Purification of UV cross-linked RNA-protein complexes by phenol-toluol extraction

DISSERTATION

zur Erlangung des akademischen Grades:

Doctor rerum naturalium

(Dr. rer. nat)

eingereicht an der
Lebenswissenschaftlichen Fakultät der
Humboldt-Universität zu Berlin

von

Erika Cristina Urdaneta Zurbarán

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst.

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin:
Prof. Dr. Bernhard Grimm.

Gutachter:

Prof. Dr. Matthias Selbach
Prof. Dr. Leonie Ringrose
Dr. Benedikt Beckmann

Tag der mündlichen Prüfung: 14. Februar, 2020

Zusammenfassung

RNA-Bindungsproteine spielen Schlüsselfunktionen bei der post-transkriptionellen Regulation der Genexpression. Durch Bindung an RNA steuern sie die RNA-Aufbereitung, den Transport, die Stabilität und die Translation. In den letzten zehn Jahren wurden bedeutende Fortschritte bei der Aufklärung bakterieller post-transkriptioneller Mechanismen erzielt. Es wird immer deutlicher, dass diese Regulierungsebene auch bei der Pathogenese und Antibiotikaresistenz eine wichtige Rolle spielt. Die Analyse von RNA-Protein-Komplexen (RNPs) auf Proteomebene wurde durch die *(m)RNA-interactome-capture* Technologie vorangetrieben, die den Teil des Proteoms isoliert, welcher mit polyadenylierter (m)RNA vernetzt ist. Dies hat zur Identifizierung von Hunderten von neuen RBPs in einer Vielzahl von eukaryontischen Arten, vom Menschen bis zur Hefe, geführt. Allerdings fehlt die Poly-Adenylierung in der funktionellen RNA von Bakterien und anderen Klassen von -eukaryontischen- regulatorischen RNAs. Ziel dieser Arbeit war es, diese Einschränkung durch die Entwicklung einer neuartigen und unvoreingenommenen Methode zur Aufreinigung von UV-vernetzten RNPs in lebenden Zellen zu überwinden: PTex (Phenol-Toluol-Extraktion). Das Reinigungsprinzip basiert ausschließlich auf den physikalisch-chemischen Eigenschaften von vernetzten RNPs gegenüber ungebundenen Proteinen oder RNA; es ist dabei unparteiisch gegenüber spezifischen RNAs oder Proteinen und ermöglicht somit erstmals eine systemweite Analyse von nicht-poly-(A)-RNA-interagierenden Proteinen sowohl in eukaryontischen (HEK293) als auch in prokaryontischen (*Salmonella* Typhimurium) Zellen.

Abstract

RNA binding proteins play key functions in post-transcriptional regulation of gene expression. By binding to RNA, they control RNA editing, transport, stability and translation. In the last decade, significant advances have been made in the elucidation of bacterial post-transcriptional mechanisms. It is becoming increasingly clear that this layer of regulation also plays an important role in pathogenesis and antibiotic resistance. The analysis of RNA-protein complexes (RNPs) at the proteome level has been driven by the (m)RNA interactome capture technology which isolates the proteome cross-linked to poly-adenylated (m)RNA. This has resulted in the identification of hundreds of novel RBPs in a diversity of eukaryotic species ranging from humans to yeast. However, poly-adenylation is absent in functional RNA from bacteria and other classes of -eukaryotic- regulatory RNAs. This work was aimed to overcome that limitation by developing a novel and unbiased method for the purification of UV-cross-linked RNPs in living cells: PTex (Phenol Toluol extraction). The purification principle is solely based on physicochemical properties of cross-linked RNPs versus unbound proteins or RNA, and it is impartial towards specific RNA or proteins; enabling for the first time a system-wide analysis of non-poly(A) RNA interacting proteins in both eukaryotic (HEK293) and prokaryotic (*Salmonella* Typhimurium) cells.

Declaration

Hereby, I declare that:

1. I do not hold a doctoral degree.
2. I am familiar with the regulations on which the doctoral procedure is based.
3. The work entitled ***"Purification of UV cross-linked RNA-protein complexes by phenol-toluol extraction"***, is an original report of the research I conducted at the Molecular Infection Laboratory, IRI Life Sciences, Humboldt-Universität zu Berlin, which has been written entirely by myself and has not been submitted for a different or similar degree at any other higher education institution in Germany or elsewhere.
4. To the best of my knowledge and intentions, all external sources have been appropriately referenced across the manuscript. The literature is provided in a dedicated chapter (Bibliography).
5. The experimental design as well as the interpretation of the data here generated were made by myself, with the support and guidance of my supervisor.
6. All experiments were executed by myself, except for the helpful contributions of lab members and external collaborators, all of which have been clearly stated within the text and to whom acknowledgments have been extended.
7. Parts of this work have been published (refer to Appendix).

Berlin, 9 December 2019

Erklärung

Hiermit erkläre ich das:

1. Ich habe keinen Doktortitel.
2. Ich bin mit den Vorschriften vertraut, auf denen das Promotionsverfahren basiert.
3. Die Arbeit mit dem Titel ***"Purification of UV cross-linked RNA-protein complexes by phenol-toluol extraction"***, ist ein Originalbericht der Forschung, die ich am Molecular Infection Laboratory, IRI Life Sciences, Humboldt-Universität zu Berlin, durchgeführt habe und die vollständig von mir selbst geschrieben wurde und die an keiner anderen Hochschule in Deutschland oder anderswo für einen anderen oder ähnlichen Grad eingereicht wurde.
4. Nach bestem Wissen und Gewissen wurden alle externen Quellen im Manuskript angemessen erwähnt. Die Literatur wird in einem eigenen Kapitel (Bibliographie) angeführt.
5. Das experimentelle Design sowie die Interpretation der hier generierten Daten wurde von mir selbst erstellt, mit der Unterstützung und Anleitung meines Vorgesetzten.
6. Alle Experimente wurden von mir selbst durchgeführt, mit Ausnahme der hilfreichen Beiträge von Labormitgliedern und externen Mitarbeitern, die alle im Text klar angegeben sind und auf die die Anerkennung ausgedehnt wurde.
7. Teile dieser Arbeit wurden veröffentlicht (siehe Anhang).

Berlin, 9. Dezember 2019

Dedication

In memoriam
of my beloved grandma Emma
and my friend Stela Domador

To all the women in my family and friends
kind, strong, resilient and persevering
your example motivates me every day

Acknowledgements

This story would not have started if it had not been for the ZIBI Summer School 2014, thank you Juliane Kofer!

To IRI-Life Sciences and IRI-Graduate School, especially to Stefanie Scharf.

To Benedikt Beckmann, who believed in me even when I was barely able to speak English. Thank you for your guidance, never-ending support, and all the memes that cheered me up when stress was taking its toll.

To my lab-mates from Beckmann and Reber labs, special thanks to Julie, Davide, Sebastian, and Abin. This time would have felt twice as long and half as fun without having you around.

This work was possible thanks to the support of a group of collaborators whom were always willing to share their expertise and resources:

- AG Selbach: Matthias Selbach and Carlos Vieira.
- AG Ohler: Uwe Ohler, Hans-Hermann Wessels, Michaela Kolbe, Ilija Bilic and Antje Hirsekorn.
- Granneman lab: Sander Granneman, Ira Iosub, Stuart McKellar, Liangcui Chu and Pedro Arade.
- AG Medenbach: Jan Medenbach and Rebecca Moschall.

To IRI-Life Sciences, EMBO, Joachim-Herz foundation, and to the Commission for the Advancement of Women (Humboldt-Universität zu Berlin) for the

financial support to attend courses, conferences, and scientific visits.

To my parents Emma and Nicanor.

I have been blessed with not one but three wonderful families: my Venezuelan family sending their love from all around the world. My German family who welcomed me with open hearts and arms; and my friends whom have become brothers and sisters. Thank you for your comfort and support.

Tío Alberto, aquí está lo prometido!

To my husband Matthias, thank you for your tireless support, unlimited comfort and generous love. You make my life brighter and my feet warmer!

Glossary and Acronyms

Glossary

Chaotropic agent	Chaotropic agents are co-solutes which can disrupt the hydrogen bonding network between water molecules and can thereby reduce the stability of the native state of proteins by weakening the hydrophobic effect [1].
Hydrophobic effect	Results from the propensity of water molecules to form many relatively strong hydrogen bonds with each other. The favorable self-interaction of water produces a strong tendency for a protein to bury those parts of its surface that are not sufficiently hydrophilic, i.e., that are not themselves polar enough to replace one of the waters in a water-water hydrogen bond causing the collapse of the protein backbone into a dense globule and the burial of hydrophobic amino acid side chains in the core of the protein [2].
Intensity-based absolute quantification (iBAQ)	The sum of peak intensities of all peptides matching to a specific protein, divided by the number of theoretically observable peptides. iBAQ is used as an accurate proxy for protein levels [3].
pK_a	Is the negative base 10 logarithm of the acid dissociation constant (K_a) of a solution; pK_a= -Log$_{10}K_a$ (https://www.thoughtco.com/). This magnitude quantifies the tendency of molecules to dissociate when in contact with water.

Ultraviolet (UV) radiation	Ultraviolet radiation covers the wavelength range of 100 nanometres (nm) to 400 nm. It is the most powerful type of optical radiation. UV radiation is not visible to the human eye and cannot be perceived by the other senses either. UV radiation is divided into the following wavelength ranges: UV-A (400-315 nm wavelength), UV-B (315-280 nm wavelength) and UV-C (280-100 nm wavelength) http://www.bfs.de/EN/topics/opt/uv/introduction/introduction_node.html.

Acronyms

4-SU 4-thiouridine. 8, 9, 36, 58, 74, 88

6-SG 6-thioguanosine. 9

BCP 1,3-bromo-chloro-propane. 13, 45

clHuR cross-linked HuR. xxiii, 42–44, 47, 51, 52

CLIP cross-linking and immunoprecipitation. 7, 9

clRNP UV cross-linked RNA-protein complex. 17, 73

clRNPs UV cross-linked RNA-protein complexes. xx, xxi, xxiv, 12, 19, 34, 41, 42, 44, 45, 47, 49, 51–54, 57, 75, 76

EJC exon junction complex. 1

GO gene ontology. 36

HuR*RNA HuR-RNA complexes. 43, 46, 47

iBAQ intensity-based absolute quantification. 68

IDRs intrinsically disordered regions. 6, 85

LFQ label-free quantification. 35, 36, 54

MS mass spectrometry. xx, 10, 17, 34, 35, 53, 54, 69, 75, 76, 79

pI isoelectric point. 56, 60

PNK T4 polynucleotide kinase assay. 70

RBD RNA-binding domain. 1, 4

RBDs RNA-binding domains. xxi, xxv, 3, 4, 6, 7, 82, 83, 86

RBP RNA-binding protein. 1, 4, 7, 42, 73, 78

RBPome RNA-bound proteome. 71, 89

RBPomes RNA-bound proteomes. 10

RBPs RNA-binding proteins. xix, 1, 3–7, 10, 41

RIC (m)RNA-interactome capture. 10, 18, 44, 51, 52

RNPs ribonucleoprotein complexes. xxiii–xxv, 3, 7, 8, 12, 44, 57, 60, 74, 80, 89

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Ribonucleoprotein complexes and their role within the cell

Cells determine the final protein output of their genetic program by controlling the transcription, localization, translation and turnover rates of their mRNAs [4]. From transcription to decay, mRNAs are in constant contact with a diversity of proteins. The eukaryotic cell is the best example to illustrate how the life cycle of an mRNA is orchestrated by proteins (Fig. 1.1).

Ribonucleoprotein complexes are formed by the interaction of RNA molecules (mRNA and ncRNA), RNA-binding proteins (RBPs) and in many cases accessory proteins and/or metabolites with a variety of functions [5; 4]. From human to bacteria, these complexes are essential for cell processes such as protein synthesis, RNA stability, transport, localisation, mRNA/ncRNA activity and RNA decay [6; 7; 8; 4; 9] (Fig. 1.1. The ribosome, spliceosome, the exon junction complex (EJC), stress granules, RNase P and the exosome are classic examples of ribonucleoprotein complexes [7; 10; 11].

Conventionally, an RBP is a protein with the capacity of binding RNA molecules via a well defined RNA-binding domain (RBD), or a modular set of them [12]. RNA-binding proteins play different roles in the RNP by modulating or stabilizing the RNA structure, e.g. enabling its catalytic conformation during pre-mRNA editing [13; 5].

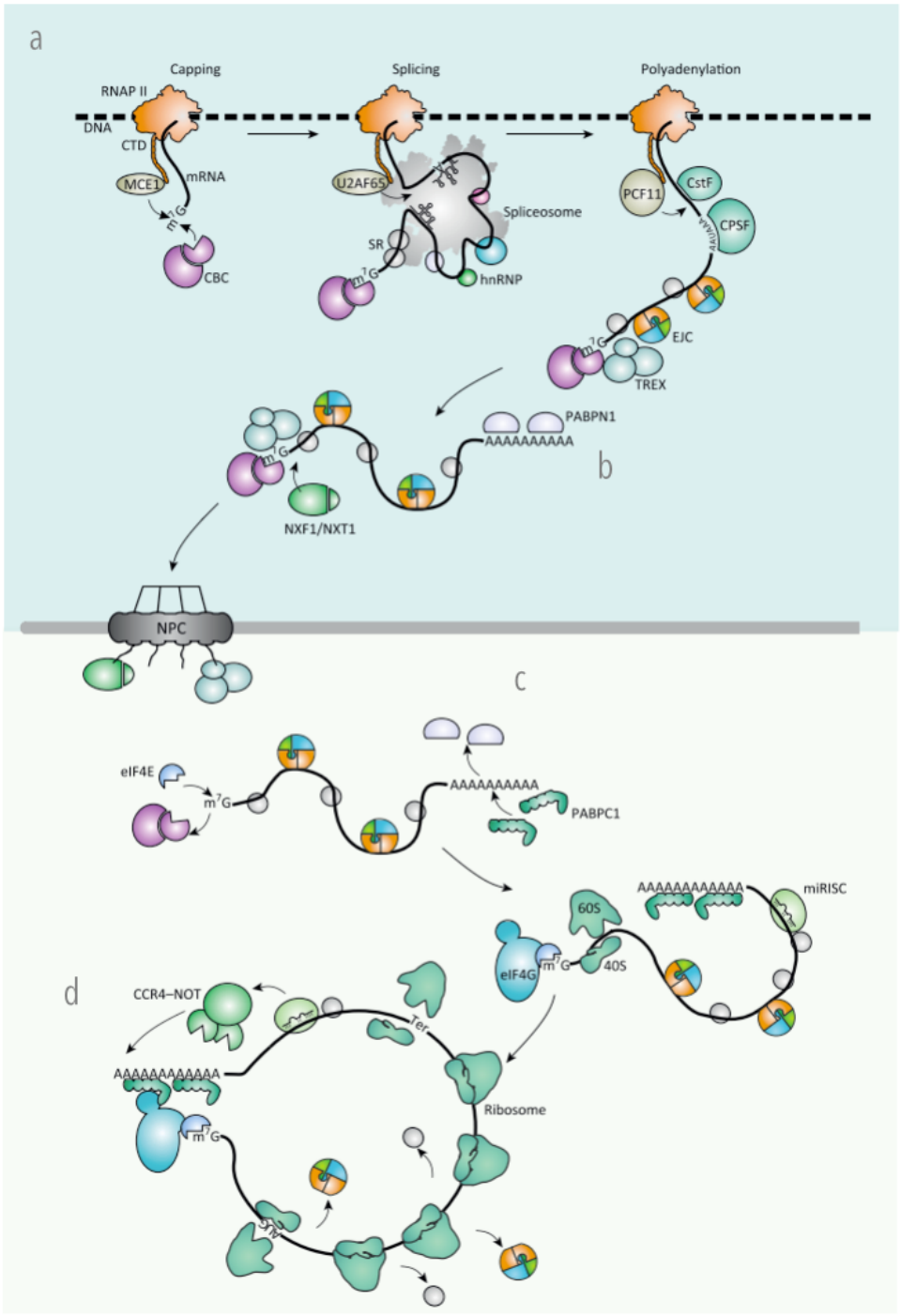

Figure 1.1: The life cycle of mRNA is orchestrated by a diversity of proteins. (a) during transcription by RNA polymerase II, nascent mRNA precursors encounter the capping machinery, cap which prevents a premature degradation. **(b)** pre-mRNAs are then spliced and polyadenylated, requisite for **(c)** their transport to the cytoplasm where mRNAs are **(d)** translated into proteins. Finally, cells can adjust to a changing environment by regulating the mRNA turnover, e.g. by disassembling mRNPs and degrading the mRNA. Modified from [4].

Recent technical advances have allowed the uncovering of hundreds of new RBPs, many of which exhibit unconventional RNA-binding domains and functions [14; 15; 16]. Likewise, the study of regulatory RNAs has also expanded, non-coding RNAs (ncRNAs) might guide the localization of RBPs, then regulating the translation and epigenetic modulation of genes [6; 17; 18]. These findings open-up the possibility that RBPs not just regulate RNA, but instead, can been regulated by it [12] (Fig. 1.2).

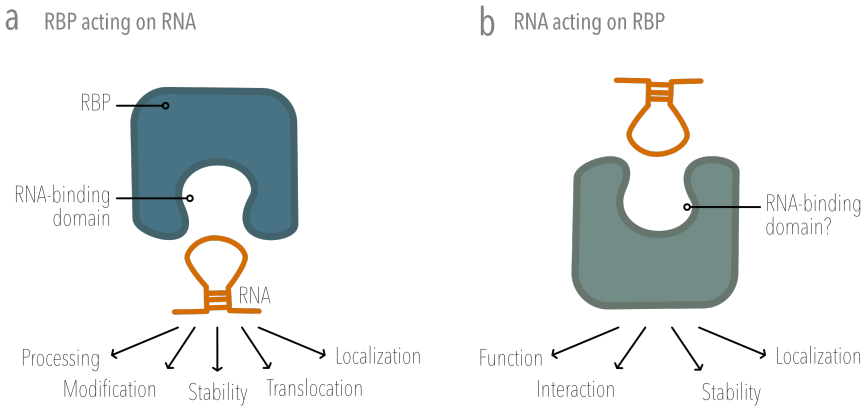

Figure 1.2: Functional cross-talk between proteins and RNA. (a) An RNA-binding protein (RBP) can interact with RNA through defined RNA-binding domains to regulate RNA metabolism and function. **(b)** Inversely, the RNA can bind to the RBP to affect its fate and function. Modified from [12].

1.1.1 Ribonucleoprotein complexes in disease and infection

Due the wide range of actions of RNPs, it is not surprising that perturbations on its functions are being associated with metabolic disorders, hereditary diseases and cancer [15; 19; 20]. It has been shown that mutations on RBPs can cause hereditary neuromotor diseases and other neuro-degenerative disorders, such as the amyotrophic lateral sclerosis (ALS), frontotemporal dementia, spinal muscular atrophy or Alzeimer's [21]. In addition, within the mRNA interactome from HeLa cells, 86 proteins were found to be involved in human Mendelian's diseases according to the OMIM database [15].

In the last decade, significant advances have been made in the elucidation of bacterial post-transcriptional mechanisms. It is becoming increasingly clear

that this layer of regulation also plays an important role in pathogenesis and antibiotic resistance [22; 23; 24; 9].

Pathogenic bacteria sense and respond to diverse microenvironmental stresses encountered during the infection process, leading a fine-tuned expression of virulence-associated genes. They rely on diverse molecular mechanisms to divert and ultimately highjack the immune response of the host as well as to resist the harsh intracellular environment.

Recently, an RBP from *Listeria monocytogenes* (protein Lmo2686 or Zea) has been found to be involved in modulating the IFN response in the host [25]. During infection of mammalian cells, Zea interacts with RIG-I and modulates RIG-I-dependent type I interferon (IFN) response. Interestingly, Zea is absent of any canonical RNA-binding domain (RBD), suggesting a non-canonical mode of RNA binding [25]. Despite the increasing interest on investigating RBPs in bacteria, the proteome-wide exploration of the RBP repertoire in bacteria has proved to be tremendously challenging due to technical difficulties (discussed later in this chapter).

1.2 RNA-binding proteins and their RNA-recognizing domains

The defining characteristic for an RNA-binding protein (RBP) is its capacity to bind RNA. The RNA-binding activity is mediated by an ever increasing variety of RNA-binding domains (RBDs) [7; 26]; such domains can also mediate protein-protein interactions and, in many cases, are subject to regulation by post-translational modification [7].

The specificity of RBPs is determined mainly by the type of the RNA-recognition domains (Fig. 1.3); combinations in these motifs determine the affinity of the RBP to its cognate RNA, without limitations in their versatility to assemble and disassemble in response to the cellular needs [27]. Interestingly, the RBD composition related to the general RBP architecture and even its localization has functional implications, e.g. the proteins Dicer and RNase III share the

Figure 1.3: Schematic representation of the distribution and increasing variety of RNA-binding domains. The number and distribution of RNA-binding domains varies greatly. Repetitions of the same or different motifs can be found within one RBP. Interestingly, the biological function and RNA specificity/affinity can differ among RBPs displaying the same RBDs (e.g. HuR, hnRNPL, PTBP1). (*) Although not strictly RBDs, catalytic domains are in close relationship with RNA-binding sites [26]. Modified from [27; 7].

display of an endonuclease catalytic domain followed by a double-stranded RNA-binding domain (dsRBD) (Fig. 1.3). However, although both proteins recognise dsRNA, only Dicer has evolved to interact specifically with RNA species that are produced in the RNA interference pathway through additional domains that recognise the unique structural features of these RNAs [27].

Analysis of the sequences and domain architecture of 120 yeast RBPs, revealed that these proteins contain an average of 1.9 domains, and a third harbor more than one [28]. On the other hand, the amount of RBPs in mammal and yeast cells might be an underestimation since only 40-50% of the RBPs contains well known RNA binding motifs [15; 28; 5].

During the analysis of the mRNA-interactome from the human cell lines HeLa and HEK293, characteristics of new RNA binding domains were established [14; 15; 19], particularly the double specificity of domains previously annotated as DNA or protein binders (SAF-A/ B, Acinus, PIAS and WD) and proteins without recognizable domains but with the common feature of having highly disordered regions [15].

A comprehensive analysis of RNA-binding domains (RBDmap) identified 1,174 binding sites within 529 HeLa cell RBPs, discovering numerous RNA-binding domains (RBDs) [26]. The authors reported that catalytic centers or protein-protein interaction domains are in close relationship with RNA-binding sites; and that nearly half of the RNA-binding sites map to intrinsically disordered regions (IDRs). They also found that RGG and YGG boxes are widely involved in RNA binding [26].

Recently, a comparison of evolutionarily conserved RNA interacting regions between human and fly applying the CAPRI technique (Crosslinked and Adjacent Peptides-based RNA-binding domain Identification), resulted in the confirmation and expansion of the RNA-binding capacity of intrinsically disordered regions (IDRs), classified as i) positively charged domains RG[G] or RS motifs, ii) aromatic amino acid-rich pattern with YG or NGF repeats, iii) single amino acid-repeat motifs containing glutamine, alanine or histidine stretches; additionally to the discovery of novel RNA-binding domains such as AAA ATPase, DZF, SPRY and P-loop containing nucleoside triphosphate hydrolase (Fig. 1.4) [29].

Intrinsically disordered regions (IDRs) have been found in conserved RBPs from yeast to humans, such regions harbor [K]- and [R]-rich tripeptide repeat motifs, which seem to be conserved across evolution [16]. This feature was used for the development of a support vector machine (SVM)-based method (TriPepSVM), for the classification of RBPs and non-RBPs [30]. Among the 990 predicted RNA-binders in humans, we found 200 proteins with ATP-binding capacity, 13 of which are proteins of the AAA ATPase family [30]. Interestingly, it has been shown that RNA- and single nucleotide binding overlapping occur in established RBPs [31]. Another finding was the enrichment in kinases, proteins harboring WD40 domains and bromodomain folds [30].

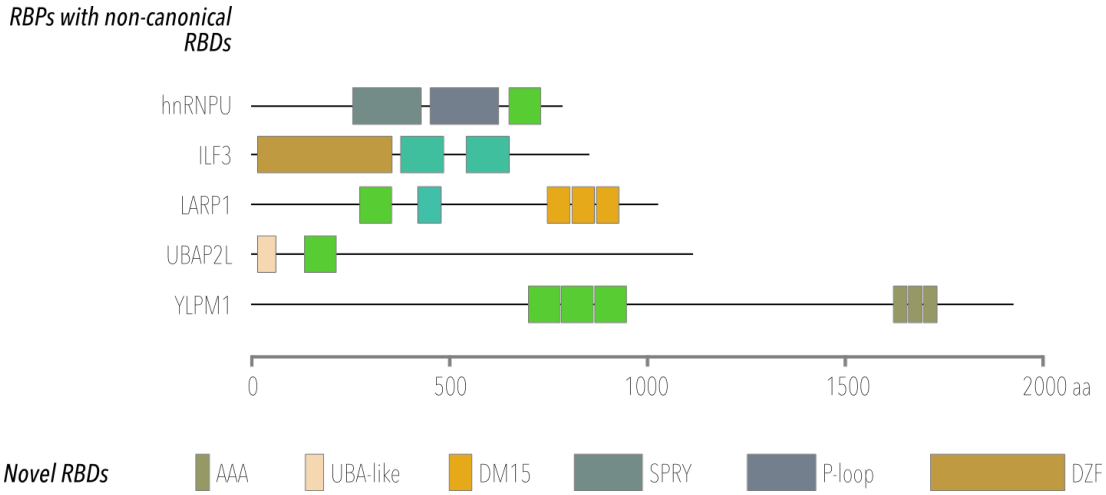

Figure 1.4: Novel RBDs conserved between human and fly. RNA-binding domains (RBDs) recently identified include AAA ATPase (AAA), domain associated with zinc fingers (DZF), SPla and RYanodine Receptor (SPRY), P-loop containing nucleoside triphosphate hydrolase (P-loop) among many other [29].

1.3 Methods to study ribonucleoprotein complexes

The increasing interest in studying the composition and function of ribonucleoprotein complexes (RNPs) has fostered the development of efficient protocols for RNPs purification and analysis. Currently, there are two main strategies to investigate RNPs, one is focused on analysing the transcriptome associated with a particular RBP (RNA-centric; Fig. 1.5a) and the other is aimed to purify and identify the repertoire of RNA-binding proteins in a cell-wide manner (Protein-centric; RNA-bound proteome or RBPome, Fig. 1.5b).

1.3.1 RNA-centric approaches

Tenenbaum and colleagues established a technique consisting in the immunoprecipitation of RNPs and analysis of the RNA by cDNA chips [33]. This technique allowed the identification of the RNA binding sites in Pumilo, miRNPs and AGO proteins [33; 34; 35]. However, the ability of RNA and proteins to associate non-specifically during cell lysis was reported [36].

To overcome these obstacle, the cross-linking and immunoprecipitation (CLIP) method [37] introduced the *in vivo* covalent bond formation between proteins and RNA by radiating cells with short-waved UV light (λ = 254 nm) [38], followed

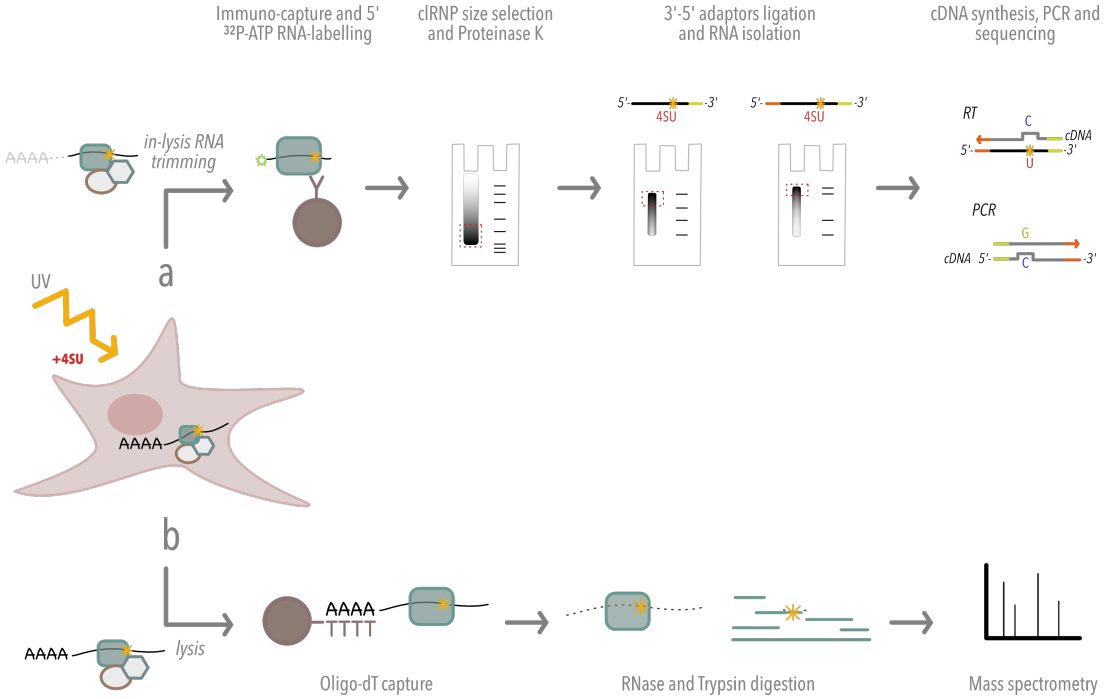

Figure 1.5: Classic methods to purify ribonucleoprotein complexes. Current methods have as starting point the covalent cross-linking of RNA-protein complexes by UV-irradiation of living cells. Often nascent RNA in cells are also labelled, e.g. with 4-thiouridine (4-SU), as in the **(a)** PAR-CLIP protocol [32] where cells are then lysed, selected RBP immunoprecipitated, and the bound RNA is radioactively marked to allow the visualization of the complexes. Complexes are size selected by SDS-PAGE and Western blot, and the purified RNA transcribed into cDNA for sequencing. T-to-C conversions in the cDNA serve as beacon to identify cross-linking sites in the RNA with one nucleotide resolution. While PAR-CLIP (and CLIP methods in general) are RNA-centric methods, with the **(b)** mRNA-interactome capture [14; 15] the proteome-wide identification of RBPs is possible. After UV-cross-linking, cell lysis is performed under highly denaturing conditions and the cross-linked complexes purified via the polyA-tails in their bound mRNA. Complexes are then released from the beads, RNase digested and proteins analysed by mass spectrometry.

by immunoprecipitation of the RBP and sequencing of the bound RNA (Fig. 1.6).

CLIP has been combined with *in vivo* incorporation of photoactivable nucleotide analogs (4-thiouridine (4-SU) and 6-thioguanosine (6-SG)) into nascent RNA transcripts, a technique known as PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation) [32]. Crosslinking of photoreactive nucleoside-labeled cellular RNAs to interacting RBPs is then achieved by UV light radiation at λ= 365 nm. The major feature of this technique is that the precise position of cross-linking can be identified by mutations residing in the sequenced cDNA; 4-SU results in a thymidine to cytidine transition, whereas 6-SG promotes guanosine to adenosine mutations [32].

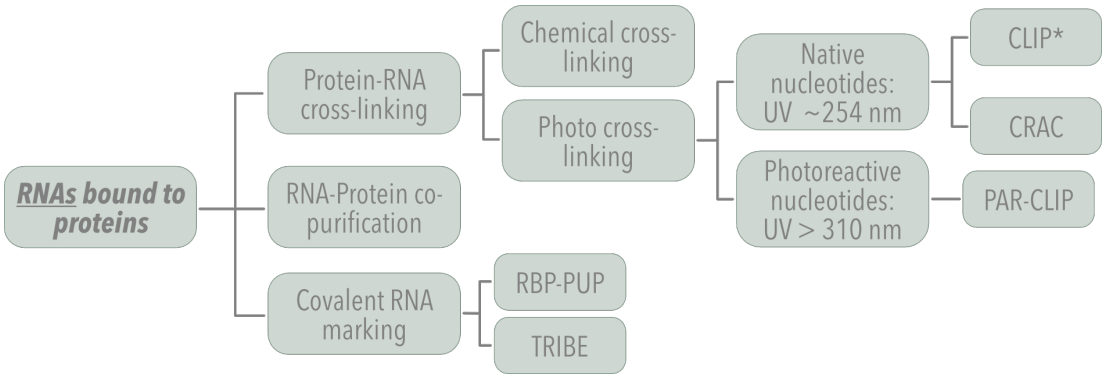

Figure 1.6: Classification of methods aimed to study the RNAs bound to a protein. RNA-protein complexes are stabilised by UV cross-linking [37], or by introducing functional groups in the RNA (TRIBE [39]; RBP-PUP [40]). (*) variants of CLIP includes: HITS-CLIP, high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation [41]; CRAC [42]; iCLIP [43]; eCLIP, enhanced-CLIP [44], among others; Modified from [45].

Alternative approaches involve the fusion of the RBP of interest to either polyU polymerase (RBP-PUP [40]), or to the catalytic domain of the ADAR RNA editing enzyme (TRIBE [39], Fig. 1.6). The main advantage of these methods is that protein-RNA complexes do not need to be biochemically isolated. However, these techniques require the construction and expression of chimeric proteins, and therefore cannot be applied to native RBPs [45].

CLIP and PAR-CLIP have become the most commonly used methods to analyse the RNAs bound to a protein. Many adaptations of CLIP have been developed (recently reviewed in [46; 47; 48; 45], Fig. 1.6). However, a requisite

for all CLIP-like methods is the previous knowledge or validation of the RBP to be investigated and are limited by i) the number of RBPs which can be simultaneously investigated, ii) the availability of antibodies with high affinity and specificity or iii) the efficient and correct expression of affinity-tagged versions of the protein of interest.

1.3.2 Protein-centric approaches

In the past, RBPs have been characterized using gel electrophoresis of UV cross-linked nuclear extracts or using RNA affinity purification coupled with mass spectrometry and/or immunodetection [49; 33; 50]. A breakthrough in the field was made in 2012 when teams led by Markus Landthaler (MDC-Berlin, DE) and Matthias Hentze (EMBL-Heidelberg, DE) independently developed a method to purify mRNA-protein complexes in a proteome-wide manner [14; 15].

The (m)RNA-interactome capture (RIC) technique involves *in vivo* UV-cross-linking and cell lysis in highly denaturing conditions followed by hybridisation and capture of the complexes via the polyadenylated tails present in the mRNA of the cross-linked complexes with oligo-dT beads. By inducing covalent bonds between interacting RNA and proteins it is possible to increase the stringency of washes during the purification process, thereby reducing the background and enhancing the signal-to-noise ratio in downstream applications. Cross-linked proteins are then released from the complexes by RNase treatment and analysed by mass spectrometry (MS). Likewise the PAR-CLIP method, RIC can be also combined with 4-SU RNA labelling, in this case the cross-linking efficiency varied in a small subset of RBPs [15] (Fig. 1.5b).

To date, the RNA-bound proteomes (RBPomes) of a wide range of organisms, covering almost all kingdoms of life, have been discovered. An integration of all data sets available counted for 1914 RBPs in *Mus musculus*, 1393 in *Homo sapiens*, 1272 in *Saccharomyces cerevisiae*, *Drosophila melanogaster* (777), *Arabidopsis thailiana* (719), *Caenorhabditis elegans* (593), *Danio rerio* (227), *Trypanosoma brucei* (155), *Leishmania donovani* (79), and *Plasmodium falciparum* (64) [12]. Despite the extended impact of the technique, its application is restricted to the presence of long polyadenylated stretches in the bound-RNA, a feature which is present almost exclusively in the mRNA of

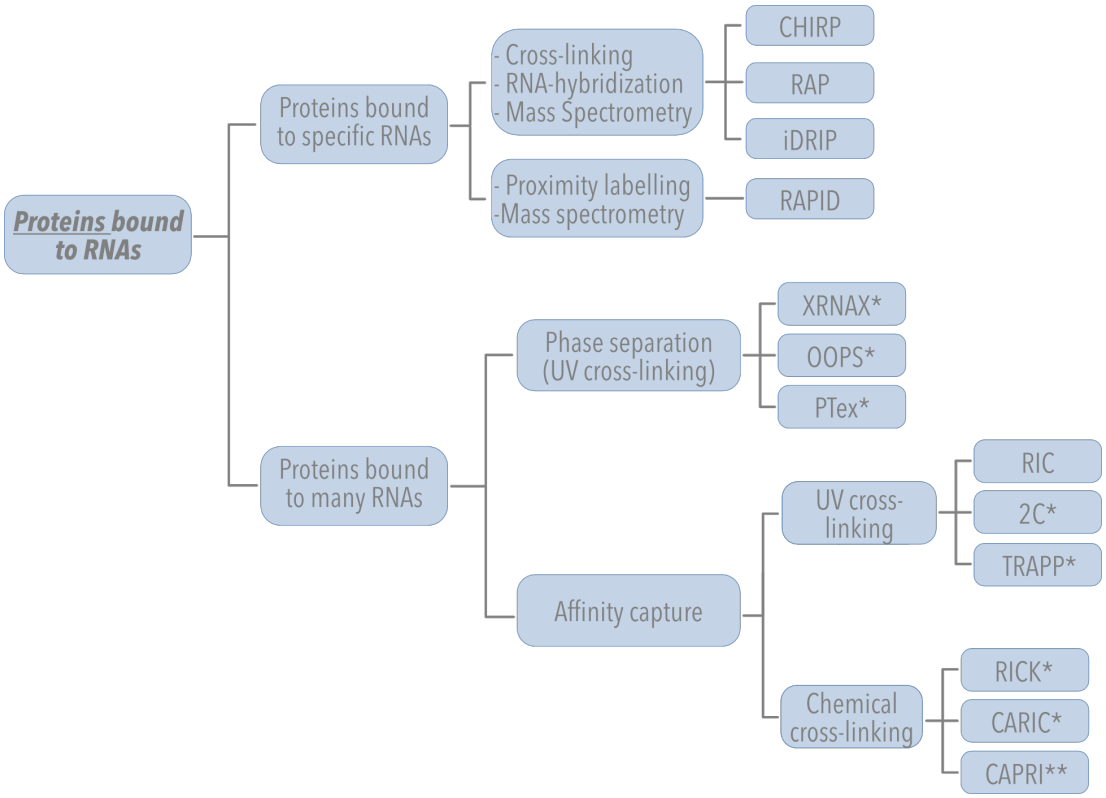

Figure 1.7: Classification of methods devoted to purify and identify RNA-binding proteins. Methods to purify proteins bound to specific RNAs include: CHIRP [51], RAP [52], iDRIP [53] and RAPID [54]. Whereas methods to purify a wider range of RNPs are based either on affinity purification, e.g. by UV cross-linking and oligo-dT capture, such as the mRNA-interactome capture technique (RIC, independently developed by [14; 15]) and its variants [55; 29]; Chemical cross-linking, CLICK chemistry or fusion of functional domains as in RICK [56], CARIC [57] and Prox. CLIP [58]; or a combination of both as in CAPRI [29], which combines the specificity of UV-cross-linking and the wider cross-linking effect of formaldehyde to purify not only mRNA-binding proteins but also other protein partners within the complexes. Phase separation upon organic extractions have been recently introduced as unbiased approaches to purify RNA-protein complexes: XRNAX [59], OOPS [60], PTex [61]. Methods marked with (*) are unbiased towards the type of RNA involved in the complex. (**) combination of UV light and chemical cross-linking. Modified from [45].

eukaryotes.

Alternative methods to purify RNPs independently of oligo-dT capture have been developed (1.7). By incorporating 5-ethynyluridine (EU) into nacent RNA, followed by cyclo-addition of biotinylated azide and CLICK chemistry *in vivo*, the RICK (RNA Interactome using Click Chemistry, [56]) and CARIC (Click Chemistry-Assisted RNA Interactome Capture [57]), the isolation of a wider, polyA independent, range of RNPs can be achieved. However, as those methods require the efficient incorporation of modified nucleotides, the applicability of these approaches is largely restricted to cell culture [45].

Recently, unbiased methods to purify UV cross-linked RNA-protein complexes (clRNPs) based on phenol-toluol or acidic-phenol extractions have been independently developed (1.7): PTex (Phenol Toluol Extraction, [61], this work), OOPS (Orthogonal Organic Phase Separation, [60]), and XRNAX (protein-crosslinked RNA extraction, [59]). The purification process is based on the different physiochemical properties of clRNPs *vs.* unbound RNA and proteins, allowing the global recovery of complexes in a cell-wide manner while bypassing the requirement for *in vivo* labelling of RNA [45; 62].

1.4 Behind the scenes: the chemistry of phenolic extractions and UV cross-linking

1.4.1 Phenolic extractions for the purification of nucleic acids

Although purification of nucleic acids by liquid-liquid organic extractions was established already in 1956 [63], it was only after Chomczynski and Sacchi introduced the "single step" method that it became the *de facto* standard for RNA isolation [64]. Exploiting the differences in solubility of proteins and nucleic acids in aqueous (polar) and organic (non-polar) solvents, isolation of RNAs is achieved by phenolic extractions of cellular lysates using guanidinium thiocyanate under acidic pH [64; 65; 66].

At cellular conditions, the hydrophobic and non-polar amino acids within a

polypeptide tend to collapse towards the core of the protein (known as the hydrophobic effect [2]); however, when denatured with organic solvents as phenol or detergents, the collapse of the protein is reversed and the hydrophobic amino acids become accessible to the solvent, allowing for the solubility of the protein [2]. This feature is further enhanced by using a denaturing solution containing the chaotropic agent guanidinium thiocyanate [64]. On the contrary, nucleic acids remain polar (depending on the pH of the solution) and dissolve in the aqueous phase (Fig. 1.8, [62]).

The difference in density between the aqueous and organic phases (Table 1.1) allows its separation by applying short centrifugations, during which the protein-containing phenolic phase descends to the bottom of the tube while the rich in nucleic acids aqueous phase is placed on the top. Chloroform or 1,3-bromo-chloro-propane (BCP) have been introduced to sharpen the phase separation due to their even higher density (Table 1.1) which reduces carry-over of one of the two phases when pipetting [65; 62].

Property	Chloroform	BCP	Phenol	Toluol
Solubility in water (g/L at 25 °C)	7.95	2.24	82.8	0.52
Relative density (water= 1)	1.48	1.6	1.06	0.87

Information obtained from https://pubchem.ncbi.nlm.nih.gov.

Table 1.1: Physico-chemical properties of chemicals used for RNA isolation. Modified from [62].

The pH of the organic and aqueous solutions during the phenolic extractions can determine partitioning of RNA and DNA. At physiological and acidic pH (7.0 and 4.8, respectively), nucleobases are in their neutral form, while the 2′OH group of RNA is ionised (Fig. 1.8). Only the phosphodiester bond has a pK_a of 6.0 - 7.0. Thus, the phosphate group of the backbone of DNA and RNA is neutral only at pH 4.8 [62]. At pH <5, the DNA shifts from a negatively charged to a neutral molecule, and the decrease in polarity then promotes its enrichment in the organic, non-polar phase [62].

On the other hand, RNA has an additional negative charge due to its 2′OH group with a pK_a of 13.0 (Fig. 1.8). Unlike in DNA, the nucleobases of RNA molecules are not all paired via hydrogen bonds in a double helix, leaving

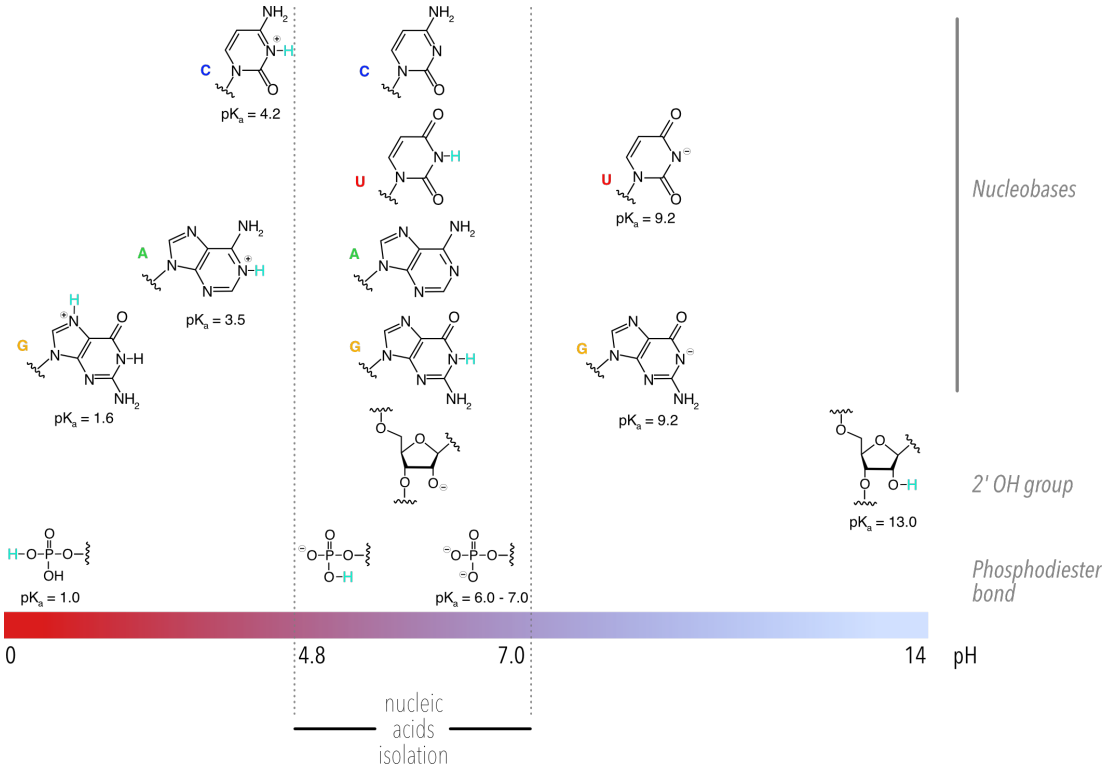

Figure 1.8: Protonation of nucleotides pK_a and pH. Protonation of sites in RNA and their respective pK_as [69]. In the pH range of buffers used for RNA isolation (and for PTex, section 2.9), only the phosphodiester backbone is protonated at pH 4.8. Nucleobases remain in their neutral form and also the 2′OH of the sugar remains ionised throughout the protocol. Modified from [69; 61].

unpaired bases free to interact with surrounding water molecules, thereby increasing the overall polarity of RNA and its enrichment in the aqueous environment [67; 62]. Modifications of the single step method for RNA isolation have allowed the simultaneous recovery of RNA in the aqueous phase and of proteins from the organic phase [68; 67].

1.4.2 UV-induced RNA-protein cross-linking

UV-induced covalent cross-linking of RNA and proteins has been in used for decades [70; 71]. The first example of a uracil covalently bound to cysteine was reported by Smith and Aplin [72] using NMR spectroscopy and mass spectrometry, resulting in the formation of the molecule 5-S-cysteine-6-hydrouracil (Fig. 1.9a).

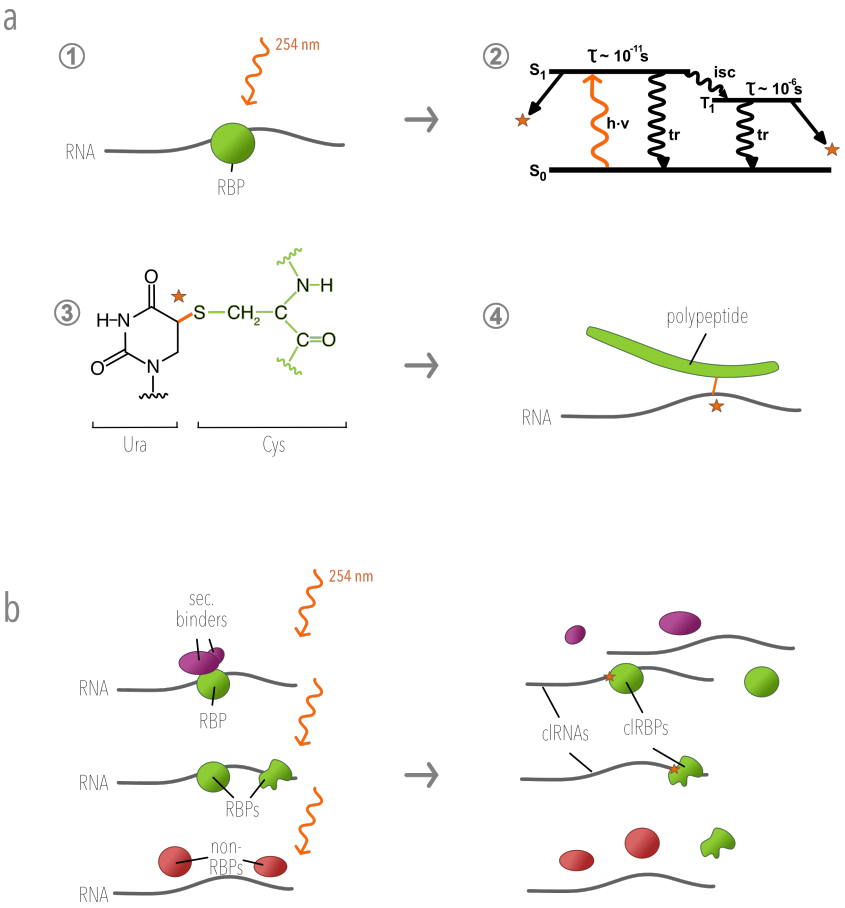

Figure 1.9: UV cross-linking of RNA-protein complexes. (a) Biophysical and chemical basis of UV cross-linking: (1) irradiation of RNA directly interacting with proteins using low energy short wavelength UV light (254 nm). (2) Simplified Jablonski diagram: excitation of nucleobases of RNA (h*ν) from the ground state S_0 to an energetically elevated singulet (S_1) state. Inter-state conversion (isc) to a triplet (T_1) state is possible. The lifetime of S_1 or T_1 states are 10 ps and 1 μs, respectively before falling back to the ground state either through thermal relaxation (tr) or by formation of a cross-link to an adjacent amino acid (orange star). Modified from [73]. (3) Example of a cross-link between uracil and cysteine by formation of 5-S-cysteine-6-hydrouracil as determined by [72]. (4) Denaturing of successfully cross-linked RNP results in a hybrid molecule consisting of a nucleic acid and a polypeptide part. **(b)** RBPs (green) can directly interact with RNA in contrast to secondary binders (violet), e.g. protein of RNP complexes interacting solely via protein:protein interactions or non-RBPs (red). UV irradiation at 254 nm can result in covalent cross-links between RBPs and RNA (denoted by an orange star). Note however that UV-induced cross-linking is inefficient and that the majority of the biological sample will remain non-cross-linked. Modified from [62].

Many combinations of amino acids/nucleobases cross-links have been investigated (reviewed in [70]). When cross-linked to poly-U, the most reactive amino acids are cysteine, tyrosine, phenylalanine, arginine, lysine and tryptophane [74; 70]. In comparison, RNA is more reactive than DNA, and pyrimidines more efficiently cross-linked than purines [75] when comparing nucleobases addition to cysteine (poly rU $>$ poly rC $>$ poly dT $>$ poly rA) [76; 62].

The chemistry behind UV-induced covalent cross-linking resides in the capacity of RNA and DNA molecules to absorb energy from short wavelength UV light (λ = 254 nm). However, when using a low energy UV source, the excitation of nucleobases to a higher energetic state S_1 or T_1 is short-lived. The excited state is then reverted to the ground state in micro- to picoseconds by thermal relaxation, or, if a suitable amino acid is in direct vicinity, by the formation of a covalent bond (Fig. 1.9a, [62]). Due to the short time of excitation, cross-links are most likely to form exclusively between components which are in direct contact at the time of irradiation ("zero distance cross-linker") [73; 70; 77; 62] (Fig. 1.9b).

Covalent cross-linking of proteins and RNA by UV radiation has advantages and disadvantages, all of which we have largely discussed in [62]. In the following section a summary of the advantages and disadvantages is given.

Advantages

- *In-vivo* RNA-protein interactions can be "frozen" by directly irradiating living cells with UV light.

- Unlike other cross-linkers such as formaldehyde [29], UV light will only cross-link proteins which are in direct contact with RNA; entire multi-protein complexes are usually not cross-linked (Fig. 1.9b).

- Covalent bonds are thermo-stable and most likely not reversible, therefore, resistant to denaturing conditions which allows the application of stringent conditions during purification.

Disadvantages

- UV cross-linking is an inefficient process as only up to 5% of a given RBP can be cross-linked to RNA [15; 16; 78; 61].

- Downstream applications, especially mass spectrometry, can be impaired due to the added molecular mass from the covalently-attached molecule, requiring RNase or protease digestion prior to assays such as electrophoresis, mass spectrometry, or cDNA synthesis. Note however that the additional mass at the cross-linking site serves as a beacon to map RNA-protein interactions at single nucleotide/amino acid resolution [32; 6; 29].

- The irreversible nature of covalent bonds can be also a disadvantage. Although single studies have suggested reversibility of DNA-protein cross-links by acidic or basic conditions [75], there are no reports that such events occur in RNA-protein covalent bonds to date.

1.5 Combining UV cross-linking and organic extractions to investigate RNPs

UV-cross-linking of RNA-protein complexes results in the formation of chimeric molecules: partially RNA and partially protein. Knowing that RNA and proteins exhibit well defined and differential physicochemical properties when in contact with hydrophobic-organic compounds, the immediate question to be asked is how would a covalent UV cross-linked RNA-protein complex (clRNP) behave during phenolic extractions?

Displaying physicochemical features of nucleic acid and proteins simultaneously, it is reasonable to assume that such clRNPs will accumulate at the phase boundary between the aqueous (polar) upper phase and the organic

(hydropohobic) lower phase [62], region known as interphase. Accumulation of clRNPs in the interphase during organic extractions was indeed observed long ago [71; 79], these reports however, were focused on the analytical investigation of the efficiency of UV irradiation. In this work however, the differential behaviour of clRNPs in comparison to free RNA and free protein during organic extractions are exploited for the unbiased purification of cross-linked RNPs.

1.6 Aim of the Study

Although the implementation of the (m)RNA-interactome capture has resulted in the identification of hundreds of novel RBPs in several eukaryotic species [12], it has the limitation of being a mRNA-based purification technique, therefore, only the RNPs bound to polyadenylated RNAs are recovered.

Polyadenylation is the major feature of mRNA in eukaryotic cells, however it is absent in other classes of eukaryotic RNA (ncRNAs, tRNAs, pre-mRNAs and rRNA); meanwhile in bacterial RNA -if present- it is limited to a maximum of 20 residues [80]. Likewise, polyadenylation is also absent in the archaeon *Haloferax vulcanii*, while in the hyperthermophile *Sulfolobus solfataricus*, RNA polyadenylation is used as a mark for degradation by the archaea exosome complex [81].

In the last decade, significant advances have been made in the elucidation of bacterial post-transcriptional mechanisms and the role of regulatory RNAs [9]. It is becoming increasingly clear that this layer of regulation also plays an important role in pathogenesis and antibiotic resistance [25; 24; 22; 23; 82; 83]. However, because of technical difficulties, exploration of the RBP repertoire in bacteria has proved to be tremendously challenging. As a result, in the model organism *Salmonella* Typhimurium only a limited number of RBPs have been implicated in pathogenesis [84; 85].

Accordingly, the aim of this work was to establish an alternative protocol for the purification of covalently cross-linked RNPs, unbiased towards the type of RNA involved in the complex, and potentially modifiable to be applied in

prokaryotic cells.

Here I show how the partitioning behaviour of clRNPs among phases depends on the physicochemical properties of the complexes themselves and of the composition of the aqueous-organic mix, a composition which was modified from the original single step method for RNA isolation [64] and which resulted in the successful development of PTex (phenol-toluol extractions). As this protocol relies solely on the differential physicochemical properties of RNA and proteins *vs.* clRNPs, the unbiased recovery and analysis of RNPs was possible, not only in mammalian -human and mouse- cultured cells and tissue, but importantly, for the very first time it allowed the proteome-wide purification of clRNPs from the human pathogen *Salmonella* Typhimurium.

Chapter 2

Methods

2.1 Standard buffers, solutions and culture media

Cross-linking buffer (CLB) 10 mM Tris pH 7.4, 50 mM KCl, 1 mM EDTA, 1 mM DDT.

DMEM Dulbecco's Modified Eagle, glucose 4.5 g/L (Gibco, 41966–029), supplemented with 10% bovine serum (Gibco, 10270–106), 100 U/mL penicillin, 0.1 mg/mL streptomycin (Gibco, 15140–122).

DPBS Phosphate-saline buffer, pH 7.4 (Gibco, 10010–015).

LB Luria-Bertani medium, 10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl.

LPM plus Low phosphate/magnesium medium plus, 5 mM KCL, 7.5 mM $(NH4)_2SO_4$, 0.5 mM K_2SO_4, 38 mM (0.3% v/v) Glycerol, 1% Casaminoacid, 8 µM $MgCl_2$, 1 mM KH_2PO_4, 1 M Tris to pH 7.6. Modified from [86].

NP-T buffer — Sodium-phosphate-tween buffer, 50 mM NaH_2PO_4, 300 mM NaCl, 0.05% Tween, pH 8.0.

PBST — Phosphate-saline buffer supplemented with Tween 20 (10 mM phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, pH 7.4 0.1% tween 20).

PBST-M — PBST with 5% milk.

RNase buffer — 50 mM Tris, 1 mM $MgCL_2$, pH 8.0.

Solution D — Denaturing solution, 5.85 M guanidine isothiocyanate, 31.1 mM sodium citrate, 25.6 mM N-lauryosyl-sarcosine, 1% 2-mercaptoethanol, pH 4.8).

TE — Tris-EDTA buffer, 20 mM Tris, 1 mM EDTA, pH 7.6.

TED buffer — TE buffer supplemented with 0.03% DDM (n-Dodecyl β-D-maltoside).

2.2 Mammalian cell culture and *in-vivo* cross-linking

HeLa and HEK293 cells were grown to 80-90% confluence using DMEM at 37 °C with 5% CO_2 on 78 cm^2 dishes. Cells were immediately washed with cold DPBS and irradiated (on-ice) with 0.15–1.5 J/cm^2 of UV light (λ= 254 nm; CL-1000 ultraviolet cross-linker (Ultra-Violet Products Ltd). Cells were detached with cold DPBS, centrifuged in 2 mL tubes (2-8x10^6 cells/tube) and stored at -20 / -80 °C (hereafter +UV). Non-irradiated cells were prepared accordingly for non cross-link controls (-UV).

RNA was isolated from HEK293 cells before and after exposure with UV light by phenol extraction [64] and analysed with a Bioanalyzer 2100 using the RNA 6000

Chip Kit (Agilent, 5067-1513). HEK293 and HeLa cellas were kindly provided by Prof. Dr. Markus Landthaler (Max-Delbrück Center for Molecular Medicine, Berlin, DE) and Prof. Dr.Andreas Hermann, (Humboldt-Universität zu Berlin, Berlin, DE), respectively.

2.3 Bacterial cell culture and *in-vivo* cross-linking

Salmonella enterica subsp. *enterica* serovar Typhimurium strain SL1344 Hfq::x3FLAG [85] was grown on LB medium to OD_{600}= 3.0. Fractions of 20 mL were pelleted (20,000×g, 8 min, 37 °C) and dissolved in 2 mL of water for UV irradiation. Cells were cross-linked on ice with 5 J/cm2 UV light (λ= 254 nm), snap-frozen and stored at -80 °C. Bacterial suspensions with a density of 7.5 ODs/mL were used for the experiments shown in Fig. 3.16c.

Salmonella Typhimurium SL1344 used for the mapping of RNA-protein interactions (Fig. 3.17) was grown in LB medium to OD_{600} 2.0. Half of the cultures were cross-linked in a Vari-X-linker (UVO3, www.vari-x-link.com), using UV light (λ= 254 nm) lamps for 90 s (aproximately 1.5 J/cm_2). Fractions of 10 mL from each, cross-linked and non-cross-linked cultures, were harvested by filtration [87] and kept at -80 °C. These samples were kindly prepared and provided by Stuart McKellar and Prof. Dr. Sander Granneman (Edinburgh University, UK). Bacteria was detached from filters with cold PBS, followed by centrifugation at 20,000 xg, 4 °C in 2 mL aliquots (equivalent to 3 ODs), pellets were snap frozen and kept on dry ice until being used.

Salmonella Typhimurium SL1344 strains FLAG(or HTF)-tagging the proteins CsrA [85], YigA, ClpX, DnaJ [30], AhpC, SipA, YihI [61] and GpmB were grown in 100 mL of LB medium to OD_{600}= 2.0. Two different cross-linking strategies were adopted: i) direct irradiation of LB cultures (on Petri dishes, ice-bed) with 5 J/cm^2 of UV light (λ= 254 nm), followed by centrifugation at 4 °C for 10 min, or ii) cultures centrifuged at room temperature for 10 min at 15,000 xg, and pellets dissolved in 0.1 volumes with water before irradiating with 5 J/cm^2 [30; 61]. Cross-linked suspensions were pelleted, snap-frozen and used for immunoprecipitation and PNK assay (section 2.19). Bacterial culture, cross-linking and immunoprecipitations shown in Fig. 3.17d-f were performed

by Davide Figini (Beckmann lab, IRI Life Sciences, Humboldt-Universität zu Berlin, DE).

Salmonella Typhimurium strain SL1344 Hfq::x3FLAG shown in Fig. 3.16b was cultivated in LPM plus medium to OD_{600}= 1.2. For UV cross-linking, aliquots of 2.5 mL were placed on the centre of Petri dishes and radiated with 0.25, 0.5 and 1.0 J/cm^2 (on ice-pad). Cells were pelleted (1 mL/tube) at 4000 xg, 3 min, 4 °C. Hot-PTex was done with 1 bacterial pellet per condition (equivalent to 1.2 ODs).

2.4 Construction of bacterial strains

The *Salmonella* Typhimurium SL1344 strains used in this work were constructed using the Lambda Red system developed by [88]. The system is based on two plasmids: pKD46, a temperature-sensitive plasmid that carries gamma, beta and exo genes (the bacteriophage λ red genes) under the control of an Arabinose-inducible promoter, and a plasmid carrying either the 3xFLAG::KmR (FLAG) or the 6xHis-TEV-3xFLAG::TetRn (HTF) cassettes, pSUB11 and pJet1.2-Hfq-HTF-TetR [89], respectively, as described in [61; 30]. Briefly:

Salmonella Typhimurium SL1344 transformed with the plasmid pKD46 was grown in LB containing Ampicillin (100 µg/mL) and L-Arabinose (100 m) at 28-30 °C to an OD_{600} of 0.8. Cells were incubated on ice for 15 min, and washed by three cycles of centrifugation (3000 xg, 5 min, 4 °C) and solubilisation in ice-cold water. Cells were resuspended in 300 µL of water and electroporated with 200 ng of PCR product.

Cells were let to recover for one hour in LB medium at 37 °C on a tabletop Thermomixer at 600 rpm, plated on LB agar with Kanamycin (50 µg/mL) or Tetracyclin (100 µg/mL) overnight. Resulting colonies resuspended in PBS were struck on plates containing Ampicillin or Kanamycin(or Tetracyclin) and incubated at 37-40 °C. Colonies that showed resistance to Kanamycin (or Tetracyclin) but not to Ampicillin were selected for further analysis. Expression of the FLAG-tag was verified by Western blot. *Salmonella* Typhimurium SL1344 strains used in this work are listed in Table 2.2.

Strain name	Genotype	Study
SL1344*	rpsL hisG	[85]
JVS-04317*	SL1344 csrA-3xFLAG KanR	[85]
AhpC-HTF	SL1344 ahpC::6xHis-TEV-3xFLAG::TetRn	[61]
SipA-HTF	SL1344 sipA::6xHis-TEV-3xFLAG::TetRn	[61]
GpmB-FLAG**	SL1344 gpmB::FLAGx3::KanR	This study
ST-BB-2025**	SL1344 clpX::FLAGx3::KanR	[30]
ST-BB-2024**	SL1344 dnaJ::FLAGx3::KanR	[30]
ST-BB-2004**	SL1344 yigA::FLAGx3::KanR	[30; 61]
YihI-FLAG**	SL1344 yihI::FLAGx3::KanR	[61]

Table 2.2: Bacterial strains used in this study. Indicated strains were kindly provided by Prof. Dr. Jörg Vogel* (Würzburg University, DE) or Davide Figini** (Beckmann lab, IRI Life Sciences, Humboldt-Universität zu Berlin, DE).

Additionally, the T7 promoter and a sxl-target DNA sequence 5-GAT CCG GTC ATA GGT GTA AAA AAA GTC TCC ATT CCT ATA GTG AGT CGT ATT AA-3 were cloned into pUC19 using the restriction enzymes BamHI and HindIII. The resulting plasmid was named pUC19-sxl-target [61].

2.5 Mouse tissue preparation

Mouse-brains were kindly provided by Prof. Dr. Florian Heyd and Dr. Marco Preussner (FU, Berlin). Tissues (260 mg) were disrupted by cryogenic grinding, half of it was subjected to UV irradiation with 0.75 J/cm^2 (λ= 254 nm). Samples of 32.5 mg (±UV) were dissolved in 600 µL DPBS for Hot-PTex extractions (section 2.9).

2.6 mRNA interactome capture

UV cross-linked mRNA-protein complexes were purified by the mRNA interactome capture (RIC) technique, according to [15] using the *in vivo* cross-linked cells prepared before (section 2.2). RIC samples were used with the aim of quantitatively determine the enrichment of such complexes across phases during the exploratory and PTex extractions.

2.7 *In vitro* transcription of Sxl-RBD4 target RNAs

RNA fragments with lengths from 13 to 191 nucleotides and containing the poly(U)$_7$ target sequence for the Sxl-RBD4 recombinant protein were synthesised as described in [61], as follows:

- The 13 nt RNA 5'-GAG UUU UUU UAC A-3' was synthesised by Biomers (Ulm, Germany).

- The 30 nt RNA was synthesised by hybridising two complementary sequences containing the T7 RNA polymerase promoter 5-TAA TAC GAC TCA CTA TAG-3 and the template sequence 5-GGT CAT AGG TGT AAA AAA ACT CTC CAT TCC TAT AGT GAG TCG TAT TAA-3 [61], followed by T7 run-off transcription [90].

- RNAs of 87 and 191 nt length were produced by *in vitro* T7 transcription using DNA restriction fragments obtained by enzimatic restriction of the pUC19-sxl-target plasmid (section 2.4) with the enzymes HindIII and EcoRI (87 bp), or HindIII and PvuI for (191 bp).

T7 RNA polymerase and restriction enzymes were purchased from New England Biolabs. Plasmids and DNA were purified using the NucleoBond Xtra Midi kit (Macherey-Nagel, 740410.100), NucleoSpin Gel or PCR clean-up (Macherey-Nagel, 740609.50). RNAs were isolated by acidic phenol extraction following the single step method [64].

2.8 Exploratory organic extractions

A series of simplified organic extractions were designed to determine the individual contribution of the reagents: phenol, toluol, pH, and physiological *vs.* denaturing conditions, on the partitioning of the different molecules during liquid-liquid organic extractions [62]. A schematic representation of the approach is resumed in the following figure:

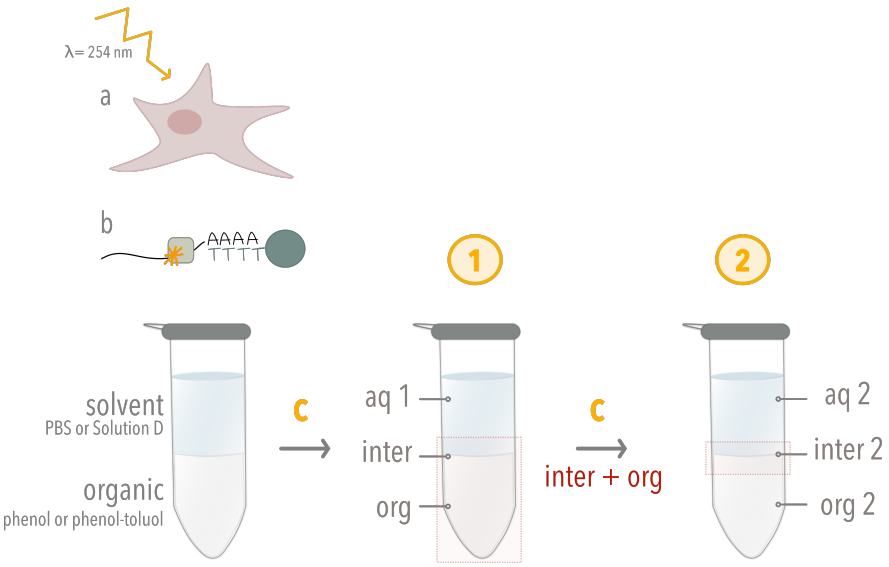

Figure 2.1: Development of the PTex method: exploratory extractions. UV-irradiated cell pellets **(a)** or mRNA-interactome samples **(b)**, prepared as in 2.2 and 2.6, were mixed with either denaturing solution D or neutral buffer (DPBS), and mixed with phenol or phenol-toluol-BCP, at a ratio 2:2:1. After centrifuging (20,000 xg, 4 °C, 3 min; C in yellow), the upper aqueous phase (aq 1) was transferred to a new tube for ethanol precipitation and the inter-organic phases mixed with ethanol and water (1), and centrifuging as before. (2) The resulting phases (aq 2, inter 2, org 2) were precipitated with ethanol (as in section 2.11).

2.9 The PTex protocol

PTex (for Phenol-toluol extractions) was designed as a modular protocol consisting in three consecutive organic extractions [61; 62]:

- **Step 1:** cellular pellets (2-8×10^6 cells), as in section 2.2, or RIC samples (section 2.6), were dissolved in 600 µL of DPBS and mixed with neutral phenol, toluol, and BCP (200 µL each) for 1 min at 21 °C and 2000 r.p.m (ThermoMixer), and centrifuged at 20,000 ×g 3 min, 4 °C.

- **Step 2:** 500 µL of the upper aqueous phase (aq 1) was transferred to a new 2 mL tube and mixed for 1 min with 300 µL of solution D, 600 µL neutral phenol and 200 µL BCP, and centrifuged as in Step 1 to allow phase separation.

- **Step 3:** With the help of a syringe (blunt-needle), 70% of the aqueous and organic phases were removed, keeping the resulting interphase (inter 2)

in the same tube. Water (400 µL), ethanol (200 µL), phenol (400 µL) and BCP (200 µL), were added, mixed and centrifuged as before. Finally, 70% from the aqueous and organic phases were removed. Interphase (or all phases when indicated) was (were) subjected to ethanol precipitation as in section 2.11.

2.10 Hot-PTex

A modification of the PTex protocol was established to allow the purification of UV cross-linked RNA-protein complexes from difficult tissues [61], briefly: mouse-brain tissue (section 2.5) or bacterial pellets (section 2.3) were dissolved in 400-600 µL of DPBS, mixed with 600 µL of phenol-toluol-BCP (2:2:1), and 0.5 g of zirconium beads, during 5 min at 65 °C (2.000 r.p.m, Eppendorf ThermoMixer). Following centrifuging at 20,000 xg, 3 min, 4 °C, the aqueous phase was transferred to a new tube were subsequent extractions, steps 2 and 3 as indicated above, were performed at 65 °C.

2.11 Precipitation of PTex samples and evaluation of intermediary steps

Unless otherwise stated, all PTex samples and intermediary phases were precipitated with nine volumes of ethanol p.a., at -20 °C for at least 30 min [61; 62]. Supernatants were carefully decanted and pellets let to dry under the hood for 10 min. Samples were solubilised with Laemmli buffer, water or the appropriate buffer according to the downstream application.

2.12 Comparison of protein precipitation methods

With the aim of selecting the most adequate precipitation method in terms of yield and solubility of the proteins, HEK293 +UV cell pellets were subjected to

Step 1 of the PTex protocol, the resulting lysates were used as probe for testing three different precipitation methods (as described in [62]):

- Ethanol precipitation: samples were mixed with 9 volumes of ethanol p.a., incubated at -20 °C during 30 min and centrifuged 30 min at 20,000 xg and 4 °C. Pellets were washed once with cold ethanol 70% followed for 10 min centrifuging.

- 2-propanol precipitation: samples mixed with 3 volumes of 2-propanol were incubated 10 min at room temperature and centrifuged 20,000 xg, 20 min at 4 °C. Pellets were washed as described above.

- TCA precipitation: cold trichloroacetic acid was added to samples in ratio 0.25:1, followed by 10 min incubation on ice. Samples were centrifuged during 10 min at 20,000 xg, and pellets washed once with cold acetone and centrifuged again.

Supernatants were carefully decanted and pellets left to dry under the hood for ∼10 min. Samples were dissolved with 100 µL of one of the following: water, TE (20 mM Tris, 1 mM EDTA, pH 7.6), or TED buffer (TE supplemented with 0.03% DDM [*n*-Dodecyl β-D-maltoside]), at 56 °C during 20 min. Samples were spun down at 1000 xg, for 2 min to separate soluble from insoluble material. Protein quantification was determined by measuring the absorbance at λ280 nm in a Nanodrop 2000. Remaining samples were electrophoresed and HuR protein detected by Western blot as detailed in section 2.15.

2.13 Analysis of individual PTex steps

Intermediary steps from the individual PTex extractions were analysed to i) identify the migration pattern of the different molecules across phases during the extractions, and ii) evaluate the carry-over of free- DNA, RNA or protein in the PTex samples (interphase 3). Independent PTex extractions were performed using the following molecules or cells as input:

- Synthetic RNA: 30-50 nt RNAs 5-labelled with γ^{32}P-ATP (provided by Dr. Hans-Hermann Wessels and Uwe Ohler (Max-Delbrück Center for Molecular Medicine, Berlin, DE).

- Linear DNA: 200 ng pUC19 *lacZ* gene-containing fragment (817 bp, generated by DrdI, NEB).

- HEK293 cell suspension (2-3x10^6) spiked-in with 0.25 µg of Sxl-RBD4 per tube (non UV-cross-linked).

Detection of DNA was performed by PCRs using primers targeting endogenous chromosomal DNA, *il-3* gene (574 bp, forward primer: 5-GAT CGG ATC CTA ATA CGA CTC ACT ATA GGC GAC ATC CAA TCC ATA TCA AGG A-3 and reverse primer: 5-GAT CAA GCT TGT TCA GAG TCT AGT TTA TTC TCA CAC-3), or the *lacZ* gene present in the pUC19 linear fragment (324 nt, forward 5- AGA GCA GAT TGT ACT GAG-3and M13-reverse 5-CAG GAA ACA GCT ATG ACC). Amplificated fragments were resolved in Agarose gels 1.0-1.5%.

Likewise, RNA samples were electrophoresed in TBE-Urea PAGE 12% gels, and the radioactivity was detected by phosphoimaging in a Life Science FLA-5100 imaging system (Fujifilm); while the proteins Sxl-RBD4, endogenous HuR or ACTB in HEK293/Sxl-RBD4 spiked-in samples were analysed by Western blot with its corresponding antibodies as described in section 2.15.

2.14 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) were performed to demonstrate that the increased molecular mass of proteins corresponded with the presence of RNA-protein complexes [61; 62]. PTex samples from mRNA-interactome capture (15 µg, section 2.6), or HEK293 cell pellets (section 2.2), were dissolved in 150 µL of ultra-pure water or buffer TED, mixed with RNaseA (1-2 ng) and incubated at 37 °C; aliquots of 20 µL were taken at different time points between 0 and 60 min. Aliquots were immediately mixed with 5 µL of 6x Laemmli buffer, heated at 95 °C for 5 min and used for protein electrophoresis

and Western blotting as in section 2.15.

2.15 Western blot and densitometry analysis

Samples were resolved in SDS-PAGE gradient gels 4–20% (TGX stain free, BioRad), or Bis-Tris-MOPS 4-12% (NuPAGE, Invitrogen), followed by transfer onto nitrocellulose membranes 0.2 µ (Turbo-blot, BioRad). Membranes were blocked during 30 min with PBST-M. Specific proteins were targeted incubating 0.1-2.0 µg/mL of the corresponding antibody by incubation at 4 °C overnight or during 2 h at room temperature.

Primary antibodies against the following proteins were used: HuR (proteintech, 11910-1-AP), hnRNPL (proteintech 18354-1-AP), ABCF2 (proteintech, 10226-1-AP), CCT7 (15994-1-AP), FUS (abcam, ab124923), GAPDH (proteintech, 10494-1-AP), alpha-enolase (ENO1, proteintech,11204-1-AP), PTBP1 (abcam, ab133734), PABPC1 (proteintech, 10970-1-AP), ACTB (proteintech, 66009-1-Ig), Histone H3 (abcam, ab21054), GroEl (abcam, ab90522), FLAG-tag (Sigma, A8592), or Sxl-RBD4 (DHSB, anti-Sxl hybridome culture supernatant M114, kind gift from Dr. Jan Medenbach).

Binding was detected using the respective secondary antibodies anti-mouseHRP (proteintech, SA00001-1), anti-mouse-AlexaFluor680 (Invitrogen, A32729) anti-rabbitHRP (proteintech, SA00001-2), or anti-rabbit-AlexaFluor488 (Invitrogen, A32731) and Clarity ECL Western blotting Substrate for chemiluminescence when required (Biorad). Imaging was performed with a ChemiDocMP imaging system (BioRad).

Densitometry analysis of the images from the HuR and Sxl Western blots were used to determine the yield of PTex recovery, by comparing the intensity of cross-linked *vs.* non-crosslinked bands using ImageJ [91] (Fig.2.2)

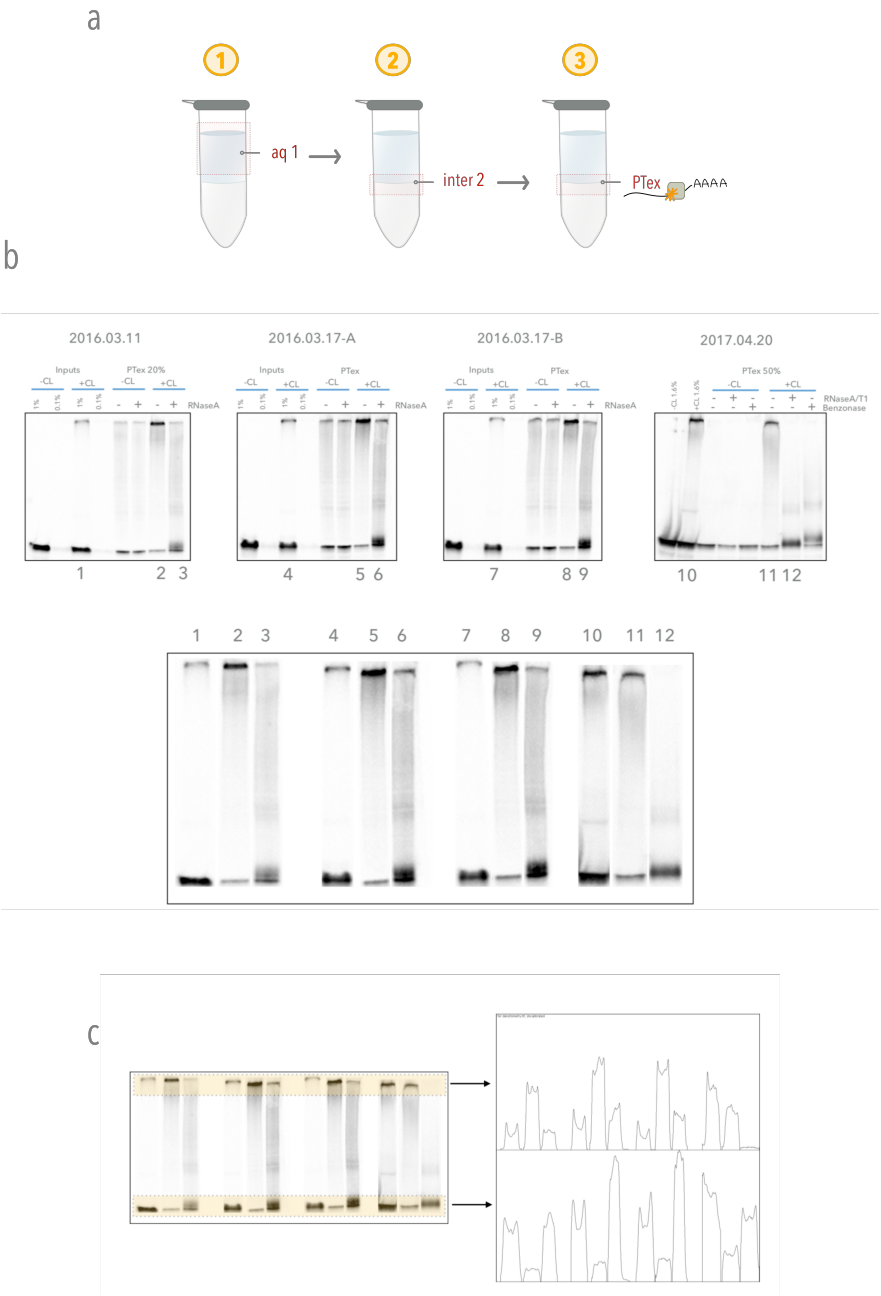

Figure 2.2: PTex quantification by densitometry. **(a)** PTex was performed with HEK293 cells +/- UV. PTex pellets were electrophoresed and Western blotted against the protein HuR. **(b)** Unsaturated images from the Western blots were placed next to each other in a 8-bit TIF file. Only the PTex +UV, +/- RNaseA were used for the analysis. **(c)** Cross-linking efficiency was calculated as the percentage of HuR shifted to the gel pockets in the inputs (lanes 1,4,7 and 10). PTex enrichment corresponds to the difference between the cross-linked (upper) signal divided by the total signal of the lane (lanes 2, 5, 8, and 11). PTex yield was calculated as the percentage of total cross-linked protein recovered (lanes 3, 6, 9, and 12) compared to the input cross-linked signal. Modified from [62].

2.16 RNase treatment prior PTex

HEK293 cell suspensions (2-3x10^6 cells/mL, ± UV) were treated with 2000 U/mL benzonase (Merck, 70664) in the recommended buffer (50 mM Tris, 1 mM $MgCL_2$, pH 8.0) during one hour at 37 °C and 1000 r.p.m (ThermoMixer, Eppendorf), a set of untreated cells were kept as control. After PTex, samples were analysed by Western blot to detect RBP and non-RBP control proteins as previously described (sections 2.9 and 2.15).

2.17 *In vitro* cross-linking assays

In order to asses the limitations of the PTex method, an *in vitro* assay was designed to evaluate the minimal length of the RNA within a complex required to accomplish an efficient purification. *In vitro* binding was allowed incubating 40 µg of Sxl-RBD4 with *in vitro*-transcribed RNA (13, 30, 87 and 191 nt, 1.7-10 µM, section 2.7) in 100 µL cross-linking buffer at 4 °C for 30 min. Samples were cross-linked with 0.25 J/cm2 of UV-254 nm, on ice. Afterwards, 98% of each sample was used for PTex extraction, keeping 2% of the sample for input control. SDS-PAGE were performed as before and the protein Sxl-RBD4 detected by Western blot.

Recombinant *Drosophila melanoganster* Sxl protein fragment (amino acids 122-301, *Dme* RBD4) was kindly donated by Prof. Dr. Jan Medenbach and Dr. Rebecca Moschall (Regensburg University, DE). Detailed protocol for the preparation and purification of the Sxl-RBD4 is available in [92].

2.18 Quantification of PTex

mRNA interactome capture samples from +UV HEK293 cells were prepared as described in section 2.6 (hereafter RIC samples), and used as a pre-purified "only-clRNPs" starting material for PTex [61]. RIC samples from five biological replicates were extracted and precipitated using the PTex protocol. PTex pellets were washed once with 5 mL of cold ethanol to remove traces of phenol which

could interfere with the quantification, and dissolved in 20-50 µL of water.

RNA and Protein quantification were done by spectroscopy and densitometry analysis of Western blot images. Absorbance was measure at λ280 nm and λ260/280 nm in a Nanodrop 2000. For quantification by densitometry, 2% and 45% of RIC and PTex samples, respectively, were digested with RNaseA (0.1 µg/µL, 37 °C, 40 min), electrophoresed in Bis-Tris-MOPS gels 4-12% (NuPage, Invitrogen), transferred to nitrocellulose membranes and blotted to detect HuR (as in section 2.15). Unsaturated TIF 8-bit images were used to quantify the yield and specific-clRNP enrichment of PTex by densitometry analysis using ImageJ [91] as described above.

2.19 Immunoprecipitation and RNA labelling by T4 Polynucleotide Kinase (PNK)

The FLAG-tagged proteins, CsrA [85], YigA, ClpX, DnaJ, UbiJ [30], and HTF-tagged AhpC, SipA, YihI [61] were immunoprecipitated using anti-FLAG magnetic-beads as described in [85; 61; 30]. Extensively washed beads were next mixed with 100 µL T4 PNK (0.1 U/µL, in PNK buffer, NEB) and 5.5 µCi γ^{32}P-ATP, at 37 °C for 30 min. Replicates of the samples were incubated in the absence of the enzyme as a control for auto-phosphorylation. Beads were washed twice in PNK buffer, and proteins eluted with loading buffer (using a magnetic rack), electrophoresed in Bis-Tris-MOPS gels 4-12% (NuPage, Invitrogen) and transferred to nitrocellulose membranes. Radioactive signal was detected by phosphoimaging in a Life Science FLA-5100 imaging system (Fujifilm). Membranes were then blotted with anti-FLAG antibody as a control to detect the exact position of the proteins.

2.20 Mass spectrometry sample preparation

HEK293 and *Salmonella* PTex samples were subjected to mass spectrometry (MS). UV cross-linked RNA-protein complexes from HEK293 cells (8x10^6 cells per sample per UV treatment) and *Salmonella* Typhimurium SL1344 (3 ODs per sample) were purified with PTex and Hot-PTex, respectively [61]. Samples were

processed further and subjected to MS by Carlos Vieira and Prof. Dr. Matthias Selbach (Max-Delbrück Center for Molecular Medicine, Berlin, DE) as described in [61]: after ethanol precipitation, PTex pellets were dissolved in ABC buffer (2 M urea, 50 mM ammonium bicarbonate) and RNA digested with Benzonase for 30 min at 37 °C. Input controls were prepared from a minor fraction of initial input material, lysed, denatured in 1% SDS and 0.1 M DTT Phosphate Buffer Solution (PBS) by boiling for 10 min at 95 °C, and RNA digested with Benzonase as before. Silver staining was performed as described in [61].

For MS, proteins were precipitated with methanol-chloroform extraction [93] and resuspended in 8 M urea and 0.1 M Tris pH 8 solution. Proteins were reduced (10 mM DTT, 30 min), alkylated (55 mM iodoacetamide, 30 min, in the dark), and digested with lysyl endopeptidase (LysC, Wako) and Trypsin (Promega). Peptides were desalted with C18 Stage Tips [94] and separated on monolithic column (100 µm ID x 2,000 mM, MonoCap C18 High Resolution 2000 [GL Sciences], or on an in-house made C18 15 cm microcolumns (75 µm ID packed with ReproSil-Pur C18-AQ 3 µm resin, Dr. Maisch GmbH). Detailed columns and instrument operation settings, and gradients run time are available in [61].

Raw files were analyzed with MaxQuant software (v1.5.1.2) [95] using the label free quantification (LFQ) algorithm [96] with default parameters and match between runs option on. Database search was performed against the human reference proteome (UNIPROT version 2014-10, downloaded in October 2014) or the *Salmonella* Typhimurium reference proteome (UNIPROT version 2017, downloaded in August 2017) with common contaminants. False discovery rate (FDR) was set to 1% at peptide and protein levels.

2.21 Bioinformatic analysis of PTex-purified proteins

Bioinformatic analysis of the LC-MS/MS generated data was performed by Timon Hick and Dr. Benedikt Beckmann (Beckmann lab, IRI Life Sciences, Humboldt-Universität zu Berlin, DE). The complete analysis is available as an R notebook in the supplementary information from [61]. In short, after label-free

quantification (LFQ) intensities were normalised to trypsin (which is constant in all samples); potential contaminants, reverse and peptides only identified by modification were excluded from analysis.

Fold changes were calculated by subtraction of the $\log_2$ values of LFQ intensity for proteins from UV cross-linked samples and non-cross-linked samples. Only proteins which were found in all replicates were processed further [61]. Enrichment (+UV/-UV) was calculated as described before [97]: P-values were calculated from an Ebayes moderated t-test using the limma package [98] followed by Benjamini-Hochberg False Discovery Rate (FDR) correction. Only proteins with an adjusted p-value of 0.01 or smaller in all 3 cross-linking intensities were considered being enriched. gene ontology (GO) analysis was performed using PANTHER V.11 [99]. Domain enrichment was done using DAVID [100] searching the SMART [101] database.

For the analysis of *Salmonella*, iBAQ-normalised values were used instead. Similarly, potential contaminants, reverse and peptides only identified by modification were excluded from the analysis. Fold changes were calculated by subtraction of the $_2$ values of iBAQ intensity for proteins from UV cross-linked samples and non-cross-linked samples [61]. Only proteins which were found in both replicates were taken into account (258 proteins; 172 with a $_2$ fold-change >0). Domain and GO terms were analysed using DAVID [100].

2.22 Using PTex to simplify the PAR-CLIP protocol (pCLIP)

PAR-CLIP protocol was performed following the methods established by [32; 102], and described in [61]: HEK293 cells stably expressing FLAG-tagged HuR (ELAVL1) were grown until 80-90% confluence. For the last 16 h of incubation, 200 mM 4-thiouridine (4-SU) was added. Living cells were irradiated with 0.15 J/cm^2 365 nm UV light, snap-frozen on dry ice and stored at -80 °C until use. Cells were collected on different days, representing biological replicates.

Cells (~1.2 x 10^8 cells/replicate) were lysed on ice for 10 min with 3 mL lysis buffer (50 mM Tris-Cl pH 7.5 (Life Tech., 15567027), 100 mM NaCl (Life Tech. AM9760G), 1% (v/v) Nonidet P40 substitute (Sigma 74385), 0.5% (v/v) Sodium deoxycholate (AppliChem No. A1531) containing 0.04 U/mL RNasin (Promega, N2515) and 2x Complete Protease Inhibitor (Roche, 11697498001). Lysates (1.5 mL/replicate) were cleared by centrifugation (20,000 x g, 10 min, 4 °C), digested with 8 U/mL TURBO DNase (ThermoFisher, AM2238) and 2 U/µL RNase I (ThermoFisher, AM2294) at 37 °C for 4 min (replicate 1) or 3 min 15 sec (replicates 2 and 3).

FLAG-tagged HuR was immunoprecipitated with 10 µg of anti-FLAG monoclonal antibody (Sigma, F1804) bound to 40 µL of Protein G Dynabeads (Life Tech, 10004D). After extensive washes with high-salt buffer (50 mM Tris-HCL pH 7.5, 1 mM EDTA, 1% Nonidet P40, 0.1% (v/v) SDS, 1 M NaCl) beads were incubated with 1 U/µL of T4 polynucleotide kinase (T4-PNK, NEB) and 0.5 µCi/µL of ^{32}P-ATP. After radiolabelling, samples were splitted into four tubes and underwent three versions of CLIP as described in [61]:

2.22.1 PAR-CLIP classic

Briefly, clHuR-RNA complexes were resolved by 4-12% Nu-PAGE MOPS (invitrogen) transferred to nitrocellulose and excised at a defined size-range (50 to 60 kDa). Proteins were digested from the membrane with proteinase K, and the RNA was recovered by acidic phenol/chloroform extraction with subsequent ethanol precipitation. Resulting RNAs were ligated with 3' adapter (5'App-NNN NTG GAA TTC TCG GGT GCC AAG G-3'InvdT) gel-slice isolated, ligated with the 5adapter (5'-GUU CAG AGU UCU ACA GUC CGA CGA UCN NNN-3') and purified by elution from PAGE-Urea gels followed by ethanol precipitation [32; 102].

2.22.2 PAR-CLIP on-beads

As an alternative, the ligation of the 3′/ 5′ adapters can be achieved directly on the beads used in the affinity capture of the selected clRNP [103; 104]. On-beads adapters ligation was done by incubating the FLAG-clHuR-RNA beads with the 3′ adapter in the presence of Rnl2(1–249) K227Q ligase and

PEG-8000 overnight at 4 °C. After washes, the 5adapter was ligated using the Rnl1 enzyme, 2 h at 37 °C as described in [104]. Followed by protein electrophoresis (4-12% Nu-PAGE MOPS, Invitrogen), blotting to nitrocellulose membranes and band excision at the defined size-range (50 to 60 kDa). Proteins were digested from the membrane with proteinase K. RNA was then recovered by acidic phenol/chloroform extraction and ethanol precipitation [61].

2.22.3 pCLIP

Beads capturing clHuR-RNA complexes were subjected to 3′/ 5′ adapters ligation as above. Immediately after the 5′ adapter ligation beads were magnetically captured and resuspended in 600 µL solution D. After a short denaturation (95 °C during 10 min), beads were separated with a magnet and the clHuR-RNA complexes recovered in the supernatant were subjected to the last two steps of the PTex protocol (20-25 min). PTex recovered interphases were precipitated with 9 volumes of ethanol on dry ice during 30 min, followed by 30 min centrifuging at 20,000 xg, 4 °C. Ethanol-precipitated pellets were digested with proteinase K and RNA isolated by phenol/chloroform extraction [61].

2.23 Library preparation and RNA sequencing

RNAs obtained from the three PAR-CLIP procedures (classic, on-beads, and pCLIP) were retro-transcribed into cDNA with the reverse transcription primer 5′-GCC TTG GCA CCC GAG AAT TCC A-3′ and the minimal PCR cycles were determined for each case. cDNA libraries were created by PCR using the forward primer 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACG TTC AGA GTT CTA CAG TCC GA-3′ and the Illumina adapter index RPI 1-6, 8, 10-11. Bands obtained at around 150 bp were excised from 2% agarose gels and purified using the Zymoclean Gel Recovery Kit (Zymo, D4002). DNA concentration and library quality was determined by Qubit Fluorometer dsDNA HS assay (Life Tech, Q32854) and BioAnalizer DNA HS Kit (Agilent 2100 Bioanalyzer; Agilent, 5067-4626). Libraries were sequenced in a NextSeq 500 [61].

2.24 CLIP data processing

The PAR-CLIP data was processed and annotated by Dr.Hans-Herman Wessels and Prof. Uwe Ohler (MDC-Berlin, DE) using the PARpipe pipeline (https://github.com/ohlerlab/PARpipe) around the PAR-CLIP data tailored peak caller PARalyzer [105] as described previously [106]. PAR-CLIP alignments were visualized using Gviz [107]. the entire procedure is deeply explained in [61].

Chapter 3

Results

In recent years, methodological advances fostered the discovery of hundreds of RNA-binding proteins (RBPs) and their RNA-binding sites. However, all those methods rely on either capturing proteins associated with polydenylated RNAs [14; 15], or incorporating labels into the RNA (e.g. click chemistry) to allow the capture of the complexes [56; 57]. Therefore, the identification of RBPs bound to other RNA classes (pre-mRNA, tRNA, rRNA) and especially, those from organisms in which polyadenylation of the mRNA is generally restricted, as in bacteria and archea, has been profoundly neglected.

Proteins and nucleic acids show differences in their mobility when in contact with organic-aqueous mixtures. That is the foundation of the protocol for RNA isolation developed by Chomczynski and Sacchi more than thirty years ago [64; 65], where cells lysed in a highly chaotropic denaturing solution are subjected to organic extraction with phenol and chloroform. In this context, the hydrophobic regions of proteins promote their partitioning to the organic phase, DNA enriches in the interphase (due to the acidic pH of the solution), while the RNA remains in the aqueous phase [66]. This differential behavior between proteins and nucleic acids depends only on their physicochemical properties.

Answering the question whether covalently cross-linked RNA-protein complexes behave as proteins, as RNA, or whether they adopt a different partitioning pattern led to the development of phenol-toluol extraction (PTex) as a method for the unbiased purification of UV cross-linked RNA-protein complexes (clRNPs). In this chapter, the development of PTex, its scope, limitations and applications

are presented. Parts of the results here described have been published in our works [61; 62; 30] (Appendix B).

3.0.1 Phenol *vs.* phenol-toluol extractions

To start developing the method, I focused on understanding how the different chemical reagents influence the partitioning of the clRNPs during organic extractions. Therefore I evaluated protein phase partitioning using either (i) phenol, or (ii) phenol-toluol as organic phase, with either a neutral buffer (PBS, pH 7.4), or with a denaturing solution (solution D, pH 4.8, methods 2.1) as for the aqueous phase (Fig. 3.1a).

In order to trace the mobility of proteins and clRNPs across phases, I selected two proteins: the protein HuR (ELAVL1, 35 kDa), for being a well established RNA-binding protein [108] which cross-links efficiently at 254 nm UV light; and beta-actin (ACTB, 42 kDa) as an example of an abundant protein which does not bind RNA.

Western blots revealed an accumulation of HuR in the interphase and the appearance of a strong signal at the upper part of the gel (Fig. 3.1bd, cd, be). Due to the UV irradiation, the RNA-protein molecules increased its molecular weight resulting in a reduced mobility when electrophoresed, explaining the accumulation of HuR signal at the upper end of the gel (pockets). Importantly, the signal was specifically reversed after RNase digestion with RNase A (Fig. 3.2).

The UV-dependant increase of the molecular weight of clHuR served as a beacon for tracking the mobility of the clRNPs between phases during the exploratory extractions. As partially lysed cells, membranes, and other cellular debris accumulated in the interphase, a major fraction of free and cross-linked HuR (clHuR) was found in this phase, which in turn masked the real migration of the clRNPs across phases [62].

Nonetheless, two different behaviours became evident: clHuR appeared in the aqueous phase only in extractions using the combination of phenol-toluol with

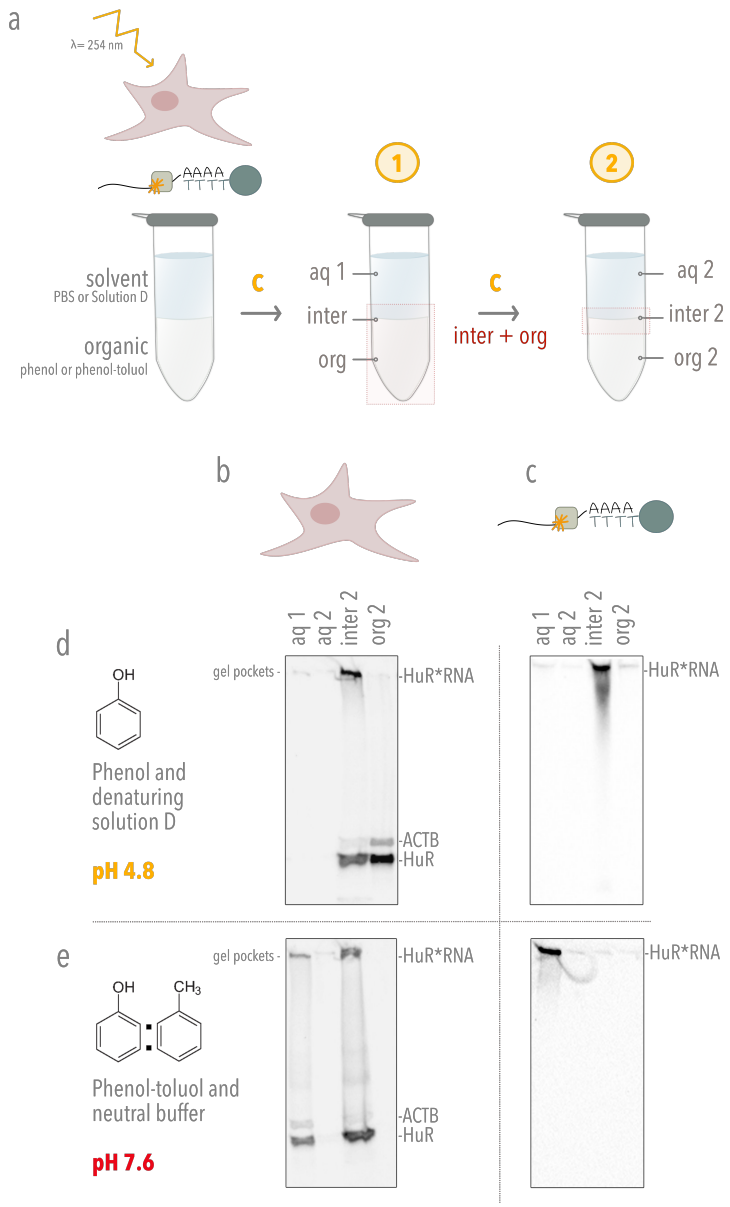

Figure 3.1: Development of the PTex method. **(a)** Scheme of the exploratory extractions: UV-irradiated cell pellets dissolved in either denaturing solution D (pH 4.8) or neutral buffer (PBS, pH 7.4), were mixed with phenol or phenol-toluol, respectively, and centrifuged (20,000 xg, 4°C, 3 min, yellow c letter). Three phases are visible: aqueous-, inter-, and organic phase (aq1, inter1, org1). Next, the upper aqueous phase was removed, and the inter- and organic phases were mixed and re-extracted with ethanol and water (aq2, inter2, org2). **(b)** Western blotting shows that phase composition alters the partitioning of HuR-RNA complexes (HuR*RNA) during the extractions: in the context of cell lysates **(b)** or RIC samples **(c)**, phenolic extraction carried out under denaturing conditions at pH 4.8 **(d)** promotes the migration of HuR*RNA to the interphase 2 **(db, dc)**, while ACTB and unbound-HuR can be detected in the organic phase 2 **(db)**. On the contrary, when a mixture of phenol-toluol and neutral buffer is used **(e)**, the HuR-RNA complexes accumulate in the aqueous phase **(ec, eb)**. ACTB and unbound-HuR can be also detected in this phase **(eb)**. Notice that clHuR has a reduced electrophoretical mobility (signal at the pockets of the gel) that can be reverse by RNase A treatment (as shown in Fig. 3.2). Modified from [62].

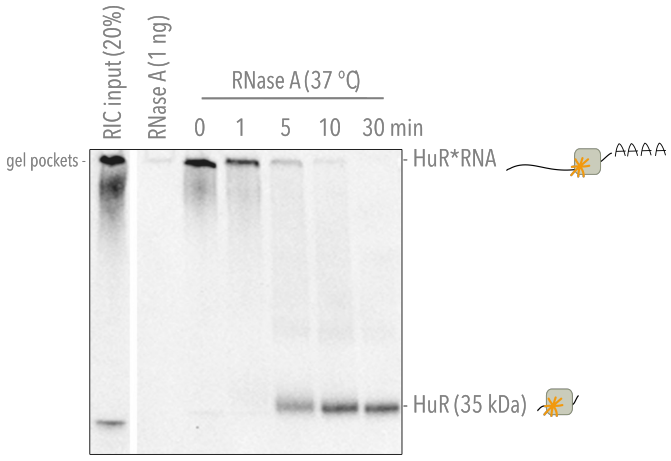

Figure 3.2: Electrophoretic mobility shift assay. Western blot detecting the protein HuR (ELAVL1; 35 kDa). Interphase containing UV cross-linked RNA-protein complexes obtained by extracting 15 µg of a pre-purified (m)RNA-interactome capture (RIC) solution with phenol and solution D. Samples were RNA digested with RNase A and electrophoresed. RNase-dependant restoration of the molecular weight of HuR indicates that the signal at the pockets corresponds with cross-linked HuR (clHuR). Modified from [62].

neutral buffer (Fig. 3.1be, ce); whereas during phenolic extractions under denaturing conditions, clHuR accumulated in the interphase while unbound HuR and ACTB were detected in the organic phase (Fig. 3.1bd, cd). Importantly, the finding that clHuR was differentially found in the aqueous- or inter- phase depending on the composition of the phases was corroborated when performing the extractions using a pre-purified clRNPs solution (RIC samples, methods 2.6), (Fig. 3.1cd, ce).

3.0.2 PTex: a three step method to purify UV cross-linked RNA-protein complexes

Although the presence of RNPs in the interphase of organic extractions was reported long ago [71], only recently it was exploited for the purification of clRNPs [59; 60; 61]. However, the high amount of cellular debris accumulated in the interphase of such extractions interferes with the purification process.

Therefore, I implemented a different approach: taking advantage of the previous observation that proteins and complexes partitioned to the aqueous phase (away from cellular debris) upon organic extraction with phenol-toluol at

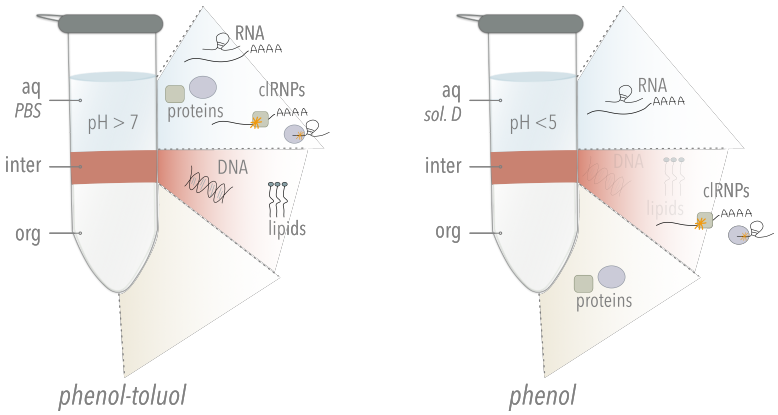

Figure 3.3: Schematic of the separation principle of biphasic organic extractions in PTex. Left: a first extraction with phenol-toluol-1,3-bromo-chloro-propane (1:1:1) and a neutral buffer results in an accumulation of soluble molecules (mainly proteins and RNA) in the upper aqueous phase (aq) while DNA and lipids are retained at the interphase (inter). Right: a second extraction with phenol and guanidine thyocianate (pH 4.8) promotes the shifting of UV cross-linked RNA-protein complexes to the interphase, while non-cross-linked RNAs remain in the aqueous phase (aq), and non-cross-linked proteins migrate to the lower organic phase (org). Modified from [61].

neutral pH (Fig. 3.1bd, be), I incorporated this chemical composition as the first step in the PTex protocol [61; 62] (Fig. 3.3, and 3.4).

The PTex method was then designed combining the different partitioning behavior of the molecules, as it was determined by the exploratory organic extraction described in the previous section. When mixing UV-cross-linked cell suspensions with phenol, toluol, 1,3-bromo-chloro-propane (BCP), and a neutral buffer, the clRNPs migrated to the aqueous phase (away from cell contaminants, Fig. 3.3 left). A second extraction with phenol and a highly denaturing aqueous phase promoted the enrichment of clRNPs in the interphase, while unbound RNAs remained in the aqueous phase and free-denatured proteins migrated to the organic phase (Fig. 3.3 right). Finally, a third extraction with phenol, BCP, ethanol and water further removed the excess of free proteins and RNA.

This migration pattern was exhibited by HuR, known for its UV cross-linking efficiency (usually >1% of the cellular protein can be cross-linked to RNA (Fig. 3.4a), and hnRNPL, which is marginally UV cross-liked (Fig. 3.4b) as only up to 0.1% of the cellular protein get cross-linked (Dr. Oliver Rossbach, University Gießen, DE).

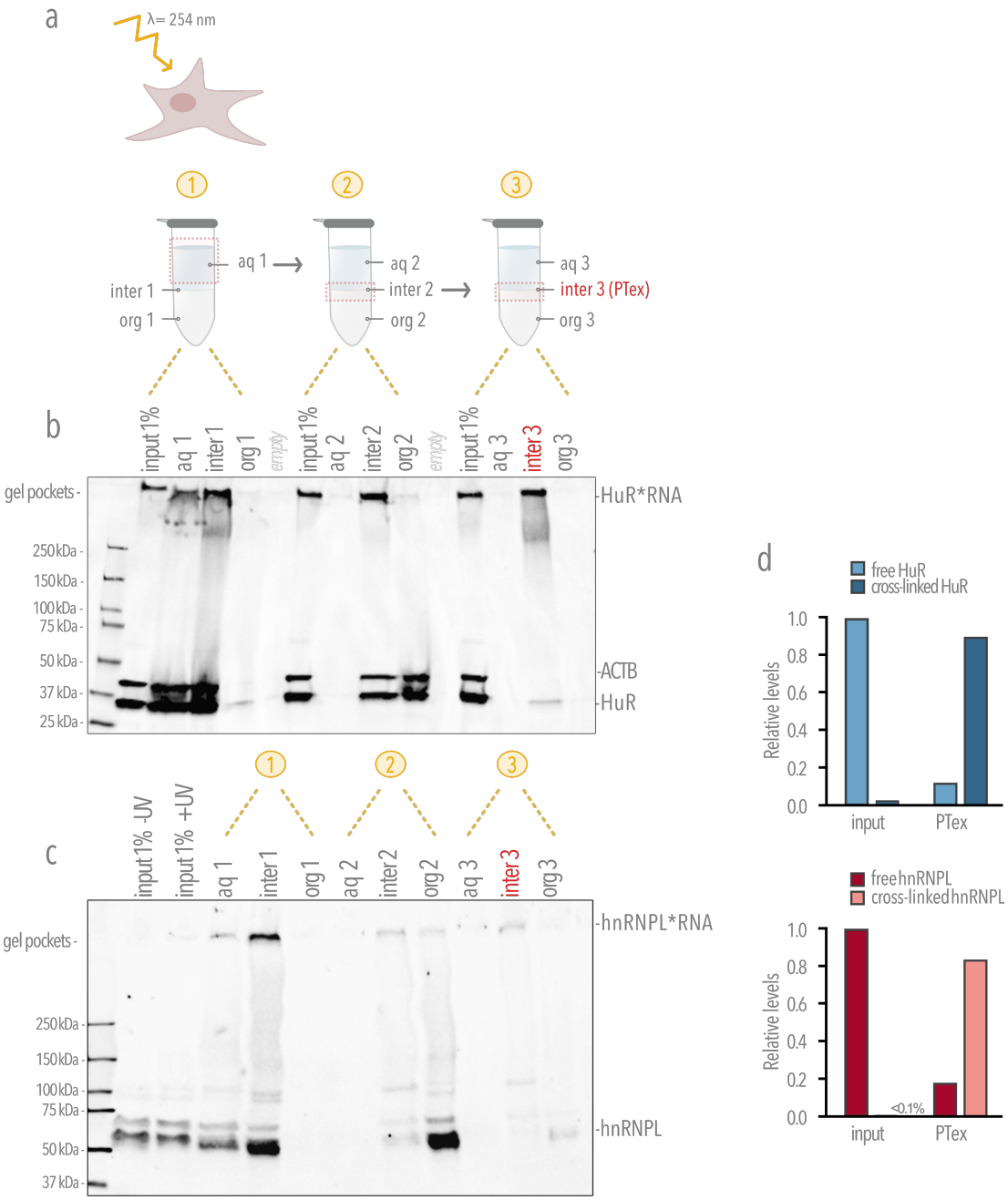

Figure 3.4: PTex, step-by-step. **(a)** UV-cross-linked HEK293 cells extracted by PTex step-by-step where all intermediary phases were precipitated and analysed by Western blot to detect two well-known RNA-binding proteins, HuR and hnRNPL. ACTB, an abundant non-RBP protein was used as control. **(b)** Western blot against HuR (ELAVL1, 35 kDa) shows that UV-cross-linked HuR-RNA complexes (gel pockets) are successfully enriched by PTex whereas the non-RBP ACTB (42 kDa) is completely removed (step 3, interphase) (aqueous phase= aq; interphase= inter; organic phase= org). **(c)** Western blot of the PTex intermediary steps showing hnRNPL (two isoforms: 64.1 and 50.5 kDa). **(d)** Quantification of (b) and (c): relative enrichment of cross-linked HuR and hnRNPL by PTex calculated as described in [62]. Modified from [62].

PTex succeeded in enriching clRNPs regardless the difference in their cross-linking efficiency (Fig. 3.4d). Important to notice is the removal of unbound protein, specially in the case of hnRNPL, where more than 99% of the protein is not cross-linked, as unbound protein can be considered contamination or background in downstream applications. PTex-purified samples were largely depleted in unbound HuR and hnRNPL and consisted of >80% of RNA-bound protein (Fig. 3.4d, interphase 3 in b and c).

Additionally, Fig. 3.5 corroborates that the accumulation of clRNPs (particularly clHuR) in the pockets of the gel corresponds with UV-cross-linked HuR-RNA complexes, and more importantly, that PTex samples can be subjected to enzymatic digestion of RNA, indicating that the method leaves the enriched RNPs amenable to downstream applications.

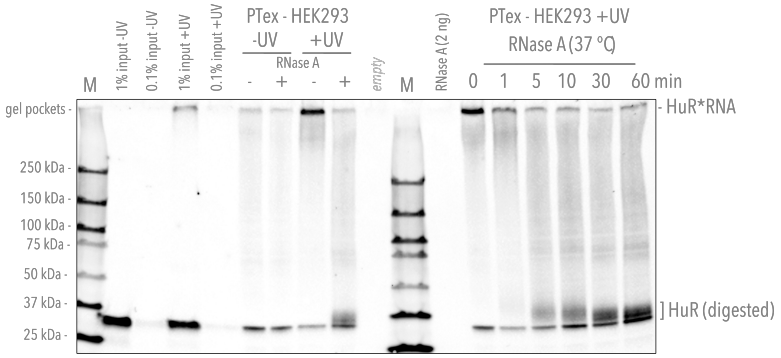

Figure 3.5: Electrophoretic mobility shift assay of PTex-purified cross-linked HuR. Western blot detecting the protein HuR (ELAVL1; 35 kDa). After UV cross-linking (+UV), a second band appears at the pocket of the gel (compare input +UV with -UV). This signal disappeared when incubating the PTex purified complexes (clHUR) with RNase A. The shifted complexes are depleted in a RNase-dependent manner (time-course, right). After 60 min incubation with RNaseA, the HuR signal is restored at a slightly higher molecular weight than unbound HuR. Depending on the RNase, a few nucleotides will remain cross-linked to the protein, causing a slower mobility in SDS-PAGE due to the additional mass of the leftover RNA. Modified from [61].

3.0.3 PTex performance, scope, and limitations

Migration of clRNPs, free-RNA, DNA, and proteins during PTex

Next, I determined the distribution of the diverse cellular molecules from total HEK293 cells during the PTex procedure by analysing all phases from the intermediary steps by Western blotting and PCR (methods 2.13). As shown before, in the PTex fraction (interphase 3), clHuR was highly enriched. On the

contrary, unbound HuR and abundant cellular proteins unrelated to RNA-binding such as beta-actin and histone H3 were not detectable (Fig. 3.6a, Fig. 3.7a,b).

To further demonstrate the efficient removal of non-cross-linked proteins, a recombinant RNA-binding domain of the *Drosophila melanogaster* Sex-lethal protein (*Dme* Sxl-RBD4, [92]), was spiked-in into the cell suspension after UV-cross-linking. PTex efficiently removed ∼99% of the unbound Sxl-RBD4, as determined by densitometry of the Western blot images compared to the input (Fig. 3.6).

Similarly, DNA and unbound RNA were efficiently removed as demonstrated by PTex using ^{32}P-5' labeled RNA (30 and 50 nt), HEK293 cells, genomic DNA from HEK293 cells, and linearised pUC19 plasmid as inputs (Fig. 3.6 b-d). Depletion of DNA was tested by PCR targeting genomic DNA (exon 5 of the IL3 gene) or plasmid DNA (lacz gene). In all cases the PTex samples were >90% depleted of DNA and RNA when comparing with input controls.

Additionally, PTex enrichment of classic and unconventional RBPs was tested by Western blot against the proteins: polypyrimidine tract binding protein 1 (PTBP1), fused in sarcoma (FUS), and the more recently identified RNA-binding enzymes glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and enolase (Eno1) [16]. Fig. 3.7a shows that all RBPs tested were enriched by PTex in a UV-irradiation-dependent manner whereas the highly abundant DNA-binding histone H3 was depleted.

Since faint signals appeared in the non-cross-linked samples when blotting against RBPs, cell suspensions were subjected to RNase digestion before applying PTex. Western blots detecting the proteins PTBP1, FUS, ENO1, histone H3, and ACTB demonstrated that RNase digestion prior to PTex abolished the enrichment for RBPs (Fig. 3.7b), consistent with the selective enrichment for RBPs in complex with RNA. It seems that some RNA-protein associations can withstand the denaturing conditions in the step 2, even in the absence of covalent bonds, and therefore being captured by PTex.

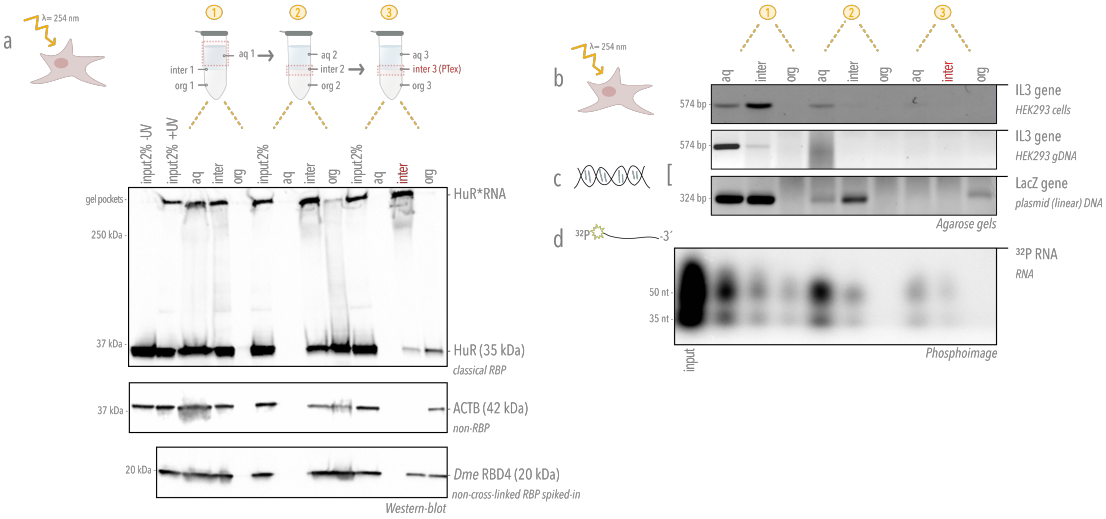

Figure 3.6: Phenol-toluol extractions as a fast method to purify cross-linked RNPs. **(a)** Schematic of organic extractions used in PTex and Western blot analysis of intermediary steps: UV radiation (254 nm) of HEK293 cells induces *in vivo* covalent bonds between RNA and proteins in direct contact. Western blot against HuR (ELAVL1, 35 kDa) demonstrates that UV-cross-linking-stabilised HuR-RNA complexes (signal at the gel pockets, denoted as HuR*RNA). HuR*RNA is largely enriched after PTex (inter 3), while the non-RNA-binderbeta-actin (ACTB) was efficiently removed (absence of signal in inter 3). A recombinant protein from *D. melanogaster* (*Dme* RBD4, 20 kDa) used as a 100% non-cross-linked RBP control (spiked-in) was ~99% removed by PTex. **(b-d)** DNA and free-RNA are also efficiently depleted by PTex: DNA carry-over was tested by PCR with specific primers against exon 5 of the interleukin 3 (IL3) gene, and the LacZ gene present in the pUC19 plasmid. Efficient removal of genomic DNA was determined as the absence of PCR product in the interphase 3 of PTex samples derived from either **(b)** UV-cross-linked HEK293 cells, or **(c)** pre-purified genomic or plasmidic DNA. "Contamination" with unbound-RNA was tested by electrophoresis of 5'- end radioactive-labeled RNA (35 and 50 nt) subjected to PTex **(d)** [61]. Modified from [61].

Summarising, PTex highly enriches for cross-linked RNPs while efficiently depleting non-RBPs, non-cross-linked proteins, and nucleic acids.

PTex: RNA-length requirements and quantification

PTex enriches for UV cross-linked RNA-protein complexes, however, how long the bound-RNA must be in order to be efficiently purified? And equally important, what is the efficiency of PTex in terms of RBPs yielded? and which parts of the procedure affect the recovery of clRNPs?

To determine the minimal RNA length required for RNPs enrichment, a set of *in vitro* transcribed RNAs varying between 13 and 191 nucleotides were produced [109; 61]. The transcribed RNAs contain one copy of an Uracil stretch of seven nucleotides at their 5' ends; a sequence which is recognised by *Dme* Sxl-RBD4

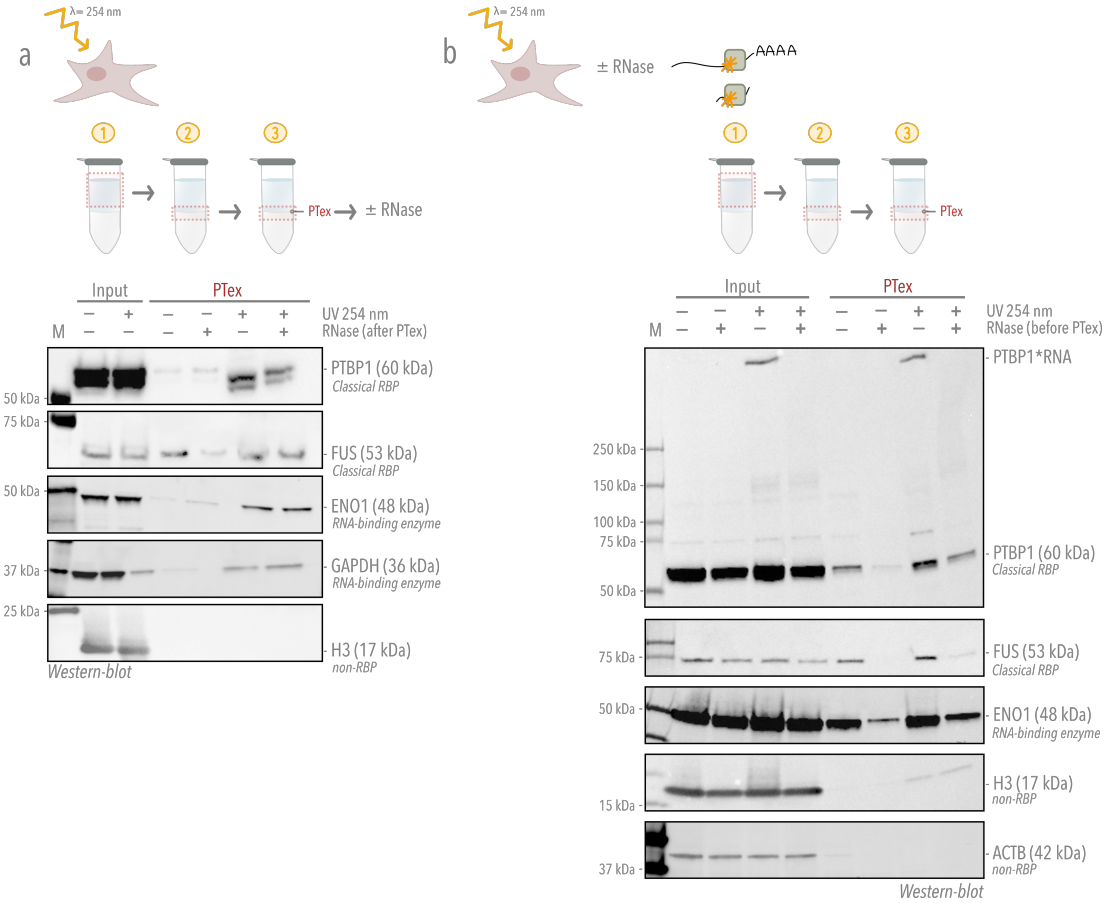

Figure 3.7: PTex enriches for classical and unconventional RBPs in an RNA dependent-manner. PTex enrichment was tested by Western blotting against the well-established RBPs polypyrimidine tract binding protein 1 (PTBP1), fused in sarcoma (FUS), and the more recently identified RNA-binding enzymes glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and enolase (Eno1)[16], and the non-RNA-binders ACTB and histone H3. RNase treatments were applied after **(a)** or before **(b)** PTex. All RBPs tested were enriched by PTex in a UV/RNA dependent manner whereas the highly abundant ACTB and DNA-binding histone H3 was depleted. Nuclease treatment of living cells before PTex abolished the recovery of RBPs **(b)**, demonstrating that clRNP enrichment by PTex is selective for RBPs complexed with RNA. Modified from [61].

(20 kDa). After binding and UV cross-linking *in vitro* (see 2.17), samples were PTex-purified and analysed by Western blotting (Fig. 3.8a). PTex efficiently recovered cross-linked *Dme* Sxl-RBD4 bound to RNA as short as 30 nt. Interestingly, a signal corresponding to unbound *Dme* Sxl-RBD4 in the PTex samples appears only in those samples which were in contact with RNA. The Western blot signal shows a steady increase dependant of the RNA length, as larger the RNA as higher the amount of protein detected. Such a signal was almost undetectable in the non-RNA control (-RNA), (Fig. 3.8a).

Answering the question about the yield rendered by PTex was indeed challenging: while cross-linked HuR generated a distinctive signal at the height of the gel pocket (Fig. 3.6a), there is no *a priori* indication about the amount of protein which is efficiently cross-linked to RNA by UV light; more over, not all RBPs are functioning as RBPs in a given time [61; 62]. To tackle this problem, mRNA interactome capture samples from HEK293 cells were used as a clRNPs 100% cross-linked to poly-A RNA. Using this samples as input for PTex allowed the quantification of the purification of HuR by densitometry of Western blot images (Fig. 3.8b). PTex recovered ∼30% of the initial amount of cross-linked protein.

Additionally, the recovery of Sxl was also calculated by densitometry of the Fig. 3.8a. Here, PTex yielded ∼50% of the total cross-linked protein. As PTex purification efficiency varies among individual proteins (Fig. 3.8c-f), protein and RNA quantification based on its absorption at 260 nm and 280 nm were also determined by spectroscopy of the PTex samples *vs* inputs (Fig. 3.8g). Consistently, PTex recovered 27% of the RNA and 33% of the proteins from HEK293 RIC samples.

Another aspect to consider is, how the precipitation and solubilisation steps affect the recovery of the complexes? To answer this question, fractions of the same cell lysate were precipitated using three of the most common protein/nucleic-acids precipitation methods: ethanol, 2-propanol and trichloroacetic acid (TCA), then the pellets dissolved in either water, TE or TED buffers were analysed by spectrophotometry (absorbance at λ280 nm) to quantify the protein content (see section 2.12).

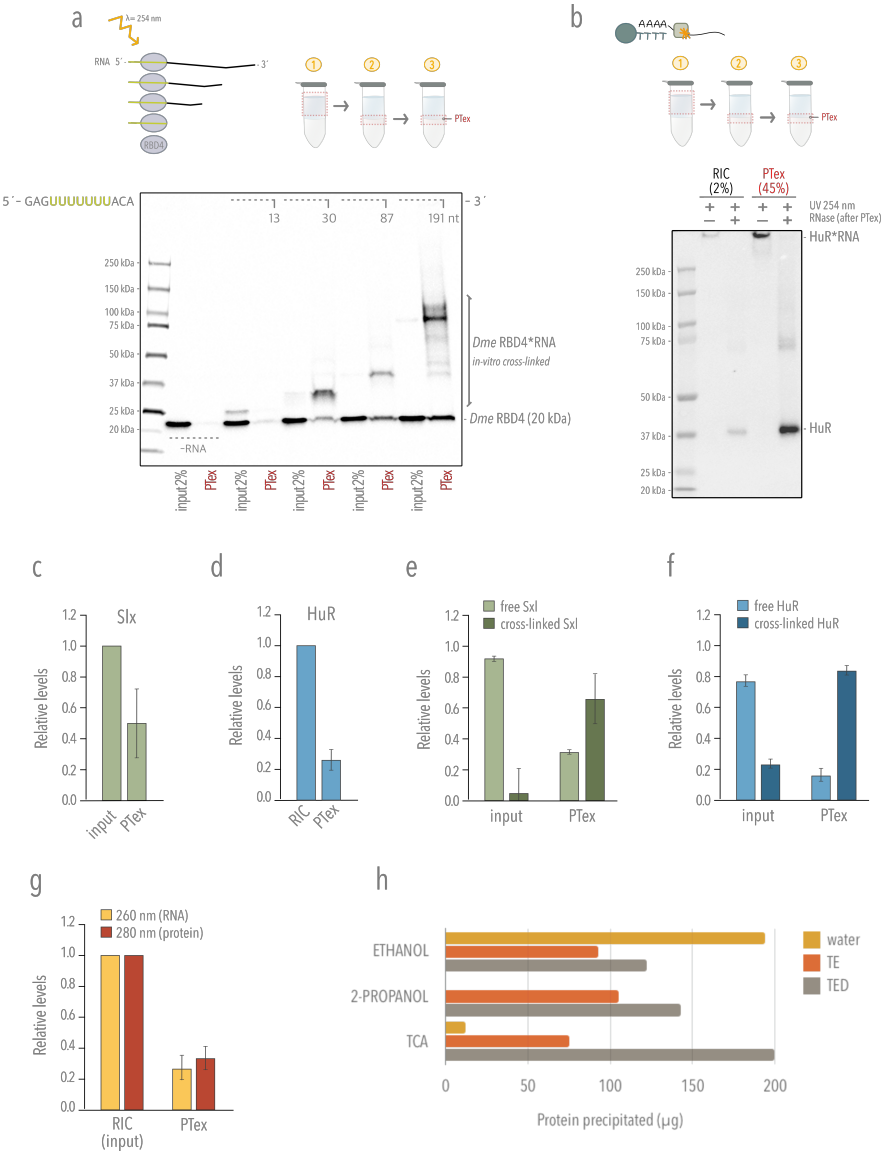

Figure 3.8: PTex performance (a) *Drosophila melanogaster* RBP Sxl RBD4 was *in vitro* bound and cross-linked with RNA of different lengths containing the Sxl U7 recognition site. A minimum of 30 nt are required for an efficient recovery of clRNPs by PTex. **(b)** (m)RNA-interactome capture (RIC) samples from HEK293 cells were used as 100% cross-linked input material and the recovery of clHuR *vs.* unbound HuR by PTex was assessed. **(c)** Quantification of a (n=3, error bars represent SD). **(d,f)** Relative enrichment of cross-linked Sxl (from 3.8a) and clHuR (from Fig. 2.2); n=3, error bars represent SD. **(e)** Quantification of b (n=3, error bars represent SD). (g) Protein and RNA quantification of PTex samples from (b) determined by UV spectroscopy (n=3, error bars represent SD). **(h)** Comparison of protein precipitation methods: lysates from HEK293 cells were precipitated by either ethanol, isopropanol or trichloroacetic acid (see methods section 2.12). Pellets resuspended with 100 µl of water, TE or TED, incubated 20 min at 56 °C and centrifuged 1000 xg 2 min. Supernatant= soluble fraction; pellet= insoluble fraction. TE= 20 mM Tris, 1 mM EDTA, pH 7.8; TED= TE supplemented with 0.03% *n*-Dodecyl β-D-maltoside (DDM). Mean from two independent experiments. [62]. Modified from [61; 62].

Although all three methods yielded similar total protein recovery (Fig. 3.8h), precipitation by TCA produced a rather insoluble pellet (Fig. 3D in [62]). Even though the precipitation methods tested yielded a low protein recovery (input consisted in 5.4 mg of protein), alcoholic precipitations resulted in an easier to dissolve pellet with a better preserved RNA [62]. Additionally, the precipitation method and the solubilisation strategy must be carefully considered for the preparation of PTex samples for mass spectrometry as RNA and traces of the reagents could compromise the integrity of the LC columns [61].

3.1 Beyond the HEK293 mRNA-bound proteome

3.1.1 Sample preparation for protein mass spectrometry

Once the scope and limitations of the PTex protocol were established, the next step was to apply PTex to explore the RNA-protein interactions in HEK293 cells. PTex samples were prepared from HEK293 cells irradiated with 0.015, 0.15, and 1.5 J/cm^2 of UV light at 254 nm wavelength (Fig. 3.9a). Mass spectrometry and label free quantification (LFQ) were performed using non-crosslinked PTex samples and total protein preparations as input controls (Fig. 3.10, [61]).

The quality of the HEK293 PTex samples used for mass spectrometry (MS) was assessed by protein electrophoresis and silver staining (Fig. 3.9b). Additionally, the effects of prolonged UV radiation was also tested by Western blotting (Fig. 3.9c). Comparing the total protein signal in inputs and whole cell lysates (Fig. 3.9b and c, respectively), the increased radiation energy used for UV cross-linking evidently caused an overall decrease of protein content. Judging by the reduced signals from HuR and hnRNPL at 1.5 J/cm^2, sufficiently high energy radiation can negatively affect the RBP recovery by PTex.

Important to notice is, however, that the shifted clRNPs signal at 1.5 J/cm^2 was resistant to DNase I but not resistant to a subsequent Benzonase treatment (Fig. 3.9c (gel pockets)). Although UV-induced protein-protein cross-linking is possible under certain conditions [73; 77; 110; 111] (discussed in [62]), the aggregates detected after UV radiation are *bona fide* UV cross-linked

RNA-protein complexes.

3.1.2 Mass spectrometry data analysis of PTex-purified clRNPs

A challenging task in biology is to reliably quantify cell-wide differences between physiological or induced states. In this particular case, the challenge was to accurately determine the very proteins enriched by PTex in a UV-dependent fashion. Since LFQ intensities are calculated assuming that all peptides were present, ionized, and detected with equal efficiency, and are normalised across samples, the LFQ intensity values were further normalised against trypsin to establish a base-line for the enrichment analysis (details in [61]).

Additionally, a stringent criterion for curating the mass spectrometry (MS) data was applied: proteins which were not identified by MS in all 3 PTex replicates were removed. Ratios of cross-linked over non-cross-linked (+UV/-UV) LFQ intensities (from the PTex experiments) were calculated for the remaining proteins, and a moderated t-test was used for multiple testing (Benjamini-Hochberg) (Fig. 3.10a, [61]).

As a result, 3037 proteins were identified as significantly enriched in a UV-dependent manner (FDR 0.01) (Fig. 3.10b). Importantly, PTex enrichment was unbiased towards protein chemical features or expression levels, such as cellular abundance, protein molecular mass, isoelectric point, or hydrophobicity (Fig. 3.10c-e).

As a proxy for the validation of novel RBPs, two proteins identified as RNA binders by PTex were challenged: the ATP-binding cassette sub-family F member 2 (ABCF2, member of the AAA+ ATPase family), and the T-complex protein 1 subunit eta (CCT7, member of the chaperonin CCT/TRiC complex involved in telomere maintenance [112]). Western blots of PTex samples show that both proteins are enriched after UV-irradiation *in vivo* and PTex purification (Fig. 3.10f).

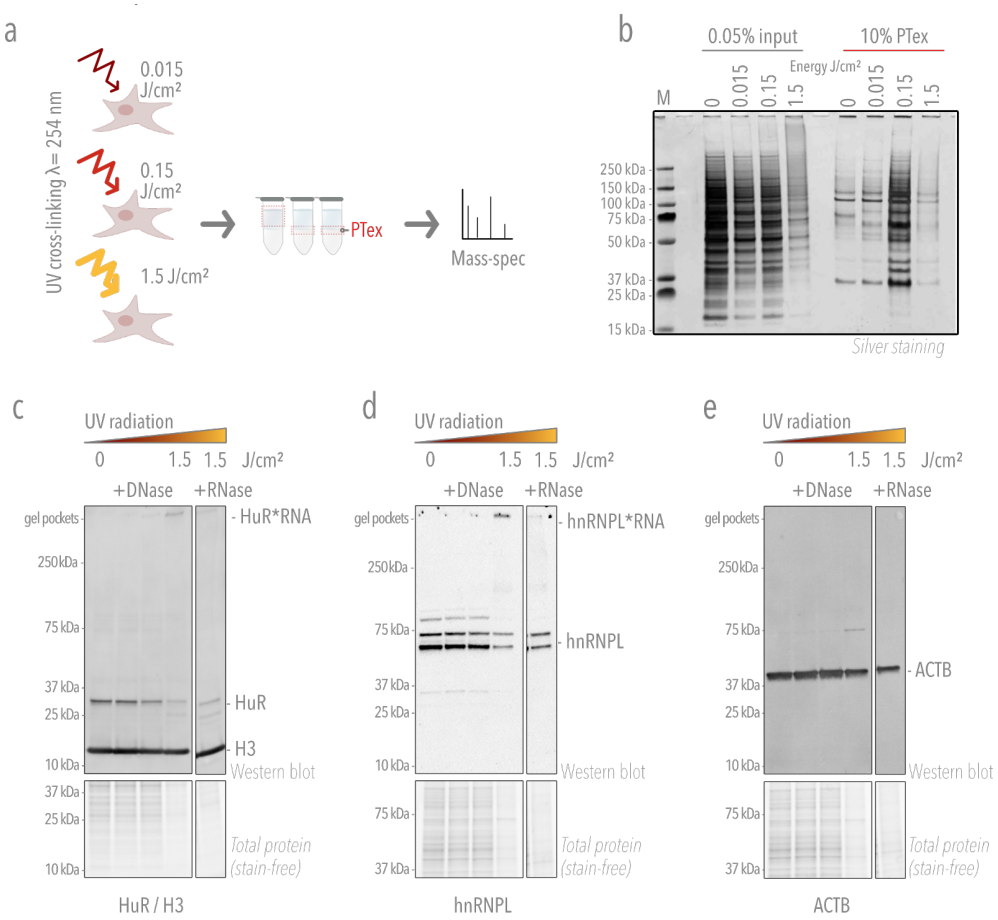

Figure 3.9: PTex sample preparation for mass-spectrometry. **(a)** HEK293 cells UV-cross-linked using 0 (noCL), 0.015 (dark red), 0.15 (red) and 1.5 (dark yellow) J/cm^2 254 nm UV light in triplicates, are purified by PTex and precipitated by ethanol. **(b)** SDS-PAGE (silver staining) of PTex pellets as used for mass-spectrometry (details on PTex sample preparation for MS in [61]. **(c)** Western blots of RBPs from UV-irradiated HEK293 cells (hnRNPL and HuR) showing that i) UV-cross-linking is rather inefficient as the main fraction of tested RBPs remains non-cross-linked. ii) cross-linking induced a shift in the molecular weight of HuR and hnRNPL due to the covalently bound RNA, as demonstrated by disappearance of the signal at the pockets after treatment with Benzonase but not with DNaseI. iii) UV radiation with 1.5 J/cm^2 results in a general loss of protein (gels in the lower panel), evident also in **(b)** and the Western blot detecting the non-RBP ACTB (compare 1.5 J/cm^2 with lower dosages [61; 62]. However, histone H3 (used as additional loading control), its relative abundance seems not to be affected by the UV radiation applied. SDS-PAGE and silver staining in figure **b** was performed by Carlos Vieira (Selbach lab, MDC-Berlin). Modified from [62].

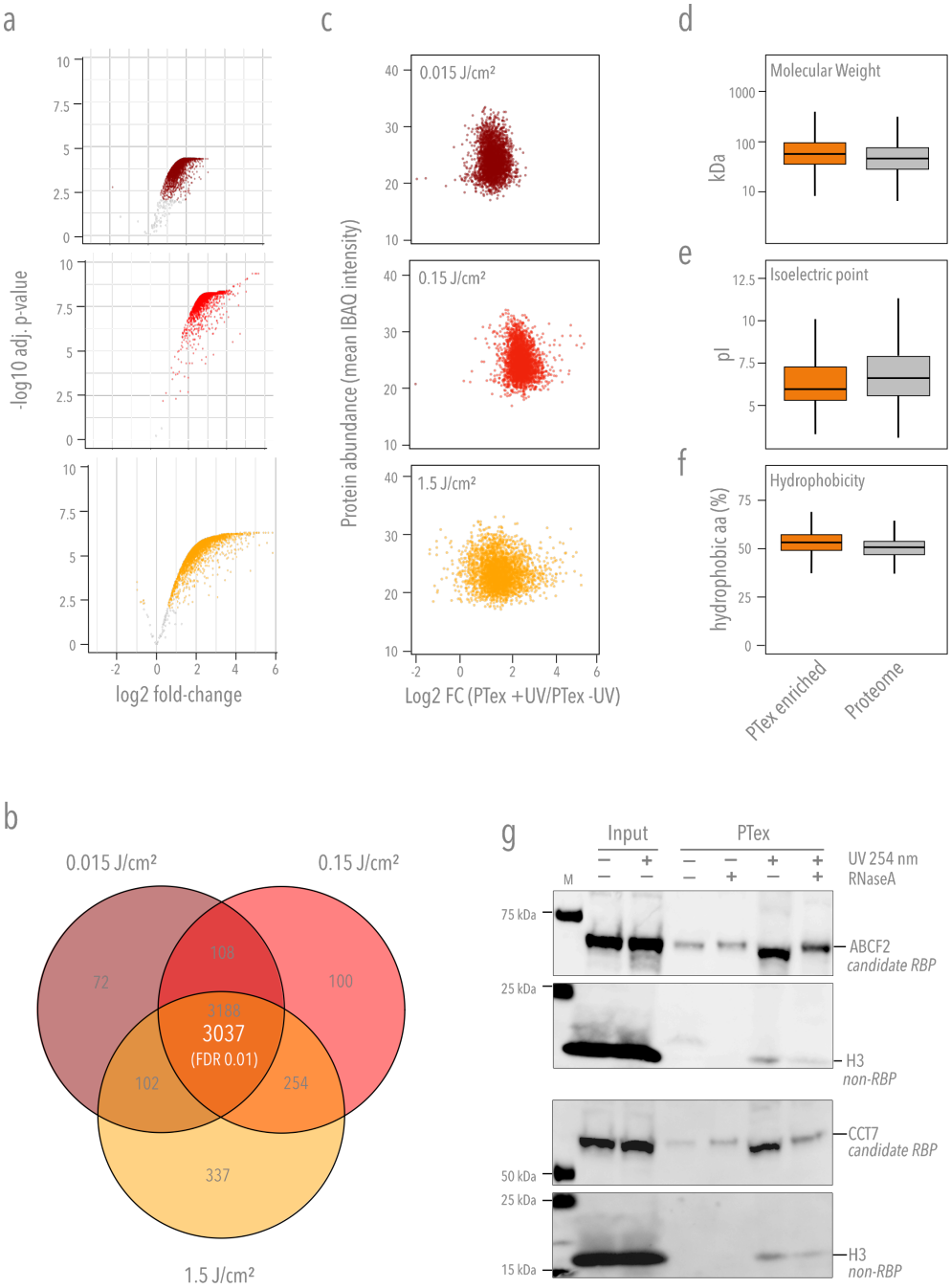

Figure 3.10: A global snapshot of RNPs in HEK293 cells. **(a)** Volcano plots of proteins enriched by PTex (FDR 0.01) of HEK293 cells UV-cross-linked with 0 (noCL), 0.015 (dark red), 0.15 (red) and 1.5 (dark yellow) J/cm^2 254 nm UV light, analysed by label-free mass spectrometry. **(b)** Protein abundance (IBAQ intensities of input samples) does not correlate with PTex enrichment (log2-fold change of intensities [+UV/-UV]). **(c)** Protein enrichment by PTex is independent of general features such as molecular weight, isoelectric point (pI) or hydrophobicity, boxplot centre line represents median, bounds are first and third quantile, and whiskers extend to 1.5 times the inter-quantile range [61]. **(e)** PTex of individual predicted RNA-associated proteins. ATP-binding cassette sub-family F member 2 (ABCF2) and T-complex protein 1 subunit eta (CCCT7) have not been reported to bind RNA. Both are enriched after PTex in a UV-irradiation-dependent fashion, indicating that they indeed associate with RNA *in vivo* [61]. Modified from [61].

It has previously been shown that extended UV exposure increases the likelihood of cross-linking events [16]. Therefore, a gradual increase in enrichment of RNA-interacting proteins after PTex with higher UV dose was expected. This was partially true for the energies used in this study: the enrichment on RNA-associated proteins steadily increased from 0 to 0.15 J/cm^2. However, irradiation with 1.5 J/cm^2 caused a decrease in recovery (Fig. 3.11a).

Additionally, analysis of the integrity of RNA purified from cells exhibited the highest RNA degradation in the case of RNA from HEK293 cells irradiated with 1.5 J/cm^2 254 nm UV light (Fig. 3.11b). In overall, the results from Figures 3.8a, 3.9, and 3.11, indicate that prolonged UV exposure can trigger breaks in the RNA and/or induce the degradation of proteins, leading to an impaired recovery of UV cross-linked RNA-protein complexes by PTex.

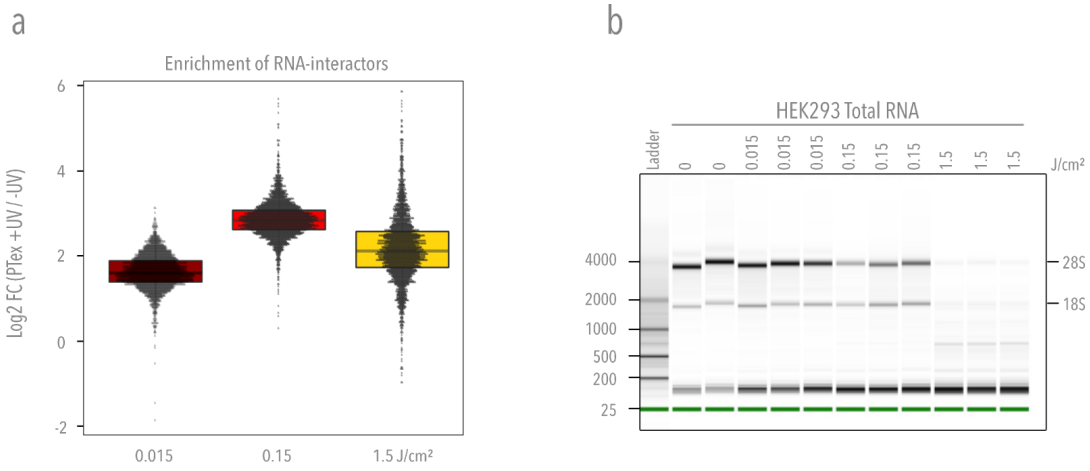

Figure 3.11: Effects of prolonged UV exposure on ribonucleoprotein complexes enrichment and RNA integrity. **(a)** UV exposure *vs.* enrichment of RNA-associated proteins in HEK293 cells. Increasing enrichment of RNA-associated proteins was detected when irradiating cells from 0 to 0.15 J/cm^2 of 254 nm UV-light. A higher exposure (1.5 J/cm^2) caused a reduction of the overall enrichment. Boxplot centre line represents median, bounds are first and third quantile, and whiskers extend to 1.5 times the inter-quantile range. n = 3 biologically independent experiments. **(b)** Analysis of RNA integrity by Agilent Bioanalyzer 2100 Total RNA Pico Chip. Total RNA was purified from HEK293 non-irradiated cells or cells radiated with 0.015, 0.015, 0.15, or 1.5 J/cm^2 254 nm UV light. Ribosomal RNA peaks are used as indicative of RNA integrity.

3.1.3 HEK293 RNA-interacting proteins

Given that the whole proteome of HEK293 consists on 10,500 proteins [113], and that the count for well-established and recently identified eukaryotic RBPs has raised from 700 to around 2000 in the last years (reviewed in [50] and [12]), finding more than 3000 proteins significantly enriched by PTex was at first unexpected. However, it is important to consider that, although the majority of the current methods used to purify clRNPs rely on UV-cross-linking, many only recover a subset of RNA interactors (e.g. poly-A-binding proteins); therefore, broadening the spectrum of RNA-associated proteins via an unbiased approach is not unrealistic.

In order to assess the sensitivity and specificity of PTex, a series of analysis were performed (Fig. 3.12):

- **Gene ontology (GO) analysis:** proteins enriched by PTex are involved in all aspects of RNA biology (Fig. 3.12a), whereas protein unrelated to RNA biology such as transporters and (trans-)membrane proteins were depleted (Supplementary Table S2 in [61]).

- **Analysis of protein domains:** RNA-binding domains were significantly enriched in PTex-purified proteins (Fig. 3.12b). The domains found cover a wide range of classic and novel RNA-recognizing motif, e.g. the classic RNA recognition motifs (RRM), helicase folds (DEXDc, HELICc), and K homology (KH); and the newly described domains, WD40 fold [116; 117; 15], AAA ATPase, tetratrico peptide repeat region (TPR) [118], Ski complex [119], translation terminator Nro1 [120], and the CH domain [16].

- **Poly-A-independent enrichment:** PTex specifically purified 72/80 ribosomal proteins (42/47 of the large and 30/33 of the small subunits), and 19/20 tRNA synthetases, demonstrating the poly-A-independent purification nature of PTex.

- **Comparison with the mRNA-interactome of HEK293:** although the mRNA-interactome capture from HEK293 was determined using *in vivo* RNA-labelling with 4-thiouridine (4-SU) and 365 nm UV-light for

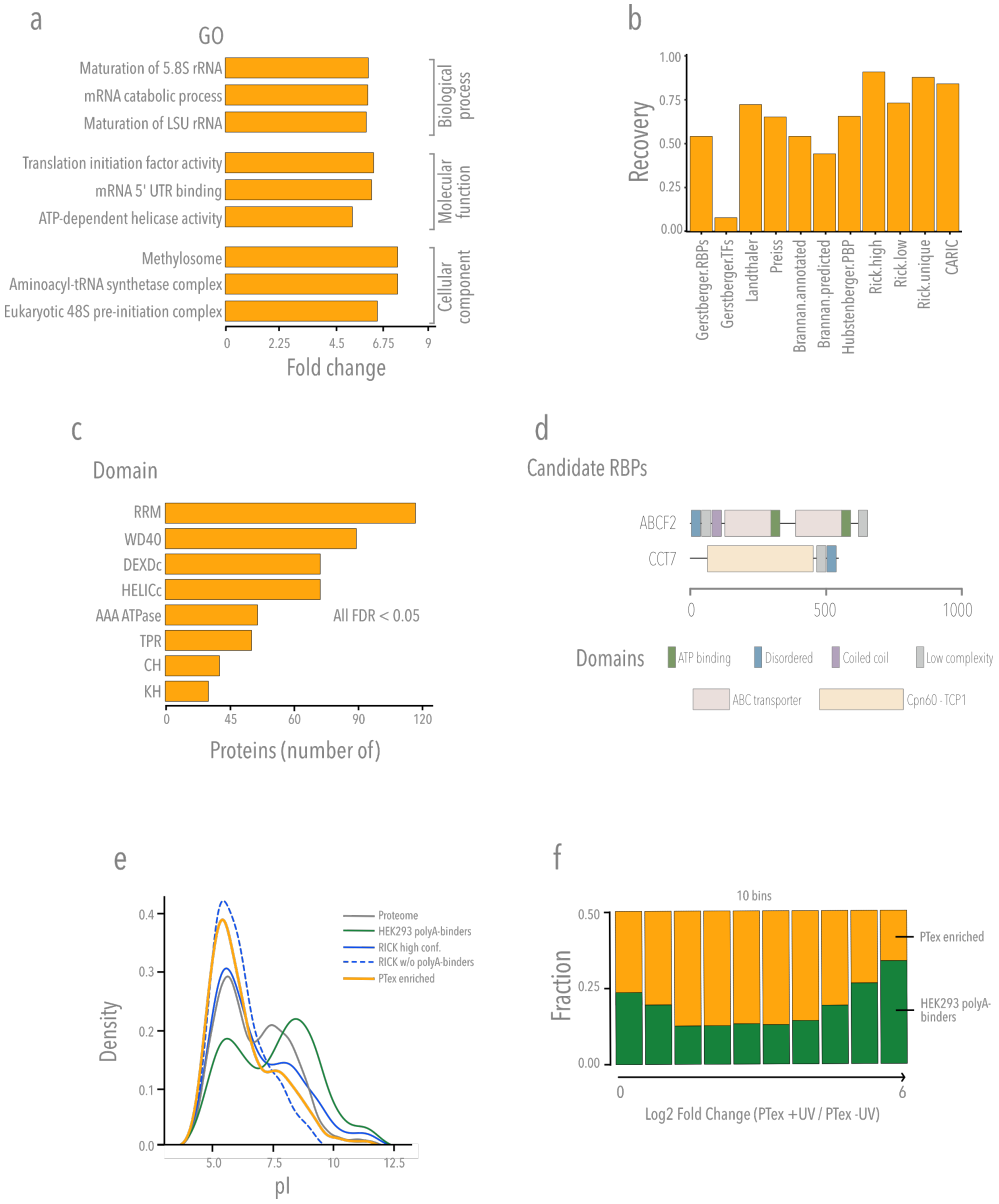

Figure 3.12: Features of RNA-interacting proteins found by PTex (a) Top 3 enriched GO terms (CC, MF, BP). **(b)** PTex-purified proteins overlap with well-described RBPs but not with transcription factors. Recovery of Gerstberger RBPs and transcripiton factors (TFs) reviewed in [50], a recent review on RBPs by [12], HEK293 poly-A binders [14], RBPs found by RNA interactome using click chemistry (RICK[56]; CARIC[57]), P-body components [114] and a recent prediction of candidate RBPs (SONAR, [115]). **(c)** Enriched protein domains in PTex-purified proteins from HEK293 cells. **(d)** Selected candidate RBPs exhibit intrinsically disordered regions and coiled-coils, ATP-binding domains, and the low complexity characteristic in many novel RBDs [16; 26; 29; 30]. **(e)** mRNA-binding proteins display a bimodal isoelectric point (pI) distribution pattern with peaks at pH 5.5 and 9.5 [14; 15]. RNA-interactors in general peak at pI 5-6 as found by PTex and RICK [56]. **(f)** Distribution of previously identified HEK293 mRNA-binding proteins (green; [14]) in PTex; each bin represents 10% of the 3037 PTex proteins from lowest to highest enrichment. Modified from [61].

cross-linking [14], PTex recovered 70% of the poly-A RNA-binding proteins previously identified in [14] (Fig. 3.12c).

- **Overlap with other poly-A independent methods [56; 57]:** PTex largely overlaps with RBPs recently found in HeLa cells by two unbiased techniques. PTex matches 94% of the high confidence RBPs and 86% of the non-poly-A RBPs identified by RICK [56] and 84% of the complexes found in CARIC [57] (Fig. 3.12c).

- **DNA-binding proteins:** proteins involved in replication and response to DNA damage (e.g. DDX54 [121]) were found enriched by PTex, however, transcription factors were underrepresented (Fig. 3.12c), demonstrating that PTex does not select for DNA-specific binding proteins. Importantly, the boundaries between RNA- and DNA-binding are blurring since nuclear DNA-binders have been found to interact with RNA [31].

- **Isoelectric point:** proteins enriched by the mRNA-interactome capture exhibit an overal higher isoelectric point (pI) [26; 16]. On the contrary, proteins identified by poly-A-independent techniques (RICK [56] or PTex [61]) showed a pI <6 (Fig. 3.12d). At cellular pH, proteins with a low pI have an negative net charge; therefore, unspecific interactions with negatively charged RNA are unlikely to occur due to the electrostatic repulsion.

- **Comparative distribution of mRNA-binding proteins:** the aim of this test was to rule out that PTex-specific purified proteins are more efficiently recovered than the established mRNA-binding proteins (indicative of carry-over of proteins unrelated to RNA interactions). A comparison of the distribution of the HEK293 mRNA-binding proteins [14] in the PTex enrichment shows that both groups of RBPs are similarly enriched along the dynamic range of PTex (from no enrichment to fold change log2= 6, Fig. 3.12e).

The results above described demonstrate that PTex specifically purifies ribonucleoprotein complexes.

3.2 Applications

3.2.1 pCLIP

As described in the introduction, the CLIP methods are the current gold standard for analysing the transcriptome associated to a particular protein (reviewed in [47; 48; 46]). However, removal of the unbound RNA remains a challenging aspect of the method and probably the most time-consuming. As PTex can efficiently deplete free-RNA and unbound proteins, a modified PAR-CLIP protocol was designed (Fig. 3.13).

HuR (ELAVL1) was selected as proof-of-concept due to its well known interactions with mRNA and pre-mRNA in several CLIP studies and its well-documented binding motif (5'- UUUUUU -3') [106]. HEK293 cells stably expressing a FLAG-tag copy of HuR were *in vivo* labelled with 4-thiouridine (4SU), followed by UV irradiation at 365 nm.

Three variants of the CLIP method were performed i) classical PAR-CLIP (PAR-CLIP-classic) [32; 102], ii) PAR-CLIP using ligation of adapters directly on the beads (PAR-CLIP-on-beads) [103; 104], and iii) a PTex derived CLIP version, pCLIP [61], in which the second and third steps of PTex were adapted to remove the unbound RNA instead of PAGE/membrane band excision (Fig. 3.13).

The libraries obtained by pCLIP contained a larger fraction of longer reads than the PAR-CLIP classic/PAR-CLIP-on-beads libraries (Fig. 3.14a). The three approaches were able to identify the canonical 5'- UUUUUU -3' motif and similar profiles of HuR-bound RNA clusters map to intronic and 3'UTR regions [106] (Fig. 3.14b-d).

The clusters were successfully mapped to the same 3'UTR loci when comparing HuR binding sites in tubulin and in splicing factor Srsf6 mRNAs (Fig. 3.14e,f). Regardless the low-read-coverage, these results demonstrate that PTex can be integrated into more complex workflows such as (PAR-)CLIP and that PTex has the potential to simplify CLIP-type approaches by enriching for clRNPs or by removing unbound RNA transcripts [61].

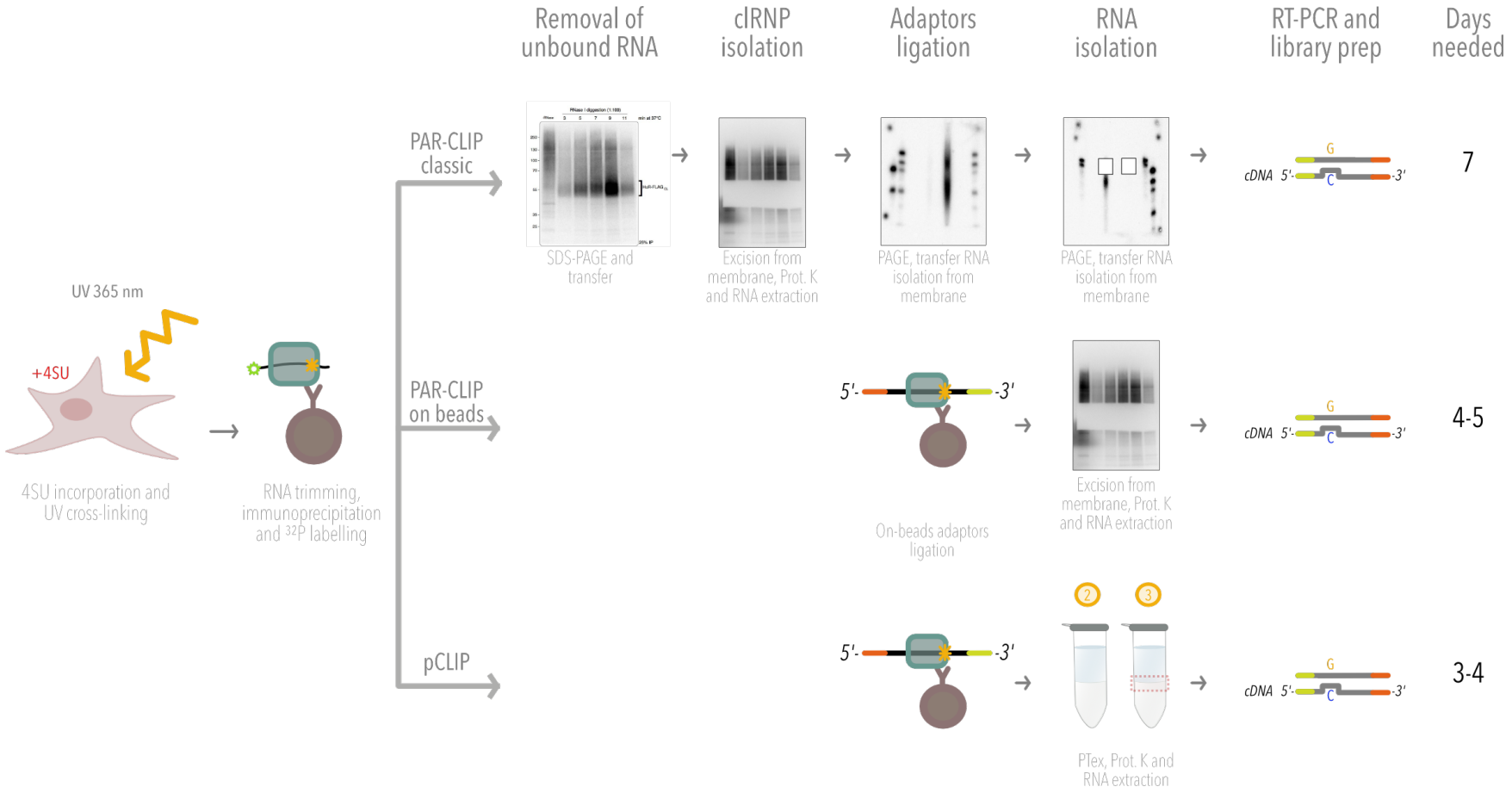

Figure 3.13: Schematic of the incorporation of PTex into the PAR-CLIP protocol classical PAR-CLIP (PAR-CLIP-classic) [32; 102], PAR-CLIP using ligation of adapters directly on the beads (PAR-CLIP-on-beads)[103; 104], were used to compare the performarce of pCLIP, version in which the second and third steps of PTex were adapted to remove the unbound RNA as alternative to PAGE/membrane band excision.

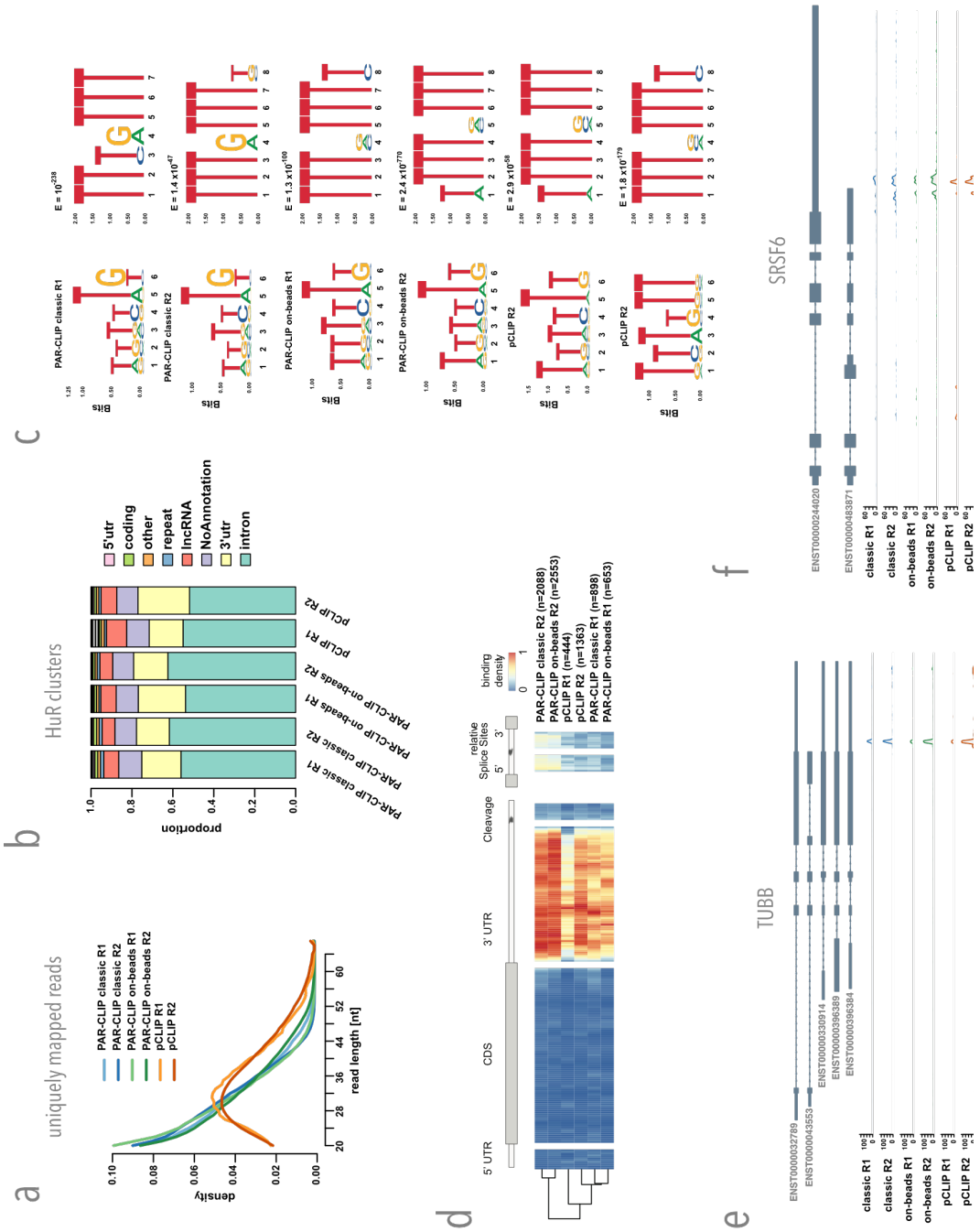
a
uniquely mapped reads
PAR-CLIP classic R1
PAR-CLIP classic R2
PAR-CLIP on-beads R1
PAR-CLIP on-beads R2
pCLIP R1
pCLIP R2
density
read length [nt]
b
HuR clusters
5'utr
coding
other
repeat
lncRNA
NoAnnotation
3'utr
intron
proportion
c
PAR-CLIP classic R1
PAR-CLIP classic R2
PAR-CLIP on-beads R1
PAR-CLIP on-beads R2
pCLIP R2
Bits
d
5' UTR
CDS
3' UTR
Cleavage
relative Splice Sites
binding density
PAR-CLIP classic R2 (n=2088)
PAR-CLIP on-beads R2 (n=2553)
pCLIP R1 (n=444)
pCLIP R2 (n=1363)
PAR-CLIP classic R1 (n=898)
PAR-CLIP on-beads R1 (n=653)
e
TUBB
ENST00000032789
ENST00000043553
ENST00000330914
ENST00000396389
ENST00000396384
classic R1
classic R2
on-beads R1
on-beads R2
pCLIP R1
pCLIP R2
f
SRSF6
ENST00000244020
ENST00000483871
classic R1
classic R2
on-beads R1
on-beads R2
pCLIP R1
pCLIP R2

Figure 3.14: pCLIP: a fast PAR-CLIP variant employing phenolic extraction. (a) Read length distribution of uniquely mapping reads utilised for determination of binding sites (cluster) of HuR (ELAVL1). PAR-CLIP samples were processed using PARpipe (see methods in [61]). **(b)** Relative proportion of PARalyzer-derived cluster annotation. **(c)** Heatmap of relative positional binding preference for intron-containing mRNA transcripts for each of the six HuR PAR-CLIP samples. Sample-specific binding preferences were averaged across selected transcripts (see methods in [61]). The relative spatial proportion of 5'UTR, coding regions and 3'UTR were averaged across all selected transcript isoforms. For TES (regions beyond transcription end site), 5' splice site, and 3' splice site, fixed windows were chosen (250 nt for TES and 500 nt for splice sites). For each RBP, meta-coverage was scaled between 5'UTR to TES. The 5' and 3' intronic splice site coverage was scaled separately from other regions but relative to each other. **(d)** *De novo* motif discovery for PARalyzer derived clusters using ZAGROS (left) and DREME (right). For Zagros, a T-rich motif was found. As ZAGROS does not return E-values, cluster sequences were analysed using DREME. For all but classic PAR-CLIP R2 a T-rich motif scoring the highest was found. For classic PAR-CLIP R2 however, the T-rich motif scored second with a similar E-value to a less frequent primary motif (Supplementary Fig. 16 in [61]). **(e,f)** Genome browser shots of TUBB and SRSF6 example genes showing reproducible 3'UTR binding sites. Track y-axes represent uniquely mapping read count. Modified from [61].

3.2.2 Purification of RBPs from challenging samples

Tissues in general are challenging samples to process, brain tissue particularly so due to its high lipidic content. During the organic extractions, the hydrophobic character of lipids promoted its accumulation at the interphase during the first step of the PTex protocol, however, the higher load in lipids contained in the brain sample saturated the system (data not shown). As solubilisation can be enhanced by temperature, the PTex extractions were performed at 65°C (Hot-PTex, Fig. 3.15a).

To analyse the efficiency of the extraction, a Western blot against HuR was performed (Fig. 3.15b). The signal at the pockets/smear, and especially the restoration of the HuR signal after RNase digestion demonstrate that HuR-RNA complexes are largely enriched after PTex in a UV-dependant manner (Fig. 3.15b). An important contribution of Hot-PTex is that it allows the unbiased purification of RNPs from animal tissues without the requirement for RNA labeling, as it is the case for RICK[56] or CARIC[57] approaches.

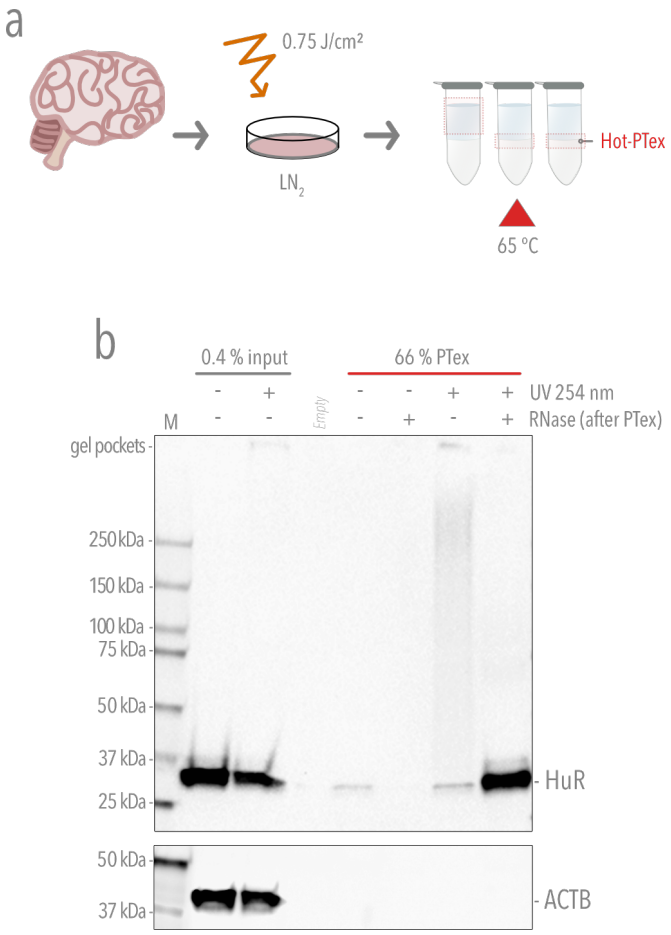

Figure 3.15: PTex purifies clRNPs from mouse-brain tissue. (a) Mouse brain tissue (obtained from Prof. F. Heyd Lab) was deep-freeze with liquid N_2 and pulverised with a mortar (pestle). Half of the powder (∼130 mg) was irradiated with 0.75 J/cm^2. Cross-linked and non-cross-linked tissue powders were resuspended in 2.4 mL of DPBS. PTex samples were prepared using 600 µL of the suspension (∼35 mg/sample; two samples per condition). **(b)** PTex pellets (-UV/+UV) were resuspended in 20 µL RNase buffer, adding 0.2 µg of RNaseA in the indicated samples. After incubating 1 h at 37°C, reaction was stopped by adding 10 µL Lämmli buffer 4x, 95°C 5 min. Samples were then electrophoresed on SDS-PAGE 4-20%, transferred onto nitrocellulose membrane 0.2 µm and incubated with anti-human HuR (1:5000) or anti-human ACTB (1:2500) overnight at 4°C. Antibody binding was developed with the corresponding secondary antibody HRP and Clarity ECL substrate for Western blot (BioRad) or anti-mouse Alexa647 (Invitrogen). Modified from [61].

3.2.3 Insights into the first RBPome of *Salmonella* Typhimurium

Unlike in eukaryotes, polyadenylation of RNA in prokaryotic cells is a rare event [80], therefore capturing mRNA-protein complexes cannot be performed via oligo-dT-based methods. The unbiased nature of PTex however, allowed the purification of RNA-crosslinked complexes from the pathogen *Salmonella* Typhimurium (Fig. 3.16).

Salmonella strains harbouring the chromosomally FLAG-tagged proteins Hfq [85] and GpmB were used to test PTex in bacteria. Hfq was selected as a positive control since it is an abundant and well known RNA-binder [85; 122]. On the other hand, GpmB is a putative phosphoglycerate mutase which was identified as a non-RNA binder by a supported vector machine algorithm for the de novo prediction of RNA-binding motifs (TripepSVM) [30], and therefore, used as a non-RNA binder control together with antibodies to detect the chaperone GroEl, also a non-RNA binder protein.

Application of the modified PTex protocol, Hot-PTex, yielded purified Hfq-RNA complexes from *Salmonella* Thyphimurium Hfq-FLAG in an UV-dependent manner (Fig. 3.16a). One of the major characteristics of Hfq is that it is physiologically active as a homo-hexamer; a complex that has been shown to resist highly denaturing conditions, e.g. during SDS-PAGE [85; 61] (Fig. 3.16b,c). Hfq shifted bands are visible after UV radiation with 0.5 J/cm^2 (Fig. 3.16b,c). Additionally, a band corresponding to the monomer with a slightly higher molecular mass is attributable to almost-complete RNase digestion of the bound-RNA (Fig. 3.16c). The physiologically relevant Hfq hexamer was detected in the Hot-PTex samples from cells radiated with 5 J/cm^2, indicating that the hexamer is still bound to remaining RNA fragments [61] (Fig. 3.16c).

RNA-protein interactions with high affinity and/or stability can resist the denaturing conditions of the purification steps even in the absence of UV-cross-linking, leading to an increased signal of the monomeric form of the RBP in Western blots even in non-UV-radiated samples (Fig. 3.7a); signal which is otherwise reduced/absent when applying RNase treatment before PTex (Fig. 3.7b). Fig. 3.16 shows an increasing signal from the monomeric form of Hfq

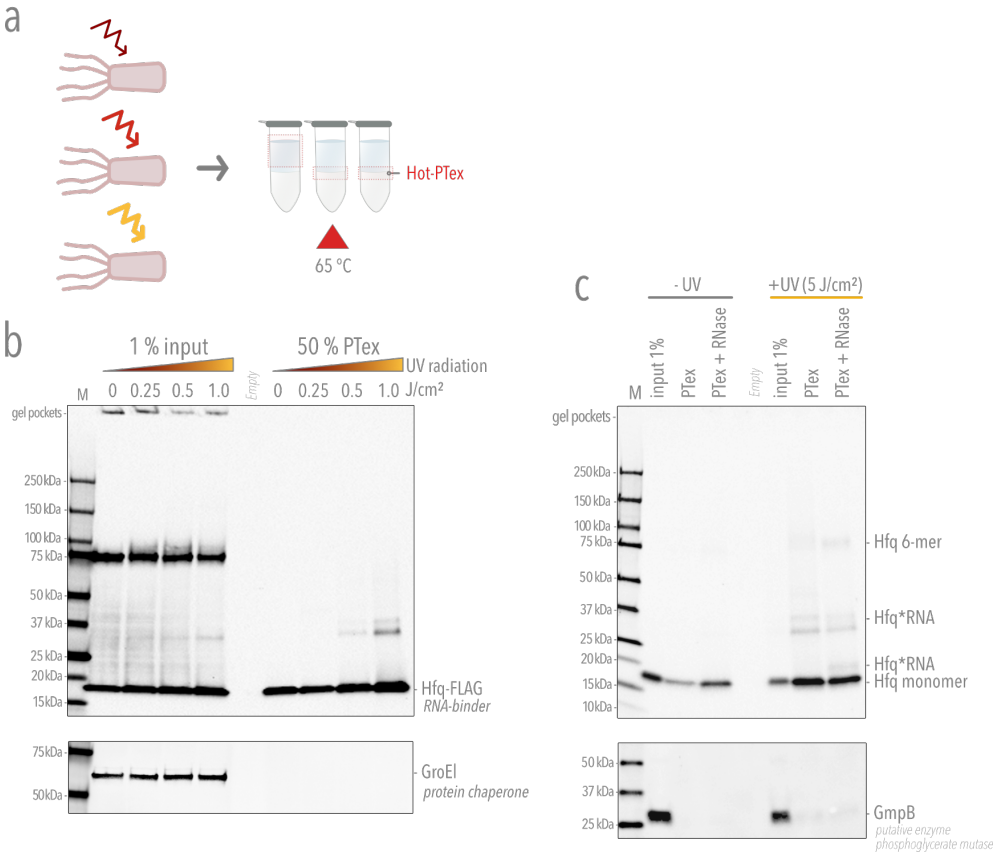

Figure 3.16: Hot-PTex allows the purification of Hfq-RNA complexes from *Salmonella* Typhimurium (a) *Salmonella* Typhimurium SL1344 Hfq-FLAG was UV-cross-linked with 0.25, 0.5, 1.0 or 5 J/cm^2 of UV-light at 254 nm. Hot-PTex was performed to purify bacterial RNPs. **(b)** Western blot using an anti-FLAG antibody demonstrates recovery of Hfq monomers linked to RNA after radiation with 0.5 J/cm^2. No bands are detected when incubating the membrane with antibodies against the chaperone GroEl. **(c)** Higher UV dosages increases the cross-linking events, note that the physiologically active Hfq hexamer partially withstands SDS-PAGE conditions [85] and that this complex is also enriched after PTex (Hfq 6-mer). Similar to GroEl, the protein GpmB is not enriched after Hot-PTex.

after Hot-PTex correlated with the increasing doses of UV-cross-linking applied. However, Western blots detecting the proteins GroEl and GpmB show the distinctive depletion in non-RNA binders (Fig. 3.16), demonstrative of the RNA-dependency of the Hot-PTex purification.

In a different experiment, Hot-PTex samples from *Salmonella* Typhimurium SL1344 (wild-type) cross-linked with the vari-x-link for 90 seconds ($\sim$1.5 J/cm^2, section 2.10) were used to map the RNA-associated proteins by mass spectrometry (Fig. 3.17a). In this case, intensity-based absolute quantification (iBAQ) intensities were used to compare recovered proteins from UV-irradiated versus non-radiated cells (biological duplicates).

Here, the overall RBP enrichment was lower in *Salmonella* compared to enrichment achieved with PTex from HEK293 cells, even for known bacterial RBPs as Hfq. Nonetheless, 172 proteins were identified as RBPs (Supplementary Table S3 in [61]), 59 of which are known RNA-associated proteins: 33 ribosomal proteins, components of the RNA polymerase complex (subunit α, σ factor RpoD, DksA) and 4 out of the 5 established mRNA-binding proteins of *Salmonella* (Hfq, ProQ, CspC/CspE) [61; 85; 84; 123] (Fig. 3.17b).

Additionally, proteins with known RNA-binding domains (RBDs) were found, e.g. the oligonucleotide/oligosaccharide-binding fold (OB-fold) present in RpsA, RpsL, RplB, CspC, CspE, Pnp, RNaseE, Ssb, NusA; and domains which were also detected in RBPs when screening eukaryotic cells: the afforementioned AAA ATPase fold in the ATP-dependent protease ATPase subunits HslU, ClpB and ClpX, or thioredoxin domains as in AhpC, Thiol:disulfide interchange protein (DsbA) and Bacterioferritin comigratory protein (Bcp) [61; 15; 16]. Similar to previous reports [16; 124], the glycolytic enzymes Pgk and Pgi were also found to be associated with RNA (Fig. 3.17b).

Gene ontology terms analysis of the identified proteins (Fig. 3.17c) shows that the most significant terms are those associated with RNA biology, such as "translation" and "ribosome". Interestingly, although RNA has been thought not to localise in the periplasmic space, the term "cell outer membrane" was also significantly represented as many membrane-associated proteins were enriched (refer to Table S3 in [61]).

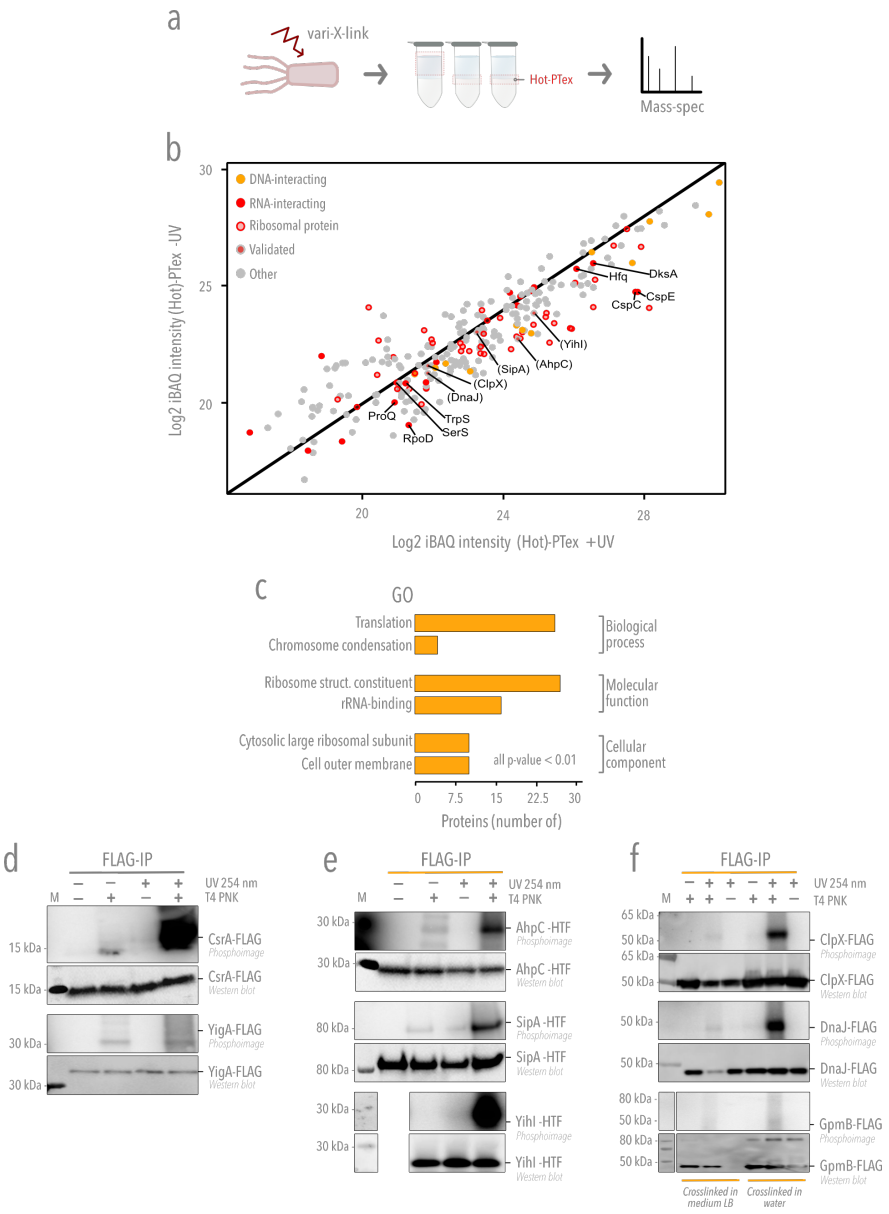

Figure 3.17: The first RNA-bound proteome from *Salmonella* Typhimurium (a) *Salmonella* Typhimurium SL1344 Hfq-FLAG was UV-cross-linked in a vari-x-link for 90 seconds (1.5 J/cm^2). Hot-PTex was performed to purify bacterial RNPs which were later analysed by protein mass spectrometry. **(b)** 172 proteins are enriched after UV-cross-linking (PTex +UV). Among the identified proteins are ribosomal proteins (transparent red), known RBPs (red) and DNA-binders (orange), and proteins not known to associate with RNA before were selected for validation (in parentheses). **(c)** RNA-associated GO terms are significantly enriched among the identified proteins. **(d-f)** Validation of RBP candidates: *Salmonella* strains expressing FLAG- or HTF-tagged proteins were cross-linked (or not) and immunoprecipitated. RNA-association is confirmed by radioactive labeling of RNA 5' ends by polynucleotide kinase (T4 PNK) using phosphoimaging; a signal is exclusively detectable after UV-cross-linking and radiolabeling of precipitated RNA. CsrA-FLAG (pos. ctr.), YigA-FLAG (neg. ctr.), AhpC-HTF, SipA-HTF, YihI-FLAG [61], ClpX and DnaJ [30] bind RNA *in vivo*. Under the conditions tested, GpmB-FLAG did not show RNA-binding capacity. Modified from [61; 30].

Outer membrane proteins were also identified as RNA-binders in *Escherichia coli* by a similar method [60]. Here again, distinguishing between RBPs with canonical functions on RNA and proteins which are associated to RNA for e.g. structural reasons is not possible ([61]). However, secreted outer membrane vesicles (OMVs) from Gram-negative bacteria have been shown to contain RNA, yet the sorting mechanism has not been discovered (reviewed in [125]). It is reasonable to speculate that, if such a sorting pathway exists, the proteins involved could get cross-linked and therefore purified by PTex.

From the 172 proteins identified in this experiment, 113 had not yet been reported to interact with RNA. Thus, in order to validate these indeed as new RNA-binding proteins , a set of proteins were selected, fused with C-terminus tags (FLAG or His-TEV-FLAG), UV-crosslinked, immunoprecipitated, and the bound-RNA radioactively labelled with γ^{32}P-ATP (T4 polynucleotide kinase assay "PNK" as described in [126], section 2.19).

The selected proteins were: YihI, a putative GTPase-activating protein which was speculated to play a role in ribosome biogenesis [127], the alkyl hydroperoxide reductase c22 protein (AhpC) [128], the cell invasion protein SipA [129; 130]; and the proteins also identified as RNA-binders in [30], the subunit ClpX from the Clp protease (involve in regulating expression of the flagellum), the hyperosmotic and heat shock responsive chaperone DnaJ [30], and the predicted non-RNA-binder GpmB, mentioned before. YigA and CsrA-FLAG strains were used as negative and positive controls, respectively [30]. As expected, all of the selected proteins except for GpmB and YigA exhibited RNA-binding activities *in vivo* (Fig. 3.17d-f) thus corroborating PTex as a protocol for the identification of new RBPs.

A recurrent mention in the literature is the removal of the culture medium before UV cross-linking [14; 26], as excess of medium causes a lower cross-linking efficiency due to the co-absorption of UV-light (Dr. Benedikt Beckmann, personal communication). However, the established protocol for UV-cross-linking of bacterial cells as a preparation for PNK described in [85], indicates that the UV radiation is performed by directly radiating the cell cultures (LB medium).

An experiment comparing two cross-linking strategies, UV radiating cultures *vs.* pellets, showed that, at least for the RBPs ClpX and DnaJ (Fig. 3.17f) and UbiG and CsrA (Fig. 5 in [30]), cross-linking bacterial cells directly in the LB medium was inefficient and produced near-to-negative results; whereas cross-linking pellets resulted in clear positive signals. Furthermore, a follow up experiment where cells were radiated directly on a translucent defined culture medium (LPM plus), the characteristic UV-dependant shifted signal of Hfq was detected in Western blot (Fig. 3.16b, methods 2.3, 2.1).

Summarising, whether the low enrichment of Hot-PTex is a consequence of an inefficient cross-linking or rather of an excessive background, the results presented here constitute the first RNA-bound proteome (RBPome) ever described for *Salmonella* Typhimurium. Although additional modifications in UV-cross-linking and/or cell lysis can improve the sensitivity, PTex is a suitable unbiased tool for cell-wide RBP purification in bacteria.

Chapter 4

Discussion

4.1 Development of the PTex approach

RNA-protein UV cross-linking is a inefficient process: as shown for the RBP hnRNPL, only 1 out of 1000 molecules can be cross-linked to RNA *in vivo* (Fig. 3.4c, [62]). Low cross-linking efficiency has also been reported in other human and yeast cell cultures where even higher UV dosages resulted in only ∼1-10% of cross-linking events [71; 15; 42; 16; 124; 97]. Likewise, not all of the (m)RNA molecules are bound to a given RBP, therefore, the excess of free-RNA can saturate the oligo-dT beads in mRNA interactome capture approaches, affecting its clRNPs capturing efficiency and elevating the noise-to-signal ratio when analysing the bound-transcriptome. Consequently, any method aimed to purify RNPs must overcome the challenge of removing 99.9% of the non cross-linked background protein.

Exploiting the phase-separation principle [64; 65; 66], PTex overcome those obstacles by i) introducing the mixture phenol-toluol in the step 1, which allows the separation of RNA, proteins, and clRNPs from other cellular molecules, and ii) due to the physicochemical differences between proteins and RNA. In step 2, by applying denaturing and chaotropic conditions, PTex can remove free RNA and free proteins from clRNPs in an unbiased manner. PTex is also an easy to perform modular method which can be completed in less than 3 hours.

Importantly, it has been demonstrated that PTex can be applied in samples

Method	Starting material (cells)	Study
PAR-CLIP	2-9x10^8	[32]
RIC	2.85x10^8	[15]
RBR-ID	N/A	[131]
RICK	2x10^7	[56]
CARIC	3.6x10^8	[57]
XRNAX	8-10x10^7	[59]
OOPS	3.2-8.8x10^6	[60]
PTex	2-8x10^6	this study [61; 62]

Table 4.1: Starting material required by the current RNPs purification methods. Modified from [61].

UV-radiated with either 254 or 365 nm, native, or *in vivo* RNA labelling and from a variety of sources: cell culture, brain tissue, bacterial cells (this work, [61], and [62]). Additionally, PTex recovers complexes with a cross-linked RNA as short as 30 nt without requiring any particular sequence specificity, allowing to investigate RNA-protein interactions involving a wider diversity of RNAs than what has been possible before.

Similar to PTex, the methods RBR-ID [131], RICK [56] and CARIC [57] also offer unbiased purification of ribonucleoprotein complexes [45]. However, it is important to underline that all the aforementioned methods (RBR-ID, RICK and CARIC) rely on *in vivo* incorporation of 5-ethynyl uridine (EU) into the RNA, followed by click-chemistry and affinity capture via streptavidine beads. While pulse/chase experiments allow to determine newly-transcribed RNA species by those methods, incorporation of nucleotide analogs into cells can be challenging [61].

Although PTex can be performed on cells labelled with 4-SU, UV irradiation at 254 nm wavelength is sufficient to interrogate the RBPome of a given cell line, advantageous when analysing biological material in which uptake of nucleotide analogs is either insufficient or cost-intensive [61]. Another advantage of PTex is that it requires lower amounts of input material in comparison with other methods (Table 4.1). However, since UV radiation is used in all the available methods to induce RNA-protein cross-links, the possible bias caused by the different cross-linking efficiencies for individual proteins remains a general issue among all methods mentioned [61].

4.1.1 Unbiased purification of UV cross-linked RNA-protein complexes by liquid-liquid phase separation

At the time of executing this project, two similar approaches to purify clRNPs employing phenolic extractions were also being developed in the labs of Kathryn S. Lilley (University of Cambridge, UK, [60]) and Jeroen Krijgsveld (Heidelberg University, DE, [59]). Conceptual similarities and technical differences of these methods are discussed in [45] and summarised in Fig. 4.1.

PTex [61], OOPS [60] and XRNAX [59] have as starting point the formation of covalent cross-links between interacting proteins and RNAs by radiating cells *in vivo* with UV light. Additionally, the three methods rely on the physicochemical properties of the complexes and its distinctive mobility across phases during organic phenolic extractions to achieve the purification of clRNPs.

XRNAX and OOPS consist on 3-4 rounds of acidic guanidinium-thiocyanate phenol-chloroform (AGPC) extractions, clRNPs accumulate in the interphase along with DNA due to the low pH of the extraction environment. Interphases are collected and treated with DNAse or RNase. UV cross-linked RNA-protein complexes are then purified either by affinity capture on silica columns (XRNAX), or by subsequent extraction with AGPC (OOPS); and subjected to mass spectrometry (MS).

PTex instead, applies first a Phenol-Toluol extraction to separate DNA and lipids from RNA, proteins, and clRNPs, which accumulate in the upper aqueous phase; the aqueous phase is then extracted with AGPB (acidic guanidinium-thiocyanate-phenol-BCP) in order to enrich clRNPs in the interphase away from free proteins and RNAs which migrate to the organic and aqueous phases, respectively.

As those three methods do not depend neither on the existance of a polyA tail of the target RNA nor on the deployment of modified nucleotides, all three can therefore be applied to any cross-linked material for the identification of proteins that interact with all classes of RNAs [45].

Figure 4.1: Comparison of phase-separation methods for the unbiased purification of clRNPs. Three methods converged in the application of liquid-liquid phase separation by phenolic extractions: XRNAX [59], OOPS [60] and PTex [61]. Following *in vivo* UV radiation XRNAX and OOPS consist of several rounds of acidic guanidinium-thiocyanate-phenol-chloroform (AGPC) extractions. Interphases are collected and subjected to DNAse or RNase, and clRNPs are then partially digested and captured on silica (XRNAX), or are re-extracted by AGPC (OOPS) and subjected to mass spectrometry (MS). Distinctively, PTex achieves the separation of DNA and lipids from RNA, proteins, and clRNPs by applying first an extraction with phenol-toluol. clRNPs are purified via a second extraction with AGPB (acidic guanidinium-thiocyanate-phenol-BCP), in which clRNPs partition to the interphase while free proteins and RNA migrate to the organic and aqueous phases, respectively. A third extraction helps to remove further the excess of RNA and proteins as well as the guanidine thiocyanate. Modified from [132]

4.1.2 Limitations of PTex

In the following, the limitations of PTex will be discussed from a technical and conceptual point of view:

- Enrichment vs. recovery: PTex recovers 25-30% of the initially cross-linked RNPs, therefore, considerations have to be taken when working with scarce material.

- Efficient cell lysis: as the migration of the clRNPs during the first step of PTex is sensitive to high concentrations of salts and detergents [62], cell lysis is achieved by dissolving the lipidic membrane of the cells with phenol and toluol. Therefore, lysis can be impaired when working with microorganisms with a rigid, non-lipidic cell wall, e.g. fungi (including yeast), and Gram(+) bacteria.

- Lower RNA-bound limit of 30 nt: PTex might not be suitable for investigating complexes containing RNA with less than 30 nt in length as in the case of mature miRNA [133; 134]. A systematic comparison of the purification efficiency of PTex with different protein masses and RNA lengths will be required.

- Starting material and saturation: For the current protocol a starting amount of 2-8x10^6 (HEK293) cells per tube is recommended, as a higher input resulted in an impaired purification due to the saturation of the phases.

- Phase removal and contamination: Phase handling and avoidable contamination between phases mainly depend on the skills of the experimenter. Contamination of phases was more likely to happened when removing individual phases using a pipet tip (1000 or 200 µL). In contrast, a 1 mL syringe with blunt needle, so far, has rendered the lowest contamination. However, carry-over of a small percentage of unbound -RNA and proteins in the interphase is inevitable since the interphase is indeed the converging volume between the aqueous and organic phases.

- RNA-binding *vs.* RNA-associated: It has been previously discussed that proteins cross-linked to RNA are not automatically RNA-binding proteins in the classical sense [97]. RNA-binding protein (RBP) is a term historically used to define proteins with a direct function in RNA biology, e.g. RNases, helicases, etc. Since UV-induces covalent bonds of RNA and protein physically close to each other ("zero distance"), independently of the potential role of the protein, PTex-purified proteins are referred as "RNA associated" in order to avoid over-interpretation.

These limitations have also been discussed and summarized in our recent work [62].

4.1.3 Modularity of PTex allows for its improvement

The simplicity and modular architecture of PTex gives space for its improvement:

- Inefficient protein precipitation can affect the overall yield of PTex. Protein precipitation can be enhanced by adding a non-related protein into the precipitation tube, as a higher protein concentration facilitates the aggregation and sedimentation during the centrifuging step (Prof. Dr. Fátima Gebauer (CRG, Barcelona, ES), personal communication).

- The saturation effect mentioned above can be avoided by reducing the input material and/or increasing the area of the interphase by performing the extractions in tubes with a wider diameter.

- Impaired lysis of Gram(+) bacteria can be solved with the enzymatic digestion of the peptidoglycan layer by lysozime or lysostaphin treatment before applying PTex (Prof. Dr. Sander Granneman, personal communication).

- The introduction of RNase treatment prior to PTex can control for UV-independent complexes. Comparing the fold-change of the different proteins in -UV, +UV and +UV/+RNase PTex samples could lead to the identification and classification of RNA-protein complexes according to its binding affinity.

- Following the same idea, analysis of the mass spectrometry data could be improved by including a whole cell lysate (input) control.

4.2 Beyond the mRNA-bound proteome of HEK293 cells

As PTex is not restricted to poly-A RNA, the complete RNA-bound proteome of the HEK293 human cell was successfully determined [61]. PTex-purified samples from HEK293 cells were analysed and the protein fraction of the clRNPs determined by mass spectrometry. Following an analysis pipeline established before [97] and applying a rigorous selection criteria, it was possible to confidently identify 3037 RNA-associated proteins significantly enriched upon UV cross-linking [61].

The mRNA-interactome capture previously described for HEK293 [14] counts 790 well established RBPs [14]. PTex successfully covered more than 70% of them. From the >2000 proteins newly identified by PTex, 729 novel RBPs were also identified by OOPS, bringing the PTex coverage of OOPS to 83%. Ultimately, 1722 PTex-exclusive proteins are enriched in RNA-related molecular functions (Fig. 4.2), yet its biological relevance remains to be elucidated.

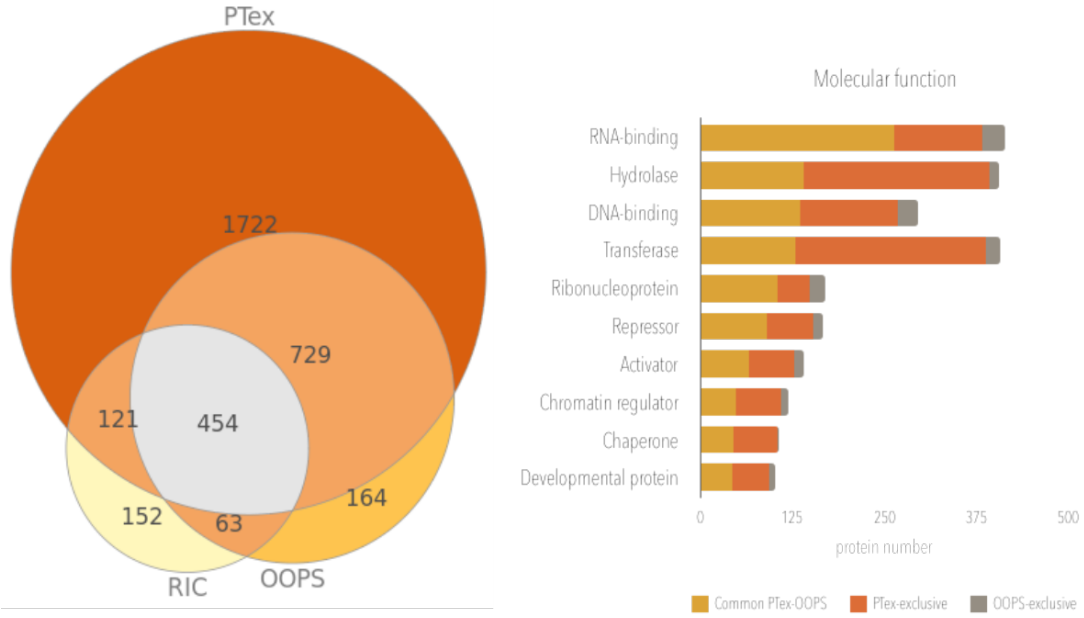

Figure 4.2: PTex covers more than 70% of the well-established and novel RBPs. (left) Venn diagram shows that PTex covers 73% of the mRNA-RBPs identified in [14] and shares >700 of the novel RBPs identified by a similar method [60]. **(right)** Proteins exclusively identified by PTex are enriched in RNA-related molecular functions. Venn diagram was created using the online tool http://www.venndiagrams.net; list of proteins classified according to their molecular function was retrieved from www.uniprot.org.

The slight differences in the number and identity of the identified proteins can be attributable to the culture and UV-radiation conditions (notice that RIC of HEK293 cells applied to 4SU-fed cells radiated with UV at 365 nm), physiological state of the cells, material loses during the procedure, among others.

The high number of PTex-purified RBPs rises a question: why have these proteins not been identified as RNA-interacting before? Until recently, the vast majority of RBP discoveries were restricted to a specific class of RNA: mRNA [12]; although heterogeneous in sequence, mRNA only represents ∼5% of the total RNA within a cell [61].

The RNA exosome from eukaryotes [135; 11] is a strong example to underline the differences between the proteins recovered by the interactome capture (poly-A RNPs) and those from PTex (RNA-interactors). The core exosome complex is built-up from ten protein subunits (Exo-10) and only one of them, the Rrp44, has a catalytic function over RNA (exo- and endoribonuclease activity, 4.3a). The remaining nine proteins (Exo-9) form a barrel structure

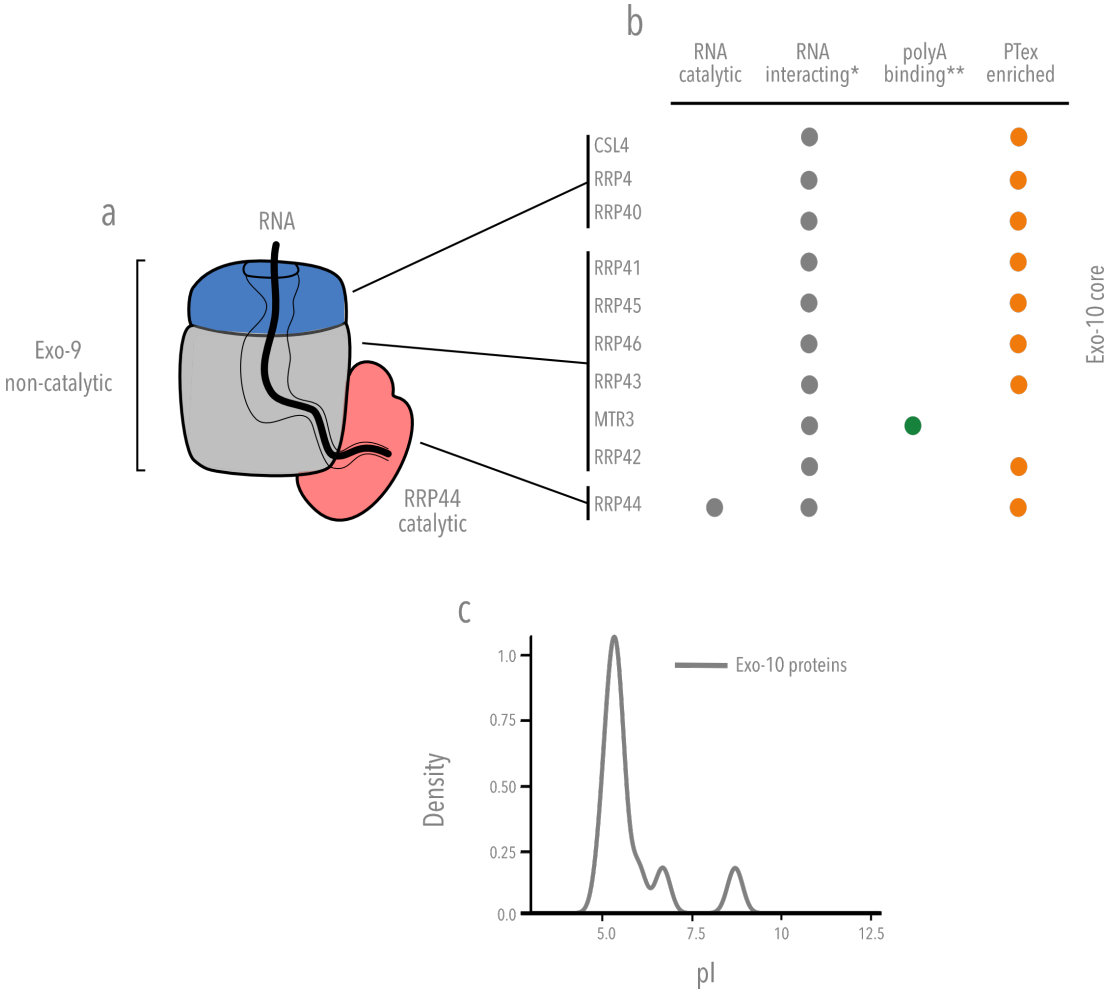

Figure 4.3: Proteins of the RNA exosome are prototype PTex proteins. (a) The RNA exosome core consists of an arrangement of nine structural proteins forming a barrel (Exo-9), and the catalytic protein Rrp44 (RNase, Exo-10). Modified from [11]. **(b)** *All exosome subunits are labeled "RNA-binding" (Uniprot.org); **green = identified via poly-A selection in [14]; orange = enriched in PTex. **(right)** Isoelectric points of the human exosome core proteins. Modified from [61].

through which the RNA is being channelled and ultimately degraded by Rrp44. Even though the Exo-9 proteins do not have RNA degrading or modifying activities [136; 11], they are in direct contact with the RNA molecule, as demonstrated by high resolution structure studies [137; 138; 139; 140].

PTex was able to enrich for 9 out of the 10 subunits (Fig. 4.3b) [61], while in the interactome capture of HEK293 cells, only Rrp44 was found [14] (Fig. 4.3b). Additionally, 7 of the Exo-10 protein have an isoelectric point below pH 6 (Fig. 4.3c). Interestingly, RNA of 30-33 nt or 9-10 nt length has been found inside the central exosome channel in *in vitro* and CRAC analyses [141]; as showed above, 30 nt is the lower length limit for efficient recovery by PTex (Fig. 3.8a).

Therefore, since all ten subunits are in direct contact with RNA, they are susceptible to UV-cross-linking.

The exosome example is also important to highlight that RNA-protein cross-linking is not a synonym of function. Whether due to structural arrangements, transient functional states, or stochastic interactions, if an RNA molecule is in direct contact with a protein in a given time, UV-cross-linking is possible. Henceforth, using the term "RNA-interacting" or "RNA-associated" to denote PTex purified proteins can help to disambiguate from "RNA binding proteins", a term historically used for proteins with known RNA-acting function.

4.2.1 PTex RNA-interacting proteins harbor conventional and newly identified RNA-binding domains

Proteins identified by PTex are rich in conventional (Fig. 1.3) and newly discovered RNA-binding domains (RBDs), the latter including domains recently validated *in vivo* in human and fly [29] (Fig. 1.4), such as the AAA ATPases or ATP-binding, as well as the prevalence of low-complexity, disordered and/or coliled-coil regions in -or near to- RNA-binding sites [29].

As shown in Fig. 3.12d, the PTex identified novel RNA-interacting protein ABCF2 contains AAA domains embedded in two ABC transporter domains (which was also classified as novel RBD by [29]), in addition to the coiled-coils, disordered and low complexity regions at its N- and C- ends. The protein CCT7 also exhibits disordered/low complexity regions at its C- end.

Additionally, RBPs newly validated *in vivo* by CAPRI [29] were also identified by PTex, e.g. LARP1, ILF3, HNRNPU, GTPBP4, among others listed in Table 4.2. Similar domain features were found by the other two methods employing phenolic extractions which investigated additional cell types and cellular compartments [60; 59].

Domain	Protein	Function
Protein kinase domain	SRPK1	Serine/arginine protein kinase specific for the SR (serine/arginine-rich domain) family of splicing factors.
Cyclophilin-type peptidyl-prolyl cis-trans isomerase	PPIA	Isomerisation of proline.
Protein phosphatase 2A, regulatory B subunit	PPP2R5E	Regulatory subunit of Protein phosphatase 2A (PP2A) is a major intracellular protein phosphatase.
Ars2	SRRT	Acts as a mediator between the cap-binding complex (CBC) and RNA-mediated gene silencing (RNAi). Associated with Alagille Syndrome 1.
Pinin/SDK/MemA protein	PNN	Transcriptional activator binding to the E-box 1 core sequence, mRNA surveillance and transport. Associated with Melanoma Metastasis and Premenstrual Tension.
Alpha crystallin/Heat shock protein	HSPB1	Act as chaperones and protect against heat shock.
Transketolase-like, pyrimidine-binding domain	TKT	Transketolase EC:2.2.1.1 (TK) catalyzes the reversible transfer of a two-carbon ketol unit from xylulose 5-phosphate to an aldose receptor. This enzyme, together with transaldolase, provides a link between the glycolytic and pentose-phosphate pathway.

Table 4.2: PTex and CAPRI share RBPs with unconventional conserved RNA-binding domains. List of some of the recently described conserved RBDs, identified by both CAPRI and PTex, and their biological importance. Protein names are given only for the human homologous. Modified from [29].

4.3 Applications of PTex

4.3.1 pCLIP

PTex is a versatile method with a modular structure which can be customised and incorporated into complex workflows. In the previous sections, it was demonstrated that PTex-purified complexes can be subjected to enzymatic digestion of RNA (Fig. 3.2, 3.5) or proteins (data not shown). This is particularly important for conducting high-throughput approaches as follow-up. PTex was proven helpful in removing the unbound RNA while enriching for clRNPs all in one. pCLIP reduced the time required to obtain cDNA libraries from 7 to 3 days when comparing to the classic PAR-CLIP protocol.

4.3.2 Purification of RBPs from challenging samples

Another challenging aspect of UV induced cross-linking is to penetrate denser material. The already low efficiency of UV irradiation is further decreased when radiating tissue samples or liquid cultures such as yeast [16; 124; 97]. We used PTex to directly purify cross-linked HuR from mouse brain samples [61]. This is of particular interest since *in vivo* RNA labeling of whole animals or some unicellular species has not been efficiently conducted. *in vivo* RNA labeling however is a prerequisite for other RNP purification techniques such as PAR-CLIP [32], RBR-ID [131], RICK [56], or CARIC [57].

4.3.3 The first RBPome of *Salmonella* Typhimurium

The lack of long poly-A streches in RNA of bacteria and archaea, a requirement for implementing the interactome capture techniques, has delayed the discovery of the RNA-bound proteome (RBPome) of these domains of life. PTex allowed for the first time to unbiasedly screen for proteins cross-linked to RNA in *Salmonella* Typhimurium [61] while RBPs in *E. coli* RBPs were purified by OOPS [60].

A comparison of the proteins enriched by the two methods resulted in a core of 109 RNA-associated proteins (Fig. 4.4a); among the expected ribosomal proteins, the here validated RBPs AhpC, ClpX, DnaJ and YihI are also conserved between *S.* Typhimurium and *E. coli*. The remaining novel RBP SipA is a pathogenicity island 1 effector protein

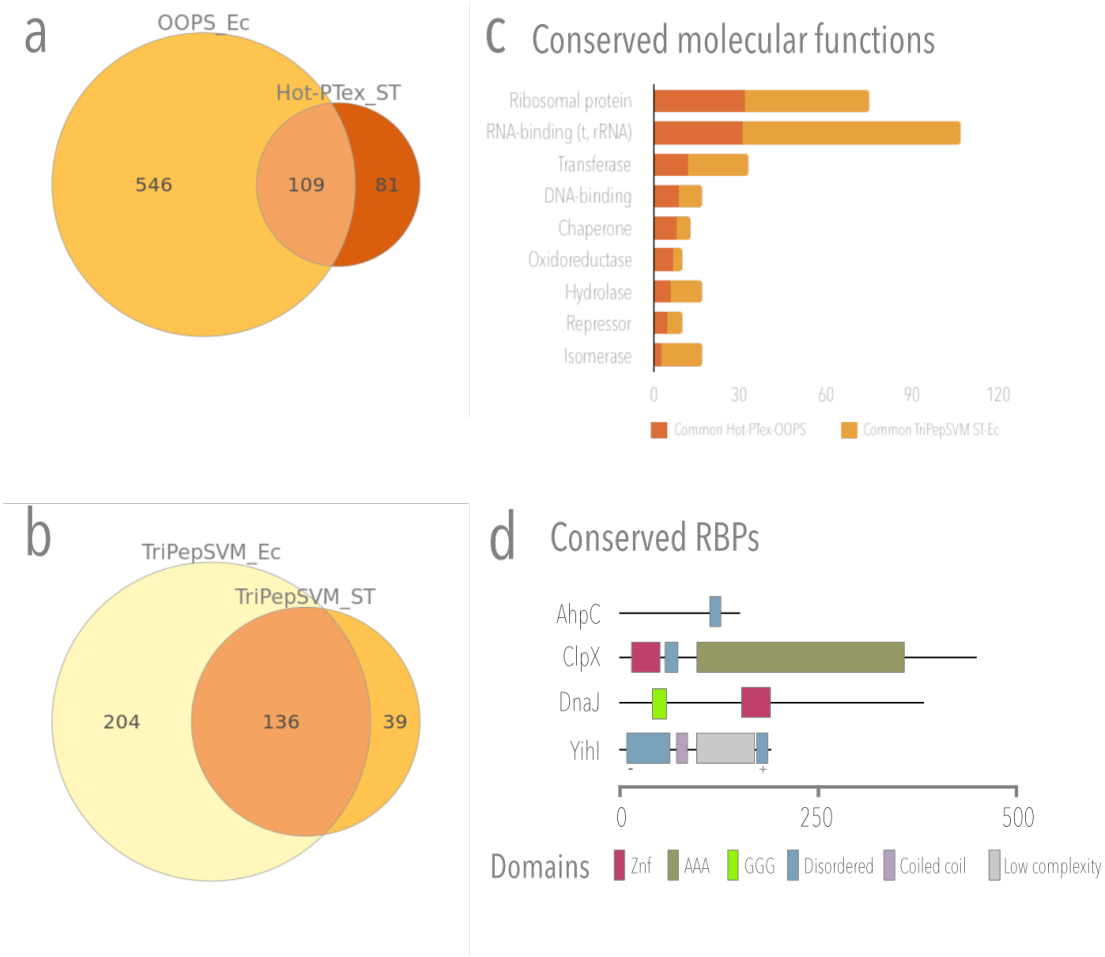

Figure 4.4: Conserved RBPs in *Salmonella* Typhimurium and *Escherichia coli* a PTex and OOPS share a core of 109 RBPs. **b** Similarly, 136 predicted RBPs are conserved between the two bacteria; **c** most of which have RNA-related molecular functions. **d** Conserved RBPs include the proteins here validated as RNA-binders *in vivo*, proteins which harbor mfeatures of the newly identified conserved RBDs [16; 30].

from *Salmonella*, which is not expected to be found in *E. coli*.

As mentioned before, intrinsically disordered regions (IDRs) rich in [K]- and [R]-rich tripeptide repeat motifs have been found in a variety of RBPs from yeast to humans [16]. This feature was taken upon in a support vector machine (SVM)-based method (TriPepSVM) for the classification of RBPs and non-RBPs [30]. Trained to recognised human RNA-binder features in bacterial proteomes, we have found hundreds of RBP candidates in *Salmonella* Typhimurium and *Escherichia coli* [30] (Fig. 4.4b).

Similarly, 136 proteins are conserved between the two bacterial species. The core RNA-interacting proteins classified theoretically by TriPepSVM, and experimentally by PTex and OOPS are enriched in GO terms related with RNA biology (Fig. 4.4c).

Additionally, the validated RBPs also contain some classic and unconventional RNA-binding domains recently discovered in Eukaryotas, e.g. zinc-finger (Znf), AAA ATPase (AAA), Glycine repeats (GGG) and disordered regions (Fig. 4.4d).

4.3.4 YihI, an unusual activator of the Double era-like (Der) GTPase

Der (double era-like GTPase) is an essential GTPase consisting of two GTPase-binding motifs (GD) followed by a KH-like domain, which becomes active upon interaction with YihI [127]. YihI over-expression affects cell growth, causing accumulation of rRNA precursors and an aberrant ribosome profile that was similar to that of Der-depleted cells, suggesting that Der and YihI are involved in the 50S ribosome assembly [127].

These authors shown that yihI deletion caused a shorter lag phase in the mutant strain in comparison with the wild-type, indicating that YihI may be a negative regulator for ribosome assembly, therefore, YihI was proposed as a GAP (GTPase-activating protein)-like protein that modulates Der function to negatively regulate cell growth at the beginning of exponential growth [127].

Der and YihI are highly conserved in Eubacteria. YihI from *E. coli* is a polypeptide of 169 amino acid residues (19 kDa) [127]. YihI from *Salmonella* contains 171 amino acids and shares 83% identity. Peculiarly, YihI is a dipolar protein where the N-terminal half (residues 1–88) contains 25 positively charged residues and only 7 negatively charged residues, while the C-terminal half (residues 89–169) harbors 28 negatively charged residues and 9 positively charged residues [127].

Interestingly, there is a sequence of 10 consecutive acidic residues (DDDEEEEEDE, residues 148-157 in *E. coli*) conserved among the YihI homologs, consistent with the tri-peptide RNA-binding motifs, EEE and DDD conserved in Eukaryotes [16] and predicted in *E. coli* and *S.* Typhimurium [30] (Fig. 4.4d). Deletion of this motif has been proven to disrupt the interaction with Der [127].

Due to its dipolar structure, YihI forms homodimers *in vivo*, however, it has been proved that the interaction between Der and YihI follow a stoichiometry of 1:1, and that the residues 146-169 in YihI and the C-terminal KH domain (or residues 201-297 in the GD2) in Der are required for the interaction [127]. Additionally, the yihI gene is activated

only at the initial growth lag-phase and its gene product seems to negatively regulate the cell growth at this lag phase [127]. Additionally, the KH-like domain in Der was shown to inhibit the two GTPase domains (GD), and that YihI was able to revert this inhibitory effect through its interaction with the KH-like domain [127].

The proposed model indicates that, in the presence of YihI at lag phase, newly synthesized rRNAs and ribosomal proteins form the p50S (a precursor of 50S ribosome); however, the completion of 50S is inhibited by the GTPase activating function of YihI on the GTP-bound Der. In the absence of YihI at exponential phase, the active form of Der completes the 50S ribosome biogenesis [127].

As an extension of this model, it seems plausible that the dimeric state of YihI can be disrupted by an interaction with RNA (possible its own accumulated mRNA) during the lag-phase, inducing the association of RNA-YihI complex and Der (via KH and GD2); as a consequence of the conformational rearrangement of the GD domains, the GTPase function of Der can be activated.

In this work, YihI was tested for *in vivo* RNA-binding only during stationary phase. Stationary phase is characterised, among others, by the decrease in protein synthesis, it is possible that the regulatory function of YihI is also active at this phase in order to prevent the maturation of new 50S ribosomes.

4.4 Outlook

PTex could be incorporated in more complex workflows, e.g. to reduce the amount of antibody needed to capture an RBP in CLIP-type experiments. Additionally, the fast and ease-of-use of PTex is an advantage for the implementation of complex high-throughput screenings.

In mRNA-interactome capture or in silica-based purification methods, mRNPs co-purify with not bound free-mRNAs [26; 142]. The excess of free-mRNAs are known to generate noise in RNA sequencing analysis. As PTex has the ability to remove the majority of the free-RNAs, the signal-to-noise ratio, and thereby, the identification of RNA binding motifs can be improved.

PTex can be useful for investigating RNPs in species that genetic manipulation or *in*

vivo RNA labelling are not possible.

With the establishment of PTex it became possible the systematic wide-screenning of RBPs in bacteria; e.g., by comparing RBPomes from Gram positive and Gram negative species we will be able to determine how these two evolutionarily distinct groups [143] control adaptive responses. This is the aim of a current collaboration with Dr. Sander Granneman (Edinburgh University, UK).

Additionally, together with the lab of Dr. Sebastién Ferreira-Cerca (Regensburg Univerity, Germany) we recently tested PTex on the archea *Haloferax volcanii* (data not shown), demonstrating that PTex can be used to expand global RNP analysis to species in all three branches of the tree of life.

Finally, combining RNA labelling strategies, e.g. with 4-SU, it will be possible to analyse cross-species RNA-protein interactions in *in vivo*.

Conclusions

With three consecutive organic extractions, PTex highly enriches for ribonucleoprotein complexes (RNPs) while efficiently depleting non-RBPs, non-cross-linked proteins and nucleic acids. PTex allowed the recovery of RNA-protein complexes from *in vivo* UV cross-linked HeLa and HEK293 cells, brain tissue from mice and the bacterial pathogen *Salmonella* Typhimurium. The singularity of the PTex approach lies in the introduction of the phenol-toluol mixture in the first step of purification, aimed to remove lipids and DNA and to prepare the cross-linked complexes for the further purification steps.

This work showed that as much as one third of the HEK293 proteome can associate with RNA *in vivo*. The underlying biological implications need to, however, be carefully addressed. It is important to underline that spacial proximity is not synonym of function, at least not in its classical meaning, unless the function is indeed to promote or maintain structures.

PTex represents a tool for the fast an unbiased recovery of RNPs from a variety of sources. In the recent decade, eukaryotic proteomes have been extensively scrutinised for RNA-binding proteins. However the two other kingdoms of life, archea and prokarya, had been neglected for technical reasons. With PTex, the first RNA-bound proteome from *Salmonella* Typhimurium is available, and uncovering the RBPome in other prokaryotes and archea is now in reach.

Bibliography

[1] Giovanni Salvi, Paolo De Los Rios, and Michele Vendruscolo. Effective interactions between chaotropic agents and proteins. *Proteins: Structure, Function and Genetics*, 61(3):492–499, 2005. ISSN 08873585. doi: 10.1002/prot.20626.

[2] Jeremy L. England and Gilad Haran. Role of Solvation Effects in Protein Denaturation: From Thermodynamics to Single Molecules and Back. 46(4):564–574, 2011. doi: 10.1016/j.cortex.2009.08.003.Predictive.

[3] Björn Schwanhüusser, Dorothea Busse, Na Li, Gunnar Dittmar, Johannes Schuchhardt, Jana Wolf, Wei Chen, and Matthias Selbach. Global quantification of mammalian gene expression control. *Nature*, 473(7347):337–342, 2011. ISSN 00280836. doi: 10.1038/nature10098.

[4] Niels H. Gehring, Elmar Wahle, and Utz Fischer. Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. *Trends in Biochemical Sciences*, 42(5):369–382, May 2017. ISSN 0968-0004. doi: 10.1016/j.tibs.2017.02.004. URL http://www.ncbi.nlm.nih.gov/pubmed/28268044.

[5] Sarah F. Mitchell and Roy Parker. Principles and Properties of Eukaryotic mRNPs. *Molecular Cell*, 54(4):547–588, may 2014. ISSN 1097-4164. doi: 10.1016/j.molcel.2014.04.033. URL http://www.ncbi.nlm.nih.gov/pubmed/24856220.

[6] Katharina Kramer, Timo Sachsenberg, Benedikt M. Beckmann, Saadia Qamar, Kum-Loong Boon, Matthias W. Hentze, Oliver Kohlbacher, and Henning Urlaub. Photo-cross-linking and high-resolution mass spectrometry for assignment of rna-binding sites in rna-binding proteins. *Nature Methods*, 11(10):1064–1070, Oct 2014. ISSN 1548-7105. doi: 10.1038/nmeth.3092. URL http://www.ncbi.nlm.nih.gov/pubmed/25173706.

[7] Tina Glisovic, Jennifer L Bachorik, Jeongsik Yong, and Gideon Dreyfuss. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*, 582 (14):1977–86, jun 2008. ISSN 0014-5793. doi: 10.1016/j.febslet.2008.03.004. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2858862{&}tool=pmcentrez{&}rendertype=abstract.

[8] a Gregory Matera, Rebecca M Terns, and Michael P Terns. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature reviews. Molecular cell biology*, 8(3):209–20, mar 2007. ISSN 1471-0072. doi: 10.1038/nrm2124. URL http://www.ncbi.nlm.nih.gov/pubmed/17318225.

[9] Erik Holmqvist and Jörg Vogel. RNA-binding proteins in bacteria. *Nature Reviews Microbiology*, 2018. ISSN 1740-1526. doi: 10.1038/s41579-018-0049-5. URL http://www.nature.com/articles/s41579-018-0049-5.

[10] Ji Young Youn, Boris J.A. Dyakov, Jianping Zhang, James D.R. Knight, Robert M. Vernon, Julie D. Forman-Kay, and Anne Claude Gingras. Properties of Stress Granule and P-Body Proteomes. *Molecular Cell*, 76(2):286–294, 2019. ISSN 10974164. doi: 10.1016/j.molcel.2019.09.014. URL https://doi.org/10.1016/j.molcel.2019.09.014.

[11] Debora L. Makino, Felix Halbach, and Elena Conti. The rna exosome and proteasome: common principles of degradation control. *Nature reviews. Molecular cell biology*, 14(10):654–660, Oct 2013. ISSN 1471-0080. doi: 10.1038/nrm3657. URL http://www.ncbi.nlm.nih.gov/pubmed/23989960.

[12] Matthias W. Hentze, Alfredo Castello, Thomas Schwarzl, and Thomas Preiss. A brave new world of RNA-binding proteins. *Nature Reviews Molecular Cell Biology*, 19(5):327–341, 2018. ISSN 14710080. doi: 10.1038/nrm.2017.130. URL http://dx.doi.org/10.1038/nrm.2017.130.

[13] Clarisse van der Feltz, Kelsey Anthony, Axel Brilot, and Daniel a Pomeranz Krummel. Architecture of the spliceosome. *Biochemistry*, 51(16):3321–33, apr 2012. ISSN 1520-4995. doi: 10.1021/bi201215r. URL http://www.ncbi.nlm.nih.gov/pubmed/22471593.

[14] Alexander G. Baltz, Mathias Munschauer, Björn Schwanhäusser, Alexandra Vasile, Yasuhiro Murakawa, Markus Schueler, Noah Youngs, Duncan Penfold-Brown, Kevin Drew, Miha Milek, Emanuel Wyler, Richard Bonneau, Matthias Selbach, Christoph Dieterich, and Markus Landthaler. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Molecular Cell*, 46(5):674–690, Jun 2012. ISSN 1097-4164. doi: 10.1016/j.molcel.2012.05.021. URL http://www.ncbi.nlm.nih.gov/pubmed/22681889.

[15] Alfredo Castello, Bernd Fischer, Katrin Eichelbaum, Rastislav Horos, Benedikt M. Beckmann, Claudia Strein, Norman E. Davey, David T. Humphreys, Thomas Preiss, Lars M. Steinmetz, Jeroen Krijgsveld, and Matthias W. Hentze. Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell*, 149(6):1393–1406, jun 2012. ISSN 1097-4172. doi: 10.1016/j.cell.2012.04.031. URL http://www.ncbi.nlm.nih.gov/pubmed/22658674.

[16] Benedikt M Beckmann, Rastislav Horos, Bernd Fischer, Alfredo Castello, Katrin Eichelbaum, Anne-Marie Alleaume, Thomas Schwarzl, Tomaž Curk, Sophia Foehr, Wolfgang Huber, et al. The rna-binding proteomes from yeast to man harbour conserved enigmrbps. *Nature communications*, 6, 2015.

[17] Tim R Mercer and John S Mattick. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology*, 20(3):300–7, mar 2013. ISSN 1545-9985. doi: 10.1038/nsmb.2480. URL http://www.ncbi.nlm.nih.gov/pubmed/23463315.

[18] Leah Sabin, M. Joaquina Delás, and Gregory J Hannon. Dogma Derailed : The Many Influences of RNA on the Genome. *Molecular cell*, 49(5), 2014. doi: 10.1016/j.molcel.2013.02.010.Dogma.

[19] Alfredo Castello, Bernd Fischer, Matthias W Hentze, and Thomas Preiss. RNA-binding proteins in Mendelian disease. *Trends in genetics : TIG*, 29(5):318–27, may 2013. ISSN 0168-9525. doi: 10.1016/j.tig.2013.01.004. URL http://www.ncbi.nlm.nih.gov/pubmed/23415593.

[20] Bruno Pereira, Marc Billaud, and Raquel Almeida. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends in Cancer*, 3(7):506–528, 2017. ISSN 24058033. doi: 10.1016/j.trecan.2017.05.003. URL http://dx.doi.org/10.1016/j.trecan.2017.05.003.

[21] Tara Vanderweyde, Katie Youmans, Liqun Liu-Yesucevitz, and Benjamin Wolozin. The Role Stress Granules and RNA Binding Proteins in Neurodegeneration. *Gerontology*, 59(6):1–13, 2013. doi: 10.1159/000354170.The.

[22] Kai Papenfort and Joerg Vogel. Regulatory RNA in bacterial pathogens. *Cell Host and Microbe*, 8(1):116–127, 2010. ISSN 19313128. doi: 10.1016/j.chom.2010.06.008.

[23] Elke Van Assche, Sandra Van Puyvelde, Jos Vanderleyden, and Hans P. Steenackers. RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Frontiers in Microbiology*, 6(MAR):1–16, 2015. ISSN 1664302X. doi: 10.3389/fmicb.2015.00141.

[24] Noriyuki Yoshida and Tominori Kimura. Pathogen-associated regulatory non-coding RNAs and oncogenesis. *Frontiers in bioscience (Landmark edition)*, 22: 1599–1621, 2017. ISSN 1093-4715. URL http://www.ncbi.nlm.nih.gov/pubmed/28410134.

[25] Alessandro Pagliuso, To Nam Tham, Eric Allemand, Stevens Robertin, Bruno Dupuy, Quentin Bertrand, Christophe Bécavin, Mikael Koutero, Valérie Najburg, Marie-Anne Nahori, Frédéric Tangy, Fabrizia Stavru, Sergey Bessonov, Andréa Dessen, Christian Muchardt, Alice Lebreton, Anastassia V Komarova, and Pascale

Cossart. An RNA-Binding Protein Secreted by a Bacterial Pathogen Modulates RIG-I Signaling. *Cell host & microbe*, pages 1–13, 2019. ISSN 1934-6069. doi: 10.1016/j.chom.2019.10.004. URL http://www.ncbi.nlm.nih.gov/pubmed/31761719.

[26] Alfredo Castello, Bernd Fischer, Christian K. Frese, Rastislav Horos, Anne-Marie Alleaume, Sophia Foehr, Tomaz Curk, Jeroen Krijgsveld, and Matthias W. Hentze. Comprehensive identification of rna-binding domains in human cells. *Molecular cell*, 63(4):696–710, Aug 2016. ISSN 1097-4164. doi: 10.1016/j.molcel.2016.06.029. URL http://www.ncbi.nlm.nih.gov/pubmed/27453046.

[27] Bradley M Lunde, Claire Moore, and Gabriele Varani. RNA-binding proteins: modular design for efficient function. *Nature reviews. Molecular cell biology*, 8(6): 479–90, jun 2007. ISSN 1471-0072. doi: 10.1038/nrm2178. URL http://www.ncbi.nlm.nih.gov/pubmed/17473849.

[28] Sarah F Mitchell, Saumya Jain, Meipei She, and Roy Parker. Global Analysis of Yeast mRNPs. *Nat Struct Mol Biol*, 20(1):127–133, 2013. doi: 10.1038/nsmb.2468.Global.

[29] Amol Panhale, Florian M. Richter, Fidel Ramírez, Maria Shvedunova, Thomas Manke, Gerhard Mittler, and Asifa Akhtar. CAPRI enables comparison of evolutionarily conserved RNA interacting regions. *Nature Communications*, 10(1), 2019. ISSN 20411723. doi: 10.1038/s41467-019-10585-3. URL http://dx.doi.org/10.1038/s41467-019-10585-3.

[30] Annkatrin Bressin, Roman Schulte-Sasse, Davide Figini, Erika C. Urdaneta, Benedikt M. Beckmann, and Annalisa Marsico. Tripepsvm - de novo prediction of rna-binding proteins based on short amino acid motifs. *bioRxiv*, 2018. doi: 10.1101/466151. URL https://www.biorxiv.org/content/early/2018/11/08/466151.

[31] Thomas Conrad, Anne-Susann Albrecht, Veronica Rodrigues de Melo Costa, Sascha Sauer, David Meierhofer, and Ulf Andersson Ørom. Serial interactome capture of the human cell nucleus. *Nature communications*, 7:11212, Apr 2016. ISSN 2041-1723. doi: 10.1038/ncomms11212. URL http://www.ncbi.nlm.nih.gov/pubmed/27040163.

[32] Markus Hafner, Markus Landthaler, Lukas Burger, Mohsen Khorshid, Jean Hausser, Philipp Berninger, Andrea Rothballer, Manuel Ascano, Anna-Carina Jungkamp, Mathias Munschauer, Alexander Ulrich, Greg S. Wardle, Scott Dewell, Mihaela Zavolan, and Thomas Tuschl. Transcriptome-wide identification of rna-binding protein and microrna target sites by par-clip. *Cell*, 141(1):129–141, Apr 2010. ISSN 1097-4172. doi: 10.1016/j.cell.2010.03.009. URL http://www.ncbi.nlm.nih.gov/pubmed/20371350.

[33] S a Tenenbaum, C C Carson, P J Lager, and J D Keene. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *Proceedings of the National Academy of Sciences of the United States of America*, 97(26):14085–90, dec 2000. ISSN 0027-8424. doi: 10.1073/pnas.97.26.14085. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=18875{&}tool=pmcentrez{&}rendertype=abstract.

[34] André P Gerber, Stefan Luschnig, Mark a Krasnow, Patrick O Brown, and Daniel Herschlag. Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (12):4487–92, mar 2006. ISSN 0027-8424. doi: 10.1073/pnas.0509260103. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1400586{&}tool=pmcentrez{&}rendertype=abstract.

[35] Markus Landthaler, Dimos Gaidatzis, Andrea Rothballer, P O Y U Chen, Steven Joseph Soll, Lana Dinic, Tolulope Ojo, Markus Hafner, Mihaela Zavolan, and Thomas Tuschl. Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. pages 2580–2596, 2008. doi: 10.1261/rna.1351608.transcripts.

[36] Stavroula Mili and Joan A Steitz. Evidence for reassociation of RNA-binding proteins after cell lysis : Implications for the interpretation of immunoprecipitation analyses Evidence for reassociation of RNA-binding proteins after cell lysis : Implications for the interpretation of immunopr. *RNA*, pages 1692–1694, 2004. doi: 10.1261/rna.7151404.mRNA.

[37] Jernej Ule, Kirk B. Jensen, Matteo Ruggiu, Aldo Mele, Aljaz Ule, and Robert B. Darnell. Clip identifies nova-regulated rna networks in the brain. *Science (New York, N.Y.)*, 302(5648):1212–1215, Nov 2003. ISSN 1095-9203. doi: 10.1126/science.1090095. URL http://www.ncbi.nlm.nih.gov/pubmed/14615540.

[38] J. R. Greenberg. Ultraviolet light-induced crosslinking of mRNA to proteins. *Nucleic Acids Research*, 6(2):715–732, Feb 1979. ISSN 0305-1048. URL http://www.ncbi.nlm.nih.gov/pubmed/424311.

[39] Aoife C McMahon, Reazur Rahman, Hua Jin, James L Shen, Allegra Fieldsend, Weifei Luo, Michael Rosbash. Hijacking an RNA-editing enzyme to identify cell-specific targets of RNA-binding proteins. *Physiology & behavior*, 176(3):139–148, 2017. doi: 10.1016/j.physbeh.2017.03.040.

[40] Christopher P. Lapointe, Daniel Wilinski, Harriet A.J. Saunders, and Marvin Wickens. Protein-RNA networks revealed through covalent RNA marks. *Nature*

Methods, 12(12):1163–1170, 2015. ISSN 15487105. doi: 10.1038/nmeth.3651. URL http://dx.doi.org/10.1038/nmeth.3651.

[41] Chaolin Zhang and Robert B. Darnell. Mapping in vivo protein-rna interactions at single-nucleotide resolution from hits-clip data. *Nature biotechnology*, 29(7):607–614, Jun 2011. ISSN 1546-1696. doi: 10.1038/nbt.1873. URL http://www.ncbi.nlm.nih.gov/pubmed/21633356.

[42] Sander Granneman, Grzegorz Kudla, Elisabeth Petfalski, and David Tollervey. Identification of protein binding sites on u3 snorna and pre-rrna by uv cross-linking and high-throughput analysis of cdnas. *Proceedings of the National Academy of Sciences of the United States of America*, 106(24):9613–9618, Jun 2009. ISSN 1091-6490. doi: 10.1073/pnas.0901997106. URL http://www.ncbi.nlm.nih.gov/pubmed/19482942.

[43] Julian König, Kathi Zarnack, Gregor Rot, Tomaz Tomaž Curk, Melis Kayikci, Blaz Zupan, Daniel J Turner, Nicholas M Luscombe, Jernej Ule, Julian Konig, Kathi Zarnack, Gregor Rot, Tomaz Tomaž Curk, Melis Kayikci, Blaz Zupan, Daniel J Turner, Nicholas M Luscombe, and Jernej Ule. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat Struct Mol Biol*, 17(7):909–915, jul 2010. ISSN 1545-9993. doi: 10.1038/nsmb.1838.iCLIP. URL http://dx.doi.org/10.1038/nsmb.1838http://www.nature.com/nsmb/journal/v17/n7/abs/nsmb.1838.html{#}supplementary-informationhttp://www.nature.com/nsmb/journal/v17/n7/abs/nsmb.1838.html.

[44] Eric L. Van Nostrand, Gabriel A. Pratt, Alexander A. Shishkin, Chelsea Gelboin-Burkhart, Mark Y. Fang, Balaji Sundararaman, Steven M. Blue, Thai B. Nguyen, Christine Surka, Keri Elkins, Rebecca Stanton, Frank Rigo, Mitchell Guttman, and Gene W. Yeo. Robust transcriptome-wide discovery of rna-binding protein binding sites with enhanced clip (eclip). *Nature methods*, 13(6):508–514, Jun 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3810. URL http://www.ncbi.nlm.nih.gov/pubmed/27018577.

[45] Donny D. Licatalosi, Xuan Ye, and Eckhard Jankowsky. Approaches for measuring the dynamics of RNA–protein interactions. *Wiley Interdisciplinary Reviews: RNA*, (July):1–23, 2019. ISSN 1757-7004. doi: 10.1002/wrna.1565.

[46] Emily C. Wheeler, Eric L. Van Nostrand, and Gene W. Yeo. Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. *Wiley Interdisciplinary Reviews: RNA*, (Figure 1):e1436, 2017. ISSN 17577004. doi: 10.1002/wrna.1436. URL http://doi.wiley.com/10.1002/wrna.1436.

[47] Lin Chenyu and Wayne O Miles. Beyond CLIP: advances and opportunities to measure RBP–RNA and RNA–RNA interactions. *Nucleic Acids Research*, pages

1–12, 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz076. URL http://fdslive.oup.com/www.oup.com/pdf/production{_}in{_}progress.pdf.

[48] Muthukumar Ramanathan, Douglas F. Porter, and Paul A. Khavari. Methods to study RNA–protein interactions. *Nature Methods*, 16(3):225–234, 2019. ISSN 15487105. doi: 10.1038/s41592-019-0330-1. URL http://dx.doi.org/10.1038/s41592-019-0330-1.

[49] Gideon Dreyfuss, Yang Do, and Stephen A Adam. Characterization of Heterogeneous Nuclear RNA-Protein Complexes In Vivo with Monoclonal Antibodies. *Molecular and Cellular Biology*, 4(6):1104–1114, 1984.

[50] Stefanie Gerstberger, Markus Hafner, and Thomas Tuschl. A census of human rna-binding proteins. *Nature reviews. Genetics*, 15(12):829–845, Dec 2014. ISSN 1471-0064. doi: 10.1038/nrg3813. URL http://www.ncbi.nlm.nih.gov/pubmed/25365966.

[51] Ci Chu and Howard Y Chang. *ChIRP-MS: RNA-Directed Proteomic Discovery*, pages 37–45. Springer New York, New York, NY, 2018. ISBN 978-1-4939-8766-5. doi: 10.1007/978-1-4939-8766-5_3. URL https://doi.org/10.1007/978-1-4939-8766-5{_}3.

[52] Colleen A McHugh and Mitchell Guttman. *RAP-MS: A Method to Identify Proteins that Interact Directly with a Specific RNA Molecule in Cells*, pages 473–488. Springer New York, New York, NY, 2018. ISBN 978-1-4939-7213-5. doi: 10.1007/978-1-4939-7213-5_31. URL https://doi.org/10.1007/978-1-4939-7213-5{_}31.

[53] Anand Minajigi, John Froberg, Chunyao Wei1, Hongjae Sunwoo1, Barry Kesner1, David Colognori1, Derek Lessing1, Bernhard Payer1,#, Myriam Boukhali2, Wilhelm Haas2 and Jeannie T. Lee1. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Physiology & behavior*, 176(3):139–148, 2017. doi: 10.1016/j.physbeh.2017.03.040.

[54] Muthukumar Ramanathan, Karim Majzoub, Deepti S Rao, Poornima H Neela, Brian J Zarnegar, Smarajit Mondal, Julien G Roth, Hui Gai, Joanna R Kovalski, Zurab Siprashvili, Theo D Palmer, Jan E Carette, and Paul A Khavari. RNA–protein interaction detection in living cells. *Nature Methods*, 15:207, feb 2018. URL https://doi.org/10.1038/nmeth.4601http://10.0.4.14/nmeth.4601https://www.nature.com/articles/nmeth.4601{#}supplementary-information.

[55] Joel I Perez-perri, Birgit Rogell, Thomas Schwarzl, Frank Stein, Yang Zhou, and Mandy Rettel. Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nature Communications*, 4, 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-06557-8. URL www.nature.com/naturecommunications.

[56] Xichen Bao, Xiangpeng Guo, Menghui Yin, Muqddas Tariq, Yiwei Lai, Shahzina

Kanwal, Jiajian Zhou, Na Li, Yuan Lv, Carlos Pulido-Quetglas, Xiwei Wang, Lu Ji, Muhammad J. Khan, Xihua Zhu, Zhiwei Luo, Changwei Shao, Do-Hwan Lim, Xiao Liu, Nan Li, Wei Wang, Minghui He, Yu-Lin Liu, Carl Ward, Tong Wang, Gong Zhang, Dongye Wang, Jianhua Yang, Yiwen Chen, Chaolin Zhang, Ralf Jauch, Yun-Gui Yang, Yangming Wang, Baoming Qin, Minna-Liisa Anko, Andrew P. Hutchins, Hao Sun, Huating Wang, Xiang-Dong Fu, Biliang Zhang, and Miguel A. Esteban. Capturing the interactome of newly transcribed RNA. *Nature Methods*, 15(3):213–220, Mar 2018. ISSN 1548-7105. doi: 10.1038/nmeth.4595. URL http://www.ncbi.nlm.nih.gov/pubmed/29431736.

[57] Rongbing Huang, Mengting Han, Liying Meng, and Xing Chen. Transcriptome-wide discovery of coding and noncoding RNA-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 115(17):E3879–E3887, Apr 2018. ISSN 1091-6490. doi: 10.1073/pnas.1718406115. URL http://www.ncbi.nlm.nih.gov/pubmed/29636419.

[58] Daniel Benhalevy, Dimitrios G. Anastasakis, and Markus Hafner. Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. *Nature Methods*, 15(12):1074–1082, 2018. ISSN 15487105. doi: 10.1038/s41592-018-0220-y. URL http://dx.doi.org/10.1038/s41592-018-0220-y.

[59] Jakob Trendel, Thomas Schwarzl, Rastislav Horos, Ananth Prakash, Alex Bateman, Matthias W. Hentze, and Jeroen Krijgsveld. The Human RNA-Binding Proteome and Its Dynamics during Translational Arrest. *Cell*, 176(1-2):391–403.e19, Jan 2019. ISSN 1097-4172. doi: 10.1016/j.cell.2018.11.004. URL http://www.ncbi.nlm.nih.gov/pubmed/30528433.

[60] Rayner M. L. Queiroz, Tom Smith, Eneko Villanueva, Maria Marti-Solano, Mie Monti, Mariavittoria Pizzinga, Dan-Mircea Mirea, Manasa Ramakrishna, Robert F. Harvey, Veronica Dezi, Gavin H. Thomas, Anne E. Willis, and Kathryn S. Lilley. Comprehensive identification of RNA-protein interactions in any organism using orthogonal organic phase separation (OOPS). *Nature Biotechnology*, 37(2):169–178, Feb 2019. ISSN 1546-1696. doi: 10.1038/s41587-018-0001-2. URL http://www.ncbi.nlm.nih.gov/pubmed/30607034.

[61] Erika C Urdaneta, Carlos H Vieira-Vieira, Timon Hick, Hans-Herrmann Wessels, Davide Figini, Rebecca Moschall, Jan Medenbach, Uwe Ohler, Sander Granneman, Matthias Selbach, and Benedikt M Beckmann. Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nature Communications*, 10(1):990, 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-08942-3. URL https://doi.org/10.1038/s41467-019-08942-3.

[62] Erika C Urdaneta and Benedikt M Beckmann. Fast and unbiased purification of

rna-protein complexes after uv cross-linking. *bioRxiv*, 2019. doi: 10.1101/597575. URL https://www.biorxiv.org/content/early/2019/04/03/597575.

[63] K. S. KIRBY. A new method for the isolation of ribonucleic acids from mammalian tissues. *The Biochemical Journal*, 64(3):405–408, Nov 1956. ISSN 0264-6021. URL http://www.ncbi.nlm.nih.gov/pubmed/13373784.

[64] P. Chomczynski and N. Sacchi. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162(1):156–159, Apr 1987. ISSN 0003-2697. doi: 10.1006/abio.1987.9999. URL http://www.ncbi.nlm.nih.gov/pubmed/2440339.

[65] P. Chomczynski and K. Mackey. Substitution of chloroform by bromo-chloropropane in the single-step method of RNA isolation. *Analytical Biochemistry*, 225(1):163–164, Feb 1995. ISSN 0003-2697. doi: 10.1006/abio.1995.1126. URL http://www.ncbi.nlm.nih.gov/pubmed/7539982.

[66] Piotr Chomcznski and Nicoletta Sacchi. The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on. *Nature protocols*, 1(2):581–585, 2006. doi: 10.1038/nprot.2006.83.

[67] P Zumbo. Phenol-chloroform Extraction. URL http://physiology.med.cornell.edu/faculty/mason/lab/zumbo/files/PHENOL-CHLOROFORM.pdf.

[68] Soroth Chey, Claudia Claus, and Uwe Gerd Liebert. Improved method for simultaneous isolation of proteins and nucleic acids. *Analytical Biochemistry*, 411(1):164–166, Apr 2011. ISSN 1096-0309. doi: 10.1016/j.ab.2010.11.020. URL http://www.ncbi.nlm.nih.gov/pubmed/21094121.

[69] Pallavi Thaplyal and Philip C. Bevilacqua. Experimental approaches for measuring pKa's in RNA and DNA. *Methods in Enzymology*, 549:189–219, 2014. ISSN 1557-7988. doi: 10.1016/B978-0-12-801122-5.00009-X. URL http://www.ncbi.nlm.nih.gov/pubmed/25432750.

[70] M. D. Shetlar. Cross-Linking of Proteins to Nucleic Acids by Ultraviolet Light. *Photochem. Photobiol. Rev.*, (5):105–197, 1980.

[71] A. Favre, G. Moreno, M. O. Blondel, J. Kliber, F. Vinzens, and C. Salet. 4-Thiouridine photosensitized RNA-protein crosslinking in mammalian cells. *Biochemical and Biophysical Research Communications*, 141(2):847–854, Dec 1986. ISSN 0006-291X. URL http://www.ncbi.nlm.nih.gov/pubmed/2432896.

[72] K. C. Smith and R. T. Aplin. A mixed photoproduct of uracil and cysteine (5-S-cysteine-6-hydrouracil). A possible model for the in vivo cross-linking of deoxyribonucleic acid and protein by ultraviolet light. *Biochemistry*, 5(6):2125–

2130, Jun 1966. ISSN 0006-2960. URL http://www.ncbi.nlm.nih.gov/pubmed/5963455.

[73] E. I. Budowsky, M. S. Axentyeva, G. G. Abdurashidova, N. A. Simukova, and L. B. Rubin. Induction of polynucleotide-protein cross-linkages by ultraviolet irradiation. Peculiarities of the high-intensity laser pulse irradiation. *European Journal of Biochemistry*, 159(1):95–101, Aug 1986. ISSN 0014-2956. URL http://www.ncbi.nlm.nih.gov/pubmed/2427338.

[74] M. D. Shetlar, J. Carbone, E. Steady, and K. Hom. Photochemical addition of amino acids and peptides to polyuridylic acid. *Photochemistry and Photobiology*, 39(2):141–144, Feb 1984. ISSN 0031-8655. URL http://www.ncbi.nlm.nih.gov/pubmed/6709720.

[75] P. R. Paradiso, Y. Nakashima, and W. Konigsberg. Photochemical cross-linking of protein . nucleic acid complexes. The attachment of the fd gene 5 protein to fd DNA. *The Journal of Biological Chemistry*, 254(11):4739–4744, Jun 1979. ISSN 0021-9258. URL http://www.ncbi.nlm.nih.gov/pubmed/374415.

[76] K. C. Smith and D. H. Meun. Kinetics of the photochemical addition of [35S] cysteine to polynucleotides and nucleic acids. *Biochemistry*, 7(3):1033–1037, Mar 1968. ISSN 0006-2960. URL http://www.ncbi.nlm.nih.gov/pubmed/5690558.

[77] I. G. Pashev, S. I. Dimitrov, and D. Angelov. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends in Biochemical Sciences*, 16(9):323–326, Sep 1991. ISSN 0968-0004. URL http://www.ncbi.nlm.nih.gov/pubmed/1835191.

[78] Benedikt M. Beckmann, Alfredo Castello, and Jan Medenbach. The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Archiv : European journal of physiology*, 468(6):1029–1040, Jun 2016. ISSN 1432-2013. doi: 10.1007/s00424-016-1819-4. URL http://www.ncbi.nlm.nih.gov/pubmed/27165283.

[79] R. Brimacombe, W. Stiege, A. Kyriatsoulis, and P. Maly. Intra-RNA and RNA-protein cross-linking techniques in Escherichia coli ribosomes. *Methods in Enzymology*, 164:287–309, 1988. ISSN 0076-6879. URL http://www.ncbi.nlm.nih.gov/pubmed/3071669.

[80] N Sarkar. Polyadenylation of mRNA in prokaryotes. *Annual review of biochemistry*, 66:173–197, 1997. ISSN 0066-4154 (Print). doi: 10.1146/annurev.biochem.66.1.173.

[81] Victoria Portnoy, Elena Evguenieva-Hackenberg, Franziska Klein, Pamela Walter, Esben Lorentzen, Gabriele Klug, and Gadi Schuster. RNA polyadenylation in Archaea: Not observed in Haloferax while the exosome polynucleotidylates RNA

in Sulfolobus. *EMBO Reports*, 6(12):1188–1193, 2005. ISSN 1469221X. doi: 10.1038/sj.embor.7400571.

[82] Ekaterina Dzyubak and Mee Ngan F. Yap. The expression of antibiotic resistance methyltransferase correlates with mRNA stability independently of ribosome stalling. *Antimicrobial Agents and Chemotherapy*, 60(12):7178–7188, 2016. ISSN 10986596. doi: 10.1128/AAC.01806-16.

[83] Man Yu and Stewart J. Levine. Toll-like receptor, rig-i-like receptors and the nlrp3 inflammasome: key modulators of innate immune responses to double-stranded rna viruses. *Cytokine & growth factor reviews*, 22(2):63–72, Apr 2011. ISSN 1879-0305. doi: 10.1016/j.cytogfr.2011.02.001. URL http://www.ncbi.nlm.nih.gov/pubmed/21466970.

[84] Alexandre Smirnov, Konrad U. Förstner, Erik Holmqvist, Andreas Otto, Regina Günster, Dörte Becher, Richard Reinhardt, and Jörg Vogel. Grad-seq guides the discovery of proq as a major small rna-binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, 113(41):11591–11596, Oct 2016. ISSN 1091-6490. doi: 10.1073/pnas.1609981113. URL http://www.ncbi.nlm.nih.gov/pubmed/27671629.

[85] Erik Holmqvist, Patrick R Wright, Lei Li, Thorsten Bischler, Lars Barquist, Richard Reinhardt, Rolf Backofen, and Jörg Vogel. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *The EMBO Journal*, 35(9):991–1011, 2016. ISSN 0261-4189. doi: 10.15252/embj.201593360. URL http://emboj.embopress.org/lookup/doi/10.15252/embj.201593360.

[86] Brian K. Coombes, Nat F. Brown, Yanet Valdez, John H. Brumell, and B. Brett Finlay. Expression and secretion of Salmonella pathogenicity island-2 virulence genes in response to acidification exhibit differential requirements of a functional type III secretion apparatus and SsaL. *Journal of Biological Chemistry*, 279(48):49804–49815, 2004. ISSN 00219258. doi: 10.1074/jbc.M404299200.

[87] Rob van Nues, Gabriele Schweikert, Erica de Leau, Alina Selega, Andrew Langford, Ryan Franklin, Ira Iosub, Peter Wadsworth, Guido Sanguinetti, and Sander Granneman. Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. *Nature Communications*, 8(1):12, Apr 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-00025-5. URL http://www.ncbi.nlm.nih.gov/pubmed/28400552.

[88] S. Uzzau, N. Figueroa-Bossi, S. Rubino, and L. Bossi. Epitope tagging of chromosomal genes in salmonella. *Proceedings of the National Academy of Sciences of the United States of America*, 98(26):15264–15269, Dec 2001. ISSN

0027-8424. doi: 10.1073/pnas.261348198. URL http://www.ncbi.nlm.nih.gov/pubmed/11742086.

[89] Jai J. Tree, Sander Granneman, Sean P. McAteer, David Tollervey, and David L. Gally. Identification of bacteriophage-encoded anti-srnas in pathogenic escherichia coli. 55, 2014. doi: 10.1016/j.molcel.2014.05.006. URL https://doi.org/10.1016/j.molcel.2014.05.006.

[90] Benedikt M. Beckmann, Arnold Grünweller, Michael H. W. Weber, and Roland K. Hartmann. Northern blot detection of endogenous small rnas (approximately14 nt) in bacterial total rna extracts. *Nucleic acids research*, 38(14):e147, Aug 2010. ISSN 1362-4962. doi: 10.1093/nar/gkq437. URL http://www.ncbi.nlm.nih.gov/pubmed/20504856.

[91] Curtis T Rueden, Johannes Schindelin, Mark C Hiner, Barry E DeZonia, Alison E Walter, Ellen T Arena, and Kevin W Eliceiri. *BMC Bioinformatics*, (1):529. ISSN 1471-2105. doi: 10.1186/s12859-017-1934-z.

[92] Rebecca Moschall, Daniela Strauss, Marina García-Beyaert, Fátima Gebauer, and Jan Medenbach. Sister-of-sex-lethal is a repressor of translation. *RNA (New York, N.Y.)*, 24(2):149–158, Feb 2018. ISSN 1469-9001. doi: 10.1261/rna.063776.117. URL http://www.ncbi.nlm.nih.gov/pubmed/29089381.

[93] D. Wessel and U. I. Flügge. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry*, 138(1):141–143, Apr 1984. ISSN 0003-2697. URL http://www.ncbi.nlm.nih.gov/pubmed/6731838.

[94] Juri Rappsilber, Yasushi Ishihama, and Matthias Mann. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and lc/ms sample pretreatment in proteomics. *Analytical chemistry*, 75(3):663–670, Feb 2003. ISSN 0003-2700. URL http://www.ncbi.nlm.nih.gov/pubmed/12585499.

[95] Jürgen Cox and Matthias Mann. Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26(12):1367–1372, Dec 2008. ISSN 1546-1696. doi: 10.1038/nbt.1511. URL http://www.ncbi.nlm.nih.gov/pubmed/19029910.

[96] Jürgen Cox, Marco Y. Hein, Christian A. Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. *Molecular & cellular proteomics : MCP*, 13(9):2513–2526, Sep 2014. ISSN 1535-9484. doi: 10.1074/mcp.M113.031591. URL http://www.ncbi.nlm.nih.gov/pubmed/24942700.

[97] Benedikt M. Beckmann. RNA interactome capture in yeast. *Methods (San Diego,*

Calif.), 118-119:82–92, Apr 2017. ISSN 1095-9130. doi: 10.1016/j.ymeth.2016.12.008. URL http://www.ncbi.nlm.nih.gov/pubmed/27993706.

[98] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47, Apr 2015. ISSN 1362-4962. doi: 10.1093/nar/gkv007. URL http://www.ncbi.nlm.nih.gov/pubmed/25605792.

[99] Huaiyu Mi, Xiaosong Huang, Anushya Muruganujan, Haiming Tang, Caitlin Mills, Diane Kang, and Paul D. Thomas. Panther version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic acids research*, 45(D1):D183–D189, Jan 2017. ISSN 1362-4962. doi: 10.1093/nar/gkw1138. URL http://www.ncbi.nlm.nih.gov/pubmed/27899595.

[100] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*, 37(1):1–13, Jan 2009. ISSN 1362-4962. doi: 10.1093/nar/gkn923. URL http://www.ncbi.nlm.nih.gov/pubmed/19033363.

[101] Ivica Letunic and Peer Bork. 20 years of the smart protein domain annotation resource. *Nucleic acids research*, 46(D1):D493–D496, Jan 2018. ISSN 1362-4962. doi: 10.1093/nar/gkx922. URL http://www.ncbi.nlm.nih.gov/pubmed/29040681.

[102] Henrike Maatz, Marcin Kolinski, Norbert Hubner, and Markus Landthaler. Transcriptome-wide identification of rna-binding protein binding sites using photoactivatable-ribonucleoside-enhanced crosslinking immunoprecipitation (par-clip). *Current protocols in molecular biology*, 118:27.6.1–27.6.19, Apr 2017. ISSN 1934-3647. doi: 10.1002/cpmb.35. URL http://www.ncbi.nlm.nih.gov/pubmed/28369676.

[103] Douglas F. Porter, Yvonne Y. Koh, Brett VanVeller, Ronald T. Raines, and Marvin Wickens. Target selection by natural and redesigned puf proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 112(52):15868–15873, Dec 2015. ISSN 1091-6490. doi: 10.1073/pnas.1508501112. URL http://www.ncbi.nlm.nih.gov/pubmed/26668354.

[104] Daniel Benhalevy, Hannah L. McFarland, Aishe A. Sarshad, and Markus Hafner. Par-clip and streamlined small rna cdna library preparation protocol for the identification of rna binding protein target sites. *Methods (San Diego, Calif.)*, 118-119:41–49, Apr 2017. ISSN 1095-9130. doi: 10.1016/j.ymeth.2016.11.009. URL http://www.ncbi.nlm.nih.gov/pubmed/27871973.

[105] David L. Corcoran, Stoyan Georgiev, Neelanjan Mukherjee, Eva Gottwein, Rebecca L. Skalsky, Jack D. Keene, and Uwe Ohler. Paralyzer: definition of rna binding sites from par-clip short-read sequence data. *Genome biology*, 12(8):R79, Aug 2011. ISSN 1474-760X. doi: 10.1186/gb-2011-12-8-r79. URL http://www.ncbi.nlm.nih.gov/pubmed/21851591.

[106] Neelanjan Mukherjee, Nicholas C. Jacobs, Markus Hafner, Elizabeth A. Kennington, Jeffrey D. Nusbaum, Thomas Tuschl, Perry J. Blackshear, and Uwe Ohler. Global target mrna specification and regulation by the rna-binding protein zfp36. *Genome biology*, 15(1):R12, Jan 2014. ISSN 1474-760X. doi: 10.1186/gb-2014-15-1-r12. URL http://www.ncbi.nlm.nih.gov/pubmed/24401661.

[107] Florian Hahne and Robert Ivanek. Visualizing genomic data using gviz and bioconductor. *Methods in molecular biology (Clifton, N.J.)*, 1418:335–351, 2016. ISSN 1940-6029. doi: 10.1007/978-1-4939-3578-9_16. URL http://www.ncbi.nlm.nih.gov/pubmed/27008022.

[108] Svetlana Lebedeva, Marvin Jens, Kathrin Theil, Björn Schwanhäusser, Matthias Selbach, Markus Landthaler, and Nikolaus Rajewsky. Transcriptome-wide analysis of regulatory interactions of the RNA-binding protein HuR. *Molecular Cell*, 43(3): 340–352, Aug 2011. ISSN 1097-4164. doi: 10.1016/j.molcel.2011.06.008. URL http://www.ncbi.nlm.nih.gov/pubmed/21723171.

[109] R. Singh, J. Valcárcel, and M. R. Green. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science (New York, N.Y.)*, 268(5214):1173–1176, May 1995. ISSN 0036-8075. URL http://www.ncbi.nlm.nih.gov/pubmed/7761834.

[110] Francesco Itri, Daria M. Monti, Bartolomeo Della Ventura, Roberto Vinciguerra, Marco Chino, Felice Gesuele, Angelina Lombardi, Raffaele Velotta, Carlo Altucci, Leila Birolo, Renata Piccoli, and Angela Arciello. Femtosecond UV-laser pulses to unveil protein-protein interactions in living cells. *Cellular and Molecular Life Sciences : CMLS*, 73(3):637–648, Feb 2016. ISSN 1420-9071. doi: 10.1007/s00018-015-2015-y. URL http://www.ncbi.nlm.nih.gov/pubmed/26265182.

[111] Francesco Itri, Daria Maria Monti, Marco Chino, Roberto Vinciguerra, Carlo Altucci, Angela Lombardi, Renata Piccoli, Leila Birolo, and Angela Arciello. Identification of novel direct protein-protein interactions by irradiating living cells with femtosecond UV laser pulses. *Biochemical and Biophysical Research Communications*, 492(1): 67–73, Oct 2017. ISSN 1090-2104. doi: 10.1016/j.bbrc.2017.08.037. URL http://www.ncbi.nlm.nih.gov/pubmed/28807828.

[112] Adam Freund, Franklin L. Zhong, Andrew S. Venteicher, Zhaojing Meng, Timothy D. Veenstra, Judith Frydman, and Steven E. Artandi. Proteostatic control of

telomerase function through tric-mediated folding of tcab1. *Cell*, 159(6):1389–1403, Dec 2014. ISSN 1097-4172. doi: 10.1016/j.cell.2014.10.059. URL http://www.ncbi.nlm.nih.gov/pubmed/25467444.

[113] Tamar Geiger, Anja Wehner, Christoph Schaab, Juergen Cox, and Matthias Mann. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular & cellular proteomics : MCP*, 11 (3):M111.014050, Mar 2012. ISSN 1535-9484. doi: 10.1074/mcp.M111.014050. URL http://www.ncbi.nlm.nih.gov/pubmed/22278370.

[114] Arnaud Hubstenberger, Maïté Courel, Marianne Bénard, Sylvie Souquere, Michèle Ernoult-Lange, Racha Chouaib, Zhou Yi, Jean-Baptiste Morlot, Annie Munier, Magali Fradet, Maëlle Daunesse, Edouard Bertrand, Gérard Pierron, Julien Mozziconacci, Michel Kress, and Dominique Weil. P-body purification reveals the condensation of repressed mrna regulons. *Molecular cell*, 68(1):144–157.e5, Oct 2017. ISSN 1097-4164. doi: 10.1016/j.molcel.2017.09.003. URL http://www.ncbi.nlm.nih.gov/pubmed/28965817.

[115] Kristopher W. Brannan, Wenhao Jin, Stephanie C. Huelga, Charles A. S. Banks, Joshua M. Gilmore, Laurence Florens, Michael P. Washburn, Eric L. Van Nostrand, Gabriel A. Pratt, Marie K. Schwinn, Danette L. Daniels, and Gene W. Yeo. Sonar discovers rna-binding proteins from analysis of large-scale protein-protein interactomes. *Molecular cell*, 64(2):282–293, Oct 2016. ISSN 1097-4164. doi: 10.1016/j.molcel.2016.09.003. URL http://www.ncbi.nlm.nih.gov/pubmed/27720645.

[116] Wenxing Jin, Yi Wang, Chao-Pei Liu, Na Yang, Mingliang Jin, Yao Cong, Mingzhu Wang, and Rui-Ming Xu. Structural basis for snrna recognition by the double-wd40 repeat domain of gemin5. *Genes & development*, 30(21):2391–2403, Nov 2016. ISSN 1549-5477. doi: 10.1101/gad.291377.116. URL http://www.ncbi.nlm.nih.gov/pubmed/27881601.

[117] Marcin Wegrecki, Wegrecki Marcin, Jose Luis Neira, and Jeronimo Bravo. The carboxy-terminal domain of erb1 is a seven-bladed ß-propeller that binds rna. *PloS one*, 10(4):e0123463, Apr 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0123463. URL http://www.ncbi.nlm.nih.gov/pubmed/25880847.

[118] Qiang Wang, Kathryn Hobbs, Bert Lynn, and Brian C. Rymond. The clf1p splicing factor promotes spliceosome assembly through n-terminal tetratricopeptide repeat contacts. *The Journal of biological chemistry*, 278(10):7875–7883, Mar 2003. ISSN 0021-9258. doi: 10.1074/jbc.M210839200. URL http://www.ncbi.nlm.nih.gov/pubmed/12509417.

[119] Felix Halbach, Peter Reichelt, Michaela Rode, and Elena Conti. The yeast ski complex: crystal structure and rna channeling to the exosome complex. *Cell*,

154(4):814–826, Aug 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.07.017. URL http://www.ncbi.nlm.nih.gov/pubmed/23953113.

[120] Delphine Rispal, Julien Henri, Herman van Tilbeurgh, Marc Graille, and Bertrand Séraphin. Structural and functional analysis of nro1/ett1: a protein involved in translation termination in s. cerevisiae and in o2-mediated gene control in s. pombe. *RNA (New York, N.Y.)*, 17(7):1213–1224, Jul 2011. ISSN 1469-9001. doi: 10.1261/rna.2697111. URL http://www.ncbi.nlm.nih.gov/pubmed/21610214.

[121] Miha Milek, Koshi Imami, Neelanjan Mukherjee, Francesca De Bortoli, Ulrike Zinnall, Orsalia Hazapis, Christian Trahan, Marlene Oeffinger, Florian Heyd, Uwe Ohler, Matthias Selbach, and Markus Landthaler. Ddx54 regulates transcriptome dynamics during dna damage response. *Genome research*, 27(8):1344–1359, Aug 2017. ISSN 1549-5469. doi: 10.1101/gr.218438.116. URL http://www.ncbi.nlm.nih.gov/pubmed/28596291.

[122] Stanislaw A. Gorski, Jörg Vogel, and Jennifer A. Doudna. Rna-based recognition and targeting: sowing the seeds of specificity. *Nature reviews. Molecular cell biology*, 18(4):215–228, Apr 2017. ISSN 1471-0080. doi: 10.1038/nrm.2016.174. URL http://www.ncbi.nlm.nih.gov/pubmed/28196981.

[123] Charlotte Michaux, Erik Holmqvist, Erin Vasicek, Malvika Sharan, Lars Barquist, Alexander J. Westermann, John S. Gunn, and Jörg Vogel. RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. *Proceedings of the National Academy of Sciences*, 114(26):201620772, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1620772114. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1620772114.

[124] Ana M. Matia-González, Emma E. Laing, and André P. Gerber. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nature Structural & Molecular Biology*, 22(12):1027–1033, Dec 2015. ISSN 1545-9985. doi: 10.1038/nsmb.3128. URL http://www.ncbi.nlm.nih.gov/pubmed/26595419.

[125] James A. Tsatsaronis, Sandra Franch-Arroyo, Ulrike Resch, and Emmanuelle Charpentier. Extracellular vesicle rna: A universal mediator of microbial communication? *Trends in microbiology*, 26(5):401–410, May 2018. ISSN 1878-4380. doi: 10.1016/j.tim.2018.02.009. URL http://www.ncbi.nlm.nih.gov/pubmed/29548832.

[126] Caroline Tawk, Malvika Sharan, Ana Eulalio, and Jörg Vogel. A systematic analysis of the rna-targeting potential of secreted bacterial effector proteins. *Scientific reports*, 7(1):9328, Aug 2017. ISSN 2045-2322. doi: 10.1038/s41598-017-09527-0. URL http://www.ncbi.nlm.nih.gov/pubmed/28839189.

[127] Jihwan Hwang and Masayori Inouye. A bacterial gap-like protein, yihi, regulating the gtpase of der, an essential gtp-binding protein in escherichia coli. *Journal of molecular biology*, 399(5):759–772, Jun 2010. ISSN 1089-8638. doi: 10.1016/j.jmb.2010.04.040. URL http://www.ncbi.nlm.nih.gov/pubmed/20434458.

[128] FS Jacobson, RW Morgan, MF Christman, and Ames BN. An alkyl hydroperoxide reductase from salmonella typhimurium involved in the defense of dna against oxidative damage. purification and properties. *J Biol Chem*, 264:1488–96, 1988.

[129] Daoguo Zhou, Mark S Mooseker, and Jorge E Galán. Role of the s. typhimurium actin-binding protein sipa in bacterial internalization. *Science*, 283:2092–2095, 1999. doi: 10.1126/science.283.5410.2092.

[130] EJ McGhie, RD Hayward, and V Koronakis. Control of actin turnover by a salmonella invasion protein. *Molecular Cell*, 13:497–510, 2004. URL https://doi.org/10.1016/S1097-2765(04)00053-X.

[131] Chongsheng He, Simone Sidoli, Robert Warneford-Thomson, Deirdre C. Tatomer, Jeremy E. Wilusz, Benjamin A. Garcia, and Roberto Bonasio. High-Resolution Mapping of RNA-Binding Regions in the Nuclear Proteome of Embryonic Stem Cells. *Molecular Cell*, 64(2):416–430, Oct 2016. ISSN 1097-4164. doi: 10.1016/j.molcel.2016.09.034. URL http://www.ncbi.nlm.nih.gov/pubmed/27768875.

[132] Tom Smith, Eneko Villanueva, Rayner M L Queiroz, Charlotte S Dawson, Mohamed Elzek, Erika C Urdaneta, Anne E Willis, Benedikt M Beckmann, Jeroen Krijgsveld, and Kathryn S Lilley. ScienceDirect Organic phase separation opens up new opportunities to interrogate the RNA-binding proteome. *Current Opinion in Chemical Biology*, 54:70–75, 2020. ISSN 1367-5931. doi: 10.1016/j.cbpa.2020.01.009. URL https://doi.org/10.1016/j.cbpa.2020.01.009.

[133] Luke a Yates, Chris J Norbury, and Robert J C Gilbert. The long and short of microRNA. *Cell*, 153(3):516–9, apr 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.04.003. URL http://www.ncbi.nlm.nih.gov/pubmed/23622238.

[134] Gunter Meister. Argonaute proteins: functional insights and emerging roles. *Nature Reviews. Genetics*, 14(7):447–459, Jul 2013. ISSN 1471-0064. doi: 10.1038/nrg3462. URL http://www.ncbi.nlm.nih.gov/pubmed/23732335.

[135] Jonathan Houseley and David Tollervey. The many pathways of rna degradation. *Cell*, 136(4):763–776, Feb 2009. ISSN 1097-4172. doi: 10.1016/j.cell.2009.01.019. URL http://www.ncbi.nlm.nih.gov/pubmed/19239894.

[136] Esben Lorentzen, Jérôme Basquin, and Elena Conti. Structural organization of the rna-degrading exosome. *Current opinion in structural biology*, 18(6):709–713, Dec

2008. ISSN 1879-033X. doi: 10.1016/j.sbi.2008.10.004. URL http://www.ncbi.nlm.nih.gov/pubmed/18955140.

[137] Elizabeth V. Wasmuth, Kurt Januszyk, and Christopher D. Lima. Structure of an rrp6-rna exosome complex bound to poly(a) rna. *Nature*, 511(7510):435–439, Jul 2014. ISSN 1476-4687. doi: 10.1038/nature13406. URL http://www.ncbi.nlm.nih.gov/pubmed/25043052.

[138] Eva Kowalinski, Alexander Kögel, Judith Ebert, Peter Reichelt, Elisabeth Stegmann, Bianca Habermann, and Elena Conti. Structure of a cytoplasmic 11-subunit rna exosome complex. *Molecular cell*, 63(1):125–134, Jul 2016. ISSN 1097-4164. doi: 10.1016/j.molcel.2016.05.028. URL http://www.ncbi.nlm.nih.gov/pubmed/27345150.

[139] John C. Zinder, Elizabeth V. Wasmuth, and Christopher D. Lima. Nuclear rna exosome at 3.1?å reveals substrate specificities, rna paths, and allosteric inhibition of rrp44/dis3. *Molecular cell*, 64(4):734–745, Nov 2016. ISSN 1097-4164. doi: 10.1016/j.molcel.2016.09.038. URL http://www.ncbi.nlm.nih.gov/pubmed/27818140.

[140] Jan Michael Schuller, Sebastian Falk, Lisa Fromm, Ed Hurt, and Elena Conti. Structure of the nuclear exosome captured on a maturing preribosome. *Science (New York, N.Y.)*, Mar 2018. ISSN 1095-9203. doi: 10.1126/science.aar5428. URL http://www.ncbi.nlm.nih.gov/pubmed/29519915.

[141] Clémentine Delan-Forino, Claudia Schneider, and David Tollervey. Rna substrate length as an indicator of exosome interactions. *Wellcome open research*, 2:34, Jul 2017. ISSN 2398-502X. doi: 10.12688/wellcomeopenres.10724.2. URL http://www.ncbi.nlm.nih.gov/pubmed/28748221.

[142] Claudio Asencio, Aindrila Chatterjee, and Matthias W Hentze. Silica-based solid-phase extraction of cross-linked nucleic acid–bound proteins. *Life Science Alliance*, 1(3):e201800088, jun 2018. ISSN 2575-1077. doi: 10.26508/lsa.201800088. URL http://www.life-science-alliance.org/lookup/doi/10.26508/lsa.201800088.

[143] Jenna Morgan Lang, Aaron E. Darling, and Jonathan A. Eisen. Phylogeny of Bacterial and Archaeal Genomes Using Conserved Genes: Supertrees and Supermatrices. *PLoS ONE*, 8(4), 2013. ISSN 19326203. doi: 10.1371/journal.pone.0062510.

Appendix

4.5 Published Articles

4.5.1 Purification of Cross-linked RNA-Protein Complexes by Phenol-Toluol Extraction

Erika C Urdaneta, Carlos H Vieira-Vieira, Timon Hick, Hans-Herman Wessels, Davide Figini, Rebecca Moschall, Jan Medenbach, Uwe Ohler, Sander Granneman, Matthias Selbach, Benedikt M Beckmann. NATURE COMMUNICATIONS | 2019 | 10:990 | https://doi.org/10.1038/s41467-019-08942-3

Abstract

Recent methodological advances allowed the identification of an increasing number of RNA-binding proteins (RBPs) and their RNA-binding sites. Most of those methods rely, however, on capturing proteins associated to polyadenylated RNAs which neglects RBPs bound to non-adenylate RNA classes (tRNA, rRNA, pre-mRNA) as well as the vast majority of species that lack poly-A tails in their mRNAs (including all archea and bacteria). We have developed the Phenol Toluol extraction (PTex) protocol that does not rely on a specific RNA sequence or motif for isolation of cross-linked ribonucleoproteins (RNPs), but rather purifies them based entirely on their physicochemical properties. PTex captures RBPs that bind to RNA as short as 30 nt, RNPs directly from animal tissue and can be used to simplify complex workflows such as PAR-CLIP. Finally, we provide a global RNA-bound proteome of human HEK293 cells and the bacterium *Salmonella* Typhimurium.

4.5.2 TriPepSVM: de novo prediction of RNA-binding proteins based on short amino acid motifs

Annkatrin Bressin*, Roman Schulte-Sasse*, Davide Figini*, **Erika C Urdaneta**, Benedikt M Beckmann and Annalisa Marsico. NUCLEIC ACIDS RESEARCH | (2019) | https://doi.org/10.1093/nar/gkz203

**Joint First Authors.*
ECU contributed to Figure 5B,C.

Abstract

In recent years hundreds of novel RNA-binding proteins (RBPs) have been identified leading to the discovery of novel RNA-binding domains (RBDs). Furthermore, unstructured or disordered low-complexity regions of RBPs have been identified to play an important role in interactions with nucleic acids. However, these advances in understanding RBPs are limited mainly to eukaryotic species and we only have limited tools to faithfully predict RNA-binders from bacteria. Here, we describe a support vector machine (SVM)-based method, called TriPepSVM, for the classification of RNA-binding proteins and non-RBPs. TriPepSVM applies string kernels to directly handle protein sequences using tri-peptide frequencies. Testing the method in human and bacteria, we find that several RBP-enriched tri-peptides occur more often in structurally disordered regions of RBPs. TriPepSVM outperforms existing applications, which consider classical structural features of RNA-binding or homology, in the task of RBP prediction in both human and bacteria. Finally, we predict 66 novel RBPs in Salmonella Typhimurium and validate the bacterial proteins ClpX, DnaJ and UbiG to associate with RNA in vivo.

4.5.3 Fast and unbiased purification of RNA-protein complexes after UV cross-linking

Erika C Urdaneta and Benedikt M Beckmann | METHODS | In Press | Available online 3 October 2019 | https://doi.org/10.1016/j.ymeth.2019.09.013

Abstract

Post-transcriptional regulation of gene expression in cells is facilitated by formation of RNA-protein complexes (RNPs). While many methods to study eukaryotic (m)RNPs rely on purification of polyadenylated RNA, other important regulatory RNA classes or bacterial mRNA could not be investigated at the same depth. To overcome this limitation, we developed Phenol Toluol extraction (PTex), a novel and unbiased method for the purification of UV cross-linked RNPs in living cells. PTex is a fast (2–3 h) and simple protocol. The purification principle is solely based on physicochemical properties of cross-linked RNPs, enabling us to interrogate RNA-protein interactions system-wide and beyond poly(A) RNA from a variety of species and source material. Here, we are presenting an introduction of the underlying separation principles and give a detailed discussion of the individual steps as well as incorporation of PTex in high-throughput pipelines.

4.5.4 Organic phase separation opens up new opportunities to interrogate the RNA-binding proteome

Tom Smith, Eneko Villanueva, Rayner M. L. Queiroz, Charlotte S. Dawson, Mohamed Elzek, **Erika C. Urdaneta**, Anne E. Willis, Benedikt M. Beckmann, Jeroen Krijgsveld and Kathryn S. Lilley. CURRENT OPINION IN CHEMICAL BIOLOGY | (2020) 54:70–75 | https://doi.org/10.1016/j.cbpa.2020.01.009

Abstract

Protein–RNA interactions regulate all aspects of RNA meta- bolism and are crucial to the function of catalytic ribonucleo-proteins. Until recently, the available technologies to capture RNA-bound proteins have been biased toward poly(A) RNA- binding proteins (RBPs) or involve molecular labeling, limiting their application. With the advent of organic–aqueous phase separation–based methods, we now have technologies that efficiently enrich the complete suite of RBPs and enable quantification of RBP

dynamics. These flexible approaches to study RBPs and their bound RNA open up new research avenues for systems-level interrogation of protein–RNA interactions.