Abstract
Nous avons étudié les aspects structuraux et fonctionnels de deux poly(U) polymérases (PUPs) afin de mieux comprendre le fonctionnement de ces enzymes: l'enzyme Cid1 PUP de Schizosaccharomyces pombe et l'enzyme RET1 de Trypanosome brucei. Ces protéines sont impliquées dans la polyuridylation de plusieurs types d'ARNs comme les ARNms ou les ARNgS. Nous avons déterminé la structure atomique par cristallographie aux rayons X de l'enzyme Cid1 liée à son substrat UTP, un UTP non hydrolysable et un pseudo-produit ApU qui correspondrait à la première étape de l'uridylation d'un ARNm polyadenylé. Récemment, nous avons obtenu des cristaux contenant un fragment fonctionnel de la protéine RET1 et la résolution de la structure est la prochaine étape. Nos études fournissent une vision atomique de la sélectivité envers l'UTP ainsi que la reconnaissance de l'ARN par les enzymes impliquées dans la polyuridylation en détaillant les mécanismes moléculaires de cette famille d'enzymes et de leurs complexes.

Reference
MUNOZ TELLO, Paola Andréa. Structural and functional studies of poly(U) polymerases

URN : urn:nbn:ch:unige-408396
DOI : 10.13097/archive-ouverte/unige:40839
Structural and Functional Studies of Poly(U) Polymerases

THÈSE

Présentée à la faculté des sciences de l’Université de Genève pour obtenir le grade de Docteur ès Sciences, mention Biologie

par

Paola MUNOZ TELLO
de
Cali (COLOMBIE)

THÈSE N° 4716

GENÈVE
Atelier d’impression Repromail
2014
Doctorat ès sciences
Mention biologie

Thèse de Madame Paola Andrea MUNOZ TELLO

intitulée :

"Structural and Functional Studies of Poly(U) Polymerases"

La Faculté des sciences, sur le préavis de Monsieur S. THORE, professeur assistant et directeur de thèse (Département de biologie moléculaire), Monsieur Th. SCHALCH, professeur assistant (Département de biologie moléculaire), Monsieur A. DZIEMBOWSKI, professeur (Institute of Biochemistry and Biophysics, Department of Genetics and Biotechnology, Warsaw University, Poland) et Madame C. ARRAIANO, professeure (Instituto de Tecnologia Quimica e Biologica, Universidade Nova de Lisboa, Portugal), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 22 septembre 2014

Thèse - 4716 -

Le Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "informations relatives aux thèses de doctorat à l'Université de Genève".
Acknowledgements

First of all, I would like to thank the members of the jury, Professor Cecilia Arraiano, Professor Andrzej Dziembowski and Professor Thomas Schalch, for accepting to evaluate my thesis research.

I would like to express my most sincere gratitude and appreciation to Professor Stéphane Thore for accepting me as a PhD candidate in his group and for his guidance, support, patience, enthusiasm and advices throughout all my time in the lab.

I am grateful to all the members of the Thore lab that have pass through these years for the good atmosphere, advices and support; with a special thanks to Lionel, Fabi and Caroline, without whom the PhD in the lab would have not been the same. Lionel, you have been a great support and made my stay a pleasant and memorable one. Fabi, we have had our ups and downs but you made me grow as a person and scientist and I really appreciate that. Caroline, our discussions about Science and life are memorable and thanks for teaching me that radioactivity is not as scary as it seems although you should not underestimate its danger. Thank you for all the laughs, the patience, the cries, the really good advices and helpful discussions we have had during all these years.

I would also like to thank all the members of my PhD TAC committee, Professor Thanos Halazonitis, Professor Robbie Loewith and Professor Angela Kramer for the helpful discussions and advices.

I would like to extend my special thanks to the past and present members of the Molecular Biology Department, the Faculty of Biology of the University of Geneva and the PhD International Program for sharing information, equipment, consumables and advices during all my PhD; it has been a pleasure to be able to exchange ideas with all of you.

Finally, I would like to thank all my family and friends from Cali, Toulouse and all around the world, without whose love, affection and encouragement, this work would not have been possible; you have been an invaluable support during this time for me. The last paragraph of acknowledgements will be written in Spanish especially for my family.
Este es el agradecimiento mas especial e importante pues no fue fácil el camino para llegar hasta aquí para nuestra familia. Agradezco inmensamente a mis padres Gloria María Tello y José Omar Muñoz, mi hermano Andrés Muñoz, mi tía Elizabeth Tello y Dios por darme la oportunidad de vivir mis sueños. La verdad es que sin ustedes nunca habría podido llegar tan lejos ni soportar tantos años lejos de la familia (al otro lado del charco…) en busca de un mejor futuro. Yo se que fue un sacrificio para todos nosotros que apenas ahora está dando sus frutos. Pero no solo ellos sino toda mi familia porque siempre creyeron en mi y en lo lejos que podría llegar como científica. Gracias por todos los rosarios, plegarias, rezos y velitas puestas en mi pidiéndole a Dios lo mejor para mi futuro. Todos estos años fueron llenos de esfuerzo tanto de su parte como de la mía y valieron mas que la pena pues gracias a ello, llegué aquí donde estoy ahora con los pies en alto y el mayor reconocimiento de estudios: un doctorado.

Mami, no estudie Medicina como tu querías, pero igual termine siendo Doctora.

Los quiero y adoro muchísimo!
# Table of Contents

ACKNOWLEDGEMENTS  

TABLE OF CONTENTS  

1 ABSTRACT  

1.1 FRENCH VERSION  

1.2 ENGLISH VERSION  

2 INTRODUCTION  

2.1 RNA LIFE  

2.1.1 TYPES OF RNAs  

2.2 RNA PROCESSING AND MODIFICATION  

2.3 3’-END ADDITION OF NON-TEMPLATED NUCLEOTIDES  

2.3.1 POLYADENYLATION  

2.3.2 POLYURIDYLATION  

2.4 THE NON-CANONICAL RIBONUCLEOTIDYL TRANSFERASES FAMILY  

2.5 SUBSTRATES OF POLYURIDYLATION IN THE DIFFERENT CELL COMPARTMENTS  

2.5.1 POLYURIDYLATION IN THE NUCLEUS  

2.5.2 POLYURIDYLATION IN THE ORGANELLES (MITOCHONDRIA)  

2.5.3 POLYURIDYLATION IN THE CYTOPLASM  

2.6 KNOWN FUNCTIONS OF POLYURIDYLATION  

2.6.1 IN THE NUCLEUS  

2.6.2 IN THE ORGANELLES  

2.6.3 IN THE CYTOPLASM  

3 GOALS OF THE THESIS WORK  

3.1 PROJECT I: FUNCTIONAL AND STRUCTURAL STUDIES OF CID1 POLY(U) POLYMERASE FROM S. POMBE  

3.2 PROJECT II: STRUCTURAL STUDIES OF MITOCHONDRIAL RNA PROCESSING COMPLEXES IN T. BRUCEI  

4 RESULTS  

4.1 [Sections related to results]  

4.2 [Additional sections related to results]  

5 [Any additional sections or tables related to results]
4.1 FUNCTIONAL AND STRUCTURAL STUDIES OF CID1 POLY(U) POLYMERASE FROM S. POMBE 41
4.1.1 Cid Overexpression and Purification 41
4.1.2 CID1 Crystallization 44
4.1.3 Data Collection and Phasing 45
4.1.4 Publication I: "FUNCTIONAL IMPLICATIONS FROM THE CID1 POLY(U) POLYMERASE CRYSTAL STRUCTURE" 50
4.1.5 CID1 Substrate/RNA Complex Crystallization 66
4.1.6 Publication II: "A CRITICAL SWITCH IN THE ENZYMATIC PROPERTIES OF CID1 PROTEIN DECIPHERED FROM ITS PRODUCT-BOUND CRYSTAL STRUCTURE" 71
4.1.7 MATERIALS AND METHODS 85
4.2 STRUCTURAL STUDIES OF MITOCHONDRIAL RNA PROCESSING COMPLEXES IN T. BRUCEI 88
4.2.1 gRNA Maturation Complexes or RET1 TUTase DSS1 Exonuclease Complex (RDS complex) 88
4.2.2 Polyadenylation/Uridylation Complex 99
4.2.3 MATERIALS AND METHODS 101

5 GENERAL DISCUSSION 107
5.1 STRUCTURAL DETERMINANTS FOR CID1 POLY(U) POLYMERASE UTP SELECTIVITY AND RNA RECOGNITION 108
5.2 RDS COMPLEX STRUCTURAL AND FUNCTIONAL STUDIES: LIGHT INTO A TRYPANOSOMAL TUTASE, RET1 113

6 ABBREVIATIONS 119

7 REFERENCES 121
Abstract
1.1 French Version

L’acide ribonucléique, plus connu sous le nom d’ARN, est une molécule très versatile et instable qui peut subir une dégradation rapide et spontanée. Afin d’éviter par exemple une dégradation prématurée des ARNs, celui-ci est soumis à des modifications post-transcriptionnelles importantes. En réalité, la très grande majorité des processus en lien avec l’ARN nécessite de telles modifications comme l’export vers le cytoplasme ou le bon fonctionnement de la traduction (Muthukrishnan et al. 1975; Munroe and Jacobson 1990; Gallie 1991; Caponigro and Parker 1995; Meyer et al. 2004; Alexandrov et al. 2006). La dégradation des ARNs est un mécanisme important dans la régulation de l’expression génique qui élimine, par exemple, les ARN messagers (ARNms) transcrits de façon erronés ainsi que les ARNms dont la cellule n’a plus besoin. Chez les eucaryotes, la dégradation des ARNs commence par la déadenylation suivi par l’enlèvement de la coiffe en 5’ et finalement la digestion du messager par plusieurs exonucléases (Decker and Parker 1993). Récemment, l’influence majeure de l’uridylation en 3’ comme étape régulatrice dans certaines voies de dégradation des ARNs a poussé les chercheurs à se focaliser sur l’action des poly(U) polymérasases (PUP), les enzymes responsable de cette modification. L’uridylation des ARN mitochondriaux chez les trypanosomes par les terminal uridyl transférases (TUTase), un groupe de PUP, est critique pour la biogénèse des ARNs guide (ARNg) nécessaires au processus d’édition des ARNs ou « RNA editing » ainsi que pour la bonne traduction des ARNms qui nécessite l’ajout de longues queues polyadenine et polyuridine. Ces nombreuses observations soulignent l’importance de la polyuridylation dans ces espèces. Chez les eucaryotes, la polyuridylation de différents types d’ARNs dans les différents compartiments cellulaires est extrêmement importante pour la régulation de l’expression génique mettant en évidence l’expansion de ce phénomène (Blum et al. 1990; Blum and Simpson 1990; Li et al. 2005; Yu et al. 2005; Horwich et al. 2007; Mullen and Marzluff 2008; Kurth and Mochizuki 2009; Rissland and Norbury 2009; Ameres et al. 2010; Aphasizheva and Aphasizhev 2010; Kamminga et al. 2010; Schmidt et al. 2011). La protéine Cid1 est le premier membre des PUP découvert chez les eucaryotes supérieures et montre une préférence pour l’utilisation in vitro et in vivo de l’UTP par rapport à l’ATP, un phénomène actuellement pas complètement caractérisé (Rissland and Norbury 2009).
Durant la période couverte par cette thèse, nous avons étudié les aspects structuraux et fonctionnels de deux PUPs afin de mieux comprendre le fonctionnement de ces enzymes: l’enzyme Cid1 PUP de *Schizosaccharomyces pombe* et l’enzyme RET1 (RNA editing TUTase 1) de *Trypanosome brucei*. Ces protéines sont toutes les deux impliquées dans la polyuridylation de plusieurs types d’ARNs comme les ARNms ou les ARNg.

Nous avons déterminé la structure atomique par cristallographie aux rayons X de l’enzyme Cid1 liée à son substrat UTP, un UTP non hydrolysable et un pseudo-produit ApU qui correspondrait à la première étape de l’uridylation d’un ARNm polyadenylé. Grâce à des analyses mutationnelles, nous proposons un cycle catalytique basé sur le modèle atomique de l’enzyme où des mouvements locaux et globaux sont nécessaires pour la translocation et l’accommodation de l’ARN substrat au sein de l’enzyme. Nos études mettent en évidence les mécanismes moléculaires pour la sélectivité de l’UTP par Cid1 en identifiant les résidus clés à cet effet. De plus, nous avons identifié les bases moléculaires pour la sélection d’un substrat de type ARN (et pas un ADN) ainsi que certaines des propriétés de liaison de Cid1 à l’ARN, caractéristiques essentielles pour la biogénèse des microARNs (miARN), la régulation des ARNm des histones, et plus généralement, la régulation de la dégradation des ARNs.

Mon deuxième projet est focalisé sur les analyses biochimiques et structurales d’un nouveau complexe identifié chez la mitochondrie de *Trypanosome brucei* composé de la TUTase RET1, de l’exonucléase DSS1 et de trois autres protéines sans aucun domaine discernable appelées RDS1, RDS2 et RDS3. Ce complexe est impliqué dans la biogénèse et la polyuridylation des ARNg. Plusieurs expressions et purifications de sous complexes ont été testées afin d’obtenir des complexes pures nécessaires pour des études structurales par rayons X. Récemment, nous avons obtenu des cristaux contenant un fragment fonctionnel de la protéine RET1 et avons pu collecter des données de diffraction à 3.5Å de résolution. La résolution de la structure est la prochaine étape.

Nos études fournissent une vision atomique de la sélectivité envers l’UTP ainsi que la reconnaissance de l’ARN par les enzymes impliquées dans la polyuridylation en détaillant les mécanismes moléculaires de cette famille d’enzymes et de leurs complexes.
1.2 English Version

Ribonucleic acid, most commonly known as RNA, is a very labile molecule prone to rapid and spontaneous decay. In order to avoid this premature degradation, RNA molecules undergo several post-transcriptional modifications required for the proper RNA function such as export to the cytoplasm or translation efficiency (Muthukrishnan et al. 1975; Munroe and Jacobson 1990; Gallie 1991; Caponigro and Parker 1995; Meyer et al. 2004; Alexandrov et al. 2006). RNA degradation is an important mechanism for regulation of gene expression leading to the disposal of erroneously transcribed messenger RNAs (mRNAs) or mRNAs that are not required anymore by the cell. In eukaryotes, general mRNA degradation pathway begins with poly(A) tail removal, followed by decapping, and finally mRNA digestion by exonucleases (Decker and Parker 1993). In recent years, the major influence of 3'-end uridylation as a regulatory step within several RNA degradation pathways has driven attention towards poly(U) polymerase (PUP) enzymes. Uridylation in trypanosomal mitochondria by RET1 and RET2 terminal uridyl transferases (TUTases) is important for guide RNA (gRNA) biogenesis (necessary for the RNA editing process) and for the proper translation of mRNAs via long poly(A)/(U) tail addition. These two examples highlight the importance of this modification in this species. In higher eukaryotes, polyuridylation of several types of RNAs in different cell compartments is crucial for regulation of gene expression supporting the idea of a widespread phenomenon critical for RNA metabolism (Blum et al. 1990; Blum and Simpson 1990; Li et al. 2005; Yu et al. 2005; Horwich et al. 2007; Mullen and Marzluff 2008; Kurth and Mochizuki 2009; Rissland and Norbury 2009; Ameres et al. 2010; Aphasizheva and Aphasizhev 2010; Kamminga et al. 2010; Schmidt et al. 2011). The protein Cid1 is the first member of PUP enzymes found in higher eukaryotes and shows a preference for UTP over ATP both in vivo and in vitro, a feature not fully understood (Rissland and Norbury 2009).

During the course of this thesis, we carried out functional and structural studies of two PUP enzymes to reveal novel aspects of poly(U) polymerases: the cytoplasmic Cid1 PUP from *Schizosaccharomyces pombe* and the mitochondrial RET1 from *Trypanosome brucei*. These proteins are both involved in the polyuridylation of coding and non-coding RNAs.
We have determined the crystal structures of Cid1 protein bound to its substrate UTP, a non-hydrolysable form of UTP and its pseudo-product ApU. Supplemented by point mutations, our atomic models are used to propose a catalytic cycle where local and global movements of the enzyme are needed for RNA accommodation and translocation. Our study provides molecular insights on Cid1’s UTP selectivity by identifying residues critical for UTP recognition. Furthermore, we highlight the molecular basis for selectivity of RNA (rather than DNA) substrates as well as Cid1 RNA binding properties, a feature with critical implications for miRNAs, histone mRNAs and, more generally, cellular RNA degradation.

As a second project, we focused on biochemical and structural analyses of a newly identified complex in trypanosomal mitochondria composed of RET1 TUTase, DSS1 exonuclease and the proteins RDS1, RDS2, and RDS3, which have a currently unknown function. This complex is involved in guide RNA biogenesis and polyuridylation. Several expressions and purifications of sub-complex combinations were attempted towards crystallization and structural studies. Recently, we crystallized a functional fragment of the RET1 protein and collected diffraction data up to 3.5Å resolution.

Globally, our studies provide atomic details into polyuridylation enzyme properties like UTP selectivity or RNA recognition, bringing new molecular insights regarding the function of this family of enzymes.
Introduction
1.3 RNA Life

The three essential molecules for known terrestrial life are deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and proteins. DNA is the most stable molecule in a cell and is composed of four bases, i.e. adenine (A), guanine (G), cytosine (C) or thymidine (T); a sugar ring (deoxyribose) and a phosphate moiety (Fig. 1). Deoxyribonucleotides store genetic information from one generation to the next. In 1953, James Watson and Francis Crick published the first model of the DNA double helix structure, leading to a revolution in genetics and biology (Watson and Crick 1953; Watson and Crick 1953). The DNA of an organism can be huge and tightly packed inside the nucleus. But there are some variations of DNA size and packing architecture between organisms. In order to pass the genetic information to the translation machinery in the cytoplasm, the DNA has to be transcribed into RNA. In other words, translation of the DNA sequence into its corresponding amino acid sequence to form protein polymer, the essential player for the catalytic and structural functions of a cell, is linked through the RNA molecule. Conversely to DNA, several copies of a RNA molecule are present in the cell and are mainly composed of single stranded nucleic acids containing 4 major bases: A, C, G and uracil (U), a ribose ring and a phosphate moiety (Fig. 1). RNA molecules can also form double stranded structures through self-complementary sequences, which can be crucial for the stability and regulation of this molecule (Buratti and Baralle 2004; Cooper et al. 2009; Cruz and Westhof 2009; Sharp 2009).

RNA is produced by the RNA polymerases during a process known as transcription. Eukaryotic transcription is initiated by these enzymes together with the necessary transcription elongation factors. The DNA template is read by the RNA polymerases that simultaneously synthesize a RNA copy of the transcription unit (Hahn 2004). When the end of the transcription unit is reached, the RNA polymerase falls off the DNA template and the newly formed strand of RNA is released (Richard and Manley 2009). In eukaryotes, there are three RNA polymerases named RNA polymerase I, II and III. Each of them is responsible for the transcription of a specific set of RNAs and able to recognize a specific set of promoters. They were identified through their sensitivity to α-amanitin (Roeder and Rutter 1969; Kedinger et al. 1970). RNA polymerase I is insensitive to this drug and localizes in the nucleolus while RNA polymerase II and III are found in the nucleoplasm and are strongly inhibited or inhibited at high concentrations of α-amanitin, respectively.
Figure 1: Nucleotide chemical structure. In DNA, a nucleotide (nt) is composed of a base (Cytosine, Thymine, Adenine or Guanine), a deoxyribose and a phosphate moiety. In RNA, the nucleotide is composed of a base (Cytosine, Uracil, Adenine or Guanine), a ribose and a phosphate moiety (the figure was created by the use of ChemDraw software).

Unlike DNA, RNA molecules are very labile and undergo fast and spontaneous decay due to the presence of the 2'-hydroxyl group in the sugar ring. In order to be functional and avoid degradation, the large variety of produced RNA molecules are either bound by proteins or need to carry specific chemical modifications. These modifications play a critical role in the organization and regulation of gene expression in the cell due to their essential influence on the steady-state levels of RNA.

1.3.1 Types of RNAs

In all prokaryotic and eukaryotic organisms there are at least three primary types of RNAs with different functions during protein synthesis: messenger RNAs (mRNAs), ribosomal RNAs (rRNAs) and transfert RNAs (tRNAs; Fig. 2). There are, additionally, several other types of RNAs playing an important role in the regulation of gene expression in different organisms such as small nucleolar RNAs (snoRNAs), small nuclear RNAs (snRNAs), microRNAs (miRNAs),...
small interfering RNAs (siRNAs), long non-coding RNA (lincRNAs) and PIWI-interacting RNAs (piRNAs) for the most represented. The different types of RNAs and their main functions are briefly described in the following sections (sections 2.1.1.1 to 2.1.1.4).

![Figure 2: Involvement of the three primary types of RNA in the protein synthesis. An mRNA molecule is translated into protein by the action of tRNAs together with the ribosome. The major components of the ribosome are the rRNAs. Adapted from (A.J.E. Griffiths et al., 1993).](image)

**1.3.1.1 mRNAs**
mRNAs are the only RNAs encoding for proteins and are transcribed by RNA polymerase II. They represent ~5% of the total RNAs in a cell and are the most heterogeneous in size and sequence. As key mediators between the genetic information encoded by the DNA and the amino acids sequence of the protein, mRNA levels and localizations are tightly regulated.

**1.3.1.2 rRNAs**
rRNAs are transcribed by RNA polymerase I and III. These RNAs along with ribosomal proteins compose the translation machinery of a cell, the ribosome. There are three species in prokaryotes the 5S and 23S, which associate with the 50S large ribosomal subunit and the 16S rRNA, which is found in the 30S ribosomal small subunit. In eukaryotes, the 60S large subunit is composed of three RNA molecules: the 5S, the 5.8S and the 28S, while the 40S small subunit contains the 18S rRNA specie. The rRNAs are the major structural component
within ribosomes and form the peptidyl transferase center, whose activity is responsible for peptide bond formation during translation (Simonovic and Steitz 2009).

### 1.3.1.3 tRNAs

Ternary complexes are transcribed by RNA polymerase III and there is at least 20 tRNA species corresponding to the 20 amino acids necessary for protein production. They actively participate in the synthesis of proteins as they are used to decode the sequence of the mRNA. They carry an amino acid on the 3’-CCA end and contact the mRNA using their anticodon loop, complementary to three nucleotides of the mRNA sequence (Sprinzl and Cramer 1979; Green and Noller 1997). Once in the ribosome, their anticodon region forms a base pair with the mRNA codon, while their 3’-CCA end positions the amino acid next to the newly synthesized polypeptide. They are the translators of the cell.

### 1.3.1.4 Other Types of RNAs

A large number of novel RNA species have been described in recent years, underscoring the complexity of functionality of RNAs. Indeed, many transcribed RNAs do not encode for a polypeptide chain but are nevertheless crucial for gene expression regulation and transcriptome metabolism. Some of these other types of RNAs are described in the following sections (sections 2.1.1.4.1 to 2.1.1.4.5)

#### 1.3.1.4.1 snRNAs

SnRNAs are found only in eukaryotes and are transcribed by RNA polymerase II or III. These small uridine-rich RNAs are always present as RNA/proteins complexes named small nuclear ribonucleoproteins (snRNPs) and play an essential role in pre-mRNA splicing in the nucleus (Maniatis and Reed 1987; Guthrie and Patterson 1988; Mowry and Steitz 1988). One of them, the U6 snRNA is uridyylated in the nucleus, a modification that is essential for its function (Trippe et al. 1998; Trippe et al. 2006).
1.3.1.4.2 snoRNAs

snoRNAs are transcribed by RNA polymerase II or III. They play a crucial role in RNA biogenesis and guide the methylation and/or pseudouridylation of rRNAs, tRNA and snRNAs. They are often embedded in small nucleolar ribonucleoproteins (snoRNPs), essential for the modification of the targeted RNA (Kiss-Laszlo et al. 1996; Ganot et al. 1997; Maden and Hughes 1997; Lafontaine and Tollervey 1998; Jady and Kiss 2001).

1.3.1.4.3 miRNAs

miRNAs are single stranded RNAs of approximately 22 nt and are crucial for gene regulation. They are generated from imperfect hairpin RNA transcripts and bind to a complementary sequence found usually in the 3’ UTR of an mRNA. The miRNA binding event will most of the time lead to translational repression and decay of the targeted mRNA. However, miRNAs have also been involved in translational activation (Filipowicz et al. 2008) and heterochromatin formation (Kim et al. 2008).

1.3.1.4.4 siRNAs

siRNAs derive from perfect long double stranded RNAs and are about 20 to 25 nt in size. As for miRNAs, they are essential regulators of gene expression and therefore their proper biogenesis and maturation are crucial for the cell (Babiarz et al. 2008; Tam et al. 2008; Watanabe et al. 2008).

1.3.1.4.5 piRNAs

piRNAs are around 24 to 29 nt in size and are specific to germ lines. Transcribed by RNA polymerase II, these RNAs interact with PIWI proteins and are mainly responsible for the silencing of transposable elements in germ lines (Girard et al. 2006; Vagin et al. 2006; Houwing et al. 2007; Kirino and Mourelatos 2007).
1.4 RNA Processing and Modification

RNA processing is a general term used to describe the events necessary for RNA maturation (Abelson 1979). Processing of RNAs can occur within the RNA molecule but also at the 5’- and 3’-ends. These processing events are an important determinant of the biological fate of RNAs, necessary for gene regulation and crucial for the proper function of the RNA.

RNA modifications regroup diverse enzymatic reactions such as pseudouridylation (change of uridine into pseudouridine: Ψ), methylation (addition of a methyl group to a nucleotide), RNA editing (Adenine-to-Inosine (A-to-I), Cytidine-to-Uridine (C-to-U) or uridine insertion/deletion in trypanosomal mitochondrial mRNAs). Such modifications can be critical for the main function of the RNA as well as for the encoded protein as some of them can modify the mRNA translation process (Benne et al. 1986; Shaw et al. 1988; Flomen et al. 2004; Nishikura 2006; Ge and Yu 2013). Indeed, the nucleotide composition of a given RNA determines its chemical and structural properties together with its defined open reading frame in the case of mRNAs. These modifications will either facilitate RNA maturation or target the RNA for degradation. One of the most characterized single nucleotide modifications is RNA editing. In higher eukaryotes, the most important editing event is the A-to-I editing on double stranded RNA (dsRNA) structures catalyzed by the Adenosine Deaminases that Act on RNA (ADARs) and the A-to-I event on tRNAs catalyzed by Adenosine Deaminases that Act on tRNA (ADAT; Fig. 3; Savva et al. 2012). The newly formed inosine nucleotide is recognized as a guanosine by the translation machinery and thus will alter the sequence of the encoded protein (Fig. 3). Additionally, editing reactions can influence the splicing pattern by recruiting splicing factors at the editing site. Editing can also trigger regulatory pathways such as silencing if the editing event happens within the UTR of mRNAs or on non-coding RNAs (Flomen et al. 2004; Nishikura 2006). Another type of editing is the insertion/deletion of uridines within mitochondrial mRNA molecules specific to trypanosomal species (Benne et al. 1986; Shaw et al. 1988). This editing event is crucial for mRNA function since it will restore open reading frames (ORFs), add or delete start or stop codons and correct frame shifts (Benne et al. 1986; Feagin et al. 1988; Shaw et al. 1988). In this case, a complex called the 20S editosome is responsible for the editing of the mitochondrially encoded mRNAs (Pollard et al. 1992; Corell et al. 1996).
Figure 3: A-to-I editing. Adenosine is deaminated by ADARs or ADATs (depending on the RNA target) to form an inosine, which will be recognized as a guanosine by the translation machinery in the cell (the figure was created by the use of ChemDraw software).

In addition to internal nucleotide modification, RNA molecules can also be modified at their extremities and such modifications are essential for proper RNA function. One of the best characterized modifications occurs at the 5'-end and is known as the 5' cap (Shatkin 1976). This event consists of a guanine nucleotide methylated in position 7 connected to the mRNA via an unusual 5' to 5' triphosphate bond catalyzed by the capping enzyme complex (Shatkin 1976; Shuman 1995). This modification is crucial for mRNA maturation, stability and translation efficiency (Merrick 2004; Meyer et al. 2004; Li and Kiledjian 2010). 5' decapping event, catalyzed by the Dcp1-Dcp2 complex in concert with the Lsm1-7 complex, will lead to the decay of such mRNA by Xrn1 exonuclease, further highlighting the critical role of 5'-capping in mRNA stability (Fig. 4; Decker and Parker 1993; Hsu and Stevens 1993; Dunckley and Parker 1999; Lykke-Andersen 2002; van Dijk et al. 2002).

Over 150 RNA modifications have been estimated so far (Czerwoniec et al. 2009). The work presented in this thesis focuses solely on proteins involved in a 3'-end modification, which was underestimated for over 40 years but has critical implications in miRNA biogenesis and more generally, in cellular RNA degradation.
Cytoplasmic mRNA degradation occurs by two general pathways, both of which are initiated by deadenylation. Shortening of the poly(A) tail is carried out by the Pan2/Pan3 complex or the Ccr4/Pop2/Not complex. Following deadenylation, mRNAs can be subjected to 3’ to 5’ degradation by the exosome. mRNAs can also be decapped by the Dcp1/Dcp2 decapping enzyme and then subjected to 5’ to 3’ degradation by Xrn1 or 3’ to 5’ decay by the exosome or DIS3L2.

1.5 3’-End Addition of Non-Templated Nucleotides

RNA 3’-end processing or modification plays an important role in determining their biological fate (Munroe and Jacobson 1990; Meyer et al. 2004; Alexandrov et al. 2006). One major type of modification encountered by mRNAs is the addition of non-templated nucleotides (Munroe and Jacobson 1990; West et al. 2006; Doma and Parker 2007; Martin and Keller 2007; Rissland and Norbury 2009). The functional consequence of this nucleotide addition is essentially to protect newly transcribed mRNAs from degradation. More generally, tail addition to RNAs regulates cellular RNA content by influencing RNA steady-state levels. Nuclear polyadenylation is essential to degrade various classes of non-coding RNAs (ncRNAs) in the nucleus (Deutscher 1990; Keller 1995; Manley 1995; LaCava et al. 2005). However, once in the cytoplasm, RNAs carrying a 3’-poly(A) tail are protected from 3’-to-5’
exonucleases. Polyuridylation is another 3’ modification that involves the addition of uridines at the 3’-end of RNA molecules. This modification is found on various types of RNAs such as mRNAs, small RNAs, miRNAs or gRNAs (Blum et al. 1990; Blum and Simpson 1990; Li et al. 2005; Yu et al. 2005; Horwich et al. 2007; Mullen and Marzluff 2008; Kurth and Mochizuki 2009; Rissland and Norbury 2009; Ameres et al. 2010; Aphasizheva and Aphasizhev 2010; Kamminga et al. 2010; Schmidt et al. 2011). This modification is known to have a major impact in multiple aspects of RNA turnover and metabolism (Blum et al. 1990; Blum and Simpson 1990; Mullen and Marzluff 2008; Rissland and Norbury 2009; Aphasizheva and Aphasizhev 2010; Schmidt et al. 2011).

1.5.1 Polyadenylation

Eukaryotic mRNAs start to be modified during their transcription, where capping and polyadenylation take place at their 5’- and 3’-ends respectively (except for histone mRNAs and some viral mRNAs; Parker and Song 2004). Pre-mRNAs are first cleaved by the cleavage and polyadenylation machinery at the polyadenylation site located near what will become the 3’-end. This cleavage is followed by addition of the poly(A) tail by nuclear poly(A) polymerases (nPAPs). This event will determine the 3’ untranslated region (UTR) of the RNA, which is crucial for the regulation of gene expression processes (Wilkie et al. 2003). Mutations and changes in the length of this region will immediately affect a variety of proceedings such as mRNA stability, mRNA localization and mRNA translation efficiency (van der Velden and Thomas 1999; Conne et al. 2000; Bashirullah et al. 2001; Babendure et al. 2006; Mayr and Bartel 2009). Once the mRNAs are exported to the cytoplasm, they may undergo several additional modifications such as methylation, editing, deadenylation, decapping and polyuridylation, which are again influencing the stability or degradation of the RNA (Benne et al. 1986; Blum et al. 1990; Shyu et al. 1991; Decker and Parker 1993; Seiwert et al. 1996; Dunckley and Parker 1999; Lykke-Andersen 2002; Yu et al. 2005; Mullen and Marzluff 2008; Rissland and Norbury 2009; Kamminga et al. 2010; Schmidt et al. 2011). Polyadenylation regulates RNA degradation, which is one of the most important gene expression mechanisms not only for the removal of mRNAs that should not be translated anymore, but also for the disposal of incorrectly transcribed mRNAs that have escaped the nuclear surveillance mechanisms. General RNA degradation pathway is well conserved
throughout eukaryotes, from yeast to mammals, and has two major directions: the 5′-3′ degradation by Xrn1 exoribonuclease and the 3′-5′ degradation catalyzed by the exosome complex (Fig. 4; for recent review: Schoenberg and Maquat 2012). However, before degrading the mRNA bodies, the cell machinery must first identify the mRNA to be degraded. The cellular cues initiating mRNA degradation are still poorly understood for mRNAs encoding house-keeping genes, while physiological inputs that trigger mRNA decay such as pro-inflammatory responses, heat shock or differentiation are far better characterized (Carballo et al. 1998; Lai et al. 2000). Deadenylation is generally the rate-limiting event in the cytoplasmic mRNA degradation and is catalyzed by the PAN2/PAN3 complex followed by the CCR4/NOT complex (Shyu et al. 1991; Decker and Parker 1993). Once the poly(A) tail has been removed, the Dcp1-Dcp2 decapping complex will withdraw the 7-methylguanylate cap (m7G) from the 5′-end of the mRNA allowing the trimming of this RNA in a 5′-to-3′ manner by Xrn1 exonuclease (Decker and Parker 1993; Hsu and Stevens 1993; Dunckley and Parker 1999; Lykke-Andersen 2002; van Dijk et al. 2002). Following deadenylation, the cytoplasmic exosome complex may cut down deadenylated RNAs as the 3′-5′ mRNA decay pathway (Anderson and Parker 1998; Wang and Kiledjian 2001).

### 1.5.2 Polyuridylation

Recently, another player in the mRNA decay pathways has come into focus: the cytoplasmic non-canonical nucleotidyl transferases (ncNTrs). These enzymes add uridine residues at the 3′-end of RNA molecules, which can be coding RNAs or ncRNA. Even though this modification has been known since the late fifties, its significance had been underestimated (Canellakis 1957; Wilkie and Smellie 1968; Wilkie and Smellie 1968). In the middle of the eighties, the importance of uridylation increased with the discovery and the characterization of the uridine insertion/deletion editing mechanism in the mitochondria of kinetoplastids. The process was subsequently shown to be crucial for translation efficiency of local mRNAs (Benne et al. 1986; Blum et al. 1990; Seiwert et al. 1996). During the last decade, evidences showed that polyuridylation also existed in higher eukaryotes. Norbury and colleagues showed its implication during the S-phase of the cell cycle (Rissland et al. 2007). Further studies demonstrated that polyuridylation was a critical step for the degradation of non-polyadenylated mRNA encoding histone proteins in mammals (Mullen and Marzluff 2008).
This new enzymatic step occurring at the 3’-end of RNAs added another level of complexity to the known mRNA decay pathways further highlighting the importance of regulation for this 3’-end modification (Mullen and Marzluff 2008; Rissland and Norbury 2009; Schmidt et al. 2011). Polyuridylation has also been found to occur on other types of RNA molecules such as miRNAs, siRNAs and piRNAs (Li et al. 2005; Yu et al. 2005; Horwich et al. 2007; Kurth and Mochizuki 2009; Ameres et al. 2010; Kamminga et al. 2010). Finally, studies from the Aphasizhev laboratory on ncNTr family members present in trypanosomal species demonstrated additional roles for these enzymes, not only in the RNA editing process but also during the processing of gRNAs and during mRNA translation in the mitochondria (Aphasizhev et al. 2003; Aphasizheva et al. 2011).

In these following paragraphs, we focus on the latest development about the terminal polyuridylation by non-canonical polymerases, a longtime underestimated 3’-end post-transcriptional modification withstand by various types of RNAs influencing their half-life and functions.

### 1.6 The Non-Canonical Ribonucleotidyl Transferases Family

Enzymes performing terminal polyuridylation belong to the Polymerase β-like nucleotidyl transferases superfamily and more precisely to the family of template-independent polymerases that covalently add nucleotides to the 3’-end of RNA molecules. This family is generally divided in three subgroups: (i) The canonical RNA-specific nucleotidyl transferases (rNTrs) group, which corresponds to the nuclear poly(A) polymerases α, β and γ. These are found in eukaryotes and share similar enzymatic and RNA-binding domains. (ii) The non-canonical rNTrs comprise the Gld-2-, Trf4/5-, Cid1-type poly(A) or poly(U) polymerases, 2’-5’-oligo(A) synthetases and the terminal uridylyl transferases (TUTases). They differ from the canonical rNTrs by their divergent nucleotide recognition motif (NRM) but possess a similar catalytic domain. (iii) The third group is the one of the CCA-adding enzymes, which also have a conserved catalytic domain but contain the most divergent NRM among this family. Canonical rNTrs and CCA-adding enzymes have been characterized in depth in past reviews, thus we will only focus on the non-canonical rNTrs (Martin and Keller 2007).
Every member from the non-canonical rNTrs group is characterized by an enzymatic domain made of two lobes denominated the catalytic and the central domains. The catalytic domain is made of four or five β-strands, which carries a conserved catalytic triad. Specifically, a DxD or DxE motif (aspartate (Asp) or glutamate (Glu) residues separated by one hydrophobic residue) is located in the second β-strand of the catalytic domain. A single Asp or Glu is positioned in the third β-strand of the catalytic domain completing the triad. Thus, the consensus signature of the catalytic site is hG[GS]x(7-13)Dh[DE]h (with h for hydrophobic residues, uppercase letters for invariant, and x for any residue). The central domain contains the NRM, which corresponds to a 10-15 amino acid long loop forming one end of the nucleotide triphosphate binding pocket. As mentioned above, there are three types of NRM corresponding to the three subgroups of the Pol β-NTrs family (canonical-, non-canonical- and CCA-rNTrs associated types). Sub-classification of the rNTrs was attempted based on sequence conservation within the NRM.

Furthermore, a RNA binding domain (RBD) is also found in all canonical and a few non-canonical rNTrs. It shares structural homology with the RNA recognition motif (RRM) protein family and has for function to bind RNA substrates in a non-sequence specific manner (Zhelkovsky et al. 1995; Martin and Keller 1996). The RRM domains are located anywhere in the sequence, i.e. near the C-terminus for the canonical PAPs or at the N-terminus in some non-canonical rNTrs. The RBD is absent in numerous non-canonical NTrs enzymes, indicating that either these proteins can act on any RNA or that their activity is restricted by binding to a yet to be identified protein partner that targets them to specific RNAs. In at least one case, the enzyme ZCCHC11 is targeted to specific pre-miRNA species through interactions with the Lin28 proteins (Yu et al. 2007; Heo et al. 2009; Chang et al. 2012).

From a phylogenetic point of view, several models have been proposed to explain the evolution of the Pol β-NTrs family. The hypothesis of Aravind and Koonin (Aravind and Koonin 1999) is that the Pol β-NTrs family members have rapidly and independently diverged from a common ancestor presenting a very general and non-specific nucleotidyl transferase activity. The different family members would have acquired distinct functional domains to occupy vacant evolutionary niches. Then, horizontal gene transfer and lineage-specific gene loss could have explained the actual distribution of the different groups in the three domains of life. Some evidences like the discovery of the archaeal and bacterial
minimal nucleotidyl transferases (MNT family) and the restricted phylogenetic distribution of most of the Pol β-NTrs family members support this model (Aravind and Koonin 1999). However, it has recently been shown that a bacterial poly(A) polymerase that possesses the RBD of a CCA-adding enzyme is able to act as a CCA-rNTrs (Betat et al. 2004). This suggests that the CCA-adding enzymes could be the ancestors of the poly(A) polymerases and possibly the founders of all the remaining rNTrs, which would have adopted different RNA binding domains mediating different target specificity.

Within the non-canonical rNTrs, there are two main groups of poly(U) polymerases based on their substrates: the Cid1-like family and the RNA editing enzymes. Cid1 (Caffeine-induced death suppressor) protein from Schizosaccharomyces pombe is the pioneer of cytoplasmic poly(U) polymerases (Aravind and Koonin 1999). Many other proteins are part of this group with limited sequence homology but highly similar enzymatic properties such as the trypanosomal protein RNA editing TUTase1 (RET1), which modifies both the gRNAs and the mRNAs in a process that still awaits full characterization (Blum and Simpson 1990; Aphasizhev et al. 2003; Aphasizheva and Aphasizhev 2010; Aphasizhev and Aphasizheva 2011; Aphasizheva et al. 2011). Seven proteins from this sub-group are found in Human although their precise action also required a more detailed characterization (Fig. 5; Trippe et al. 2006; Rissland et al. 2007; Heo et al. 2009).

![Figure 5: Schematic representation of human TUTase proteins.](image)

The TUTases catalytic motif is composed of the nucleotidyl transferase domain (NTr domain, dark aquamarine box) and the PAP-associated domain (magenta box). Light aquamarine box indicates a conserved inactive NTr domain (sequence variations). Green and yellow boxes represent zinc finger and zinc knuckles, respectively. Violet box corresponds to the RNA recognition motif (RRM).
The RNA editing enzymes are responsible for mitochondrial mRNA editing by U-insertion/deletion in kinetoplastids (Rusche et al. 1997; Aphaiszhev et al. 2002; Panigrahi et al. 2003; Wang et al. 2003; Stuart et al. 2005; Aphaiszhev and Aphaiszheva 2008). Two main proteins have been studied in details: RNA editing TUTase 2 (RET2) and the mitochondrial editosome-like complex associated TUTase 1 (MEAT1). RET2 and MEAT1 are found with the 20S editosome complex of trypanosomes and are crucial for the U insertion-type of editing in this organism (Ernst et al. 2003; Wang et al. 2003). Crystal structures of RET2 and MEAT1 showed conserved domain organization except for the middle domain (MD; Deng et al. 2005; Stagno et al. 2010). The lack of sequence similarity within this middle domain suggests divergent functions for these specific modules only found in trypanosomal TUTases.

1.7 Substrates of Polyuridylation in the Different Cell Compartments

Polyuridylation at the 3’-end of RNAs varies based on their localization as well as on the type of RNAs to be modified. Until a few years ago, this process had been mostly reported in the mitochondria of the parasitic protist Trypanosome (Benne et al. 1986; Blum et al. 1990). More recently, non-canonical rNTrs were found in the cytoplasm of eukaryotic organisms and were shown to modify a wide range of non-translated and translated RNAs (Abelson 1979; Mullen and Marzluff 2008; Rissland and Norbury 2009; van Wolfswinkel et al. 2009; Ibrahim et al. 2010; Kamminga et al. 2010; Schmidt et al. 2011; Zhao et al. 2012). Details of the different polyuridylation substrates in the cell nucleus, cytoplasm and mitochondria are described hereafter and summarized in figure 6 (Fig. 6).
1.7.1 Polyuridylation in the nucleus

Until now, the only known substrate of uridylation reported in the nucleus is the U6 snRNA (Fig. 6). This RNA is uridylated by U6 Terminal Uridyl Transferase (TUTase), which is essential for cell survival in mammals (Trippe et al. 2006). U6 TUTase is responsible for the addition or restoration of four uridine residues at the 3’-end of U6 snRNA since 3’-end of U6 snRNA is constantly subjected to exonucleases activity (Trippe et al. 1998; Trippe et al. 2006). These four U residues form an intramolecular double strand with a stretch of adenines within the U6 snRNA molecule, which is important for mRNA splicing (Trippe et al. 2006).
1.7.2 Polyuridylation in the organelles (mitochondria)

Uridylation events in the organelles have only been reported in mitochondria (Aphasizheva and Aphasizhev 2010; Aphasizhev and Aphasizheva 2011). It is apparently absent from the chloroplastic compartment, although proteins from the rNTrs family are present such as the poly(A) polymerase (Zimmer et al. 2009). One possible reason is the close evolutionary conservation of the RNA processing pathways found in the chloroplast and in bacteria (Raynal and Carpousis 1999).

Poly(U) tails have been reported mostly on kinetoplastid organisms. Two main substrates are targeted in these organisms: gRNAs and mRNAs (Fig. 6). The gRNAs are specific to kinetoplastid species and are crucial for cell survival as they are in charge of guiding the editosome machinery to its mRNA targets in these organisms (Blum et al. 1990). RET1, the first characterized ncNTrs, is responsible for the polyuridylation of gRNAs and this step allows the proper maturation of the gRNA (Aphasizhev et al. 2003; Aphasizheva and Aphasizhev 2010). Currently, it is not yet fully understood how RET1 enzyme recognizes its gRNA substrate nor how the pre-gRNA processing step takes place (Aphasizhev et al. 2003; Aphasizheva and Aphasizhev 2010). RET1 was more recently shown to be involved in polyuridylation of mRNAs where it works in concert with the kinetoplast poly(A) polymerase 1 (KPAP1). These two proteins add alternating A/U residues specifically to the 3’-end of mitochondrially-encoded mRNAs. Their action is coordinated by the kinetoplast polyadenylation / uridylation factor 1 and 2 (KPAF1 and KPAF2) complex (Aphasizheva et al. 2011). As for the cytoplasmic RNAs, our molecular understanding of the factor targeting specifically RET1 and KPAP1 activities to the mRNAs is poor and awaits further structural and biochemical characterization (Aphasizheva and Aphasizheva 2011; Aphasizheva et al. 2011).

It is noteworthy that RNAs with poly(U) tails have also been observed in human mitochondria under certain conditions (Slomovic et al. 2005; Slomovic and Schuster 2008; Borowski et al. 2010; Szczesny et al. 2010). In spite of this, how this process is achieved in this compartment and its implication(s) for human mitochondrial RNA metabolism still remain to be characterized.
1.7.3 Polyuridylation in the cytoplasm

Cytoplasmic polyuridylation occurs on a variety of RNA molecules ranging from polyadenylated to non-polyadenylated RNA molecules including mRNAs, small RNAs, miRNAs or piRNAs (Fig. 6; Mullen and Marzluff 2008; Rissland and Norbury 2009; van Wolfswinkel et al. 2009; Ibrahim et al. 2010; Kamminga et al. 2010; Schmidt et al. 2011; Zhao et al. 2012).

As mentioned previously, several eukaryotic mRNAs were shown to be uridylated in the cytoplasm by the poly(U) polymerase Cid1 (Fig. 6; Rissland et al. 2007; Rissland and Norbury 2009). Until now, only a hand full of mRNAs has been identified to be specifically uridylated such as act1, urg1 and adh1 (Rissland and Norbury 2009). More recently, one study looked at the 3'-end sequence of mRNAs at a genome-wide level and revealed that U tails are apparently attached to short poly(A) tracks rather than to the mRNA body (Chang et al. 2014). Interestingly, while some mRNAs like the one encoding the poly(A) binding protein 4 are polyuridylated in more than 25% of the cases, about 80% of mRNAs have an uridylation frequency comprised between 2 and 5%. Overall, the functional relevance of these low-level of uridylation or the identification of the actors responsible for the cytoplasmic mRNA polyuridylation is currently poorly known. Their characterization will likely bring new lights on our understanding of this particular process.

Non-polyadenylated mRNAs are also uridylated in the cytoplasm (Fig. 6). This is the case of the histone mRNAs, which are polyuridylated by Cid1 orthologous enzymes TUTase1 and TUTase3 (also named PAPD1 and Trf4-2 respectively; Mullen and Marzluff 2008). It is not fully clear, however, how these enzymes would switch between cell compartments or how they could select which substrates to uridylate and which one to adenylate in vivo (Mullen and Marzluff 2008). One should not exclude that several pools of PAPD1 differentially located in the cell may exist. Several years later another Cid1 orthologous enzyme, ZCCHC11, was also shown to polyuridylate histone mRNAs following inhibition or complete DNA replication (Schmidt et al. 2011). More data are definitely required to fully apprehend ncNTrs role during regulated histone mRNA degradation in particular regarding the factors bringing together the histone mRNAs and the ncNTrs.

A variety of ncRNAs in a variety of organisms have been shown to carry mono- or multiple non-templated uridine residues at their 3'-end (Fig. 6; Li et al. 2005; Ibrahim et al. 2006;
Ibrahim et al. 2010; Heo et al. 2012; Thornton et al. 2012). However, the functional consequences associated with these modifications are highly diverse. 3’ uridylation of various miRNAs has been observed in multiple sequencing studies suggesting a wide role of uridylation during miRNA biogenesis pathway (Landgraf et al. 2007; Morin et al. 2008; Choi et al. 2012). These mono- or polyuridylation events have been found in both pre-miRNAs and mature miRNAs (Ibrahim et al. 2010; Newman et al. 2011; Heo et al. 2012; Thornton et al. 2012). In C. elegans and Human, polyuridylation of the pre-let-7-miRNA has been reported and is performed by the proteins PUP-2 and ZCCHC11 respectively (Hagan et al. 2009; Lehrbach et al. 2009; Thornton et al. 2012). Association between the pre-miRNA and the Lin28 protein induces a conformational change in the pre-miRNA loop, which possibly favors modification by ZCCHC11 (Nam et al. 2011; Loughlin et al. 2012). However, the presence of a single 3’-overhanging nucleotide appears critical for the uridylation process therefore excluding the so-called "group I" or canonical miRNAs from being subject to uridylation (Heo et al. 2012). Furthermore, in the same study, the ZCCHC6 enzyme was found responsible for the mono-uridylation of group II let7 pre-miRNAs and this modification is independent of Lin28 protein (Heo et al. 2012). ZCCHC11 has also been involved in the uridylation of specific mature miRNA such as miR-26 (Jones et al. 2009). Further biochemical and biophysical studies are needed in order to identify the specific enzymes responsible for the uridylation of other miRNAs in higher organisms as well as the target-specific effects induced by this 3’-end modification. 3’ uridylation of piRNAs have been observed in zebrafish and drosophila but the enzymes responsible for this modification are currently unknown (Ameres et al. 2010; Kamminga et al. 2010). Another class of ncRNAs subjected to 3’ uridylation is the siRNAs (Fig. 6). In nematodes, this particular type of RNA substrate is modified by the protein CDE-1 (Cosuppression defective 1) and HESO-1 (HEN1 SUPPRESSOR 1; Ren et al. 2012). In plants and lower eukaryotes, MUT68 has been implicated in this event (Ibrahim et al. 2010). This last enzyme is shown to act on atypical substrates, i.e. the product of the miRNA-directed mRNA cleavage (Ren et al. 2014). In this case, the uridine nucleotides are apparently added to the 5’-fragment of the cleaved mRNA when it is still bound by the AGO1 complex (Ren et al. 2014). Further studies will help determining the generality of this mechanism as the responsible enzyme HESO-1 does not seem to be conserved in higher eukaryotes.
1.8 Known Functions of Polyuridylation

3’ uridylation is a widespread phenomenon in eukaryotes with multiple biological consequences depending on the substrate to be modified (coding RNA or ncRNA), the eventual targets of the uridylated RNA and the cell compartment. In the following section, we will review the main known consequences of uridylation in the nucleus (Fig. 7), in the mitochondria (Fig. 8) and in the cytoplasm (Fig. 9).

1.8.1 In the nucleus

The highly specific uridylation of U6 snRNA by U6 TUTase is necessary for the U6 snRNA stability and function (Trippe et al. 2006). As we mentioned before, this is the only substrate known to be uridylated in this compartment and the main function of this uridylation event is the stabilization of this snRNA crucial for RNA splicing. Mammalian U6 snRNA uridylation in vivo has been reported with up to 20 nucleotides added at the 3’-end of the RNA molecule (Rinke and Steitz 1985; Lund and Dahlberg 1992). It is important to note that U6 snRNA is also subjected to adenylation and this event inhibits its uridylation (Fig. 7; Chen et al. 2000). Further analysis by Chen and colleagues observed U6 snRNA transcripts where the adenine residue at position 102 was replaced with a uridine, strongly suggesting a role of uridylation in the regeneration of the U6 snRNA 3’ end following its shortening by exonucleases (Chen et al. 2000). Biochemical studies identified the TUTase responsible for U6 snRNA 3’ uridylation and this enzyme was named U6 TUTase (Trippe et al. 2006). SiRNA-mediated silencing of the U6 TUTase lead to U6 snRNA decay confirming the necessity of uridylation for U6 snRNA stability (Trippe et al. 2006). So far, this is the only uridylation event known in the nucleus (Fig. 7).
1.8.2 In the organelles

So far, no polyuridylation event has been found in the chloroplast of plants and algal cells. In this organelle, RNA surveillance mechanism is rather close to the one found in bacteria where polyadenylation of the 3’-end of mRNAs is the main destabilizing factor (Lisitsky et al. 1996; Lisitsky et al. 1997; Blum et al. 1999).

Polyuridylation of RNAs in the mitochondria has been observed and has two major impacts depending on the type of RNA modified. As I mentioned before, this event is specific to kinetoplastid organisms and is found on gRNAs and on mRNAs. For gRNAs, polyuridylation is the final maturation step and is necessary for the editing process of mitochondrially-encoded mRNAs whereas uridylation together with polyadenylation are necessary for the proper translation of edited and never edited mRNA (Fig. 8).

In order to be matured, pre-gRNAs need to pass through an exonucleolytic process followed by stabilization by the gRNA Binding Complex (GRBC) and RET1 uridylation (Blum and Simpson 1990; Weng et al. 2008; Aphasizheva and Aphasizhev 2010). Mature gRNA is thus composed of a 5’ phosphate from the transcription followed by an anchor region complementary to a target unedited mRNA, a guiding region that directs the editing of its mRNA target and a final poly(U) tract at the 3’-end. In RET1-depleted cells, gRNAs are stable but not able to perform their editing function suggesting a crucial role of the oligo(U) tail in the editing event in the mitochondria. This oligo(U) tract may stabilize the gRNA-mRNA interaction.
hybrid through binding with the purine-rich pre-edited region (Blum and Simpson 1990). The uridylated gRNA bound to its mRNA target recruits the 20S editosome. This gRNA-mediated mRNA editing in kinetoplastid trypanosomes is crucial for the parasite survival, as these editing events are needed for the proper establishment of the coding sequence of the mitochondrial mRNAs (Blum and Simpson 1990).

After editing, mRNAs need to be further modified at the 3'-end in order to be translationally competent in trypanosomal mitochondria. The addition of a long 3’ A/U tail is necessary to make the mRNAs favorable for translation (Fig. 8; Aphasizheva et al. 2011). This nucleotide addition is performed by the RET1/KPAP1 complex, which adds around 200 alternated adenines and uridines to the 3'-end of the targeted mRNA and is modulated by the KPAF1/KPAF2 protein complex (Aphasizheva et al. 2011). The polyadenylated/polyuridylated mRNA is then recognized by the translation machinery to give birth to the encoded protein (Aphasizheva et al. 2011).

1.8.3 In the cytoplasm

The functional outcome of polyuridylation in this compartment offers new insights into RNA turnover and small RNA biogenesis (Fig. 9). Most of cytoplasmic uridylated mRNAs and sRNAs are targeted for decay thus strongly implying a destabilizing role for uridylation. RNA
cRACE studies in fission yeast revealed a role of uridylation in a new deadenylation-independent decapping-mediated degradation pathway (Rissland and Norbury 2009). Furthermore, a single uridine at the 3’-end of a RNA molecule is sufficient to be recognized by the Lsm1-7 complex, known to link 3'-end deadenylation and 5'-end decapping, therefore supporting the relationship between uridylation and mRNA degradation (Song and Kiledjian 2007). Recently discovered, Dis3L2 protein in fission yeast was shown to preferentially degrade mRNAs with 3'-end uridylation and its deletion together with Lsm1 led to the accumulation of uridylated RNAs (Malecki et al. 2013). Mammalian Cid1 orthologs ZCCHC11 and ZCCHC6 proteins together with the tumor suppressor protein Lin28 uridylate pre-let7-miRNA in humans, thereby influencing positively and negatively the miRNA production (Schmidt et al. 2011). Interestingly, the mammalian DIS3L2 exonuclease was also shown to specifically degrade uridylated pre-let7-microRNA discriminating them from 3’-unmodified RNAs (Lubas et al. 2013). Further studies between DIS3L2 exonuclease and TUTases will be necessary to better understand their respective function and more importantly, the functional link existing between these enzymatic activities.

Upon inhibition of DNA replication or conclusion of S-phase, histone mRNAs must be rapidly degraded in order to avoid their accumulation and their interference with other cellular pathways (Osley 1991). Histone mRNAs are not polyadenylated but possess a stem-loop structure at their 3’-end crucial for pre-mRNA processing, export and proper translation (Pandey and Marzluff 1987; Gallie et al. 1996; Battle and Doudna 2001). Studies aiming to understand the mechanism by which histone mRNA degradation was triggered found that histone mRNAs were targeted to decay by uridylation (Mullen and Marzluff 2008; Schmidt et al. 2011). Indeed, knockdown of ZCCHC11 enzyme stimulated the accumulation of histone-encoded mRNAs (Schmidt et al. 2011). Interestingly, as for polyadenylated mRNAs in fission yeast, uridylation of histone mRNAs was shown to promote decapping followed by 5’-3’ degradation (Mullen and Marzluff 2008). More recently, 3’-5’ degradation of histone mRNAs by ERI1 exonuclease has been reported (Hoefig et al. 2013). Lsm1-7 promotes the degradation of the 3’-end stem-loop of the mRNA by recruiting ERI1 to the stem-loop and simultaneously blocks translation of these mRNAs. In this case, Lsm1-7 complex binds to the uridylated histone mRNAs and to ERI1 (Hoefig et al. 2013).
In humans, ZCCHC11 in concert with Lin28 polyuridylates pre-let7-miRNA, which will then be degraded by DIS3L2 exonuclease. ZCCHC6 alone is responsible for the monouridylation of group II pre-miRNA, which will be further processed by Dicer. In the case of mature miRNA, ZCCHC11 monouridylates some miRNA leading to indirect consequences to the miRNA targeted mRNAs. In plants, zebrafish and flies, methylation and polyuridylation have antagonistic effects. Methylated siRNAs, piRNAs and miRNAs will be stabilized whereas those polyuridylated will be degraded. Finally, in S. pombe, polyuridylation of mRNAs by Cid1 PUP leads to decapping and decay.

Studies in plant, animal and fly species have demonstrated an antagonistic role of uridylation and 2'-O-methylation in these organisms (Li et al. 2005; Yu et al. 2005; Kurth and Mochizuki 2009; Ameres et al. 2010; Kamminga et al. 2010). HEN1 (HUA ENHANCER 1) and its homologs methylate sRNAs in plants, piRNAs in animals and Ago2-associated siRNAs in flies protecting these RNAs against 3' uridylation (Li et al. 2005; Yu et al. 2005; Kurth and Mochizuki 2009; Ameres et al. 2010; Kamminga et al. 2010). In zebrafish hen1 mutants, piRNAs derived from retrotransposons are found uridylated and their levels are decreased suggesting a sensitivity of these uridylated piRNAs to degradation. Interestingly, a mild
repression of retransposons is observed in these mutants thus highlighting a destabilizing role for uridylation of piRNAs and a stabilizing role for methylation, which is the likely cause of transposons silencing in the *hen1* mutant (Kamminga et al. 2010). In *C. elegans*, CSR-1 is an Ago protein necessary for proper chromosome segregation rather than regulation of mRNA levels (Claycomb et al. 2009; Gu et al. 2009; van Wolfswinkel et al. 2009). CDE-1, a *C. elegans* PUP, uridylates unmethylated siRNAs of the CSR-1 pathway (van Wolfswinkel et al. 2009). Mutation of this CDE-1 enzyme leads to accumulation of CSR-1 siRNAs, which promotes erroneous chromosome segregation and defective gene silencing (van Wolfswinkel et al. 2009). Uridylation is then a destabilizing factor against CSR-1 siRNAs, which regulates CSR-1-dependent and specific siRNA levels in this organism. In *Chlamydomonas reinhardtii*, MUT68 was first known to adenylate 5’ cleavage fragments of mRNAs targeted by the RNA-induced silencing complexes (RISC), thereby promoting their decay (Ibrahim et al. 2006). Further studies showed an important role of MUT68 in miRNA and siRNA degradation through 3’ uridylation (Ibrahim et al. 2010). Taken together, these data highlights the crucial role of sRNA uridylation for diverse biological processes in several organisms and that defects in the regulation of this phenomenon can have important consequences on the gene expression of the RNA targets (Fig. 9).

Polyuridylation has also been reported to stabilize RNAs and thus prevent their degradation. In *Arabidopsis thaliana*, uridylation of oligoadenylated mRNAs has been suggested to prevent their 3’ trimming and rather establish a preferential 5’-to-3’ mRNA degradation manner (Sement et al. 2013). Indeed, URT1 (UTP:RNA uridylyl transferase 1) was shown to uridylate oligo(A)-tailed mRNAs *in vivo* and its absence contributed to the degradation of oligoadenylated mRNAs highlighting a new role of uridylation in mRNA stability. ZCCHC11 enzyme, besides its role on histone mRNA and pre-miRNA decay, has also been implicated in indirect mRNA stabilization by uridylation of mature miRNAs. ZCCHC11-dependent uridylation of mature cytokine-targeting miRNAs is known to lead to the stabilization of cytokine transcripts and hence regulates cytokines gene expression. Mature miR-26 can bind interleukin IL-6 mRNA in its 3’ UTR and targets this cytokine-encoding mRNA to degradation (Jones et al. 2009). Upon miR-26 uridylation by ZCCHC11, the miRNA is unable to bind the 3’ UTR of the mRNA and thus the transcript is stabilized even though no degradation of the miRNA is observed. In this case, uridylation of miRNAs indirectly stabilizes the mRNA
targeted by the unmodified miRNA by preventing the formation of the miRNA/mRNA complex and thus blocks the miRNA directed silencing function. This is further confirmed by ZCCHC11 knockdown experiments where several cytokine mRNAs are downregulated in the absence of uridylation (Minoda et al. 2006; Jones et al. 2009).

These data together supports a crucial role of cytoplasmic polyuridylation in the regulation of gene expression and stability control of coding and non-coding RNAs in eukaryotes. Understanding in detail the enzymatic properties of PUPs is therefore of primordial importance and is the subject of the present thesis research.
Goals of the Thesis Work

The aim of this thesis work was to understand the molecular mechanisms by which polyuridylating enzymes discriminate UTP over ATP, CTP and GTP. In parallel, we wished to gather insights of how these enzymes recognize their RNA target at the atomic level to help explain their ability to modify such a variety of RNAs. For this purpose, we used X-ray crystallography as our main tool in order to reveal novel aspects of poly(U) polymerases along with their protein/protein and protein/RNA interactions. In the first project, we obtained structural information on protein/substrate, protein/RNA and protein/substrate/RNA interactions, which were complemented by mutational and functional studies of the different components of these complexes. In the second project, we tried to study protein/protein and protein/RNA interactions attempting to elucidate the specific functions of the different subunits of the RDS complex (RET1 TUTase DSS1 exonuclease complex), a major uridylation complex in Trypanosomal species.
1.9 Project I: Functional and Structural Studies of Cid1 Poly(U) Polymerase from *S. pombe*

Cid1, a template independent cytoplasmic RNA poly(U) polymerase from *S. pombe* and pioneer of PUP enzyme family, is able to use ATP and UTP as a substrate but will preferentially use UTP *in vivo* (Rissland and Norbury 2009). This enzyme lacks any known RRM suggesting that it needs a protein partner for RNA target selection even though no partners have been identified so far (Rissland *et al.* 2007). The molecular mechanism by which this enzyme was able to discriminate UTP over ATP was not described previously. Hence, in order to understand its UTP selectivity and how the enzyme was selecting and recognizing its RNA target, we aimed to solve the crystal structure of the enzyme alone and in secondary and ternary complexes. Followed by point mutations, several biochemical analyses were performed to better characterize its function and selectivity *in vitro*.

1.10 Project II: Structural Studies of Mitochondrial RNA Processing Complexes in *T. brucei*

DNA in the mitochondria of trypanosomes is densely packed in a structure called the kinetoplast, composed of approximately 40 to 50 maxicircles, which yield to rRNA and mRNA precursors, and thousands of minicircle encoding mostly gRNAs. These gRNAs are necessary for mRNA editing however their biogenesis is not completely understood. Our collaborators in the group of Prof. Ruslan Aphasizhev characterized the complex responsible for the gRNA biogenesis called RDS complex. This complex is composed of the TUTase RET1, the exonuclease DSS1 and three other proteins lacking any recognizable motif named RDS1, RDS2 and RDS3. Our purpose in this project was to solve the atomic structure of the RDS complex. However, purification of a stable complex has not yet been achieved. Therefore, we are now pursuing to obtain atomic structures of the individual subunits with a special emphasis on RET1 alone, RET1 with UTP and/or with its RNA substrate and as a protein/protein complex with DSS1. We aim to solve the crystal structure of RDS1, RDS2 and RDS3 proteins either as a whole complex or as several sub-complexes along with RET1 and DSS1 in order to further investigate the functions of these proteins.
Results

The results obtained during this PhD research are presented in the two following sections 4.1 and 4.2. The first section is focused on “Functional and structural studies of Cid1 poly(U) polymerase from *S. pombe*” project, and the second section describes the “Structural studies of mitochondrial RNA processing complexes in *T. brucei*” project. Within the first section, two publications are included and complemented with unpublished work performed on Cid1 enzyme. The second section describes the ongoing work on the different proteins of the RDS complex from trypanosomes in collaboration with Ruslan Aphasizhev laboratory in Boston and a master student in our laboratory, Marius Long.
1.11 Functional and Structural Studies of Cid1 Poly(U) Polymerase from S. pombe

1.11.1 Cid Overexpression and Purification

Early studies of Cid1 full-length (FL) enzyme showed that the protein is not stable during purification (Rissland et al. 2007). We then decided to try two different constructs named Cid1S (amino acids 40 to 366) and Cid1L (amino acids 40 to 377; Fig.10A) based on secondary structure predictions performed by the Phyre program (Kelley and Sternberg 2009). We cloned Cid1S and Cid1L into two of our expression vectors with either a heptahistidine tag (pST2) or a nonahistidine tag followed by an MBP (Maltose Binding Protein) tag (pST18). A TEV (Tobacco Etch Virus) protease cleavage site is found between the tag and the recombinant protein ORF in order to be able to remove the tag prior to crystallization. We initially tested expression of the His-tagged/MBP version of Cid1S and Cid1L in small-scale, in several different conditions (different E. coli strains, media, induction temperatures and times; Fig. 10B). We observed no expression in Luria Broth (LB), Terrific Broth (TB) and autoinduction (AI) media when we used the Rosetta 2 strain of E. coli (Fig. 10B). When using E. coli strain BL21-Star, all the three different media had good expression but the best condition for Cid1 expression was AI (Figure 10C).

We then checked the solubility after lysis followed by a small-scale purification with Ni-NTA beads of the two enzyme constructs (Fig. 11A, 11B and 11C). We could obtain low quantity of soluble protein for Cid1S whereas the amount was higher for Cid1L (Fig. 11B and 11C respectively). To be sure that the eluted protein is indeed our recombinant Cid1, we took one of the elution fractions and cleaved the tag with the TEV protease overnight (O/N) at 16°C (Fig. 11D). Both of the constructs are TEV cleavable but since we obtain higher amounts of protein for Cid1L, the large-scale purification optimizations were performed with this construct. For structural biology studies, the protein purification should yield at least 1 mg of 95% pure protein.

The protein large-scale purification was straightforward and the details of buffers and purification method are described in materials and methods chapter 4.1.4. Throughout the whole purification, we assessed the purity of the enzyme in a 10% SDS-PAGE gel (Fig. 12A). In the size exclusion chromatography step, we observe two peaks (Fig. 12B). The first peak
corresponds to the aggregated population of tagged protein and contaminants and the second peak corresponds to Cid1L monomer (Fig. 12B). At the end of the purification, we obtain 10 mg of Cid1L enzyme with more than 95% purity out of 10 g of lysed cells. The protein was further concentrated to 9.5-10mg/ml for setting up factorial crystallization screens.

**Figure 10: Cid1 truncated versions scheme and expression tests.** (A) Cid1 enzyme constructs Cid1L (40-377 Aa.) and Cid1S (40-365 Aa.) tested. The tag is in light blue, the TEV cleavage site is in dark blue and Cid1 enzyme is in grey with the nucleotide transferase domain in aquamarine and the PAP-associated domain in magenta. (B) Cid1L and Cid1S expression tests in Rosetta 2 E. coli strain with three different media (LB, TB and AI) loaded onto a 15% SDS-PAGE gels. (C) Cid1L and Cid1S expression tests in BL21-Star E. coli strain with three different media (LB, TB and AI) loaded onto a 15% SDS-PAGE gels. The arrow corresponds to Cid1 tagged protein in panels B and C. LB and TB expressions were induced at 18°C with 0.1mM IPTG (Isopropyl-β-D-
thiogalactopyranosine). AI expressions were performed at 30°C in panels B and C. Ni for none-induced sample, O/D for overlay, O/N for overnight, h for hour and Mw for protein molecular weight.

**Figure 11: Cid enzyme truncated versions solubility, purification and TEV cleavage trials.** (A) Cid1S and Cid1L small-scale expressions in LB and AI media lysed with the emulsiflex and loaded into a 10% SDS-PAGE gel to check the presence of Cid1 in the soluble fraction. (B) Cid1S expressions in LB or AI media batch purification. FT for Flow-through, W for wash and E1, 2, 3 and 4 for elutions 1, 2, 3 and 4. (C) Cid1L expressions in LB or AI media batch purification. (D) 16°C TEV overnight cleavage of Cid1L and Cid1S constructs expressed in LB or AI media. Ni for none-induced sample, T for total fraction, P for pellet insoluble fraction, S for soluble fraction, BT for sample before TEV cleavage and AT for sample after TEV cleavage.
1.11.2 Cid1 Crystallization

We first screened ~480 individual crystallization conditions of the protein alone testing two different protein concentrations (9.5mg/ml and 4.5mg/ml) at 18°C. Several crystallization hits were obtained over these conditions but only one gave single crystals (Fig. 13A). Morpheus screen condition containing 0.1M MES/Imidazole pH 6.5, 30% Glycerol/PEG4000 and 0.09M of a mix of NaI, NaBr and NaF gave single crystals in the middle of urchins (Fig 13A). After one round of optimization of the crystallization condition, we checked if these crystals were indeed Cid1L-containing crystals. For this purpose, we fished three big rod crystals which we washed three times on three different reservoir drops and loaded onto a 15% SDS-PAGE gel together with the whole drop, last wash of the crystals and the protein sample (Fig 13B). We can clearly see on the SDS-PAGE gel that the crystals contain Cid1L protein (Fig 13B).
Changing the buffer, precipitant and additive concentrations and combinations still gave us very small crystals with no defined edges in sitting drop vapor diffusion plates (Fig 13C). During optimization trials, we noticed that leaving the protein together with the crystallization condition in an eppendorf gave us bigger crystals with sharper edges. This means that there is no need of equilibration for crystal growth. So, we decided to try two other different crystallization methods: macro-seeding and pseudo-batch (the solution on the reservoir and the drop are at the same concentration, so there is no equilibration between the reservoir and the drop as in the normal vapor diffusion method; Fig. 13D and 13E). In the case of macro-seeding, we managed to obtain bigger crystals but their shape was not sharp enough. But when we performed pseudo-batch method, the crystals grew bigger and with sharp edges (Fig. 13E). These crystals were then tested for diffraction using our in-house source and the Beamline ID14-4 of the European Synchrotron Radiation Facility (ESRF) together with crystals soaked with 2.5mM of UTP or ATP supplemented with 5mM of MgCl₂.

### 1.11.3 Data Collection and Phasing

Cid1 crystals diffracted up to 2.4Å and 2.28Å resolution for the native and UTP-soaked crystals respectively (Fig. 14A and 14B). Diffraction data reduction and scaling were performed as described in materials and methods chapter 4.1.4.

The data reduction and scaling only give us the intensities and amplitudes of the spots. To obtain the exact position of the atoms inside the crystal lattice, we need to calculate the phases. In order to do so, we used the molecular replacement (MR) method where the phases of an atomic model with at least 25% sequence identity to our protein would be used for phase calculation. Unfortunately, molecular replacement experiments with TbtUT4 as a model were not successful since this protein only has 20% sequence identity with Cid1.
We then expressed Cid1 seleno-methionine (SeMet) derivative. All the methionines in the protein will be replaced by seleno-methionine, which will have 18 electrons more than methionine allowing the phasing by anomalous method. We implemented the same purification method as for the WT protein and yield to around 2mg of pure SeMet-derivative monomeric protein for 5g of cell pellet (Fig. 15A and 15B). We performed crystallization trials in the same conditions as for the WT protein as well as general grid screens and micro-seeding (Fig. 15C). All the crystals obtained were tested at the synchrotron but unfortunately we could not obtain enough anomalous signal (SigAno <1) to identify the position of the methionines (Fig. 15D).

Figure 14: Cid1 enzyme high resolution diffraction patterns. (A) Cid1 native enzyme high resolution diffraction with a snap shot of the crystal used for data collection. (B) Cid1 enzyme soaked with 2.5mM UTP supplemented with 5mM MgCl₂ high resolution diffraction pattern.
In order to delineate the phases with a different method, we tried heavy atom soakings of the native crystals. Once inside the crystals, heavy atoms can be very destructive depending on the concentration and time of soaking. Several rounds of soakings with the different conditions were tried and summarized on table 1 (Table 1). In the first round of heavy atom soakings with KAuCl₄ compound, the crystals cracked very fast meaning that this heavy atom interacts with the protein. Lower concentrations of KAuCl₄ compound and O/N soakings as well as a 7 days KPtCl₄ platinum soak allowed us to obtain one useful derivative out of which we could solve the structure using a single-wavelength anomalous diffraction (SAD) phasing strategy (Fig 16A and 16B). These preliminary phases allowed us to build a partial atomic model of Cid1, which was then used to obtain a solution using the molecular replacement technique into a native dataset with a higher resolution. We tried the same procedure with the datasets collected from UTP-soaked Cid1 crystals. In some cases, the density was good enough to start building the enzyme (Fig 16C). Details of the structure solving along with the mutational and functional analysis of Cid1 based on the crystal structure were published in the journal Structure of Cell press and are shown in the chapter below (section 4.1.4).
### Table 1: Different heavy atom compounds soaking concentrations and times tested for SAD phasing.

<table>
<thead>
<tr>
<th>Heavy atom compound</th>
<th>Soaking concentration</th>
<th>Soaking time</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{K}_2\text{PtCl}_4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{K}_2\text{PtBr}_4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{K}_2\text{Pt(CN)}_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{C}_8\text{H}_8\text{HgO}_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{C}_9\text{H}_9\text{HgNaO}_2\text{S}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{C}_2\text{H}_5\text{HgCl}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{HgCl}_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{NaAu(CN)}_2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{KAuCl}_4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{K}_2\text{ReCl}_6$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{GdCl}_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{NdCl}_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{YbCl}_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samarium Acetate</td>
<td>$0.5\text{mM, 1mM, 2mM, 5mM}$</td>
<td>1h, 5h, 8h, O/N, 2 days, 7 days</td>
</tr>
<tr>
<td>Samarium Nitrate</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 16: Data analysis of crystals soaked with a platinum and gold compound. (A) Diffraction pattern of a K₂PtCl₄ compound and its XSCALE.LP table. (B) Diffraction pattern of a KAuCl₄ compound and its XSCALE.LP table. For panels A and B, XSCALE.LP table showing the subset of resolution limit, the I/σ, the R_{meas}, the R_{merge} and the Anomalous Corr and SigAno, which shows the anomalous signal of the heavy atoms. (C) Electron density map from the KAuCl₄ derivative obtained after heavy atom localization, phasing and density modification procedures.
1.11.4 Publication I: “Functional Implications from the Cid1 Poly(U) Polymerase Crystal Structure”

We solved the atomic structure of Cid1 enzyme in complex with its substrate UTP and identified several residues critical for the UTP selectivity and RNA recognition of the enzyme. The structure reveals multiple residues from the nucleotide recognition motif (NRM) and from several positively charged patches of the enzyme that are shown to stabilized the UTP molecule. Indeed, histidine 336 of the NRM loop is in direct contact with the uracil base of the UTP, a feature of PUPs, which was not shown before. Our atomic model, complemented with point mutations and functional studies, is used to propose a catalytic cycle where local and global movements of the enzyme are needed for RNA accommodation and translocation. Our research underlines the molecular insights on Cid1’s UTP selectivity and RNA recognition with crucial implications for the PUP targeted RNAs along with the PUP-dependent polyuridylation process.
Functional Implications from the Cid1 Poly(U) Polymerase Crystal Structure

Paola Munoz-Tello,1 Caroline Gabus,1 and Stéphane Thore1,*
1Department of Molecular Biology, University of Geneva, 30 Quai Ernest Ansermet, Geneva 1211, Switzerland
*Correspondence: stephane.thore@unige.ch

SUMMARY

In eukaryotes, mRNA degradation begins with poly(A) tail removal, followed by decapping, and the mRNA is then degraded by 3′-exonucleases. In recent years, the major influence of 3′-end uridylation as a regulatory step within several RNA degradation pathways has generated significant attention toward the responsible enzymes, which are called poly(U) polymerases (PUPs). We determined the atomic structure of the Cid1 protein, the founding member of the PUP family, in its UTP-bound form, allowing unambiguous positioning of the UTP molecule. Our data also suggest that the RNA substrate accommodation and product translocation by the Cid1 protein rely on local and global movements of the enzyme. Supplemented by point mutations, the atomic model is used to propose a catalytic cycle. Our study underlines the Cid1 RNA binding properties, a feature with critical implications for miRNAs, histone mRNAs, and, more generally, cellular RNA degradation.

INTRODUCTION

In bacteria, the addition of the poly(A) tail at the 3′-end of mRNA molecules is a destabilizing factor for the mRNAs (for a recent review, see Régner and Hajnsdorf, 2009). In eukaryotes, a similar effect on mRNA stability has been attributed to 3′-end uridylation (Song and Kiledjian, 2007; Mullen and Marzluff, 2008). This addition of uridylic nucleotides to RNA substrates has been involved in several distinct RNA metabolic pathways (Shen and Goodman, 2004; Norbury, 2010). Recently, the general mRNA degradation pathway in the Schizosaccharomyces pombe was shown to be modulated by the protein Cid1 (Rissland and Norbury, 2009). The polyuridylation of polyadenylated mRNAs is crucial for triggering the UTP-dependent degradation pathway in Trypanosoma brucei mitochondria, and this modification is performed by RNA editing terminal uridylyl transferase 1 (RET-1; Ryan and Read, 2009). The uridylation of U6 snRNA 3′-end by U6 snRNA-specific terminal uridylyl transferase (TUTase) occurs in molecules lacking the cyclic 2′-3′-phosphate group. This post-transcriptional modification is required for the maintenance of the 3′-end of the U6 snRNA (Hirai et al., 1988; Chen et al., 2000). Another example is the regulation of mammalian let-7 miRNA production, which is, at least partially, mediated by the Zcchc11-dependent uridylation of the let-7 precursor (Heo et al., 2009). Finally, several reports have recently shown that 3′-end uridylation happens on other types of poly(A)-minus RNAs, such as histone mRNAs, subsequently targeting these particular RNAs for degradation (Mullen and Marzluff, 2008; Schmidt et al., 2011).

In the past decade, an entire family of polymerase enzymes, called terminal uridylyl transferase (TUT) or poly(U) polymerase (PUP), has been identified in the metazoa (Martin and Keller, 2007). Several of its members have been shown to uridylate mRNA substrates in vivo (Kwak and Wickens, 2007; Martin and Keller, 2007). One of the first characterized enzymes with the above-mentioned destabilizing activity is the protein Cid1 (caffeine-induced death suppressor), a cytoplasmic PUP (Wang et al., 2000). Its action was first identified in the S-M phase transition control. Deeper analysis revealed that the Cid1 protein was not adding a poly(A) tail but rather a poly(U) stretch in vivo (Rissland et al., 2007). Moreover, the half-life of mRNAs with uridine residues added to the poly(A) tail, such as act1, was longer in cid1 mutants compared to the wild-type, indicating that the Cid1-dependent uridylation was likely to promote mRNA degradation (Rissland et al., 2007; Rissland and Norbury, 2009). The Cid1 uridylation of polyadenylated mRNAs like act1, adh1, and urg1 promotes decapping, interceded by the Lsm1-7 complex, which leads to their decay (Rissland and Norbury, 2009). In vitro, these enzymes can add both poly(A) or poly(U) stretches. However, in vivo, they exert their function in a uridine-specific fashion (Rissland et al., 2007). Members of this family, Cid1 being the most fully characterized, exhibit highly processive activity, leading to the addition of long U-stretches (Rissland et al., 2007). However, the length of the uridine tail observed in vivo is far shorter (between 1 to 10 nucleotides). The reason for these observed differences is presently unclear. One possible mechanism would consist of a balance between the PUP activity and an exonuclease trimming mechanism allowing the control of the poly(U) length as observed with the U6 snRNA (Chen et al., 2000) or for the poly(A) tail (Kim and Richter, 2006).

Here we report the crystal structure of Schizosaccharomyces pombe Cid1 protein in complex with UTP. The structure reveals the selective pocket for the pyrimidine nucleotide and highlights the specific interaction between the uracil base and multiple amino acids of the protein. Multiple residues of the nucleotide recognition motif (NRM) and from various parts of the protein are shown to stabilize the UTP. Mutations of the corresponding residues lead to functional consequences as reflected by variable polymerase activity in vitro. In addition, superpositions of our UTP bound structure and several related Pol J family

1Department of Molecular Biology, University of Geneva, 30 Quai Ernest Ansermet, Geneva 1211, Switzerland
*Correspondence: stephane.thore@unige.ch
Table 1. Data Collection

<table>
<thead>
<tr>
<th>Crystal</th>
<th>KAUCl4-Soaked</th>
<th>UTP-Soaked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.0494</td>
<td>0.9811</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Maximum resolution (Å)</td>
<td>43.8–2.6</td>
<td>45.2–2.28</td>
</tr>
<tr>
<td></td>
<td>(2.67–2.6)</td>
<td>(2.3–2.28)</td>
</tr>
<tr>
<td>Space group</td>
<td>P2₁,2₁,2₁</td>
<td>P2₁</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
<td>52.8, 75.7, 78.4</td>
<td>53.5, 76.9, 81.9, l = 81.17</td>
</tr>
<tr>
<td>No. of observations</td>
<td>190,902</td>
<td>83,724</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>18,538</td>
<td>29,999</td>
</tr>
<tr>
<td>Redundancy</td>
<td>10.3</td>
<td>2.8</td>
</tr>
<tr>
<td>Data completeness (%)</td>
<td>99.2 (91.9)</td>
<td>98.74 (93.9)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>28.13 (1.91)</td>
<td>11.5 (2.64)</td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td>RMSD from ideal geometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.112</td>
<td></td>
</tr>
<tr>
<td>Mean B (Å²)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein chain</td>
<td>30.9</td>
<td></td>
</tr>
<tr>
<td>Nucleotide/magnesium chains</td>
<td>26.1</td>
<td></td>
</tr>
<tr>
<td>Waters</td>
<td>31.7</td>
<td></td>
</tr>
<tr>
<td>Residues in favored region of the Ramachandran plot (%)</td>
<td>603 (87.3)</td>
<td></td>
</tr>
<tr>
<td>Residues in allowed region of the Ramachandran plot (%)</td>
<td>17 (2.7)</td>
<td></td>
</tr>
</tbody>
</table>

Values in bracket are for the highest resolution shell.

Structure of the Cid1 Protein

The Cid1 protein was first characterized as a factor involved in the S-M phase transition process (Wang et al., 1999) and later was classified as a NT from the polymerase β superfamily based on sequence conservation (Figure 1A; Wang et al., 2000). The overall protein fold is related to the polymerase β fold with two characteristic domains, catalytic (CAT) and central (CD) domains, separated by a trench where the UTP is bound (Figure 1B).

The CAT domain contains the typical features of the DNA polymerase β enzyme, i.e., a five-stranded β sheet with two α helices on one side and the catalytic aspartic acid triad facing the CD domain (residues 101, 103, and 160; Figure 1B). The CD domain is composed of six α helices connected by multiple loops of various lengths (Figure 1B). The CD domain contains the NRM as shown in Figure 1B (see below). The CAT domain and the CD domain are linked by two loops (res. 56–58 and res. 163–165; Figure 1B).

Since the Cid1 protein has been classified in the group of non-canonical template-independent NT by Martin and colleagues (Martin and Keller, 2007), we used a 17-mer long RNA finishing with a Cytidine to test the polymerizing activity of our engineered protein. Our construct displays similar activity level compared to the full-length enzyme (Figure 1C). We also introduced two point mutations at positions 101 and 103, demonstrating that the...
poly(U) addition is due to our protein and not to a co-purifying contaminant.

The Cid1/UTP/Mg2+ Complex Structure

The Cid1/UTP complex structure was obtained from an over-night soaked crystal. The UTP is buried at the bottom of the cleft found between the CAT and the CD domains (Figure 2A; Figure S1 available online). Multiple interactions are observed between the nucleotide triphosphate and the protein and rely on residues conserved in the PUP/TUT group of proteins (Figures 1A and 2B). Eight direct hydrogen bonds are found between the protein and the UTP molecule. The nucleobase and the sugar ring are recognized by every edge capable of engaging in a hydrogen bond (Figure 2C). Noticeably, the Cid1 protein is capable of sensing the presence of the two carbonyl oxygen atoms of the uridine base via the Asn171 and the His336 residues (Figure 2C). Until the present work, the contacts describing the pyrimidine recognition mode are based on the recognition of the two carbonyl oxygen atoms of the uridine base by the Cid1 protein.
by the TUT enzymes were focused on water-mediated interactions (Deng et al., 2005; Stagno et al., 2007b, 2010). Although these interactions allow the TUT proteins to select UTP, the mechanistic basis of uridine selectivity by the Cid1 protein has been left unclear. In particular, the capacity of the Cid1 protein to use both UTP and ATP for polymerization is unexplained. We show that mutations of the His336 into its equivalent residues in TUT or PAP proteins (H336L or H336N respectively and H336A) decrease the overall activity of the protein compared to the wild-type enzyme (WT; Figure 2D). The results of experiments performed in the presence of both UTP and ATP suggest that mutant proteins still incorporate UMP preferentially over AMP (P.M.-T. and S.T., data not shown; Rissland et al., 2007). We therefore conclude that the Histidine 336 certainly participates in the selection process since we observe reduced polymerizing activity when mutated. But this residue is not sufficient to fully explain the atypical property of the Cid1 protein.

In addition to the nucleobase edges, the 3'-oxygen of the sugar ring is stabilized via H-bonds with the Thr172 and the Asn171 residues of Cid1. However, the protein does not recognize the 2' position of the ring, differently than observed in the TUT structures, in which the equivalent residue (serine) contacts both the 3' and the 2'-oxygens (Deng et al., 2005; Stagno et al., 2007b, 2010). The triphosphate moiety is stabilized by hydrogen bonds with Lys193 and Lys197, both highly conserved in the Pol β family (Figure 1A), as well as with Ser90, Ser211 and Tyr212. A single magnesium ion is coordinated on one side by the catalytic amino acid Asp103 and, on the other side, by the phosphate-bound oxygen atoms, a situation encountered in the TUT4/UTP co-crystal structure. Another magnesium ion is found above the uridine base in a position that is likely to become occupied by the incoming RNA substrate. This ion is stabilized via π interaction forming an organomagnesium compound (Dougherty, 1996).

The Nucleotide Recognition Motif
Poly(A) and poly(U) polymerase enzymes have been shown to contain a highly conserved motif in the CD domain. This sequence, called the NRM, has been used to derive the nucleotide-binding specificity of a given polymerase (Martin and Keller, 2006).
The NRM sequences are divided into two groups: (1) group 1 for the canonical NT, which contains PAP enzymes, and (2) group 2 for the non-canonical NT to which the TUT enzymes and the Cid1 protein belong (Figure 1A). In the co-crystal structures of RET2, MEAT1 and TUT4 trypanosomal proteins, no direct contacts were observed between the amino acids of the NRM motif and the nucleotide base. Furthermore, the function of individual residues in the NRM loop is still unclear. In particular, the question of how the NRM sequence distinguishes between the PAP and the PUP/TUT groups is unanswered. The consposition of the NRM loops (our structure and from other PUP or PAP structures) indicates an excellent match within the PUP/TUT group and highlights a couple of clear differences with the PAP group (Figure 3A).

A specific hydrogen bond between the NE2 position of the Histidine residue 336 in the NRM sequence and the O4 carbonyl oxygen of the uracil base is found in our Cid1/UTP structure (Figures 2C and 3B). This additional bond ensures that the two oxygen atoms from the nucleobase are selected by specific interactions with the side chains of residues 171 and 336 (Figure 2C). It is striking to note that the Cid1 protein and the two human poly(U) polymerases Zcchc11 and Zcchc6, which have demonstrated poly(U) polymerase activity in vivo, also have a Histidine residue at the equivalent position (Figures 1A and 3A; Rissland et al., 2007; Heo et al., 2009). It would be interesting to see whether the two human proteins also have the capacity to use ATP as substrate in vitro and are able to select UTP over ATP as the Cid1 protein.

The next residue of particular interest in the NRM sequence and which has a differential localization in the two groups corresponds to the Phenylalanine 332 (Figures 3A and 3B). In the yeast poly(U) polymerase, this residue is an Alanine, while a larger residue corresponds to the Phenylalanine 332 (Figures 3A and 3B). In our structure, the Phenylalanine 332 is buried into a hydrophobic pocket, directly facing the uridine moiety and interact through hydrogen bonds (Stagno et al., 2007a). The Arg121 of TUT4, equivalent to Arg139 in the Cid1 protein, has a proven role in the catalytic activity of the TUT4 protein. The strong sequence conservation and similarities existing between the poly(U) polymerase involved in cellular RNA regulation (Trippie et al., 1998; Perumal and Reddy, 2002; Shen and Goodman, 2004) and the TUT enzymes associated with RNA processing mechanisms in mitochondria (Aphasizhev and Aphasizheva, 2011). Despite the very limited sequence homology existing between the Cid1 protein and the trypanosomal enzymes TUT4, RET 2, and MEAT1 (Figure 1A), their folds are highly related (Figure 4A). Structural superpositions point out multiple loops in the CAT and the CD domains that may have different orientations. Based on their electrostatic properties, these loops are likely to participate in the selectivity/stabilization of the RNA primer (Figure 4B).

The loop between β strand 4 and 5 (Region 1 in Figure 1A) is located above the UTP molecule in our structure (Figure 4C). This protein fragment is highly conserved in all the NT sequences (Figure 1A) and is also positively charged (Figure 4B). In the TUT4/UpU structure, the corresponding amino acids (residues 118-124 in the TUT4 sequence) are in proximity to the second uridine moiety and interact through hydrogen bonds (Stagno et al., 2007a). The Arg121 of TUT4, equivalent to Arg139 in the Cid1 protein, has a proven role in the catalytic activity of the TUT4 protein, which is likely to be linked to its role in substrate selection/stabilization (Stagno et al., 2007a, 2007b). Furthermore, upon UTP association, the corresponding residues in the MEAT1 protein structure move up to 7 Å toward the CD domain (Figures 4B and 4C; Stagno et al., 2010). The β sheet 4, which immediately follows this loop, contains the conserved Lysine in position 144 as well as several hydrophobic residues (Figure 1A). Mutation of Arg126 in the TUT4 protein (equivalent to Lys144 in Cid1) abolished TUT activity (Stagno et al., 2007b). In fact, the substitution of Arg126 to Alanine disrupted the salt bridge between Arg126 and Asp136 leading to inactivation of the TUT4 protein. The strong sequence conservation and the above-mentioned observations clearly indicate that this protein region will play a major role in regulating the activity of the Cid1 protein. We mutated the two most conserved residues, Arg139 and Lys144, to assess their importance in Cid1 activity. The Arg139 mutation strongly reduces the polymerization indicating that, as observed in the TUTase group, this
Figure 3. The NRM

(A) Superposition of the NRM loop structures from the Cid1 protein (shown in green), the yeast PAP (Protein Data Bank [PDB] code: 2Q66; in cyan), the trypanosomal TUT4 (PDB code: 2IKF; in orange) and the RET2 (PDB code: 2B51; in violet) enzymes. Important side chains are shown with ball-and-stick representation, with the corresponding protein color and label (labels in bracket correspond to PAP, TUT4 or RET2 sequence).

(B) Global view of the Cid1 structure along the axis of helix 4, showing its critical location between the CAT, the CD and the NRM loop. The Cid1 protein is colored as in panel 1B with the helix 4 shown in yellow. The catalytic Aspartates, Phe332 and His336 residues are shown in ball-and-stick representation (carbon colored in orange and magenta, respectively).

(C) Residue Phe332 is buried in a large hydrophobic cavity, lined by residues represented in ball-and-stick, labeled and colored in cyan. The pocket is adjacent to helix 4, which is proposed to communicate the association of the triphosphate group.

(D) Poly(U) polymerase assays. The Cid1 WT or the indicated mutants were incubated with radioactively labeled 17-mer oligonucleotides. The lane labeled Ct shows the RNA primer after incubation without enzyme. The three mutants (F332L, F332A, and F332D) have different levels of activity reflecting the importance...
Differently from the TUT4 protein, mutation of Lys144 resulted in a limited decrease of the polymerization level when compared to the WT enzyme suggesting that the catalytic mechanism of the Cid1 protein can function despite the absence of a stabilizing partner for Asp160 (Figure 4D).

A second loop, corresponding to residues 202–206, is found upstream of the conserved Tyr212, on which the UTP molecule of the modification. If the hydrophobic character is kept (F332L and to a lower degree F332A), the enzyme is still active. Once a charged residue is introduced (F332D), the activity is almost abolished.

The region 2 in our sequence alignment is located above the active site. The same region is involved in the poly(A) stabilization in the scPAP/RNA complex structure (shown as a cyan cartoon; PDB code: 2Q66) via hydrogen bonds and van der Waals interaction with residues 226 and 227. In the Cid1 structure, these residues are replaced by Proline and Tyrosine, respectively (shown as green ball-and-stick). Both residues are suitable for interactions with the growing poly(U) tail. The model of TUT4 (in orange) is shown to highlight the structural differences observed for this particular region between the TUT and the PAP enzymes. Magnesium ions are gray.

Atoms are shown as spheres, and colored according to their type. Magnesium ions are gray.
The absence of the substrate their case, two of the catalytic aspartates were bound to the phosphate oxygens and Aspartate 103 (Figure 2D). A similar moiety. Two magnesium ions are found in the pocket; one is coordinated by the catalytic Asp103 residue and the phosphate groups, as well as His336 and Asn171 and the uracil base of the UTP, most likely at the position of the RNA polymerase (Figure 4C and 4D). Balbo and Bohm, 2007). The structural similarities observed between the Cid1 and PAP loops suggest a similar mode of association between the Cid1 protein and its growing product. Migrations of Tyr205 into various residues highlight the role of this residue. It appears that a hydrophobic residue is needed in this loop to keep Cid1 active (Figures 4D and 4F). If the enzyme contains a charged residue like an Aspartate, then its ability to polymerize is significantly decreased (Figure 4E and Figures S2A and S2B). The decreased activity is observed with three RNA primers (a 17 nucleotide primer, a 15-mer poly(A) and a short poly(A) and a short UpU substrate).

**DISCUSSION**

The present structure of fission yeast Cid1 protein reveals the molecular basis for UTP binding by this particular enzyme. The active site is located at the bottom of the cleft separating the CAT and CD domains, with the three aspartic acid residues facing the phosphate moiety of the bound UTP (Figures 2A and 2B). Stabilization of the UTP is performed by several H-bonds, principally between Lys193 and 197 and the phosphate groups, as well as His336 and Asn171 and the uracil moiety. Two magnesium ions are found in the pocket; one is coordinated by the catalytic Asp103 residue and the phosphate-bound oxygen atoms and the other positioned above the uracil base of the UTP, most likely at the position of the RNA primer. In our structure, a single magnesium ion coordinates the phosphate oxygens and Aspartate 103 (Figure 2D). A similar situation was encountered in the RET2 structure although, in their case, two of the catalytic aspartates were bound to the magnesium ion (Deng et al., 2005). The absence of the substrate DNA certainly prevented the stable association of the second product. The affinity constant for the UTP compound of 1 and 45 μM have shown that the TUTase proteins TUT4 and RET1 have an affinity constant for the UTP compound of 1 and 45 μM respectively (Aphashicheva et al., 2004; Stagno et al., 2007b). The proposed hinge region between the central and the catalytic domains suggests that the Cid1 protein translocates its RNA substrate by adjusting the accessibility of the active site-bound UTP (Figures 3B and 3D). We suggest that Pro32 plays a critical role in the communication between the CAT domain closure and the selection of the nucleobase by the NRM loop. These movements can be observed on their own, as seen in the apo and UTP-bound RET2 structures, or through a protein partner. The recruitment of the Zcc1 protein by the Lin28A protein could reflect such a need (Hagan et al., 2005). When compared with the yeast PAP enzyme, the Cid1 protein lacks the additional domain found above the active site cleft, which globally reorients after catalysis. The reorganization of the PAP RNA binding sites is, at least partially, mimicked by movements of the two Cid1 domains. This conformational flexibility is likely to dictate the ratio between the processive and the distributive character of the enzyme. These phenomena are noteworthy to understand since the length of the poly(U) polymerase-added poly(U) tail determines the fate of the substrate RNA. With the help of our atomic model and of our activity assays of mutated proteins, a more comprehensive analysis of the PUP-dependent polyuridylation process is achieved.

**EXPERIMENTAL PROCEDURES**

**Expression, Purification, and Mutagenesis**

The nucleotide sequence of the Cid1 protein encoding residues 40-377 was amplified by PCR from Schizosaccharomyces pombe genomic DNA kindly provided by M. Bühler, FMI, Basel and sequenced. The ORF was cloned into pET32a (derived from pET24a) for expression in E. coli as a fusion protein with N-terminal His6 and MBP tag followed by a TEV cleavage site. The recombinant protein was overexpressed in Escherichia coli BL21-Star cells, grown in TB media at 37°C for 6 hrs followed by overnight induction at 18°C with 0.1mM isopropyl-β-D-thiogalactopyranoside (IPTG). Induced cells were harvested by centrifugation, resuspended in buffer A (20 mM Phosphate buffer pH 7.5, 300mM NaCl, 25mM Imidazole, 0.1% Triton X-100, DNase 1ug/ml, Lysosyme 1 μg/ml, 5 mM β-mercaptoethanol), supplemented with protease inhibitors PhenylMethylSulfonyl Fluoride (PMSF) 1mM, leupentin 1 μg/ml, and pepstatin 2 μg/ml.

Cells were lysed using an emulsiflex system (AVESTRA) and cleared by centrifugation at 15,000 rpm for 35 minutes at 4°C. The soluble fraction was subjected to an initial affinity purification using a chelating Hitrap FF crude column (GE Healthcare) charged with Ni2+ ions. The protein was eluted with 250 mM imidazole and desalted against 20 mM phosphate buffer pH 7.5, 300 mM NaCl, 25 mM imidazole, and 5 mM β-mercaptoethanol. The Tobacco Etch Virus protease was added at a ratio of 1:50 (TEV:protein) and the reaction was performed at 16°C for 10 hours. Cleaved protein was separated from the tag, the protease and the contaminants by reapplication to the Hitrap column. After concentration, remaining impurities were removed by gel filtration chromatography with a Superdex 200 column (GE Healthcare), in a buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM β-mercaptoethanol, and 10% (v/v) glycerol. Samples were concentrated (Amicon) to 9.5 mg/ml.
and used immediately or stored at -80°C. All protein samples eluted from the gel filtration column as monomers and were homogeneous when analyzed by SDS-PAGE.

All Cid1 point mutants were prepared by site-directed mutagenesis using the Quick-Change mutagenesis kit from Stratagene according to manufacturer instructions. Point mutations were verified by DNA sequencing prior to large-scale expression. The oligonucleotides used for the mutations of the catalytic Aspartate residues 101 and 103 (called Cid-AMA), the Arginine 139, the Lysine 144, the Tyrosine 205, the Phenylalanine 332 and the Histidine 336 are shown in Table S1 of supplemental data.

Protein Crystallization and X-Ray Data Collection
Cid1 was crystallized in two steps. Crystals were produced using the sitting-drop vapor diffusion technique at 18°C, with a reservoir solution containing 0.1 M Imidazole/MES (equal ratio) pH 6.1, 12%–20% Glycerol, 6%–10% polyethylene glycol (PEG) 4000 and 90–144 mM of a halogen salt (NaF, NaCl, NaBr). The protein concentration was 9.5 mg/ml. Due to a high level of nucleation, a “pseudo-batch” procedure was performed at 18°C to obtain diffraction quality crystals. In this case, crystallization conditions were further refined. Equal volumes of the gel filtration buffer and the crystallization condition mentioned before were mixed in the reservoir. Crystals appeared in two days and grew to a maximum size of around 0.15 x 0.15 x 0.1 mm in 1 week. Crystals of Cid1 in complex with UTP were obtained by soaking native crystals for 14 hours with 2.5 mM of UTP supplemented with 5 mM of MgCl₂. Crystals were passed into polypropylene cryo oil (Hampton Research) prior to flash freezing in liquid nitrogen.

Diffraction data for the UTP-soaked protein crystal were collected at Beamline ID14-4 at the European Synchrotron Radiation Facility (ESRF, Grenoble). Diffraction was observed to about 2.3 Å resolution. Native crystals were soaked with numerous heavy metal compounds at a range of concentrations and incubation times. We collected a highly redundant data set from a crystal soaked for 7 days in the crystallization solution supplemented with 1 mM KAuCl₄. A single anomalous diffraction (SAD) data set to a resolution of 2.7–2.8 Å was collected at the Beamline PX11 at the Light Source (SLS, Villigen-Paul Scherrer Institute).

Structure Determination and Refinement
The structure of Cid1 was determined by SAD method using the anomalous signal of the gold atom. Briefly, the data set was indexed with the XDS package (Kabsch, 2010). Heavy atom location was performed with the Shelx suite using a maximum size of around 0.15 Å. Crystallographic solution finding methods included the Aufab program (Brüger et al., 2009), followed by molecular replacement using the Phaser program (McCoy et al., 2007). The final model was refined to 2.3 Å resolution with 99.7% of the residues in the preferred and allowed regions of the Ramachandran plot. The final model was refined to 2.3 Å with 98.7% of the residues in the preferred and allowed regions of the Ramachandran plot.

Poly(U) Polymerase Assays
For each experiment, two quantities of enzymes were used (50 and 150 ng) and the reaction volume was 3 μl. The reaction mixture contained 50 μl of 10 mM Tris-Cl, pH 7.9, 50 mM KCl, 1 mM MgCl₂, 0.5 mM UTP for 10 min at 30 °C. Reactions were stopped with an equal volume of 100 mM EDTA and 2% sodium dodecyl sulfate (SDS). Extended RNA primers were extracted with phenol–chloroform, precipitated with ethanol and resuspended in 90% formamide loading buffer except for the reactions with the UpU substrate which were directly inactivated. Products were denatured at 95°C and separated by gel electrophoresis in a 12% (or 25% for the UpU) polyacrylamide/7M urea gel.

Accessibility Numbers
Atomic coordinates and structure factors for the Cid1/UTP model have been deposited in the Protein Data Bank (http://www.pdb.org) with the accession code 4EPT.

Supplemental Information
Supplemental Information includes two figures, one table, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cell.2012.04.006.

Acknowledgments
We wish to acknowledge Marc Bühl for the generous gift of Schizosaccharomyces pombe genomic DNA. We also acknowledge the staff of X06SA at the Swiss Light Source and of ID14-4 at the ESRF for developing excellent beamlines and supports with data collection. We wish to thank all members of the Stéphane Thore lab for critical reading, helpful discussions and technical advice. We are grateful for financial contributions from the Ernest Bonnici and the Ernst and Lucie Schmidheiny foundations together with the R’equip subsidy from the Swiss National Science Foundation (n 31003A_128787), used to acquire theBiostructural platforms equipments. This work was partially supported by individual research grants from the Swiss National Science Foundation to S.T. (n 31003A_124909 and n 31003A-140924).

Received: February 15, 2012
Revised: April 17, 2012
Accepted: April 17, 2012
Published online: May 17, 2012

References


Supplemental Information

Functional Implications

from the Cid1 Poly(U) Polymerase Crystal Structure

Paola Munoz-Tello, Caroline Gabus, and Stéphane Thore

Inventory of Supplemental Information

I. Supplementary Data

A. Figure S1. Fo – Fc difference electron density map of the bound UTP molecule, related to Figure 2.

B. Figure S2. Poly(U) polymerase assays with different RNA primers, related to Figure 4.

C. Table S1. Cid1 mutagenesis primers.

II. Supplementary Experimental Procedure

Cloning the Cid1 gene from genomic DNA.
Supplemental Data

Figure S1, related to Figure 2:

View of the Fo – Fc difference density map of the bound UTP molecule. The Cid1 protein is shown in green (CD domain) and orange (CAT domain). The NRM loop is colored in magenta. The NRM amino acids and the UTP molecule are shown in ball-and-stick representation. Two magnesium ions are depicted as gray sphere. The difference electron density map was calculated using the refined protein chains and the magnesium ions as models and is shown as gray mesh (contoured at 2.7σ).
**Poly(U) polymerization assays with different RNA primers.** (A) UTP polymerization by mutants R139A, Y205A and K144A compared to the wild-type Cid1 protein with a UpU RNA template. (B) UTP polymerization by mutants Y205F, Y205V and Y205D compared to the wild-type Cid1 protein using the same primer as in panel A. (C) UTP polymerization by mutants R139A, Y205A and K144A compared to the wild-type Cid1 protein with a polyA tract (15 nt). (D) UTP polymerization by mutants Y205F, Y205V and Y205D compared to the wild-type Cid1 protein using the same primer as in panel C. Equivalent amount of proteins were used in all the reactions (50 and 150 ng) and enzymatic conditions were as described in the experimental procedures section. Products were separated on a 25 and 12% polyacrylamide/7M urea gel (panel A and B and panel C and D, respectively).
<table>
<thead>
<tr>
<th>Cid1</th>
<th>Forward Primer (5’ → 3’)</th>
<th>Reverse Primer (5’ → 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMA</td>
<td>CTTAAAATTCGGCTATGCTTTGTGGCTTATG</td>
<td>GATAATGGAATTTGCTCCCCTTTGTAAAA</td>
</tr>
<tr>
<td>R139A</td>
<td>TTTTACAAGGGGCAAGATCCCTATTATC</td>
<td>GATAATGGAATTTGCTCCCCTTTGTAAAA</td>
</tr>
<tr>
<td>K144A</td>
<td>CAAGAATTCCATATACTGAGATTACATGATACG</td>
<td>CGTAGACGTAAATGCGATATGGAATTCTTG</td>
</tr>
<tr>
<td>Y205A</td>
<td>GCAAATCAACTCTCGCTTTTGAACAATCTTTCC</td>
<td>GAAAGAGTTCAAAAGCAAGAAGATTGATTTTG</td>
</tr>
<tr>
<td>Y205D</td>
<td>GCAAATCAACTCTCGCTTTTGAACAATCTTTCC</td>
<td>GAAAGAGTTCAAAAGCAAGAAGATTGATTTTG</td>
</tr>
<tr>
<td>Y205F</td>
<td>GCAAATCAACTCTCGCTTTTGAACAATCTTTCC</td>
<td>GAAAGAGTTCAAAAGCAAGAAGATTGATTTTG</td>
</tr>
<tr>
<td>F332A</td>
<td>GCGATTGAGATCTGCGAGATTTTCCACATAATG</td>
<td>CATTATGTAATCTCTGCAAGATCTCTCAATCG</td>
</tr>
<tr>
<td>F332D</td>
<td>GCGATTGAGATCTGCGAGATTTTCCACATAATG</td>
<td>CATTATGTAATCTCTGCAAGATCTCTCAATCG</td>
</tr>
<tr>
<td>F332L</td>
<td>GCGATTGAGATCTGCGAGATTTTCCACATAATG</td>
<td>CATTATGTAATCTCTGCAAGATCTCTCAATCG</td>
</tr>
<tr>
<td>H336A</td>
<td>CTGGAGGATCTTGAAGATCTTCCACATAATG</td>
<td>CCTATGTAATCTCTGCAAGATCTCTCAATCG</td>
</tr>
<tr>
<td>H336L</td>
<td>CTGGAGGATCTTGAAGATCTTCCACATAATG</td>
<td>CCTATGTAATCTCTGCAAGATCTCTCAATCG</td>
</tr>
<tr>
<td>H336N</td>
<td>CTGGAGGATCTTGAAGATCTTCCACATAATG</td>
<td>CCTATGTAATCTCTGCAAGATCTCTCAATCG</td>
</tr>
</tbody>
</table>

Table S1: Cid1 mutagenesis primers.
Supplemental Experimental Procedures

Cloning the Cid1 gene from the schizosaccharomyces pombe genomic DNA

The DNA construct coding for S. pombe Cid1 was synthesized by nested PCR, using S. pombe genomic DNA (kind gift from M. Bühler). The two introns found in the Cid1 gene were removed using polymerase chain reaction and the following primers:

Forward primer (Nterm): 5' - CAGTACATGTCGACAAAGGAATTACGAAAGTTTTGC - 3'
Forward primer (intron1): 5' - CTAAACGAATATCCCTGATGCTGATTGTTAG - 3'
Forward primer (intron2): 5' - GCTTATAGCTGAAGGATTGAAGGAAAATTTTTAC - 3'
Reverse primers (intron1): 5' - CTACCAATTCAGCATCGGGATATTCTGTAA AG - 3'
Reverse primer (intron2): 5' - GTAAAAATTTTCTCTTAAATCTCTTTGAGCTATAAAC - 3'
Reverse primer (Cterm): 5' - GCCGCTCGAGTCAGGCCTCATTAAATGAATC - 3'

The primers were designed in order to suppress the predicted disorder N-terminal (region 1-39) and C-terminal (378-405) regions of the protein. The PCR product was cloned into pST18 (derived from pET42) for expression in E. coli as a fusion protein with N-Terminal His₉ and MBP tag followed by a TEV cleavage site.
**1.11.5 Cid1 Substrate/RNA Complex Crystallization**

In order to better understand how Cid1 enzyme catalytic cycle is taking place, we tried to obtain the crystal structures of several binary and ternary complexes. Different compounds and concentrations were used for soaking and co-crystallization of the native crystals (Fig. 13E). The different trials performed are summarized on table 2 (Table 2). Furthermore, multiple crystallization conditions were also tested (Fig. 17). Unfortunately, all these trials were unsuccessful since none of the electron density maps calculated using the dataset of these soaked crystals had any visible positive density in the active site indicating the presence of a bound ligand (Fig. 18).

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Soaking concentration</th>
<th>Soaking times</th>
<th>Co-crystallization concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>10mM, 5mM, 2.5mM</td>
<td>15min, 2h, 4h, 6h, 8h, 16h, 2 days</td>
<td>1mM, 0.75mM, 0.5mM, 0.25mM</td>
</tr>
<tr>
<td>ATP/ApU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP/ApCpp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP/UpA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP/AAAAAU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ApCpp</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ApCpp/ApU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMPnPP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMPnPP/UTP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UMPnPP/UpU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTP/UpU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UpA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UpU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ApU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTP/AAAAAU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAAAAAAAAAAU</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>AAAAAAAAAAAAAA</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>UUUUUUUUUUUUUUUU</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>17-mer RNA</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

*Table 2: Different ligand soakings and co-crystallization trials. The X corresponds to trials not tested.*

Based on a publication where the authors could observed how a phosphodiester bond was made by collecting and analyzing data from crystals soaked for different times (Nakamura et al. 2012), we tested if the protein was still active in several crystallization buffers and then tried soakings at different times and concentrations in that particular condition (Fig. 19). In
spite of our efforts, we could not observe any catalytic activity inside the crystal or any other ligand in the active site.

Figure 17: Different crystallization conditions tested for co-crystallization and soaking of ligands.

Figure 18: Cid1 electron density maps zoomed into a none-ligand bound active site. Tyrosine 205 and Histidine 336 are shown in the active site to orient the protein. The positive and negative $F_0 - F_c$ density maps are shown as mesh and colored in green and red respectively. $2F_0 - F_c$ density map is colored in blue.
Figure 19: Cid1 in vitro polymerization assays with radiolabeled ApU RNA in different crystallization buffers. Experiments were performed at 150ng of Cid1 and a radiolabeled 2-mer RNA substrate with 10mM of UTP (U) or ATP (A). Each reaction mix was incubated for 20min at 30°C. - , no protein control; Buf1, NEB buffer used for polymerization reaction (Materials and Methods section 4.1.4); Buf2, 0.1M CHES pH 9.0 and 6% PEG 3350; Buf3, 0.1M Imidazole/MES pH 6.1, 15% Glycerol/PEG 4000 and 0.06 M NaI, NaBr, NaF; Buf4, 0.1M Hepes pH 7.5, 6% 2-propanol, 0.2M NaCl and Buf5, 4% Tacsimate pH 8.0, 9% PEG 3350. All the buffers were supplemented with 5mM MgCl₂.

We then decided to mutated one residue from the catalytic triad (residue D160) in an attempt to stabilize binary or ternary Cid1 complexes as successfully done in the PAP/ATP/RNA ternary complex (Balbo and Bohm 2007). Protein expression and purification were performed as for the WT protein and D160A mutant elutes as a monomeric protein in the gel filtration column (Fig. 20). The WT Cid1 crystallization condition was supplemented with 5mM of TCEP leading to larger crystals (Fig. 21A) and new crystallization conditions were also tested (Fig. 21B). The same binary and ternary complexes performed for WT Cid1 were tested in soaking and co-crystallization trials (Table 2). We tested approximately 200 crystals and only one of them had a positive density inside the active site, corresponding to an ApU molecule (Fig. 22). Details of the structure solving along with the mutational and functional analysis of Cid1 based on this crystal structure were published in the journal Nucleic Acids Research and are shown in the following chapter (section 4.1.6).
Figure 20: Gel filtration chromatographic profile of the D160A Cid1 protein mutant purification.

Figure 21: Different D160A mutant crystal hits. (A) D160A mutant crystallization conditions supplemented with TCEP. (B) Additional D160A crystallization conditions used for soaking or co-crystallization.
Figure 22: Cid1 electron density map zoomed into a ligand-bound active site. Tyrosine 205 and Histidine 336 are shown in the active site to orient the protein. The positive and negative $F_0 - F_c$ density maps are shown as mesh and colored in green and red respectively. $2F_0 - F_c$ density map is colored in blue.
In order to trap other Cid1 binary and ternary complexes, we mutated one of aspartatic residues of the catalytic triad (D160A) to inactivate the enzyme as successfully performed for the poly(A) polymerase/ATP/RNA ternary complex (Balbo and Bohm 2007). We then solved the crystal structure of Cid1 in complex with a non-hydrolysable UTP (UMPnPP) and a pseudo-product ApU. This pseudo-product mimics the post-catalysis reaction state since Cid1 targets are polyadenylated. This atomic model gave us deeper molecular insights on the likely catalytic mechanism and substrate preference of Cid1 enzyme. We first confirmed the RNA selectivity of the enzyme by showing a direct contact of aspartate 101 with the 2’-OH group of the RNA substrate. We also highlighted the enzyme’s strong preference for uridylated substrates and emphasized the importance of several other key residues for RNA accommodation and Cid1’s activity. Taken together, our study allows us to propose a switching mechanism of Cid1 activity, which changes from a distributive to a processive mode, potentially allowing the addition of two distinct U-tails with different sizes.
A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure

Paola Munoz-Tello, Caroline Gabus and Stéphane Thore*

Department of Molecular Biology, University of Geneva, Geneva, 1211, Switzerland

Received September 18, 2013; Revised November 14, 2013; Accepted November 16, 2013

ABSTRACT

The addition of uridine nucleotide by the poly(U) polymerase (PUP) enzymes has a demonstrated impact on various classes of RNAs such as microRNAs (miRNAs), histone-encoding RNAs and messenger RNAs. Cid1 protein is a member of the PUP family. We solved the crystal structure of Cid1 in complex with non-hydrolyzable UMPNPP and a short dinucleotide compound ApU. These structures revealed new residues involved in substrate/product stabilization. In particular, one of the three catalytic aspartate residues explains the RNA dependence of its PUP activity. Moreover, other residues such as residue N165 or the β-trapdoor are shown to be critical for Cid1 activity. We finally suggest that the length and sequence of Cid1 substrate RNA influence the balance between Cid1’s processive and distributive activities. We propose that particular processes regulated by PUPS require the enzymes to switch between the two types of activity as shown for the miRNA biogenesis where PUPS can either promote DICER cleavage via short U-tail or trigger miRNA degradation by adding longer poly(U) tail. The enzymatic properties of these enzymes may be critical for determining their particular function in vivo.

INTRODUCTION

Cytoplasmic messenger RNA (mRNA) homeostasis is a complex phenomenon in which the balance between mRNA stability and mRNA degradation is tightly monitored. Eukaryotic species, from plants to mammals, rely heavily on the poly(A) tail to control the mRNA degradation process (1–2). Recently, a long-known family of proteins, named poly(U) polymerases (PUPS), was shown to be an unexpected player in mRNA homeostasis in various eukaryotes (3–6). The PUPS add poly(U) tails to mRNAs, which appear to modulate the rates and/or the directionality of the mRNA degradation process (4,7–9). Furthermore, members of the PUP family also regulate the production of histone mRNAs, microRNAs, siRNAs or U6 snRNA in the nucleus (6,10–17).

The PUP enzymes contain a nucleotidyl transferase domain, which has been characterized structurally (18–23). These structures show that the nucleotide triphosphate and the substrate RNA are bound in a space between the catalytic domain and the central domain (21,24). The PUP Cid1 protein was identified in Schizosaccharomyces pombe and was shown to have the capacity to use both uridine triphosphate (UTP) and adenosine triphophate (ATP) as substrates for polymerization in vitro, while apparently being specific for UTP in vivo (25–26). This is surprising, as other members of the PUP family were depicted to be highly specific for UTP (18–20,27–28). Residue H336 found in the nucleotide recognition motif (NRM) loop has been shown to be mainly responsible for determining the specificity of the nucleotide recognized in vitro and in vivo (21–23). Additional charged residues were shown to be involved in substrate RNA association, although their exact function was not entirely characterized (21–22).

To further understand the mechanism behind RNA recognition and elongation, we determined the structure of Cid1 bound to a non-hydrolyzable nucleotide UMPNPP and bound to its minimal pseudo-product ApU. These structures revealed new key residues for the substrate/product recognition. In particular, one of the three catalytic aspartate residues interacts with the 2'-OH group of the pseudo-product-bound dinucleotide, thus explaining the RNA specificity displayed by this class of enzymes. Moreover, residue N165 interacts with the adenosine base in our pseudo-product-bound crystal structure. This interaction is essential for the enzymatic reaction. Finally, we demonstrate that the polymerizing capacities of Cid1 are modulated by a highly flexible loop named the β-trapdoor (residues 310–322) and by the residue K144 when UTP is used as a substrate (22). Our study thus highlights the structural basis behind several unique...
properties of the PUP enzyme family: properties mediated by a combination of specific residues and conserved secondary structures. We speculate that the specific type of activity that PUP enzymes display may determine their involvement in particular biochemical pathways and require factors to regulate their polymerizing activity.

MATERIALS AND METHODS

Expression and purification

*A. pombe* Cid1 protein (residues 40–377) was expressed and purified as described previously (21). Briefly, Cid1 was expressed in *Escherichia coli* BL21 Star cells (Novagen) as a His-MBP fusion protein and purified using HiTrap FF crude column (GE Healthcare). After cleavage with the Tobacco Etch Virus protease (ratio of 1/50) at 16°C, the sample was reloaded onto the HiTrap column to remove the tag, the protease and the contaminants. The purified protein was finally applied on a Superdex 200 column (GE Healthcare), in a buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl, 1 mM MgCl₂, and 10% (v/v) glycerol. Fractions containing the protein were pooled and concentrated (30 000 MWCO Amicon) and 10% (v/v) glycerol. Fractions containing the protein were loaded onto a HiTrap FF crude column (GE Healthcare). After cleavage with the Tobacco Etch Virus protease (ratio of 1/50) at 16°C, the sample was reloaded onto the HiTrap column to remove the tag, the protease and the contaminants. The purified protein was finally applied on a Superdex 200 column (GE Healthcare), in a buffer containing 20 mM Hepes (pH 7.5), 150 mM NaCl, 1 mM β-mercaptoethanol and 10% (v/v) glycerol. Fractions containing the protein were pooled and concentrated (30 000 MWCO Amicon) to 10 mg/ml and stored at −80°C. All protein samples eluted from the gel filtration column as monomers. The same expression and purification procedure was performed for all the Cid1 mutants.

**Mutagenesis**

Cid1 point mutants were prepared by site-directed mutagenesis using the Quick-Change mutagenesis kit from Stratagene according to manufacturer’s instructions. The Cid1 Δ310–322 deletion mutant was prepared by Nested-polymerase chain reaction. All the constructs were verified by DNA sequencing. The oligonucleotides used for the mutations are shown in Supplementary Table S1.

**Protein crystallization and X-ray data collection**

D160A mutant protein was used at 10 mg/ml. Crystals were produced using the sitting-drop vapor diffusion technique at 18°C, with a reservoir solution containing 0.1 M imidazole/MES (equal ratio) (pH 6.1), 12–20% glycerol, 6–10% polyethylene glycol (PEG) 4000, 126 mM of a halogen salt (NaI, NaBr, NaF in equal ratio) and 6–10% polyethylene glycol (PEG) 4000, 12–20% glycerol, 10 mM TCEP [tris(2-carboxyethyl)phosphine] as an additive. Crystals appeared in 2 days and grew to a maximum size of around 0.25 × 0.25 × 0.2 mm in 1 week. Crystals of D160A mutant were soaked for 16 h with 2.5 mM UMPNPP and UTP or with 2.5 mM of ApU and ATP, in both cases supplemented with 5 mM of MgCl₂. Crystals were passed into the crystallization solution supplemented with 10% glycerol before flash freezing in liquid nitrogen.

Diffraction data for the UMPNPP-soaked and ApU-soaked protein crystals were collected using the Beamline ID14-4 at the European Synchrotron Radiation Facility (ESRF, Grenoble). Complete data sets for UMPNPP- and ApU-soaked crystals were collected to 1.9 and 1.94 Å resolution, respectively, and the statistics are reported in Table 1.

**Structure determination and refinement**

The structures of the D160A mutant in complex with UMPNPP or in complex with ApU were determined by molecular replacement with the Cid1 protein structure as a search model (PDB code 4EP7). The data sets were indexed with the XDS package (29). Phasing was done using the program Phaser from the CCP4 package (30) and the models were refined using the routine phenix.refine from the Phenix program (31). A large unidentified density was visible in both 2F国防 F and the F₀ – F electron density maps where an UMPNPP or an ApU molecule was built (Supplementary Figure S1A and B). Atomic coordinates for the UMPNPP and ApU molecules were taken from the HIC-Up database (32) and fitted in the empty electron density map using the program COOT (33). The final Cid1/UMPNPP and Cid1/ApU models were refined to a resolution of 1.9 and 1.94 Å, respectively.

Both crystal structures contain two molecules of Cid1 in the asymmetric unit. The Cid1/UMPNPP structure consists of residues 38–377 from the Cid1 protein, 2 UMPNPP compounds, 4 magnesium ions and 3 bromide.

<table>
<thead>
<tr>
<th>Table 1. Data collection and refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
</tr>
<tr>
<td>Crystal</td>
</tr>
<tr>
<td>Cid1</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Temperature (K)</td>
</tr>
<tr>
<td>Maximum resolution (Å)</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>Unit cell dimensions (Å)</td>
</tr>
<tr>
<td>Number of observations</td>
</tr>
<tr>
<td>Number of unique reflections</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>Data completeness (%)</td>
</tr>
<tr>
<td>R(s)</td>
</tr>
<tr>
<td>Refinement</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Number of atoms</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Ligand</td>
</tr>
<tr>
<td>Magnesium ions</td>
</tr>
<tr>
<td>Bromine ions</td>
</tr>
<tr>
<td>Water molecules</td>
</tr>
<tr>
<td>Rwork (%)</td>
</tr>
<tr>
<td>Rfree (%)</td>
</tr>
<tr>
<td>RMSD from ideal geometry</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
</tr>
<tr>
<td>Bond angles (degrees)</td>
</tr>
<tr>
<td>Mean B (Å²)</td>
</tr>
<tr>
<td>Protein chain</td>
</tr>
<tr>
<td>Nucleotide/ions</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>Residues in favored region of the Ramachandran plot (%)</td>
</tr>
<tr>
<td>Residues in allowed region of the Ramachandran plot (%)</td>
</tr>
</tbody>
</table>
ions present in the crystallization condition. The Cid1/ ApU model contains the same Cid1 sequence (residues 38–377 for molecule A and residues 39–377 for molecule B). 2 ApU dinucleotides, 3 magnesium ions and 3 bromide ions. For both models, loops containing residues 109–115 and 309–322 were not visible in our electron density maps.

Poly(U) polymerase assays
RNA primers (U15, A15 or ApU) were labeled using T4 polynucleotide kinase (NEB) and γ-32P-ATP using manufacturer’s protocol. Labeled RNA substrates (10 pmol) were incubated with 5 or 15 pmol of Cid1 proteins [wild-type (WT) or mutant] in 25 μl of 10 mM Tris-Cl (pH 7.9) containing 50 mM NaCl, 10 mM MgCl2, 1 mM DTT and 0.5 mM UTP (or ATP when indicated) for 40 min at 37°C. Reactions were stopped with an equal volume of 100 mM EDTA and 2% sodium dodecyl sulfate. Extended primers were extracted with phenol-chloroform, precipitated with ethanol and resuspended in 90% formamide loading dye. Products were denatured at 95°C for 2 min and separated by gel electrophoresis in a 12% (or 25% for the ApU substrate) polyacrylamide gel.

Electrophoretic mobility shift assay
The indicated amount of proteins was mixed with the labeled RNAs in 20 mM Hepes (pH 7.5) containing 150 mM NaCl, 5 mM MgCl2, 1 mM β-mercaptoethanol, 10% glycerol and incubated at 37°C for 10 min. Following incubation, 2 μl of loading dye (0.1% bromophenol blue, 0.1% xylene cyanol, 180 mM Tris-boric acid and 50% glycerol) was added, and the samples were immediately loaded on a 10% native polyacrylamide gel in 0.5 x TBE buffer (45 mM Tris-boric acid, 1 mM EDTA). The gel was then run at 300 V for 120 min at 4°C and exposed in a cassette with an X-ray film (Fujifilm) for 30 to 45 min. Following quantification with ImageQuant software (GE Healthcare), affinity constants were calculated using the non-linear regression with one site-specific binding option from GraphPad Prism 6 program (GraphPad).

For comparing directly the RNA binding efficiency of various mutated Cid1 proteins, we used a unique concentration of protein (WT or mutant) and measured complex formation with a given probe (U15 or A15). Then, we quantified and normalized the amount of complex obtained with a given probe (U15 or A15). We used these two fixed concentrations of protein to maximize complex formation based on WT Cid1 Kd value for a given probe and to take into account the lower solubility observed with mutant proteins K144A and F332A. After loading and running the sample as indicated previously, we quantified our signal, normalized it against WT signal and plotted it as histogram. Individual binding reactions were performed three times in triplicates.

RESULTS AND DISCUSSION

Overall structure of the Cid1/UMP PPP and Cid1/ApU complexes
We mutated one residue from the catalytic triad (residue D160) in an attempt to stabilize the UMP PPP- or the ApU-bound Cid1 complexes as successfully done in the poly(A) polymerase (PAP)/ATP/RNA ternary complex structure (24). Here, we report the atomic structures of the D160A Cid1 mutant in complex with the non-hydrolysable UMP PPP and the D160A Cid1 mutant in complex with a minimal pseudo-product (ApU) (Figure 1A and B; see X-ray statistics in Table I). Both ligands were built using unbiased Fo-Fc Fourier difference electron density map calculated with the protein model only after a round of simulated annealing (Supplementary Figure S1).

UMP PPP mode of binding is similar to our Cid1/UTP complex (21) confirming that D160A mutation does not induce any conformational changes in the overall protein fold or in the substrate nucleotide recognition by the protein. The only difference is the absence of the first magnesium ion necessary for the nucleolytic attack, as previously observed in the PAP/ATP/RNA crystal structure (Figure 1A) (24).

The bound ApU molecule should be considered as a mimic of the post-catalysis reaction state as defined by Stagno et al. (27). As such, ApU is considered as a pseudo-product of the Cid1-mediated reaction. The pseudo-product ApU is representative of the in vivo situation encountered by Cid1 protein as, in S. pombe, Cid1 natural mRNA substrates are polyadenylated (Figure 1B) (26). In the ApU-bound Cid1 structure, the uridine nucleotide is recognized as in the UMP PPP and the UTP-bound Cid1 structures (21–23) (Figure 1C). The adenine base of the bound ApU molecule is stacked on the uridine base, a situation reminiscent of the TUT4/UplU structure (Figure 1C) (27). Several Cid1 residues are involved in the stabilization of the adenosine nucleotide. First, its ribose 2′-OH group establishes a direct hydrogen bond with the catalytic residue D103 (Figures 1C and D). Although Cid1 was previously classified as an RNA-specific nucleotidyl transferase, the basis for such specificity was not described (8). We suggest that the catalytic residue D103 stabilizes the proper type of substrate near the catalytic site by detecting the presence of a 2′-hydroxyl group (Figure 1D). This mechanism of RNA detection is probably shared across the entire family, as a similar interaction is found in the TUT4/UplU crystal structure (27). Second, residue D103 and the ApU molecule have another water-mediated interaction with the N3 position of the adenine, likely used to stabilize the incoming RNA substrate before the nucleotidyl transfer reaction (Figure 1C and D). The adenine is further stabilized via additional water-mediated interaction between its N6 position and residue E333 (Figure 1C and D). None of these water-mediated interactions confer nucleotide specificity on the 3′-end of the substrate RNA, as expected for a template-independent polymerase (8). In the ApU complex structure, the N1 position of the bound adenine is recognized via a specific hydrogen bond with the OD1 position of residue N165. In
**Trypanosoma brucei** TUT4/UpU structure, the equivalent residue is an arginine (R141), which cannot contact the uracil base because it forms a salt bridge with residue E300 (E333 in Cid1) (27). However, if we superpose the TUT4/UpU and Cid1/ApU structures, the N3 atom of the uracil base in the trypanosomal structure is located at nearly the same place as the N1 atom of the adenosine in our structure (Figure 2A). Therefore, we suggest that N165 is a key residue for binding to the 3' end of the RNA substrate either before or after the Cid1 uridylation reaction. This proposal is valid whether this substrate RNA contains a 3' end adenosine nucleotide, as expected for the native poly(A)-containing substrates in *S. pombe*, or whether the RNA went through at least one round of polyuridylation (Figure 2A). Our proposal can be extrapolated to any RNA 3' end base, although it will require the nucleotide tautomeric form with a hydrogen-bonded N1 atom to establish a proper bond with N165 OD1 position. Finally, the TUT/PUP superposition also shows that residue R139 in Cid1 is not in contact with the pseudo-product ApU (Figure 2B). We presume that the two enzyme groups may have evolved toward slightly different RNA recognition/stabilization modes, most likely reflecting their specific enzymatic properties.

### RNA binding properties

Cid1 has been classified as a non-canonical template-independent nucleotidyl transferase (3,7–8,21–23,35). Previous studies have also shown that the overall activity of Cid1 is unexpectedly high when presented with U15 (26). To further understand the Cid1 enzymatic properties, we generated various Cid1 mutants driven by the analysis of our Cid1/ApU crystal structure. We then measured their binding properties for two different substrates (A15 and U15) in the absence of any polymerizing activity to avoid interference from the newly synthesized tail (Figure 3 and Supplementary Figure S2).
substances—mimics of a short poly(A) tail (A15) or a newly added U-tail (U15; Figure 3A and B; Supplementary Figure S2A and B). The affinity constant calculated between the Cid1 WT and the U15 probe was ~1 and 20 μM with the A15 probe (Figure 3C). These values were identical for the D160A mutant, reinforcing the fact that this mutation does not alter the enzyme’s RNA binding capacity. Overall, we observe a 20-fold difference between the affinity constants of the various tested Cid1 binding capacity. Several residues have been shown to be important for the Cid1 activity such as F332, K144 or H336 were finally tested for their potential effect on the RNA binding properties. Strikingly, the F332 mutant had a reduced capacity for binding RNAs (Figure 3D). We previously proposed that F332 was critical for the UTP clamping mechanism mediated by the ratchet helix 4 (21). We suggest that the importance of the domain movement during the nucleotide transfer likely extends to RNA template trapping mechanism as well (Figure 3D) (21). We then looked into the RNA binding capacity of the K144A mutant protein, we observed that its association with A15 RNA was strongly reduced, whereas it was unchanged for a 15-mer U stretch (Figure 3D). Therefore, K144 residue seems important for the stabilization of our short incoming substrate RNA. These results are slightly different from the data published by Gilbert’s laboratory, possibly due to the differences in the RNA used for the binding studies (a 40-mer RNA in their study) (22). The poly(A) polymerase equivalent residue K145 is found to contact the 2′-OH group of the nucleotide 2 in the PAP:ATP:RNA crystal structure (24). Therefore, the Cid1 protein may use a substrate stabilization strategy closer to the one of the PAP enzymes. Altogether, it appears that the β-sheet structure on top of the active site, named β-trapdoor, plays a role during UTP exchange, although its role in RNA association has not been assessed (22).
that Cid1 enzyme is highly specialized with residues critical for the nucleotide triphosphate association (H336), residues specialized for the RNA primer association (K144) and residues involved in the nucleotidyl transferase enzymatic cycle (β-trapdoor, D103, N165 or F332).

**Polymerization activity of Cid1 mutants**

To complete our RNA binding study, we measured the polymerization activities of the same set of mutants. As we observed significant differences between Cid1 affinity for a U-tail as compared with an A-tail, we looked at the polymerizing activity in the presence of either UTP or ATP. The N165A mutant is unable to add more than one nucleotide per substrate regardless of the provided nucleotide triphosphate (UTP or ATP) even though its RNA binding capacity is unimpaired, at least for the 15-mer primer (Figures 3C, 4A and B). If N165 residue is changed into a charged residue (N165D), the enzyme is able to add up to 3–4 uridines or up to 20 adenines to the 15-nt RNA template but is still almost unable to use the ApU substrate (Figure 4, mutant N165D). Overall, these experiments demonstrate the critical role of N165 residue in the polymerizing reaction. Being located at the beginning of helix 4, N165 is likely important during the swivel motion occurring throughout the catalytic cycle of the enzyme (21). The rotation of the catalytic domain around the helix 4 can open the space between the domains. We suggest that N165 residue is involved in pulling on the RNA primer after the addition of the UMP. On displacement of the RNA primer, the UTP binding cavity is emptied allowing a new UTP molecule to take its place, and a new cycle to start.

Another region of interest near the NRM loop is the β-trapdoor (22). Deletion of the β-trapdoor residues does not affect the overall PUP activity of the enzyme for either of the RNA primers as long as we use UTP as a substrate (Figure 4A and C). We can observe the typical biphasic profile with, on one hand, short products of 7–9 added nucleotides corresponding to a distributive activity and, on the other hand, highly elongated products that did not enter the gel (Figure 4A and B). However, when we look at activity of the Δ310–322 mutant protein in the presence of ATP, we do not observe any larger products suggesting that Cid1 without the β-trapdoor is almost exclusively a distributive PAP enzyme (Figure 4B and D). This is even more striking with the short substrate ApU (Figure 4D). Because these observed differences between PAP and PUP activities cannot be linked to a difference in the RNA binding properties of the truncated protein (Figure 3C; Supplementary Figure S2E and F), we...
conclude that the β-trapdoor has an important role during the catalytic cycle. This may be to favor NTP exchange after each polymerization cycle by promoting domain movements (22).

Finally, we tested the polymerizing activity of the K144A mutant. This single-point mutant has a PUP activity comparable with the WT protein, but its PAP activity is drastically reduced as demonstrated by the almost complete absence of extended products (compare Figure 4A and B). As observed for the Δ310–322 deleted protein, its activity is now almost exclusively distributive when presented with ATP (Figure 4A and B). As the K144A mutant has an impaired A15 RNA association, we propose that the change in the polymerizing profile is linked with the poly(A) RNA binding capacities of the protein.

CONCLUSIONS

Our atomic models provide missing information regarding the association of RNAs with the uridylylating enzyme
Cid1. We identify multiple determinants behind the RNA binding properties of Cid1 and decipher the roles played by several residues or protein fragments during Cid1’s polymerization activity. When the 3’-end of the template RNA is not a uridine, Cid1 has a low RNA binding affinity and therefore a highly distributive activity reflecting its classification as a template-independent nucleotidyl transferase (8). Our binding tests show that on addition of 6–7 uridine nucleotides, the enzyme’s affinity for the 3’-end of its template becomes at least 20 times higher as shown by our binding tests. Following this switch in its overall affinity, Cid1’s activity becomes highly processive. We show that residue K144 is involved in this switch of activity. The Cid1 catalytic cycle is complex and its completion critically depends on several key residues: (i) the f-trap controls at least partially the UTP exchange after the reaction; (ii) N165 residue, which is hydrogen bonded to the 3’-penultimate nucleotide in our atomic model, may be involved in the displacement of the elongated RNA product after UMP addition. Finally, the overall catalytic cycle depends on the previously proposed domain movement around helix 4 and residue F332 to proceed (21).

Until now, the length of the characterized poly(U) tail due to PUP activities in vivo is generally short (between 1 and 5 nt) (4,5,37,38). We propose that Cid1 enzymatic activity is influenced by its higher affinity for the homouridine-containing tails. To prevent Cid1 from activity is influenced by its higher affinity for the homouridine-containing tails. To prevent Cid1 from generating long U-tails that can trigger RNA degradation, factor Lin28 (11,40). Lastly, as shown by the differential use of particular protein residues like residue R139 or K144, TUT and PUP enzymes have evolved differently (19,21). It has recently been proposed that DSSI protein, a uridine-specific exonuclease, plays a critical role during the guide RNA maturation process in trypanosome, a process where RET1 TUT enzyme is critical (41,42). As our results present shed new light on the properties of the polyuridylylating enzyme Cid1, it will be exciting to see whether a similar switch exists in the RET1 enzyme (43–45).

ACCESSION NUMBERS

Atomic coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 4NKT for the Cid1-UMPNP complex and 4NKU for the Cid1-APU complex.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors acknowledge the staff of ID14-4 at the ESRF for developing excellent beamlines and supporting optimal data collection. They also thank all the members of Stephane Thore Lab for their critical reading, helpful discussion and technical advice. They thank the University of Geneva for gratefully allowing the installation of the biostuctural platform together with the generous financial support of the Boninchi foundation, the Schmidheiny foundation and the Swiss National Science Foundation R’equip grant (310030_128787).

FUNDING

Swiss National Science Foundation [31003A_140924 and 31003A_124909 to S.T.]. Funding for open access charge: Swiss National Science Foundation.

Conflict of interest statement. None declared.

REFERENCES


Supplementary Information

A critical switch in the enzymatic properties of the Cid1 protein deciphered from its product-bound crystal structure

Paola Munoz-Tello¹, Caroline Gabus¹ and Stéphane Thore¹ *

¹ Department of Molecular Biology, University of Geneva, Geneva, 1211, Switzerland

* To whom correspondence should be addressed. Tel: 4122 379 6127; Fax: 4122 379 6868; Email: stephane.thore@unige.ch
Supplementary Tables

Table SI. Primers for mutagenesis

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Primer forward (5'→3')</th>
<th>Primer reverse (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>R139A</td>
<td>TTTTACAAAGGGCAAGCTCCATTATC</td>
<td>GATAATGGGAATTGCTGCCCTTTGTAAA</td>
</tr>
<tr>
<td>K144A</td>
<td>CAAGAATTCCATTATCGCAATACAG</td>
<td>CATTATCGAGATGTAAATGCGATAATGGGAATTCTTG</td>
</tr>
<tr>
<td>D160A</td>
<td>CGTTTCATGTGCTATTGGATTTAAC</td>
<td>GTTAAATCCAAATGCACATTGAAAG</td>
</tr>
<tr>
<td>N165A</td>
<td>GATTTGGATTTACAGCTGCTAGCTATTC</td>
<td>GATAGCCTAGACGCGTATTAATGGAAATCCAATTC</td>
</tr>
<tr>
<td>F332D</td>
<td>GCGATTTGAAGATCCATCTGCAGGATTTCACAATAG</td>
<td>CATTATGGAATCTCGTCAGGATCTCAAATGC</td>
</tr>
<tr>
<td>H336A</td>
<td>CTTTCGAGATTCGCTATAATGGTAG</td>
<td>CCTACCCACATTAAGCTGAAATCTCGAAAG</td>
</tr>
<tr>
<td>Δ310-322</td>
<td>GGATGGAAGTCCAGCTGACAGGTATA</td>
<td>CGCAAGAATACCTGTACGCTGAAATCTGAG</td>
</tr>
<tr>
<td></td>
<td>TCTTGCGCGC</td>
<td>GTCCATCC</td>
</tr>
</tbody>
</table>
Supplementary Figures and Legends

Figure S1: View of the Fo – Fc difference density maps of the UMPNPP and ApU compounds bound to Cid1.

(A) Fo – Fc difference density map of the UMPNPP nucleotide bound to Cid1. The Cid1 protein is shown in lime (CD domain) and salmon pink (CAT domain) with helix 4 in yellow. The NRM loop residues are colored in light magenta, the aspartate triad residues are colored in cyan and the major residues interacting with the ligands are shown in slate. Side chains and ligands are shown using a ball and stick representation (carbon as indicated before, nitrogen in blue, oxygen in red). The difference map was calculated using the refined protein as a model and is shown as a gray mesh (contoured at 2.5σ). (B) Fo – Fc difference density map of the minimal product ApU bound to Cid1. The difference map was calculated using the refined protein as a model and is shown as a gray mesh (contoured at 2.5σ). Side chains and ligand are shown as in panel A.
Figure S2: Characterization of the RNA binding properties of the WT and mutant Cid1 proteins by electrophoretic mobility shift assay (EMSA).

(A) EMSA showing the association between D160A mutant and an A₁₅ probe. (B) EMSA showing the interaction between D160A mutant and U₁₅ RNA. (C) EMSA showing the binding between N165A mutant and A₁₅ RNA. (D) EMSA illustrating the binding between N165A mutant and a U₁₅ probe. (E) EMSA displaying the interaction of Δ310-322 mutant and A₁₅ RNA. (F) EMSA showing the association of Δ310-322 mutant and a U₁₅ probe. Individual reactions in all the panels contained no protein or the indicated amounts.
1.11.7 Materials and Methods

1.11.7.1 Media Compositions

1.11.7.1.1 Luria Broth (LB) Media
For 1L of LB media: 10 g tryptone, 5 g yeast extract, 10 g NaCl and autoclaved deionized water to make 1L supplemented with the appropriate antibiotics. Autoclave for 20min at 120°C.

1.11.7.1.2 Terrific Broth (TB) Media
For 1L of TB media: 12g tryptone, 24g yeast extract, 4ml of glycerol and autoclaved water to make 1L after addition of 200ml of K salts and the appropriate antibiotics.
For 2L of K salts: 250.8g K2HPO4, 46.2g KH2PO4 and water to make 2L. Autoclave for 20min at 120°C.

1.11.7.1.3 Autoinduction (AI) Media
Several media are needed to obtain the final AI media.

10X ZY (1L):

<table>
<thead>
<tr>
<th>100g</th>
<th>Tryptone</th>
</tr>
</thead>
<tbody>
<tr>
<td>50g</td>
<td>Yeast extract</td>
</tr>
<tr>
<td>qsp 1L</td>
<td>Water</td>
</tr>
</tbody>
</table>

50X 5052 (1L):

<table>
<thead>
<tr>
<th>250g</th>
<th>Glycerol</th>
</tr>
</thead>
<tbody>
<tr>
<td>25g</td>
<td>Glucose</td>
</tr>
<tr>
<td>100g</td>
<td>Lactose</td>
</tr>
<tr>
<td>qsp 1L</td>
<td>Water</td>
</tr>
</tbody>
</table>

1M MgSO₄ (100ml):

<table>
<thead>
<tr>
<th>24.65g</th>
<th>MgSO₄•7H₂O</th>
</tr>
</thead>
<tbody>
<tr>
<td>qsp 100ml</td>
<td>Water</td>
</tr>
</tbody>
</table>

20X NPS (2L):

<table>
<thead>
<tr>
<th>132g</th>
<th>(NH₄)₂SO₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>272g</td>
<td>KH₂PO₄</td>
</tr>
<tr>
<td>284g</td>
<td>Na₂HPO₄</td>
</tr>
<tr>
<td>qsp 1L</td>
<td>Water</td>
</tr>
</tbody>
</table>

Autoclave for 20min at 120°C.

For 1L of AI media add 100ml of 10X ZY, 1ml of 1M MgSO₄, 20ml of 50X 5052, 50ml of 20X NPS supplemented with the appropriate antibiotics and autoclaved water to make 1L.
1.11.7.1.4 Seleno-Methionine (SeMet) Derivative Media

Several media are needed for SeMet-derivative protein expression.

Medium A (1L):

<table>
<thead>
<tr>
<th>100ml</th>
<th>M9 medium (10X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10ml</td>
<td>Trace elements solution (100X)</td>
</tr>
<tr>
<td>20ml</td>
<td>20% (w/v) Glucose</td>
</tr>
<tr>
<td>1ml</td>
<td>1M MgSO₄</td>
</tr>
<tr>
<td>0.3ml</td>
<td>1M CaCl₂</td>
</tr>
<tr>
<td>1ml</td>
<td>Biotin (1mg/ml)</td>
</tr>
<tr>
<td>1ml</td>
<td>Thiamin (1mg/ml)</td>
</tr>
<tr>
<td></td>
<td>Appropriate antibiotic(s)</td>
</tr>
</tbody>
</table>

10X Trace elements solution (1L):

<table>
<thead>
<tr>
<th>5g</th>
<th>EDTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83g</td>
<td>FeCl₃•6 H₂O</td>
</tr>
<tr>
<td>84mg</td>
<td>ZnCl₂</td>
</tr>
<tr>
<td>13mg</td>
<td>CuCl₂•2 H₂O</td>
</tr>
<tr>
<td>10mg</td>
<td>CoCl₂•6 H₂O</td>
</tr>
<tr>
<td>10mg</td>
<td>H₃BO₃</td>
</tr>
<tr>
<td>1.6mg</td>
<td>MnCl₂•6 H₂O</td>
</tr>
</tbody>
</table>

10X M9 medium (1L):

<table>
<thead>
<tr>
<th>80g</th>
<th>Na₂HPO₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>40g</td>
<td>KH₂PO₄</td>
</tr>
<tr>
<td>5g</td>
<td>NaCl</td>
</tr>
<tr>
<td>5g</td>
<td>NH₄Cl</td>
</tr>
</tbody>
</table>

For all the media, add water to make 1L. For medium A use autoclaved water. M9 medium autoclave for 20min at 120°C. Trace elements need to be filtered with a 0.22μm filter.

1.11.7.2 Protein Expression and Batch Purification Tests

Small-scale cultures of *E. coli* BL21-Star or Rosetta 2 cells (Novagen), previously transformed with either Cid1L or Cid1S construct, were grown in LB or TB media at 37°C for 6 hours. The cells were then induced overnight at 18°C with 0.1mM IPTG or diluted in AI media (1/50) for overnight expression at 30°C. Expression of the recombinant proteins was analyzed by SDS-PAGE. Cells expressing Cid1L or Cid1S were harvested using centrifugation at 4 000 rpm and resuspended in 50mM Phosphate buffer pH 7.5 containing 300mM NaCl, 20mM imidazole, 0.1% Triton X-100, DNase 1 (1μg/ml), Lysozyme (1μg/ml), 5mM β-mercaptoethanol (β-MT), and a cocktail of protease inhibitors (1mM PhenylMethylSulfonyl Fluoride, 1μg/ml of leupeptin and 2μg/ml of pepstatin). Cells were lysed using an Emulsiflex system (AVESTIN) and cleared using centrifugation at 15 000 rpm for 30 minutes at 4°C. The soluble fraction
was subjected to an initial affinity purification using 1ml slurry of Ni-NTA beads (GE Healthcare). The protein was eluted from the beads with a buffer containing 250mM imidazole and desalted against 50mM Phosphate buffer pH 7.5, 300mM NaCl, 20mM imidazole, and 5mM β-mercaptoethanol. The TEV protease was added at a ratio of 1/50 (TEV/protein) and the sample was incubated at 16°C overnight. The fractions containing the total and soluble fractions, the elution peak and the sample after protease treatment were analyzed on a SDS-PAGE.

1.11.7.3 Seleno-Methionine Derivative Protein Expression, Purification and Crystallization

Cid1L is overexpressed in E. coli B834 cells, which are auxotroph for methionine, grown in LB media at 37°C for 8 hours then diluted 1/50 and grown overnight at 37°C. The overnight culture is diluted 1/100 and grown for 6 hours, centrifuged at 4 000 rpm for 15 min at room temperature (RT) and resuspended in 1L of Medium A (without methionine) described previously for overnight growth at 37°C. The next morning, the cells were harvested by centrifugation at 4 000 rpm for 15 min at RT, resuspended in 1L of Medium A supplemented with 50mg of seleno-methionine and incubated at 37°C for 30 min followed by 4 hours induction at 37°C with 1mM IPTG. Induced cells were harvested by centrifugation at 4 000 rpm for 15 min at 4°C and resuspended in a lysis buffer containing 50mM Phosphate buffer pH 7.5, 300mM NaCl, 0.1% Triton X-100, 20mM Imidazole, DNase 1 (1μg/ml), Lysozyme (1μg/ml), 5mM β-MT, and a cocktail of protease inhibitors (1mM PhenylMethylSulfonyl Fluoride, 1μg/ml of leupeptin and 2μg/ml of pepstatin). Cells were lysed using the emulsiflex system (AVESTIN) and cleared by centrifugation at 15 000 rpm for 30 min at 4°C. SeMet-derivative Cid1L was then purified and crystallized as for Cid1 WT protein at 9.5mg/ml. We also tested a second crystallization technique called micro-seeding, which consists in crushing the obtained crystals in order to seed a new set of crystal trials. At least two or three rounds of micro-seeding needed to be performed in order to obtain larger crystals at 18°C in a reservoir solution containing 0.1 M HEPES pH7.0, 20% PEG 6000, 0.2M Ammonium Chloride.
1.12 Structural Studies of Mitochondrial RNA Processing Complexes in *T. brucei*

1.12.1 gRNA Maturation Complexes or RET1 TUTase DSS1 Exonuclease Complex (RDS complex)

1.12.1.1 RDS Complex Expression and Purification

Ruslan Aphasizhev group in Boston identified through RET1 TUTase pull-down assays and mass spectrometry analysis four proteins implicated in gRNA biogenesis (unpublished data). The first one is DSS1, a 80 kDa 3’ exonuclease known to be an essential protein influencing the RNA editing reaction and the RNA stability in trypanosomal mitochondria (Penschow *et al.* 2004). The other 3 proteins named RDS1, RDS2 and RDS3 (standing for RET1 TUTase DSS1 complex protein), are three new proteins without any discernable motif of 189, 180 and 123 kDa respectively. These five proteins compose the RDS complex in trypanosomes. RNAi tests on RET1 TUTase as well as DSS1 exoribonuclease show accumulation of pre-gRNAs in the mitochondria suggesting that this RDS complex might be necessary for gRNA biogenesis (unpublished data). In order to better understand the function of this newly identified RDS complex, we started a collaborative project with their laboratory aiming to solve the crystal structure of the factors, either as sub-complexes or single subunits. It is noteworthy that RET1 TUTase is also involved in the uridylation of mitochondrially-encoded mRNAs as part of a polyadenylation/uridylation complex, which will be discussed in the section 4.2.2 (Aphasizheva and Aphasizhev 2010; Aphasizheva *et al.* 2011).

We received several clones from our collaborators ready for *E. coli* protein expression trials of RET1, DSS1, RDS1, RDS2 and RDS3 full-length (FL) proteins. These plasmids contained either one single protein tagged with a N- or C-terminal His-tag or two ORFs with one protein tagged and the other one untagged for co-expression and co-purification assays. All the expression and purification tests performed with the different plasmids are summarized in table 3 (Table 3). Unfortunately, in our co-expression trials we could only express one of the proteins no matter what combination we would try and obtained very poor expression levels (Table 3). Since the size of these proteins is rather high (more than 100kDa each), we decided to try to express and purify the proteins separately. We would then *in vitro*...
reconstitute the complex from these individual subunits. RDS1, RDS2 and RDS3 expressions and purifications were still unsuccessful most probably due to their large size. For DSS1 protein, we obtained very good expression levels but the protein was not soluble (Fig. 23A). We tried to solubilize it by adding a N-terminal GST tag and testing different lysis buffers however unsuccessfully (Table 4, Fig. 23A). RET1 TUTase construct had good expression levels and small-scale purification seemed promising (Table 3; Fig. 23B).

On RET1 small-scale elutions we performed a size exclusion chromatography in order to analyze the oligomeric state of the enzyme as well as mass spectrometry analysis in order to confirm the presence of RET1. Mass spectrometry analysis revealed that both of the upper bands seen on the gel of the elution correspond to RET1 protein (Fig. 24). We realized that these two protein forms are found in several different oligomeric states and that they are indiscernible (Fig. 25A and 25B). We then focused ourselves on RET1 enzyme purification optimization to remove these different populations to eventually setup crystallization screens.

![Figure 23: Batch purification of DSS1 and RET1 FL enzymes (A) SDS-PAGE gels of the batch purifications and solubility tests performed on DSS1-expressing cells. (B) SDS-PAGE gel showing the different batch purification](image)
steps of RET1 FL enzyme. In all the panels, the arrow indicates the protein of interest and T is Total fraction, S is Soluble fraction, FT is Flow-Through, W is Wash and E1, E2 and E3 are the three eluted fractions.

<table>
<thead>
<tr>
<th>Protein(s) and tag(s)</th>
<th>Expression</th>
<th>Soluble</th>
<th>TEV cleavage</th>
<th>Ni-NTA affinity purification</th>
<th>Size exclusion chromatography</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AI 30°C O/N</td>
<td>LB or TB 18° O/N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R2</td>
<td>BL21 ★</td>
<td>R2</td>
<td>BL21 ★</td>
<td></td>
</tr>
<tr>
<td>DSS1-6xHis / RET1 untagged</td>
<td>NO</td>
<td>FEW</td>
<td>FEW</td>
<td>FEW</td>
<td>NO for BL21 ★ YES for R2 RET1</td>
</tr>
<tr>
<td>6xHis-RDS2 / RET1 untagged</td>
<td>-</td>
<td>Only RDS2</td>
<td>Only RDS2</td>
<td>-</td>
<td>NO</td>
</tr>
<tr>
<td>RET1 untagged / RDS1-6xHis</td>
<td>Only RET1</td>
<td>Only RET1</td>
<td>Only RET1</td>
<td>Only RET1</td>
<td>Purifiable without tag</td>
</tr>
<tr>
<td>6xHis-N-DSS1 / RET1 untagged</td>
<td>Only DSS1</td>
<td>NO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RET1 untagged</td>
<td>FEW</td>
<td>YES</td>
<td>FEW</td>
<td>FEW</td>
<td>YES</td>
</tr>
<tr>
<td>DSS1-6xHis</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>6xHis-N-DSS1</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td>Few binds column or beads</td>
</tr>
<tr>
<td>6xHis-N-RDS3</td>
<td>FEW</td>
<td>FEW</td>
<td>FEW</td>
<td>FEW</td>
<td>FEW</td>
</tr>
<tr>
<td>6xHis-N-RDS2</td>
<td>FEW</td>
<td>FEW</td>
<td>FEW</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>RDS1-6xHis</td>
<td>NO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GST-8xHis-RET1L and DSS1</td>
<td>YES</td>
<td></td>
<td>YES</td>
<td></td>
<td>NO</td>
</tr>
<tr>
<td>RET1L and GST-8xHis-DSS1</td>
<td>YES</td>
<td></td>
<td>NO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Expression and purification trials for the different proteins forming the RDS complex. R2 corresponds to Rosetta 2 E. coli strain and N is the TEV cleavage site.
Figure 24: 10% SDS-PAGE gel used for mass spectrometry analysis of RET1 FL Ni-NTA purification elutions. The first two bands correspond to RET1 enzyme whereas the third band corresponds to a contaminant.

<table>
<thead>
<tr>
<th>Buffer 1</th>
<th>Buffer 2</th>
<th>Buffer 3</th>
<th>Buffer 4</th>
<th>Buffer 5</th>
<th>Buffer 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM Phosphate buffer pH 7.5</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
</tr>
<tr>
<td>300mM NaCl</td>
<td>1M NaCl</td>
<td>1M NaCl</td>
<td>1M NaCl</td>
<td>1M NaCl</td>
<td>1M NaCl</td>
</tr>
<tr>
<td>20mM imidazole</td>
<td>20mM imidazole</td>
<td>20mM imidazole</td>
<td>20mM imidazole</td>
<td>20mM imidazole</td>
<td>20mM imidazole</td>
</tr>
<tr>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
</tr>
<tr>
<td>0.1% Triton X-100</td>
<td>0.015mM DDM</td>
<td>2mM CHAPS</td>
<td>2mM CHAPS</td>
<td>0.1% Triton X-100</td>
<td>0.015mM DDM</td>
</tr>
<tr>
<td>5% Glycerol</td>
<td>5% Glycerol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: DSS1 solubility test buffers.

Figure 25: Gel Filtration chromatography analysis of RET1 FL after Ni-NTA purification elutions. (A) Gel Filtration chromatogram. (B) 10% SDS-PAGE gel analysis of the different gel filtration peaks. Bgf, sample before gel filtration and Agg, aggregates.
1.12.1.2 RET1 Terminal Uridyl Transferase

Based on secondary structure predictions, we performed three different constructs of RET1 TUTase named RET1L, RET1S and RET1S2 to improve its behavior during the various purification steps (Fig. 26). We cloned them into our pST0 vector containing a GST-8xHis tag, which contain a TEV cleavage site for tag removal (Fig. 37). We then performed expression tests with RET1L and RET1S (Table 5) followed by small-scale Ni-NTA affinity purifications (Fig. 27). RET1L being poorly expressed (Fig. 27A), we continued with a large-scale purification of RET1S construct. As for Cid1 enzyme, we performed the same purification protocol as a first trial: i) Day 1: Ni-NTA affinity purification followed by desalting to remove imidazole followed by O/N TEV cleavage ii) Day 2: Reloading on the Ni-NTA column to remove the tag and contaminants followed by a size exclusion chromatography step. Unfortunately, during the reloading process we realized that RET1S protein has some affinity for Ni\(^{2+}\) ions since it was able to bind the Ni-NTA column without a tag. We then changed our purification method using a GST column for affinity purification instead and optimized it as shown in the purification scheme in materials and methods section 4.2.3.3 (Fig. 38). Indeed, we had to add a heparin chromatographic step in order to remove the nucleic acid contamination of our preparation (Fig. 28). It is noteworthy that this heparin step allowed, after TEV cleavage, to avoid a second reloading onto the GST column as the GST tag remained in the 100mM NaCl flow-through (FT) of the purification and the uncleaved RET1S protein eluted at 1M NaCl salt wash. RET1S untagged version eluted from the heparin column with few contaminants at ~480mM NaCl salt and was directly injected into a Superdex 200 gel filtration column where it eluted as a dimeric protein (Ve\text{lution}=65ml). The final SDS-PAGE gel summarizing the different purification steps is shown in figure 29 (Fig. 29A). From our large-scale purifications, we obtained 5mg of protein out of 10g of cell pellet.

In order to assess the activity of the C-terminally truncated RET1S construct, we performed polyuridylation activity assays at 37°C with a 17-mer unlabeled RNA supplemented with ATP or UTP, MgCl\(_2\) and equimolar amounts of the tested proteins. We used Cid1 enzyme as a positive control for polyuridylation activity. Activity test of the RET1S construct confirmed that the enzyme was active and uses only UTP as a substrate unlike Cid1 that uses also ATP (Fig. 29B).
We tested over 1500 crystallization conditions at 18°C and 4°C at a protein concentration of 3.5 mg/ml. We could not go higher in protein concentration because RET1S construct was very unstable and heavily precipitated upon concentration. We crystallized the protein at 18°C in a reservoir solution containing 25% of Sokalan HP56, 0.2M of ammonium acetate and 0.1M HEPES-NaOH pH 7.0 but failed optimizing the condition (Fig. 29C). We tested these small needles without obtaining any diffraction.

**Figure 26: RET1 enzyme constructs.**

<table>
<thead>
<tr>
<th>Media</th>
<th>E. coli strain</th>
<th>Temperature</th>
<th>[IPTG]</th>
<th>Time</th>
<th>Expression</th>
<th>Solubility</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Rosetta 2</td>
<td>30°C</td>
<td>O/N</td>
<td>Few</td>
<td>Soluble</td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td>BL21★</td>
<td>30°C</td>
<td>O/N</td>
<td>Yes</td>
<td>Soluble</td>
<td></td>
</tr>
<tr>
<td>LB</td>
<td>Rosetta 2</td>
<td>18°C</td>
<td>0.1mM</td>
<td>O/N</td>
<td>Few</td>
<td>Soluble</td>
</tr>
<tr>
<td>LB</td>
<td>BL21★</td>
<td>18°C</td>
<td>0.1mM</td>
<td>O/N</td>
<td>Yes</td>
<td>Soluble</td>
</tr>
<tr>
<td>LB</td>
<td>BL21★</td>
<td>37°C</td>
<td>1mM</td>
<td>4h</td>
<td>Insoluble</td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td>Rosetta 2</td>
<td>18°C</td>
<td>0.1mM</td>
<td>O/N</td>
<td>Few</td>
<td>Soluble</td>
</tr>
<tr>
<td>TB</td>
<td>BL21★</td>
<td>18°C</td>
<td>0.1mM</td>
<td>O/N</td>
<td>Yes</td>
<td>Soluble</td>
</tr>
<tr>
<td>TB</td>
<td>BL21★</td>
<td>37°C</td>
<td>1mM</td>
<td>4h</td>
<td>Yes</td>
<td>Insoluble</td>
</tr>
</tbody>
</table>

*Table 5: RET1 expression tests. In bold are the best expression conditions for all the RET1 constructs.*
Figure 27: RET1L and RET1S small-scale purifications. (A) RET1L 10% SDS-PAGE gel showing the different batch purification steps. (B) RET1S 10% SDS-PAGE gel showing the different batch purification steps. In all the panels, the arrow indicates the protein of interest and T is Total fraction, S is Soluble fraction, FT is Flow-Through, W is Wash and E1, E2 and E3 are the three eluted fractions. Ni, none-induced; I, induced.

Figure 28: 1% agarose gel showing the nucleic acids contamination of the RET1 eluted sample. Bef Hep, sample before heparin purification; Aft Hep, sample after heparin purification.
Figure 29: RET1S purification, polymerization activity assay and crystallization. (A) 10% SDS-PAGE gel showing the different purification steps and gel filtration chromatogram showing the monomeric state of RET1S, the aggregates and the GST tag. Tot, total fraction; Sol, soluble fraction; FT, flow through; BT, before TEV; AT, after TEV; He, heparin step; Bgf, Before GF; Agf, after GF; Prot, pure protein. (B) Poly(U) and poly(A) polymerization assay. 181μM of a 17-mer RNA were incubated for 40 min at 37°C with either RET1S or Cid1 supplemented with 5mM MgCl2 and 10mM of UTP or ATP. The smeared signal observed corresponds to the poly(U)-containing product of the reaction. Ct, RNA alone with UTP. Prot, protein alone. (C) RET1S small thin needle-like crystals.

Finally, we decided to test our RET1S2 construct for expression and purification in a large-scale using the optimized purification method of RET1S shown in figure 38 of materials and methods section 4.2.3.3 (Fig. 38). The protein was well expressed and soluble, and eluted as a monomer in the Superdex 75 column (Veletion=56ml; Fig. 30A). We checked the polyuridylylating activity of RET1S2 and confirmed the activity of this construct (Fig. 30B). As for RET1S, we performed crystallization screens 18°C and 4°C at 4mg/ml. Two crystallization hits were revealed and optimized using the sitting drop vapor diffusion technique (Fig. 31A and 31B). The first hit leads to very small see urchins, which after several rounds of optimization and the use of the pseudo-batch technique (previously described for Cid1), formed needle-like RET1S2 crystals. These crystals grew to full size in a week in a reservoir.
solution containing 0.1M Bicine/Tris pH 8.5, 25-30% Glycerol/PEG 4000, 0.9M NPS (0.3M sodium nitrate, 0.3M disodium hydrogen phosphate and 0.3M ammonium sulfate). The second RET1S2 initial hit contained 0.1M Bicine pH 9.0, 20% polyacrylate 2100 and 0.2M NaCl forming very small needle-like see urchins. In order to optimize these crystals we combined the micro-seeding technique together with the pseudo-batch technique (Fig. 31B). RET1S2 crystals grew to full size in a week and a half giving plate-like crystals in a reservoir solution containing 20% polyacrylate 2100, 0.1M Bicine pH 8.6, 0.2M NaI supplemented with non-diluted crystal seeds (Fig. 31B). Several crystals with different cryoprotection solutions were tested for diffraction at PXIII beamline at the Swiss Light Source (SLS) synchrotron. Needle-like crystals diffracted very poorly (less than 15Å) whereas plate-like crystals diffracted up to 3.5Å resolution (Fig. 32A). We processed the diffraction data and RET1S2 plate-like crystals belonged to the space group P2(1)2(1)2 with the following cell parameters a=127.70, b=193.18, c=59.44, α=β=γ=90. A summary of the processing and scaling statistics are shown in figure 32B (Fig. 32B). We are now aiming to solve the structure by molecular replacement using RET2 enzyme (PDB code: 2B51) as a first model (Fig. 32B and 33A) and also trying to obtain better quality crystals. RET2 TUTase has only 26% sequence identity with RET1 (Deng et al. 2005). Thus, seleno-methionine derivative RET1S2 protein expression, purification and crystallization are being set up in case the MR fails to bring up the phases. Other models will also be tried for the MR such as TUT4 (33% identity, PDB code: 2IKF) and Cid1 (25% identity; PDB code: 4EP7), which might help to obtain phase information and subsequently calculate electron density maps in order to solve the structure of the RET1S2 fragment (Fig. 33B and 33C; Stagno et al. 2007; Munoz-Tello et al. 2012).
Figure 30: RET1S₂ purification and polymerization assay. (A) 10% SDS-PAGE gel showing the different purification steps and gel filtration chromatogram showing the monomeric state of RET1S₂. (B) Poly(U) and poly(A) polymerization assay. 181μM of a 17-mer RNA were incubated for 40 min at 37°C with either RET1S₂ or Cid1 supplemented with 5mM MgCl₂ and 10mM of UTP or ATP. Prot stands for RET1S₂ protein alone.
Figure 31: RET1S₂ crystal hits optimization. (A) Needle-like RET1S₂ crystals optimized using the pseudo-batch technique. (B) Plate-like RET1S₂ crystals optimized using micro-seeding and pseudo-batch techniques.

Figure 32: RET1S₂ diffraction data. (A) Diffraction pattern of RET1S₂ plate-like crystals shown in the figure. (B) RET1S₂ data processing final table showing the statistics of the data reduction and processing.
1.12.2 Polyadenylation/Uridylation Complex

In trypanosomes, the polyadenylation/uridylation complex composed of KPAP1 poly(A) polymerase, RET1 TUTase and KPAF1/KPAF2 factors is required for mRNA translation (Aphasizheva et al. 2011). This complex adds a long A/U tail to the 3'-end of the mRNA that is then recognized by the translation machinery. In order to better understand the function of each of these proteins within the complex, we tried to solve its atomic structure by X-ray crystallography. KPAF1 and KPAF2 factors are mainly composed of pentatricopeptide repeats (PPR), which form a helix-loop-helix structure and have been shown to be important for RNA recognition and enhance the catalytic activity of RET1 and KPAP1 (Aphasizheva et al. 2011).

Another PhD student in our laboratory is currently studying the structural characterization of KPAF1 and KPAF2 complex. We focused our efforts on the catalytic part of the polyadenylation / polyuridylation complex meaning RET1 (structural work discussed in section 4.2.1.2) and KPAP1.

Based on secondary structure predictions, we clone a truncated version of KPAP1 protein with either a N-Terminal 9xHis-MBP tag; GST-8xHis tag or a 7xHis tag and performed small-scale expression tests (Fig. 34). All the constructs were expressed but the GST-tagged and His-tagged versions were not soluble. We then performed a large-scale purification of the MBP-tagged KPAP1 protein (Fig. 35). At the affinity purification step, we realized that the protein was not very stable since the MBP tag was getting self-cleaved (Fig. 35A). Nevertheless, we pooled the elution fractions together for an O/N TEV cleavage at 4°C (Fig. 35B).
followed by a reloading into the Ni-NTA column in order to remove the contaminants and the tag (Fig. 35C). Most of the protein was either precipitated or binding the nickel column (Fig. 35C). The few protein left was then concentrated and loaded into a gel filtration column to evaluate the oligomeric state of the untagged protein (Fig. 35D). Unfortunately, the protein was found in the aggregated fractions. We tried to optimize the purification buffers but the protein was still eluting as an aggregated population (Table 6). Further optimizations of the purification method as well as the protein construct are needed in order to obtain a non-aggregated population of KPAP1 enzyme.

Figure 34: KPAP1 (29-436 Aa.) protein construct with three different purification tags.

Figure 35: KPAP1 enzyme purification. (A) 10% SDS-PAGE gel showing the Total (T) and Soluble (S) fractions, the flow-through (FT) of the Ni-NTA column and the Ni-NTA elution of KPAP1. We observe self-cleavage of the MBP-KPAP1 tagged protein. (B) Before TEV (BT) and After TEV (AT) cleavage samples loaded on a 10% SDS-
Page 101

Table 6: Different buffers tested for KPAP1 enzyme purification.

We also tried to obtain KPAF1 and/or KPAF2 together with RET1 TUTase but we did not manage to obtain a stable complex. Trials with the whole complex will be performed once we obtain a stable KPAP1 protein.

1.12.3 Materials and Methods

1.12.3.1 RDS complex proteins cloning

We received several ready-to-use expression and co-expression vectors from our collaborators in Boston containing the RDS complex protein components. An example of plasmid organization is presented in figure 36 (Fig. 36).
New DSS1, KPAP1 and RET1 constructs were designed based on secondary structure predictions obtained from the online server Phyre (Kelley and Sternberg 2009). Phyre is an online server that performs a local alignment of the query sequence against a fold library. The output of this server is a structure prediction model, which is based on the alignment of the query sequence with all the sequences present in the library. This model allows us to visualize the potential disordered regions of the protein sequence that could impair the crystallization of the protein, and remove them for structural purposes. The laboratory has developed a series of pET42-derived expression vectors, called, pST vectors, with identical multi-cloning sites, allowing us to use a single pair of PCR primers to clone our proteins of interest at the N-terminal of different tags. An example of pST vector with a GST-8xHis tag, named pST0, is shown in figure 37 (Fig. 37). The primers used for the PCR amplification of the protein fragments are shown in table 7 (Table 7). The constructs were cloned in the corresponding vectors using Ncol and Xhol restriction enzymes (New England Biolabs) O/N at 37°C followed by T4 DNA Ligase (New England Biolabs) ligation for 3h at RT. The ligated vectors were then transformed into XL1 E. coli cells, extracted and sequenced.
Figure 37: Schematic representation of the expression plasmid pST0 used for the overexpression of GST-8x His tag RET1 constructs in E. coli cells. Only the unique cutter enzymes are represented. CoIE1 origin: the high-copy-number origin of replication. MCS: Multiple Cloning Site.

| For KPAP1 (29-436aa) | 5' - gctattCAATGGACaaaagcgacagccagtgg - 3' |
| Rev KPAP1 (29-436aa) | 5’-cccaatCTCGAGtcatcgttcctgaagtagagtcgc - 3' |
| For RET1L and RET1S | 5' - gctattGGCGCGCCtatggtaagtaagtaccaccgc - 3' |
| Rev RET1L | 5’ - ccgattGGCGCGCCtcaagtaaccttgctgatgccgc - 3’ |
| Rev RET1S | 5’ - ccgattGGCGCGCCtcatgctgcggcgaagacgcac - 3’ |
| For DSS1 (56-743aa) | 5’ - gctattGGCGCGCCtatgcttgacaaggaactgcttc - 3' |
| Rev DSS1 (53-743aa) | 5’ - ccgattGGCGCGCCtcatgctgcggcgaagacgcac - 3’ |
| For RET1S2 | 5’ - cccgattCCATGGgtgtgcggttgtattc - 3' |
| Rev RET1S2 | 5’ - cccgattCTCGAGtcatcgttcctgaagtagagtcgc - 3’ |

Table 7: Primers used for cloning.

1.12.3.2 Protein Expression and Purification Tests

Small-scale cultures of E. coli BL21-­‐Star or Rosetta 2 cells (Novagen), previously transformed with the appropriate protein construct, were grown in LB or TB media at 37°C for 6 hours. The cells were then induced overnight at 18°C with 0.1mM IPTG or diluted in Al media (1/50) for overnight expression at 30°C. Expression of the recombinant proteins was analyzed by SDS-­‐PAGE. Cells expressing the protein were harvested using centrifugation at 4 000 rpm and resuspended in 50mM Tris buffer pH 8.5 containing 300mM NaCl, 20mM imidazole, 0.1% Triton X-­‐100, DNase 1 (1μg/ml), Lysozyme (1μg/ml), 5mM β-­‐MT, and a cocktail of protease inhibitors (1mM PhenylMethylSulfonyl Fluoride, 1μg/ml of leupeptin and 2μg/ml of pepstatin). Cells were lysed using an Emulsiflex system (AVESTIN) and cleared using
centrifugation at 15,000 rpm for 30 minutes at 4°C. The soluble fraction was subjected to an initial affinity purification using 1 ml slurry of Ni-NTA beads (GE Healthcare). The protein was eluted from the beads with a buffer containing 250 mM imidazole and desalted against 50 mM Tris buffer pH 8.5, 300 mM NaCl, 20 mM imidazole, and 5 mM β-MT. The TEV protease was added at a ratio of 1/50 (TEV/protein) and the sample was incubated at 16°C or 4°C overnight. The fractions containing the total and soluble fractions, the elution peak and the sample after protease treatment were analyzed on a SDS-PAGE. When obtaining enough material, eluted proteins were loaded on a Superdex 200 analytical column for checking the oligomeric state of the protein or complexes.

**1.12.3.3 Large-Scale RET1 TUTase Expression, Purification and Crystallization**

RET1 constructs were overexpressed in BL21-Star *E. coli* cells, grown in TB media at 37°C for 6 hours followed by overnight induction at 18°C with 0.1 mM IPTG. Induced cells were harvested by centrifugation at 4,000 rpm for 30 min at 4°C and resuspended in a lysis buffer A (Table 8) supplemented with DNase 1 (1 μg/ml), Lysozyme (1 μg/ml), and a cocktail of protease inhibitors (1 mM PhenylMethylSulfonyl Fluoride, 1 μg/ml of leupeptin and 2 μg/ml of pepstatin).

Cells were lysed using the emulsiflex system (AVESTIN) and cleared by centrifugation at 16,000 rpm for 30 min at 4°C. The soluble fraction was subjected to several purification steps summarized in figure 38 and the different buffers used during the purification are shown in table 8 (Fig. 38; Table 8). Briefly, the soluble fraction is loaded on a GST column and the tagged RET1 protein is eluted with Elution buffer A containing 10 mM of reduced glutathione. The eluted fractions are then set for TEV protease cleavage at 4°C O/N. The next morning, the sample is diluted to Wash buffer Hep to loaded on a heparin column to remove nucleic acids contamination and the tag. The heparin is eluted with the High Salt buffer Hep and the peak corresponding to RET1 is subjected to a gel filtration step in GF buffer A to check the oligomeric state of the protein prior to crystallization trials.
### Table 8: Buffers used for RET1 optimized purification.

<table>
<thead>
<tr>
<th>Lysis buffer A</th>
<th>Wash buffer A</th>
<th>Wash buffer Hep</th>
<th>High Salt buffer Hep</th>
<th>Elution buffer A</th>
<th>GF buffer A</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM Tris pH 8.0</td>
<td>50mM MES pH 6.5</td>
</tr>
<tr>
<td>500mM NaCl</td>
<td>500mM NaCl</td>
<td>100mM NaCl</td>
<td>1M NaCl</td>
<td>500mM NaCl</td>
<td>500mM NaCl</td>
</tr>
<tr>
<td>0.1% Triton X-100</td>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td>10mM L-Glutathione reduced</td>
<td>2mM TCEP</td>
</tr>
<tr>
<td>5mM β-MT</td>
<td>5mM β-MT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 38: RET1 optimized purification scheme.** The lysate containing RET1 GST-8xHis tagged protein was loaded 2 times on a 2x5ml GST column. The GST affinity purification was followed by a cleavage reaction, overnight at 4°C, by TEV protease in order to remove the tags. After tag removal, a heparin column is loaded 3 times in order to remove the nucleic acid contamination and separated the GST-8xHis tag from the protein. As last step, to further polish and exchange the buffer of the sample, the protein was injected on a size exclusion chromatography previously equilibrated with the GF buffer. The protein is then concentrated and ready for setting crystallization screens.

RET1S₂ protein was crystallized in two steps. Crystals were produced using the sitting-drop vapor diffusion technique at 18°C, with a reservoir solution containing 20% polyacrylate 2100, 0.1M Bicine pH 8.6 and 0.2M NaI. The same pseudo-batch technique used for Cid1 enzyme was performed for RET1S₂ combined with micro-seeding in order to obtain diffracting plate-like crystals. Crystals appeared in two days and grew to maximum size of 0.1 x 0.04 x 0.2 mm in a week and a half. Crystals were passed into the reservoir solution supplemented with 15% glycerol prior to freezing in liquid nitrogen.
1.12.3.4 In Vitro Polymerization Assay

A 17-mer RNA was used for all the RET1 polymerization assays. 181μM of unlabeled RNA were incubated with 70μM or 180μM of RET1 constructs, or Cid1 (positive control), supplemented with 5mM MgCl₂ and 10mM of UTP or ATP for 40min at 37°C. The reactions were then stopped with an equal volume of 100mM EDTA and 2% sodium dodecyl sulfate (SDS). Extended RNA primers were then extracted with phenol/chloroform, precipitated with 80% ethanol and resuspended in 90% formamide loading buffer. Products were denatured at 95°C for 3 min and separated by gel electrophoresis in a 12% polyacrylamide/7M urea gel.

1.12.3.5 Large-Scale KPAP1 Expression and Purification

KPAP1 was overexpressed in BL21-Star E. coli cells, grown in TB media at 37°C for 6 hours followed by overnight induction at 18°C with 0.1mM IPTG. Induced cells were harvested by centrifugation at 4 000 rpm for 30 min at 4°C and resuspended in a lysis buffer containing 50mM Tris buffer ph 8.0, 500mM NaCl, 20mM imidazole, 0.1% Triton X-100 and 5mM β-MT supplemented with DNase 1 (1μg/ml), Lysozyme (1μg/ml), and a cocktail of protease inhibitors (1mM PhenylMethylSulfonyl Fluoride, 1μg/ml of leupeptin and 2μg/ml of pepstatin).

Cells were lysed using the emulsiflex system (AVESTIN) and cleared by centrifugation at 16 000 rpm for 30 min at 4°C. The soluble fraction was subjected to a first affinity purification using chelating HiTrap FF crude column (GE Healthcare) charged with Ni²⁺ ions. The protein was eluted with a buffer containing 250mM of imidazole and desalted against a buffer containing 50mM Tris buffer ph 8.0, 500mM NaCl, 20mM imidazole and 5mM β-MT. The eluted fractions were then set for TEV protease cleavage at 4°C O/N. Cleaved protein was reloaded on the Ni-NTA column to remove the tag and contaminants followed by a final gel filtration step to check the oligomeric state of the enzyme.
General Discussion
1.13 Structural Determinants for Cid1 Poly(U) Polymerase UTP Selectivity and RNA Recognition

Cid1 enzyme belongs to the polymerase β superfamily and was uncovered through its resistance to the combination of hydroxyurea and caffeine, both drugs disturbing S-M phase control, during a screen aiming to identify new players of this checkpoint (Aravind and Koonin 1999; Wang et al. 1999; Wang et al. 2000). Nevertheless, no link between Cid1 and the S-M checkpoint has been revealed so far. Cid1 was initially considered as a poly(A) polymerase until further biochemical analysis uncovered its preferential polyuridylation activity in vivo and in vitro (Wang et al. 2000; Read et al. 2002; Rissland et al. 2007; Aphasizheva et al. 2011). Recent studies have implicated Cid1 enzyme in a new RNA degradation pathway independent of deadenylation, which starts with polyuridylation of the mRNA followed by decapping and mRNA body decay (Rissland and Norbury 2009). Cid1 protein lacks any RNA binding domain and is highly processive in vitro. Surprisingly, only one to five uridines have been found in vivo on Cid1 target mRNAs even though longer untemplated U-tails can be observed in vitro (Rissland and Norbury 2009). These evidences together with the fact that in vivo Cid1 will preferentially use UTP while in vitro this protein was able to use both ATP and UTP as a substrate brought us to investigate the molecular detail of Cid1 activity. Discovered more than ten years ago, the molecular mechanism by which Cid1 exerts its polyuridylation activity was poorly understood. Thus, structural studies will shed light on how this enzyme accommodates its RNA target as well as let us obtain some insights into how its catalytic activity is performed.

We decided to work with a truncated version of Cid1 based on secondary structure predictions since earlier studies showed that the full-length version of the enzyme was unstable during purification (Fig. 10A; Rissland et al. 2007). We then solved the crystal structure of Cid1 together with its substrate UTP and obtained critical molecular insights on the enzyme’s specificity and function (Munoz-Tello et al. 2012). Cid1/UTP complex allowed us to identify the crucial residues for UTP recognition, highlighting the UTP selectivity of the enzyme. As for other members of the polymerase β family, Cid1 is composed of a catalytic (CAT) domain with the three catalytic aspartic residues and a central domain (CD). The UTP molecule is bound at the bottom of the cleft between the two domains, which compose the
active site. The UTP molecule is stabilized via several direct and water mediated interactions, which are similar to other TUTase crystal structures (Deng et al. 2005; Stagno et al. 2007; Stagno et al. 2010). Two magnesium ions are stabilized in the active site. The first one is stabilizing the triphosphate moiety of the UTP together with the catalytic residue 103 and the second one is found on top of the uracil base, taking the place of the substrate RNA molecule. Surprisingly, histidine residue 336 is hydrogen bonded to the uracil base of the UTP molecule, a stabilization feature that was not shown before this date in any of the known PUP structures, and is the main residue responsible for UTP selection. Indeed, deletion of this residue strongly reduces the specific uridylylating activity of Cid1 enzyme.

When we superpose the NRM region of the PUPs and the PAPs enzymes, we observe that histidine 336 (according to Cid1 numbering) is conserved in all PUPs whereas most of the PAPs have an asparagine at this position likely underlying the importance of this residue in the UTP selectivity (Fig. 39). This was further confirmed by other published crystal structures of the Cid1 enzyme where the authors managed to show a complete abolishment of the uridylylating activity upon mutation of this histidine (Yates et al. 2012). In fact, the authors show that mutating this histidine into an asparagine (which is present at that position in poly(A) polymerases) turned the enzyme into a solely poly(A) polymerase (Yates et al. 2012).

These data have to be taken with care, as in our hands and in the experiments reported by another competing group, i.e. Lunde and colleagues, we only observed reduction of the polyuridylylating activity upon H336 mutation rather than a complete abolition. Further biochemical studies will be needed to be able to definitely turn a PUP into a PAP in vivo and to fully understand the role of this histidine residue on UTP selection in the protein family. In the NRM loop we also find the phenylalanine 332 (F332), which is buried into a hydrophobic pocket and is necessary for the proper folding of the protein. The hydrophobic pocket is composed of residues found in helix 4 and an extended loop of the CD domain. Interestingly, this helix 4 corresponds to the helix F in the yeast PAP enzyme, which is required for the movement of the CAT domain during catalysis (Balbo and Bohm 2007). Furthermore, a similar cavity is observed in the trypanosomal TUT4 protein suggesting a conserved movement of the two domains forming the active site cleft during catalysis in the entire polymerase family (Stagno et al. 2007). When we mutate this hydrophobic residue (F332) into an alanine, we have a very strong decrease of Cid1’s activity underlying the importance of phenylalanine 332 in the proper positioning of the NRM and the CAT domain during
catalysis. Furthermore, other residues found on positively charged patches of the enzyme such as arginine 139 and tyrosine 205 seem to play an important role for RNA recognition. Upon mutation of these residues, Cid1 activity is strongly decreased involving these positively charged patches in the RNA translocation and accommodation during the catalytic cycle of Cid1. Based on these results we propose a catalytic cycle where binding of the triphosphate moiety of the nucleotide leads to a movement of the CAT domain towards the CD domain of the enzyme making helix 4 rotate and thus lock phenylalanine 332 into the hydrophobic pocket immediately below the helix 4 (Munoz-Tello et al. 2012). The uracil base of the UTP is then recognized by the NRM loop giving the time to the RNA molecule to bind and allowing triphosphate hydrolysis for the addition of a single uridine. Upon pyrophosphate hydrolysis, the active site becomes unstable and the CAT and CD domains move apart from each other. This leads to the translocation of the RNA molecule and a new cycle of polymerization will be performed. The conformational changes needed upon Cid1 polyuridylation partially mimic the yeast PAP reorganization upon RNA binding and this flexibility likely influences the processive and distributive mode of action of the enzyme.

![Figure 39: Superposition of the Nucleotide Recognition Motif (NRM) of several eukaryotic poly(U) polymerases (PUPs) and poly(A) polymerases (PAPs). PUP enzymes NRM are framed in yellow whereas PAP enzymes NRM are framed in purple. The PUP enzymes aligned are S. pombe Cid1 (SpCid1), C. elegans PUP1 (CePUP-1) and PUP-2 (CePUP-2), Human ZCCHC11 (HsTUT4) and A. thaliana AT1 (AtAT1). The PAP enzymes used for the alignment are S. cerevisiae TRF4 (ScTRF4), Human PAPD5 (HsPAPD5), S. pombe Cid14 (SpCid14) and Cid12 (SpCid12) and Human GLD2 (HsGLD2). The black triangle represents the conserved residues in the NRM loop. Phe332 (Cid numbering) is very well conserved in PUPs but not in PAPs. His336 (Cid numbering) in PUPs is conserved at this position whereas in PAPs this residue corresponds to an Asparagine highlighting the importance of this Histidine in PUP enzymes. An asterisk underlines other important residues of the NRM.](image)

Our Cid1 structure bound to an ApU molecule shed further lights on the catalytic mechanism and substrate preference of this enzyme. Previous studies have shown the implication of Cid1 enzyme in the polyuridylation of polyadenylated RNAs (Rissland et al. 2007; Rissland and Norbury 2009). Thus, we consider the ApU dinucleotide observed in our crystal structure
as a pseudo-product corresponding to the post-catalysis state of the polyuridylation reaction. The uridine base of the ApU molecule is stabilized as for the uracil part in the Cid1/UTP complex (Munoz-Tello et al. 2012; Munoz-Tello et al. 2014). The adenine base of the pseudo-product is stabilized via several water-mediated interactions with Cid1 but also with a direct hydrogen bond with asparagine 165, the first residue of the helix 4. We further analyzed by mutational studies the importance of asparagine 165. As I mentioned before, helix 4 is required for the global movements of the enzyme upon catalysis (Munoz-Tello et al. 2012). Cid1/ApU binding mode is different to the one found in TbTUT4/UpU complex where a movement of the loop between β-sheets 4 and 5 is necessary for UpU direct stabilization (Stagno et al. 2007). In our case, the same loop does not move when we compare our two structures. This asparagine 165 is identified as a key residue contacting the substrate in the Cid1/ApU crystal. This amino acid does not influence the RNA binding properties of Cid1 but is required for Cid1’s enzymatic activity since mutations of this residue into an alanine lead to the complete abolishment of the enzyme’s activity. The reason behind this might be its involvement in the swivel motion of the enzyme during the catalytic cycle and strategic position at the beginning of helix 4 (Munoz-Tello et al. 2012; Munoz-Tello et al. 2014). Furthermore, our Cid1/ApU complex delineated the molecular bases for selectivity of RNA rather than DNA substrates by showing a direct contact of the aspartate 103 of the catalytic triad with the 2'-OH group of the RNA substrate in our atomic model. This binding mode resembles to the previously described RNA specificity of TUT4 RNA uridylyl transferase from Trypanosome brucei suggesting a conserved mechanism of RNA detection among the entire family (Stagno et al. 2007). Based on electromobility shift assays (EMSA), we showed that Cid1 WT or mutated at the position 160 (active site) or 165 (residue binding to the last base of the substrate) or deleted of its β-trapdoor has a substantially higher affinity for poly(U) than for poly(A) substrates (Munoz-Tello et al. 2014). The β-trapdoor is a very flexible region found above the active site of Cid1, which was suggested to be involved in keeping the UTP inside the cleft (Yates et al. 2012). But when we deplete this structured-region in Cid1 enzyme, we only observed a strong effect on the poly(A) activity of Cid1 thus implicating the β-trapdoor in the polyadenylation activity of the enzyme rather than in its polyuridylation activity. It is important to note that if we mutate the lysine 144, found in the β-sheet 4 of Cid1, the binding to a small stretch of As is impaired whereas its binding to a U15 RNA is not affected by the mutation. This suggests that lysine 144 is required
for the stabilization of short incoming RNA molecule. Yates and collaborators do not observe this effect in their RNA binding assays most likely because they used a stretch of 40 nucleotides (Yates et al. 2012). In the poly(A) polymerase, lysine 144 equivalent residue is lysine 145, which contacts the 2’-OH group of the -2 nucleotide in the PAP ternary crystal structure (Balbo and Bohm 2007). Thus, Cid1 may stabilize its substrate in a manner closer to PAP enzymes. The equivalent residue of lysine 144 in the trypanosomal TUT4 structure is the aspartate 126, which once mutated strongly affects the polyuridylation activity of the enzyme (Stagno et al. 2007). Taken together, these results suggest a mechanism, involving lysine 144, which allows Cid1 to switch from a distributive to a processive mode of action, potentially allowing the addition of two distinct size lengths of U-tails. These second structure allowed us to further complete our catalytic cycle model highlighting the importance of several additional key residues such as the Asparagine 165 for the complete description of the catalytic reaction: (i) the β-trapdoor partially modulates the exchange of UTP after each cycle; (ii) asparagine 165, which is in direct contact with the 3’-penultimate nucleotide in our structure, may be implicated in the translocation of the RNA molecule after catalysis. Finally, as mentioned before, the overall activity of the enzyme clearly depends on global movements of the enzyme around helix 4 and phenylalanine 332 in order to pursue to the next polyuridylation cycle (Munoz-Tello et al. 2012).

Our atomic structures shed new light into the UTP selectivity of Cid1 enzyme allowing us to propose a model for catalysis where local and global movements of the enzyme are crucial for RNA translocation and accommodation. Yates and collaborators, Lunde and collaborators and our group obtained critical information about the RNA recognition mode of the enzyme although a protein structure with a longer RNA target will definitely clarify this matter (Lunde et al. 2012; Munoz-Tello et al. 2012; Yates et al. 2012; Munoz-Tello et al. 2014). Unfortunately, our attempts to obtain diffracting crystals of the RNA complex were not successful. We also acquired insights into the catalytic switch of Cid1 from a distributive to a processive enzyme and underlined its preference for uridyalted substrates. Nevertheless, several questions still remain unanswered and will need further structural, biochemical and genetic studies. We still do not understand how this enzyme selects its targets and whether a cellular mechanism exists to regulate the Cid1-mediated U-tail addition, although we obtained some hints of the Cid1 enzymatic mechanism with our ApU complex structure.
We also do not know why certain mRNAs need to be regulated by uridylation in fission yeast or in other organisms. We propose that Cid1 enzymatic activity is influenced by its strong affinity for the uridine-containing tails. Factors such as Lin28 in humans, Dis3L2 exonuclease or other unknown factors might be necessary to further regulate Cid1’s activity (Heo et al. 2009; Malecki et al. 2013). Indeed, U-tails of up to 15 nt are found in dis3L2 depleted S. pombe cells reinforcing the idea that other factors might regulate Cid1’s activity in vivo (Malecki et al. 2013). It is tempting to speculate that Cid1, as for its human orthologs ZCCHC11, might need a RNA binding protein or other factors to guide the enzyme towards a specific RNA target and regulate its function in the cell (Heo et al. 2009). Deep Cid1 protein partner identification studies will be needed to further clarify this matter.

1.14 RDS Complex Structural and Functional Studies: Light into a Trypanosomal TUTase, RET1

To better understand the role of PUP enzymes in eukaryotes, we started a collaborative work with Ruslan Aphasizhev’s laboratory in Boston on a PUP named RET1, which is responsible for the polyuridylation of gRNAs necessary for the RNA editing reaction and required for the proper mRNA translation (Blum and Simpson 1990; Ernst et al. 2003). The kinetoplast DNA in the mitochondria of trypanosomes contains approximately 40 to 50 maxicircles, which encode rRNA and mRNA precursors, and thousands of minicircles encoding gRNAs. These gRNAs are required for mRNA editing and their biogenesis is not completely understood. Through RET1 TUTase pull-down assays and mass spectrometry analysis, our collaborators discovered a complex containing RET1 plus four proteins, i.e. DSS1 exonuclease and 3 other proteins without any discernable motif named RDS1, RDS2 and RDS3, essential for gRNA biogenesis (unpublished data). So far, the significance of this newly characterized complex in gRNA synthesis and the protein functions are not fully understood except that they are all required for proper gRNA biogenesis. Interestingly, RET1 is also involved in mRNA translation through the formation of long A/U tails, which are required at the 3’-end of mitochondrially-encoded mRNAs in order to be translated (Aphasizheva et al. 2011). The addition of this A/U tail is coordinated by the KPAF1/KPAF2
complex and catalyzed by KPAP1/RET1 complex (Aphasizheva et al. 2011). KPAF1 and KPAF2 proteins contain 20 and 5 PPR repeats respectively. Their main known function is to enhance the poly(A) and poly(U) activity of KPAP1 and RET1 (Aphasizheva et al. 2011). The implications of RET1 in both of these pathways awaits for complete characterization.

In order to better understand the function of this new complex involved in gRNA biogenesis and maturation, we aimed to solve the atomic structure of this complex either as a reconstituted complex or through structural studies of individual subunits. Firstly, we tried to obtain several co-expression and co-purifications of these proteins without any success. We then decided to take each protein alone for expression, purification and crystallization. In this case, trials with DSS1, RDS1, RDS2 and RDS3 proteins were still unsuccessful, most probably because of the large protein size, which might not be very well tolerated in E. coli (Table 3). Expression with protein chaperones or using insect cell Multi-BAC system might lead to successful results (Berger et al. 2004; Prasad et al. 2011). Nevertheless, we were able to express, purify and crystallize a truncated version of RET1 (named RET1S2), which still possess a polyuridylating activity (Fig. 29B and 30B). The enzyme gave plate-like crystals of 2μm thickness and around 200μm length (Fig. 31B). We tested these plates for diffraction at PXI beamline of the Swiss Light Source (SLS) and obtained a complete data set diffracting up to 3.5Å (Fig. 32A). So far, we were able to process the data and scale it but we have just started the analysis for phasing (Fig. 32B). Molecular replacement method using RET2 (PDB code: 2B51), Cid1 (PDB code: 4EP7) and TUT4 (PDB code: 2IKF) will be tried first in order to obtain the phases to calculate an electron density map using our diffraction data (Fig. 33; Deng et al. 2005; Stagno et al. 2007; Munoz-Tello et al. 2012). These protein models will need to be modified and adapted for molecular replacement. For example, RET2 enzyme possesses an additional non-conserved domain, which will need to be removed in order to pursue the phasing experiment. Another option will be to use all the three PDB files as an ensemble for molecular replacement. Indeed, we will superpose the three models and remove the most variable parts leaving only the main structured parts of each protein and use them for phasing. At the same time, seleno-methionine derivative of the truncated RET1 protein is being prepared and will be tested for crystallization trials in case our molecular replacement attempts with RET2, TUT4 and Cid1 enzymes remain unsuccessful. Interestingly, RET1 is suspected to bind zinc ion through the presence of four conserved
cysteines in the N-terminal region. RET1S2 crystals contain zinc ion(s) as demonstrated by
the fluorescence scan performed on RET1S2 crystal. This signal could also be helpful for
phasing but higher diffracting crystals will be needed as the tested one were not recorded at
the optimal wavelength. Lastly, if these other phasing methods fail to give us a proper
density map for building RET1S2, soakings with heavy atoms will be required for single
anomalous scattering phasing method as used for the determination of the Cid1/UTP
structure (Munoz-Tello et al. 2012). Since RET1 is a PUP with two different type of substrates
(gRNAs and mRNAs) it will be interesting to try to obtain the crystal structure of RET1 in
complex with each of these RNAs to better understand the RNA recognition mode of the
enzyme as well as how the RNA will be accommodated within the protein. It is interesting to
note that gRNAs are supposed to have two stem-loops within the RNA, which differs from
mRNAs. Obtaining crystals with both a gRNA and a mRNA will help us understand how this
PUP enzyme is able to recognize such a different type of substrates for polyuridylation. As
for Cid1 and TUT4 PUPs, co-crystallization and soaking trials with UTP alone and with UpU or
UMP molecules will be tried in order to try to block a binary and/or ternary complex in the
active site of RET1S2. If needed, RET1S2 enzyme will be mutated into an inactive protein to
further trap RNA molecules in the active site as successfully done by us with Cid1/ApU
complex and the PAP/ATP/RNA ternary complex (Balbo and Bohm 2007; Munoz-Tello et al.
2014).

As I mentioned before, RET1 is required for proper mRNA translation in the mitochondria of
trypanosomes through the formation of long A/U tails together with KPAP1 poly(A)
polymerase and enhancer factors KPAF1 and KPAF2 (Aphasizheva et al. 2011). We then tried
to obtain a KPAP1/RET1 stable complex or KPAP1 protein alone (Fig. 34 and Fig. 35; Table 6).
Unfortunately, we were not able to successfully purify KPAP1 poly(A) polymerase in E. coli
cells. Several new constructs based on secondary structure predictions and changing the
position of the tag from N-terminal to C-terminal will need to be tested to try to purify
KPAP1 enzyme as tried for RET1. If the truncated versions of KPAP1 are purified, activity
tests with a mRNA molecule will have to be performed to confirm that our construct is still
active. To fully characterize this polyadenylation/polyuridylation complex, another PhD
student in our laboratory is studying the structural and functional bases of KPAF1 and KPAF2
complex.
In both of these projects, RET1 structure will bring further light into how this enzyme is bound by other proteins and the ways in which RET1 protein partners influence the activity and substrate recruitment of the enzyme. Thus, the structures may help understanding how this TUTase is able to make complexes with different RNA substrates and proteins in different pathways in trypanosomal mitochondria. For this purpose, obtaining protein/protein structures of RET1 with any of the above-mentioned enzymes would help to answer at least partially this question. Furthermore, atomic models of RET1 with either a gRNA or a mRNA may help us understand the complex mode of RNA binding of RET1 bringing some interesting molecular insights into PUP enzymes RNA selectivity and conformational flexibility.

Polyuridylation was for a longtime an underestimated 3’-end modification; most probably because sequencing techniques were focused on polyadenylated RNAs. With the development of new and adapted techniques to detect 3’ uridylation, this event is starting to gain strength with impacting roles in RNA degradation and stability (Choi et al. 2012; Clamer et al. 2014). RNA sequencing analysis of mammalian cells, not depending on oligodT primers but rather using 3’ ligated linkers specific for small RNAs of 200nt or less, showed a widespread tendency of 3’-end uridylation of small RNAs (Choi et al. 2012). Interestingly, besides the already known uridylated targets, they also found this 3’ modification on transcriptional start-site-associated RNAs along with spliced introns suggesting a larger role of polyuridylation in RNA metabolism in mammals, and despite the fact that PUPs are mostly localized in the cytoplasm. Optimized RNA sequencing methods in different backgrounds, such as DNA replication inhibition and stress conditions, and refinements in these methods are necessary to understand the global biological consequences of uridylation in RNA metabolism and might also be the start for uncovering new nCNRs among eukaryotes. With RNA-Seq development, more and more RNAs are found to be uridylated in several organisms but the enzymes responsible for this process are still unknown. The identification of polyuridylyating enzymes becomes now critical for obtaining a larger picture of polyuridylation in eukaryotes, its evolution and functional implications in the cell.

It is important to note that RNA uridylation in the cytoplasm has been shown to induce tumorogenesis in mammals. Uridylation at the 3’-end of the tumor suppressor pre-let-7 microRNA by cytoplasmic ZCCHC11 and ZCCHC6 enzymes blocks let-7 miRNA maturation,
which in turn stimulates tumor growth (Heo et al. 2009). Lin28 is a factor of pluripotency in stem cells and once expressed, it helps the maintenance of an undifferentiated and proliferative state by blocking the expression of let-7 miRNA by recruiting ZCCHC11 for uridylation-mediated decay (Yu et al. 2007; Heo et al. 2009; Chang et al. 2012). In adult somatic cells, Lin28-let-7 pathway is normally silenced even though we still observe expression of LIN28A or LIN28B in a wide variety of human cancers (Viswanathan et al. 2009; Piskounova et al. 2011). Inhibition of this oncogenic pathway blocks the tumorigenicity of cancer cells (Piskounova et al. 2011). It has recently been shown that modified let-7 microRNAs are degraded by DIS3L2 exonuclease (Chang et al. 2013). Furthermore, DIS3L2, which preferentially trims uridylated cytoplasmic RNAs, has been found mutated in patients with Perlman syndrome and in some cases this mutation lead to the development of Wilm’s tumor at early stages of child’s growth (Chang et al. 2013). Even though RNA uridylation has been linked with tumor growth, the biological significance of such event is still poorly understood and as such is being studied. In order to better understand tumorogenesis, it is necessary to identify the RNA targets as well as the protein partners that recruits either the RNA or the poly(U) polymerase. Such information will allow the in depth studies of the link between TUTases and diseases. Furthermore, structural and biochemical studies of substrate recognition by TUTases and ncPAPs will provide a rational foundation for therapeutic purposes. In kinetoplastid organisms, this information will bring new insights into U-insertion/deletion, gRNA biogenesis and translational efficiency required for parasite survival and thus, helps for the design of new trypanocides important to treat various trypanosomal diseases including the fatal human sleeping sickness.

In summary, our structural and functional studies on Cid1 PUP and the ongoing work on RET1 TUTase help to our understanding of PUPs flexibility and substrate interactions, completing the work done on TUTases implicated in the RNA editing process taking place in the mitochondria (Deng et al. 2005; Stagno et al. 2007; Stagno et al. 2007; Stagno et al. 2010). Nevertheless, in order to complement our structural results, functional tests based on in vivo studies would bring a better understanding of these enzymes. In particular, studies of Cid1 H336A, K144A or even the β-trapdoor deleted mutants in cells lacking Cid1 WT would allow a precise understanding of their role during substrate recognition and more generally polyuridylation function. Furthermore, our studies combined with other PUPs structural and
bioinformatic studies provide crucial structural and architectural information on PUP’s mode of action providing then a strong base for testing inhibitors for therapeutic benefits. It is important to note that 3’ uridylation is involved in several key aspects of the RNA biology and all the proteins implicated in this process in eukaryotes are not fully known, bringing into focus the importance of further studies concerning this event and its players. Several research groups nowadays started to focus their work on identifying new ncNTrs along with their targets and possible protein partners. Thus, we will most likely hear a lot more about PUPs and their influence in RNA metabolism and turnover during the coming years.
Abbreviations

- A  Adenine
- ADAR  Adenosine Deaminase that Act on RNA
- ADAT  Adenosine Deaminase that Act on tRNA
- AI  Autoinduction
- AT  After TEV cleavage
- BT  Before TEV cleavage
- C  Cytosine
- CAT  Catalytic Domain
- CD  Central Domain
- CDE-1  Cosuppression defective 1
- Cid1  Caffeine-induced death suppressor 1
- DNA  Deoxyribonucleic acid
- dsRNA  double-stranded RNA
- ESRF  European Synchrotron Radiation Facility
- FL  Full-Length
- G  Guanine
- GRBC  gRNA binding complex
- gRNA  Guide RNA
- GST  Glutathione-S-transferase
- h  Hours
- HESO-1  HEN1 SUPPRESSOR 1
- IPTG  Isopropl-β-D-thiogalactopyranosine
- KPAF1  Kinetoplast Polyadenylation / uridylation Factor 1
- KPAF2  Kinetoplast Polyadenylation / uridylation Factor 2
- KPAP1  Kinetoplast poly(A) polymerase
- LB  Luria Broth
- lincRNA  Long non-coding RNA
- MBP  Maltose Binding Protein
- MD  Middle Domain
- MEAT 1  Mitochondrial Editosome-like complex Associated TUTase 1
- min  Minutes
- miRNA  Micro RNA
- MNT  Minimal Nucleotidyl Transferase
- MR  Molecular Replacement
- mRNA  Messenger RNA
- Mw  Molecular weight
- ncNTrs  Non-canonical ribonucleotidyl transferase
- ncRNA  Non-coding RNA
- nPAP  Nuclear Poly(A) Polymerase
- NRM  Nucleotide Recognition Motif
- nt  Nucleotide
- O/N  Overnight
- ORF  Open Reading Frame
- PAP: Poly(A) Polymerase
- PEG: Polyethylene Glycol
- piRNA: PIWI-interacting RNA
- Pol: Polymerase
- PPR: Pentatricopeptide repeats
- PUP: Poly(U) Polymerase
- RBD: RNA binding domain
- RDS: RET1 TUTase DSS1 exonuclease
- RET1: RNA editing TUTase 1
- RET2: RNA editing TUTase 2
- RNA: Ribonucleic acid
- RNP: Ribonucleoprotein
- rNTrs: Ribonucleotidyl transferase
- RRM: RNA recognition motif
- rRNA: Ribosomal RNA
- SAD: Single-wavelength anomalous diffraction
- SeMet: Seleno-methionine
- siRNA: Small interfering RNA
- SLS: Swiss Light Source
- snoRNA: Small nucleolar RNA
- snoRNP: Small nucleolar ribonucleoprotein
- snRNA: Small nuclear RNA
- snRNP: Small nuclear ribonucleoprotein
- sRNA: Small RNA
- T: Thymidine
- TB: Terrific Broth
- TEV: Tobacco Etch Virus
- tRNA: Transfer RNA
- TUTase: Terminal Uridyl Transferase
- U: Uracil
- URT1: UTP:RNA uridylyl transferase 1
- UTR: Untranslated region
- WT: Wild-type
- ZCCHC11: Zinc finger, CCHC containing 11
- ZCCHC6: Zinc finger, CCHC containing 6
- β-MT: β-mercaptoethanol


Anderson, J. S. and R. P. Parker (1998). The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SK12 DEVH box protein and 3' to 5' exonucleases of the exosome complex. EMBO J 17, 1497-1506.


Hsu, C. L. and A. Stevens (1993). Yeast cells lacking 5'--->3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. Mol Cell Biol 13, 4826-4835.


Martin, G. and W. Keller (1996). Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. EMBO J 15, 2593-2603.


Mullen, T. E. and W. F. Marzluff (2008). Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev 22, 50-65.


