow for DEGs with significant variations only in the translome, red for DEGs with opposite significant variations, and green for DEGs with significant homodirectional changes. The spearman correlation between the fold changes of the whole set of genes or the restricted set of DEGs is shown in the bottom right corner. (c) Barplot highlighting the number of DEGs falling in each the uncoupling classes described before. Genes are further divided according to the verse of the expression change (either upregulation or downregulation). Bars are coloured following the same colour scheme adopted in panel C. (d) Correlation of the  $\log_2$  transformed fold changes derived from microarray hybridizations and quantitative RT-PCR on a set of five genes considered at both polysomal and total RNA levels, are displayed as blue dots. The regression line is drawn in black. (All the genome-wide system data analysis has been done in collaboration with Dr. Toma Tebaldi, Centre for Integrative Biology, Laboratory of translational genomics, University of Trento).

In order to validate the microarray data, five genes with disparate transcriptional and translational variations were selected for quantitative real-time (qRT) PCR analyses. The five genes represented different classes of DEGs: DCST2 (downregulated in the translome), ELFN1, MARVELD3 and DHH (upregulated in the translome), L1CAM (upregulated in the transcriptome). All these genes are either integral or indirectly bound to the cellular membrane. The qRT-PCR data confirm microarray changes for the selected genes, with a correlation of 0.71 (p-value 0.02) (Figure 1.12D and Figure 1.13).

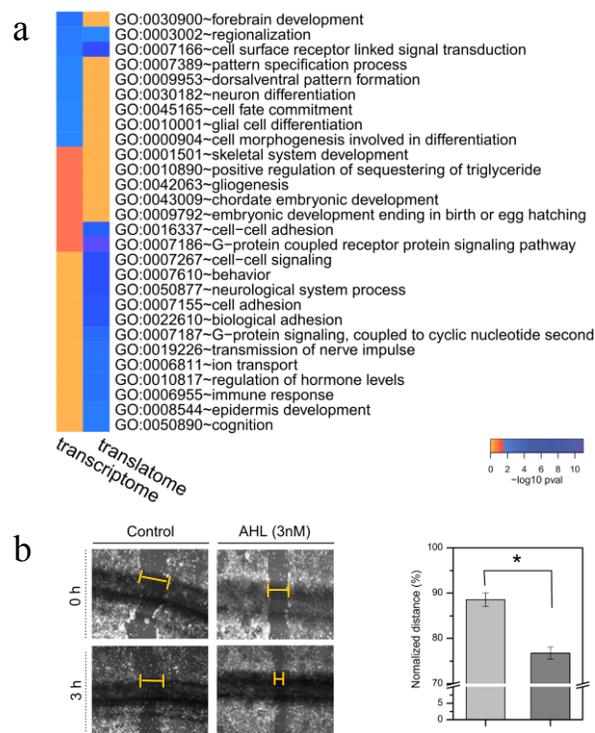
Overall, the microarray results suggest that treatment with a sublytic concentration of  $\alpha$ HL triggers in the host cell an elaborate gene expression reshaping, characterized by a strong uncoupling between transcriptome and translome variations and by the emergence of numerous genes upregulated or downregulated purely at the translational level.



**Figure 1.13 qRt-PCR of selected gene for microarray validation.** Comparison between log<sub>2</sub> FC obtained from arrays (in red the polysomal RNA and in blue the total RNA) and the log<sub>2</sub> qRT-PCR FC (in orange the polysomal RNA and in cyan the total RNA) on total RNA.

**1.2.5 Biological themes of uncoupling: ontological enrichments reveal a translational-specific response to sublytic  $\alpha$ HL toward the plasma membrane.** In order to understand the biological answer to  $\alpha$ HL treatment from a broader perspective, the focus of the analysis switched from movements of single genes to more general biological

annotations. To this end, ontological enrichment analysis was performed on the lists of transcriptome and translome DEGs identified in the previous section. Enrichments were calculated using annotations and statistical tests provided by the DAVID resource<sup>79</sup> (see Methods). The comparison of the top enriched terms for transcriptome and translome DEGs is shown in Figure 1.14A for the Biological Process (BP) branch of GO. From the heatmap of enriched terms is clear that the two level of gene expression supervised distinct cellular responses, i.e. different sets of genes are involved in disjointed biological processes.



**Figure 1.14 The uncoupling between transcriptome and translome reflects enrichments of specific biological themes. (a)** Heatmap showing the top enriched Biological Process GO terms associated to the lists of translome (poly) and transcriptome (tot) DEGs. Significant enrichments are labelled in shades of blue, non significant enrichments are labelled in orange. **(b)** Effect of sublytic doses of  $\alpha$ HL on cell motility of SH-SY5Y cells after 9h treatment by scratch test (wound healing assay). Confluent monolayers of SH-SY5Y cells were scratch wounded by pipet tip, washed and allowed to regenerate in the presence or absence of 3nM  $\alpha$ HL. The length of wound was measured after 3 h. The normalized distance data are shown as the mean  $\pm$  SD of triplicate wells. The statistical difference was assessed by Student's t-test:  $P < 0.01$ . Orange bars: diameters of the scratch. Black sign: marker for the scratch.

In general, the number of enriched terms associated to translome DEGs is higher than the number of terms associated to transcriptome DEGs. This could reflect a more selective and targeted functional response orchestrated by translational regulation, but the possibility that this could derive from the higher number of translome DEGs cannot be neglected.

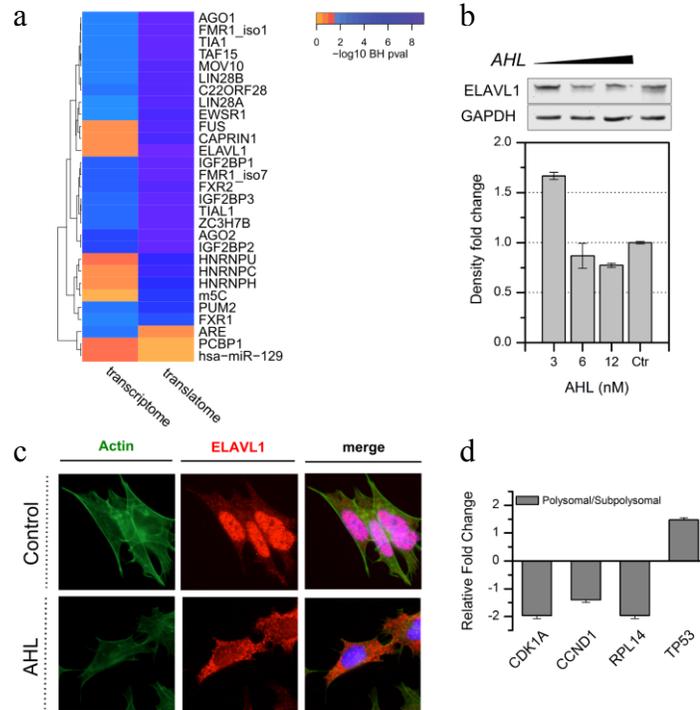
Among the term with higher enrichment, signal transduction, cell adhesion and immune response were the most relevant as translational specific terms. By looking in a deeper detail to the movements at the transcriptional and translational level of the corresponding genes belonging to these terms, we found that up-regulated translational genes are involved in transduction and immune response, while the down-regulated genes are involved in membrane processes. In accordance, the majority of the enriched categories are clearly related to membrane processes, as confirmed by Molecular Function (MF) terms and Cellular Components (CC) Terms (data not shown). This observation and the possible effect on cell adhesion, another biological term emerging in Figure 1.14A, prompted us to check for any changes in the functionality of cell adhesion. We used a cell motility assay to confirm the enrichment data and found that after  $\alpha$ HL treatment, SH-SY5Y increased their mobility (Figure 1.14B), suggesting a profound reshaping of the overall membrane organization.

In conclusion, this functional analysis reinforce the message of a strong uncoupling between polysomal-associated and total RNA variations induced by sublytic doses of PFTs and the presence of semantically coherent response that target the plasma membrane driven by unknown translational control of gene expression.

**1.2.6 Post-transcriptional trans-acting factors involved in  $\alpha$ HL translational specific response.** Our results suggested the activation of translational mechanisms guiding the choice and the regulation of membrane-specific genes upon native  $\alpha$ HL injury. To test this hypothesis, we employed a computational analysis followed by *in vivo* validation. As outlined in the introduction, RNA Binding proteins (RBPs) and ncRNAs

are the main classes of trans-factors driving translational regulation of gene expression<sup>80,81</sup>. RBPs have been called upon to be responsible for buffering transcriptional changes of their targets (resulting in genes exclusively changing in the transcriptome) and altering the translational efficiency of their targets (resulting in genes exclusively changing in the translome)<sup>18,19</sup>. A first way to detect which *trans*-factors could mediate the translational specific response of the host cells to PTFs is to look for known target transcripts among the populations of transcriptome and translome DEGs. This is possible using experimental annotation of interactions between human UTRs and post-transcriptional trans-factors, collected in the AURA database<sup>82</sup>. We calculated the overrepresentation, or enrichments, of targets of specific RBPs or miRNAs, the results are outlined in Figure 1.15A.

Regarding possible translational controls exerted by miRNAs, our results did not exhibit single miRNAs to be significantly associated to transcriptome or translome DEGs. Nevertheless, an enrichment of the binding sites of AGO1, an ubiquitously expressed protein able to associate to miRNAs and siRNAs<sup>83</sup>, could be observed.



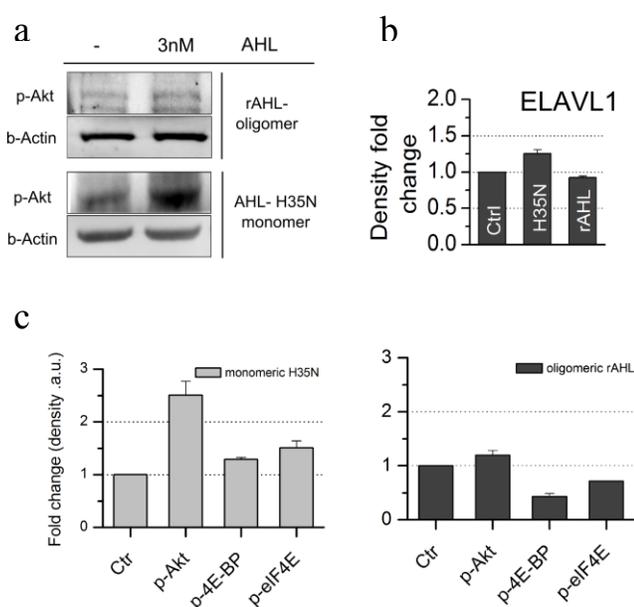
**Figure 1.15 Sublytic doses of  $\alpha$ HL recruit post-translational trans-acting factors.** (a) Heatmap showing the top enriched post-transcriptional regulators (mainly RNA binding proteins and miRNAs) associated to the lists of translatome (poly) and transcriptome (tot) DEGs. The enrichment is based on experimentally annotated interactions collected in the AURA database. Significant enrichments are labelled in shades of blue. (b) Representative western-blot analysis and quantification of ELAVL<sub>1</sub> expression in cytoplasmic lysates of SH-SY5Y treated with sublytic (3nM and 6 nM) and lytic doses (12 nM). GAPDH was used as a loading control. The ELAVL<sub>1</sub> expression levels were quantified. The untreated samples, normalized for GAPDH, were set to unity. All experiments has been performed in triplicate and the error bars indicate the corresponding s.e.m. (c) Immunofluorescence detection of ELAVL<sub>1</sub> (in red) in SH-SY5Y cells before (upper panel) and after 3 nM  $\alpha$ HL treatment for 2h. The green staining labels the actin filaments bound to phalloidin-Alexa-488. DAPI staining to visualize nuclei is colored in blue. (d) Evaluation of ELAVL<sub>1</sub> targets balance among polysomal and sub-polysomal RNA fractions. After treatment of SH-SY5Y with 3nM  $\alpha$ HL for 2 h, the polysomal and sub-polysomal RNA fractions (RNPs, 40S, 60S and 80S fractions in Figure 1.11A) were collected using sucrose gradient fractionation, as described in Materials and Methods. Presented are the average relative ratios of control over treated samples in the comparison between polysomal-associated and sub-polysomal RNAs. Three independent sub-polysomal/polysomal RNA preparations were obtained and analyzed with qRT-PCR.

By looking at classical RBPs, two groups of proteins (HNRNPU, HNRNPC, HNRNPH and ELAVL<sub>1</sub>, FUS, CAPRN1) emerged to be specifically enriched when considering translatome DEGs (Figure 1.15), suggesting an involvement in translational specific response of the host cells to sublytic  $\alpha$ HL injury. Strikingly, the well known and studied RBP, ELAVL<sub>1</sub> (a.k.a. HuR), was predicted to be the RBP most selectively associated to the

translatome DEGs. The Hu/ELAV family member ELAVL1 binds to the 3' untranslated region (UTR) of several genes, regulating transcript stability and access to polysomal complexes. ELAVL1 was first reported as a nuclear protein involved in the regulation of splicing, but it is able to shuttle to the cytoplasm under certain conditions thereby affecting the stability, translation and poly-adenylation of its targets<sup>84,85</sup>. We analyzed the cytoplasmic level of the total ELAVL1 at different concentrations of sublytic  $\alpha$ HL by immunoblotting (Figure 1.15B) and immunofluorescence (Figure 1.15C) on SH-SH5Y. Only upon treatment with 3 nM  $\alpha$ HL, ELAVL1 increased its cytoplasmic level of 1.5 fold (Figure 1.15C). In accordance, the translocation of ELAVL1 from the nucleus to the cytoplasm was observed (Figure 1.15C), confirming the role of this factor in the translational specific response of the host (Figure 1.15B). To get further proof, we selected known ELAVL1 targets (TP53, RPL14, CDK1A, CCND1)<sup>84,86-89</sup> and addressed their relative movements between the translationally active compartment (polysomes) with respect to the translationally silent compartment (sub-polysomal, i.e. RNPs and 80S) by using qRT-PCR (Figure 1.15D). Given the enrichment results observed we expected a major down-regulation of ELAVL1 targets within the polysomes. In fact, for three out of four ELAVL1 targets (RPL14, CDK1A and CCND1) we observed a decreased polysomal/subpolysomal fold-change ratio (Figure 1.15D), meaning that the host cells relocated these mRNAs into the translationally inactive cellular compartment upon sublytic  $\alpha$ HL treatment. Interestingly, the polysomal levels of TP53 (p53 tumor suppressor transcription factor), a well studied regulator of multiple cellular responses and a known target of ELAVL1, are increased. In accordance, p53 was found to be translationally stabilized by ELAVL1 in response to UVC damage<sup>89</sup>.

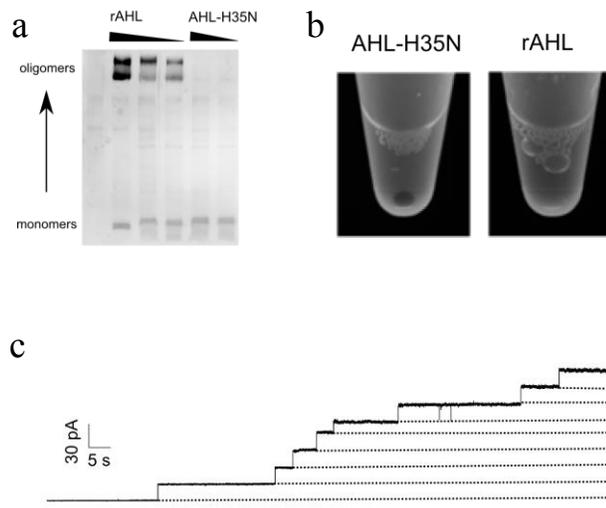
In summary, these results demonstrate the activation of a post-transcriptional gene expression regulation, a recruitment of translational trans-acting factors and a substantial exclusion of a nuclear transcriptional regulation of gene expression in response to the virulent attack of native  $\alpha$ HL.

**1.2.7 The recruitment of ELAVL1 is pore-independent.** As stated above, native PFTs at lytic concentrations form pores thanks to a dynamic process owing the presence of monomers, non-lytic pre-pores and pores on the target plasma membrane. In this study we chose sublytic doses of  $\alpha$ HL treatment due to the already mentioned relevance of these concentrations in vivo. We cannot exclude that very few orphan pores were still present on the cell membranes. Therefore, to understand if the involvement of ELAVL1, the observed activations of Akt and of the translational effectors were due to the monomeric bound  $\alpha$ HL and/or to active pores, we separated the two states using either the extracted  $\alpha$ HL oligomer or the non-lytic mutant,  $\alpha$ HL-H35N, that is able to bind to the membrane but it does not oligomerize and form pores<sup>90-92</sup> (Figure 1.16A and Figure 1.17).



**Figure 1.16 Monomers of  $\alpha$ HL are responsible for Akt activation** (a) The capability of the purified oligomer (r $\alpha$ HL-oligomer) and of the not lytic mutant ( $\alpha$ HL-H35N) to induce Akt phosphorylation in SH-SY5Y was verified by western blotting. The incubation time was 2 h and the concentration 3nM. After two hours of incubation with the monomeric H35N, an increase in the phosphorylation state of Akt is detected, while the pre-formed oligomer does not show any consistent activation of the protein. (b) Representative western-blot analysis and quantification of the phosphorylation state of Akt, eIF4E and 4E-BP in SH-SY5Y treated with 3nM of the monomeric (in light gray) and oligomeric (in dark gray) forms of  $\alpha$ HL, namely  $\alpha$ HL-H35N and the purified oligomer obtained from the recombinant wt  $\alpha$ HL. GAPDH was used as a loading control. The untreated samples, normalized for GAPDH, were set to unity. All experiments have been performed in triplicate and the error bars indicate the corresponding s.e.m. (c) In the inset ELAVL1 level after cell treatment 3nM of the monomeric  $\alpha$ HL-H35N and of the oligomeric r $\alpha$ HL are shown.

After  $\alpha$ HL oligomer purification, we ascertain its ability to form transmembrane pores in Planar Lipid Membranes (PLM) (Figure 1.17C). Therefore, we treated the cells with 3 nM of the extracted oligomers ( $r\alpha$ HL) and with  $\alpha$ HL-H35N and checked for Akt, eIF4E and 4E-BP activation (Figure 1.16B and C). Similarly to what observed for the lower sublytic concentration of native  $\alpha$ HL, the non-lytic mutant H35N activates Akt and eIF4E. Differently from the native  $\alpha$ HL,  $\alpha$ HL-H35N exerted a slight activation of 4E-BP rather than stimulating a dephosphorylation. The concomitant activation of these proteins is indeed compatible with an overall stimulation of translation induced only by the monomeric state, which may account for the high up-regulation of translational DEGs. Noteworthy, the extracted oligomer was completely ineffective as happens at higher sublytic doses of native  $\alpha$ HL (6 nM and 12 nM, Figure 1.7), inducing a slight but yet detectable depression of 4E-BP phosphorylation. Then we checked the level of cytoplasmic ELAVL1 which was previously found to play a role in cellular responses to sublytic  $\alpha$ HL. In Figure 1.16C the comparison between the native sublytic  $\alpha$ HL, the oligomeric  $r\alpha$ HL and the monomeric  $\alpha$ HL-H35N demonstrated that the effect on ELAVL1 cytoplasmic level observed in Figure 1.15B was due to the monomeric portion rather than to the residual pores present on the plasma membrane.



**Figure 1.17 Oligomerization ability: rWT- $\alpha$ HL vs H35N** (a) The oligomerization ability of the recombinant WT- $\alpha$ HL (r $\alpha$ HL) and of the mutant H35N ( $\alpha$ HL-H35N) on Rabbit Red Blood Cells membrane was checked by SDS-PAGE. The physiological ability of r $\alpha$ HL to oligomerize in the presence of membranes and to form complexes is proved by the presence of two bands at 250 kDa, corresponding to bona fide hexamers and heptamers. As expected the mutant H35N does not form any oligomers and the proteins appears in its monomeric form. (b) In both cases toxins were incubated with a RRBC suspension and incubated at 37°C for 1 hour. After centrifugation a pellet of erythrocytes appeared in the sample treated with the non-lytic mutant, demonstrating that no cytolysis occurred. (c) Current trace of the extracted  $\alpha$ HL oligomer (r $\alpha$ HL). Step-wise current increases correspond to the opening of single ion channels starting from no insertions (level 0, zero current level). Protein was added to the *cis* side at a concentration of ~ 50 nM with a constant applied voltage of +120 mV (see Methods).

These results open the possibility that the cell reacts in different ways to pores and to the membrane bound non lytic  $\alpha$ HL activating opposite stimuli on translational control of gene expression at certain monomers/oligomer proportions. Therefore, depending on the site of infection and on the local concentration of the virulent factor, the translational control is differently hijacked. These processes are indeed dynamic equilibria *in vivo*, where translational stimulation and inhibition are therefore dose-dependently balanced. Noteworthy, our results unveil that both the protein interactions with the membrane, i.e. the monomer binding and the pore formation with ion imbalance, recast the

translational control of host gene and almost completely skip the transcriptional level of the cell.

### 1.3 Conclusion

This study stemmed from the following observations: i) the role exerted by translational and transcriptional control remodelling of host gene expression upon the virulent attack of low doses of pore forming toxins has been completely neglected as well as ii) their genome-wide descriptions; iii) the role of translation impairment due to pore forming toxin was studied only partially and at relatively high protein doses; iv) the role of post-transcriptional regulators of gene expression, namely ncRNAs and RNA binding proteins has never been evaluated nor taken into consideration in previous studies; v) the role of physiologically relevant doses of pore forming toxins, i.e. sublytic<sup>52</sup> has never been clearly addressed using native pore forming toxins.

Here, we chose a host-virulent factor model system comprising human cell lines and staphylococcal  $\alpha$ HL, the archetypal pore forming toxins. We used this model as a proof of principle for future studies and to provide new useful information concerning the involvement of translational trans-factors, such as RNA binding proteins, as active players in the host-pathogen response.

*S.aureus* is a commensal pathogen bacteria of human skin involved in different diseases such as pneumonia, infective endocarditis, skin infections, arthritis, and septic shock<sup>46</sup> and has been related with the squamous cell carcinoma of the skin<sup>49</sup>. Moreover, methicillin resistant *S. aureus* strains are leading cause of hospital and community infections<sup>93</sup>, raising the study of its pathogenicity as a worldwide priority. As a gram positive it secretes a wide variety of virulence factors, among the others  $\alpha$ HL has been proved to be the protein factor directly causing several pathologies<sup>26-28,30</sup>.

Few studies addressed the effects of so called low doses of native  $\alpha$ HL or of non-lytic monomeric  $\alpha$ HL (using the mutant H35N) on the host cells, showing quite different cellular impact ranging from cellular proliferation<sup>12</sup>, caspase activation via mitochondrial

pathway and oligonucleosomal DNA fragmentation<sup>61</sup>, to Akt inhibition<sup>13</sup> and histone methylation<sup>36</sup>. The great majority of the studies concerning the cellular effects of PFTs focuses on host-cell response to pore formation rather than to “sublytic” concentrations, i.e. in a condition where very few pores are inserted into the membrane and most of the protein is bound to the membrane as monomers. This fact is important because up to now an accurate estimation of the toxin concentration in tissues during severe or chronic infections has not been addressed<sup>52</sup>. The *in vivo* concentration of virulence factors likely depends on the site of infection and on the distance between the target cell and the bacteria producing the toxin<sup>94</sup>. Even if not strictly known, it can be supposed that physiological concentration of toxin are almost sublytic *in vivo*<sup>52</sup>. Therefore, the effects of short exposure and/or sublytic doses of pore forming toxins to host cells may be those relevant for understanding the physiopathology of these virulence factors *in vivo*. Given the mechanism of action of PFTs, the balance between monomers and oligomers depends on several factors, not all well understood. In our study we aimed at using the native  $\alpha$ HL, rather than the non-lytic mutant, because it is the most pertinent for getting information on possible translational control impairment relevant to human infections. We defined two cell-specific windows of low doses whose effect was not cytolytic. We employed these concentrations to probe the changes of well known modulators of translational initiation, to quantify the *de novo* protein synthesis, to obtain genome-wide information about the movement of transcripts at the translational level by polysomal profiling and finally to demonstrate the involvement of the RBP ELAVL1.

The initiation of translation is one of the most important mechanisms modulating a rapid, reversible and spatially controlled cellular response during certain stimuli or during any physiological process occurring within the cell<sup>62</sup>. The activation of initiation factors is finely tuned as a response of downstream signaling platforms and plays fundamental role during cell remodeling. A well known pathway controlling the activation of key components of protein synthesis is the Akt/mTOR axis<sup>66</sup>. Akt is the major mTor upstream activator able to integrate membrane signals<sup>95</sup>. In such a way, Akt and mTor give rise to a complex control of gene expression, playing an evolutionary conserved role to

supervise metabolic activities such as protein synthesis in response to extracellular stimuli. The most important downstream targets of mTORC<sub>1</sub> include the eukaryotic initiation factor 4E-binding protein-1 (4E-BP1) and the 40S ribosomal protein S6 (RPS6). Upon phosphorylation, 4E-BP1 dissociates from the cap-binding protein eukaryotic initiation factor 4E (eIF4E). Once free, eIF4E binds the cap of each mRNA and recruits other initiation factors (such as eIF4G, eIF4A and eIF3) at the 5' cap, allowing translation initiation to proceed<sup>96</sup>. At the same time mTORC<sub>1</sub> activates S6K1, responsible for the phosphorylation of RPS6 and of the co-helicase initiator factor 4B (eIF4B)<sup>62,64,97</sup>. In recent studies<sup>9,11</sup> translation has been addressed at lytic concentrations of PFTs and an overall inhibition of protein synthesis has been observed to be eIF2A dependent<sup>11</sup>, but the role of cellular control mechanisms mediating these effects has not been considered. Recently, infection of *Pentomophyla* producing pore forming toxins in *Drosophila* gut was studied. The inhibition of translation by the impairment of the mTor pathway has been observed, but only the 4E-BP dephosphorylation was addressed<sup>9</sup>. We filled the gap obtaining a complete description of the possible activation and/or inactivation of translational effectors (Akt, 4E-BP, eIF4E, eIFB, RPS6K, RPS6) in three cell lines. After two hours of incubations of SH-SY5Y and HeLa with the lowest concentration in the sublytic window, we found that RPS6K, its downstream target RPS6 and eIF4B, as well as eIF4E and Akt were activated. As stated above, all these stimuli are compatible with translation stimulation, even if the function of eIF4E phosphorylation remains controversial<sup>98,99</sup>. Surprisingly, this overall effect was counteracted by the dephosphorylation of 4E-BP, which locks eIF4E impeding its binding to mRNA and inhibiting cap-dependent translation<sup>62</sup>. To understand these concurrent opposite effects and any possible impairment of translation, we first analyzed the polysome content in the cells and quantified the *de novo* synthesis after cell injury. We found that the incorporation of a methionine analog, homo-azidoalanine, was slightly decreased when compared to the control, but this effect did not conversely reflect into changes in the polysome vs ribosome ratio. Again, this result can be interpreted as the sum of different and probably opposite messages to the translational machinery. Indeed mTor did not account for the whole translational

control and 4E-BP1 does not control the entire transcriptome, as already demonstrated in yeast<sup>100</sup>. Therefore, we wondered if: i) this effect could reflect into concomitant mRNA-specific up-loading and/or down-loading from the protein synthesis machinery in response to  $\alpha$ HL exposure and/or ii) the two opposite messages on translation are caused specifically by monomers or by the few pores which may still form at sublytic doses.

An effective way to address the first point is to unearth translational specific reshaping of gene expression by physically separating the transcripts associated to polysomes and involved in active translation from the non translating mRNAs. Polysomal profiling by sucrose gradient fractionation of cellular extracts was the technique of choice<sup>77</sup>. Therefore, we compare the transcriptome and translome profiles to uncover any selective up-loading into polysomes of mRNAs, to understand the global up or down-regulation of mRNAs and to disclose any underlying translational controls damage during the membrane attack. The transcripts have been sequentially quantified by high-throughput assays. Parallel comparison of polysomal profiling (quantifying the translome) with conventional total mRNA profiling (quantifying the transcriptome)<sup>20</sup>, gave us the possibility to detect 782 genes selectively regulated only at the translational level, while the transcriptional level was substantially poorly modified. From the microarray analysis we conclude that the translational response to  $\alpha$ HL is globally independent from the transcriptional response, as resumed by the low spearman correlation value associated to all the genes (0.19) and to DEGs (0.28). According to the number of DEGs, cells replied consistently to the  $\alpha$ HL treatment by changing the mRNAs uploaded onto polysomes rather than inducing *de novo* mRNAs transcription. 95% of DEGs have independent variations, only 5% show a significant concordance between transcriptional and translational regulation. Given the majority of significant variations (78%) to be exclusively translational, a bypassing of transcription and a predominant role of translational regulation seems to be implicated in mediating response to sublytic doses of PTFs. Interestingly, in line with our results, scarce transcriptional variations have been observed as the host response of HeLa cells to pore-independent treatments with  $\alpha$ HL<sup>36</sup>. So far our results represent the

first global portrait of transcriptional and translational changes during pore forming interaction with cells. In fact, only few pore-dependent transcriptional changes have been addressed so far. The activation of a reporter gene has been studied in lymphoblastoid cell lines in response to low doses of  $\alpha$ HL and found to induce the production of mediators that contribute to the initiation and propagation of inflammatory lesions<sup>35</sup>. Moreover, analyzing by Serial Analysis of Gene Expression (SAGE) a portion of the transcriptional profile after  $\alpha$ HL pore formation in keratinocytes, immediate early genes appear to be upregulated and to mediate cellular proliferation after damage recovery<sup>12</sup>. We found that among the translational DEGs 64% were down-regulated and 36% up-regulated. These results are in accordance with the opposite signals previously described and demonstrate that  $\alpha$ HL did not induce a simple complete block of translation.

Functional enrichment analysis confirmed that the two sets of transcriptome and translome DEGs are involved in disjointed biological functions. Among the terms with higher enrichment, signal transduction, cell adhesion and immune response were the most relevant as translational specific terms. This signal transduction and protein G involvement<sup>101</sup> as well as the obvious immune host cell response<sup>52</sup> are indeed expected biological processes perturbed by the  $\alpha$ HL treatment. Remarkably, the down-regulated genes are significantly associated to membrane processes, meaning that the reshaping of gene expression induced by events occurring at the membrane compartment promote a coherent response toward the same compartment. An up-regulation of cell-cell adhesion and motility was also observed. In line with these evidences and with the validation of the enhancement in cell motility after  $\alpha$ HL treatment, it has been proven that  $\alpha$ HL plays a role in the extracellular matrix reorganization and in mediating invasiveness of *S. aureus* in tissues<sup>102-104</sup>.

Enrichment analysis of experimentally known binding sites of the mRNAs UnTranslated Regions (UTRs) did result in many RNA binding protein (RBPs) to be significantly associated to transcriptome or translome DEGs, but none single miRNAs appeared as significant. Nevertheless, an enrichment of the binding sites of AGO1, an ubiquitously expressed protein able to associate to miRNAs and siRNAs<sup>83</sup>, could be observed. Strictly

speaking, AGO1 is not an RNA binding protein because its interaction with the mRNA is miRNA-mediated. Regardless of the mechanism, AGO1 impairs the translational efficiency and/or induces mRNAs degradation. Accordingly it is clear, by looking at transcriptional or translational DEGs known to be AGO1 target, that miRNA silencing pathways could be involved in the translational down-regulation of several genes.

By looking at the binding sites of classical RBPs, two groups of proteins (HNRNPU, HNRNPC, HNRNPH and ELAVL1, FUS, CAPRIN1) emerged to be specifically enriched when considering translome DEGs, suggesting their involvement in translational specific response of the host cells to sublytic  $\alpha$ HL injury. Among the most important trans-factors responsible for translational control, ELAVL1 (a.k.a. HuR) is the most studied. Not surprisingly, we found ELAVL1 to be enriched in translational DEGs. In fact, ELAVL1 is a pleiotropic protein<sup>105</sup> able to regulate many physiological processes. ELAVL1 is known to act on target mRNAs as stabilizer and/or a translational enhancer, binding to AU-rich element (ARE) containing mRNAs<sup>86,106</sup>. Strikingly, the fact that it was predicted to bind in the most significant way the translome DEGs, suggests that it may play a role in modulating the translational reorganization upon  $\alpha$ HL exposure. In accordance with this evidence, we demonstrated that ELAVL1 cytoplasmic level increased after sublytic treatment with  $\alpha$ HL. Since this RBP mediates stimulus-induced targeting of mRNAs from translational inactive ribonucleoprotein particles to polysomes<sup>107</sup>, we studied the changes between these subcellular compartment of four well known mRNA targets. One of the confirmed targets, for which we could observe a shift towards polysomal fractions after  $\alpha$ HL treatment, was p53. p53 is a well known and extensively studied transcription factor, whose activation was found to characterize cellular warning in response to various cellular stresses, owing a deep cellular reprogramming that induces cell cycle arrest, apoptosis, or cellular senescence to eliminate cells that have suffered irreparable damage. The network of p53 targets is also involved in angiogenesis, immune response and migration<sup>108</sup>.

The three other targets shifted from polysomes to the translationally silent compartments, as revealed by the decrease in the polysomal vs sub-polysomal fold-change. This

is indeed not obvious because the most common effect exerted by ELAVL1 is an enhancement in the stability of the mRNA targets. Nevertheless, it has been demonstrated to be involved in mRNAs destabilization under certain conditions<sup>109,110</sup> or to colocalize with translationally silent stress-granules<sup>106,111</sup>. Noteworthy, a destabilization of one ELAVL1 target was found when this RBPs acts as a co-factor of another RBPs, AUF1<sup>109</sup>. This is of outmost interested because AUF1 is a major attenuator of the inflammatory response and it is responsible for promoting mRNA degradation<sup>112</sup>. The emerging hypothesis that ELAVL1 may interact with AUF1 in order to modulate inflammatory response to pore-forming toxin needs further evaluation but undoubtedly represent a challenging and intriguing perspective for the field.

Finally, to address the second point, we used a non-lytic mutant and the purified oligomer of  $\alpha$ HL and clarified that the two opposite messages on translation, i.e. the stimulus and the inhibition suggested by the activation of initiation factors and by the dephosphorylation of 4E-BP, are caused specifically by monomers and by the few pores which may still form at sublytic doses, respectively. In accordance to our results, Chakrabarti et coworkers found an inhibition of 4E-BP in pore forming condition<sup>9</sup>. Also the recruitment of ELAVL1 was found to be monomer dependent, suggesting that specific translational control rather than only unspecific blockage of protein synthesis are acting and interplaying during sublytic attack of pore forming toxins and probably depending on the monomer to pore ratio. This fact may also depend on the lipid composition of the membrane. In this line, we demonstrated that the replenishment of cholesterol from the cell membrane prevent the activation of mTor downstream target and protect from the  $\alpha$ HL effects.

In conclusion, we performed for the first time a combined genome-wide analysis of the total and the polysomal RNA after sublytic  $\alpha$ HL stimulation of human cells. We observed a strong mRNA hijacking (i.e. selection of specific mRNAs) at the translational level and an almost complete bypass of the transcriptional level. This extensive uncoupling between transcriptome and translome variations shows the presence of a majority of membrane differentially expressed genes, proving that the cellular response to  $\alpha$ HL

stimulus is compartment-specific and characterized by a boost for translation of membrane related genes. This system approach provides new insights into the biology of an ancient fine-tuned response, portraying the first demonstration that a membrane perturbing stimulus induces a translational specific response of the target cell which recruits specific genes and trans-factors. Further studies will be required to understand which structural elements (*cis*-acting factors) on the mRNA may act in combination with *trans*-acting factors to finally decode a hypothetical membrane related translational operon.

## **1.4 Methods**

### **1.4.1 Cell culture**

Neuroblastoma cells SH-SY5Y, CHP-134 and HeLa cells were maintained in DMEM (SH-SY5Y and HeLa) or RMPI (CHP-134) supplemented with 10% Fetal Bovine Serum (FBS), 2 mM L-glutamine, penicillin 100.000 U.I./l and streptomycin 100 mg/l. Cells were cultured at 37 °C in humid atmosphere of 5% CO<sub>2</sub>. All experiments performed in the presence of αHL were performed using medium without phenol red, FBS, and antibiotics to avoid any interference with protein. If not specified differently, cells were grown to 80% of confluence before each assay.

### **1.4.2 Adenilate Kinase assay**

Cells were seeded at  $1 \times 10^4$ /well in 96 well plates (NUNC 96 fw luminescence compatible) in the presence of 100µl of medium. After 24 h cells were treated with αHL at decreasing concentration (2µM, 1µM, 200nM, 5nM, 12nM, 3nM, 0,5nM) for 4h and 24 h. Following manufacturer's' protocol, Adenilate Kinase detection reagent was added (ToxiLight® BioAssay Sample Kit – Lonza). The levels of Adenilate Kinase was measured by luminescence (TECAN infinite M200). Experiments were run in triplicate; t-test ( $P < 0.05$ ) was applied to validate differences.

### **1.4.3 Determination of hemolytic activity**

RRBCs were purchased from Zootechnica Il Gabbiano (Casole dElsa, Siena, Italy). The time course of hemolysis was determined from the turbidity at 650 nm in a 96-well microplate reader (UVmax, Molecular Devices, Sunnyvale, CA, USA) for 45 min. αHL were twofold serially diluted in the buffer used for washing (100 mM NaCl, 30 mM Tris-HCl, 0.1 mM EDTA pH 7.0 for RRBCs). Erythrocytes, at a 0.13% (v/v) concentration, were added immediately before starting kinetic measurement. The percentage of hemolysis was calculated as

$$\%HA = 100 \cdot (A_i - A_f) / (A_i - A_w) \quad (1)$$

where  $A_i$  and  $A_f$  are the absorbances at the beginning and at the end of the reaction, and  $A_w$  is the absorption after complete lysis of cells in pure water.

#### 1.4.4 Protein extraction and immunoblotting

Cells were seeded  $7 \times 10^5$  cells/well in 6-well plates till reaching 80% of confluence. After extensive washes with PBS to remove FBS, fresh DMEM *free* was added in the presence of convenient amount of  $\alpha$ HL for incubation time depending on experiment to experiment. To extract cytoplasmic proteins, cells were washed with cold PBS and harvested by scraping at 4°C. Cell suspension were lysed in a RIPA buffer containing Tris-base 50 mM, NaCl 150 mM, Igepal 1%, EDTA 1 mM, Sodium deoxycholate 0.5%, phosphatase cocktail I and II (Sigma), and protease inhibitor (Sigma). After centrifugation at 15000 for 20 min at 4°C to remove the nuclei and cellular debris, the supernatants were collected and stored at -80°C. Protein concentration was determined by the Bradford method (Sigma). Proteins were resolved using 10% or 12% SDS-PAGE in running buffer containing Tris base 25 mM, glycine 190 mM, SDS 0.1%, pH 8.3, using loading buffer 4X (SDS 8%, Tris-HCL 0.2M, glycerol 40%,  $\beta$ -mercaptoethanol 10%, phenol blu 0.004%, pH 6.8) and blotted on PVDF membrane in transfer buffer containing methanol 20%, 25 mM Tris base, 190 mM glycine, SDS 0.1% pH 8.3. Blots were probed overnight at 4°C with antibodies against the proteins of interest. After incubation with secondary antibodies conjugated to horseradish peroxidase the protein was detected using ECL prime detection reagent (GE Healthcare, Amersham). The luminescence was acquired by ChemDoc-It (BioRad) and analyzed with ImageQuant TL software (BioRad). Statistical analysis (t-test,  $p < 0,05$ ) of western blot data was performed on the densitometry values obtained with the software ImageQuant™ TL (BioRad). For immunoblotting and microscopy staining the follows primary antibodies were used: rabbit p-Akt (phosphorylation site Ser-437, dilution 1:1000), rabbit p-RPS6K (phosphorylation site Ser-441, dilution 1:1000) and rabbit Ribosomal protein S6 (1:500) were purchased from Cell Signaling; rabbit Akt

(1:1000), mouse 4E-BP1 (1:1000) and p-4E-BP1 (phosphorylation site Ser-65 and Thr-70, dilution 1:500), rabbit eIF4B (1:1000) and p-eIF4B (phosphorylation site Ser-422, dilution 1:500) rabbit p-S6K (phosphorylation site Ser-235 and Ser-236, dilution 1:1000), mouse eIF4E (1:1000) and p-eIF4E (phosphorylation site Ser-209, dilution 1:1000), mouse Hur (1:500), goat  $\beta$ -actin (1:2000) and were purchased from Santa Cruz Biotechnology Inc, CA; sheep alpha Hemolysin (1:1000) from Abcam. Secondary antibody anti-rabbit, anti-mouse or anti-goat, conjugated with HRP (Horseradish peroxidase) were used at final dilution of 1:10000 and purchased from Santa Cruz Biotechnology.

#### **1.4.5 Methyl- $\beta$ CD treatment**

SH-SY5Y cells were seeded at  $7 \times 10^5$  cells/well in 6-well culture plates (Corning, NY) and when reached 80% confluence were incubated for 1 h and 30 min at 37°C in DMEM with 10 mM of M- $\beta$ CD. The cytoplasmic portion of proteins was extracted as previously described (see Protein extraction paragraph).

#### **1.4.6 Recombinant protein expression**

Recombinant  $\alpha$ HL wild type and the mutant H35N proteins were prepared by coupled *in vitro* transcription and translation (IVTT). Briefly, radio-labelled  $\alpha$ HL was produced by coupled *in vitro* transcription and translation (IVTT) using an *E. coli* T7-S30 expression system for circular DNA (Promega) according to the manufacturer's instructions. See chapter 2 for more details. The genes of WT- $\alpha$ HL and H35N in pT7 plasmids were received from Hagan Bayley (Chemistry Research Laboratory, University of Oxford)

#### **1.4.7 Fluorescence microscopy**

Subcellular localization of ELAV1 was studied in SH-SY5Y cell lines before and after  $\alpha$ HL treatment by immunofluorescence. Cells were seeded at  $5 \times 10^4$  cell/coverslip in 6 well plates and maintained in culture for 24 h. After treatment with 3 nM  $\alpha$ HL for 2 h, cells were washed in PBS, fixed with 4% PFA and permeabilized with 0.05% triton in PBS

for 15 min. After passivation with 3% (w/v) BSA (Sigma-Aldrich) each sample was incubated with Phalloidin-488 first and then with the primary antibody for ELAV1, both at for 1h at RT. After incubation with anti rabbit Alexa-594 (Invitrogen) cells were mounted using ProLong<sup>®</sup> Gold reagent in the presence of DAPI (Invitrogen), which allows the staining of the nuclei. Epifluorescence microscope (Nikon Eclipse 90i.) was used to acquire the images. The fluorescence emission filters used covers 440-470 nm, 495-525 nm and 587-722 nm for DAPI and Alexa-594, respectively. Fluorescent secondary antibody purchased from Sigma-Aldrich.

#### **1.4.8 De-novo protein synthesis assay**

Ten thousand cells/well were grown for 24 h in 96-well plates and then treated with 3nM  $\alpha$ HL for 2 h. Protein synthesis was measured using Click-iT AHA Alexa Fluor 488 Protein Synthesis HCS Assay (Life Technologies) following the manufacturer's protocol. Cells were fixed and incubated with fluorescent alkyne to label the AHA incorporated into nascent proteins. The relative AHA incorporation was assessed by using Operetta HCS. For each sample replicates 400-500 cells were considered and the mean fluorescence intensity/well was obtained. The % of AHA incorporation was measured as follows

$$\%AHA = 100 \cdot (F^{\text{treated}} - F^{\text{background}}) / (F^{\text{untreated}} - F^{\text{background}}) \quad (3)$$

where  $F^{\text{treated}}$  is the mean fluorescence/well in the cytoplasm of cells treated with 3 nM  $\alpha$ HL,  $F^{\text{background}}$  is the mean fluorescence/well of the well background and  $F^{\text{untreated}}$  is the mean fluorescence/well in the cytoplasm of control cells.

#### **1.4.9 Polysomal profiling**

SH-SY5Y cells were seeded at a density of  $2.5 \times 10^4$  cells/cm<sup>2</sup> and maintained for 3 days in the growth medium. Once the 80% confluence was reached cells were incubated with  $\alpha$ HL 3nM for 2h. Afterwards the treated samples and the corresponding controls were

incubated for 3-4 minutes with cycloheximide 10 µg/ml at 37°C to trap the ribosomes on the mRNAs. Cells were washed with phosphate buffered saline (PBS + cycloheximide 10 µg/ml) and scraped directly on the plate with 300 µl lysis buffer (10 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM Tris-HCl, pH 7.5, 1% Triton X-100, 1% sodium deoxycholate, 0.2 U µl<sup>-1</sup> RNase inhibitor (Fermentas), cycloheximide 10 µg/ml dithiothreitol 1 mM) and transferred to an Eppendorf tube. After a few minutes of incubation on ice with occasional vortexing, nuclei and cellular debris were removed by centrifugation for 5 min at 12,000 g at 4°C. The supernatant was directly transferred onto a 15-50% linear sucrose gradient containing 30 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, and centrifuged in a Sorvall ultracentrifuge on a swinging rotor for 100 min at 180,000 g at 4°C. The fraction corresponding to the 80S peak and those corresponding to the polysomes were collected monitoring the absorbance at 254 nm. Each fraction was aliquoted, flash frozen in liquid N<sub>2</sub> and stored at -80°C for further RNA extraction.

#### **1.4.10 RNA extraction**

For polysomal and subpolysomal RNA extraction, the corresponding fractions were collected and treated with proteinase K 100 g/l and SDS 1% for 2 hours at 37°C. After phenol-chloroform extraction and isopropanol precipitation, polysomal and subpolysomal RNA samples were resuspended in 30 µl of water, quantified by reading absorbance at 260 nm with Nanodrop ND-1000 UV-VIS Spectrophotometer, aliquoted and stored at -80°C for further analysis.

For total RNA extraction,  $7 \times 10^5$  cells/ml were seeded into 6-well plates and cultured for 36 hours and then washed with PBS and treated with 3nM αHL as previously described for 2 hours. An extensive wash with PBS, Trizol Reagent (Invitrogen) was added directly to the cell monolayers. After 5 minute at room temperature, 200µl chloroform was added and centrifuged at 20000 g for 15min at 2°C. The aqueous phase containing the RNA was transferred to a new tube, 500µl isopropanol was added kept 10 min at room temperature, then centrifuged again at 20000g and 10 minutes at 2°C in order to

precipitate the ribonucleic acids. The pellet was washed with 80% ethanol, precipitated again at 20000g for 20 min, dried, redissolved in 30  $\mu$ l RNase free water, quantified by reading absorbance at 260 nm with Nanodrop ND-1000 UV-VIS Spectrophotometer, aliquoted and stored at -80°C for further analysis.

All RNA samples quality was assessed by agarose gel electrophoresis and by the Agilent 2100 Bioanalyzer platform.

#### **1.4.11 Quantitative Real time-PCR (q-PCR)**

For real-time PCR analysis, cDNA synthesis was performed using ReversAid™ reverse transcriptase (Fermentas) and random primers or SuperScript VILO cDNA kit (Invitrogen). RNA (1 $\mu$ g/reaction) was reverse-transcribed to single stranded cDNA, according to the manufacturer's instructions. q-PCR experiments were run in technical triplicate and biological triplicate using Bio-Rad CFX and 96-wells or 394-wells plates. The primers used for q-PCR are listed in Table S1. Data were analysed with the software BioRad CFX-Manager 1.6. Relative quantification of target genes was determined calculating the delta cross-threshold ( $\Delta$ Ct) and the relative  $\Delta\Delta$ Ct after normalization with the housekeeping gene Alu-J (for SYBER probe) or with the geometric mean of four different housekeeping genes, GAPDH, ACTIN, PPIA and MRPC19 (for TaqMan probe), according to the Pfaffl method (Pfaffl, 2001, Vandesompele 2000).

#### **1.4.12 Microarray**

Cells were harvested and washed twice with ice-cold PBS. For total RNA profiling, RNA was extracted using trizol-chloroform extraction and isopropanol precipitation. For polysomal RNA profiling, sucrose gradient fractions of polysomal RNA were collected and purified using phenol-chloroform extraction and isopropanol precipitation. All the RNA samples were submitted to microarray analysis using the Agilent-014850 Whole Human Genome Microarray 4x44K G4112F chip from Agilent Technologies, (Santa Clara, CA). Three biological replicates were done for each condition (control and treated at 2 hours

for total and polysomal RNA) and twelve hybridization arrays are obtained. cRNA probe generation as along with array hybridization, washing and staining were carried out according to the standard One-Color Microarray-Based Gene Expression Analysis (Quick Amp Labeling) protocol. Hybridized microarray slides were scanned with an Agilent DNA Microarray Scanner G2505C at 5µm resolution with the manufacturer's software (Agilent ScanControl 8.1.3).

The scanned TIFF images were analyzed numerically and the background corrected using the Agilent Feature Extraction Software version 10.7.7.1 according to the Agilent standard protocol GE1\_107\_Sep09. The output of Feature Extraction was analyzed with the R software environment for statistical computing (<http://www.r-project.org/>) and the Bioconductor library of biostatistical packages (<http://www.bioconductor.org/>). 9673 low signal Agilent probes, distinguished by a repeated “absent” detection call across the majority of the arrays in every condition, were filtered out from the analysis, leaving 30922 probes corresponding to 16087 genes with the HGNC gene symbol. Signal intensities across arrays were normalized with the quantile normalization algorithm. DEGs were determined adopting a double threshold based on 1) the magnitude of the change ( $\log_2$  fold change  $> 0.5$  and  $< -0.5$  for induced and repressed genes, respectively); 2) the statistical significance of the change, measured with a rank product test (p-value  $< 0.01$ ) implemented in the Bioconductor RankProd package. All microarray data are available through the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>).

#### **1.4.13 Gene Ontology Enrichment analysis**

The DAVID resource was used for enrichment analysis of the transcriptome and the translome DEGs lists, using annotations from Gene Ontology (<http://www.thegeneontology.org> website), KEGG (<http://www.genome.jp/kegg/> website), PFAM (<http://pfam.sanger.ac.uk/> website). The significance of overrepresentation was determined using a p-value threshold of 0.05.

#### **1.4.14 AURA enrichment analysis**

The AURA resource (<http://aura.science.unitn.it/>) was used for enrichment analysis of post-transcriptional regulators, based on the annotation of binding sites on UTR regions of transcriptome and translome DEGs. Enrichment was tested with the Fisher exact test. The significance of over-representation is determined at a 0.05 p-value threshold.

#### **1.4.15 FACS**

SH-SY5Y ( $7.5 \times 10^5$ ) and HeLa ( $2.5 \times 10^5$ ) cells were plated in petri dish with 5 ml DMEM medium and cultured overnight before starvation in medium without FBS and antibiotics for 24 h. For cytofluorimetric DNA analyses, after eight hours treatment, cells were harvested by gentle trypsinization and fixed by rapid submersion in ice-cold 70% ethanol. After overnight fixation at  $-20^\circ\text{C}$ , DNA was stained in an appropriate volume of staining solution containing 0.50  $\mu\text{g/ml}$  propidium iodide, 200  $\mu\text{g/ml}$  RNase, 0.1% TritonX-100 and EDTA 0.1 mM in phosphate buffer (PBS) pH 7.0 for 1 hour. Alternatively samples were stored at  $-20^\circ\text{C}$  for two weeks. A total of  $1.5 \times 10^4$  to  $3 \times 10^4$  cells per sample were collected in linear amplification mode and analyzed by FACS CANTO. A crosshair was set to the center of the G<sub>0</sub>/G<sub>1</sub> (2N) population of control cells

#### **1.4.16 Migration assay**

A scratch test was performed on SH-SY5Y cultured in 35 mm dishes. The day after the medium was changed in DMEM without FBS and antibiotics for 16h. A needle was used to scratch and remove cells from a discrete area of the confluent monolayer to form a cell-free zone into which cells at the edges of the wound can migrate. Cells were treated with 3 nM  $\alpha\text{HL}$  and movements were captured at regular intervals (2h) within 9h. Images were acquired by using an optical microscope.

#### 1.4.17 Lipid extraction

Cells at 80% confluence ( $\sim 10^8$ ) were harvested and washed twice with ice-cold PBS before trypsinization. The cell suspension was centrifuged at 2100 rpm for 10 min (RT). The pellet was resuspended in 200  $\mu$ L of mQ water and mixed with 3 mL of a solution of tetrachloromethane:methanol (2:1). Samples were then sonicated for 15 min at + 4°C with a pulse sonication (3 sec at 20W follow by 5 sec pauses). After centrifugation at 10000 rpm for 10 min at 4 °C the lower phase was transferred onto a dark glass tube and kept under a nitrogen atmosphere at - 80°C before NMR analysis.

#### 1.4.18 Planar lipid bilayer recording

This method is described extensively in chapter 2. Briefly, a two compartment (1 mL each) chamber was used as a support for the vertical bilayer device. The two compartments were separated by a 20  $\mu$ m-thick Teflon film with a central aperture ( $\sim 80$ -120  $\mu$ m) created “zapping” the film with a high-voltage spark generator. The aperture was pre-treated with a mixture pentane:hexadecane (10:1) and the bilayer was formed by flowing 1,2-diphytanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids) monolayers across both sides of the aperture. Experiments were performed under symmetrical buffer conditions (150 mM KCl, 100 mM NaCl, 2 mM MgCl<sub>2</sub> and 10 mM HEPES at pH 6.5) using electrodes Ag/AgCl with 3% agar bridges in 3.0 M KCl.  $\alpha$ HL protein monomers were added to the *cis* (grounded) side and a positive potential of +120 mV was applied. The currents were recorded by a patch clamp amplifier (Axopatch 200, Axon Instruments). A PC equipped with a DigiData 1200 A/D converter (Axon Instruments) was used for data acquisitions at a sampling frequency of 5 kHz. During data processing the signal was filtered at 0.100 kHz by a digital filter. The acquisition software was Axoscope 8 (Axon Instruments). Measurements were performed at  $20 \pm 2$  °C.

#### 1.4.19 List of TaqMan probes and primers used

All the primers used with TaqMan or SYBR Green are listed. For TaqMan qRT-PCRs, 20X TaqMan gene expression mix, 1µl cDNA and RNase-free water were mixed with 2X KAPA Probe Fast qPCR to a total volume of 10 µl. For SYBR Green analysis, reactions were carried out in a final volume of 20 µl adding 1 µl cDNA (diluted 1:5) using using the follow PCR protocol: 3 min - 95°C activation; 10 sec - 95°C denaturation, 30 sec – 55°C annealing; 5 sec 72°C extension; 40 cycles; melting ramp 65°C to 95°C.

<i>Gene symbol</i>	<i>Assay</i>	<i>ID/sequence</i>
MARVELD3	TaqMan	Hs00369354_m1
SLC17A7	TaqMan	Hs00220404_m1
SYTL1	TaqMan	Hs01070946_m1
LiCAM	TaqMan	Hs01109748_m1
DHH	TaqMan	Hs00368306_m1
ELFN1	TaqMan	Hs01062441_m1
NACA2	TaqMan	Hs01042642_s1
SERPINB9	TaqMan	Hs00244603_m1
ACTIN	TaqMan	Mm00607939_s1
GAPDH	TaqMan	Hs02758991_g1
PPIA	TaqMan	Hs04194521_s1
AluJ	SYBR Green	F: 5'-CAACATAGTGAAACCCCGTCTCT R: 5'-GCCTCAGCCTCCCAGTAG
P21	SYBR Green	F: 5'-CAGGGGACAGCAGAGGAA R: 5'-GGCGTTTGGAGTGGTAGAAA

CyclinD	SYBR Green	F: 5'-ACGAAGGTCTGCGCGTGTT R: 5'-CCGCTGGCCATGAACTACCT
TP53	SYBR Green	5'-TCAACAAGATGTTTTGCCAACTG 5'-ATGTGCTGTGACTGCTTGTAGATG
RPL14	SYBR Green	F: 5'-GATGGGCAAGAAGATTGAA R: 5'-TGGGAGAAGCTTTCAGGAGA

**Table 1.4.1. List of qRT-PCR primers.**

## 1.5 References

1. Schiavo, G. & van der Goot, F. G. The bacterial toxin toolkit. *Nature reviews. Molecular cell biology* **2**, 530–7 (2001).
2. Lemichez, E. & Barbieri, J. T. General aspects and recent advances on bacterial protein toxins. *Cold Spring Harb Perspect Med* **3**, a013573 (2013).
3. Menestrina, G. *et al.* Ion channels and bacterial infection: the case of  $\beta$ -barrel pore-forming protein toxins of *Staphylococcus aureus*. *FEBS Letters* **552**, 54–60 (2003).
4. Iacovache, I., van der Goot, F. G. & Pernot, L. Pore formation: an ancient yet complex form of attack. *Biochim. Biophys. Acta* **1778**, 1611–1623 (2008).
5. Gonzalez, M. & Bischofberger, M. Bacterial pore-forming toxins: the (w) hole story? *Cellular and molecular ...* (2008). doi:10.10007/s00018-007-7434-y
6. Bischofberger, M., Iacovache, I. & van der Goot, F. G. Pathogenic pore-forming proteins: function and host response. *Cell host & microbe* **12**, 266–75 (2012).
7. Walev, I. *et al.* Delivery of proteins into living cells by reversible membrane permeabilization with streptolysin-O. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3185–90 (2001).
8. Los, F. C. O. *et al.* RAB-5- and RAB-11-dependent vesicle-trafficking pathways are required for plasma membrane repair after attack by bacterial pore-forming toxin. *Cell host & microbe* **9**, 147–57 (2011).
9. Chakrabarti, S., Liehl, P., Buchon, N. & Lemaitre, B. Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the *Drosophila* gut. *Cell Host Microbe* **12**, 60–70 (2012).
10. Von Hoven, G. *et al.* Modulation of translation and induction of autophagy by bacterial exoproducts. *Med. Microbiol. Immunol.* **201**, 409–418 (2012).
11. Gonzalez, M. R. *et al.* Pore-forming toxins induce multiple cellular responses promoting survival. *Cell. Microbiol.* **13**, 1026–1043 (2011).
12. Haugwitz, U. *et al.* Pore-forming *Staphylococcus aureus* alpha-toxin triggers epidermal growth factor receptor-dependent proliferation. *Cellular microbiology* **8**, 1591–600 (2006).
13. Wiles, T. J., Dhakal, B. K., Eto, D. S. & Mulvey, M. A. Inactivation of Host Akt / Protein Kinase B Signaling by Bacterial Pore-forming Toxins. *Molecular Biology of the Cell* **19**, 1427–1438 (2008).
14. Husmann, M. *et al.* Differential role of p38 mitogen activated protein kinase for cellular recovery from attack by pore-forming *S. aureus* alpha-toxin or streptolysin O. *Biochemical and biophysical research communications* **344**, 1128–34 (2006).
15. Husmann, M. *et al.* Elimination of a bacterial pore-forming toxin by sequential endocytosis and exocytosis. *{FEBS} Lett.* **583**, 337–344 (2009).
16. Kao, C.-Y. *et al.* Global functional analyses of cellular responses to pore-forming toxins. *PLoS pathogens* **7**, e1001314 (2011).
17. Huffman, D. L. *et al.* Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10995–10000 (2004).

18. Vogel, C. Translation's coming of age. *Molecular systems biology* **7**, 498 (2011).
19. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–42 (2011).
20. Tebaldi, T. *et al.* Widespread uncoupling between transcriptome and translome variations after a stimulus in mammalian cells. *BMC genomics* **13**, 220 (2012).
21. Kitamura, H. *et al.* Genome-wide identification and characterization of transcripts translationally regulated by bacterial lipopolysaccharide in macrophage-like J774.1 cells. *Physiol. Genomics* **33**, 121–132 (2008).
22. Ceppi, M. *et al.* Ribosomal protein {mRNAs} are translationally-regulated during human dendritic cells activation by {LPS}. *Immunome Res* **5**, 5 (2009).
23. Colman, H. *et al.* Genome-wide analysis of host {mRNA} translation during hepatitis C virus infection. *J. Virol.* **87**, 6668–6677 (2013).
24. Klevens, R. M. *et al.* Estimating health care-associated infections and deaths in U.S. hospitals, 2002. *Public health reports (Washington, D.C. : 1974)* **122**, 160–6
25. Otto, M. Basis of virulence in community-associated methicillin-resistant *Staphylococcus aureus*. *Annual review of microbiology* **64**, 143–62 (2010).
26. Kebaier, C. *et al.* *Staphylococcus aureus*  $\alpha$ -hemolysin mediates virulence in a murine model of severe pneumonia through activation of the {NLRP3} inflammasome. *J. Infect. Dis.* **205**, 807–817 (2012).
27. Niebuhr, M., Mamerow, D., Heratizadeh, A., Satzger, I. & Werfel, T. Staphylococcal  $\alpha$ -toxin induces a higher T cell proliferation and interleukin-31 in atopic dermatitis. *Int. Arch. Allergy Immunol.* **156**, 412–415 (2011).
28. Niebuhr, M. *et al.* Staphylococcal alpha-toxin is a strong inducer of interleukin-17 in humans. *Infect. Immun.* **79**, 1615–1622 (2011).
29. Dhakal, B. K. & Mulvey, M. A. The {UPEC} pore-forming toxin  $\alpha$ -hemolysin triggers proteolysis of host proteins to disrupt cell adhesion, inflammatory, and survival pathways. *Cell Host Microbe* **11**, 58–69 (2012).
30. Powers, M. E., Kim, H. K., Wang, Y. & Bubeck-Wardenburg, J. {ADAM10} mediates vascular injury induced by *Staphylococcus aureus*  $\alpha$ -hemolysin. *J. Infect. Dis.* **206**, 352–356 (2012).
31. Fang, Y., Cheley, S., Bayley, H. & Yang, J. The heptameric prepore of a staphylococcal alpha-hemolysin mutant in lipid bilayers imaged by atomic force microscopy. *Biochemistry* **36**, 9518–22 (1997).
32. Czajkowsky, D. M., Sheng, S. & Shao, Z. Staphylococcal alpha-hemolysin can form hexamers in phospholipid bilayers. *J. Mol. Biol.* **276**, 325–330 (1998).
33. Bhakdi, S., Muhly, M., Korom, S. & Hugo, F. Release of interleukin-1 beta associated with potent cytotoxic action of staphylococcal alpha-toxin on human monocytes. *Infection and immunity* **57**, 3512–9 (1989).
34. Menestrina, G. Ion channels and bacterial infection: the case of  $\beta$ -barrel pore-forming protein toxins of *Staphylococcus aureus*. *FEBS Letters* **552**, 54–60 (2003).
35. Dragneva, Y. *et al.* Subcytotoxic attack by staphylococcal alpha-toxin activates {NF-kappaB} and induces interleukin-8 production. *Infect. Immun.* **69**, 2630–2635 (2001).

36. Hamon, M. A. *et al.* Histone modifications induced by a family of bacterial toxins. *Proc. Natl. Acad. Sci. {U.S.A.}* **104**, 13467–13472 (2007).
37. Alto, N. M. & Orth, K. Subversion of cell signaling by pathogens. *Cold Spring Harbor perspectives in biology* **4**, a006114 (2012).
38. Lax, a J. & Grigoriadis, a E. Pasteurella multocida toxin: the mitogenic toxin that stimulates signalling cascades to regulate growth and differentiation. *International journal of medical microbiology : IJMM* **291**, 261–8 (2001).
39. Gurcel, L., Abrami, L., Girardin, S., Tschopp, J. & van der Goot, F. G. Caspase-1 activation of lipid metabolic pathways in response to bacterial pore-forming toxins promotes cell survival. *Cell* **126**, 1135–45 (2006).
40. Wu, S. *et al.* A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nature medicine* **15**, 1016–22 (2009).
41. Clarke, J., Wu, H., Jayasinghe, L. & Patel, A. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature ...* **4**, (2009).
42. Orth, J. H. C. *et al.* Pasteurella multocida toxin activation of heterotrimeric G proteins by deamidation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7179–84 (2009).
43. Babb, R. C., Homer, K. a, Robbins, J. & Lax, A. J. Modification of Heterotrimeric G-Proteins in Swiss 3T3 Cells Stimulated with Pasteurella multocida Toxin. *PloS one* **7**, e47188 (2012).
44. Travaglione, S., Fabbri, A. & Fiorentini, C. The Rho-activating CNF1 toxin from pathogenic E. coli: A risk factor for human cancer development? *Infectious agents and cancer* **3**, 4 (2008).
45. Valeva, A. *et al.* Pro-inflammatory feedback activation cycle evoked by attack of Vibrio cholerae cytolysin on human neutrophil granulocytes. *Medical microbiology and immunology* **197**, 285–93 (2008).
46. Ferry, T., Perpoint, T., Vandenesch, F. & Etienne, J. Virulence determinants in Staphylococcus aureus and their involvement in clinical syndromes. *Current infectious disease reports* **7**, 420–8 (2005).
47. Malachowa, N. & DeLeo, F. R. Staphylococcus aureus survival in human blood. *Virulence* **2**, 567–9
48. Prosperi, M. *et al.* Molecular epidemiology of community-associated methicillin-resistant Staphylococcus aureus in the genomic era: a cross-sectional study. *Scientific reports* **3**, 1902 (2013).
49. Kullander, J., Forslund, O. & Dillner, J. Staphylococcus aureus and squamous cell carcinoma of the skin. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **18**, 472–8 (2009).
50. Kloft, N. *et al.* Pore-forming toxins activate MAPK p38 by causing loss of cellular potassium. *Biochemical and biophysical research communications* **385**, 503–6 (2009).
51. Berube, B. J. & Bubeck Wardenburg, J. Staphylococcus aureus  $\alpha$ -toxin: nearly a century of intrigue. *Toxins* **5**, 1140–66 (2013).

52. Aroian, R. & Goot, F. G. Van Der. Pore-forming toxins and cellular non-immune defenses ( {CNIDs} ). 57–61 (2007). doi:10.1016/j.mib.2006.12.008
53. Menestrina, G., Serra, M. D. & Prévost, G. Mode of action of beta-barrel pore-forming toxins of the staphylococcal alpha-hemolysin family. *Toxicon : official journal of the International Society on Toxinology* **39**, 1661–72 (2001).
54. Valeva, A. *et al.* Evidence that clustered phosphocholine head groups serve as sites for binding and assembly of an oligomeric protein pore. *J. Biol. Chem.* **281**, 26014–21 (2006).
55. Thay, B., Wai, S. N. & Oscarsson, J. Staphylococcus aureus  $\alpha$ -toxin-dependent induction of host cell death by membrane-derived vesicles. *PloS one* **8**, e54661 (2013).
56. Inoshima, I. *et al.* A Staphylococcus aureus pore-forming toxin subverts the activity of ADAM10 to cause lethal infection in mice. *Nature medicine* **17**, 1310–4 (2011).
57. Song, L. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science (New York, N.Y.)* **274**, 1859–66 (1996).
58. Olson, R., Nariya, H. & Yokota, K. Crystal structure of staphylococcal LukF delineates conformational changes accompanying formation of a transmembrane channel. *Nature Structural & ...* **6**, 134–140 (1999).
59. Montoya, M. & Gouaux, E. Beta-barrel membrane protein folding and structure viewed through the lens of alpha-hemolysin. *Biochimica et biophysica acta* **1609**, 19–27 (2003).
60. Miles, G., Jayasinghe, L. & Bayley, H. Assembly of the Bi-component leukocidin pore examined by truncation mutagenesis. *The Journal of biological chemistry* **281**, 2205–14 (2006).
61. Saumya S. Srivastava, S. P. A. S. N. A. A. R. and K. V. M. *et al.* Membrane Bound Monomer of Staphylococcal  $\alpha$ -Hemolysin Induces Caspase Activation and Apoptotic Cell Death despite Initiation of Membrane Repair Pathway. *PloS one* **4**, e6293 (2009).
62. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731–745 (2009).
63. Spriggs, K. a, Bushell, M. & Willis, A. E. Translational regulation of gene expression during conditions of cell stress. *Molecular cell* **40**, 228–37 (2010).
64. Silvera, D., Formenti, S. C. & Schneider, R. J. Translational control in cancer. *Nature reviews. Cancer* **10**, 254–66 (2010).
65. Zoncu, R., Efeyan, A. & Sabatini, D. M. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nature reviews. Molecular cell biology* **12**, 21–35 (2010).
66. Ma, X. M. & Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nature reviews. Molecular cell biology* **10**, 307–18 (2009).
67. Schepetilnikov, M. *et al.* Viral factor TAV recruits TOR/S6K1 signalling to activate reinitiation after long ORF translation. *The EMBO journal* **30**, 1343–56 (2011).
68. Kawate, T. & Gouaux, E. Arresting and releasing Staphylococcal alpha-hemolysin at intermediate stages of pore formation by engineered disulfide bonds. *Protein Sci.* **12**, 997–1006 (2003).
69. Thompson, J. R., Cronin, B., Bayley, H. & Wallace, M. I. Rapid assembly of a multimeric membrane protein pore. *Biophysical journal* **101**, 2679–83 (2011).
70. Harris, B., Pereira, I. & Parkin, E. Targeting {ADAM10} to lipid rafts in neuroblastoma {SH-SY5Y} cells impairs amyloidogenic processing of the amyloid precursor protein. *Brain Res.* **1296**, 203–215 (2009).

71. Wilke, G. A. & Bubeck Wardenburg, J. Role of a disintegrin and metalloprotease 10 in *Staphylococcus aureus* alpha-hemolysin-mediated cellular injury. *Proc. Natl. Acad. Sci. {U.S.A.}* **107**, 13473–13478 (2010).
72. Zidovetzki, R. & Levitan, I. Use of cyclodextrins to manipulate plasma membrane cholesterol content: evidence, misconceptions and control strategies. *Biochim. Biophys. Acta* **1768**, 1311–1324 (2007).
73. Dieterich, D. C. *et al.* In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons. *Nat. Neurosci.* **13**, 897–905 (2010).
74. WARNER, J. R., KNOPE, P. M. & RICH, A. A multiple ribosomal structure in protein synthesis. *Proc. Natl. Acad. Sci. {U.S.A.}* **49**, 122–129 (1963).
75. Anand, M., Chakraborty, K., Marton, M. J., Hinnebusch, A. G. & Kinzy, T. G. Functional interactions between yeast translation eukaryotic elongation factor (eEF) 1A and eEF3. *The Journal of biological chemistry* **278**, 6985–91 (2003).
76. Ortiz, P. A. & Kinzy, T. G. Dominant-negative mutant phenotypes and the regulation of translation elongation factor 2 levels in yeast. *Nucleic Acids Res.* **33**, 5740–5748 (2005).
77. Arava, Y. Isolation of polysomal {RNA} for microarray analysis. *Methods Mol. Biol.* **224**, 79–87 (2003).
78. Arava, Y. *et al.* Genome-wide analysis of {mRNA} translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. {U.S.A.}* **100**, 3889–3894 (2003).
79. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using {DAVID} bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
80. Krol, J., Loedige, I. & Filipowicz, W. The widespread regulation of {microRNA} biogenesis, function and decay. *Nat. Rev. Genet.* **11**, 597–610 (2010).
81. Szostak, E. & Gebauer, F. Translational control by 3'-{UTR-binding} proteins. *Brief Funct Genomics* **12**, 58–65 (2013).
82. Dassi, E. *et al.* AURA: Atlas of UTR Regulatory Activity. *Bioinformatics (Oxford, England)* **28**, 142–4 (2012).
83. Höck, J. & Meister, G. The Argonaute protein family. *Genome Biol.* **9**, 210 (2008).
84. Lafarga, V. *et al.* p38 Mitogen-activated protein kinase- and HuR-dependent stabilization of p21(Cip1) mRNA mediates the G(1)/S checkpoint. *Molecular and cellular biology* **29**, 4341–51 (2009).
85. Colombrita, C., Silani, V. & Ratti, A. ELAV proteins along evolution: Back to the nucleus? *Molecular and cellular neurosciences* (2013). doi:10.1016/j.mcn.2013.02.003
86. De Silanes, I., Zhan, M., Lal, A., Yang, X. & Gorospe, M. Identification of a target {RNA} motif for {RNA-binding} protein {HuR.}. *Proc. Natl. Acad. Sci. {U.S.A.}* **101**, 2987–2992 (2004).
87. López de Silanes, I. *et al.* Role of the RNA-binding protein HuR in colon carcinogenesis. *Oncogene* **22**, 7146–54 (2003).
88. Masuda, K. *et al.* Global dissociation of HuR-mRNA complexes promotes cell survival after ionizing radiation. *The EMBO journal* **30**, 1040–53 (2011).
89. Mazan-Mamczarz, K. *et al.* {RNA-binding} protein {HuR} enhances p53 translation in response to ultraviolet light irradiation. *Proc. Natl. Acad. Sci. {U.S.A.}* **100**, 8354–8359 (2003).

90. Jursch, R. *et al.* Histidine residues near the N terminus of staphylococcal alpha-toxin as reporters of regions that are critical for oligomerization and pore formation. *Infection and immunity* **62**, 2249–56 (1994).
91. Panchal, R. G. & Bayley, H. Interactions between residues in staphylococcal alpha-hemolysin revealed by reversion mutagenesis. *J. Biol. Chem* **270**, 23072–23076 (1995).
92. Srivastava, S. S. *et al.* Membrane bound monomer of Staphylococcal alpha-hemolysin induces caspase activation and apoptotic cell death despite initiation of membrane repair pathway. *{PLoS} {ONE}* **4**, e6293 (2009).
93. Sinha, B. & Fraunholz, M. Staphylococcus aureus host cell invasion and post-invasion events. *International journal of medical microbiology : IJMM* **300**, 170–5 (2010).
94. Menestrina, G., Dalla Serra, M., Pederzoli, C., Bregante, M. & Gambale, F. Bacterial hemolysins and leukotoxins affect target cells by forming large exogenous pores into their plasma membrane: Escherichia coli hemolysin A as a case example. *Bioscience reports* **15**, 543–51 (1995).
95. Ruggero, D. & Sonenberg, N. The Akt of translational control. *Oncogene* **24**, 7426–34 (2005).
96. Santhanam, A. N. *et al.* Role of 3'UTRs in the translation of mRNAs regulated by oncogenic eIF4E--a computational inference. *PLoS one* **4**, e4868 (2009).
97. Ma, X. M., Yoon, S.-O., Richardson, C. J., Jülich, K. & Blenis, J. SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* **133**, 303–13 (2008).
98. Minich, W. B., Balasta, M. L., Goss, D. J. & Rhoads, R. E. Chromatographic resolution of in vivo phosphorylated and nonphosphorylated eukaryotic translation initiation factor {eIF-4E:} increased cap affinity of the phosphorylated form. *Proc. Natl. Acad. Sci. {U.S.A.}* **91**, 7668–7672 (1994).
99. Svitkin, Y. V, Ovchinnikov, L. P., Dreyfuss, G. & Sonenberg, N. General {RNA} binding proteins render translation cap dependent. *{EMBO} J.* **15**, 7147–7155 (1996).
100. Cridge, A. G. *et al.* Identifying eIF4E-binding protein translationally-controlled transcripts reveals links to mRNAs bound by specific PUF proteins. *Nucleic acids research* **38**, 8039–8050 (2010).
101. Aktories, K. Bacterial protein toxins that modify host regulatory {GTPases}. *Nat. Rev. Microbiol.* **9**, 487–498 (2011).
102. Kwak, Y.-K. *et al.* The Staphylococcus aureus alpha-toxin perturbs the barrier function in Caco-2 epithelial cell monolayers by altering junctional integrity. *Infect. Immun.* **80**, 1670–1680 (2012).
103. Reiss, K., Cornelsen, I., Husmann, M., Gimpl, G. & Bhakdi, S. Unsaturated fatty acids drive disintegrin and metalloproteinase ({ADAM}-dependent) cell adhesion, proliferation, and migration by modulating membrane fluidity. *J. Biol. Chem.* **286**, 26931–26942 (2011).
104. Soong, G., Chun, J., Parker, D. & Prince, A. Staphylococcus aureus activation of caspase 1/calpain signaling mediates invasion through human keratinocytes. *J. Infect. Dis.* **205**, 1571–1579 (2012).
105. Lebedeva, S. *et al.* Transcriptome-wide analysis of regulatory interactions of the {RNA-binding} protein {HuR}. *Mol. Cell* **43**, 340–352 (2011).

106. Fan, X. C. & Steitz, J. A. Overexpression of {HuR}, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of {ARE-containing} {mRNAs}. *{EMBO} J.* **17**, 3448–3460 (1998).
107. Doller, A., Schulz, S., Pfeilschifter, J. & Eberhardt, W. {RNA-dependent} association with myosin {IIA} promotes F-actin-guided trafficking of the {ELAV-like} protein {HuR} to polysomes. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt663
108. Levine, A. J., Hu, W. & Feng, Z. The P53 pathway: what questions remain to be explored? *Cell Death Differ.* **13**, 1027–1036 (2006).
109. Chang, N. *et al.* {HuR} uses {AUF1} as a cofactor to promote {p16INK4} {mRNA} decay. *Mol. Cell. Biol.* **30**, 3875–3886 (2010).
110. Scott, G. K. *et al.* Destabilization of {ERBB2} transcripts by targeting 3' untranslated region messenger {RNA} associated {HuR} and histone deacetylase-6. *Mol. Cancer Res.* **6**, 1250–1258 (2008).
111. Kedersha, N. & Anderson, P. Stress granules: sites of {mRNA} triage that regulate {mRNA} stability and translatability. *Biochem. Soc. Trans.* **30**, 963–969 (2002).
112. Gratacós, F. M. & Brewer, G. The role of {AUF1} in regulated {mRNA} decay. *Wiley Interdiscip Rev {RNA}* **1**, 457–473 (2010).



## CHAPTER 2

### Detection of 3'-end RNA uridylation with a protein nanopore

## 2.1 Introduction.

### 2.1.0 Overview

In the first chapter we focused our attention on the translational regulation of gene expression based on a fine-tuned cellular mechanisms stimulated by  $\alpha$ HL exposure.

In this second chapter we take advantage of the  $\alpha$ HL pore to realize a new tool for the detection of biologically relevant RNA signatures. Here we present a surprisingly efficient RNA-binding nanopore able to detect at the single molecule level RNAs with 3' oligo U tails.

The 3'-end uridylation of several types of RNAs has been observed as a widespread post-transcriptional modification affecting the turnover and functionality of the tagged molecules<sup>1-3</sup> and, as a consequence, the biology of gene expression. For example, the 3' addition of a variable numbers (< 20) of uridines<sup>4,5,6</sup> plays a crucial role in determining the directionality of mRNA degradation and the turnover of microRNAs<sup>7-10</sup>. Means to determine the abundance and the length of oligo U tails require improvement because the currently available sequencing technologies are not optimal for the rapid and cheap detection of this RNA modification, especially for longer U tails<sup>4,6</sup>.

Here, stochastic sensing with protein nanopores has been developed for the rapid, label-free, amplification free and low cost detection of 3' oligo U tails. To this aim, we used the pore forming protein  $\alpha$ -hemolysin ( $\alpha$ HL), which is produced by *Staphylococcus aureus* and it is known to assemble into transmembrane heptameric “mushroom-shaped” pore<sup>11</sup> (See introduction chapter 1). X-ray diffraction analysis of the pore revealed a 2.6 nm *cis* aperture leading into a wider vestibule connected to a narrow transmembrane domain<sup>12</sup>. This 14-stranded  $\beta$ -barrel has an internal volume of  $\sim 20$  nm<sup>3</sup> and is connected to the vestibule through a constriction of  $\sim 1.5$  nm diameter<sup>12</sup>. The pairs of  $\beta$ -strands from each subunit form turns that line the  $\sim 2.2$  nm diameter *trans* entrance<sup>12</sup>. Intensive studies about the interaction and translocation of nucleic acids with and through the  $\alpha$ HL nanopore have allowed the analysis of single-stranded nucleic acid length<sup>13</sup> and DNA

hairpin unzipping<sup>14</sup>. The  $\alpha$ HL pore has been extensively used as a single molecule nanopore sensor<sup>15</sup>. Nucleobase recognition with nanopores has been investigated by different approaches such as the immobilization of the single strand inside the pore<sup>16-18</sup>, the changes in both the sizes and the structures of the nanopore<sup>19-21</sup> together with enzymatic<sup>22-24</sup> and solid state<sup>25,26</sup> improvements.

In the past, the selective recognition of short sequence signatures has been achieved using DNA probes that hybridize with the complementary DNA or RNA sequence<sup>27-31</sup> or using RNA binding proteins that selectively bind the target sequence before translocation of the oligonucleotide<sup>32</sup>. However, despite the relevance of RNA-sequence signatures in gene expression regulation<sup>33</sup>, little work has been done on single-molecule sensing of specific RNA. Here, we report the selective detection of 3' oligo U tails with the  $\alpha$ HL pore and we illustrate the interaction of this peculiar RNA signature with a specific recognition site of the protein. Finally, we describe the possibility to detect the length of this biologically relevant post-transcriptional modification and the presence of non-canonical U in the sequence, proposing a method for the purification of 3'-end oligo U RNA fragments from longer RNAs.

### **2.1.1 RNA 3'-uridylation**

The 3'-polyuridylation is a template-independent post-transcriptional RNA modification. The possible presence of this modification of the human transcriptome had been hypothesized at the end of 1950s<sup>34</sup>. In 1972 Burdon and Shenkin provided some evidences for 3' U-rich RNA sequences in mammalian cells<sup>35</sup>. Nevertheless, the modification became biologically interesting only as a consequence of studies on small non coding RNA (ncRNA) biology, along with the development of mRNA 3' end sequencing methods that are not reliant on oligo(dT)-primed reverse transcription<sup>3</sup>. The earliest direct sequence evidence for post-transcriptional RNA uridylation came from a study of a beet virus RNA<sup>36</sup>. The existence of many eukaryotic U-specific transferases (e.g. Cid1, GLD-2, ZCCHC11)<sup>37,38,39</sup> is an evidence of the RNA uridylation as an important and widespread

mechanism of gene expression regulation. The 3'-polyuridylation can induce both the stabilization<sup>40,41</sup> and destabilization<sup>42,43</sup> of the target oligo-ribonucleotide. Recent data show that this post-transcriptional modification occur on many different classes of RNAs<sup>4</sup> and different length of 3' U tails (generally shorter than 20 uridines) has been observed<sup>4,6</sup>. Uridylation of mRNAs and noncoding RNAs has been mainly reported in eukaryotes<sup>6</sup>. Polyuridylation has been described to destabilize the human let-7 miRNA but on the other hand mono-uridylation seems to have a stabilizing effect on miRNAs<sup>44</sup>. Oligouridylation of mRNAs in humans has been described to enhance the binding of the Lsm1-7 protein complex with the consequent decapping and 5' to 3'degradation (by the Xrn1 exoribonuclease) or activation of the Dis3L2 exoribonuclease for a 3' to 5' mediated degradation<sup>45-48</sup>. Interestingly, mRNA uridylation by both Cid1 and URT1 prevents 3' to 5' exosome ribonucleolytic attack and the UTR1 poly(U)-polymerase seems to exert its activity on polysomal loaded mRNAs<sup>7,49</sup>. The biological meaning of adding U-tails in 3' rather than poly(C) or poly(G) is unclear. The recruitment of specific RNA-binding proteins and/or the formation of 3' hairpins are some hypothesis<sup>9</sup>. What it is clear, is that this non-template RNA modification is an important signature acting as post-transcriptional regulator of RNA biogenesis, turnover and function.

### **2.1.2 The *state-of-art* on single molecule RNA detection.**

Single molecule techniques present many advantages to understand the physical, mechanical and biological kinetic parameters (association/dissociation constants of very fast reactions) that are not possible to be observed accurately in bulk solutions. Nanopore technologies based on the heptameric  $\alpha$ HL pore have been demonstrated to be flexible single molecule platforms and has been used to recognize the identity, structure and conformations of small molecules<sup>50</sup>, proteins<sup>51</sup>, DNA and RNA<sup>52</sup>.

Although its biological impact, little work has been done on RNA sensing and sequencing by nanopores. Possible reasons could be (i) the intrinsic instability of the RNA, (ii) complex and flexible 3D structures of many RNA sequences, (iii) the massive in-

vestments on DNA sequencing with, for example, the “1,000 genome” project of the National Institutes of Health started in 2004. Recently new developments on RNA sensing by biological and solid-state nanopores were published<sup>18,32,53-55</sup>. More than ten years after the Kasianowicz’s and Akelson’s papers on homopolimeric ssRNA translocation<sup>56,57</sup>, N. Dekker’s lab shows the possibility to discriminate between ssRNA and dsRNA with solid state nanopores<sup>54</sup>. The same lab published the electrophysiological characterization of translocation events of long (~ 5-30 kilobases) single-stranded RNA molecules through small (1.5 to 8 nm in diameter) solid state nanopores<sup>53</sup>. At the same time M. Drndic’s lab<sup>58</sup> presented the discrimination of dsRNA and tRNA translocating through a 3 nm silicon nitride solid state nanopore. After these, other two papers<sup>30,31</sup> described the possibility of microRNA detection with solid and biological nanopores, respectively. These latter works used oligo-deoxyribonucleotide probes that form short hybrid double helices on the target RNA, allowing the detection of the double helix by monitoring the ionic current flowing through the nanopore. Finally, Bayley’s lab presented the possibility to achieve a clear discrimination of the four RNA nucleobases when the oligonucleotide is immobilized within the  $\alpha$ HL nanopore<sup>18</sup>. Another interesting paper recently published by the Meller’s lab describes the possibility to study complex binding kinetics of RNA and RNA binding protein in a single molecule fashion<sup>32</sup>.

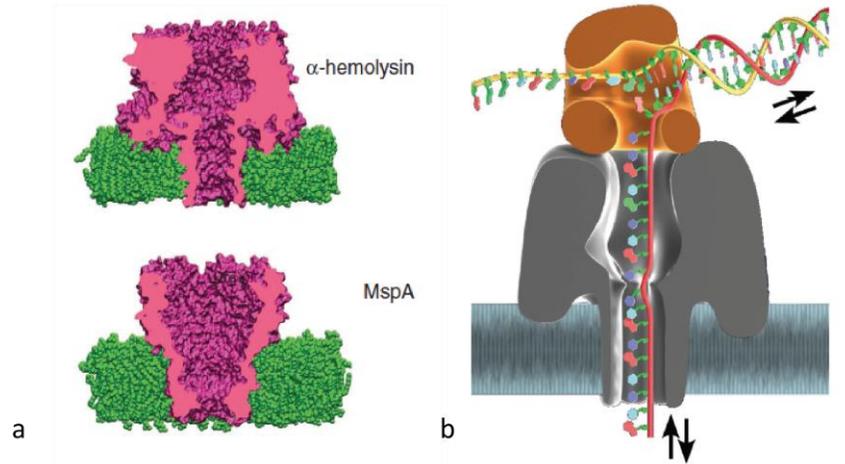
### **2.1.3 Biological nanopores for nucleic acid detection**

*Staphylococcal*  $\alpha$ HL has been widely used as a nanopore for single molecule sensing<sup>15</sup>. The principle of nanopore sensing is analogous to that of a Coulter Counter<sup>59</sup>. A potential is applied across a nanopore, which separates two chambers containing an electrolyte solution, allowing the electrophoretically passage of both an ionic current and a charged nucleic acid molecule through the pore. As nucleic acid bases pass through the main constriction of the channel, the ionic current flow changes. If the detection is achieved at the single base level, the nanopore can be used as cheap, label-free, amplification-free and low cost biological tools for the detection and sequencing of the nucleic

acids. In the past, it has been proposed that biological nanopores could be used to sequence DNA and RNA<sup>56</sup>. Biological nanopores that has been used detection or sequencing (such in the case of the  $\alpha$ HL pore or the *Mycobacterium smegmatis* porin A (MspA)<sup>60,61</sup>) have usually dimensions that allow the passage of single stranded ssDNA or ssRNA and do not permit dsDNA or structured RNA to be translocated (Figure 2.1).

$\alpha$ HL is an extremely robust biological nanopore: the pore is stable and functional at high temperature<sup>62</sup>, high pH<sup>63</sup> and high concentrations of urea<sup>64</sup> and SDS. Temporal (currently available instruments have a time resolution of  $\sim 10 \mu\text{s}$ ) and spatial limitations (multiple nucleotides are usually contributing the residual current signature in the channel) are the major challenges in the field of nanopore sequencing. In fact, a single strand DNA moves through the nanopore at velocities estimated to be  $\sim 1 \text{ nt}/\mu\text{s}$  and under a force of  $\sim 10 \text{ pN}$  at  $+120 \text{ mV}$ . Additionally a small number of ions (as few as  $\sim 100$ ) are available in the nanopore to correctly identify any given nucleotide of a ssDNA/ssRNA. Therefore it is impossible to sequence freely translocating ssDNA using  $\alpha$ HL<sup>65</sup>. In order to slow down DNA translocation, efforts have been made; such as the voltage manipulation<sup>66</sup>, the change of temperature<sup>67</sup> and the viscosity<sup>68</sup> of the solution, the mutation of charges within the channel<sup>69</sup>, the immobilization of biotinylated oligonucleotides<sup>18</sup> or sequence specific recognition and double helix unzipping<sup>27,70,71</sup>. The  $\alpha$ HL equipped with a chiral adapter ( $\beta$ -aminocyclodextrin) allowed the detection of all four DNA bases as nucleoside monophosphates<sup>21,72</sup>.

Recently, two independent groups<sup>23,24</sup>, provided the first reports that sequence information can be obtained upon ssDNA translocation through  $\alpha$ HL and MspA. Both groups used a polymerase to slow down the speed of the DNA through the nanopore in order to be able to read nucleotide-specific current levels (Fig 2.1). These studies represented an important step toward DNA nanopore sequencing.



**Figure 2.1 Biological nanopores for RNA/DNA sequencing.** **a**, Structure of the  $\alpha$ HL (top) and MspA nanopores (bottom) inserted into a lipid bilayer. **b**, DNA passing through a nanopore, with the speed control provided by a phi29 DNA polymerase (orange). The ssDNA substrate (red backbone) is inserted into the pore by an applied electric field and its motion inside or outside the pore (see the arrows) can be controlled by the applied electric field and the polymerase activity. (Modified from Schneider F.G. and Dekker C., 2012)<sup>73</sup>

#### 2.1.4 Planar (PLM) and Droplet (DIB) interface bilayers.

PLM and DIBs has been widely used to study properties of ion channels, pores and single molecule kinetics, in single or multiple channel recording. The principle behind these setups is extremely simple: when a potential is applied between two compartments isolated by a lipid bilayer, if a channel is formed all the ion current will pass through it. From the characteristics of the current trace biophysical and kinetic parameters can be extrapolated. These parameters can be related with the channel/pore itself (e.g, conductance and ion selectivity)<sup>74</sup> or with molecule interacting with the pore (e.g. structures, charges and binding rate constants)<sup>50,51,75-77</sup>.

Planar bilayers, often created by using the classical Montal-Mueller method<sup>78</sup>, have been used for decades as model systems to study the functional properties of ion channels and pores. In this method, the bilayer is usually formed by flowing a 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) lipid monolayers across both sides of an aperture<sup>78</sup> of  $\sim 100 \mu\text{m}$  in diameter in a Teflon film ( $25 \mu\text{m}$  thick) pretreated with pentane:hexadecane (10:1). As open chamber a custom Derlin vertical bilayer devices can

be used as a support for the Teflon film (but horizontal devices are also reported in literature<sup>79</sup>). The aperture in the Teflon is created “zapping” the film with a high-voltage spark generator. In PLM i) the possibility to easily work in single channel and ii) an open system that allow buffer exchanges are the most important advantages. Main drawbacks are the i) high volume (> 500  $\mu\text{L}$ ) in which the protein or the molecules that have to be analyzed are diluted and ii) the instability of the bilayer.

Another approach to study channels or pore forming proteins in simple and model systems is based on the Droplet Interface bilayer technique (DIB). The first DIB experiment was carried out in Moscow in 1966 (Tsofina et al., Nature. 1966), but began to be widely used for biophysical experiments from 2005, after a discussion between prof. David Needham (Duke University) and prof. Hagan Bayley (Oxford University) in a conference in Vancouver<sup>80</sup>. In the DIB approach, an aqueous droplet submerged under an oil-lipid mixture spontaneously acquires a lipid monolayer coat. If two of such droplets are brought into contact, the oil between the monolayers is displaced and a bilayer forms at the interface of the two droplets, giving rise to DIB. The oil phase is generally a linear or branched hydrocarbon such as hexadecane or squalene. Droplets can be created either by hand-pipetting aqueous solutions into the oil or by using microfluidic devices. DIBs are i) highly stable (droplet can stay in place for days) and ii) the volume can be very small (from 200 nL to 20 pL) but a droplet cannot be perfused (buffer exchange)<sup>80,81</sup>. Electrodes in droplet enable the application of a potential and the recording of a current ion flow through protein pores embedded in the bilayer. In both systems, the protein is usually added to the *cis* chamber/droplet which is connected to the ground. The potential is applied on the *trans* side, positive to force nucleic acid to translocate through the channel. A positive current is one in which positive charge moves through the bilayer from the *trans* to *cis* side. The advantage of this convention is that it is the same as that used when recording from a cell (patch clamp).

Single channel electrical recording in both system (PLM and DIB) requires basic electrical equipment: a patch-clamp amplifier (e.g. Axopatch 200B, Molecular Devices) with a 16-bit digitalizer (132x or 1440A, Molecular Devices) and a faraday cage to isolate the

system from external radiative noise. The proteins behavior in both system (PLM and DIB) is generally the same<sup>82</sup>.

In this chapter we used these biophysical techniques to find a biological tool able to detect, at the single molecule level, relevant post-transcriptional modifications (i.e. polyadenylation or polyuridylation) affecting the pathophysiology of many different RNAs.

## 2.2 Result and discussion

Recently, the 3' end uridylation of single stranded RNAs (ssRNAs) has been observed as a widespread template-independent post-transcriptional modification. Here, we demonstrated the selective, transient and modular binding of 3' poly-uridylated ssRNAs inside the  $\beta$ -barrel of the WT- $\alpha$ HL pore. Sensing the change in ionic current through the nanopore we described this undisclosed RNA-protein interaction.

In particular, after the observation of a surprising affinity of the WT- $\alpha$ HL with oligo U ssRNAs, suggesting an intriguing and previously not described RNA binding property of this pore forming toxin, we characterized the RNA/protein interaction as follow: 1) we defined the physical parameter of binding (binding rate constants  $k_{\text{on}}$  ( $\text{M}^{-1} \text{s}^{-1}$ ) and  $k_{\text{off}}$  ( $\text{s}^{-1}$ ) - dissociation constant ( $K_{\text{D}}$ )) and we described the RNA features important for the binding; 2) we characterized the affinity of the  $\alpha$ HL pore for ssRNAs with different length and number of uridines at the 3' terminus; 3) we identified the amino acid residues involved in the RNA binding; 4) we described the simultaneous binding of the slot of binding pockets in the homoheptameric pore (i.e. if all seven monomers interact with the RNA at once); 5) we prove that a given RNA cannot be sensed twice and 6) we demonstrated that the  $\alpha$ HL was able to discriminate the length of the uridylation. Finally, we set up a protocol for the purification and detection of 3'-end polyuridylation, with the aim to give a possible practical solution for the RNA processing of biological samples.

**2.2.1 Selective RNA sequence detection.** Since a cheap, fast and reliable technology to detect, at the single molecule level, relevant post-transcriptional modifications (i.e. polyadenylation or polyuridylation) is not available yet, we investigated the natural propensity of residues lining the lumen of the  $\alpha$ HL pore for the binding of specific nucleotides and/or sequence. We used single channel recording, PLM and DIBs to characterize the RNA/protein interaction, because these are powerful techniques for stochastic sensing.

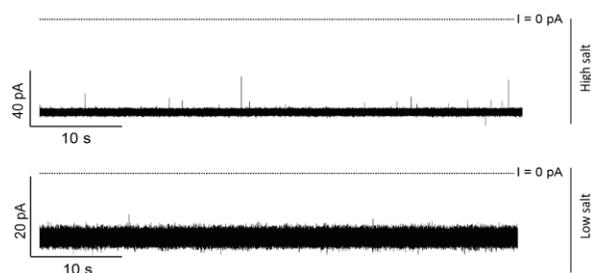
DNA translocation is affected by interactions with residues lining the lumen of the  $\alpha$ HL pore, and particular is the case of oligonucleotides shorter than the length ( $\sim 10$  nm) of the pore<sup>83,84</sup>. To examine ssRNA sequence recognition by the  $\alpha$ HL pore, we studied the electrical signatures of ssRNA 10-mers in single channel PLM analysis using WT- $\alpha$ HL heptemers extracted from SDS-PAGE gel (see Method). To obtain information concerning the selectivity against specific bases in the 3' tail, we used ssRNAs with the sequences 5'-C<sub>5</sub>X<sub>5</sub> (where X represents any nucleotide). After a single channel insertion, the oligo were added to the *cis* side of a two-compartment chamber (Figure 2.2a) filled with a *low ionic strength buffer* (150 mM KCl, 100 mM NaCl, 2 mM MgCl<sub>2</sub> and 10 mM HEPES, pH 6.5 in DMPC-treated water). We chose this buffer condition to minimize the charge screening (with still convenient current resolution in single channel). We observed long blocking events ( $> 5$  ms) only when uridines were located at the 3' end of the oligo (i.e. X=U). Nucleotides others than uridine did not produce blocking activity (Figure 2.2b). Then, we address the RNA over DNA specificity and used deoxy-uridines (dU) placed at the 3' end (X=dU). In this case long blocking events were again absent, underlining the dramatic effect of the 2'-OH group in the RNA chain on the translocation kinetics and demonstrating a preferential blocking of RNA respect to DNA in this condition. An A<sub>5</sub>-tail produced fast spikes ( $\bar{\tau}_D < 1$  ms) in agreement with previously reported data<sup>33</sup> (Figure 2.2a). For C<sub>5</sub>U<sub>5</sub>, the mean dwell time ( $\bar{\tau}_D$ ) was  $> 50$  ms and the majority of the events ( $> 90\%$ ) were longer than 5 ms (Figure 2.2b). No binding events were observed when the same ssRNA was added to the *trans* compartment applying a negative potential in different salt concentrations (Figure 2.3). The residual current ( $I_{RES\%}$ , see Methods) during the RNA blockades reflected an almost complete block of the channel ( $I_{RES\%} < 10\%$ ,  $n > 3$ ) (Figure 1c).

To characterize the RNA interaction and compare it with already published data on nucleic acid translocation, different voltages were applied in the presence of a fixed RNA concentration (2  $\mu$ M) of C<sub>5</sub>U<sub>5</sub>. The experimental dwell times and inter-event intervals, giving a quantitative information on the RNA/nanopore interaction, were fitted to a single component probability density function to obtain the associated rate constants  $k_{on}$

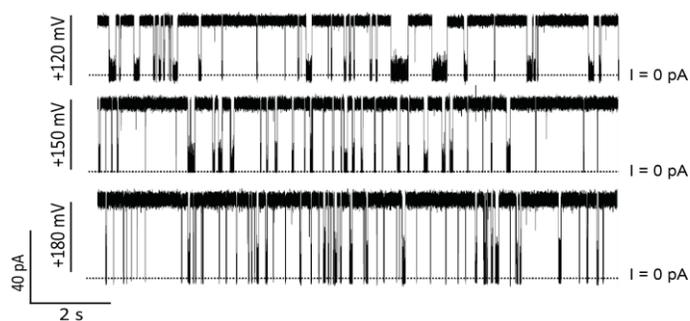
( $\text{M}^{-1}\text{s}^{-1}$ ) and  $k_{\text{off}}$  (s). The  $k_{\text{on}}$  value increases with the voltage (Figure 2.2c and 2.3) and the blockades ( $P_{\text{block}}\%$ , see Methods) linearly increased ( $R^2 = 0.99$ ) with increasing oligonucleotide concentration. This result confirmed that the pore closures were dose dependent. Moreover, when the voltage applied was increased from + 80 mV to + 120 mV we did not observe any change in the residential time (i.e. the dwell time, in other words the length of the blockades) but we noted an increase in the frequency of the events (assuming the rate constant  $k_{\text{on}}$  as a frequency because the RNA concentration does not change) (Figure 2.2c). This was an indication that the difference between the two voltages is due to the entropic barrier to drag the RNA in to the pore (as previously observed for different polymers<sup>57,85</sup>) and that the increase of the voltage from +80 mV to + 120 mV did not affect the dwell time of the events.



Previously published studies on the translocation through the  $\alpha$ HL pore<sup>52,57,65,66,87,88</sup> of RNA and DNA oligonucleotides longer than those we used, reported values of  $\bar{\tau}_D$  much smaller  $\bar{\tau}_D$  ( $\sim 1-22 \mu\text{s}/\text{nt}$ ). Moreover, recently it has been demonstrated that the translocation of long hetero-polymeric RNAs (90-6083 bases) is characterized by long  $\bar{\tau}_D$  values ( $\sim 1-10 \text{ ms}/\text{nt}$ ) and a low  $I_{\text{RES}\%}$  ( $< 1\%$ )<sup>55</sup>. In contrast, our results showed exceptionally long event only for short oligonucleotides ( $\bar{\tau}_D \sim 5-10 \text{ ms}/\text{nt}$  with 3' U tails. Therefore, we hypothesized that the observed transient blockades could be caused by specific binding of the oligo U sequence to the nanopore.



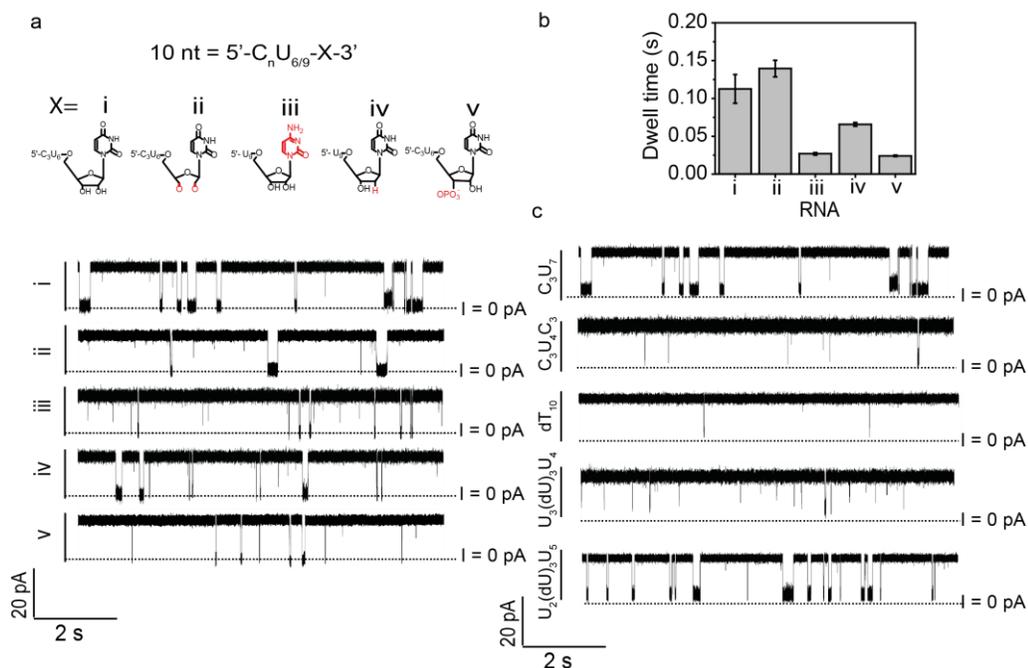
**Figure 2.3.  $C_3U_7$  does not affect the current of the WT- $\alpha$ HL pore from the *trans*.** Single channel traces acquired at negative potential in two different salt concentrations. *High salt buffer*: 1M KCl, 10mM Tris, 0.1mM EDTA, pH 7.5. Data acquired in *low salt buffer*: 150 mM KCl, 100 mM NaCl, 2 mM  $MgCl_2$  and 10 mM HEPES at pH 6.5 in DMPC water. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. Potential applied: -80mV.



**Figure 2.4 Ion current traces showing  $C_5U_5$  interacting with the WT- $\alpha$ HL pore at different voltages.** Data acquired in *low salt buffer*. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.

To test this hypothesis, we first change the features of the oligo U tail. Blockades generated by oligonucleotides containing uridines repeats with different length at their 3' termini were analyzed (Figure 2.2d and 2.2e). Only ssRNAs 10-40 bases long with a 3' tail longer than 4 uridines exhibited long binding events ( $> 5$  ms). We also found that the position of the oligo U segment is important for the current blockade. When the U signature is at the 5' terminus or in the middle of the RNA (Figure 2.2f and Figure 2.5), it is not recognized, demonstrating the specificity of U-tail orientation for the recognition. For RNAs with fixed total length, the  $\bar{\tau}_D$  increased with increasing length of the 3' U-tail, meaning a stronger binding. In addition, we observed that  $\bar{\tau}_D$  decreased with the length of the strand for both 5'-U<sub>n</sub> and 5'-C<sub>n</sub>U<sub>5</sub> ssRNAs ( $n \geq 5$ ) (Figure 2.2e). This finding contradicts previous RNA translocation studies<sup>13,56,66</sup>, in which  $\bar{\tau}_D$  was observed to increase linearly with the length of the strand<sup>13,56,66</sup>. This result underlined the possibility of a different interaction respect to what was previously described.

To gain our understanding of the binding within the nanopore, we used short (10-mer) oligonucleotides to study the effect of the nucleotide chemistry at the 3' end (Figure 2.5a). The substitution of the 3' terminal uridine (i.e. X = U) with a cytosine (i.e. X = C) in oligo U<sub>9</sub>X<sub>1</sub> reduced the residential time by  $\sim 3$  fold (from  $101 \pm 9$  ms to  $27 \pm 2$  ms). On the other hand, the single terminal 3' deoxy-uridine (i.e. X = dU) was not that disruptive of the signal (Figure 2.5). Finally, we observed that the blocking events disappeared when less than four U, upstream the last nucleotide, were present at the 3'- end (Figure 2.5c), meaning that that the physical properties of the last 5 uridines in 3' of the single stranded RNA are critical for the blockades.



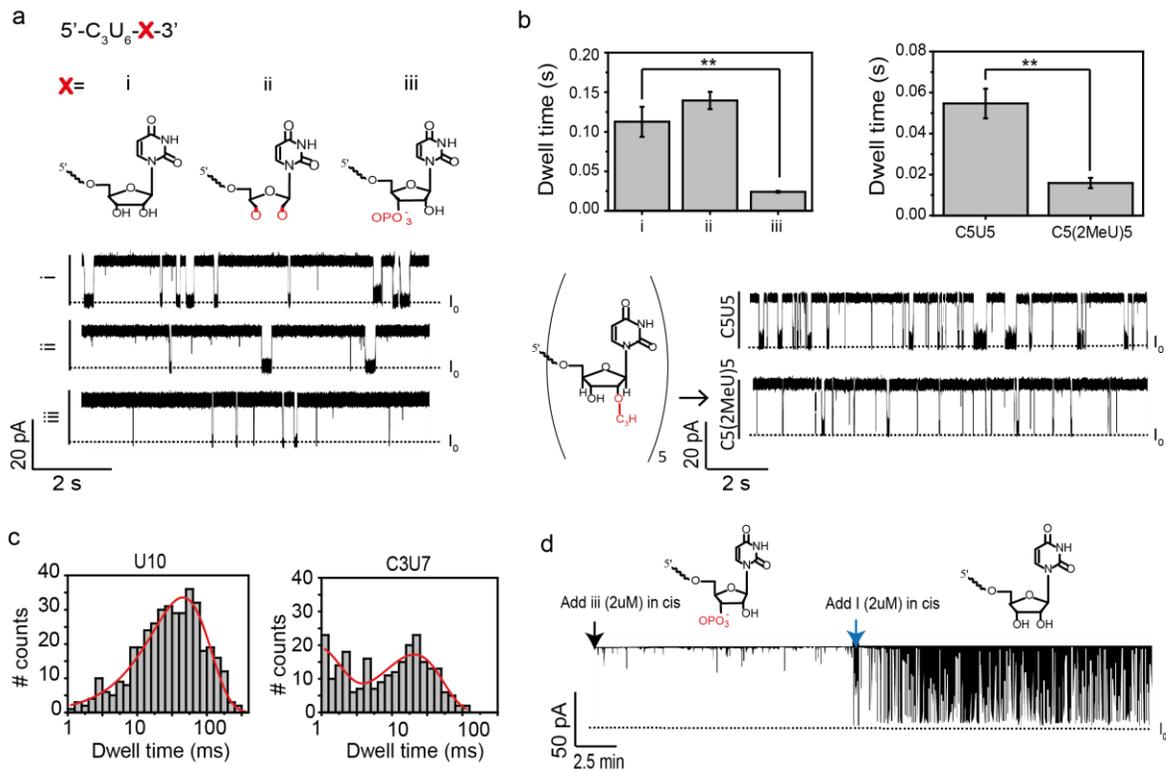
**Figure 2.5 Effect of the 3' ribose structure on the binding a. and b,** Single channel current traces showing the effect of the last nucleotide on the dwell time. **c,** Effects of the deoxynucleotides and the position of the oligo U on the binding of short 10-mer. Data were acquired in *low salt buffer*. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. Voltage applied + 80 mV.

Our results, compared to what previously obtained for the RNA/ $\alpha$ HL interaction are quite different. The first paper ever published about nucleic acid translocation through the  $\alpha$ HL pore was published in 1996 by Kasianowicz et al.<sup>56</sup> and was about the translocation of poly(U) RNA fragments. In that report, long polyuridinic acid were prepared with polynucleotide phosphorylase<sup>89</sup> and after fragmentation (~ 200 nt in length) added to the *cis* chamber in a 1 M KCl buffer under an applied potential of - 120 mV (negative to the *cis* side). Another work published in 1999 by Akeson M et al.<sup>54</sup>, described the translocation through the  $\alpha$ HL pore of poly(U) RNA fragments (~ 100 nt in length) obtained from polyuridinic acid hydrolysed in alkaline conditions. Therefore, we were surprised for the absence of long blocking events in those papers. We observed that three important differences were present with respect to our work: first, the alkaline fragmentation leave a 2'- 3' cyclic monophosphate derivative, further hydrolysed to give a mixture of 2'- and 3'-monophosphate derivatives; second, the high salt concentration of the buffer

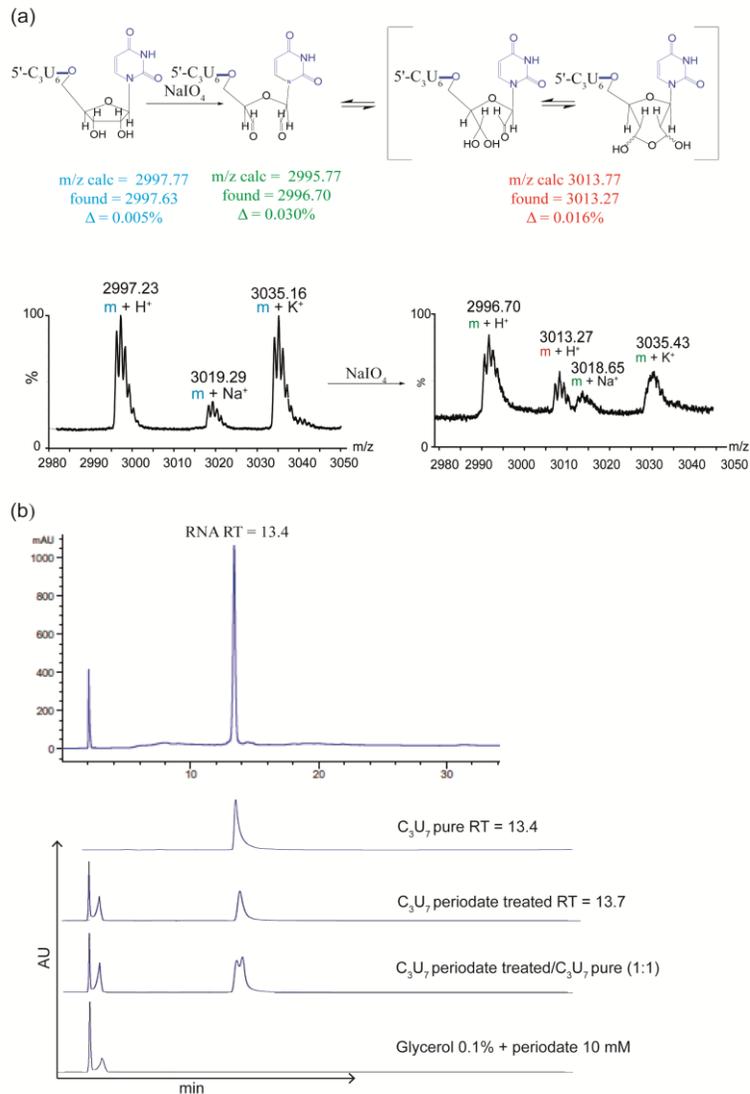
used; third, the fragments were at least ten times longer than the oligonucleotides used here. Surprisingly, so far no one else published other experiments with oligo U RNAs. Based on these observations, to compare our results with previously published works and to further characterize the blockades, we studied the effect of the structure of the 3' ribose phosphorylation. Adding a phosphate in 3', the translocation through the terminus where the additional negative charges are placed can be affected as well<sup>41,42</sup>. In fact, the phosphorylation at the 5' or 3' termini may enhance the probability of the RNA capture and the translocation direction of the oligonucleotides. Therefore, we phosphorylated the 3' end of the RNA C<sub>3</sub>U<sub>7</sub> and found that the  $\bar{\tau}_D$  decreased  $\sim 4$  fold (from  $114 \pm 2$  ms to  $30 \pm 2$  ms) (Figure 2.6a and b). This result suggested that the phosphorylation in 3' strongly affects the signal. We then define if the close ring of the ribose on the last nucleotide at the 3' was important for the binding. Interestingly, the selective and complete oxidation (Figure 2.7) of the 3'-terminal ribose into the di-aldehyde did not decrease the length of the blockades ( $\bar{\tau}_D = 140 \pm 10$  ms,  $n = 3$ ) (Figure 2.6a and b), but it affected the  $k_{on}$ . This suggested that the different conformation of the last ribose ring do not strongly affect the signal when the binding take place. Moreover, if no additional negative charges are placed the 3'- 2'- positions on the last nucleotide are only partially required for the binding, differently from all others 2'-OH in the pentameric U tail.

To conclude the characterization of the nucleic acid, we then asked if a steric hindrance in 2' could affect the blockades. The presence of a 2'-O-methylation on all 5 uridines in C<sub>5</sub>U<sub>5</sub> causes a decrease ( $\sim 30\%$ ) in the dwell time ( $16 \pm 2$  ms) (Figure 2.6b). The difference between uridine and 2'-O-methyl uridine can be interpreted in terms of a strong specificity in the docking events into a putative RNA binding pocket of the  $\alpha$ HL.

In conclusion, in order for the nanopore to detect urydilation at the 3' end of a ssRNA by means of long blockades, more than four uridines in 3' are required and the 2'-3'-OH cannot be modified unless if it is in the last 3' position.



**Figure 2.6** ssRNAs binding of the WT- $\alpha$ HL pore and effect of the ionic strength of the buffer. **a**, Current blockades produced by ssRNA decamers with different 3' sugar rings: (i) ribose with a free 3'-OH, (ii) the 2'-3'-dialdehyde (iii) ribose with a 3'-phosphate ( $n=3$ ). Recording in low ionic strength buffer at +80mV. **b**, Histograms of the mean dwell time for the three different oligos in 'a' (right) and for a ssRNA with five 2'-O-methyluridines [ $C_5(MeU)_5$ ] at the 3' end (left) (\*\*)  $P < 0.1$  Student's t-test. **Bottom panel:** Ionic current traces for  $C_5U_5$  and  $C_5(MeU)_5$  recorded at +80 mV in low ionic strength buffer. **c**, Dwell time distribution in high ionic strength buffer (1M KCl, 10 mM Tris, 0.1 mM EDTA, pH 7.5) for  $U_{10}$  (left) and  $C_3U_7$  (right). The histograms contain more than 100 events. Traces were recorded at +80 mV. **d**, A recording is shown in the presence of  $2\mu M$  of 3'-phosphorilated  $C_3U_7$  (black arrow). At the point indicated (blue arrow)  $2\mu M$  unphosphorilated  $C_3U_7$  was added. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.  $I_0$  = zero current level.



**Figure 2.7 Periodate oxidation of C<sub>3</sub>U<sub>7</sub> assayed by MALDI mass spectrometry and RP-HPLC.** (a) Top panel: The product of periodate oxidation of a ssRNA is a mixture of 3'-dialdehyde and its hydrated form (presumably cyclized to the bis-hemiacetal form) as illustrated. Bottom: MALDI mass spectrometry was performed before (left) and after (right) oxidation. The detection of a di-aldehyde after oligonucleotides peroxidation is in agreement with previous work on periodate oxidation of ssRNA<sup>9</sup>. Peaks are color-coded as reaction scheme above. (b) Top: preparative RP-HPLC for C<sub>3</sub>U<sub>7</sub> (25  $\mu$ g) reacted with periodate. Bottom, qualitative RP-HPLC chromatograms are listed (from top to bottom) for: C<sub>3</sub>U<sub>7</sub>, not modified; C<sub>3</sub>U<sub>7</sub>, modified; a mixture of C<sub>3</sub>U<sub>7</sub> reacted and not reacted (1:1); periodate only quenched with glycerol (3  $\mu$ g RNA). Purification was carried out with an Eclipse Plus C<sub>18</sub>, 3.5  $\mu$ m, 4.6 x 100 mm. RT: retention time

**2.2.2 Effect of the ionic strength on the binding.** Most of the previous work about single channel recordings to study polymers translocation has been performed in buffers at high ionic strength ( $\geq 1\text{M}$  KCl) to increase the sensitivity window of the

signal<sup>17,51,54,55</sup>. Moreover, the ionic strength can affect different parameters of the single stranded RNA (e.g. persistence length)<sup>90</sup> or the protein-RNA interaction (e.g. charge screening) and, in turn, the sensitivity of the nanopore. In the case of ssRNA, we observed that in *high ionic strength buffer* (1 M KCl, 10 mM Tris, 0.1 mM EDTA, pH 7.5 in DMPC-treated water) the  $\bar{\tau}_D$  was reduced by more than 60% depending on the oligo U length (Figure 2.6c and Table 2.1), meaning that the charge screening had an important role in the interaction under study.

Finally, when 3'-phosphorylated C<sub>3</sub>U<sub>7</sub> was added to the *cis* compartment in *high ionic strength buffer*, no blocking events (Figure 2.6d) were observed. Upon the addition of RNA with the same sequence but with a 3'-OH terminal ribose, the blockades reappeared (Figure 2.6d). These results demonstrated the combined effect of the phosphorylation state of the 3'-end and the high ionic strength in defining the blockage efficiency of the pore. This combined effect caused a complete inability to detect the RNA in solution as demonstrated by pioneering experiments of Kasianowicz and Akeson.

In summary, many factors are contributing to the RNA binding in the  $\alpha$ HL pore: a 3' U tail longer than four uridines with the last position in 3' less stringent than others, the OH in 2' of the ribose and the low buffer ionic strength.

RNA	High salt $\tau_D$ (ms)	Low salt $\tau_D$ (ms)
U <sub>20</sub>	46 ± 6	120 ± 13
U <sub>10</sub>	39 ± 5	101 ± 10
C <sub>3</sub> U <sub>7</sub>	19 ± 3	130 ± 9

**Table 2.1 Effect of the salt concentration on the oligo U binding.** *High salt buffer:* 1 M KCl, 10 mM Tris, 0.1 mM EDTA, pH 7.5. *Low salt buffer:* 150 mM KCl, 100 mM NaCl, 2 mM MgCl<sub>2</sub>, 10 mM HEPES in DMPC water, pH 6.5. The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter).  $n \geq 3 \pm$  s.d

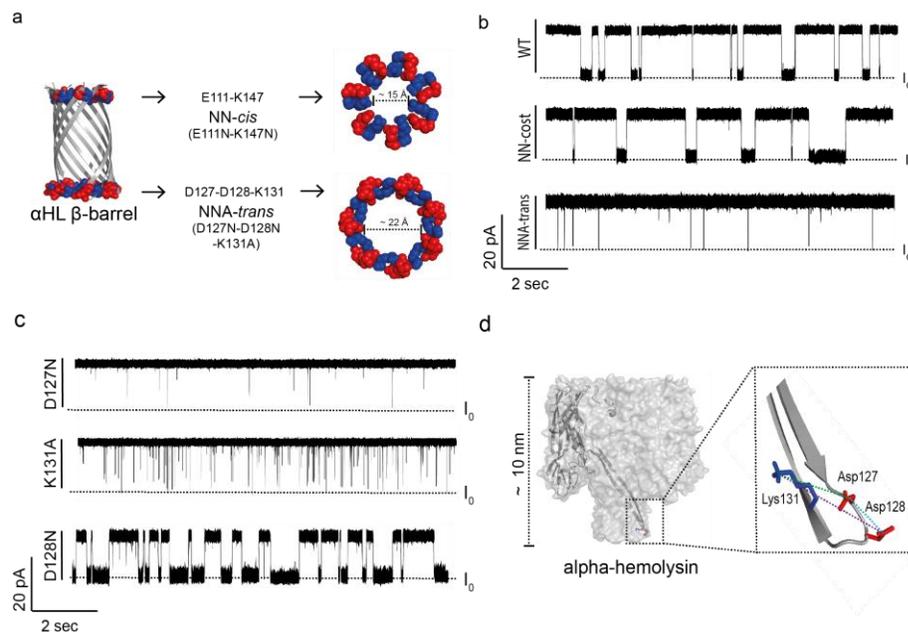
**2.2.3 The RNA binding site.** To identify the binding site of the oligo U, a mutagenesis analysis was performed on WT- $\alpha$ HL (Figure 2.8). We reasoned that if the blockades were due to eventual U-U base pairs<sup>92</sup>, the RNA should arrest in the narrowest portion of the

pore. Therefore, two mutants were produced. The two mutant were called NN-*cis* and NNA-*trans* because they are at the two opposite ends of the  $\beta$ -barrel, facing the *cis* chamber (NN-*cis*) and the *trans* chamber (NNA-*trans*), respectively. In the first mutant, the NN-*cis* mutation (E111N / K147N) was designed to neutralize the charges at the constriction of the nanopore and to widen the internal entrance of the  $\beta$ -barrel (*cis*) that is known to result in a decreased voltage threshold for nucleic acid translocation<sup>69</sup>. The NNA-*trans* mutation (D127N / D128N / K131A) was inspired by the hydrophilic, highly charged and flexible  $\alpha$ HL *trans* terminal loop (Gly<sup>122</sup> to Ile<sup>136</sup>, located between the two antiparallel  $\beta$ -strands of the  $\alpha$ HL  $\beta$ -barrel). In this region of the  $\alpha$ HL pore, the two aspartic acids (Asp-127 and Asp-128) are followed by a positively charged residue (Lys<sub>131</sub>). Moreover, when the amino acid sequence forming the  $\beta$ -barrel (from Glu-111 to Lys-147) was analyzed with BindN<sup>93</sup>, a software for RNA-binding residues prediction (<http://bioinfo.ggc.org/bindn/>), only the short region between Tyr-125 to Lys-131 included, was associated with a high binding score (see Methods). Therefore, considering that not all the residues in the barrel are faced toward the central lumen, we created the NNA-*trans*  $\alpha$ HL mutant (D127N / D128N / K131A) to neutralise these charged residues (Figure 2.8a).

The NN-*cis* mutant had no effect on the blockades events observed previously. When NNA-*trans* was used, no long blockades were observed, but very short blockades ( $\bar{\tau}_D < 1$  ms) were present (Figure 2.8b). This result means: i) that removing the charges D127N / D128N / K131A the block is prevented, making these residues interesting candidates for the RNA binding; ii) that the signal block is not due to unlikely secondary structure of the RNA occluding the pore but are related with a specific interaction in the *trans* side of the  $\beta$ -barrel.

We then studied which of the three point-mutation in NNA-*trans* is the most important for the binding of oligo U. To this aim we generated and tested different single and double point mutants (neutralizing the charges in position 127, 128 and 131 in a sequential manner, Figure 2.8c. Not all mutants tested are reported). As a probing oligonucleotide we used C<sub>3</sub>U<sub>7</sub> for its the ability to tightly interacted with the WT- $\alpha$ HL pore.

The aspartate in position 128 does not participate in the U recognition, as demonstrated by the unchanged behaviour as compared to the WT. D128N has a higher affinity than the WT- $\alpha$ HL pore ( $K_D^{(D128N)} = 2.0 \pm 0.9 \mu\text{M}$  vs  $K_D^{(wt)} = 11.6 \pm 0.6 \mu\text{M}$  at +80mV,  $n \geq 3$ ;  $K_D^{(D128N)} = 1.3 \pm 0.1 \mu\text{M}$  vs  $K_D^{(wt)} = 4.7 \pm 0.8 \mu\text{M}$  at 120 mV,  $n \geq 3$ , C<sub>3</sub>U<sub>7</sub> ssRNA). Additionally, we found that the mutated  $\alpha$ HL pore (D128N) was more sensitive to the oligo U tail. We speculated that the Asp<sup>128</sup> interfered with Asp<sup>127</sup> or with the nucleotide chain, altering the pKa and reducing the interaction. In conclusion, by the combination of site-directed mutagenesis and oligo probing, we identified the two residues that are important for the binding. We can speculate that salt bridges could be relevant for the binding (Figure 2.8d) because the distance between Asp-128 and Lys-131 allows salt bridges formation with the uridine. Moreover, we observed that other oligomeric RNA binding proteins bind 3' oligo U tails only when more than four uridines are present, holding the nucleobase in a network of hydrogen bonds formed by positive and negative charged amino acid residues placed on secondary loop structures: the ubiquitous family of Sm-like RNA binding proteins<sup>94-101</sup> is a clear example where all these characteristics are present.



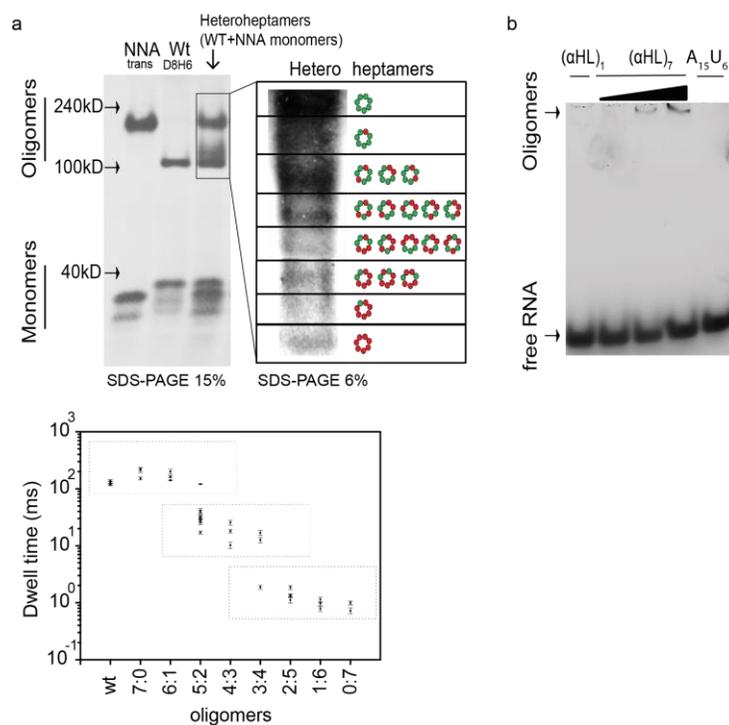
**Figure 2.8 The RNA binding pocket in the WT- $\alpha$ HL nanopore.** a, Mutated residues in the  $\beta$ -barrel of the WT- $\alpha$ HL pore. The constriction formed by the ring of residues Glu<sup>111</sup> and Lys<sup>147</sup> were mutated to Asn<sup>111</sup>

and Asn<sup>147</sup> in the NN-*cis* mutant. The residues Asp<sup>127</sup>, Asp<sup>128</sup> and Lys<sup>131</sup> were mutated to Asn<sup>127</sup>, Asn<sup>128</sup> and Ala<sup>131</sup> in the NNA-*trans* mutant. The diameters of the two entrance of the WT- $\alpha$ HL  $\beta$ -barrel are reported on the left. **b**, Current traces for 2 $\mu$ M C<sub>3</sub>U<sub>7</sub> detected by the WT- $\alpha$ HL (top), NN-*cis* (middle) and NN-*trans* (bottom), monitored at +80mV in *low ionic strength buffer*. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. **c**, Current traces of single point  $\alpha$ HL mutants reporting the blockades in the presence of 2 $\mu$ M of C<sub>3</sub>U<sub>7</sub> in *cis* +80mV in *low ionic strength buffer* acquired as in 'b'. **d**, Heptameric  $\alpha$ HL with single subunits shown as a ribbon structure. **Zoom in:** *trans* terminal loop with the three residues mutated in NNA-*trans*. The distance (C-N) between the Asp<sup>127</sup> and Lys<sup>131</sup> is  $\sim 7.6$  Å (green broken line), between Asp<sup>128</sup> and Lys<sup>131</sup> is  $\sim 12.7$  Å (violet broken line), between two aspartates is  $\sim 5.2$  Å (blue broken line) and between Lys<sup>131</sup> and the Asp<sup>127</sup> on the next monomer is  $\sim 3.2$  Å (not reported). I<sub>0</sub> = zero current level.

**2.2.4 Binding modularity.** As mentioned at the beginning, to understand if all seven monomers in the heptameric pore were bound simultaneously to the oligo U, crucial is the characterization of a hypothetical multiple (more than a monomer a time) or single (only a single monomer bind the RNA) interaction. For this reason we tried to elucidate if the RNA was trapped by a modular recognition in the homoheptameric  $\beta$ -barrel. Monomers of  $\alpha$ HL bearing at the C-terminal a D<sub>8</sub>H<sub>6</sub> tail can be used to separate hetero-heptameric  $\alpha$ HL pores by gel-shift electrophoresis. This approach has been previously used with success to determine the heptameric stoichiometry of the  $\alpha$ -hemolysin pore<sup>102,103</sup>. Purified  $\alpha$ HL pores obtained from combinations of  $\alpha$ HL(D<sub>8</sub>H<sub>6</sub>) and  $\alpha$ HL - NNA-*trans* were tested in single channel recordings with C<sub>3</sub>U<sub>7</sub>. Analysis of  $\bar{\tau}_D$  for each  $\alpha$ HL hetero-heptamer showed three populations of blockades, corresponding to strong binding ( $\bar{\tau}_D \geq 100$ ms), moderate binding ( $5 \text{ ms} \leq \bar{\tau}_D \leq 100 \text{ ms}$ ) and no binding ( $<1 \text{ ms}$ ). A strong binding was observed only with hetero-heptamers with one single mutated monomer, moderate binding with more than two NNA-*trans* monomers and no binding was observed with more than four NNA-*tans* monomers in the pore (Figure 2.9a). This result showed that most probably a minimal number of adjacent monomers are required for at least a moderate binding to occur, although we cannot exactly define the order of the mutants, because each hetero-heptamer can have different permutations.

Finally, we considered the possibility that the RNA-protein interaction is cooperative. We calculated the P<sub>block</sub> as a function of ssRNA concentration (data not shown). Our results followed a Hill plot with  $n = 0.98 \pm 0.04$ , showing that the binding is not coopera-

tive. We concluded that the RNA-binding nanopore was able to bind sequence specific RNA strands in a modular manner. As confirmed in bulk by an electro-mobility shift assay (Figure 2.9b), all the above described findings brought us to the conclusion that the  $\alpha$ HL pore works as an RNA binding nanopore with a strong affinity only for a 3' end oligouridylated ssRNAs.

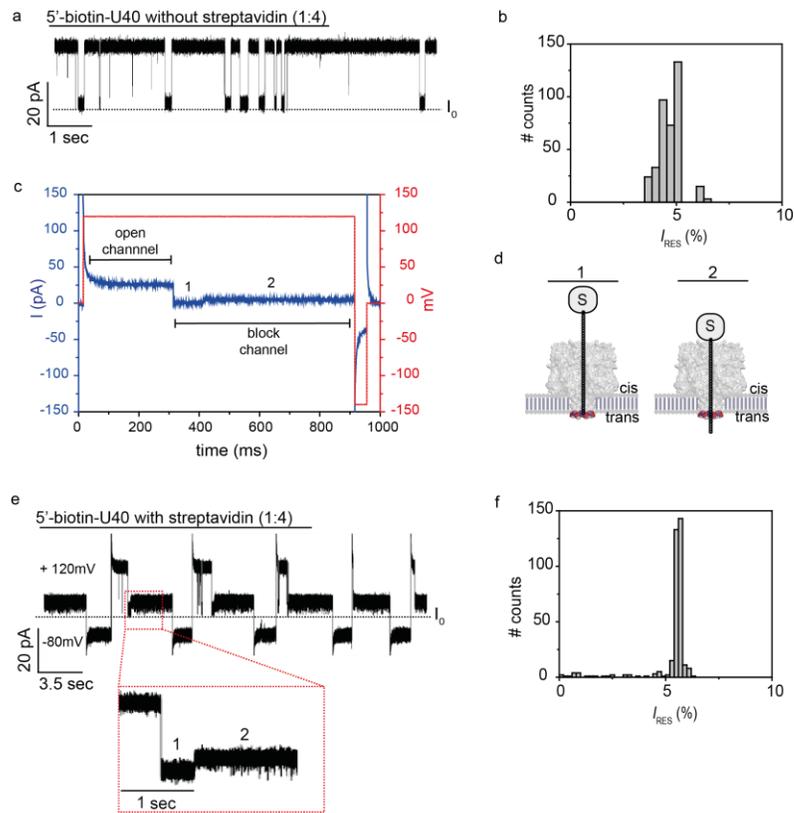


**Figure 2.9 Binding modularity. a, Upper panel:** Modularity of binding during recognition: heteroheptameric  $\alpha$ HL pores generated by gel-shift electrophoresis. WT- $\alpha$ HL(D8H6) and NNA- $\alpha$ HL monomers were prepared by coupled *in vitro* transcription and translation. The monomers were mixed 3:2 (WT:NNA) and assembled on rabbit red blood cell membranes (RRBCMs). SDS-PAGE analysis of the radiolabeled proteins revealed a ladder of oligomeric bands with different subunit ratios. The homoheptameric WT- $\alpha$ HL(D8H6)<sub>7</sub> has the fastest electrophoretic mobility, whereas the homoheptameric NNA- $\alpha$ HL the slowest. In a 6% SDS-PAGE all different heteroheptamers can be resolved (zoom in panel). Heterheptameric pores were tested in single channel recording with 2  $\mu$ M C<sub>3</sub>U<sub>7</sub> (3'-OH) ssRNA. **Bottom panel:** Mean dwell time for each of the different bands showed three populations of signals: strong binding ( $\tau_D \geq 100$  ms), moderate binding ( $1 \text{ ms} \leq \tau_D < 100$  ms) and no binding ( $< 1$ ms). The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter); +120 mV. Data acquired in *low salt buffer* (error bars:  $\pm$  s.d,  $n \geq 3$ ) **b,** Electromobility shift assay with monomeric (WT- $\alpha$ HL)<sub>1</sub> protein and increasing concentration of recombinant pre-oligomerized (WT- $\alpha$ HL)<sub>7</sub> proteins. 3  $\mu$ g starting quantity of protein for monomers and oligomers. Oligomers were diluted 1:2 progressively. After electrophoresis, the gel was stained with SyberGold. RNA/protein complex and free RNA are indicated (black arrows).

**2.2.5 Proof of translocation.** Generally speaking, for the  $\alpha$ HL pore exist a threshold voltage below which no oligonucleotide translocation events are appreciable. This threshold depends on charged amino acid residues in the pore<sup>69,85</sup>. We observed events at voltages higher than +80 mV and an increase in  $P_{\text{block}\%}$  from ~5 % at + 80mV to ~20 % at + 120 mV (at 2  $\mu$ M of  $C_5U_5$ , Figure 2.2c), in agreement with the described entropic barrier to thread the DNA strand into the pore<sup>67</sup>, meaning that the RNA was dragged into the pore by the potential applied.

Although DNA translocation through the  $\alpha$ HL has been previously demonstrated<sup>16</sup>, we could not disregard the possibility that the short ssRNA we examined, visited the  $\beta$ -barrel to produce a blockade, and then exit on the side of addition.

To demonstrate that the long current blockades arose from the translocation of the RNA, a streptavidin-(5')biotin-RNA complex was used as previously reported<sup>16,18,104</sup>. A 5' biotinylated ssRNA of 40 nt (the longest RNA used in Figure 2.2c) was incubated with streptavidin to form a tight non-covalent complex with biotin (Figure 2.10). After this treatment, ~ 90 % of the events were “permanent” (> 60 s) blockades ( $I_{\text{RES}}\% = 5.60 \pm 0.01$ ,  $n=3$ ) and required an inversion of the voltage polarity to unclog the pore. Around 20 % of the “permanent” blockades exhibited a two-step signal, where the first step (step 1) arose probably from the binding of the 3' end of the RNA in the  $\alpha$ HL pore. The second step (step 2) was a permanent current block due to the streptavidin-coupled RNA (Figure 2.10). The increase in the residual current of step 2 compared to that at step 1 is most probably due to the stretching of the RNA under the influence of the electric field, as previously observed<sup>716,105</sup>. The step 1 was not present when we performed the same experiment with a  $\alpha$ HL mutant pore (D127N-K131A) incapable of RNA binding (data not shown). These findings demonstrated that transient blockades were caused by the translocation of the RNA through the pore. The important implication of this result is that the same RNA cannot be sensed twice, making the sensor suitable for biological application.



**Figure 2.10. RNA current blockades arise from the translocation of the RNA through the WT- $\alpha$ HL pore.** **a**, Current trace with 5' biotinylated U<sub>40</sub> translocating in the WT- $\alpha$ HL pore. **b**, histogram of the  $I_{RES\%}$  of the events in 'a'. **c**, The biotinylated RNA-streptavidin complex (1:2 molar ratio, streptavidin:RNA, NEB) was introduced to the *cis* compartment and captured by the WT- $\alpha$ HL pore under an applied voltage of +120 mV (in *low salt buffer*) by using an automated voltage protocol with 900 ms capture time. Example of a single sweep of the protocol applied is reported. Each sweep was characterized by the follow voltage sequence (red line) (i) 10 ms at 0 mV, (ii) 900 ms at +120 mV, (iii) 45 ms at -120 mV, (iv) 45 ms at 0 mV. The protocol was repeated 1000 times for each experiment. Blue signal: ionic current measured. ~ 20 % of the events presented a "low amplitude step" (called step 1, presumably related to the RNA binding in the  $\beta$ -barrel) before the permanent blockades (called step 2) due to biotinylated RNA-streptavidin interaction with the pore. **d**, Schematic representation of the two steps: WT- $\alpha$ HL nanopores (PDB:7AHL) embedded in a lipid bilayer with RNA:streptavidin complex. **e**, A longer trace with manual potential inversion is reported. Zoom out. binding events (step 1) observed before permanent block (step 2) of the channel. **f**, histogram of the  $I_{RES\%}$  of the blockades after addition of the streptavidin:RNA complex. 20 kHz acquisition, 2 kHz low-pass filter. Data acquired in *low salt buffer*. 2  $\mu$ M RNA concentration.

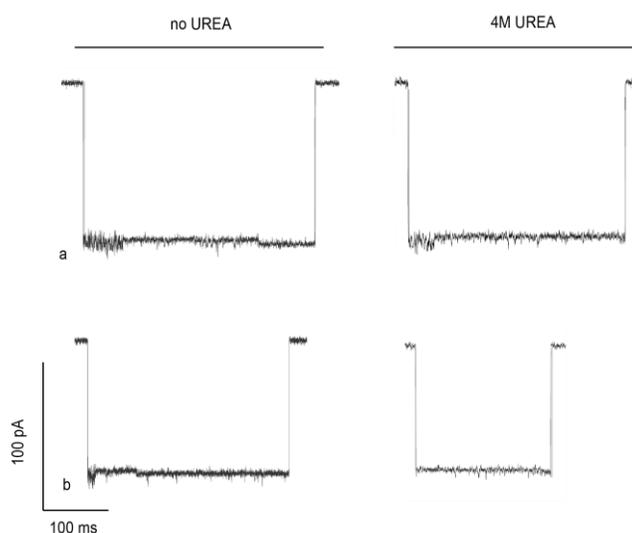
We then moved on to characterize the binding of folded ssRNA with oligo U tails, since a folded RNA can stuck in the pore mimicking the streptavidin block.

RNA can fold adopting several different secondary structures depending on the environment. Structured RNAs achieve their low energy state by traversing complex energet-

ic landscapes<sup>106</sup>. Structured polynucleotides are too large to pass through the WT- $\alpha$ HL, for this reason it is useful to unfold their structure before translocation. In the past, the effect of denaturants on ssDNA and ssRNA secondary structure was studied using nanopores in alkaline solutions<sup>63,107</sup> or in the presence of urea<sup>64</sup>.

In order to understand how secondary structures affect the recognition of the U-tail by the  $\alpha$ HL we decided to address the translocation of the pre-microRNA let-7a, known to present a characteristic secondary structure. Let-7a is a member of the Let-7 family, a class of microRNA with important biological functions such as suppression of cell proliferation and promotion of cell differentiation by targeting multiple genes including HMGA2, RAS, and Lin28<sup>108</sup>. The regulated RNA-binding protein Lin28 is also an inhibitor of let-7. Lin28 interacts with terminal uridyl transferase 4 (TUT4, also known as ZCCHC11, PAPD3, and Hs3) to induce oligo-uridylation (10–30 nt) of pre-let-7 by TUT4<sup>44,109</sup>.

We measured the translocation in *high ionic strength* of the pre-let-7a with a 3'-U tail of 7 uridines in presence and absence of denaturants. We observed that in absence of denaturant the electric signal was characterized by three distinct levels of residual current. Interesting, the same result was observed with a RNA aptamer with a 31 uridines in 3' (the Malachite green aptamer; see chapter 3 for a detailed description) characterized by a peculiar secondary structure. In the presence of 4M urea the same pattern was observed, but with lower resolution between levels in terms of residual current (Figure 2.11).



**Figure 2.11** Examples of blocking events observed with folded RNAs translocating through the  $\alpha$ HL pore. **a**, pre-let7a-U7. **b**, Malachite green aptamer-U<sub>31</sub> Data collected in *high ionic strength buffer* at + 120 mV.

Our finding showed that RNA with secondary structures and a 3' oligo U tail behaved differently from unstructured ssRNA. Intriguingly, the ionic current blockades pattern (with and without urea) was very similar to the one recently observed for protein unfolding in a nanopore<sup>51</sup>. We did not investigate further the signal from these complex RNAs, but if the signal reflects the RNA unfolding before translocation, this result is an additional indirect prove of translocation.

We concluded that every type of secondary structure increased the complexity in the detection of the RNA signature and therefore the use of short and unstructured RNA for the poly(U) detection is essential and the unique option with this system to be specific. Further experiments are needed to understand if this characteristic signal is depending or not on the 3' U-tail.

**2.2.6 The  $\alpha$ HL pore as a detector of 3'-end uridylation lenght and post-transcriptional modifications.** Finally, we tried to understand if the WT- $\alpha$ HL could be used to discriminate the number of 3' uridines in an RNA strand. All the 3' uridylated

oligonucleotides tested in *low ionic strength buffer* showed differences in  $\bar{\tau}_D$  but not in  $I_{RES\%}$  (<10%). This result was different when using *high ionic strength buffer*. For example we found that a mixture of four 10-mer RNAs with 3'-end bearing 5, 6, 7 or 10 uridines could be distinguished by their difference in residual current values (Figure 2.12a, Figure 2.13 and Table 2.2). The difference in the residual current between the two most widely dispersed well separated current peaks ( $\Delta I_{RES\%}^{OVERALL} = I_{RES\%}^{C4U6} - I_{RES\%}^{C3U20}$ ) was  $+ 25 \pm 4\%$ .

RNA	$\tau_D$ (ms) <sup>HS</sup>	$I_{RES\%}$ <sup>HS</sup>
U <sub>20</sub>	46 ± 6	7 ± 1
U <sub>10</sub>	39 ± 5	16 ± 1
C <sub>3</sub> U <sub>7</sub>	19 ± 3	24 ± 1
C <sub>4</sub> U <sub>6</sub>	10 ± 1	29 ± 1
C <sub>5</sub> U <sub>5</sub>	4.7 ± 0.3	32 ± 2

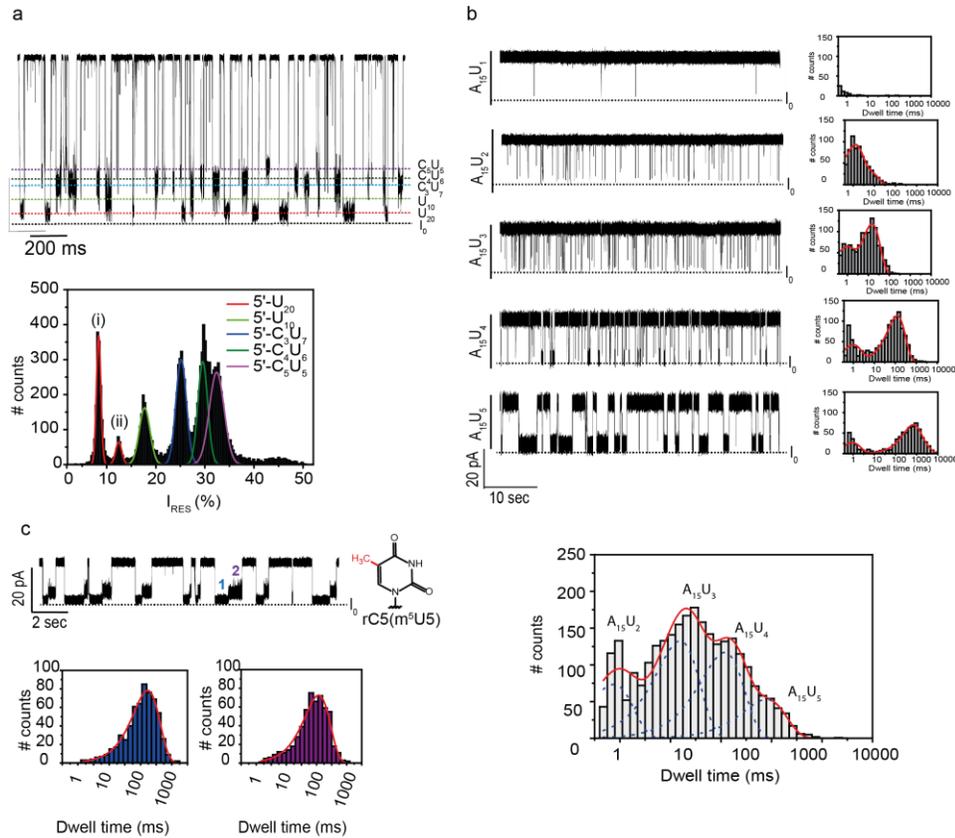
**Table 2.2 Mean dwell time and residual current (%) of different RNA oligos detect by WT- $\alpha$ HL.**  $I_o$  (open current level) = 24 ± 2 pA in LS;  $I_o$  = 130 ± 5 pA in HS. Potential applied = +120 mV. HS: high ionic strength buffer. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.  $n \geq 3$  s.d. HS: high ionic strength buffer; LS: low ionic strength buffer.

Therefore, the  $I_{RES\%}$  displayed a very good resolution, although different RNAs also show different dwell time distribution (Table 2.2, Figure 2.12 and Figure 2.13). With oligonucleotides longer than ten bases the residual current drops to values < 10% (e.g.  $I_{RES}^{U_{20}} = +7 \pm 1\%$ ) and different oligonucleotide can no longer be distinguished. Interestingly, with U<sub>20</sub>, a second small population of events at higher residual current was observed (Figure 2.13a, bottom). In conclusion, these results demonstrate that the WT- $\alpha$ HL pore can be used to detect the length of the U tail on short (< 10 bases) and unstructured ssRNAs when present in mixed samples.

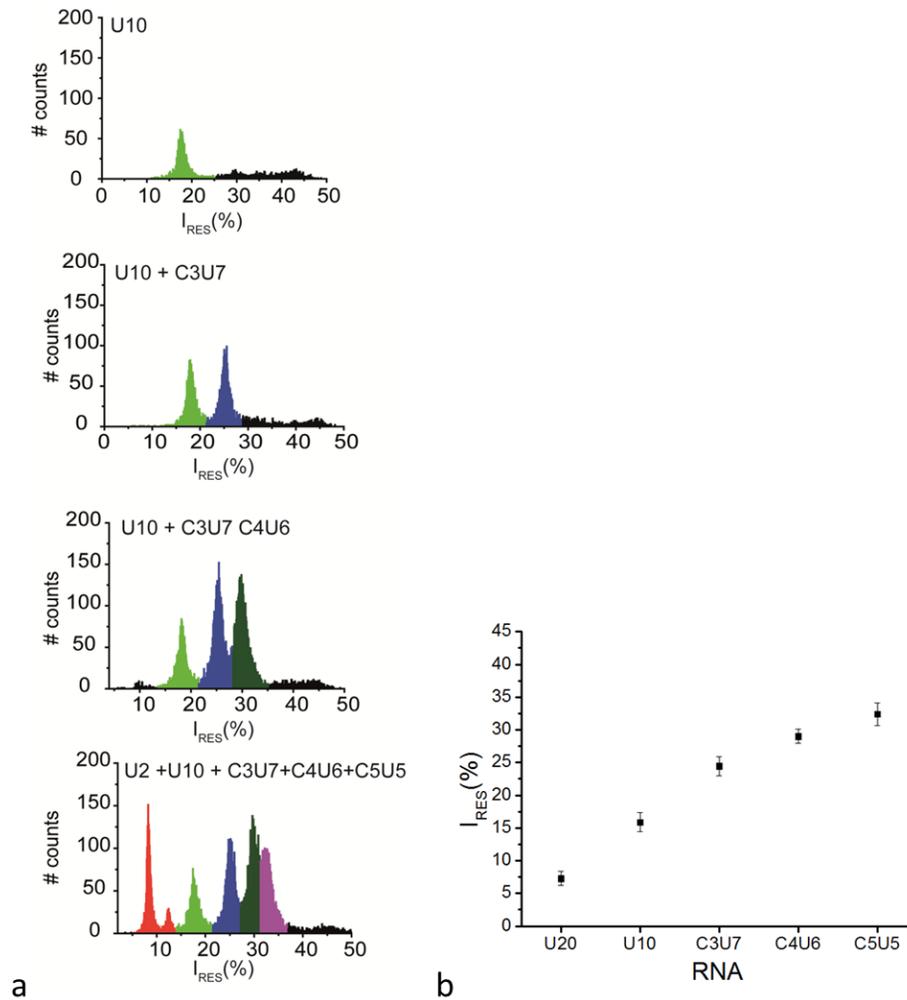
Recently, it has been described that mRNAs can be modified by poly (U) polymerases (PUPs, e.g. Cid1). PUPs add uridines (usually < 3 bases) to the 3' end of mature polyadenylated mRNAs in a manner that is independent of the poly(A) tail length<sup>7,10,38,48,110</sup>.

Given the possible biological relevance of sensing few uridines and prompted by our results, we tested whether the  $\alpha$ HL pore can recognize U tails shorter than 5 bases in a fixed background of 15 adenosines. This possibility would make the pore able to detect this type of post-transcriptional modification. To demonstrate that this is indeed possible we reasoned that under our conditions, secondary structures of ssRNA A<sub>15</sub>U<sub>n</sub> ( $n \leq 5$ ) are energetically disfavored (see Methods Table 2. 4). The D128N- $\alpha$ HL pore was able to

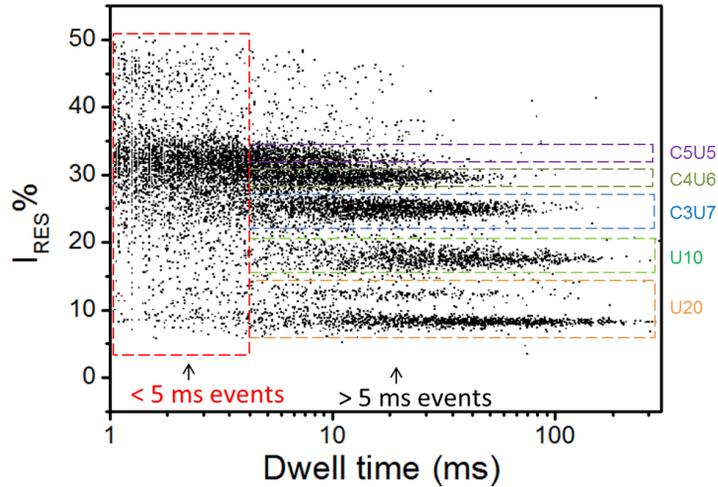
detect  $A_{15}U_X$  ( $X \leq 5$ ) ssRNA based on the  $\bar{\tau}_D$ . In *low ionic strength buffer* conditions, the detection of these oligonucleotides was clear (Figure 2.12).



**Figure 2.12 The  $\alpha$ HL pore as a sensor for different RNA-uridylylations.** **a**, Ion current trace after addition of the of  $2\mu M$  each of different ssRNAs ( $C_5U_5$ ,  $C_4U_6$ ,  $C_3U_7$ ,  $U_{10}$ ,  $U_{20}$ ). Dashed lines indicate the blockade level associated with each ssRNA in *high ionic strength buffer* at +120 mV. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. Each ssRNA produces a specific signature in residual current. **Bottom:**  $I_{RES}\%$  are shown for each ssRNA used above and a Gaussian fits were performed for each peak. The  $I_{RES}\%$  values are given in Table 2 ( $n \geq 6$ ).  $U_{20}$  shows (red Gaussians) an additional characteristic small population of events at higher residual current (indicated as 'ii') close to the main population (indicates as 'i'). **b, Top panel:** Current traces for the  $D128N$  pore with  $2\mu M$  ssRNA  $A_{15}U_X$  in *cis* ( $1 \leq X \leq 5$ ) are listed (*low ionic strength buffer*, +120mV. Signal acquired as in 'a'). At the right of each trace is a dwell time distribution for a typical 15 min trace. The events were fitted to a single component probability density function. **Bottom:** Dwell times histogram for a mixture of different ssRNA ( $A_{15}U_1/A_{15}U_2/A_{15}U_3/A_{15}U_4/A_{15}U_5$ ,  $1\mu M$  each, *cis*). Data were recorded in *low ionic strength buffer* at +120 mV. Four different populations of events can be fitted with a four component probability density function (red line). A single component probability density function has been fitted on each population (broken blue line). **c**, Characteristic signature of ionic current from a U tail composed of five noncanonical nucleobases ( $m^5U$ ). A typical current trace for  $C_5(m^5U)_5$  is reported. Two blockades steps for each level (1 and 2) were monitored at +80 mV in *low ionic strength buffer*. Step 2 always follows step 1. Bottom panels: Dwell time distribution for each single step  $\sim 500$  events for each plot were fitted to a single component probability density function. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.  $I_0 =$  zero current level.



**Figure 2.13 Progressive additions of RNAs with different U-tails length.** **a**, Residual current distributions of the blockades events for different ssRNAs sequentially added (from top to bottom) to the *cis* chamber. **b**, mean residual current ( $I_{RES}\%$ ) of the Gaussian distribution of events for each RNA is reported. Error bars.  $\pm$  SD. Standard deviation is relative to the mean residual current of at least three independent experiments. Data were acquired in *high salt buffer*. 20 kHz acquisition, 2 kHz low-pass filter; applied potential: + 120 mV. 2  $\mu$ M final RNA concentration.



**Figure 2.14** All data points scatter plot  $I_{RES}\%$  vs dwell time for a typical experiment with  $C_5U_5$ ,  $C_4U_6$ ,  $C_3U_7$ ,  $U_{10}$ ,  $U_{20}$  in the *cis* chamber. The majority of the events are longer than 5 ms (red dashed line) also in *high salt buffer*. Specific population of events for each RNA are defined by color-coded dash-line boxes.  $C_5U_5$  has a fast population of events (<10 ms, violet box). Data were acquired in *high salt buffer*. 20 kHz acquisition, 2 kHz low-pass filter; +120 mV. 2  $\mu$ M final RNA concentration.

$A_{15}U_4$  gave a  $\bar{\tau}_D$  of  $116 \pm 8$  ms ( $n=3$ ), which is  $\sim 20$  times longer than the mean dwell time observed for the WT- $\alpha$ HL in the same conditions (WT- $\alpha$ HL cannot distinguish fewer than five uridines at the 3' on a poly $A_{15}$ ; the dwell time of the population of longer events were  $+5.5 \pm 0.8$  ms ( $n=3$ ) for  $A_{15}U_4$  and  $+5.5 \pm 0.4$  ms for a  $A_{15}U_5$ ;  $n=3$ ). Well defined populations of events longer than 5 ms were detected, using the D128N pore, for  $A_{15}U_2$ ,  $A_{15}U_3$ ,  $A_{15}U_4$  and  $A_{15}U_5$  (Table 2.3).

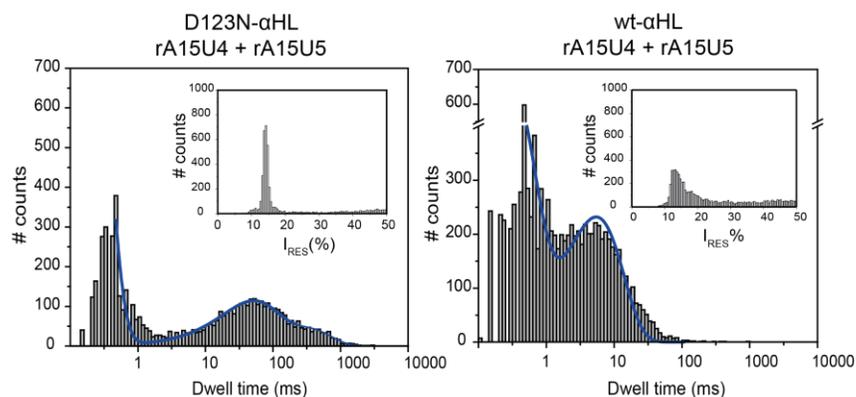
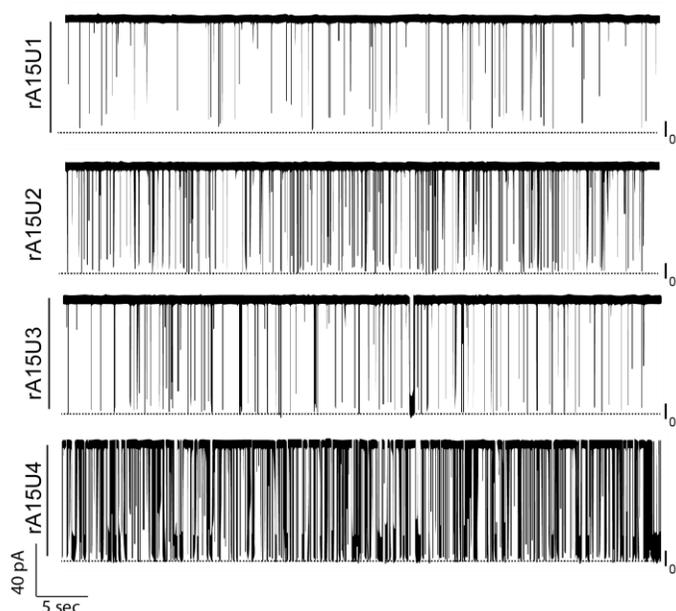
RNA	$\tau_D$ (ms) <sup>LS</sup>	$I_{RES\%}$ <sup>LS</sup>	$\tau_D$ (ms) <sup>HS</sup>	$I_{RES\%}$ <sup>HS</sup>
$A_{15}U_2$	$1.1 \pm 0.2$	$1.5 \pm 0.2$	<1	$13 \pm 2$
$A_{15}U_3$	$15.2 \pm 0.3$	$1.5 \pm 0.2$	$3.4 \pm 0.7$	$13 \pm 2$
$A_{15}U_4$	$116 \pm 8$	$1.5 \pm 0.2$	$75 \pm 4$	$13 \pm 2$
$A_{15}U_5$	$820 \pm 80$	$1.5 \pm 0.2$	$200 \pm 100$	$13 \pm 2$

**Table 2.3.** Mean dwell time and residual current (%) of different RNA oligos detect by D128N- $\alpha$ HL.  $I_o$  (open current level) =  $24 \pm 2$  pA in LS;  $I_o = 130 \pm 5$  pA in HS. Potential applied = +120 mV. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. HS: *high ionic strength buffer*; LS: *low ionic strength buffer*.

After simultaneous addition of all these different ssRNAs ( $A_{15}U_1/A_{15}U_2/A_{15}U_3/A_{15}U_4/A_{15}U_5$ ) to the *cis* chamber, four populations of events could be distinguished and a four-component probability density function fitted (Figure 2.2.11b, bottom panel). The mean dwell time value of each population corresponds to the  $\bar{\tau}_D$  of the specific U-tail length (Table 2.3). Few short events ( $\bar{\tau}_D < 1$  ms) were observed even with  $A_{15}U_1$ . For all the oligonucleotides, a second population of shorter blockades (~1-5 ms) was recorded. The oligo  $A_{15}U_2$  showed only one population of shorter events, with a mean dwell time of  $1.1 \pm 0.1$  ms (n=3). These experiments demonstrate that the D128N- $\alpha$ HL pore can detect poly(A)<sub>15</sub> with more than one uridine in 3' and it can clearly distinguish the U length of a mixture of RNAs with more than two terminal uridines.

In *high ionic strength buffer*, the differences in the dwell time between RNAs are smaller than observed with longer ( $\geq 5$ ) U tails with the WT- $\alpha$ HL (Table 2.3). In both conditions (*high and low salt buffer*) the residual current does not significantly differ between ssRNAs ( $+1.5 \pm 0.2$  pA at low salt and  $+23 \pm 2$  pA at high salt) (Figure 2.15).

### D128N detection



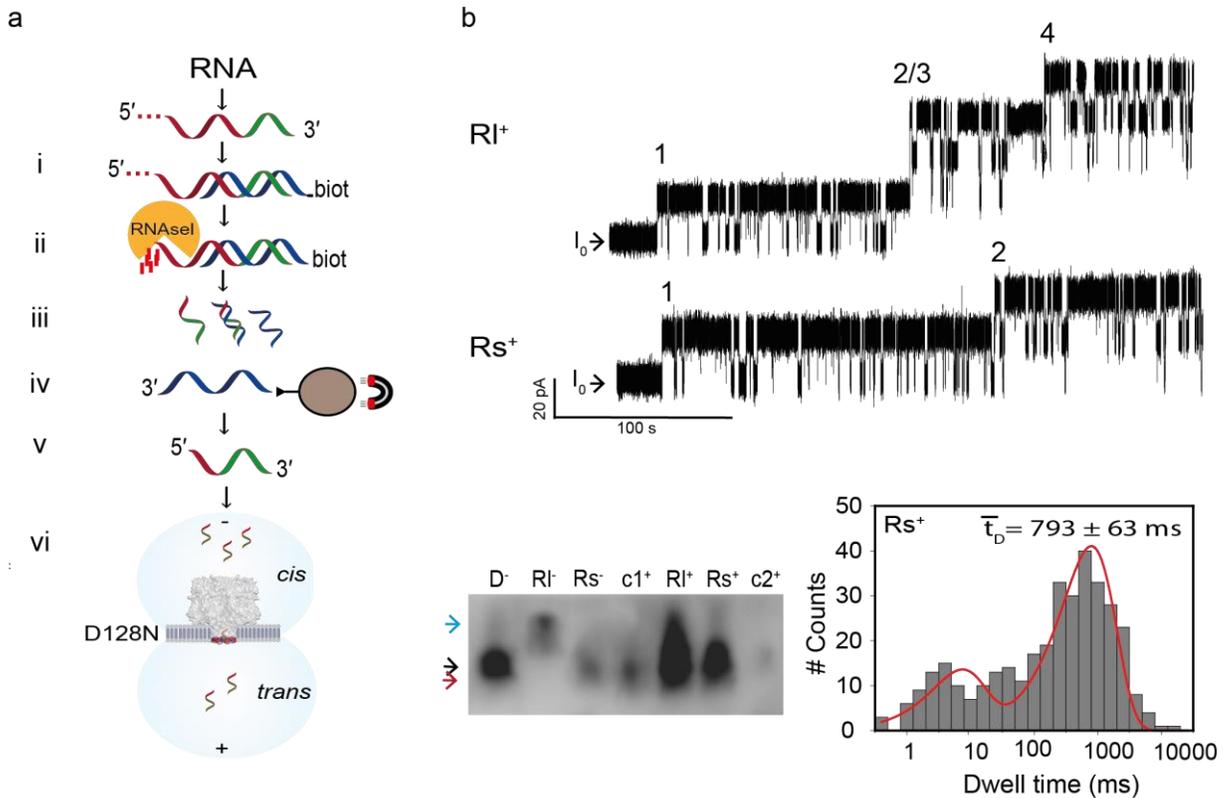
**2.15 Detection of  $A_{15}U_x$  ( $x < 5$ ) with the D128N- $\alpha$ HL pore and comparison with the WT- $\alpha$ HL pore.** **Top:** Current traces representing the blockades from  $A_{15}U_1$ ,  $A_{15}U_2$ ,  $A_{15}U_3$ , and  $A_{15}U_4$  respectively in the D128N- $\alpha$ HL pore. Acquisition: 20khz and 2khz low-pass filter; +120mV. Data acquired in *high ionic strength buffer*. 2 $\mu$ M RNA concentration. **Bottom:** Dwell time histogram of  $A_{15}U_4$  and  $A_{15}U_5$  (1  $\mu$ M each) is reported with the D128N- $\alpha$ HL pore (**left**) and WT- $\alpha$ HL pore (**right**). A double probability density function has been fitted on the data. Inset:  $I_{RES\%}$  of the RNA mixture is shown and fitted to a Gaussian. No discrimination on the  $I_{RES\%}$  is possible with the D128N- $\alpha$ HL pore.

In conclusion, the specific discrimination of longer (> 10 nt) ssRNAs with short U tails can be achieved with the D128N- $\alpha$ HL mutant, in a *low ionic strength buffer*, comparing the mean dwell times.

Other than the recognition of uridylation length, the  $\alpha$ HL pore could be used for the detection of ssRNAs containing non-canonical nucleobases. Ribonucleic acids exhibit more than 100 nucleoside structure variations<sup>111</sup>. For example, the 5-methyluridine ( $m^5U$ ) is a common RNA modification in all of the three domains of life and it mainly affects tRNAs and rRNA<sup>57</sup>. We compared  $C_5U_5$  with  $C_5(m^5U)_5$  and observed blockades with a distinctive double level structure (step 1 and step 2) for the methylated oligo (Figure 2.11c). The two steps of a same event have a mean dwell time of  $\bar{\tau}_{D1} = 143 \pm 7$  ms ( $n = 5$ ) and  $\bar{\tau}_{D2} = 56 \pm 3$  ms ( $n = 5$ ). The total residential time (step 1 + step 2) was longer than the control  $C_5U_5$  ( $\bar{\tau}_D = 56 \pm 7$  ms,  $n=16$  at + 80 mV).

These results demonstrate the ability of the nanopore to discriminate between strands of normal uridine and common noncanonical uridine nucleobases.

**2.2.7 Purification of defined RNA signature fragments.** In order to find a practical solution for the purification and sensing of 3' oligo U signatures starting from common RNA samples, we propose an easy protocol for the selective enrichment of 3' oligo U fragments with a defined length, from longer RNAs (Figure 2.16 and Figure 2.17).



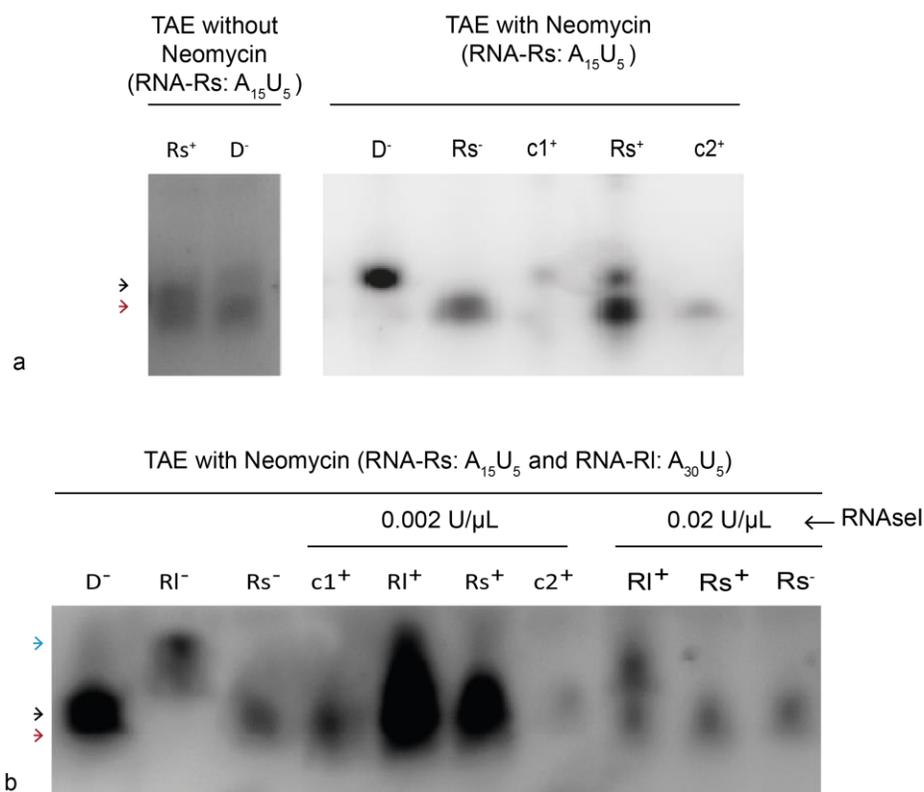
**Figure 2.16 Protocol for RNA purification and analysis.** **a**, (i) Annealing of  $A_{30}U_5$  (RI) or  $A_{15}U_5$  (Rs) with the 5'-biotinylated DNA probe  $A_{20}T_{15}$  in  $1 \times$  TAE/Mg $^{2+}$ /Neomycin buffer. (ii) The 5' overhanging ssRNA has been digested with RNaseI and (iii) purified by spin columns. (iv) The DNA probe was separated by streptavidin magnetic beads and the RNA eluted. (v) After DNase treatment, RNA fragments were purified again by miRNAeasy column and eluted in 30  $\mu$ L DMPC-treated water., mixed 1 : 1 with a buffer 300 mM KCl, 200 mM NaCl, 4 mM MgCl $_2$  and 20 mM HEPES,pH 6.5 added of of D128N- $\alpha$ HL and (vi) finally the RNA was detected in DIBs. For details of the procedure see supporting informations (S1). **b - top**, current traces showing RNA blockades (RI top, Rs bottom) and multiple D128N pore insertions (labeled with numbers at the insertion point) are reported. RI $^+$ :  $A_{30}U_5$  processed as in 'a'. Rs $^+$ :  $A_{15}U_5$  processed as in 'a'.  $I_0$ : zero current level. **Bottom - left**, 6 M Urea PAGE gel for the DNA and RNAs used. D $^-$ : DNA probe; RI $^-$ :  $A_{30}U_5$ ; Rs $^-$ :  $A_{15}U_5$ ; c1 $^+$ : control without RNA processed as in 'a'; RI $^+$ :  $A_{30}U_5$  processed as in 'a';  $A_{15}U_5$  processed as in 'a'; c2 $^+$ : control without the DNA probe processed as in 'a'. **Bottom - right**: dwell time analysis of the blocking events observed with Rs $^+$ . The mean dwell time of the longest population of events signed and a probability density function is fitted (red line).Traces were recorded in symmetric conditions (low ionic strength buffer). The signal was filtered at 1 kHz (low-pass Bessel filter) and acquired at 5 kHz.

We optimized the protocol with two synthetic RNAs ( $A_{15}U_5$ ,  $A_{30}U_5$ ). The first was used as internal standard because it has the same length of a protective DNA probe; the second was 15 nt longer than the DNA. We anealed the 3' of each RNA with a specific DNA ( $A_{20}T_{15}$ ) probe to form a hybrid  $\beta$  double helix and after digestion of the 5' overhangs, the RNA fragments were analyzed by droplet interface bilayer (DIB)<sup>80,81</sup>. We used the DIBs setup in order to be able to work a with final RNA concentration (droplet vol-

ume < 200 nL) high enough to collect a useful number of events in a sensible time (to minimize RNA degradation). Briefly, the RNAs were processed through the steps i-vi described in figure 2.16:

- (i) The ssRNA was mixed with the 5' biotinylated DNA probe A<sub>30</sub>T<sub>5</sub> the DNA probe (1 : 1) in 1 × TAE/Mg<sup>2+</sup>/Neomycin buffer. The mixed aqueous solution was incubated at 95 °C for 2 min followed by 65°C for 10 min, 50 °C for 10 min, 37 °C for 10 min, 25 °C for 10 and 30 ° at 10°C. TAE buffer was previously used for RNA:DNA hybridization<sup>112</sup>. Neomycin has been described to enhance and stabilize the β-duplex of the RNA:DNA hybrid<sup>113</sup>.
- (ii) The 5' overhanging ssRNA was digested with RNaseI (0.002 U/μL) for 15 min. The enzyme was stopped by adding SUPERaseIN RNase Inhibitor (20U/ul). The RNA/DNA mixture was purified by miRNAeasy columns.
- (iii) The DNA probe was separated by magnetic beads (dynabeads MyOne Streptavidin C1, RNase-free, Invitrogen) in a buffer containing 1 M NaCl, 5 mM Tris-HCL, 0.5 mM EDTA (pH 7.5). The RNA was eluted after incubation at 65°C for 2 min. The high salt concentration is required for the optimal binding of the DNA on the beads.
- (iv) The solution was than diluted to 200 mM NaCl, traded with DNase RQ<sub>1</sub> for 30 min and RNA fragments were purified again by miRNAeasy column and eluted in 30 μL DMPC-treated water.
- (v) Purified RNA was mixed 1:1 with a buffer 300 mM KCl, 200 mM NaCl, 4 mM MgCl<sub>2</sub> and 20 mM HEPES, pH 6.5 in DMPC-treated water, 0.2 μL of a diluted solution of D128N-αHL was added to the sample and finally the RNA was detected in DIBs. Blockades from purified RNA fragments showed the characteristic dwell time distribution previously characterized in PLM.

Critical steps in the process were the incubation time with the RNaseI and presence of Neomycin in the TBE buffer. These two parameters must be tuned carefully in order to optimize the final yield, as reported in figure 2.17.



**Figure 2.17 Effects of neomycin and incubation times on the final yield. a - left**, PAGE gel 6M Urea with A<sub>15</sub>U<sub>5</sub> (Rs) processed through point (i-vi) of figure 2.2.15 with annealing buffer TAE + 5mM Mg<sup>++</sup>. **Right**, A<sub>15</sub>U<sub>5</sub> (Rs) and A<sub>30</sub>U<sub>5</sub> (RI) processed through point (i-vi) of figure 2.16 with annealing buffer TAE + 5mM Mg<sup>++</sup> + 10 μM Neomycin, pH 7.8. Incubation times: 5 min RNaseI, 10 min DNase I. **b**, A<sub>15</sub>U<sub>5</sub> (Rs) and A<sub>30</sub>U<sub>5</sub> (RI) processed through point (i-vi) of figure 2.2.15 with annealing buffer TAE + 5mM Mg<sup>++</sup> + 10 μM Neomycin, pH 7.8. Incubation times: 15 min RNaseI, 30 min DNaseI. RI and Rs purification has been performed with two different RNaseI concentrations as indicated in the panel. (-) controls, not processed through point i-vi; 200 ng of RNA and DNA were loaded in D<sup>-</sup>, RI<sup>-</sup> and Rs<sup>-</sup>. (+) processed samples; RNA was mixed 1:1 (mol) with DNA probes (2μL - 100 μM DNA + 1.5 μL 200 μM RNA). D<sup>-</sup>: DNA probe; RI<sup>-</sup>: A<sub>30</sub>U<sub>5</sub>; Rs<sup>-</sup>: A<sub>15</sub>U<sub>5</sub>; c1<sup>+</sup>: control without RNA; RI<sup>+</sup>: A<sub>30</sub>U<sub>5</sub> processed; Rs<sup>+</sup>: A<sub>15</sub>U<sub>5</sub> processed; c2<sup>+</sup>: control without the DNA. Blue arrow: A<sub>30</sub>U<sub>5</sub>; black arrow: DNA probe; Red arrow: A<sub>15</sub>U<sub>5</sub>.

With this experiment we demonstrated that the use of the αHL pore as a sensor for this important post-transcriptional modification was feasible also in the perspective to use more complex RNAs, as common biological RNA samples are.

### 2.3 Conclusion.

Our findings show the potential of the  $\alpha$ HL pore to selectively sense a specific biologically relevant RNA signature (3'-end uridylation) at the single molecule level. Together, the results demonstrate that the  $\alpha$ HL pore is an RNA binding nanopore because:

1. 3' oligo-uridylated RNAs gave a distinctive pattern of ionic current.
2. the  $\bar{\tau}_D$  of the blockades is sensitive to structural variation of the U tail and to the ionic strength of the buffer.
3. the slot of RNA-binding pockets that line the *trans* entrance of the  $\beta$ -barrel have a modular interaction with the RNA.
4. the  $\alpha$ HL pore can discriminate, the length of the U-tails and the presence of non-canonical m<sup>5</sup>U nucleobases.

In conclusion, the strong selectivity and binding affinity ( $\sim 1 \mu\text{M}$ ) prospect the very intriguing application of this biological nanopore as a fast, simple and reliable stochastic sensor for 3'-end uridylations, a post-transcriptional RNA modification.

## 2.4 Methods.

**2.4.1 Single channel recordings.** As described in detail previously<sup>1</sup>, a 1,2-diphytanoyl-sn-glycero-3-phosphocholine (Avanti Polar Lipids) bilayer (~100  $\mu\text{m}$  diameter) was created between the two compartments (each 1 mL) of a bilayer recording chamber. Experiments were performed under symmetrical buffer conditions. We used two buffers: low ionic strength buffer (150 mM KCl, 100 mM NaCl, 2 mM MgCl<sub>2</sub> and 10 mM HEPES at pH 6.5) or high ionic strength buffer (1M KCl, 10 mM Tris, 0.1 mM EDTA, pH 7.5). All solutions were made using water (18.2 M $\Omega$ cm, Millipore) treated with 0.1% v/v DMPC (dimethyl-propyl carbonate, a safer alternative to DEPC) overnight at room temperature and then autoclaved to remove residual DMPC. The protein was added to the grounded cis compartment. The experimental voltage was applied through a pair of Ag/AgCl electrodes set in bridges of 2% agar in 3.0 M KCl. After the insertion of a single  $\alpha$ HL pore, the buffer was repeatedly replaced by manual pipetting to avoid multiple insertions. ssRNAs were introduced into the cis compartment at a concentration of 0.5  $\mu\text{M}$  to 4  $\mu\text{M}$ , and after stirring incubated in the electrolyte solution for ~5 min prior to data recording. An IV-curve and a control recording were performed prior to RNA addition. The single channel current was amplified by using a patch-clamp amplifier (Axopatch 200B, Axon Instruments), filtered with a low-pass Bessel filter (80 dB/decade) with a corner frequency of 2 kHz and then digitized with a Digidata 1320 A/D converter (Axon Instruments) at a sampling frequency of 20 kHz. The signal was not filtered further unless otherwise stated. The acquisition software was Clampex 10.2 (Molecular Devices). The measurements were conducted at  $20 \pm 2$  °C.

### 2.4.2 Data analysis

Data analysis was performed with a custom python script using scipy, cython and neo libraries<sup>117-118</sup>. The analysis was based on threshold search. For the calculation of kinetic values the traces were divided into two levels. The transition between the open and closed levels was confirmed if both the rolling median and the actual current value

crossed the threshold level. The histograms of the logarithmic open and closed times were fitted on a probability density functions (Pdf norm, 1), (single or multiple components):

$$\text{Pdf norm} = 1/\tau \exp[ti - 1/\tau \exp(ti)] \quad (1)$$

Where, Pdf norm is the normalized frequency,  $\tau$  is the mean dwell or inter event time,  $t_i$  is the duration of individual events. The rate constants  $k_{\text{on}}$  ( $M^{-1} s^{-1}$ ) and  $k_{\text{off}}$  (s) of the events were used to calculate the probability that the channel was blocked ( $P_{\text{block}\%}$ ) based on equation (2).

$$P_{\text{block}\%} = \{1 - [\tau_{\text{on}}/(\tau_{\text{on}} + \tau_{\text{off}})]\} \times 100 \quad (2)$$

Because the signature events for the oligo U tails (~5 – 5000 ms) were well separated in duration from the normal oligonucleotide translocations (~10 – 200  $\mu$ s), we set 1 ms as the cut-off to analyze the events, unless otherwise stated. We considered a binding event only when it was longer than 5 ms. A second population of shortest event (<5 ms) probably due to incomplete binding is sometimes recorded in high ionic strength buffer. In this situation the kinetic constants were calculated only on the longer population of events.

The residual current ( $I_{\text{RES}\%}$ ) of the RNA blockades was calculated with equation (3)

$$I_{\text{RES}\%} = I_{\text{O}} - I_{\text{b}}/I_{\text{O}} * 100 \quad (3)$$

Where  $I_{\text{O}}$  is the open current and  $I_{\text{b}}$  is the amplitude of the current block event during RNA binding and translocation.

Data were presented as mean  $\pm$  s.d. of at least three independent experiments, and the differences were considered statistically significant at  $P < 0.05$  using the Student's t-test.

#### **2.4.3 *In vitro* protein expression**

Radio-labeled  $\alpha$ HL was produced by coupled *in vitro* transcription and translation (IVTT) using an *E. coli* T7-S30 expression system for circular DNA (Promega) according to the manufacturer's instructions. Heptamers were purified from a 6% SDS-PAGE gel. The region of the dried gel containing  $\alpha$ HL heptamers was cut out, rehydrated and crushed in 10 mM TRIS-HCl, pH 8, containing 100  $\mu$ M EDTA. Proteins were purified from SDS by buffer exchange and centrifugation in a 100 KD cut-off filter (MICROCON, Millipore) for 10 min at 4°C, 25000 x rpm. Aliquots of the purified proteins were stored at -80°C. For planar lipid bilayer experiments the protein solution was diluted 1:20 and a portion (1-10 $\mu$ L) was added to the *cis* compartment of a bilayer apparatus (see single channel recording).

#### **2.4.4. Mutagenesis**

Mutant  $\alpha$ HL genes were prepared by using a site-directed mutagenesis kit (Quik-Change XL®, Life Technology). All mutants were made by using the WT- $\alpha$ HL gene as template. The  $\alpha$ HL gene was mutated at the constriction of the  $\beta$ -barrel (Glu111 $\rightarrow$ Asn, Lys147 $\rightarrow$ Asn). We called this mutant NN-*cis*. The mutant NNA-*trans* was prepared by changing three residues on the loop between the two  $\beta$ -strands on the *trans* side of the  $\beta$ -barrel (Asp127 $\rightarrow$ Asn, Asp128 $\rightarrow$ Asn, Lys131 $\rightarrow$ Ala). For the NNA-*trans* mutant we also prepared all the possible (six) combination in order to check which residues or combinations of mutations residues is more important for oligo U binding. The coding regions of the  $\alpha$ HL mutants were verified by sequence analysis. The  $\alpha$ HL(D8H6) was prepared cloning a D8H6 tail in the C-terminal of the WT- $\alpha$ HL. This produces a shift in a SDS-PAGE and allows the separation and the extraction of different heterooligomers as previously described<sup>102,103</sup>.

### 2.4.5 Thermodynamic analysis

The melting temperature analysis of the ssRNAs reported in Table 2.4.1 was performed using an online calculator of oligonucleotide properties (OligoCalc, <http://www.basic.northwestern.edu/biotools/OligoCalc.html#helpthermo>) with the Nearest-Neighbor method<sup>114</sup>. The calculations were performed for a simulated NaCl concentration of 1M. The analysis of RNA folding was done using the software mFold (<http://mfold.rna.albany.edu/?q=mfold/RNA-Folding-Form2.3.>). The calculations were performed for a simulated temperature of 20 °C and for a RNA concentration of 10 μM and the results are summarized in Table 2.4

<u>n</u> (5'-A15U <u>n</u> )	$\Delta G$ (sum of minimum free energies)	number of secondary structures	T <sub>m</sub> (°C)
0	-	0	+12.1
1	-	0	+15.8
2	+21.07	9	+18.2
3	+10.18	9	+20.7
4	-1.06	8	+22.8
5	-9.56	7	+24.8
6	-18.01	7	+26.5

Table 2.4 Thermodynamic analysis

### 2.4.6 Periodate oxidation

The cis-diol group on the 3' end (without phosphate) was cleaved with 10 mM sodium periodate (Sigma) to generate the 2', 3'-dialdehyde. The reaction was carried out for 40 min on ice in a total volume of 50 μL containing the 10 nmol RNA. The reaction was quenched with 5 μL of 0.1% v/v glycerol (Sigma) in RNase free water (Ambion) for 1 h on ice. The reaction mixture was used directly for single channel recording and analyzed by HPLC and MALDI-MS. The reaction gave 100% of conversion.

#### 2.4.7 HPLC

ssRNAs were purchased from Sigma or IDT Technology at the HPLC pure grade. ssRNAs were then purified by reverse phase high performance liquid chromatography (RP-HPLC) using an Eclipse Plus C18 column (3.5 $\mu$ m, 4.6x100mm). The column was eluted with a gradient of 0% to 40%, of acetonitrile 80% (Sigma) over 30 min followed by a 5 min washing. The aqueous solution was 0.1 M TEAA (triethylamine acetate, Sigma) in DMPC water (0.1%). A flow rate of 1 mL min<sup>-1</sup> was employed. The peak, detected by absorbance at 260 nm, was collected and diluted with the appropriate matrix for MALDI-MS analysis or lyophilized and dissolved in RNase free water (Ambion) for further use in single channel recording.

#### 2.4.8 MALDI-MS

Fractions of the RP-HPLC purified RNA were collected in a final volume of 0.5 ml. A small aliquot was mixed in 1:1 ratio with 3-hydroxypicolinic acid (Sigma) dissolved at saturation in 50% CH<sub>3</sub>CN (Sigma). From this combined matrix/sample solution, 2  $\mu$ L was spotted onto a steel target and allowed to co-crystallize at room temperature for 30 minutes. MALDI analysis was conducted with a Water MALDI Micro MX spectrometer with TOF detection, in positive reflection mode. A standard mixture (Invitrogen) was used as a lock mass calibrant. Laser energy and pulse width were optimized. Data were further processed using Mass Lynx 4.1.

#### 2.4.9 RNA sequences

Oligonucleotides, were synthesized and PAGE-purified by Integrated DNA Technologies or Sigma (Table 2.5). RNase-free water was used to prepare RNA samples.

RNA length (nt)	Sequence (5'-3')	3' U tail length (nt)	Suppliers
10	CCCCUUUUU	5	Sigma and IDT

10	CCCCCCCCCU	1	Sigma and IDT
10	CCCCCCUUU	3	Sigma and IDT
10	CCCCCUUUU	4	Sigma and IDT
10	CCCCUUUUU	6	Sigma and IDT
10	CCCUUUUUU	7	Sigma and IDT
10	UUUUUUUUU	10	Sigma and IDT
10	AAAAUUUUU	5	Sigma and IDT
10	CCCCCCCCC	0	Sigma and IDT
10	CCCCGGGGG	5	Sigma and IDT
10	CCCCAAAAA	0	Sigma and IDT
20	AAAAAAAAAAAAAAAAAAAA	0	Sigma and IDT
20	UUUUUUUUUUUUUUUUAAA	0	Sigma and IDT
20	AAAAAAAAACCCCCCCCC	0	Sigma
16	AAAAAAAAAAAAAAAAAU	1	Sigma
17	AAAAAAAAAAAAAAAAAUU	2	Sigma
18	AAAAAAAAAAAAAAAAAUUU	3	Sigma
19	AAAAAAAAAAAAAAAAAUUUU	4	Sigma
20	AAAAAAAAAAAAAAAAAUUUUU	5	Sigma
40	U <sub>40</sub>	40	Sigma and IDT
40	Biot- U <sub>40</sub>	40	Sigma
40	C <sub>35</sub> UUUUU	5	Sigma
25	C <sub>20</sub> UUUUU	5	Sigma
15	CCCCCCCCCUUUUU	5	Sigma
10	UUUUUUUUU(dU)	0	Sigma
10	UUUUUUUUUC	0	Sigma
10	TTTTTTTTTT	0	Sigma
10	CCUUUUCCC	0	Sigma
10	UUU(dU)(dU)(dU)UUUU	4	Sigma
10	UU(dU)(dU)(dU)UUUUU	5	Sigma
10	CCUUUUUUU-3'P	7*	Sigma
10	CCCC(dU) (dU) (dU) (dU) (dU)	0	Sigma
10	CCC(m <sup>5</sup> U) (m <sup>5</sup> U) (m <sup>5</sup> U) (m <sup>5</sup> U) (m <sup>5</sup> U)	5**	Sigma
10	CCC(2'MeU) (2'MeU) (2'MeU) (2'MeU) (2'MeU)	5***	Sigma

**Table 2.5 RNA sequences**

(\*) 3'-P: 3' ribose phosphorylation on the last nucleotide

(\*\*) m<sup>5</sup>U: 5-methyluridine

(\*\*\*) 2'MeU: 2'-O- methyluridine

Oligonucleotides, were synthesized and HPLC-purified by Integrated DNA Technologies or Sigma. RNase-free water was used to prepare RNA samples. Noncanonical or deoxy- nucleobases are in brackets. Biot: 5' biotinylated U<sub>40</sub> RNA

#### **2.4.10. BindN analysis**

The RNA-protein binding prediction is extremely challenged but many approaches are today available for an accurate estimation of the protein-nucleic acid interaction. Software are usually based on machine-learning based approaches, other on template-based methods<sup>15</sup>. We used a free online tool as BindN<sup>93</sup> (<http://bioinfo.ggc.org/bindn/>) for the detection of a sensitive region for RNA interaction in the  $\beta$ -barrel of the  $\alpha$ HL pore.

The submission of the amino acid protein sequence that spans the lipid bilayer (EYMSTLTYGFNGNVTGDDTGKIGGLIGANVSIGHTLK) gave a high score for binding only between Tyr-125 to Lys-131 included. The software is a machine-learning based tool and it predicts RNA-binding residues from multiple sequence features, including the side chain pKa values, hydrophobicity index and molecular mass of each amino acid.

## 2.5 References

1. Wickens, M. & Kwak, J. E. Molecular biology. A tail tale for U. *Science (New York, N.Y.)* **319**, 1344–5 (2008).
2. Gallouzi, I. E. & Wilusz, J. A {DIStinctively} novel exoribonuclease that really likes U. *The EMBO Journal* **32**, 1799–1801 (2013).
3. Scott, D. D. & Norbury, C. J. RNA decay via 3' uridylation. *Biochimica et biophysica acta* 1–12 (2013). doi:10.1016/j.bbagr.2013.01.009
4. Choi, Y. S. U. N. S., Patena, W., Leavitt, A. D. & McManus, M. T. Widespread RNA 3'-end oligouridylation in mammals. *RNA (New York, N.Y.)* **18**, 394–401 (2012).
5. Burroughs, a M., Kawano, M., Ando, Y., Daub, C. O. & Hayashizaki, Y. pre-{miRNA} profiles obtained through application of locked nucleic acids and deep sequencing reveals complex 5'/3' arm variation including concomitant cleavage and polyuridylation patterns. **40**, 1424–37 (2012).
6. Newman, M. A., Mani, V. & Hammond, S. M. Deep sequencing of microRNA precursors reveals extensive 3' end modification. *RNA (New York, N.Y.)* **17**, 1795–803 (2011).
7. Sement, F. M. *et al.* Uridylation prevents 3' trimming of oligoadenylated mRNAs. *Nucleic acids research* 1–13 (2013). doi:10.1093/nar/gkt465
8. Lehrbach, N. J. *et al.* LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in *Caenorhabditis elegans*. *Nature structural & molecular biology* **16**, 1016–20 (2009).
9. Rissland, O. S. & Norbury, C. J. The Cid1 poly(U) polymerase. *Biochimica et biophysica acta* **1779**, 286–94 (2008).
10. Yates, L. A. A., Norbury, C. J. J. & Gilbert, R. J. C. J. C. The Long and Short of {MicroRNA}. *Cell* **153**, 516–519 (2013).
11. Menestrina, G. Ion channels and bacterial infection: the case of  $\beta$ -barrel pore-forming protein toxins of *Staphylococcus aureus*. *FEBS Letters* **552**, 54–60 (2003).
12. Song, L. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science (New York, N.Y.)* **274**, 1859–66 (1996).
13. Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Accounts of Chemical Research* **35**, 817–825 (2002).
14. Vercoutere, W. *et al.* Rapid discrimination among individual DNA hairpin molecules at single-nucleotide resolution using an ion channel. *Nature biotechnology* **19**, 248–52 (2001).
15. Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **413**, 226–30 (2001).
16. Stoddart, D., Maglia, G., Mikhailova, E., Heron, A. J. & Bayley, H. Multiple base-recognition sites in a biological nanopore: two heads are better than one. *Angewandte Chemie (International ed. in English)* **49**, 556–9 (2010).
17. Wallace, E. V. B. *et al.* Identification of epigenetic DNA modifications with a protein nanopore. *Chemical communications (Cambridge, England)* **46**, 8195–7 (2010).
18. Ayub, M. & Bayley, H. Individual RNA base recognition in immobilized oligonucleotides using a protein nanopore. *Nano letters* **12**, 5637–43 (2012).
19. Seol, Y., Skinner, G., Visscher, K., Buhot, A. & Halperin, A. Stretching of Homopolymeric RNA Reveals Single-Stranded Helices and Base-Stacking. *Physical Review Letters* **98**, 158103 (2007).

20. Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M. & Gundlach, J. H. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20647–52 (2008).
21. Clarke, J., Wu, H., Jayasinghe, L. & Patel, A. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature ...* **4**, (2009).
22. Lieberman, K. & Cherf, G. Processive replication of single DNA molecules in a nanopore catalyzed by phi29 DNA polymerase. *Journal of the ...* 17961–17972 (2010). at <<http://pubs.acs.org/doi/abs/10.1021/ja1087612>>
23. Cherf, G. M. *et al.* Automated forward and reverse ratcheting of DNA in a nanopore at 5-Å precision. *Nature biotechnology* **30**, 344–8 (2012).
24. Manrao, E. a *et al.* Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature biotechnology* **30**, 349–53 (2012).
25. Hall, A. R. *et al.* Hybrid pore formation by directed insertion of  $\alpha$ -haemolysin into solid-state nanopores. *Nature nanotechnology* **5**, 874–7 (2010).
26. Venta, K., Shemer, G. & Puster, M. Differentiation of Short Single-Stranded {DNA} Homopolymers in Solid-State Nanopores. *ACS ...* 4629–4636 (2013). at <<http://pubs.acs.org/doi/abs/10.1021/nn4014388>>
27. Sauer-Budge, A., Nyamwanda, J., Lubensky, D. & Branton, D. Unzipping Kinetics of Double-Stranded DNA in a Nanopore. *Physical Review Letters* **90**, 238101 (2003).
28. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nature biotechnology* **19**, 636–9 (2001).
29. Iqbal, S. M., Akin, D. & Bashir, R. Solid-state nanopore channels with DNA selectivity. *Nature nanotechnology* **2**, 243–8 (2007).
30. Wanunu, M. *et al.* Rapid electronic detection of probe-specific microRNAs using thin nanopore sensors. *Nature nanotechnology* **5**, 807–14 (2010).
31. Wang, Y., Zheng, D., Tan, Q., Wang, M. X. & Gu, L.-Q. Nanopore-based detection of circulating microRNAs in lung cancer patients. *Nature nanotechnology* **6**, 668–74 (2011).
32. Lin, J., Fabian, M., Sonenberg, N. & Meller, A. Nanopore detachment kinetics of poly(A) binding proteins from RNA molecules reveals the critical role of C-terminus interactions. *Biophysical journal* **102**, 1427–34 (2012).
33. Ray, D. *et al.* A compendium of {RNA-binding} motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
34. CANELLAKIS, E. S. Incorporation of radioactive uridine-5'-monophosphate into ribonucleic acid by soluble mammalian enzymes. *Biochimica et biophysica acta* **23**, 217–8 (1957).
35. Burdon, R. & Shenkin, A. Uridylate-rich sequences in rapidly labelled RNA of mammalian cells. *FEBS letters* **24**, 1–4 (1972).
36. Jupin, I., Bouzoubaa, S., Richards, K., Jonard, G. & Guilley, H. Multiplication of beet necrotic yellow vein virus RNA 3 lacking a 3' poly(A) tail is accompanied by reappearance of the poly(A) tail and a novel short U-rich tract preceding it. *Virology* **178**, 281–4 (1990).
37. Kwak, J. E. & Wickens, M. A family of poly(U) polymerases. *RNA (New York, N.Y.)* **13**, 860–7 (2007).
38. Lunde, B. M., Magler, I. & Meinhart, A. Crystal structures of the Cid1 poly (U) polymerase reveal the mechanism for UTP selectivity. *Nucleic acids research* **40**, 9815–24 (2012).
39. Schmidt, L. L. E., West, S., Norbury, C. J. & Schmidt, M.-J. The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. *RNA (New York, N.Y.)* **44**, 39–44 (2011).

40. Sement, F. M. *et al.* Uridylation prevents 3' trimming of oligoadenylated {mRNAs}. 1–13 (2013). doi:10.1093/nar/gkt465
41. Jones, M. R. *et al.* Zcchc11-dependent uridylation of microRNA directs cytokine expression. *Nature cell biology* **11**, 1157–63 (2009).
42. Ren, G., Chen, X. & Yu, B. Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. *Current biology : CB* **22**, 695–700 (2012).
43. Katoh, T. *et al.* Selective stabilization of mammalian microRNAs by 3' adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. *Genes & development* **23**, 433–8 (2009).
44. Heo, I. *et al.* Mono-uridylation of pre-microRNA as a key step in the biogenesis of group II let-7 microRNAs. *Cell* **151**, 521–32 (2012).
45. Malecki, M. *et al.* The exoribonuclease Dis3L2 defines a novel eukaryotic RNA degradation pathway. *The EMBO journal* **32**, 1–13 (2013).
46. Lubas, M. *et al.* pathway of human cytoplasmic {mRNA}. *The EMBO Journal* **32**, 1855–1868 (2013).
47. Chang, H., Triboulet, R., Thornton, J. E. & Gregory, R. I. A role for the Perlman syndrome exonuclease Dis3L2 in the Lin28-let-7 pathway. *Nature* **497**, 244–248 (2013).
48. Yates, L. a *et al.* Structural basis for the activity of a cytoplasmic RNA terminal uridylyl transferase. *Nature structural & molecular biology* **19**, 782–7 (2012).
49. Rissland, O. & Norbury, C. 3' uridylation precedes decapping in a novel pathway of bulk mRNA turnover. *Nature structural & molecular biology* **16**, 616–623 (2009).
50. Choi, L.-S., Mach, T. & Bayley, H. Rates and Stoichiometries of Metal Ion Probes of Cysteine Residues within Ion Channels. *Biophysical journal* **105**, 356–64 (2013).
51. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nature Nanotechnology* **8**, 288–295 (2013).
52. Wanunu, M. Nanopores: A journey towards DNA sequencing. *Physics of life reviews* **9**, 125–58 (2012).
53. Van den Hout, M., Skinner, G. M., Klijnhout, S., Krudde, V. & Dekker, N. H. The passage of homopolymeric RNA through small solid-state nanopores. *Small (Weinheim an der Bergstrasse, Germany)* **7**, 2217–24 (2011).
54. Skinner, G. M., van den Hout, M., Broekmans, O., Dekker, C. & Dekker, N. H. Distinguishing single- and double-stranded nucleic acid molecules using solid-state nanopores. *Nano letters* **9**, 2953–60 (2009).
55. Cracknell, J. A., Japrun, D. & Bayley, H. Translocating Kilobase RNA through the Staphylococcal  $\alpha$ -Hemolysin Nanopore. 1–17 (2013).
56. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13770–3 (1996).
57. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophysical journal* **77**, 3227–33 (1999).
58. Complexes, R. N. A. A. *et al.* Nanopore Analysis of Individual. 9345–9353 (2011).
59. DeBlois, R. W. Counting and Sizing of Submicron Particles by the Resistive Pulse Technique. *Review of Scientific Instruments* **41**, 909 (1970).

60. Huff, J., Pavlenok, M., Sukumaran, S. & Niederweis, M. Functions of the periplasmic loop of the porin MspA from *Mycobacterium smegmatis*. *The Journal of biological chemistry* **284**, 10223–31 (2009).
61. Faller, M., Niederweis, M. & Schulz, G. E. The structure of a mycobacterial outer-membrane channel. *Science (New York, N.Y.)* **303**, 1189–92 (2004).
62. Kang, X.-F., Gu, L.-Q., Cheley, S. & Bayley, H. Single protein pores containing molecular adapters at high temperatures. *Angewandte Chemie (International ed. in English)* **44**, 1495–9 (2005).
63. Franceschini, L. & Mikhailova, E. Nucleobase recognition at alkaline pH and apparent pKa of single DNA bases immobilised within a biological nanopore. *Chemical ...* 1520–1522 (2012). doi:10.1039/c1cc16124e
64. Japrun, D., Henricus, M., Li, Q., Maglia, G. & Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the alpha-hemolysin nanopore. *Biophysical journal* **98**, 1856–63 (2010).
65. Venkatesan, B. M. & Bashir, R. Nanopore sensors for nucleic acid analysis. *Nature nanotechnology* **6**, 615–24 (2011).
66. Meller, A., Nivon, L. & Branton, D. Voltage-Driven DNA Translocations through a Nanopore. *Physical Review Letters* **86**, 3435–3438 (2001).
67. Meller, A. & Branton, D. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis* **23**, 2583–91 (2002).
68. Kawano, R., Schibel, A. E. P., Cauley, C. & White, H. S. Controlling the translocation of single-stranded DNA through alpha-hemolysin ion channels using viscosity. *Langmuir: the ACS journal of surfaces and colloids* **25**, 1233–7 (2009).
69. Maglia, G., Restrepo, M. R., Mikhailova, E. & Bayley, H. Enhanced translocation of single DNA molecules through alpha-hemolysin nanopores by manipulation of internal charge. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19720–5 (2008).
70. Mathé, J., Arinstein, a, Rabin, Y. & Meller, a. Equilibrium and irreversible unzipping of DNA in a nanopore. *Europhysics Letters (EPL)* **73**, 128–134 (2006).
71. Mathé, J., Aksimentiev, A., Nelson, D. R., Schulten, K. & Meller, A. Orientation discrimination of single-stranded DNA inside the alpha-hemolysin membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12377–12382 (2005).
72. Gu, L. Q., Cheley, S. & Bayley, H. Prolonged residence time of a noncovalent molecular adapter, beta-cyclodextrin, within the lumen of mutant alpha-hemolysin pores. *The Journal of general physiology* **118**, 481–94 (2001).
73. Schneider, G. F. & Dekker, C. DNA sequencing with nanopores. *Nature biotechnology* **30**, 326–8 (2012).
74. Menestrina, G. *et al.* Ion channels and bacterial infection: the case of  $\beta$ -barrel pore-forming protein toxins of *Staphylococcus aureus*. *FEBS Letters* **552**, 54–60 (2003).
75. Gu, L.-Q., Cheley, S. & Bayley, H. Electroosmotic enhancement of the binding of a neutral molecule to a transmembrane pore. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 15498–503 (2003).
76. Lu, S., Li, W.-W., Rotem, D., Mikhailova, E. & Bayley, H. A primary hydrogen-deuterium isotope effect observed at the single-molecule level. *Nature chemistry* **2**, 921–8 (2010).
77. Ayub, M. & Bayley, H. Individual {RNA} base recognition in immobilized oligonucleotides using a protein nanopore. *Nano letters* **12**, 5637–43 (2012).

78. Montal, M. & Mueller, P. Formation of bimolecular membranes from lipid monolayers and a study of their electrical properties. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 3561–6 (1972).
79. Hansen, J. S. *et al.* Large scale biomimetic membrane arrays. *Analytical and bioanalytical chemistry* **395**, 719–27 (2009).
80. Bayley, H. *et al.* Droplet interface bilayers. *Molecular bioSystems* **4**, 1191–208 (2008).
81. Villar, G., Heron, A. J. & Bayley, H. Formation of droplet networks that function in aqueous environments. *Nature nanotechnology* **6**, 803–8 (2011).
82. Fischer, A., Holden, M. a, Pentelute, B. L. & Collier, R. J. Ultrasensitive detection of protein translocated through toxin pores in droplet-interface bilayers. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16577–81 (2011).
83. Bond, P., Guy, A., Heron, A., Bayley, H. & Khalid, S. Molecular Dynamics Simulations of DNA within a Nanopore: Arginine– Phosphate Tethering and a Binding/Sliding Mechanism for Translocation. *Biochemistry* **3777–3783** (2011). at <<http://pubs.acs.org/doi/abs/10.1021/bi101404n>>
84. Rincon-Restrepo, M., Mikhailova, E., Bayley, H. & Maglia, G. Controlled translocation of individual DNA molecules through protein nanopores with engineered molecular brakes. *Nano letters* **11**, 746–50 (2011).
85. Wong, C. T. A. & Muthukumar, M. Polymer translocation through  $\alpha$ -hemolysin pore with tunable polymer-pore electrostatic interaction. *The Journal of Chemical Physics* **133**, 045101 (2010).
86. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single {RNA} molecules. *Biophysical journal* **77**, 3227–33 (1999).
87. Renner, S., Bessonov, A., Gerland, U. & Simmel, F. C. Sequence-dependent unfolding kinetics of DNA hairpins studied by nanopore force spectroscopy. *Journal of physics. Condensed matter : an Institute of Physics journal* **22**, 454119 (2010).
88. Panwar, A. & Muthukumar, M. Enzyme-modulated DNA translocation through a nanopore. *Journal of the American Chemical ...* **18563–18570** (2009). at <<http://pubs.acs.org/doi/abs/10.1021/ja904047q>>
89. Sulewski, M., Marchese-ragona, S. P., Johnson, K. A. & Benkovic, S. J. Mechanism of Polynucleotide Phosphorylase ? 5855–5864 (1989).
90. Chen, H. *et al.* Ionic strength-dependent persistence lengths of single-stranded {RNA} and {DNA.}. **109**, 799–804 (2012).
91. Wong, O. Y., Mulcrone, A. E. & Silverman, S. K. DNA-catalyzed reductive amination. *Angewandte Chemie (International ed. in English)* **50**, 11679–84 (2011).
92. Wahl, M. C., Rao, S. T. & Sundaralingam, M. The structure of r(UUCGCG) has a 5'-UU-overhang exhibiting Hoogsteen-like trans U.U base pairs. *Nature structural biology* **3**, 24–31 (1996).
93. Wang, L. & Brown, S. J. BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic acids research* **34**, W243–8 (2006).
94. Achsel, T., Stark, H. & Lührmann, R. The Sm domain is an ancient RNA-binding motif with oligo(U) specificity. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 3685–9 (2001).

95. Achsel, T. *et al.* A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *The EMBO journal* **18**, 5789–802 (1999).
96. Törö, I. *et al.* RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *The EMBO journal* **20**, 2293–303 (2001).
97. Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. *The Journal of biological chemistry* **278**, 1239–47 (2003).
98. Moll, J. M., Sobti, M. & Mabbutt, B. C. The Lsm Proteins: Ring Architectures for RNA Capture. (2008).
99. Fischer, S. *et al.* The archaeal Lsm protein binds to small RNAs. *The Journal of biological chemistry* **285**, 34429–38 (2010).
100. Tang, W., Kannan, R., Blanchette, M. & Baumann, P. Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484**, 260–4 (2012).
101. Mura, C. & Randolph, P. S. Archaeal and eukaryotic homologs of Hfq. *RNA Biology* **10**, 636–651 (2013).
102. Gouaux, J. E. & Braha, O. Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proceedings of the ...* **91**, 12828–12831 (1994).
103. Hammerstein, A. F., Jayasinghe, L. & Bayley, H. Subunit dimers of alpha-hemolysin expand the engineering toolbox for protein nanopores. *The Journal of biological chemistry* **286**, 14324–34 (2011).
104. Tropini, C. & Marziali, A. Multi-nanopore force spectroscopy for DNA analysis. *Biophysical journal* **92**, 1632–7 (2007).
105. Keyser, U. F., van der Does, J., Dekker, C. & Dekker, N. H. Optical tweezers for force measurements on DNA in nanopores. *Review of Scientific Instruments* **77**, 105105 (2006).
106. Woodson, S. A. Compact intermediates in RNA folding. *Annual review of biophysics* **39**, 61–77 (2010).
107. Nanopore, S., Fologea, D., Uplinger, J., Thomas, B. & Mcnabb, D. S. Slowing DNA Translocation in a. 20–23 (2005).
108. Büssing, I., Slack, F. J. & Grosshans, H. let-7 microRNAs in development, stem cells and cancer. *Trends in molecular medicine* **14**, 400–9 (2008).
109. Tut, T. Z., Tut, Z., Thornton, J. E., Chang, H. & Piskounova, E. Lin28-mediated control of let-7 microRNA expression by alternative Lin28-mediated control of let-7 microRNA expression by alternative TUTases Zcchc11 (TUT4). **11**, (2012).
110. Rissland, O. S., Mikulasova, A. & Norbury, C. J. Efficient RNA polyuridylation by noncanonical poly(A) polymerases. *Molecular and cellular biology* **27**, 3612–24 (2007).
111. Carell, T. *et al.* Structure and function of noncanonical nucleobases. *Angewandte Chemie (International ed. in English)* **51**, 7110–31 (2012).
112. Wang, P., Ko, S. H., Tian, C., Hao, C. & Mao, C. RNA-DNA hybrid origami: folding of a long RNA single strand into complex nanostructures using short DNA helper strands. *Chemical communications (Cambridge, England)* **49**, 5462–4 (2013).
113. Shaw, N. N., Xi, H. & Arya, D. P. Molecular recognition of a DNA:RNA hybrid: sub-nanomolar binding by a neomycin-methidium conjugate. *Bioorganic & medicinal chemistry letters* **18**, 4142–5 (2008).

114. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. a. Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 3746–50 (1986).
115. Zhao, H., Yang, Y. & Zhou, Y. Prediction of RNA binding proteins comes of age from low resolution to high resolution. *Molecular bioSystems* 2417–2425 (2013). doi:10.1039/c3mb70167k
116. T. E. Oliphant, *Computing in Science & Engineering*, 9 (2007) 10-20
117. R. Bradshaw, S. Behnel, D. S. Seljebotn, G. Ewing, et al., *Computing in Science Engineering*, 13 (2011) 31 -39
118. S. Garcia, et al., <http://neo.readthedocs.org>



## CHAPTER 3

A step forwards the sequential recognition of RNA by custom RNA-binding nanopores.

## 3.1 Introduction

### 3.1.0 Overview

Proteins that bind RNA specific stretches of nucleotides in a modular one-to-one sequential interaction (i.e. a unique binding pocket for each nucleotide in a sequence) are attractive for many medical and biotechnological applications. In principle, once fused to any convenient effector domain, they could enable the detection and/or the functional modulation of the bound RNA. In the second chapter of this thesis we discovered a modular RNA binding property of WT- $\alpha$ HL and used the pore as a sensor for detecting a specific post-transcriptional RNA modification, i.e. uridylation. A co-crystal structure of the RNA bound in the pore would be necessary to understand the atomic details of the transient oligo U interaction within the WT- $\alpha$ HL  $\beta$ -barrel. Nevertheless, although it is possible that each  $\alpha$ HL monomer binds a single nucleotide, a sequential binding is unlikely and the geometrical constraints may play a role in the recognition. If we consider the *trans* aperture of the  $\beta$ -barrel as a circumference, three  $\alpha$ HL monomers span an arc that can be covered by 5 nucleotides in a stretched conformation ( $\sim 30$  Å). Therefore, in the oligo U<sub>7</sub> (UUUUUUU) it is hard to imagine that each nucleotide bind sequentially on each  $\alpha$ HL monomers inside the  $\beta$ -barrel. More realistic is the case where only 3-4 nucleotides of the single stranded RNA are interacting. This means that with the  $\alpha$ HL pore is possible to detect the oligo U signature in 3', its length and composition, but not a defined and sequential 5' to 3' series of nucleotide.

In this third chapter we tried to improve the RNA binding capability of the nanopore using modular RNA binding proteins that are known to bind ssRNAs in sequential one-to-one fashion. The improvements concerned: i) increasing the complexity of target sequences, ii) refining the accuracy of the sequence recognition and iii) tuning the binding kinetics.

We used two RNA binding proteins and the WT- $\alpha$ HL as building blocks to engineer custom RNA-binding nanopores (RBnPs) for a sequential nucleobase RNA binding, recognition and possibly, unique identification.

RBnPs were obtained fusing the monomer of the WT- $\alpha$ HL with the *Haloferax volcanii*-Lsm (*HV*-Lsm) protein or with Pumilio (PUF) domains (see Appendix for the second one). Each monomer of the heptameric RBnP was designed to bind a single nucleotide of the ssRNA.

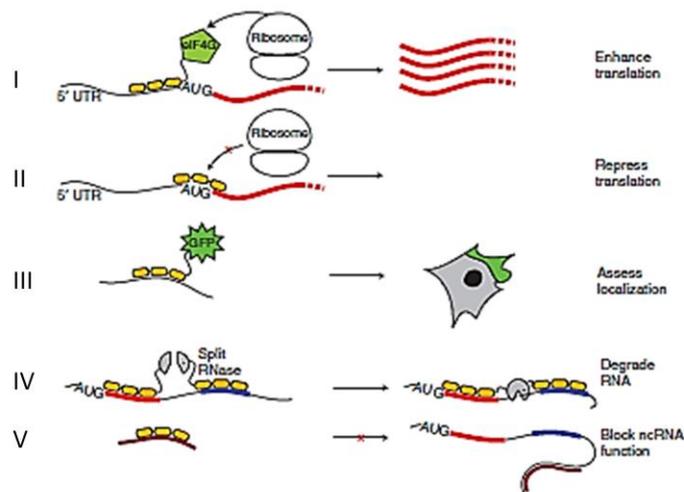
For the fusion construct with the *HV*-Lsm, we aimed to increase the binding affinity of the WT- $\alpha$ HL for oligo uridylated ssRNAs. This work not only describes the implementation of the WT- $\alpha$ HL as a oligo U sensor, but it is a step forward to the design of custom RBnPs for stochastic sensing of RNA sequences by modular and sequential base recognition, in a one-to-one protein-nucleobase interaction. Remarkably, the chimeric protein described in this chapter is the first example of functional active fusion protein based on the  $\alpha$ HL scaffold.

### **3.1.1 A new challenge: RNA detection by modular recognition.**

RNA-binding proteins (RBPs) regulate numerous aspects of co- and post-transcriptional gene expression, including RNA splicing, polyadenylation, polyuridylation, capping, modification, and turnover. Sequence-specific associations between RBPs and their target RNA are typically mediated by one or more RNA-binding domains (RBDs) that are able to recognize specific evolutionary conserved RNA sequence signatures<sup>1</sup>. An interesting review of Makay J.P (2011)<sup>2</sup> described new perspectives in functional design of RBPs and RNA binding domains (RBD). The possibility to fuse these binding proteins to effector domains with different functions (e.g. ribonuclease, ribosomal proteins, initiator factors) in order to modulate RNAs (e.g. mRNA, miRNA) splicing, localization, degradation, and finally the expression is an attractive challenge (Figure 3.1). For these purposes, the K-homology domain<sup>3</sup>, the pentatricopeptide repeat<sup>4</sup>, tristetraprolin motives<sup>5</sup>, Lsm proteins<sup>6</sup> and Pumilio repeats<sup>7</sup> are all good candidates. The

use of RBPs with tunable activities became possible thanks to the recent discoveries on RBPs structures bounded with their respective RNAs<sup>8,9</sup>. Moreover, after the discovery of RecA and TALEs<sup>10,11</sup> DNA binding proteins and their application for DNA editing, the field of custom made nucleic acid binding proteins received a lot of attention.

RBPs and RBDs physiologically bind specific ssRNA or ssDNA sequence signatures in a modular fashion. Some of them bind the nucleic acid sequence using a single domain for multiple nucleotides. Other proteins (like Lsm proteins or proteins bearing PUF domains) bind a single nucleotide with a single domain in a one-to-one interaction (Table 3.1).



**Figure 3.1** Examples of fused RNA-binding proteins or domains. (adapted from Mackay J.P at al., 2011)<sup>2</sup>; **(I)** Driving translation with RBDs fused to the eukaryotic translation initiation factor eIF4G; **(II)** inhibition of translation by RBDs bound to the start codon; **(III)** RBDs fused to a fluorescent protein (such as GFP) and used to track RNA in living cells; **(IV)** RBDs fused to a nonspecific RNase could allow the degradation of a specific target RNA; **(V)** RBDs bound to a specific noncoding RNA (ncRNA) to block its activity.

Protein	Target	Organism	Dimension of each module (aa)	Number of modules	Residues forming the binding pocket-position	Unique code of recognition for each base?
TALE	dsDNA	bacterial.	34 or more	from 6 to 20	2-12,13	YES
PPR	ssRNA	eukaryotes	35 or more	from 2 to 26	3-1,4,34	Not known
Lsm	ssRNA	bacterial, archae, eukaryotes	76 (>70)	from 6 to 8	3-37,39,63	Not Known (only U)
PUF	ssRNA	eukaryotes	36 ore more	8	3-12,13,16	YES

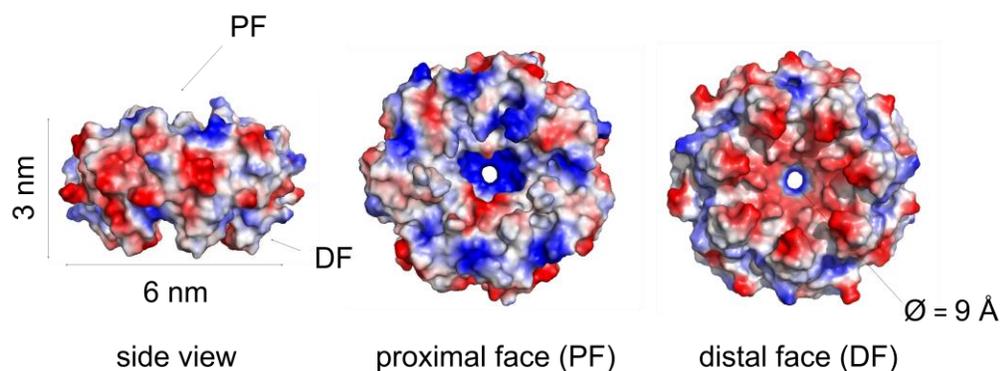
**Table 3.1 Proteins that bind ssDNA or ssRNA in a modular fashion.** TALE: TALEs Binding protein<sup>11</sup>; PPR: pentatricopeptide repeat<sup>4</sup>, PUF: Pumilio domains<sup>12</sup>, Lsm: anti-Sm like antigen<sup>13</sup>.

### 3.1.2 A doughnut shape for RNA capture: Lsm proteins

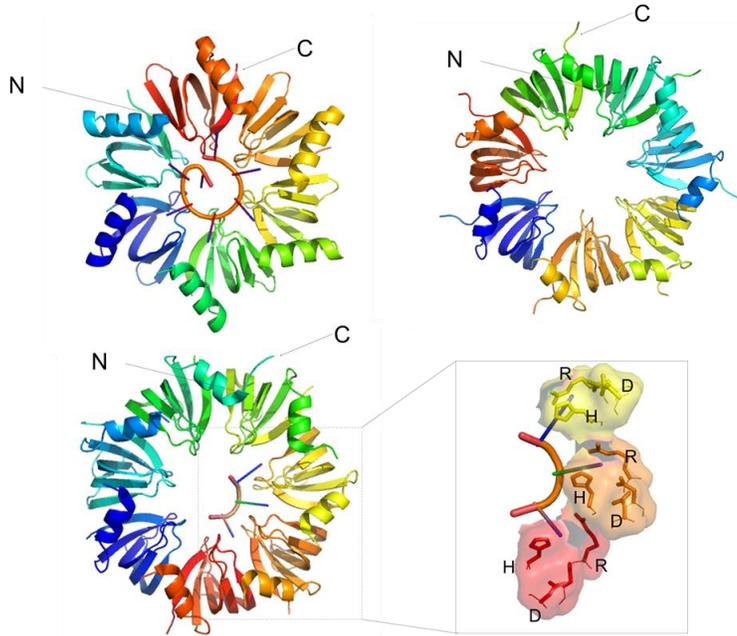
So far, Sm-like proteins (Lsm) is a protein family described across all the domains of life<sup>14,15</sup>. It is widely studied with 1,981 hits in the NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>). This family include the eukaryotic Lsm, Sm<sup>16,17</sup>, the archeal Lsm<sup>18,19</sup> and the bacterial Hfq<sup>20,21</sup>. The eukaryotic Lsm proteins are divided into Lsm and Sm. The differentiation is based on the *small nuclear ribonucleo-protein* (snRNP) complexes in which the protein is involved. Up to 16 different Lsm and 7 Sm proteins are encoded in the eukaryotic genome, 3 Lsm are found in the archaea genome and only one in prokaryotes. Lsm are proteins with pleiotropic functions that mainly act as RNA chaperones during RNA processing. Lsm are involved in splicing<sup>22</sup>, histone maturation<sup>23</sup> and telomerase maintenance<sup>24</sup>. It is interesting to note that in some cases the Lsm gene is co-transcribed with ribosomal protein coding genes<sup>25</sup>. Deletion of the *lsm* genes always revealed a pleiotropic deficient phenotype, meaning that these proteins are involved in different cellular functions. Bacterial, archaea and their eukaryotic Lsm orthologs share limited amino acid sequence similarity (< 20 % between archaea and eukaryotes), and <50% among parental proteins (< 50% in archaea), but all Sm-like proteins share a high similarity of tertiary and quaternary structures. A bipartite consensus

region (Sm1 and Sm2), where the RNA-binding pockets are placed<sup>14</sup>, can be identified in the amino acid sequence .

Individual Lsm vary in size from 8 to 25 kDa (76-240 amino acids). In the Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) 23 Lsm structures are listed (17 X-ray, 6 NMR): fourteen from eukaryote, three from archaea and two from bacteria. A characteristic feature of Lsm proteins is their toroid-shaped structure (or doughnut-like) of the oligomers. In prokaryotes and archaea, functional oligomeric complexes are formed by six or seven homomeric units. Heteromeric assemblies of distinct Lsm proteins are found in eukaryotes. The Lsm doughnut diameter has been observed to range between 58 and 75 Å with a central pore of 6-15 Å<sup>26,27</sup> (Figure 3.2). Each monomer presents a five-stranded antiparallel β-sheet with the C-terminal and the N-terminal on opposite side of the doughnut (Figure 3.3). The two faces of the doughnut are called the distal and proximal face (this last one corresponds to the N-terminal alpha helix) (Figure 3.2 and 3.3). The two conserved regions (called Sm1 and Sm2) are part of loop L3 and loop L5. The sequence motives Asp-x-His-x-Asn and Arg-Gly-Asp (where x stands for any amino acid) in Sm1 and Sm2 respectively, are the highly conserved residues that form the most well characterized RNA binding pocket<sup>26</sup>.



**Figure 3.2 Views of an archaea Lsm.** Surface charge distribution generated using PyMOL in vacuum electrostatic mode. Blue: positive charges residues, Red: negative charges residues. Structures from PDB 1I81.



**Figure 3.3. Examples of resolved Lsm structures.** a, prokaryotic Lsm PDB 1KQ2; b, eukaryotic Lsm PDB 4EMG; c, archaeal Lsm PDB 1I81. Outset: zoom on the one-to-one interaction of three uridines with three monomers; main residues are labeled. N: N-terminal. C: C-terminal.

Biochemical and structural studies concerning the Lsm-RNA interaction show distinct RNA binding sites within the oligomer: (i) in the lumen of the ring, (ii) close to the helix face (i.e. the proximal) and (iii) on the distal face<sup>26</sup>. The binding site within the lumen (formed by the amino acid sequence described above) is characteristic across Eukarya and Archaeae and is the most selective for oligo U. The loops L3 and L5 create a modular binding pocket running around the central pore of the doughnut<sup>28</sup>. The crystal structures of the archaeal Lsm-RNA complex clearly shows the RNA threaded around the inner rim of the complex<sup>19,29,19</sup>. Each binding pocket allows specific U base stacking of a simple oligo U or a consensus sequence RAU<sub>4-6</sub>GR (where R stands for a purine)<sup>19</sup>. Each monomer binds a single nucleotide. More than four U are required for Lsm binding. According with electron-density data it is unlikely that more than three uridine residues bind at the same time the internal “slot” of the binding pockets.

Two types of Lsm are found in archaea. The monomeric protein can assemble around the RNA strand (Lsm type 2) or the RNA can be threaded through the pre-formed doughnut complex from the helical face (Lsm type 1). In both cases the RNA-protein interaction is static because (i) no conformational change of the protein after RNA binding has been reported and (ii) the oligomer have none processive activity (it does not move) on the RNA strand. All the Lsm ring structures reveal cluster of positive residues lining the internal pore<sup>29,30</sup>.

### **3.1.3 The archaea *Haloferox volcanii* Lsm.**

*Haloferox volcanii* belongs to the aerobic and mesophilic species of the *Halobacteriaceae* family, a group within the phylum Euryarchaeota of the *Archaea* domain. This organism is commonly found in high-salinity aquatic environment. The complete genome was sequenced by Hartman et al. (2010)<sup>31</sup> and contains a single *lsm* gene, which encodes a protein of 76 amino acids with a molecular mass of 8.25 kDa and an isoelectric point of 3.9. The *lsm* gene overlaps by four nucleotides with a gene annotated to encode for the L37e ribosomal protein. Fisher and coworkers (2010)<sup>25</sup> demonstrated the binding of the recombinant Lsm protein with oligo U RNAs and tRNAs<sup>25</sup>. The *Haloferox* Lsm protein was found to belong to the Lsm 1 sub-family. At the moment no monomeric or oligomeric crystal structure are available.

## 3.2 Results and discussion.

Protein tools that can bind and detect in a qualitative and quantitative manner any RNA sequence of interest have many potential biotech and medical applications. In this chapter we design chimera proteins for selective detection of specific RNA sequences. After the discovery of a slot of RNA-binding pockets in the *trans*-side of the  $\alpha$ HL  $\beta$ -barrel (see Chapter 2) we genetically fused the *Haloferax volcanii*-Lsm RNA-binding (*HV*-Lsm) protein in C-terminal of the  $\alpha$ HL. Complementary to this, in the appendix of this chapter, two Pumilio domains were fused in C-terminal of the same pore forming toxin. The final aim of these two approaches was to improve the RNA binding of the nanopore.

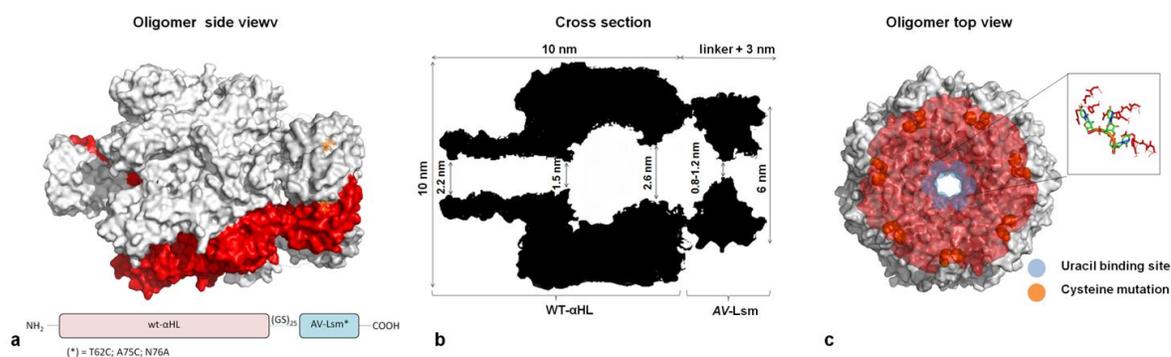
For the fusion construct with the *HV*-Lsm we planned a rational approach based on: 1) design and expression of the protein; 2) test the RNA binding on short and unstructured oligonucleotides in bulk solution (biochemical evidence); 3) test of the RNA binding on short and unstructured oligonucleotides in PLM in single channel (electrophysiological evidence); 4) test of the RNA binding in PLM, in single channel, on complex RNAs.

### 3.2.1 Design and expression.

Both Lsm and  $\alpha$ HL have a heptameric oligomeric structure. Therefore, taking advantage of this structural feature, we designed the RNA-binding nanopore assuming that the Lsm doughnut could follow the stoichiometry of the  $\alpha$ HL: when the toxin oligomerizes and inserts onto the lipid bilayer the Lsm will lodge on the cap of the  $\alpha$ HL (Figure 3.4a, and 3.4b and 3.4c).

We designed two fusion proteins. We obtained the first chimeric RNA binding nanopore (RBnP) (called LadyA) fusing at the C-term of the wild-type  $\alpha$ HL the wild-type *HV*-Lsm proteins. The two portions were connected with a flexible Serin-Glycin (SG) linker ( $\alpha$ HL-(SG)<sub>10</sub>-*HV*-Lsm). In the second RBnP (called LadyB) the WT- $\alpha$ HL was fused (Figure 3.42.1) with a longer SG linker to a mutated *HV*-Lsm protein ( $\alpha$ HL-(SG)<sub>25</sub>-*HV*-Lsm T62C/A75C/N76A) (Figure 3.4a, see Methods). For LadyB, the cysteine residues

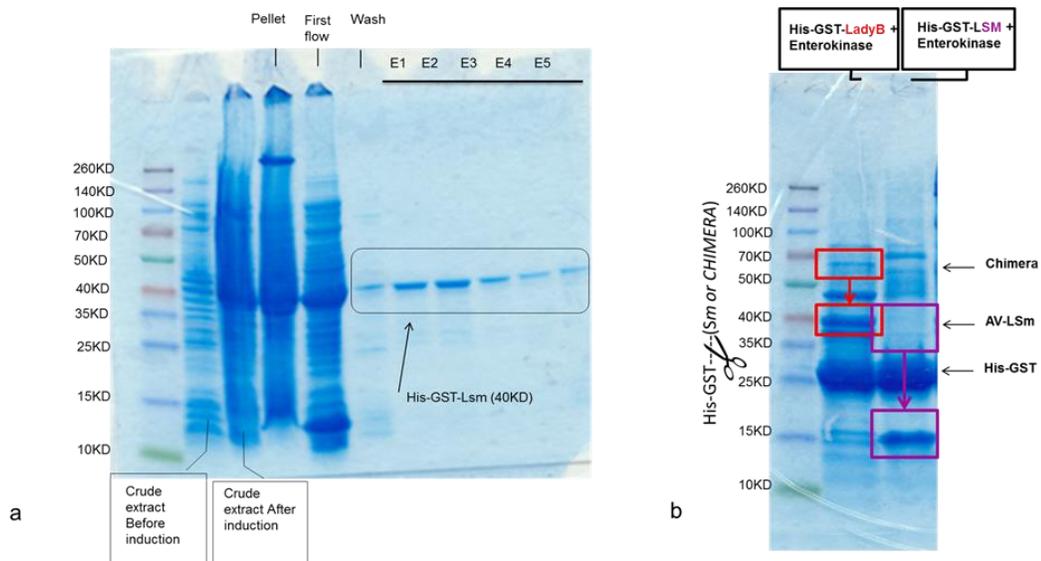
were placed at each side of the Lsm monomer to induce the formation of disulfide bonds between them and therefore improve the stability of the chimeric protein in electrical recording. The distance between the two cysteine residues on adjacent monomers was  $\sim 5.8 \text{ \AA}$  (based on PDB structure 1H64). The C75 was placed on the flexible C-terminal region of the Lsm (Figure 3.4c). In order to enhance the formation of disulfide bridges, the chimeric oligomers were oxidized in *low ionic strength buffer* (150 mM KCl, 100 mM NaCl, 2 mM MgCl<sub>2</sub> and 10 mM HEPES, pH 6.5 in DMPC-treated water) with 1.5 mM of Cu(o-phenanthroline)<sub>2</sub> (see Methods).



**Figure 3.4. LadyB design.** **a**, Side view of the oligomeric WT- $\alpha$ HL and HV-Lsm fused through a flexible (SG)<sub>25</sub> linker. A fusion monomer is highlighted in red. PDB structures were assembled by PyMol. **b**, Cross-section and dimensions of the RBnP. **c**, Top view of the RBnP with the RNA-binding pockets and cysteine highlighted in blue and orange respectively. Zoom: three RNA binding pockets and relative residues (red) forming the binding site for a single uracil (green). (PDBs file 1H64 from *Pyrococcus abyssi* for the HV-Lsm and PDB 7AHL for the WT- $\alpha$ HL oligomer.).

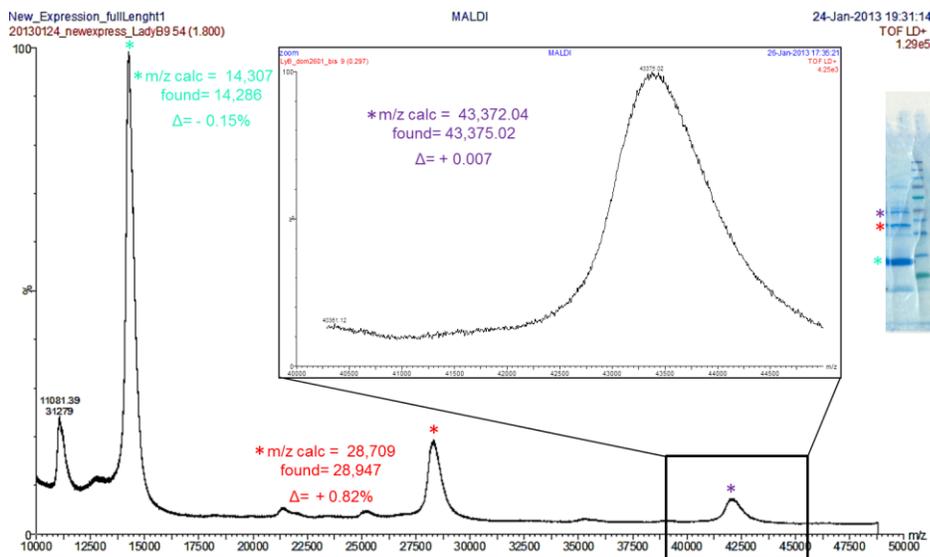
Both RBnPs (LadyA and LadyB) and the control HV-Lsm were expressed both *in vitro* and *in vivo*. The *in vivo* expression of the fusion constructs (in pGST-21a plasmid) (Figure 3.5) were carried out in E.coli as described in Methods. Proteins were expressed with a His-GST tag in N-terminal, in order to allow the purification by cobalt resins and to increase the solubility of the protein. This is mainly relevant for fusion construct with the  $\alpha$ HL. In fact, the pore forming toxin was known to be compartmentalized in the inclusion bodies during expression. As reported in figure 3.5, all the proteins expressed *in vivo* were mainly present in the soluble fractions and only part of the proteins stayed in

the inclusion bodies. This allowed a rapid purification without the use of denaturants. The His-GST tag was then removed by enterokinase digestion (Figure 3.5b)



**Figure 3.5 *In vivo* protein expression.** **a**, Coomassie staining of fusion proteins (His)<sub>6</sub>-GST HV-Lsm – a and b - and LadyB – b - expressed in E.Coli-NICO<sub>21</sub> (NEB). The supernatant was load in a gravity column (Econo-Column, Biorad) and the protein purified by cobalt resin affinity chromatography.. LadyA and LadyB were expressed in the same way as LadyB. **bc**, SDS-PAGE after enterokinase digestion. The digestion was performed to remove the His-GST tag and was carried out overnight at 22°C.

Only in the case of LadyB, the expression of the full length protein was confirmed by MALDI-TOF (Figure 3.6). We used the lysozyme and the residual His-GST tag as internal standard. The calculated m/z was found to be as expected ( $\Delta\% = 0.007$ ). Moreover we try to define the MW also by trypsin digestion of the PAGE extracted monomeric protein and ESI-MS. We reach 86% of sequence coverage, but some trypsin digested fragment in C-terminal could not be detected in ESI-MS at positive mode. The reason of this could be the presence of many negative charged amino acid residues in the C-terminal of the Lsm sequence, where also the two cysteine residues are placed. Moreover, we cannot exclude that some expressed protein were not full-length due to the proteolysis during the optimization of the expression conditions.



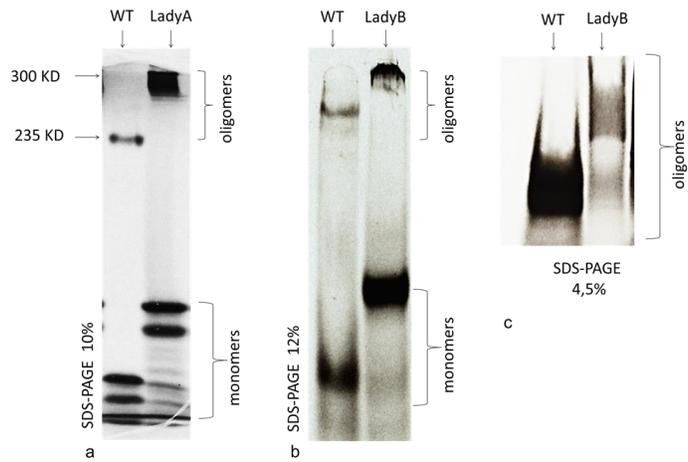
**Figure 3.6 MALDI-MS of LadyB monomers.** Resin purified protein at 1 mg/ml was mixed in 1:1 ratio with sinapinic acid (see Methods). MALDI analysis was conducted on a Water MALDI Micro MX spectrometer with TOF detection, in positive linear mode. **Left:** SDS-PAGE showing multiple bands after purification/digestion: lysozyme residues (green), His-GST tag residues (red) and LadyB (violet). Green and red were used as internal standard.

**3.2.2 Oligomerization and activity.** To understand if the presence of the RBP portion affected the overall pore-forming activity, we characterized the pore forming ability of the chimera. First, we tested the oligomerization efficiency on RRBCM; second, we determined the hemolytic activity in comparison to the WT- $\alpha$ HL.

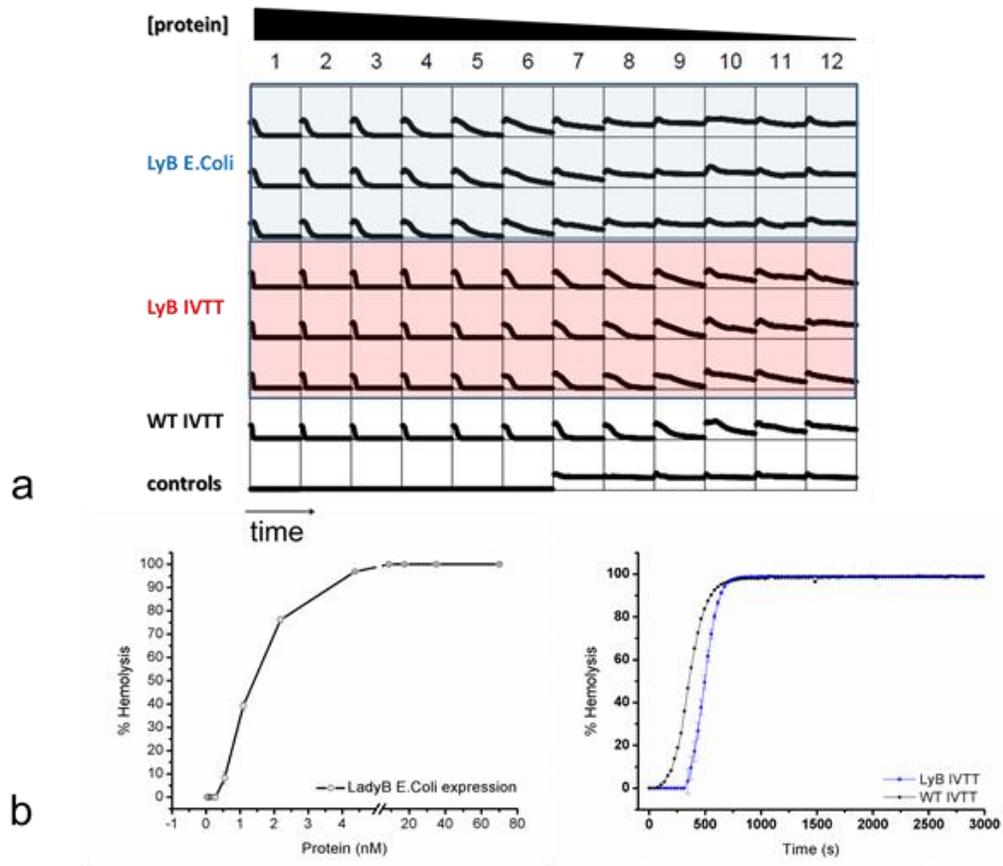
The oligomerization abilities of the two chimera proteins respect to the WT- $\alpha$ HL pore, were compared by SDS-PAGE gel. As reported in figure 3.7 both the chimeric proteins were able to oligomerize and to form SDS-stable pores.

The hemolytic assay on rabbit red blood cells we did not show any remarkable difference in both hemolytic activity and hemolytic kinetics between the WT and LadyB (Figures 3.8).

In conclusion, the fusion constructs LadyA and LadyB showed a comparable oligomerization efficiency and lytic activity to the WT- $\alpha$ HL. This implies that the Lsm structure on the  $\alpha$ HL cap did not affect the pore forming function of the scaffold..



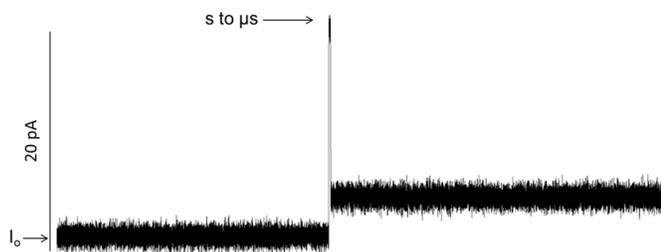
**Figure 3.7 SDS-PAGE of *in vitro* expressed chimeric proteins.** **a**, Monomers and oligomers of WT- $\alpha$ HL and LadyA and LadyB - **b** - in SDS-PAGE gel at 12% and **c**, 4.5%. LadyA: WT- $\alpha$ HL-(SG)<sub>10</sub>-wt-AVLsm. LadyB: WT- $\alpha$ HL-(SG)<sub>25</sub>-AVLsmT62C,A75C,N76A. Proteins expression *in vitro* were prepared by coupling *in vitro* transcription and translation (IVTT) using an E. coli T7-S30 expression. Chimera and WT- $\alpha$ HL oligomers were formed on rabbit red blood cells membranes and purified before loading on the gel.



**Figure 3.8 Hemolytic assay.** **a**, Hemolytic kinetics and activity comparison between LadyB (expressed *in vitro* and *in vivo*) and WT- $\alpha$ HL (*in vitro* expressed). **b**, Hemolytic activity (left panel) of the Lady B and the kinetic of hemolysis (right panel), in blue the LadyB and in black the WT.

**3.2.3 Electrophysiological characteristics of LadyA and LadyB.** In theory, any modification introduced in the overall pore organization can modify the electrical stability and electrophysiological characteristics of the pore once it is inserted into a lipid bilayer. To understand the feasibility of using these chimeric nanopores for RNA sensing, we tested their electrophysiological properties in planar lipid bilayers (PLM).

Both RBnPs were able to open channels in 1,2-diphytanoyl-sn-glycero-3-phosphocholine bilayer (Figure 3.8 and 3.9). LadyA was not stable and was found to be permanently blocked at applied potential higher than +60 mV (Figure 3.8).

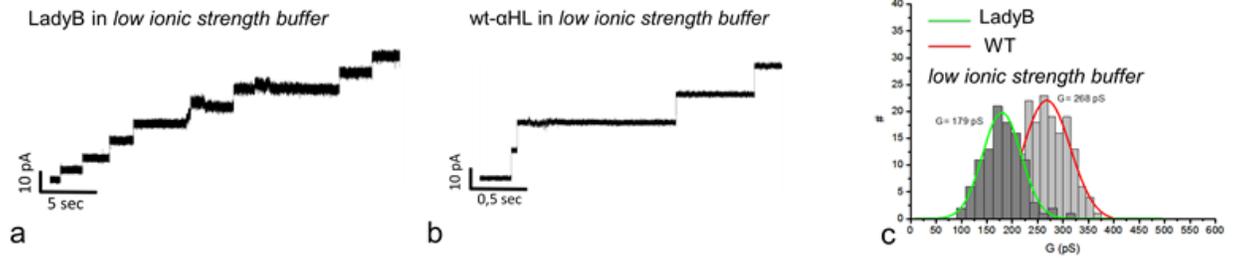


**Figure 3.9. Insertion of LadyA oligomers and permanent block of the channel after insertion.** The permanent block occurs after few second/hundreds of microseconds after insertion. A 10 s trace is reported. Data were acquired in *low ionic strength buffer*, acquired at 20 kHz and filtered at 5 kHz.

Once inserted into the bilayer, the LadyB nanopore was stable at applied potential  $< +120$  mV, but it showed permanent closures and gating at high voltages ( $> +120$  mV) of either polarity. LadyB was in general more stable than the version A. For these reasons, in all experiments in single channel recording we worked only with LadyB, never using applied potential above  $+100$  mV.

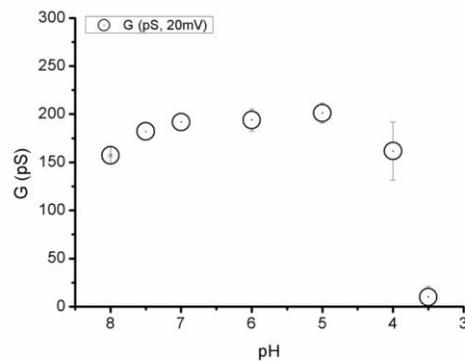
In order to describe the electrophysiological properties of the fusion LadyB protein we compared the conductance (G), the rectification factor of the I/V curve and the ion selectivity of the channel with respect to the WT- $\alpha$ HL (Figure 3.10, 3.12a and 3.12b, respectively).

LadyB presented a slight rectification at negative potentials ( $I_+/I_- = 1.07 \pm 0.18$ ,  $n=15$ ) (Figure 3.12ab) and a mean unitary conductance of  $179 \pm 75$  pS ( $n=150$ ) lower than the WT- $\alpha$ HL ( $268 \pm 90$  pS,  $n=200$ ) (Figure 3.10c). This result can suggest a different dimension (LadyB smaller) of the two pores, although the conductance can be affected by other parameters such as the charge distribution in the channel.



**Figure 3.10 Pore insertion in PLM and conductance distribution of LadyB and WT- $\alpha$ HL.** **a,c**, Insertion of LadyB oligomers in *low ionic strength buffer*. **b**, channels formation by WT- $\alpha$ HL in *low ionic strength buffer* **d**, single pore conductance distribution of LadyB vs WT- $\alpha$ HL. Recording at +80mV, data acquired in *low ionic strength buffer*. The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter) in ‘a’ and at 1 kHz in **b, c**.

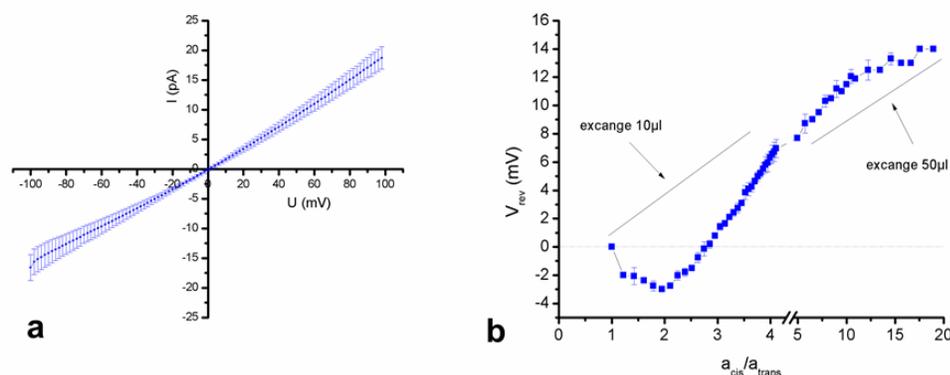
All experiments were performed in *low ionic strength buffer*. Higher salt concentration affects negatively the stability of the pore (data not shown). Moreover, LadyB was stable (i.e. no block of the channel) between pH 5 and 8 (Figure 3.11). The pH, salt concentration and voltage were all critical parameters for the stability in PLM (Figure 3.9, 3.11 and 3.12).



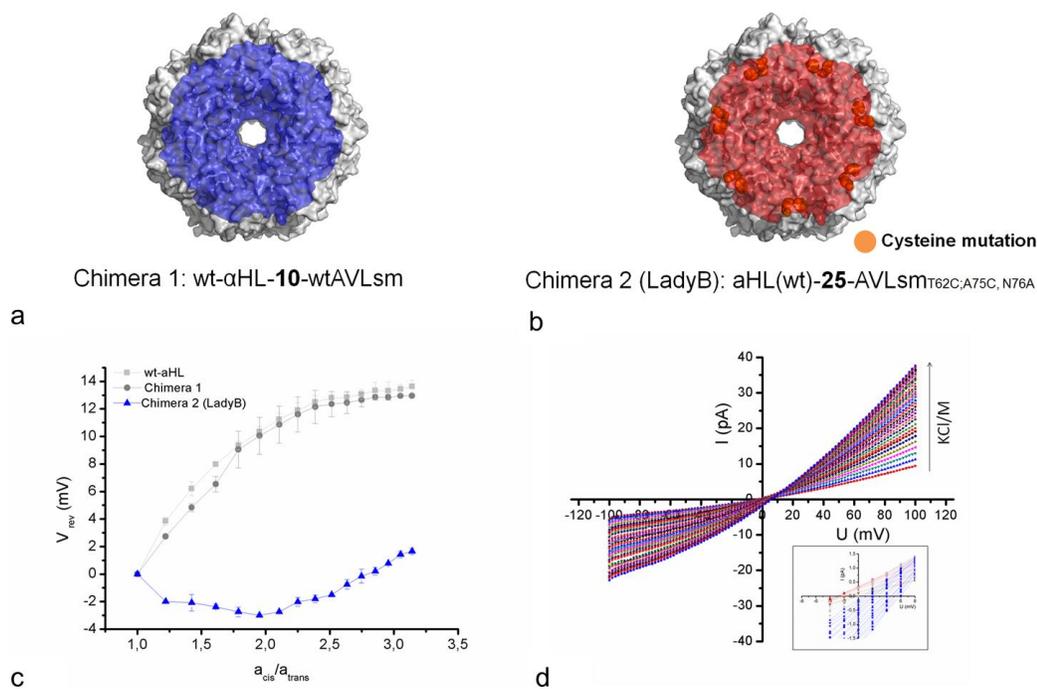
**Figure 3.11 Mean conductance of a single-channel in PLM at different pH. G: conductance.** The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.

To characterize the ion selectivity, the reversal voltage ( $V_{rev}$ ) was measured, and the permeability ratio ( $P^{k+}/P^{Cl-}$ ) was calculated using the Goldman-Hodgkin-Katz (GHK) equation<sup>32</sup> (see Methods). Similarly to the WT- $\alpha$ HL, which was found, as expected, to be slightly anion selective<sup>33</sup>, LadyB showed an overall anionic selectivity at high ionic asymmetric conditions (Figure 3.12b), but presented a characteristic behavior at  $a_{cis}/a_{trans}$

values lower than 4 (Figure 3.12b and 3.13); where the  $V_{rev}$  potential had opposite sign (negative  $V_{rev}$ ), meaning cationic selectivity. The permeability ratio  $P^+/P^-$  at  $[a]_c/[a]_t=2$  and at  $[a]_c/[a]_t=4$  were  $0.64 \pm 0.09$  and  $0.70 \pm 0.03$  respectively (Figure 3.12). From a comparison of the  $V_{rev}$  curves of the WT- $\alpha$ HL with the LadyB and LadyA we observed a peculiar behavior of the LadyB at  $a]_c/[a]_t < 4$  (Figure 3.13c and 3.13d)). In higher ionic asymmetric conditions the behavior follows the WT- $\alpha$ HL pore and, on the other hand, LadyA showed the same behavior of the WT- $\alpha$ HL pore (Figure 3.13c)



**Figure 3.12** Electrophysiological characterization of LadyB. **a**, IV curve in low ionic strength buffer (n=15). **b**, Reverse potential versus KCl activity ratio cis/trans (n=3). The  $V_{rev}$  was measured exchanging 10 or 50  $\mu$ L of buffer. The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter).



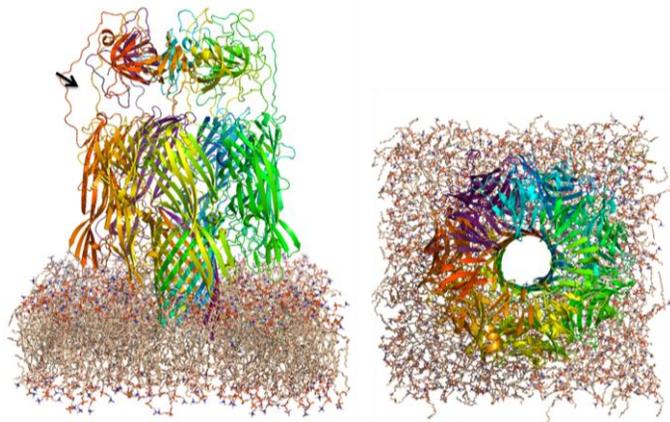
**Figure 3.13 LadyA vs ladyB: effect of linker length and the mutations T62C, A75C, N76A on the  $V_{\text{rev}}$ .** **a and b**, top view of the two fusion constructs. **c**,  $V_{\text{REV}}$  of LadyB in a short window of asymmetric KCl concentration ( $[a]_c/[a]_t < 3$ ) compared with the WT- $\alpha$ HL and the chimera 1. **d**, Shape of the current-voltage (IV) at increasing KCl concentration in *cis*. Inset: zoom in on the IV curve around the zero mV. Red lines show negative reverse potentials and blue lines positive reverse potentials. Electrophysiology experiments on  $V_{\text{REV}}$  were performed only with fusion proteins expressed *in vivo*. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.

In conclusion, the conductance of the pore, the ion selectivity and the shape of the I/V curve were slightly different from the WT- $\alpha$ HL, indicating that the Lsm doughnut partially affects the biophysical properties of the channel (see Table 3.3 for an overview of the differences between the WT- $\alpha$ HL and the LadyB).

The narrowest point in the WT- $\alpha$ HL is the constriction at the entrance of the  $\beta$ -barrel (see chapter 2). In this chimeric RBnPs, assuming that the oligomerization of the fusion protein allows the Lsm heptamer to correctly oligomerize, the main pore constriction is given by the doughnut's hole, where the RNA binding pockets are placed. According to the X-ray crystal structure of the parental archaee proteins, the Lsm constriction should be  $\sim 1.2$  nm. The lower mean conductance of the LadyB respect to the WT- $\alpha$ HL pore

could be a rough indication (see above) that the channel is indeed narrower than the WT- $\alpha$ HL. The Lsm constriction should be the region where the all the voltage will drop down if the all ions will pass through. This is true if no “secondary leaking sides” are present in the nanopore. We performed molecular modeling simulation (data not showed) and showed that 90% of the ions do not pass through the doughnut’s hole but on its side, at the interface of two oligomers as can be observed from the model of the fusion construct (Figure 3.14). This result allowed us to hypothesize that the Lsm can partially affect the channel’s electrical properties if the SG linker allows the doughnut to move away from the  $\alpha$ HL cap and the majority of ions escape from the hole of the Lsm.

Taking into account all the computational and experimental results we hypothesized that the chimeric nanopore can be “leaking” between the two fused proteins. This conclusion is in agreement with the results on RNA binding ( $K_D$ ) presented below; but so far, this hypothesis needs further experiments to be proved.

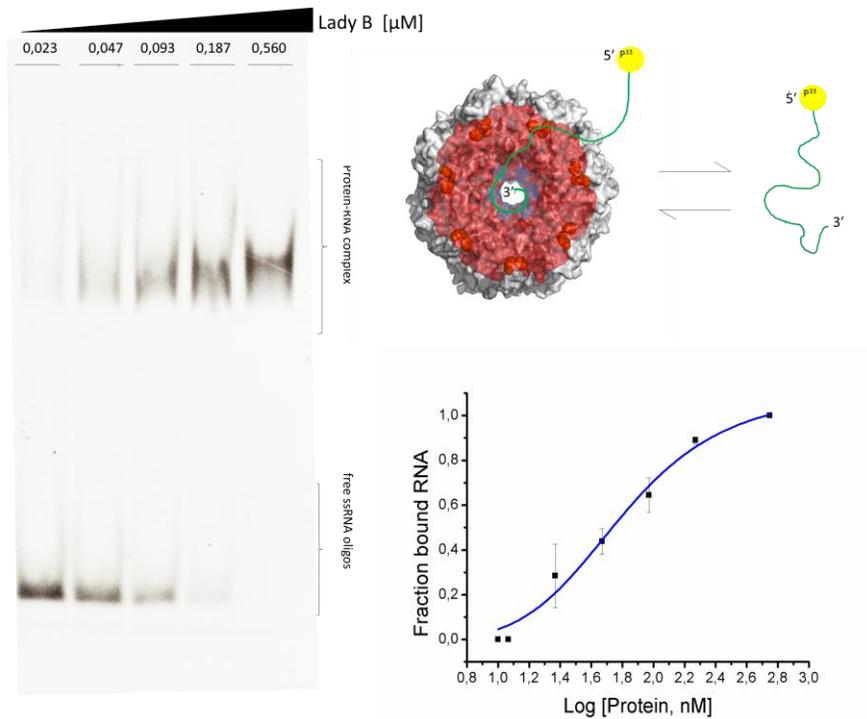


**Figure 3.14 LadyB modeling.** Structures of the Lady-B heptamers were derived by homology modelling using the MODELLER suite (version 9.10) based on the WT- $\alpha$ HL crystal structure (pdb: 7AHL) connected to the RNA binding protein (pdb: 1I81) by 25-amino acid long SG linker (thanks to Dr. Lajos Höfler for the modeling). Black arrow: leaking site between the two fused proteins.

**3.2.4 RNA binding: biochemical evidence.** As for the WT- $\alpha$ HL, we performed the electromobility shift assay (EMSA) to observe the RNA binding in bulk solution. The EMSA assay is a common affinity electrophoresis technique used to study protein-DNA or protein-RNA interactions. The speed at which different molecules (and any combinations thereof) move through the gel is determined by their size and charge, and to a lesser extent, their shape. Under the correct experimental conditions (see Methods), the interaction between the RNA and the protein is stabilized and the ratio of bound to unbound nucleic acid on the gel reflects the fraction of free and bound probe molecules as the binding reaction enters the gel.

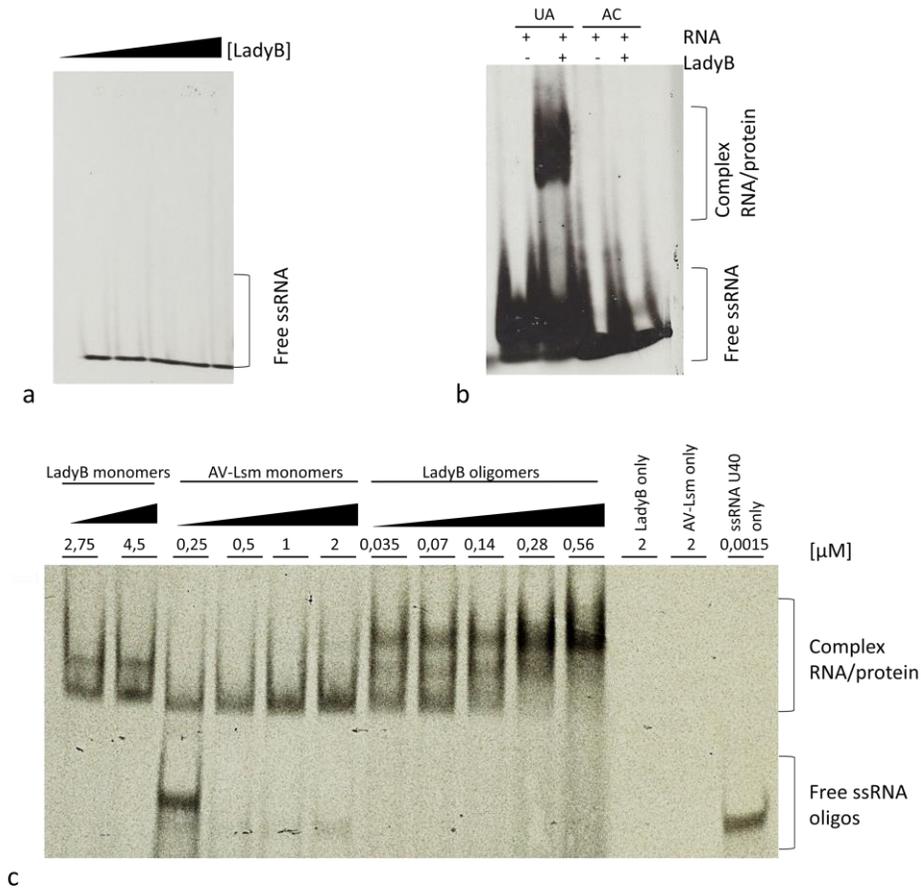
For this EMSA assay, the chimeric LadyB pore was incubated with 50 pmol of oligo U<sub>20</sub> radioactive labeled at the 5'-end with P<sub>33</sub> (see Methods).

We observe a shift of the radioactive RNA bands, meaning a binding of the pore to the ssRNA (Figure 3.15 - left panel). From a densitometric analysis of the bands the dissociation constant ( $K_D$ ) for LadyB on the U<sub>20</sub> RNA (Figure 3.15 right - bottom) was deduced to be 63 nM, in agreement with published data ( $K_D$  72 nM)<sup>25</sup> and our control on the *HV*-Lsm.



**Figure 3.15 Determination of the  $K_D$  of binding in bulk.** Left - Electromobility shift assay at increasing concentration of recombinant pre-oligomerized protein incubated with 50 pmol radioactive ( $P^{33}$ ) oligo  $U_{20}$  RNA and subsequently loaded onto a non-denaturing 0.5X TBE-PAGE gel. **Right top** - RNA/protein complex and free RNAs are shown schematically. **Right bottom** - Densitometry bands analysis is reported,  $n=3$ . The normalized fraction of RNA bound is plotted over the protein concentration in a logarithmic scale. A Hill function ( $y=V_{max} \cdot x^n / (k^n + x^n)$ ,  $V_{max}=1,11$ ;  $K=1,8$ ;  $n=5,32$ ) has been fitted.

With the EMSA assay we also proved the binding of  $U_{40}$  and the selectivity between oligonucleotides with or without the 3' U tail. LadyB was unable to bind  $A_{10}C_{10}$  but it still interacted with  $U_{17}A_3$  and the longer RNA  $U_{40}$  (Figure 3.16a and 3.16b). Finally, our EMSA control on the HV-Lsm (Figure 3.16c) reported a  $K_D$  of  $\sim 40$  nM in agreement with published results<sup>25</sup>. All these results suggested that the Lsm did not lose the RNA binding ability when fused to  $\alpha$ HL.



**Figure 3.16 EMSA assays.** Progressive protein dilution were incubated with 50 pmol of <sup>33</sup>P-labeled RNAs in a 40 μl reaction for 15 min at 30°C in a binding buffer containing 100 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 12 mM HEPES/KOH, 10% glycerol, 0.1% Triton X-100 and RNase inhibitor, pH 7.6. Starting concentrations of 2 μM for recombinant/purified AV-Lsm monomers (equal to 0.28 μM of Lsm homoheptameric complexes) and 0.56 μM of chimera oligomers were used (assuming 50% oligomerization efficiency) **a**, incubation of LadyB with A<sub>10</sub>C<sub>10</sub> RNA. **b**, incubation of LadyB with oligo U<sub>17</sub>A<sub>3</sub> and A<sub>10</sub>C<sub>10</sub> RNA. **c**, incubation of increasing concentration of LadyB monomers, AV-Lsm monomers, LadyB oligomers with U<sub>40</sub> RNA.

As for the WT-αHL, we proved the ability of the LadyB to bind the RNA by biochemical experiments using short and unstructured ssRNA.

### 3.2.5 RNA binding: electrophysiological characterization.

Given the obtained biochemical evidences we moved toward biophysical experiments to prove the RNA sensing in single molecule.

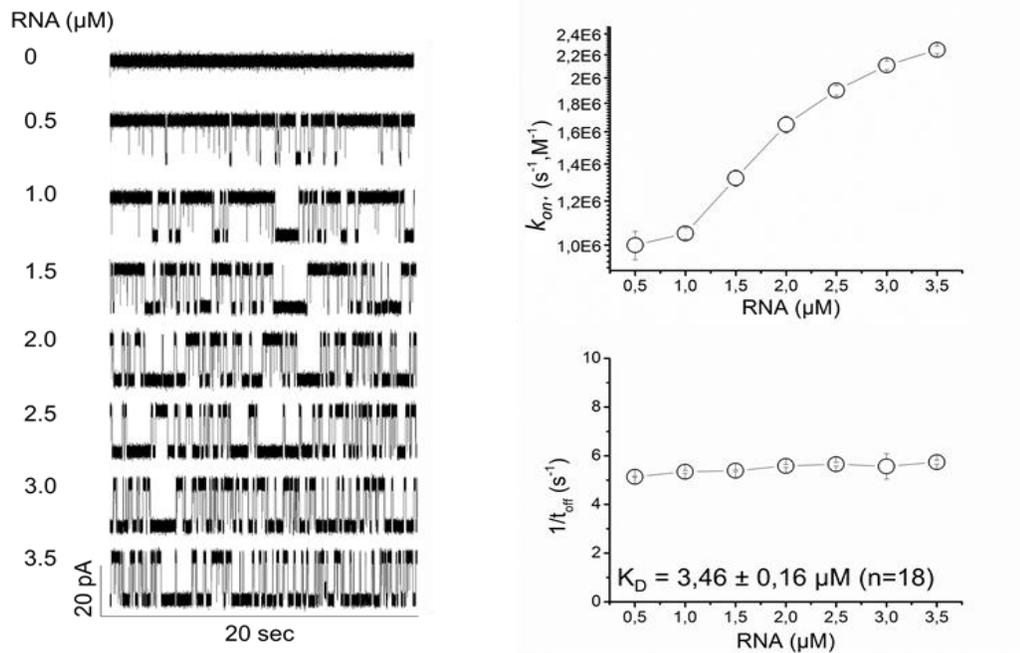
According to electron density analysis on parental Lsm (no crystal structure for the *HV*-Lsm is available) each Lsm monomer bind a single uridine. For a correct binding, the RNA strand must be long and flexible enough to allow the nucleotide accommodation in the protein pockets<sup>29,34</sup>. For this reason we used mainly short ( $\geq 10$  nt and  $\leq 60$  nt) and unstructured ssRNA to characterized the binding of the fusion constructs.

We adopted the same approach used for the WT- $\alpha$ HL sensor (see Chapter 2) and employed the planar bilayer system in single channel recording to study the interaction of ssRNAs with the chimeric nanopore. The RNA was added from the same side of the protein (*cis*, grounded) under an applied constant potential of + 80 mV in *low ionic strength buffer*. In these conditions the protein was stable and the interaction with the RNA was characterized from the blockades observed in the current traces.

For each RNA we obtained the mean dwell time ( $\bar{\tau}_D$ ), the residual current level ( $I_{\%RES}$ ) and the relative amplitude of the blocking current during RNA binding ( $I_b$ ) (as reported in the second chapter for the WT- $\alpha$ HL).

A U<sub>20</sub> ssRNA was recognized with a ion current signal comparable with the characteristic WT- $\alpha$ HL signal behavior ( $I_{\%RES}$  LadyB =  $4,10 \pm 0,54$  pA,  $n = 7$ ;  $\bar{\tau}_D$  LadyB =  $182 \pm 7$  ms,  $n = 20$ ).

We observed that the rate constant ( $k_{on}$ ) of the events increased with the RNA concentration (Figure 3.17 – left, top and right), confirming that the long blocking events ( $> 5$  ms) were dose dependent similarly to what observed for the WT- $\alpha$ HL. Moreover, the mean dwell time has been observed to remain constant (Figure 3.17 – right, bottom).

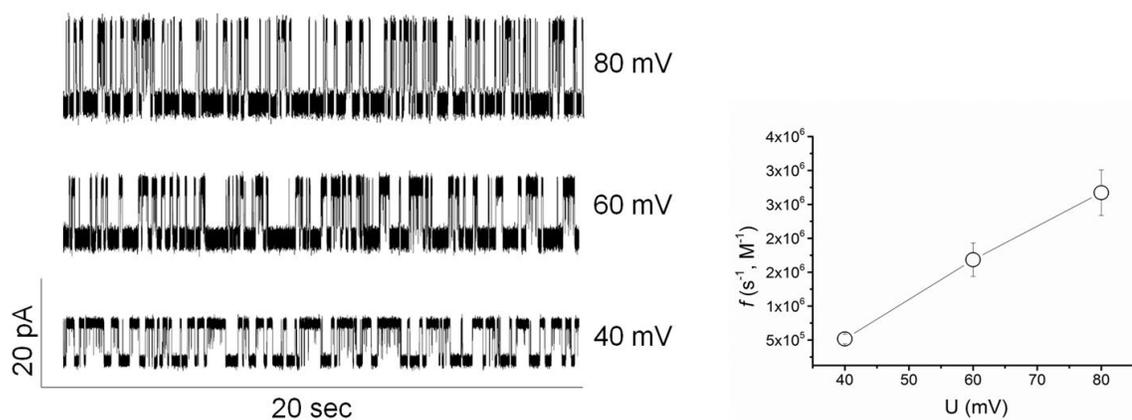


**Figure 3.17 Detection of a polyU<sub>20</sub> ssRNA.** Single-channel ionic current traces. U<sub>20</sub> ssRNA was added to the cis chamber at increasing concentration. Examples of typical ionic current traces for each concentration point are reported.  $k_{on}$ (left top) and  $k_{off}$  values (left bottom) of the events at increasing RNA concentration are reported on the left.

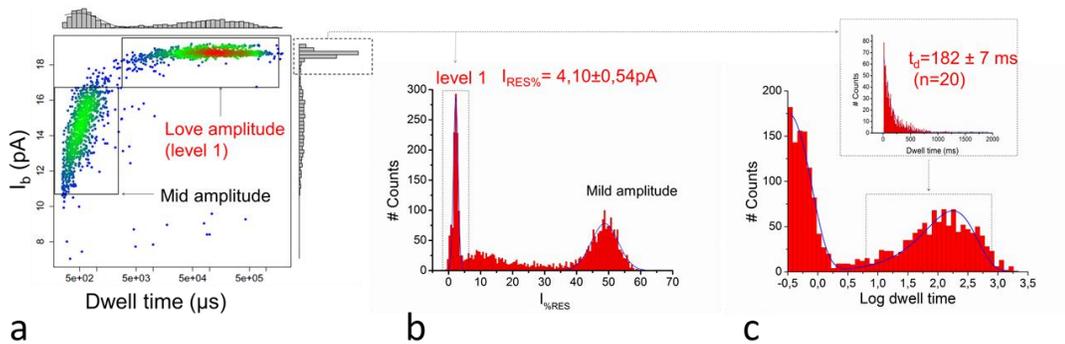
The  $k_{on}$  increased with voltage (Figure 3.18). In the WT- $\alpha$ HL the RNA capture has been demonstrated to occur above a voltage threshold of + 80 mV. At this potential  $\sim 1$  molecule every 1,000 reaching the entrance of the pore is translocated<sup>35</sup>. Interestingly, with the LadyB we observed events with an applied potential of +40 mV (with a  $k_{on}$  of  $5 \times 10^5$ ) (Figure 3.18). This result suggested that the LadyB has a lower entropic threshold for RNA translocation. We speculated that the charged residues at the constriction of the LadyB doughnut affect the capture of the RNA through electrostatic interaction extended to the bulk solution. This perspective is not unrealistic because the Debye length under the experimental conditions we used (20°C,  $\sim 250$ mM KCl) is  $\sim 0.60$  nm and the entrance constriction of the doughnut in the chimeric pore is not embedded in a vestibule like for the WT- $\alpha$ HL. In this way, the Lsm could increase the local concentration of RNA by slowing down the entropic barrier to enter the pore.

Although we did not observe any differences with the WT- $\alpha$ HL in the distribution kinetics of the  $k_{off}$ , the affinity of the chimeric nanopore was  $\sim 5$  times higher than the WT pore. The difference in  $K_D$  was observed to be related mainly with the  $k_{on}$  rather than to the  $k_{off}$ , although the residential time resulted to be  $\sim 30\%$  longer (see Table 3.3 for an overview of the differences between the WT- $\alpha$ HL and the LadyB for RNA binding and channel behavior). As introduced above, the dwell time did not have a double distribution of the events (Figure 3.19), indicating that only a single type of binding is present, with the same characteristics of the WT- $\alpha$ HL (low  $I_{RES\%}$  and long residential time (Figure 3.19c ,d and e). The longer  $\bar{\tau}_D$  observed with the LadyB could be explained by a different electric field across the channel and/or with the Lsm ability to slowdown the RNA translocation: all conditions that could allow a better binding in the  $\alpha$ HL  $\beta$ -barrel.

Although important differences are observed in the kinetics of the blocking events, the binding seems mainly related with the recognition in the  $\beta$ -barrel of the  $\alpha$ HL, helped by the RBP; rather than an exclusive additional binding on the Lsm.

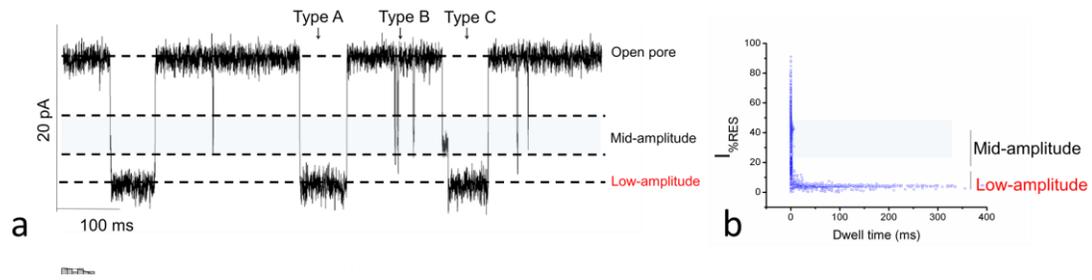


**Figure 3.18 Voltage dependence of the blockades for U2o translocation.** Example of a trace for each potential applied (+40mV, +60V, +80mV) and  $k_{on}$  ( $s^{-1}, M^{-1}$ ) plotted on a linear scale. Data obtained with a poly U<sub>20</sub> ssRNA in *low ionic strength buffer*. The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter).



**Figure 3.19. Behavior of the blocking events.** **a**, Scatter plot of dwell time vs amplitude of the blocking current ( $I_b$ ) is presented. Mid and low-amplitude population are indicated (the scatter plot color code reflect the point density). **b**, Histograms of  $I_{\text{RES}}\%$  and dwell time (**c**) distribution with the level 1 reported in the broken line boxes. The dwell time is reported in  $\text{Log}_{10}$  scale and as a single exponential distribution (e bottom and top respectively) and events  $>5 \text{ ms}$  has been fitted on a single probability density function. Recording in *low ionic strength buffer*.at  $+80 \text{ mV}$ . Filter  $2 \text{ kHz}$  (low-pass Bessel filter). Acquisition at  $20 \text{ kHz}$ .

The ionic current in the presence of the blockades was slightly different from the WT- $\alpha\text{HL}$  (Figure 3.20, see chapter 2 for the WT- $\alpha\text{HL}$  behavior). Three levels in the closure event were observed: once the open pore underwent any closure event, a step characterized by rapid mid amplitude appeared before the canonical (as reported for the WT- $\alpha\text{HL}$ ) low amplitude level (level 1,  $I_{\text{RES}}\% < 10 \%$ ). This type of behavior was previously reported for the translocation of longer DNA oligonucleotides<sup>36</sup>, but we never observed this behavior with short ssRNA oligo U in the WT- $\alpha\text{HL}$ , therefore this is a specific feature of the LadyB.



**Figure 3.20** A zoom in on the signals blockades. **a**, Snapshot of a 3 second trace with the tree detected levels (open, mid amplitude and low-amplitude level) Type of events in single-channel recording: Type A are only level 1 events, Type B are short mid amplitude events and Type C level 1 events with a short mid amplitude step just before the deeper step. **b**, Scatter plot of the residual current blockades (%) and dwell time for current blockades caused by a polyU<sub>20</sub> ssRNA in *low ionic strength buffer* at + 80 mV. The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter).

Since the mid amplitude level was very fast ( $\bar{\tau}_D < 1$  ms) and close to our cut-off filter of acquisition (2 kHz), we decided to further analyze the RNA binding specificity considering exclusively the low amplitude level (level 1). Possible reasons accounting for the mid amplitude level may be the misfolding of Lsm monomers, the motion of the doughnut on the  $\alpha$ HL, or the RNA binding to the Lsm protein. In other word, this uncertainty in the origin of the current signal of the mid-level hampered the accurate characterization of the mid-amplitude events.

To understand the specificity of the U recognition observed previously, we employed four oligo with different sequences were tested against the LadyB single channel: A<sub>10</sub>C<sub>10</sub>, U<sub>17</sub>A<sub>3</sub>; U<sub>20</sub>; C<sub>3</sub>U<sub>7</sub>.

With RNA lacking 3' U tail (e.g. A<sub>10</sub>C<sub>10</sub>), events longer than 5 ms were not detected (Figure 3.21).

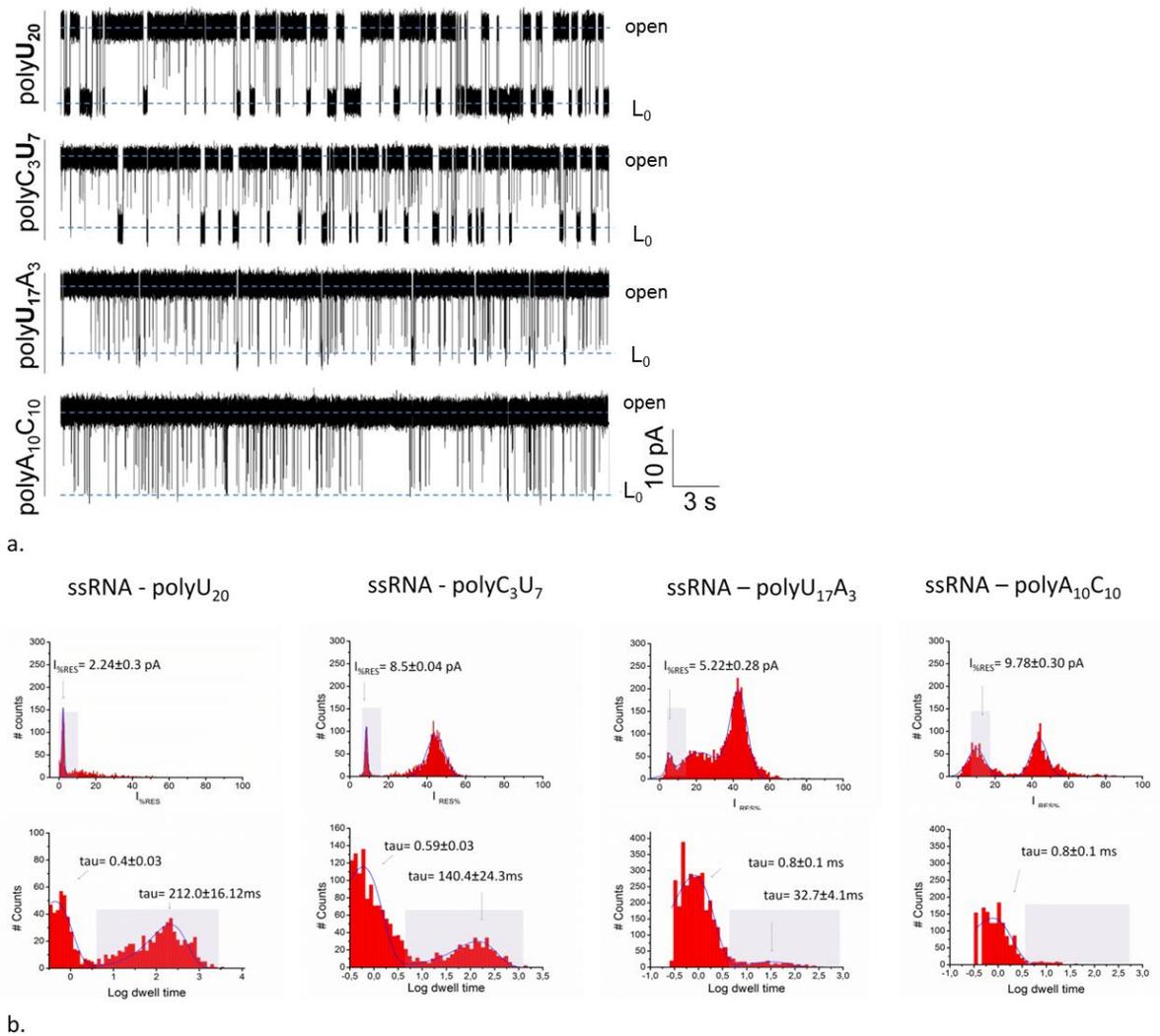
As reported for the the WT- $\alpha$ HL the position of the tail affects the signal. Only if the oligo U was at the 3'-end the oligo it was recognized with long blockades (Table 3.2 and figure 3.21)

RNA	$\bar{\tau}_D$ (ms)	$I_{\%RES}$ (%)
U <sub>20</sub>	182 ± 7	2.2 ± 0.3
C <sub>3</sub> U <sub>7</sub>	143 ± 11	8.5 ± 0.1
U <sub>17</sub> A <sub>3</sub> *	95 ± 12	5.2 ± 0.3
A <sub>15</sub> C <sub>5</sub>	9.8 ± 0.3	0.8 ± 0.1

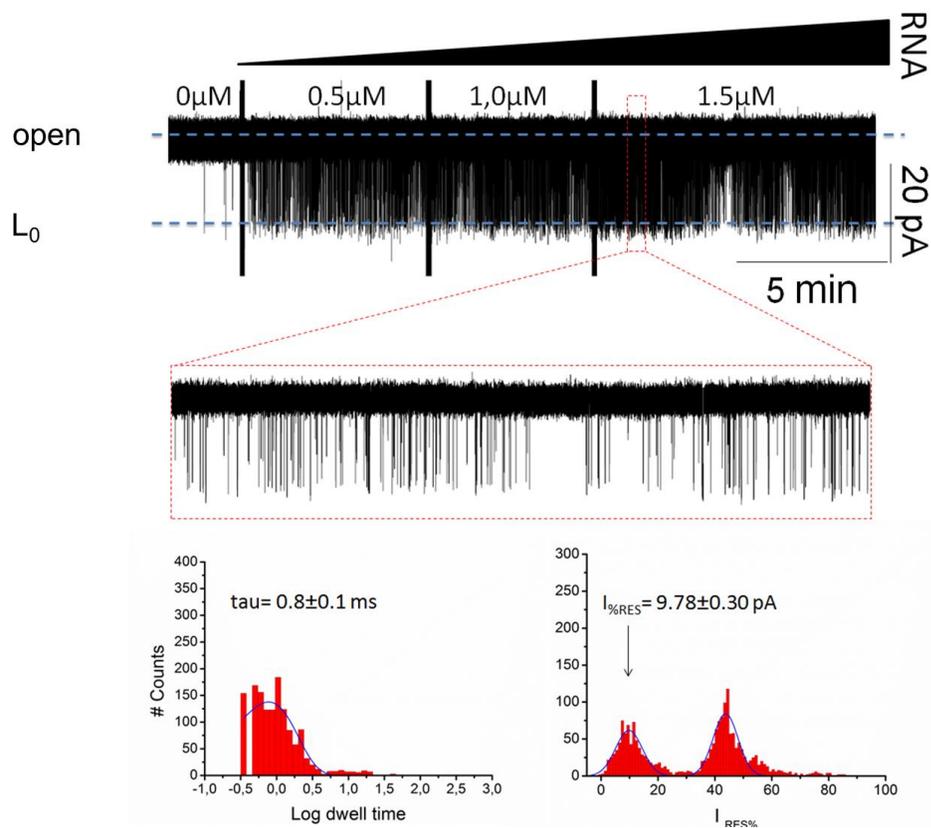
**Table 3.2. Dwell time and residual current for different RNA interacting with the LadyB.** (\*): only population of longer events has been considered;  $k_{on}$ :  $0.9 \cdot 10^6$  at  $2\mu\text{M}$ ;  $K_D$ :  $100\mu\text{M}$ .  $n=3$ .

With oligo U in 5' the  $\bar{\tau}_D$  was reduced  $\sim 50\%$  (Figure 3.21 and Table 3.2), and we observed deep level 1 events ( $< 10\%$ ;  $I_{RES\%}$ ) (Figure 3.21), with a  $k_{on}$  increasing with voltages and RNA concentrations (data not shown). This population of events was not observed with the WT pore.

A  $\bar{\tau}_D$  of  $< 22\mu\text{s}/\text{base}$  for a poly(A) and  $< 2\mu\text{s}/\text{base}$  for a poly(C) has been demonstrated by Akeson (1999)<sup>37</sup> and a  $\bar{\tau}_D$  of  $300\mu\text{s}$  for longer ( $>100\text{ nt}$ ) poly(U) has been demonstrated by Kasianowicz (1996)<sup>38</sup>. Moreover, Deamer and Branton<sup>39</sup> described how the blockade duration decreases shortening the DNA/RNA strand and with oligonucleotides shorter than  $12\text{ nt}$  it decrease exponentially<sup>40</sup>. With the LadyB the mean lifetime for a poly A<sub>10</sub>C<sub>10</sub> was  $\sim 800\mu\text{s}$  (Acquisition  $20\text{ kHz}$ , filter  $5\text{ kHz}$ ), at least  $\sim 4$  times longer than expected for such short oligonucleotides (Figure 3.21 and 3.22). These blockades were depending on the RNA concentration because they increased with the RNA abundance in the *cis* chamber (Figure 3.22).



**Figure 3.21. Effect of the U on ssRNA detection.** **a**, Single-channel recordings LadyB in the presence of ssRNA with different sequences **b**, Statistical analysis of the  $I_{\%RES}$  and  $t_D$  for each of the four oligonucleotides reported in the panel **a**. The plots are a representative of a 6 min trace analysis and the values are the average of at least 3 independent experiments. In the gray panel the population of events relative to the specific U-tag recognition. Recording in *low ionic strength buffer* at +80 mV. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. All the data collected with 1.5  $\mu$ M ssRNA in *cis*.  $L_0$  = zero current level.



**Figure 3.22 RNA A<sub>10</sub>C<sub>10</sub> titration.** Analysis of a 6 min trace. Dwell time (bottom left) and I<sub>%RES</sub> (bottom right) are reported. Data were acquired in *low ionic strength buffer* at +80 mV. The signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter). L<sub>0</sub> = zero current level.

These results are compatible with the possibility that the Lsm doughnut could slow down the translocation of RNAs, probably by the effect of the additional constriction (the doughnut's hole). We suggest that probably the Lsm is not strong enough to keep the oligo U bound and the weak binding could not be clearly detected in our experimental conditions.

In conclusion, our findings demonstrated that the LadyB had different behaviors than the WT- $\alpha$ HL pore (Table 3.3) for the following parameters:

- 1) The electrophysical properties (e.g. ion selectivity, conductance) were slightly different;

- 2) The efficiency in RNA recognition ( $K_D$ ) of 3'-end uridylated ssRNAs was increased;
- 3) Blockades with a characteristic mid amplitude level were observed;
- 4) ssRNA bearing a U-tail in 5' (small population events at low residual current) and with oligonucleotides missing the U-tail (i.e. C<sub>3</sub>U<sub>7</sub>, where only the mid amplitude level was observed) gave longer blockades than observed with the WT pore.

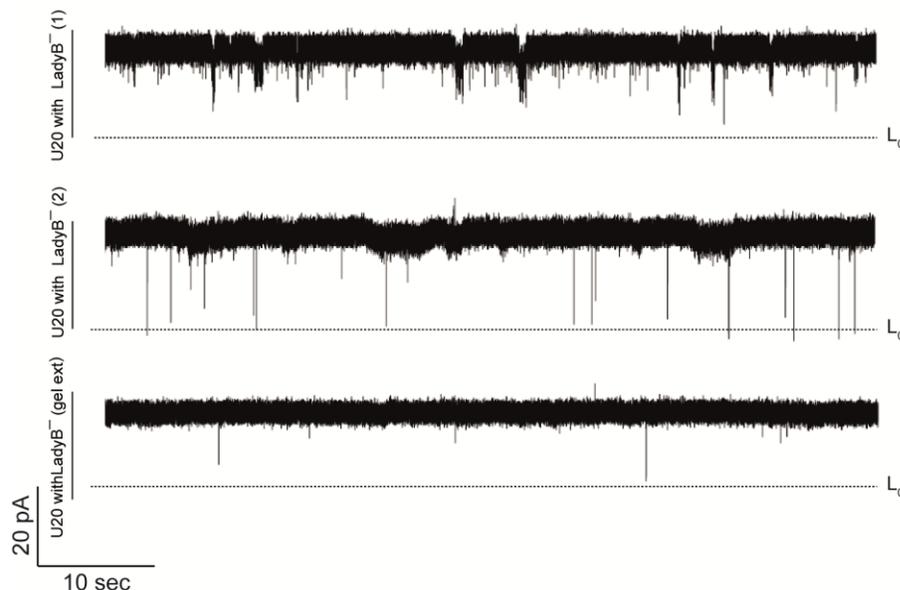
Parameter	WT- $\alpha$ HL pore	LadyB pore
Pore formation/lysis	yes	yes
G (pS)	268 $\pm$ 90 (n = 200)	179 $\pm$ 75 (n = 150)
P <sup>+</sup> /P <sup>-</sup> (a <sub>cis</sub> /a <sub>trans</sub> = 4)*	0.34 $\pm$ 0.10 (n = 5)	0.70 $\pm$ 0.03 (n = 5)
Voltage threshold	+ 80 mV	+ 40 mV
K <sub>D</sub> ( $\mu$ M)	15.0 $\pm$ 5 $\mu$ M (n>3)	3.6 $\pm$ 0.3 $\mu$ M (n>3)
k <sub>on</sub> (M <sup>-1</sup> s <sup>-1</sup> ) - U <sub>20</sub>	0.6 $\times$ 10 <sup>6</sup> $\pm$ 0.1 $\times$ 10 <sup>6</sup> (n > 3)	1.5 $\times$ 10 <sup>6</sup> $\pm$ 0.2 $\times$ 10 <sup>6</sup> (n = 3)
$\bar{\tau}_D$ (s) - U <sub>20</sub>	0.119 $\pm$ 0.013 (n > 3)	0.182 $\pm$ 0.007 (n > 3)
Mid step events	no (with ssRNAs < 40 nt)	yes
$\bar{\tau}_D$ (s) U <sub>17</sub> A <sub>3</sub>	no	0.095 $\pm$ 0.012 (n > 3)**
$\bar{\tau}_D$ (s) A <sub>10</sub> C <sub>10</sub>	no	0.0008 $\pm$ 0.0001 (n = 3)
Stability in SDS	yes	yes***
Stability in urea (2M)	yes	Yes

**Table 3.3. Characterization of the different behaviors: LadyB pore vs WT- $\alpha$ HL pore.** Data acquired in low ionic strength buffer.: + 80 mV voltage applied to calculate the rate and dissociation constants. Rate constants calculated on 1  $\mu$ M RNA oligo (U<sub>20</sub> if not reported). The signal was acquired at 20 kHz and filtered at 2 kHz (5 kHz for the A<sub>10</sub>C<sub>10</sub>). (\*): at P<sup>+</sup>/P<sup>-</sup> (a<sub>cis</sub>/a<sub>trans</sub> = 2) WT value was 0.20  $\pm$  0.06 and LadyB values were from 0.68  $\pm$  0.09 (n=3) to 1.95  $\pm$  0.03 (n=3). (\*\*): k<sub>on</sub>: 0.9  $\times$  10<sup>6</sup> at 2 $\mu$ M. (\*\*\*) the stability of the *in vivo* expressed protein has not been characterized.

### 3.2.6 Discrimination of the *Lsm*/RNA binding signal.

In order to understand the origin of the above mentioned differences in term of binding site, we mutated the *LadyB* in position D127N, D128N and K131A to silence the  $\alpha$ HL  $\beta$ -barrel. We referred to this mutant as *LadyB*<sup>-</sup>. We expressed this protein *in vitro* and we tested in PLM the binding properties toward C<sub>3</sub>U<sub>7</sub>, an RNA that showed a strong binding (Figure 3.23).

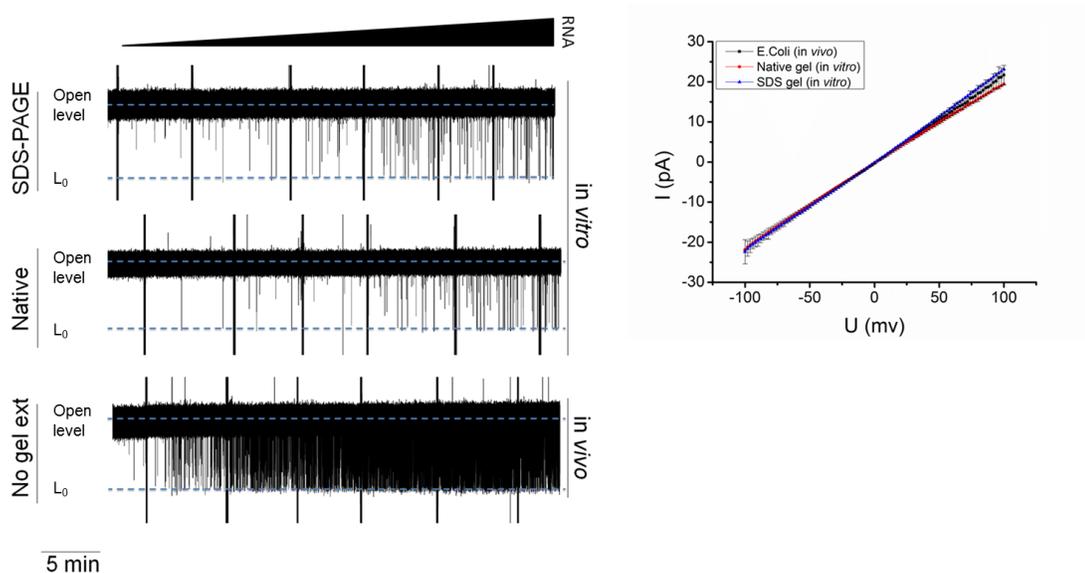
The protein extracted from SDS gel did not show any blockades, while the protein oligomerized on RRBCM showed mid-amplitude signals in the current trace when tested in PLM (Figure 3.23).



**Figure 3.23 *LadyB*<sup>-</sup> RNA sensing in single channel recording.** *LadyB*<sup>-</sup> expressed by *in vitro* transcription-translation was incubated with 2 $\mu$ M of C<sub>3</sub>U<sub>7</sub> in *cis* (panel 1 and 2 show two independent experiments). In the last panel *LadyB*<sup>-</sup> protein extracted from SDS-PAGE gel has been used. Recording conditions: 150 mM KCl, 100 mM NaCl, 2 mM MgCl<sub>2</sub>, 10 mM HEPES pH 6.5 (*low ionic strength buffer*). The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz, +80 mV. Lo = zero current level

Interestingly, we observed a significant difference in binding capability between protein expressed *in vivo* or *in vitro* and extracted from PAGE gels (Figure 3.24). The *in vitro* expression is probably not optimal for the correct expression and folding of the fusion

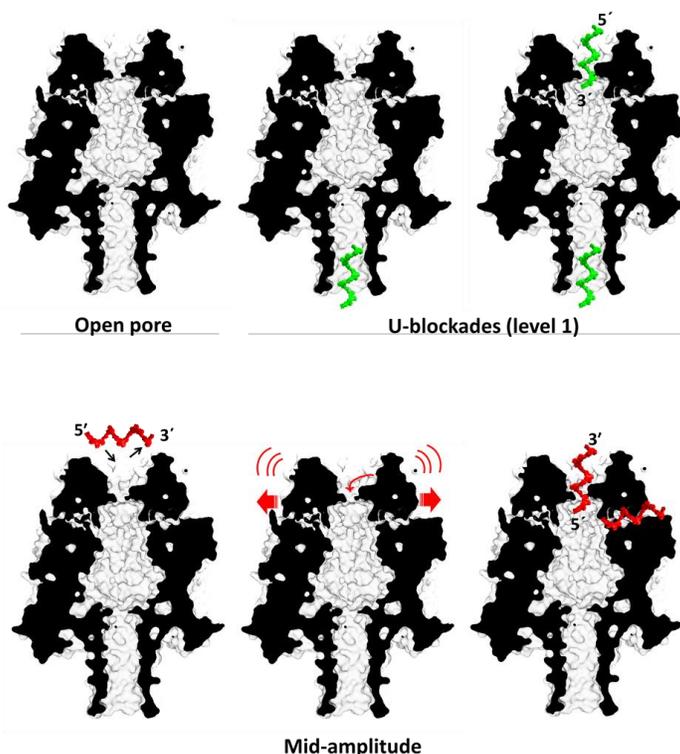
construct. Moreover, the gel extraction probably destroys the Lsm quaternary structure. Further experiments with *E. Coli* expressed protein are required to better understand the Lsm binding in single channel electrical recording. If not stated differently, the experiments above reported) we always used *in vivo* expressed LadyB protein without gel extraction.



**Figure 3.24 Single channel recording: *in vitro* vs *in vivo* protein expression.** Single-channel recording at increasing  $U_{20}$  RNA concentration ( $0.25 \mu\text{M}$  each addition) of different oligomers preparation (right) and relative IV curve (left). Recording in *low ionic strength buffer*, + 80 mV, signal was acquired at 20 kHz and filtered at 2 kHz (low-pass Bessel filter)- Lo: zero current level.

In conclusion, to rationalize all the information obtained from electrophysiological and biochemical data, we proposed a possible model of RNA binding. It is possible that the low-amplitude blockades (level 1) were caused by  $\alpha\text{HL}$  recognition, helped by a weak binding of the Lsm, owing the mid-amplitude blockades (Figure 3.25). Whereas the only mid-amplitude levels could be originated from due to unspecific binding of the RNA, from a movement of doughnut on the cap, or from RNA translocation in a  $3' \rightarrow 5'$  orientation through the Lsm or through leaking sites between the Lsm and the  $\alpha\text{HL}$ . More ex-

periments with E.Coli expressed LadyB<sup>+</sup> protein are required to better understand binding.

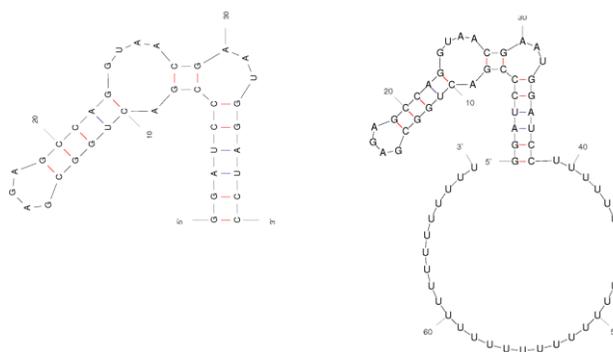


**Figure 3.25 Model of RNA binding for LadyB.** *Low amplitude events (Level 1):* Specific interaction of ssRNA with RNA-binding pockets of Lsm and WT- $\alpha$ HL. The dwell time is longer when the U-tag enters oriented in 3'. *Mid-amplitude events:* signals due to unspecific binding, noise from doughnut movement and 3'  $\rightarrow$  5' RNA translocation through the Lsm constriction or through leaking sites between the Lsm and the  $\alpha$ HL.

### 3.2.7 Detection of RNAs with complex secondary structure

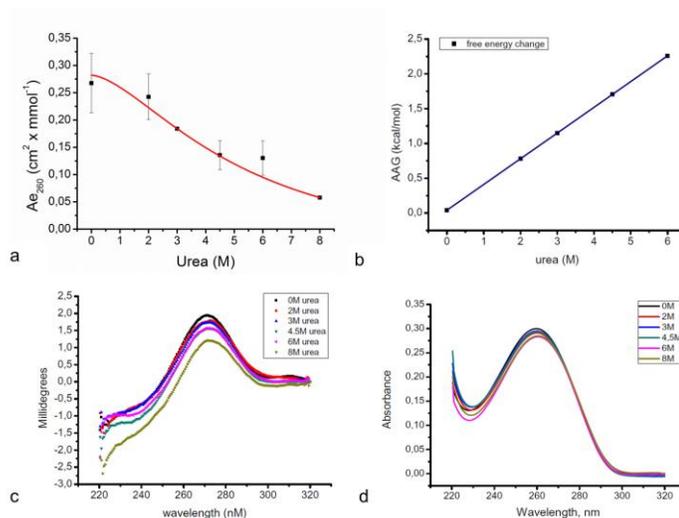
RNA can fold in complex, different and functional secondary structures. In a folded form it cannot pass through the constrictions of the RBnP. A common way to unfold RNA structure is the use of urea. In order to study the effect of urea on the LadyB binding property of folded RNAs, we used 2 M urea for the partial unfolding of an oligo formed by two parts: the malachite green RNA aptamer (MGA) and a poly(U)<sub>31</sub> tail in 3'. As MGA, we used the same MGA sequence (38 nt) that has been previously described by NMR and crystal structures<sup>41,42</sup>. According to Mfold<sup>43</sup> (<http://www.ncrna.org/software/rfold/> <http://mfold.rna.albany.edu/?q=mfold>) simulations, the U-tail should not affect destroy overall 2D structure of the aptamer (Figure 3.26).

We first demonstrated by circular dichroism (CD) experiments that the MGA can be partially unfolded in 2M urea (Figure 3.26 - left) and that the presence of 2M urea in the *low ionic strength buffer* can reduce for ~20% the energy required for unfolding (Figure 3.26 - right).



**Figure 3.26 Mfold structure prediction.** MGA without – left (-11.00 kcal/mol at 37 °C) and with the U-tag – right ( $\Delta G = -18.00$  kcal/mol at 37 °C). Ionic conditions:  $[Na^+] = 0.05$  M,  $[Mg^{++}] = 0$  M.

In order to test if this low concentration of urea can affect the translocation of folded RNA in the LadyB, we performed single channel experiments in the presence of 2M urea in both *cis* and *trans* chambers (symmetric condition).



**Figure 3.27 Circular dichroism analysis of the MGA-U RNA.** **a**, Molar circular dichroism absorption at 260nm wavelength ( $n=3$ ) and an example. **b**, Free energy changes as a function of urea concentration for MGA-U<sub>31</sub> at 25°C in EM buffer. **c**, of a circular dichroism and absorbance spectra for each concentration reported as a function of the angle with **d**, relative absorbance spectra.

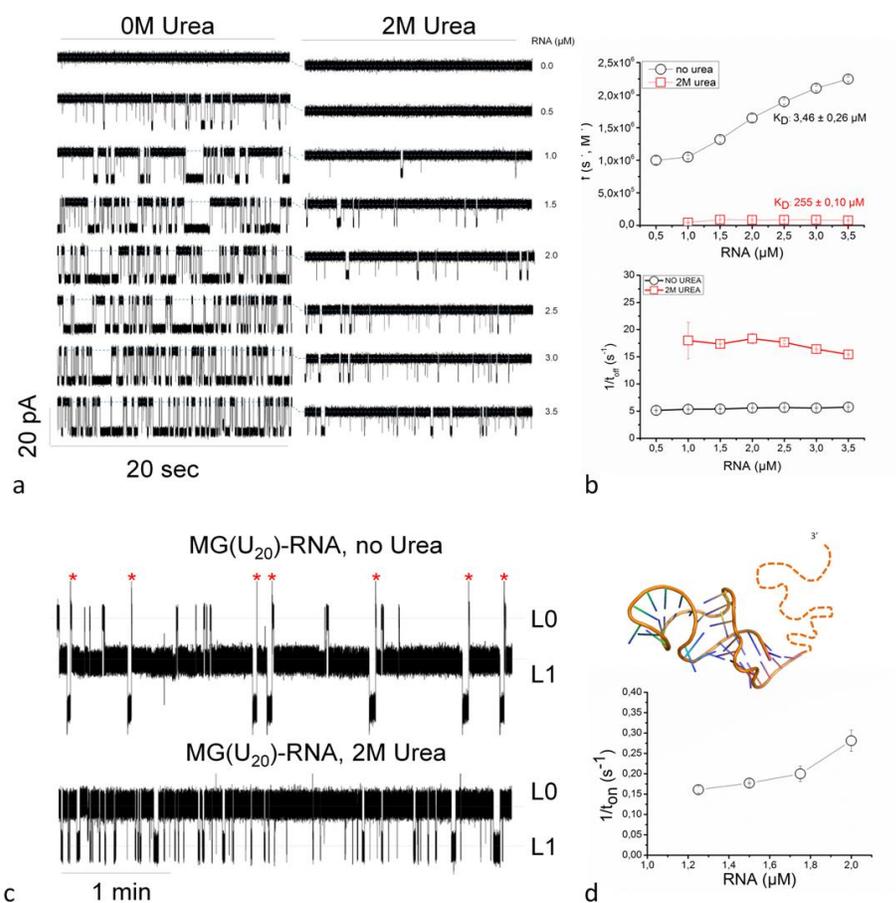
In the absence of urea the MG-RNA could not translocate through the LadyB, resulting in permanent blockades or very long events ( $> 60$  sec, Figure 3.28c - top). In the presence of 2 M urea the MG-U<sub>31</sub> did not show any irreversible blockades (Figure 3.28c - bottom) and events with a  $\bar{\tau}_D$  of  $\sim 100$  ms were observed. The “three-step” signal pattern observed with the WT- $\alpha$ HL pore in *high ionic strength buffer* at + 120 mV, was not observed in *low ionic strength buffer* with the LadyB with 2M urea at + 80 mV. This is because in this condition the resolution in current is lost.

After the results of 2 M urea on a complex RNA (i.e. MGA) we characterized the stability of the LadyB in the same denaturant conditions (2 M, in both *cis* and *trans* chamber).

The stability of the WT- $\alpha$ HL oligomers has been previously examined at various urea concentrations by polyacrylamide gel electrophoresis and circular dichroism<sup>44</sup>. The WT- $\alpha$ HL pore was found to be stable even at 8M urea<sup>44</sup>. We decided to test the stability of LadyB, and in particular of the Lsm fusion ring, in the presence of the previously use low urea concentration (2M).

The LadyB pore did not show any instability at this concentration of denaturant. Therefore, we tested if in this condition was still possible to appreciate the RNA binding of the short and unstructured RNAs (i.e. U20) previously characterized.

We observed a strong effect of the urea on the RNA interaction with the  $\alpha$ HL pore. The  $k_{on}$  of a  $U_{20}$  was drastically reduced in the presence of 2M urea, probably for its interference with the RNA/protein binding. In urea we were still able to detect the RNA binding but with a reduced sensitivity ( $K_D \sim 10 \mu\text{M}$ ). The lifetime of the observed events ( $\sim 50 \text{ ms}$ ) was  $\sim 5$  times shorter than recorded for a  $U_{20}$  without urea (Figure 3.28a and 3.28b).



**Figure 3.28 Effect of 2M urea on the recognition of ssRNAs.** **a**, No urea/2M urea comparison on the binding of a short poly  $U_{20}$  ssRNA and **b**, relative frequency of the events and mean  $k_{off}$  values for each concentration points in the two different conditions (red: 2M urea, black: 0M urea). **c**, Single-channel recordings of MG- $U_{31}$  without (top) or with 2M urea (bottom). **d**, Rate constant of the inter-events time at increasing RNA concentrations and 3D structure of the MG aptamer in solution as reported in the PDB database (PDB: IQ8N). The dashed line represents the  $U_{31}$ -tag added in 3'. Recording conditions: 150 mM KCl, 100 mM NaCl, 2 mM  $\text{MgCl}_2$ , 10 mM HEPES pH 6.5 (*low ionic strength buffer*). The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz. +80 mV.

In conclusion, urea reduced the residential time of the RNA in the LadyB pore and the frequency of the events. Urea at 2 M could unfold the MG aptamer and allowed it to translocate without irreversible block of the channel (this is an indirect prove of translocation of this RNA in the LadyB pore). The strong decrease in binding affinity toward the short and unstructured U<sub>20</sub> suggests that the weak binding contribution of the Lsm doughnut was probably eliminated in the presence of urea, but the pore itself is stable at this concentration of denaturant. In principle, the nanopore could be used to detect 3' U-tail on every kind of long and structured RNA under denaturing condition, without any pre-treatment of the sample. Nevertheless, for this approach we must to consider the reduction in sensitivity of the channel. As observed for the WT- $\alpha$ HL, in the presence of RNA with secondary structure, it is hard to understand the contribution of the specific U-binding from the unfolding signal.

### 3.3 Conclusion

Our attempt to implement the RNA-binding property of the WT- $\alpha$ HL on 3'-end uridylation was only partially successful. By both biochemical and electrophysiological approaches we demonstrated the pore forming activity of a new chimeric pore (LadyB) and the selective binding of 3' oligo U RNAs.

So far, LadyB is the first functional chimeric protein prepared using the  $\alpha$ HL as scaffold. The chimeric nanopore was active on RNA binding in bulk solution ( $K_D = 63$  nM) and slightly better than the WT- $\alpha$ HL in single channel electrical recording ( $K_{DLadyB\ C_3U_7} = 3.5 \pm 0.3$   $\mu$ M ( $n > 3$ ) at +80 mV). The different dissociation constant was due to a longer mean residential time (+30%) and to an increase in the  $k_{on}$  of the events. The dwell time displayed a single distribution, indicating that the analyzed signal originated mainly from the binding in the *trans* side of the  $\alpha$ HL  $\beta$ -barrel (D127-K131, see chapter 2), as confirmed by the result obtained with the mutant LadyB<sup>-</sup>.

The increased sensitivity of LadyB compared to the WT- $\alpha$ HL could be explained by (i) a reduction of the entropic barrier for RNA translocation and/or (ii) a slowdown of the RNA translocation.

Preliminary results with the LadyB<sup>-</sup> (LadyB D128N, D127N, K131A) pointed out a peculiar current signal behavior in the presence of RNA, with blockades events with high residual current (~80%). Remarkably, these mid-amplitude blockades have been observed specifically with the LadyB<sup>-</sup>. Together, our results demonstrate that this chimeric construct can be used as a variant of  $\alpha$ HL pore in RNA sequencing for the ability to slow down the RNA translocation. Further experiments are required to better understand the potential of the Lsm-binding in this fusion construct.

Finally, this new protein is a first step toward a modular and sequential recognition of nucleic acids by biological nanopores. Further efforts should be done to increase the sequence complexity as the length and the structure of the sequence signature recognized. For example different domains can be fused with the  $\alpha$ HL (or other pore forming toxins) to bind more complex RNA binding motif than a poly(U), increasing the complexity both in sequence and structure. This because the “nanopore world” is a constantly renewing field and adding new horizons is a never ending endeavor.

APPENDIX of CHAPTER 3

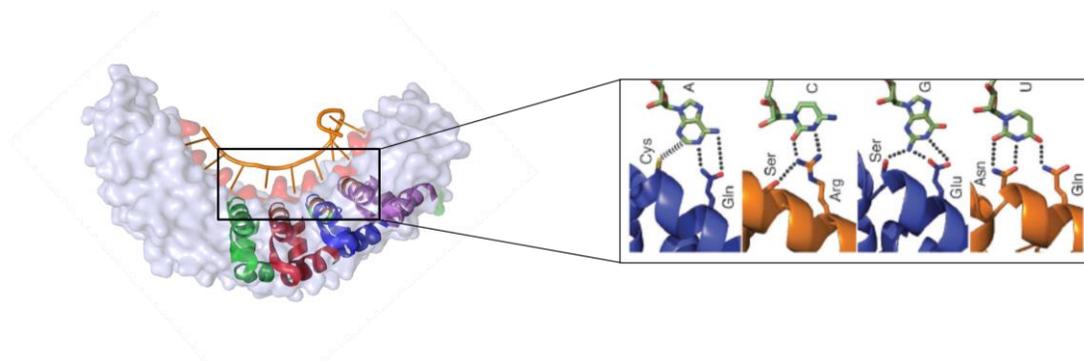
The PUF case

## A.1 Introduction.

### A.1.1 PUF repeats

The interaction between proteins and nucleic acids is characterized by combinations of electrostatic interactions, intermolecular hydrogen bonds, stacking interactions and van der Waals bonding. All these chemical interactions are used to achieve a specific recognition. The recognition of the RNA sequence and overall structural 3D features allows the discrimination<sup>45</sup>. The complex interplay between these interactions typically results in idiosyncratic but difficult to predict binding. Exceptions are the PUFs repeats. PUFs are phylogenetically evolved alpha helices arrays that bind nucleic acid in a modular fashion with one single repeat interacting with one base (Figure A.1). Only two residues within each repeat dictate the sequence specificity in a predictable manner<sup>7</sup>. PUF proteins are named after their founding members *Drosophila melanogaster* Pumilio (DmPUM). This protein acts regulating gene expression by binding to specific sequences in the 3'-untranslated region (UTR) of the maternal *hunchback* (*hbmat*) mRNA in the posterior half of the *Drosophila* embryo, thereby permitting abdominal development<sup>46</sup>. All PUF domains contain multiple repeats (typically 8) each of 36 amino acid<sup>47</sup>. The single stranded RNA runs antiparallel to the protein and binds to the inner concave surface generated by the PUF repeats<sup>48</sup>. Each repeat binds a single nucleotide of the target RNA in a modular fashion (Figure A.1). The side chains of the residues 12 and 16 binds the RNA base via hydrogen bonding or Van der Waals contact, whereas the amino acid in position 13 makes a stacking interaction with the base<sup>49</sup>. The code is relatively simple: the two side chains in position 12 and 16 (Gln<sup>12</sup> and Cys<sup>16</sup> for A; Glu<sup>12</sup> and Ser<sup>16</sup> for G; Gln<sup>12</sup> and Asn<sup>16</sup> for U) are dictating the target nucleotide sequence. Naturally occurring PUF domains have not been observed to recognize cytosine. Recently, two independent groups<sup>7,49,50</sup> reported an *in vitro* selection of PUF repeat variants that specifically recognize cytosine. Dong and coworkers reported the crystal structure of the mutated PUF domain showing that Arg<sup>12</sup> and Ser<sup>16</sup> bound cytosine. The simple code and the possibility to change the specificity by changing only two residues, makes it easy to play with the target sequence. In fact, mRNA localization<sup>51</sup>, selective splicing reaction<sup>52</sup> and transcrip-

tion activation<sup>50</sup> can be achieved fusing effector domains with a PUF sequence-specific domain. Moreover, increasing the number of repeats in the PUF domain the target sequence has been successfully increased from 8 nt up to 16 nt<sup>49</sup>.



**Figure A.1 PUF repeats.** PUF proteins contain a domain composed of eight 36 amino-acid repeats. Each repeat binds a single nucleotide of the target RNA. Amino acids side chains at position 12 and 16 of the PUF repeat binds each RNA base via hydrogen bonding or Van der Waals contact with the Watson-Crick edge, whereas the amino acid at position 13 makes a stacking interaction. The only two residues 12 and 16 are able to discriminate between the four RNA bases. PDB file 1M8Y

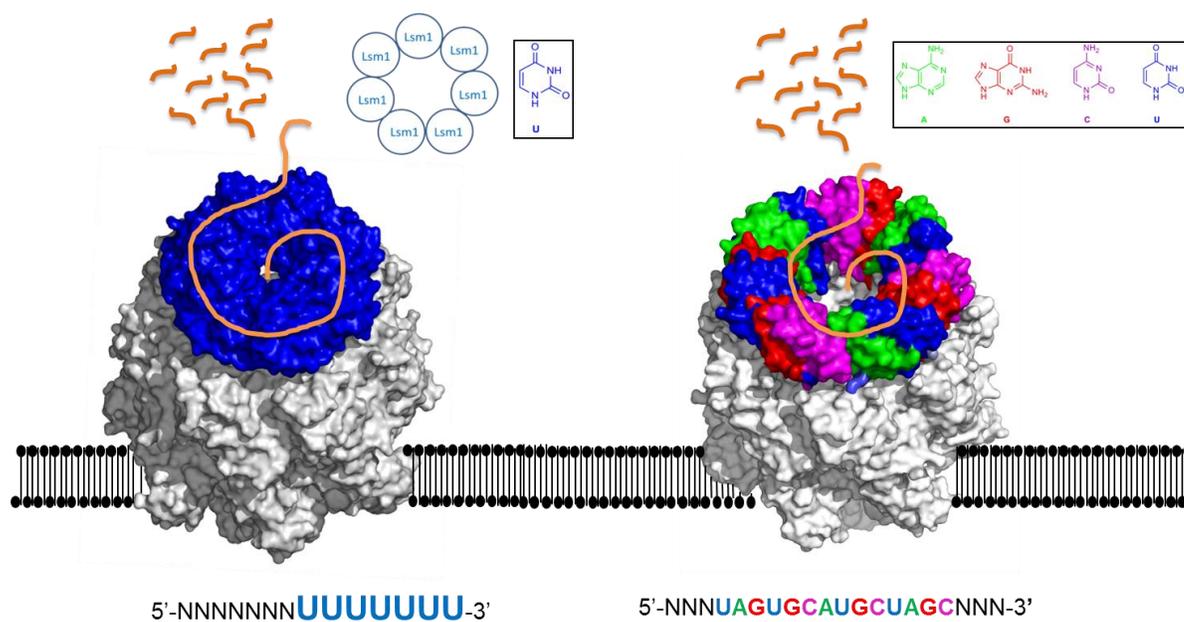
Despite its relative easy design (heptameric ring on a heptameric pore) LadyB had some drawbacks: first, the possibility that ionic leaking sides between the doughnuts and the cap of the WT- $\alpha$ HL are present and, second the limitation in the sequence specific recognition for oligo U only.

In order to overcome these issues, possible solutions were: (i) to mutate the *HV*-Lsm protein by single point mutagenesis and/or choose other more complex parental eukaryotic Lsm in order to broaden the base recognition, (ii) to develop new RNA binding nanopores (RBnPs) fusing the nanopore with known modular domains able to selectively recognize each four nucleotide.

Inspired by the recent burst of application and scientific publications related to the DNA binding TALEs repeats, we engineered the WT- $\alpha$ HL genetically fusing its C-terminus with two PUF domains, able to selectively bind adenines. We connected the two building blocks with a SG linker (PUF-RBnP) (see Methods). Once the toxin has oli-

gomerized on the lipid bilayer a ring of 14 domains should be ready for the binding of a poly(A) single stranded RNA (Figure A.2).

Whit the PUF-RBnP we aimed to design custom RBnPs for any type of RNA sequence, expanding the potential of RNA-sequence signature recognition in single molecule to its limit.



**Figure A.2. Design of LadyB and PUF-RBnP: an overview.** LadyB (left) can recognized only U-tails and it has only seven binding monomers. PUF-RBnP (right) can bind every type of nucleotide in a predictable manner. Moreover, the length of the sequence strand can be increased because the  $\alpha$ HL cap bear a total of 14 Puf domains in a ring shape.

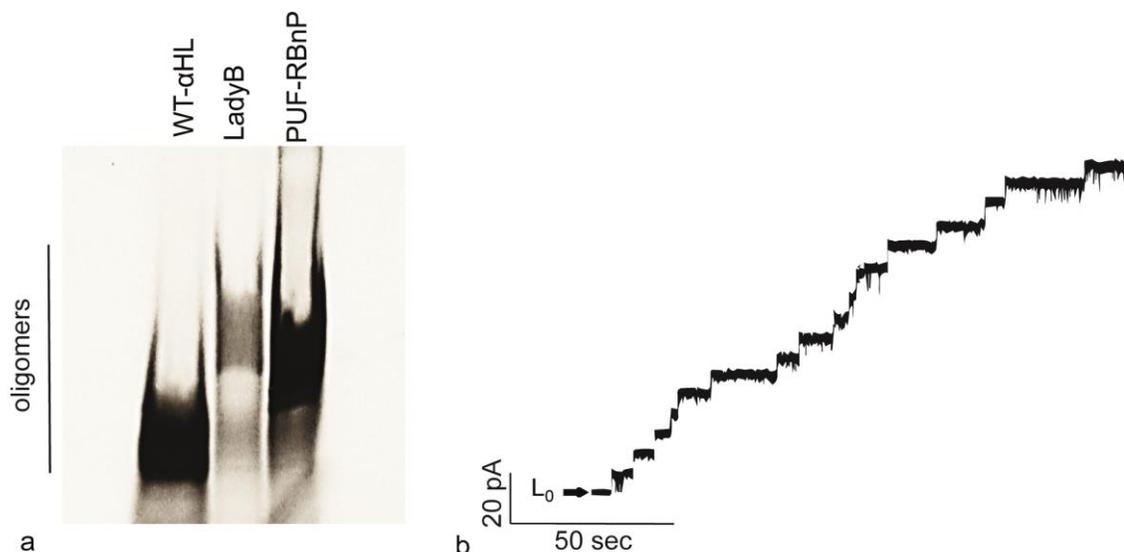
## A.2 Results and discussion.

We designed the RBnP-PUF by fusing the wild-type  $\alpha$ HL with two wild-type PUF domains that are selective for adenosine ( $\alpha$ HL(wt)-(GS)<sub>12</sub> $\alpha$ GGC(GS)<sub>12</sub>(PUF)<sub>2</sub>C74, where  $\alpha$  stands for alpha helix -helical cap domain with protective functions-). Each PUF domain is formed by two alpha helices. We introduce an additional alpha helix between the  $\alpha$ HL and the first PUF domain in order to stabilize the structure and increase the steric volume (the sequence of the additional helix has been copied from the protective cap domain of PUF proteins). Two PUF domains should fit on a  $\alpha$ HL monomer (based on pdb file combination). After oligomerization on the heptameric pore a ring of 14 PUF domains should sit on the top of the  $\alpha$ HL cap. As for the LadyB pore (seven mutated Lsm monomers) PUF domains has been modified introducing two cysteines residues in order to stabilize the ring by S-S disulfide bridges: the first has been introduced in the additional alpha helix and the second in the SG extension in C-term of the last PUF domain. In order to enhance the formation of disulfide bridges, the chimeric oligomers were oxidized in a *low ionic strength buffer* with 1.5 mM of Cu (o-phenanthroline)<sub>2</sub> (see Methods).

The monomeric protein was expressed *in vitro* (see Methods) and allowed to oligomerize on RRBCM. The oligomerization efficiency was tested by SDS-PAGE (4.5%) after assembly on RRBCM and as reported in Figure A.3a the oligomeric pores appears as a band a higher MW than the WT- $\alpha$ HL pore. The chimeric protein showed a good ability to form oligomers on RRBCM.

We then started the characterization of the nanopore following the same rational used for the LadyB (conductance distribution, stability, ability to for oligomers). Oligomers were tested in PLM in order to demonstrate the pore forming activity in synthetic bilayers (see Methods). The ability to form pores in PLM (Figure A.3b) was comparable with the WT- $\alpha$ HL pore. The single channel mean conductance was lower than the WT and comparable to the LadyB (data not shown). The channel was stable up to + 90 mV and irreversible blockades have been observed at higher potentials.

These preliminary results were very promising because channel formation and stability were not obvious goal in the design of this nanopore.



**Figure A.3 Oligomerization and insertions in a planar lipid bilayer.** **a**, SDS-PAGE gel (4.5%) with a comparison of WT, LadyB and PUF-RBnP oligomers. **b**, multiple insertion of PUF-RBnP in PLM. L<sub>0</sub>: zero current level. Buffer EM, +40mV voltage applied. The signal was filtered at 2 kHz (low-pass Bessel filter) and acquired at 20 kHz.

### A.3 Conclusion

Recent biotechnological applications used PUF and TALE repeats fused to effector domains with a well-characterized functions<sup>11,52</sup>. Fusing PUF domains with the  $\alpha$ HL, we used the  $\alpha$ HL pore as a framework for the PUF binding activity, inverting the point of view and adapting the system to RNA sensing. Here, we demonstrated that the oligomerization of multiple PUF domains on the  $\alpha$ HL cap does not affect the pore forming activity and the stability of the channel in PLM.

These preliminary results are important to set the bases for future designs of custom RBnPs able to bind different RNA-sequence signatures. A possible limitation of this approach can lie in the ring shape of the PUFs domains after oligomerization (the native linear structure has been distorted). An eventual interesting solution could be the fusion

of PUF domains on a single  $\alpha$ HL monomer followed by the purification of the functional heteroheptamer (1 : 6 -  $\alpha$ HL(PUF)<sub>x</sub> : WT).

The use in biotechnology of PUF domains is likely to explode in an analogous manner of TALE domains, especially given the role of RNA in numerous post-transcriptional gene regulatory mechanisms.

In conclusion, we demonstrated that is indeed feasible to take up new directions for the use of the modular PUF domains and developing new single molecule RNA sensing strategies.

### 3.4 Methods (Chapter 3 and Appendix)

#### 3.4.1 Gene sequences.

The gene encoding for the chimeric protein was designed using Lablife online tools (<https://www.lablife.org/ll>) and purchased from Genescript (<http://www.genescript.com/>) in a pUC-57 plasmid vector. The coding sequence of the LadyB chimeric nanopore is reported in Figure 3.29. The sequence of the Chimera 1 was identical to the LadyB, with the only exceptions of the mutations C62T; C75A; A76T that restored the WT-Lsm sequence. The coding sequence has been sub-cloned in a vector suitable for in vitro transcription-translation (pT7) and for in vivo *E. coli* expression (pGS21) using restriction enzymes Ndo I, Nco I and Hind III (NEB). The coding regions of the all plasmids were verified by sequence analysis.

```
ATGGCTCATATGGCGGACTCGGACATCAACATCAAAACGGGCACGACGGACATCGGTTCCAATACGACG
GTGAAAACGGGCGACCTGGTGACCTATGATAAAGAAAACGGCATGCATAAAAAGGTGTTTTATAGCTTC
ATCGATGACAAAAACCACAATAAAAAGCTGCTGGTGATTCGTACCAAGGGCAGCATCGCGGGCCAGTAT
CGCGTTTACTCAGAAGAAGGCGCCAATAAATCAGGTCTGGCATGGCCGTCGGCTPTTAAGGTCCAGCTG
CAACTGCCGGATAACGAAGTGGCGCAAATTAGCGACTATTACCCGCGTAACTCTATCGATACCAAGGAA
TACATGAGTACCCTGACGTACGGCTTCAACGGTAATGTTACCGGCGATGACACGGGTAAAAATTGGCGGT
CTGATCGGCGCCAACGTTAGTATTGGTCATACCCTGAAATACGTCCAGCCGGACTTTAAGACCATCCTG
GAATCCCCGACGGATAAAAAGGTTGGCTGGAAAGTCATTTTCAACAATATGGTCAACCAGAATTGGGGT
CCGTATGATCGTGACAGCTGGAACCCGGTGTACGGCAATCAACTGTTTATGAAAACCCGCAACGGTTCT
ATGAAGGCGGCCGATAAATTCTGGACCCGAATAAAGCGAGCTCTCTGCTGAGTTCGGCTTTTCAACCG
GACTTCGCAACCGTGATCACGATGGATCGTAAAGCGAGCAAGCAGCAAACCAATATGATGTGATCTAT
GAACGTGTTTCGCGATGACTACCAGCTGCATTGGACCTCTACGAACTGGAAAGGTACCAATACGAAGGAT
AAGTGGACCGACCGTTTCATCGGAACGCTACAAGATCGATTGGGAAAAGGAAGAAATGACGAAAT

GGCGGTAGCTCTGGCGGTAGTTCGGTGGTGGTTCATCGGGTGGTAGCAGCGGTGGTGGTGGTAGT
TCCGGT

AGTGGTCGTCCGCTGGATGTTCTGGAAGCATCCCTGGACGAACCGGTCACCGTGCTGCTGAAAGATGGC
AACCGTATTTTGGCGTTCTGGCCGGTTACGATCAGCACCTGAATGTGGTTCGGAAGAAGCTCTGGAT
GAAGACACGGTCCCGGGTGATATTGAACTGGAACAGGTCCAGGATTGTACGATTATTCGCGGTGATAAC
GTGGTGACGATTAAGTGTGCGTAAAAGCTT
```

**Figure 3.29** Nucleic acid sequences LadyB. wt- $\alpha$ HL is highlighted in green, violet the SG linker, blue HV-LSM T62C;A75C;N76A.

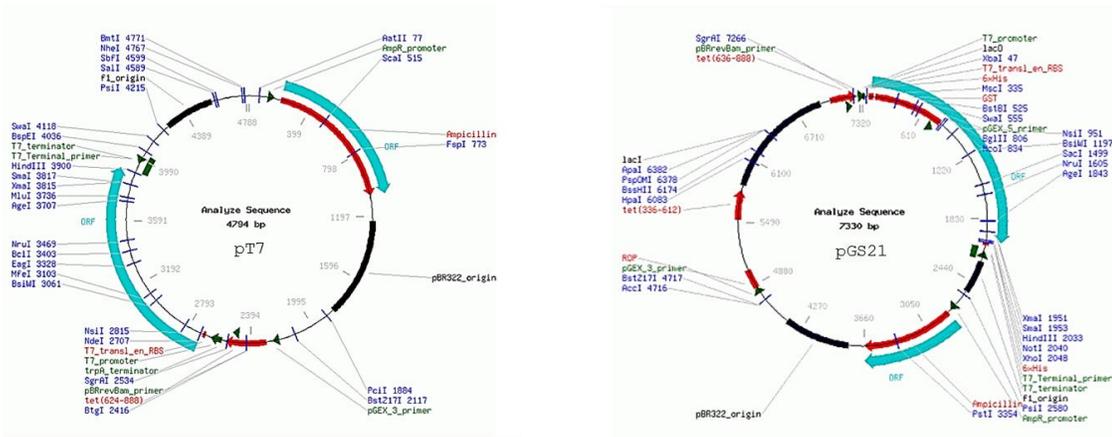
The coding sequence of the LadyB chimeric nanopore is reported in Figure 3.30 and the plasmid used for protein expression are reported in figure 3.31

```
ATGGCGGACTCGGACATCAACATCAAAACGGGCACGACGGACATCGGTTCCAATACGACG
GTGAAAACGGGCGACCTGGTGACCTATGATAAAGAAAACGGCATGCATAAAAAGGTGTTT
TATAGCTTCATCGATGACAAAACCAATAAAAAGCTGCTGGTGATTCGTACCAAGGGC
ACGATCGCGGGCCAGTATCGCGTTTACTCAGAAGAAGGCGCCAATAAATCAGGTCTGGCA
TGGCCGTCGGCTTTTAAGGTCCAGCTGCAACTGCCGGATAACGAAGTGGCGCAAATTAGC
GACTATTACCCGCGTAACTCTATCGATACCAAGGAATACATGAGTACCCTGACGTACGGC
TTCAACGGTAAATGTTACCGGCGATGACACGGGTAAAATTGGCGGTCTGATCGGCGCCAAC
GTTAGTATTGGTCATACCCTGAAATACGTCCAGCCGGACTTTAAGACCATCCTGGAATCC
CCGACGGATAAAAAGGTTGGCTGGAAAGTCATTTTCAACAATATGGTCAACCAGAATTGG
GGTCCGTATGATCGTGACAGCTGGAACCCGGTGTACGGCAATCAACTGTTTATGAAAACC
CGAACGGTTCATGAAGGCGGCCGATAACTTCCTGGACCCGAATAAAGCGAGCTCTCTG
CTGAGTTCGGCTTTTACCAGGACTTCGCAACCGTGATCACGATGGATCGTAAAGCGAGC
AAGCAGCAAACCAATATTGATGTGATCTATGAACGTGTTTCGCGATGACTACCAGCTGCAT
TGGACCTCTACGAACTGGAAGGTACCAATACGAAGGATAAGTGGACCGACCGTTCATCG
GAACGCTACAAGATCGATTGGGAAAAGGAAGAAAT
```

```
GGTGGCTCTAAGTGGTGGCTCTAGCGGTGGTAGCTCCCGCTCGCGTCTGCTGGAAGACTT
TCGT
```

```
GATATCGGTGGCTGCTCTAGCGGCGGCAGCTCTGGCGGTAGTTCCGGCGGTTCATGTCCTG
TCTCTGGCGCTGCAGATGTATGGTTGCCGTGTGATTCAAAAAGCCCTGGAATTTATCCCG
TCAGATCAGCAAAACGAAATGGTGCGCGAACTGGACGGCCAGGTTTTTCGCGCTGTGACC
CACCCGTACGGTTGTCGTGTTATTCAACGCATCCTGGAACATTTGCTGCCGGATCAAACG
CTGCCGATTCTGGAAGAACTGCATGGCGGTTGTGGTGCCTAAAAGCTT
```

**Figure 3.30** Nucleic acid sequence of  $\alpha$ HL-(PUF)<sub>2</sub> chimeric protein.  $\alpha$ HL(wt)-(GS)<sub>12</sub> $\alpha$ HLGGC(GS)<sub>12</sub>(PUF)<sub>2</sub>C<sub>74</sub>. Alpha is an alpha helix cap domain present in pumilio proteins with a protective function. wt- $\alpha$ HL is highlighted in green, violet the SG linker, blue the PUF dimer. Two Cys were placed in position 323 and 409 of the amino acid sequence.



**Figure 3.31 Plasmids.** pT7 (left) and pGS21 (right) plasmids. Open reading frames (ORF) are represented by blue arrows.

### 3.4.2 *In vivo* protein expression and purification

(His)<sub>6</sub>-GST fusion proteins (pGS-21a) were expressed in *E.Coli Nico21* (NEB). A starter culture of 30 ml LB media (Sigma) with ampicillin (100 µg/ml) was incubated for 12 h at 37°, 200 rpm. The days after 15 ml of this culture were added to a 1L LB-amp (100ug/ml). Once reaching an OD<sub>600</sub> of 0.7, cells were induced with IPTG (0.2 mM). The expression was carried out for 20-22 hours at 28°C, 200 rpm. Cell suspension was pellet down by centrifugation at 8000 g for 20 min at 4°C. The pellet (~ 2.5 g for 50 ml suspension) was resuspended in 5 mL (for 2.5g pellet) of lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>; 500 mM NaCl; 10 mM imidazole; 0.1 % TRITON, pH 8, filtered through 0.45 µm filter), added of protease inhibitor cComplete Ultra without EDTA (1 tablet for 100ml buffer, Roche), 1 mg/mL lysozyme (Sigma) and 1 uL/ml of Benzonase (250 U/ul, Novagen). The suspension was sonicated in ice with a pulse sonication (40 s at 300 W for two times) and centrifuged at 16000 g for 1 hour at 4°C. The supernatant was loaded in a gravity columns (Econo-Column®, Biorad) and the protein purified by cobalt resin affinity chromatography (His-Select Cobalt affinity gel, Sigma) equilibrated with 3-5 bed volumes of a washing buffer containing 300 mM NaCl, 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 20 mM imidazole, pH 8 (filtered through 0.45 µm filter). After loading the sample, the cobalt column was washed with 5 bed volumes of the washing buffer before eluting the chimeric protein with the elution buffer

(50 mM NaH<sub>2</sub>PO<sub>4</sub>; 300 mM NaCl; 250 mM imidazole, pH 8, filtered through 0.45 μm filter). Fractions of 1 mL were collected and analyzed by Nanodrop at 280 nm absorbance. The fractions were collected until any peak at 280 nm was detected (Nanodrop measurement). Approximately three bed volumes of elution buffer were used. Elution fractions underwent immediately concentration and buffer exchanged in a 50 kD cut-off filter (MICROCON, Millipore).

### **3.4.3 Enterkinase digestion**

The digestion of the monomeric (His)<sub>6</sub>-GST was carried out for ~ 16 hours at 22°C in the follow conditions: 1 μL of (His)<sub>6</sub>-enterokinase (Genescript) and 2 uL of 10 X cleavage buffer (Genescript) were mixed with 15 μL protein concentrate in buffer containing 100 mM NaCl, 25 mM Tris-HCl in a final volume of 20 μL. The day after the solution was reloaded on His-Select Cobalt. The (His)<sub>6</sub>-GST tag stacked on the cobalt resin and the purified proteins were concentrated with a 3 kD cut-off filter (MICROCON, Millipore). The elution buffer was exchanged with a buffer 25 mM Tris-HCl and 100 mM NaCl on the same cut-off filter. Proteins were stored at -80°C in 25% glycerol (Sigma). The yield and purity were checked by Nanodrop spectroscopy, SDS-PAGE and LC-MS after trypsin digestion.

### **3.4.4 Hemolytic assay**

Monomers obtained by *E.Coli* expression and *in vitro* transcription-translation were subjected to two-fold serial dilutions across the 12-well row of a microtiter plate to a final volume of 50 μL/well in MOBS buffer (10 mM MOPS, titrated to pH 7.4 with NaOH, 150 mM NaCl, 1 mg/mL bovine serum albumin). An equal volume of 1 % washed rabbit erythrocytes suspension (optical density = 0.1 at 595 nm) in MOBS buffer was quickly added to each well, starting with the most dilute sample. Hemolytic activity was recorded for 1 h by monitoring the decrease in light scattering at 595 nm with a Bio-Rad micro-

plate spectrophotometer, using the Microplate Manager 5.2 software. The percentage of hemolysis (%HA) was calculated (1) as

$$\%HA = 100 \cdot (A_i - A_{ib}) / A_w \quad (1)$$

$A_i$  is the absorbances of the sample,  $A_{ib}$  is the absorbance without lysis (only buffer) and  $A_w$  is the absorption after complete lysis (triton 0.05%, positive control). Experiments were run in triplicates.

### 3.4.5 *In vitro* protein expression

Proteins expression *in vitro* were prepared by coupled *in vitro* transcription and translation (IVTT) using an E. coli T7-S30 expression system for circular DNA (Promega) as reported in the technical bulletin. Briefly, for a 25  $\mu$ L reaction, premix solution (10  $\mu$ L) was combined with the amino acid mixture minus methionine (2.5  $\mu$ L), plasmid DNA (4  $\mu$ L, 400 ng/ $\mu$ L), T7 S30 extract (7.5  $\mu$ L) and [<sup>35</sup>S]L-methionine (1  $\mu$ L, MP Biomedicals, 1174 Ci/mmol, 10 mCi/mL). The reaction was incubated at 37° C for 45min and then centrifuged for 10min at 4°C at 25,000 x g to pellet any insoluble proteins. The T7-S30 extract was supplemented with rifampicin (1  $\mu$ L of 500  $\mu$ g/mL rifampicin per 150  $\mu$ L reaction). For gel electrophoresis 1:2 volume of loading buffer (Biorad) was added to the sample analysis and run on polyacrylamide gels in Tris-glycin SDS running buffer (TGS, Biorad).

### 3.4.6 Chimera oligomers preparation.

Recombinant chimeric monomers expressed *in vivo* were oligomerized on RRBCM in TRIS/NaCl buffer for 30 min at 37°C. The proteins were resuspended in 400  $\mu$ L buffer (Tris/NaCl) and then concentrated by centrifugation for 10min at 4°C at 25,000 x g using 100kD cut-off filter (MICROCON, Millipore), frozen in liquid nitrogen and stored at -80°C in 25% glycerol (Sigma). In order to enhance the formation of disulfide bridge be-

tween Lsm mutant monomers, the chimeric oligomers were oxidized with Cu (o-phenanthroline)<sub>2</sub>. Cu (o-phenanthroline)<sub>2</sub> was prepared as a mixture of CuSO<sub>4</sub> and o-phenanthroline in a ratio 1:3.5 mol/mol. The oligomers were incubated with Cu (o-phenanthroline)<sub>2</sub> for 2 hs on ice (1.5 mM final concentration) before start the experiment in single channel recording.

Gel extraction of chimeric oligomers expressed *in vitro* was performed using the standard WT- $\alpha$ HL protocol. Briefly, recombinant chimeric monomers expressed *in vitro* were assembled on RRBCM, loaded onto PAGE gel (8%) and subjected to electrophoresis at +80 mV for 2 h. The protein oligomer bands were cut from the gel and hydrated in pure water (~500  $\mu$ L). The gel was then crushed using a pestle and the slurry was filtered through once with 0.2  $\mu$ m filter (BD) to remove gel residues and the flow through concentrated in a 100 kD cut-off filter (MICROCON, Millipore). The protein solution was stored at -80°C in 25% glycerol.

#### **3.4.7 Radioactive labeling of ssRNAs.**

ssRNA U<sub>20</sub> oligos were radioactive labeled in 5' by P<sup>33</sup> using T4 Polynucleotide Kinase (T4 PNK, Promega). The reaction was carried out in a 6  $\mu$ L volume using 10 units of T4 PNK and 50 nmol of RNA. The reaction was incubated at 37°C for 1 hour and purified from unincorporated nucleotide and residual proteins with miRNeasy Mini Kit (Quiagen) as described in the technical bulletin. The labeling efficiency was checked by running 1:2 oligo dilutions in TBE-PAGE urea gel (15% - 6M urea, Biorad). After electrophoresis RNAs were visualized by autoradiography.

#### **3.4.8 EMSA assay**

Starting concentrations of 2  $\mu$ M for recombinant/purified AV-Lsm monomers (equal to 0.28  $\mu$ M of Lsm homoheptameric complexes) and 0.56  $\mu$ M of chimera oligomers were used (assuming 50% oligomerization efficiency). Progressive protein dilutions were incubated with 10 pmol of <sup>33</sup>P-labeled RNAs in a 40  $\mu$ l reaction for 15 min at 30°C in a bind-

ing buffer containing 100 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 12 mM HEPES/KOH 10 % glycerol, 0.1 % Triton X-100, and 2 μL/ml RNAse inhibitor (Neb), pH 7.6. In the absence of proteins, the binding buffer was used to compensate the volume. Samples were electrophoresed in 0.5 X TBE buffer on a native 5% polyacrylamide gel (acrylamide:N,N<sub>9</sub>-methylene bisacrylamide 80:1, in 0.5 X TBE) for 3 h at 7.6 volts/cm at 4°C. After electrophoresis, the gel was dried and bands visualized by autoradiography. The dissociation constants (K<sub>D</sub>) was calculated by densitometry analysis on the bands using the ImageQuant software (Biorad). The bound RNA was divided by the amount of free RNA and the fraction of RNA bound vs protein concentration was reported. A Hill function

$$y = V_{\max} * x^n / (k^n + x^n) \quad (2)$$

was fitted and the concentration of protein that bound the 50% of RNA was extrapolated.

### 3.4.9 Single channel recordings

See Chapter 2.

### 3.4.10 Data analysis

All the traces were recorded and analyzed in the same condition as in chapter 2.

The reverse potential ( $V_{\text{rev}}$ ) in planar lipid bilayer under asymmetric condition of ionic force was measured creating a KCl gradient through the membrane (Fig. 2c). The experiment started in symmetric concentration in both chambers (buffer 1: 100 mM KCl, 10mM Tris and 0.1 mM EDTA, pH 8.0). After recording the zero  $V_{\text{rev}}$ , the buffer was exchange (10 μL at time) of buffer 1 with a buffer 2 (buffer 2: 3M KCl, 10 mM Tris, and 0.1 mM EDTA, pH 8.0) in the *cis* side of the chamber. After each exchange we measure the voltage ( $V_{\text{rev}}$ ) between the two electrodes at zero potential applied. The junction poten-

tial at any contribution from the electrodes were measured at the end of the experiment and subtracted from the reported  $V_{rev}$ . From the reversal voltage, the permeability ratio ( $P_{K^+}/P_{Cl^-}$ ) was calculated by using the Goldman-Hodgkin-Katz (GHK) equation<sup>32</sup> (2):

$$P^{+}/P^{-} = \left( \frac{[a]_t}{[a]_c} \exp^{(eV_{rev}/FT)} - 1 \right) / \left( \frac{[a]_t}{[a]_c} - \exp^{(eV_{rev}/FT)} \right) \quad (3)$$

Where  $R$  is the gas constant,  $T$  the absolute temperature,  $F$  the Faraday constant,  $e$  is the elementary charge (at 23°C  $FT/e$  is 25 mV) and  $[a]_t$  and  $[a]_c$  are the KCl activities in the *cis* and *trans* solution respectively.

### 3.4.11 Gel trypsin digestion for ESI-MS.

The monomeric protein (2-5 mg/ml) was run on SDS-PAGE gel (MiniProtean 4-12 % Bis-Tris, Biorad). The gel band corresponding to the protein was visualized by Coomassie staining (Sigma). A clean scalpel was used to excise the band corresponding to the predicted molecular weight. The gel slice was washed in 100  $\mu$ L 25 mM ammonium bicarbonate in acetonitrile (HPLC pure, Sigma) for 30 min. This step was repeated with the same solution until band was no longer a strong blue color. Then we wash the band in 100  $\mu$ L acetonitrile for 10 min, removed it, and dry it in SpeedVac for 10 min. The sample was incubated for 30 min at 37 °C in 100  $\mu$ L of DTT water solution (10 mM), with sequential washing in 25 mM ammonium bicarbonate solution. Washing was repeated until gel pieces become white and then we incubated the sample in 100  $\mu$ L iodoacetamide (55 mM in 25 mM ammonium bicarbonate solution) in the dark for 60 min. Then, after washing with 100  $\mu$ L of 25 mM ammonium bicarbonate in acetonitrile and in 100  $\mu$ L pure acetonitrile, we dry the sample in SpeedVac for 10 min. Finally we add 2 $\mu$ L trypsin (Promega) in 25 mM ammonium bicarbonate solution to the tube for overnight digestion at 37 °C. The day after 1  $\mu$ L of formic acid was added to stop digestion and the supernatant containing peptides was removed to a clean new tube. Digested samples were submitted for tandem

mass spectroscopy analysis. Data were visualized by the MASCOT server (Oxford University).

### **3.4.12 MALDI-MS**

Resin purified protein was collected in a final volume of 0.5 mL. A small aliquot of this was mixed in 1:1 ratio with sinapinic acid 10mg/ml (Sigma) dissolved a solution of  $C_2H_3N/0.1\%$  trifluoroacetic acid (4:6). From this combined matrix/sample solution, 2  $\mu$ L was spotted onto a steel target and allowed to co-crystallize at room temperature for 30 minutes. MALDI analysis was conducted on a Water MALDI Micro MX spectrometer with TOF detection, in positive linear mode. A standard protein mixture (Invitrogen) was used as a lock mass calibrant. Laser energy and pulse width were optimized. Data were further processed using Mass Lynx 4.1.

### 3.5 References

1. Ray, D. *et al.* A compendium of {RNA-binding} motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
2. Mackay, J. P., Font, J. & Segal, D. J. The prospects for designer single-stranded RNA-binding proteins. *Nature structural & molecular biology* **18**, 256–61 (2011).
3. Musco, G. *et al.* The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nature structural biology* **4**, 712–6 (1997).
4. Kotera, E., Tasaka, M. & Shikanai, T. A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* **433**, 326–30 (2005).
5. Blackshear, P. J. *et al.* Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates. *The Journal of biological chemistry* **278**, 19947–55 (2003).
6. Wilusz, C. J. & Wilusz, J. Eukaryotic Lsm proteins: lessons from bacteria. *Nature structural & molecular biology* **12**, 1031–6 (2005).
7. Filipovska, A. & Rackham, O. Modular recognition of nucleic acids by PUF, TALE and PPR proteins. *Molecular bioSystems* **8**, 699–708 (2012).
8. Chen, Y. & Varani, G. Protein families and RNA recognition. *The FEBS journal* **272**, 2088–97 (2005).
9. Auweter, S. D., Oberstrass, F. C. & Allain, F. H.-T. Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic acids research* **34**, 4943–59 (2006).
10. Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489–4 (2008).
11. Pennisi, E. The tale of the TALEs. *Science (New York, N.Y.)* **338**, 1408–11 (2012).
12. Koh, Y. Y. *et al.* Stacking interactions in {PUF} – {RNA} complexes. **1**, 718–727 (2011).
13. Wilusz, C. J. & Wilusz, J. Lsm proteins and Hfq: Life at the 3' end. *RNA biology* **10**, 592–601 (2013).
14. Scofield, D. G. & Lynch, M. Evolutionary diversification of the Sm family of RNA-associated proteins. *Molecular biology and evolution* **25**, 2255–67 (2008).
15. Murina, V. N. & Nikulin, a D. RNA-binding Sm-like proteins of bacteria and archaea. similarity and difference in structure and function. *Biochemistry. Biokhimiia* **76**, 1434–49 (2011).
16. Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. & Séraphin, B. Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *The EMBO journal* **18**, 3451–62 (1999).
17. Zaric, B. *et al.* Reconstitution of two recombinant LSm protein complexes reveals aspects of their architecture, assembly, and function. *The Journal of biological chemistry* **280**, 16066–75 (2005).
18. Collins, B. M. *et al.* Crystal Structure of a Heptameric Sm-like Protein Complex from Archaea: Implications for the Structure and Evolution of {snRNPs}. 915–923 (2001). doi:10.1006/jmbi.2001.4693

19. Mura, C., Cascio, D., Sawaya, M. R. & Eisenberg, D. S. The crystal structure of a heptameric archaeal Sm protein: Implications for the eukaryotic snRNP core. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5532–7 (2001).
20. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nature reviews. Microbiology* **9**, 578–89 (2011).
21. Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S. a & Brennan, R. G. Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract. *Nucleic acids research* **40**, 11023–35 (2012).
22. He W. and Parker R.\_2000.pdf.
23. Schümperli, D. & Pillai, R. S. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cellular and molecular life sciences : CMLS* **61**, 2560–70 (2004).
24. Tang, W., Kannan, R., Blanchette, M. & Baumann, P. Telomerase RNA biogenesis involves sequential binding by Sm and Lsm complexes. *Nature* **484**, 260–4 (2012).
25. Fischer, S. *et al.* The archaeal Lsm protein binds to small RNAs. *The Journal of biological chemistry* **285**, 34429–38 (2010).
26. Moll, J. M., Sobti, M. & Mabbutt, B. C. The Lsm Proteins: Ring Architectures for RNA Capture. (2008).
27. Mura, C. & Randolph, P. S. Archaeal and eukaryotic homologs of Hfq. *RNA Biology* **10**, 636–651 (2013).
28. Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *The EMBO journal* **29**, 4172–84 (2010).
29. Törö, I. *et al.* RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *The EMBO journal* **20**, 2293–303 (2001).
30. Naidoo, N. *et al.* Crystal structure of Lsm3 octamer from Saccharomyces cerevisiae: implications for Lsm ring organisation and recruitment. *Journal of molecular biology* **377**, 1357–71 (2008).
31. Hartman, A. L. *et al.* The complete genome sequence of Haloferax volcanii {DS2}, a model archaeon. *PloS one* **5**, e9605 (2010).
32. Gu, L. Q. *et al.* Reversal of charge selectivity in transmembrane protein pores by using noncovalent molecular adapters. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3959–64 (2000).
33. Menestrina, G. Ionic channels formed by Staphylococcus aureus alpha-toxin: voltage-dependent inhibition by divalent and trivalent cations. *The Journal of membrane biology* **90**, 177–90 (1986).
34. Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. Crystal structures of the Pyrococcus abyssi Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. *The Journal of biological chemistry* **278**, 1239–47 (2003).
35. Meller, A., Nivon, L. & Branton, D. Voltage-Driven DNA Translocations through a Nanopore. *Physical Review Letters* **86**, 3435–3438 (2001).
36. Maglia, G., Restrepo, M. R., Mikhailova, E. & Bayley, H. Enhanced translocation of single DNA molecules through alpha-hemolysin nanopores by manipulation of internal charge.

*Proceedings of the National Academy of Sciences of the United States of America* **105**, 19720–5 (2008).

37. Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E. & Deamer, D. W. Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules. *Biophysical journal* **77**, 3227–33 (1999).
38. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13770–3 (1996).
39. Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Accounts of Chemical Research* **35**, 817–825 (2002).
40. Deamer, D. W. & Akeson, M. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in biotechnology* **18**, 147–151 (2000).
41. Vinkenborg, L. & Famulok, M. aptamers for allosteric regulation. *Nature Chemical Biology* **7**, 519–527 (2011).
42. Flinders, J. *et al.* Recognition of Planar and Nonplanar Ligands in the Malachite Green ± {RNA} Aptamer Complex. **02119**, 62–72 (2004).
43. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415 (2003).
44. Japrun, D., Henricus, M., Li, Q., Maglia, G. & Bayley, H. Urea facilitates the translocation of single-stranded DNA and RNA through the alpha-hemolysin nanopore. *Biophysical journal* **98**, 1856–63 (2010).
45. Lunde, B. M., Moore, C. & Varani, G. RNA-binding proteins: modular design for efficient function. *Nature reviews. Molecular cell biology* **8**, 479–90 (2007).
46. Macdonald, P. M. The *Drosophila pumilio* gene: an unusually long transcription unit and an unusual protein. *Development (Cambridge, England)* **114**, 221–32 (1992).
47. Wang, X., Zamore, P. D. & Hall, T. M. Crystal structure of a *Pumilio* homology domain. *Molecular cell* **7**, 855–65 (2001).
48. Wang X. *et al.*, 2002.pdf.
49. Filipovska, A., Razif, M. F. M., Nygård, K. K. a & Rackham, O. A universal code for RNA recognition by PUF proteins. *Nature chemical biology* **7**, 425–7 (2011).
50. Dong, S. *et al.* Specific and modular binding code for cytosine recognition in *Pumilio*/FBF (PUF) RNA-binding domains. *The Journal of biological chemistry* **286**, 26732–42 (2011).
51. Ozawa, T., Natori, Y., Sato, M. & Umezawa, Y. Imaging dynamics of endogenous mitochondrial RNA in single living cells. **4**, 413–419 (2007).
52. Wang, Y., Cheong, C.-G., Hall, T. M. T. & Wang, Z. Engineering splicing factors with designed specificities. *Nature methods* **6**, 825–30 (2009).



## Concluding Remarks

In the first chapter we studied the host-pathogen interaction using a non-immune response model system formed by human cell lines and the staphylococcal  $\alpha$ HL, to understand the fast and dynamic translational regulation of gene expression on human cell lines in response to sublytic doses of the pore forming toxin.

First, we defined a window of sublytic concentration for each cell line under study (CHP-134, SH-SY5Y and HeLa) and we addressed the effect of the native  $\alpha$ HL on protein factors affecting translational initiation (i.e. eIF4B, RPS6K, RPS6, 4EBP, eIF4E, Akt). Interestingly, the effect was not uniquely depressive on the protein synthesis machinery. We described an opposite effect characterized by activation (eIF4B, RPS6K, RPS6, eIF4E, Akt) and a concomitant inactivation (4E-BP) of some of these key proteins; only on cell lines with a high level of cholesterol and sphingomyelin. In fact, we observed that the depletion of cholesterol from the cell membrane block the effect of the  $\alpha$ HL. In accordance, the *de novo* protein synthesis was slightly depressed but the ribosome up-loading onto polysomes was unaffected. We then performed a genome-wide analysis of the changes occurring at the transcriptional and translational level. We observed a strong uncoupling between transcription and translation (both on DEGs and on ontological enriched terms) and by the emergence of numerous genes upregulated or downregulated purely at the translational level, confirming that the expression was selectively regulated only at the translational level, while the transcriptional level was poorly modified.

Finally, using a non-lytic mutant ( $\alpha$ HL-H35N) we found that the activation of translation effectors and of the RNA binding protein ELAVL1 were specifically due to the binding on the plasma membrane of the monomeric protein. On the other hand, the depression of translational effectors was pore-dependent, in agreement with previously published data. These results suggest a different cell reaction depending on mono-

mers/oligomer proportions and that in our condition the membrane bound and the few orphan pores were responsible for the observed opposite stimuli on translational control of gene expression.

All the above described findings demonstrate the discovering of an exclusive response to a virulent sublytic attack of the native  $\alpha$ HL and this will help to better understand the physiopathology of *Staphylococcus aureus* infection in vivo.

In the second chapter we used the  $\alpha$ HL pore to develop a stochastic sensor for the widespread post-transcriptional modification, 3'-end uridylation. We demonstrated the selective, transient and single molecule modular binding of 3' poly-uridylated ssRNA inside the  $\beta$ -barrel of the WT- $\alpha$ HL pore. In particular, using a planar bilayer system, in single channel electrophysiological recording, we characterized the kinetics and the structural features of the RNA/pore interaction. Our results showed the ability of the nanopore to discriminate the length of the 3' U-tail and the presence of non-canonical nucleobases, providing a possible biotech approach and a powerful unique single molecule technique for the detection of this specific RNA signature. Furthermore, we set up a protocol for the purification and the detection with Droplet Interface Bilayers of specific 3' oligo U fragments from longer RNA, giving a possible practical solution for the processing of biological samples in the near future. In conclusion, we described surprising properties of this common nanopore tool, showing that the  $\alpha$ HL pore can be used as a simple, reliable and efficient RNA binding nanopore.

In the third chapter we created new chimeric proteins genetically fusing the pore forming  $\alpha$ HL to the RNA binding protein Lsm or the RNA binding domain Pumilio.

With this approach we tried to increase the complexity, the accuracy and the strength of the binding recognition at the single molecule level, with the aim to design custom

RNA binding nanopores (RBnPs) for any type of RNA sequence, expanding the potential of RNA-sequence signature recognition. The fusion constructs comprised: i)  $\alpha$ HL pore forming activity and the ii) RNA binding protein affinity for the ssRNA in bulk solution. We wondered if both these properties were affected by the fusion. We then characterized the stability of each chimeric protein in single channel planar bilayer. We found that a Lsm-chimeric protein (LadyB), characterized for basic electrophysiological properties (e.g. conductance, rectification factor, ion selectivity), was able to form pores and bind oligo U RNAs. We showed peculiar differences with the WT- $\alpha$ HL pore that underlined by our studies on different RNAs. The recognition ability of Lady was different respect to the WT pore for the sensing of 3'-uridylated ssRNAs in term of dwell time, frequency of the interactions and strength of the binding, and found to be more efficient in some cases. In conclusion, we demonstrated that is indeed feasible to take up new directions in the constantly renewing field of “nanopore sensors” and the development of new single molecule RNA sensing strategies is extremely promising.