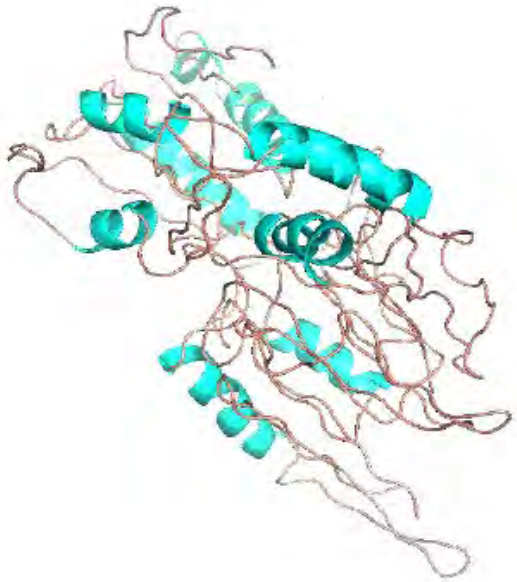
UPPSALA
UNIVERSITET

Giardia lamblia poly (A) polymerase

Structural and functional characterization



Albena Draycheva

Degree project in biology, Master of science (1 year), 2008
Examensarbete i biologi 30 hp till magisterexamen, 2008
Biology Education Centre and Department of Cell and Molecualr Biology, Uppsala University
Supervisor: Anders Virtanen

Contents

Abbreviations

amp	ampcicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
BSA	bovine serum albumin
BV	bed volume
CCAtrs	ATP/CTP:tRNA nucleotidyl transferase
CstF	cleavage stimulation factor
CTP	cytidine triphosphate
D	aspartic acid
ddH_2O	Deionized, distilled water
gPAP	giardial PAP
GRAVY	grand average hydropathicity
GSH	reduced glutathione
GST	glutathione-S-transferase
MW	molecular weight
NMD	nonsence mediated decay
Ntr	nucleotidyl transferase
PABP	poly(A) binding protein
PDB	protein data bank
pI	isoelectric point
R	arginine
RBD	RNA binding domain
T	threonine
TCA	trichloroacetic acid
UTR	untranslated region
Y	tyrosine

Summary

In the nucleus of eukaryotic cells most of the messenger RNA (mRNA) receives a 3'-extension modification of 200 adenosine residues, a poly(A) tail. It is important for the mRNA stability, nuclear export and translation initiation. It is synthesized in a two-step reaction, called polyadenylation. In the first step the precursor (pre-mRNA) is endonucleolytically cleaved, which generates a 3'-hydroxyl group. In the second step, poly(A) polymerase (PAP) processively starts adding adenine residues to the 3'-end. The protein machinery involved in this reaction is very complex and rich in components. Thus, the reconstitution of the polyadenylation reaction using recombinant proteins is hindered.

Recently, the genome sequence of the parasitic microorganism *Giardia lamblia* has revealed a simplified polyadenylation machinery, which could be adapted for *in vitro* reconstitution of the reaction with recombinant proteins.

The goal of the current project was to elicit whether the protein identified as PAP actually performs a polyadenylation reaction and further develop an *in vitro* polyadenylation system with recombinant proteins. For this reason an expression and purification protocol had to be established and later the protein was to be tested for activity. Different online protein analysis tools were used to predict specifications of giardial PAP as isoelectric point (8.08), molecular weight (77 kDa), stability of the protein, etc. Secondary structure predictions showed the catalytic site was preserved, but no RNA binding domain was present. Instead there was a hydrophobic domain. These results made it possible to improve the expression and purification protocol and to obtain more soluble giardial PAP. Different modifications of the purification system used could not overcome the low elution quality of giardial PAP. In the activity assay no polyadenylation activity could be detected.

The results from the computational analysis and the purification trials led to the hypothesis that the low solubility may be because giardial PAP is membrane bound. Further the *in silico* analysis showed lack of RNA binding domain which could be the reason for the inability to recover polymerization activity in PAP preps. This also suggested that giardial PAP probably functions together with a protein that serves as an RNA-binding domain. Not all possibilities were exhausted during the project. Thus, in the future another expression system should be tested, in order to verify the assumptions made here.

Introduction

RNA polyadenylation

RNA polyadenylation has been observed in all three domains of life: eukaryotes (Edmonds & Abrams, 1960), bacteria (O'Hara *et al.*, 1995) and archaea (Brown & Reeve, 1986). The study of the polyadenylation began in the 1960s' and continues up until the present, by establishing the essential components for the reaction and elucidating the different roles it plays in cell metabolism.

In eukaryotes, polyadenylation is carried out in the nucleus, right after transcription of messenger RNA (mRNA) by RNA polymerase II (RNA Pol II). This is one of the three processing stages to which the newly synthesised transcript is subjected before being ready for translation. It receives a 7-methylguanine cap at the 5'-end (cap), an extension of around 200 adenine residues at the 3'-end (poly(A) tail) and finally it is spliced by excision of the introns and ligation of the exons.

The polyadenylation reaction comprises two coupled steps. In the first step, the pre-mRNA is endonucleolytically cleaved at the 3'-end. This generates a 3'-hydroxyl group (3'-OH) (Wahle & Keller, 1992; Wahle & Keller, 1996). In the second step, the enzyme poly(A) polymerase (PAP) starts adding adenine residues to the newly generated 3'-OH of the RNA (Sheets & Wickens, 1989)(Figure 1) .

The poly(A) tail serves different purposes in the mRNA life cycle, mostly by facilitating the binding of poly(A) binding proteins (PABP). It stabilises the mRNA (Jackson & Standart, 1990), stimulates its nuclear export (Chekanova & Belostotsky, 2003), and facilitates translation initiation (Gallie, 1991; Gallie, 1998). Recently, it has been shown that the poly(A) tail takes part in mRNA turnover, e.g. nonsense-mediated decay (NMD) (Dzikiewicz & Szweykowska-Kulińska, 2006) .

During cleavage a multi-protein complex is assembled at the 3' untranslated region (UTR) of the mRNA (Figure 1). One of the more important multi-subunit proteins is cleavage/polyadenylation specificity factor (CPSF). It binds to the polyadenylation signal (AAUAAA-hexanucleotide) at the 3' UTR . This is approximately 30 nucleotides upstream from the cleavage site. The 160 kDa subunit of CPSF possesses a binding motif which recognises the AAUAAA sequence (Murthy & Manley, 1995). It is proposed that the 30, 73 and 100 kDa subunits are involved in the cleavage reaction, but still no exact mechanism has been presented (Mandel *et al.*, 2006; Mandel *et al.*, 2006; Zarudnaya *et al.*, 2002). What has been shown experimentally is that PAP binds to CPSF, which enhances the polymerization processivity (Kaufmann *et al.*, 2004).

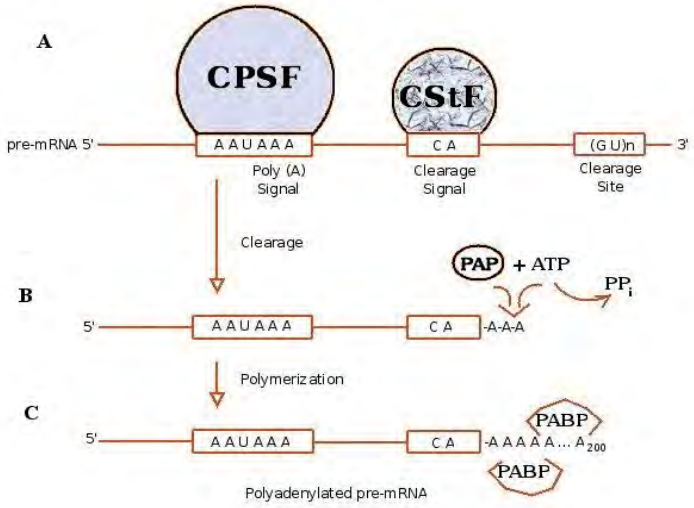


Figure 1: Schematic overview of nuclear polyadenylation.
A. The figure illustrates the general 3' UTR elements of the pre-mRNA necessary for the polyadenylation reaction. The cleavage/polyadenylation stimulatory factor (CPSF) binds to the polyadenylation signal (AAUAAA) and the cleavage stimulatory factor (CstF) binds to the cleavage signal (CA). Then CPSF performs the cleavage reaction. **B.** poly(A) polymerase starts adding adenosine residues, from ATP, to 3'-end of the pre-mRNA. **C.** During the polymerisation reaction poly(A) binding proteins (PABP) bind to the emerging poly(A) tail.

Poly(A) polymerase

The poly(A) polymerase (PAP) was first discovered in the 1960s' in a nuclear extract from calf thymus (Edmonds & Abrams, 1960). Ever since, PAPs have been found in bacteria, plants, vertebrates, etc. They have been classified extensively according to structure, substrate specificity and function.

Eukaryotic poly(A) polymerase (PAP) belongs to class I nucleotidyl transferases (Ntrs) along with DNA polymerase β (Pol β), kanamycin nucleotidyl transferase (KanNtr) and archeal ATP (CTP):tRNA nucleotidyl transferase (CCANtr) (Martin G & Keller W , 2007). The first solved structure of a template-independent polymerase was PAP from *Bos taurus* (bovine PAP) (Martin G *et al.*, 2000).

Structure and active site

Bovine PAP consists of a single peptide with three distinguished domains (Figure 2, A and Figure 3). The N-terminal domain is built up by a five-stranded β-sheet and two α-helices. It is homologous to the nucleotidyl transferases and is characterized as the catalytic domain of PAP. (Figure 2, B). The central domain follows the N-terminal and is linked to it by β-strand. It comprises four-helix bundles (Figure 2, C) that show structural homology to the allosteric activity domain of the ribonucleotidyl transferase R1. The C-terminal domain is linked to the central domain through a flexible hinge region. The domain is built up by four-antiparallel β-sheets and two α-helices (Figure 2, D), forming an α-β plaits topology. This

topology is similar to that of the RNA binding proteins and is thus regarded as the RNA-binding domain (RBD) of PAP. The C-terminal domain also contains two nuclear localization signals (NLS-1 and NLS-2) and regulatory regions of PAP (Martin *et al.*, 2000).

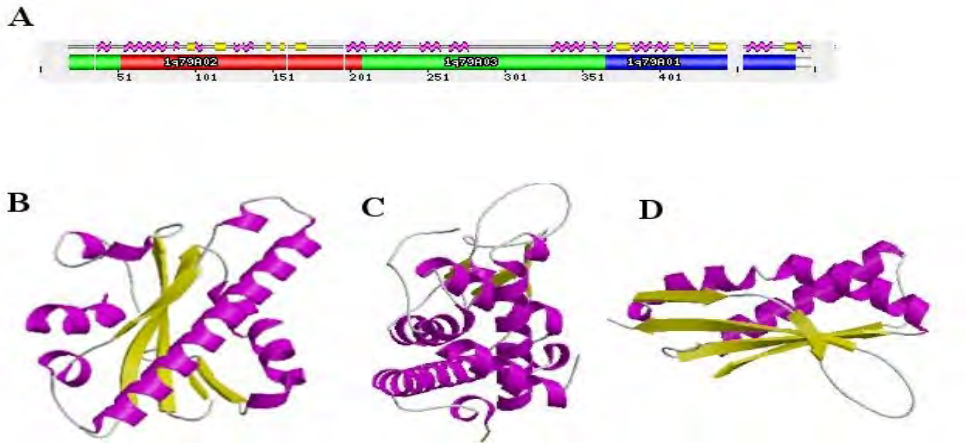


Figure 2: Bovine poly(A)polymerase domain topology (pictures were taken from CATH database for PDB entry 1q79).
According to the CATH database (Orengo et al., 1997) bovine poly(A) polymerase is built up from three domains. Their distribution along the protein sequence is shown in Panel A. 1q7902, which corresponds to the Ntr domain is shown in red (panel B); 1q7903 which corresponds to the central poly(A)polymerase domain is shown in green (panel C) and 1q7901 which corresponds to the RNA binding domain (RBD) in dark blue (panel D).

Figure 3: Domain architecture of the bovine poly(A)polymerase (PDB ID 1q79).
The nucleotidyl transferase domain (Ntr) is shown in red, the central poly(A) polymerase domain in green and the RNA binding domain (RBD) in dark blue. The active site of bovine poly(A) polymerase is located between the Ntr and the central domain. The ATP in the active site is shown in orange, green and blue (The pictures was generated with PyMol).

The structure reveals that the catalytic reaction depends on magnesium ion. The magnesium ions bind as co-substrates ($Mg{\cdot}ATP^{2-}$ and $Mg{\cdot}PPi^{2-}$) and stabilize the transition states by coordinating the 3'-hydroxyl and the α-phosphate of the incoming ATP (Balbo *et al.*, 2005). By analysing the structure, residues K228, K232 and Y237 have been identified as crucial for coordinating the γ-phosphate and T317 for recognition of the incoming adenine base (Martin *et al.*, 2004). Through alignment studies and mutational analysis, D113, D115 and D167 have been found to be core residues for the reaction catalysis (see Figure 4) (Martin & Keller, 1996).

Evolution

The phylogenetic studies of transferases from the Pol β-Ntrs group have suggested a rapid evolution according to the needs of the organism (Martin & Keller, 2007).

There are two models proposed for the evolution of PAPs. The first one suggests divergence of the proteins, e.g. the development from a simple enzyme to a more sophisticated one according to the specific needs of the organism (Aravind & Koonin, 1999). The second one relies on the interconversion theory of PAPs and CCAtrs. This hypothesis suggests that CCAtrs have obtained an RNA recognition motif (RRM) that recognizes mRNA, but not tRNA, and consequently the active site has changed just to perform the step of adding adenine (Yue *et al.*, 1996).

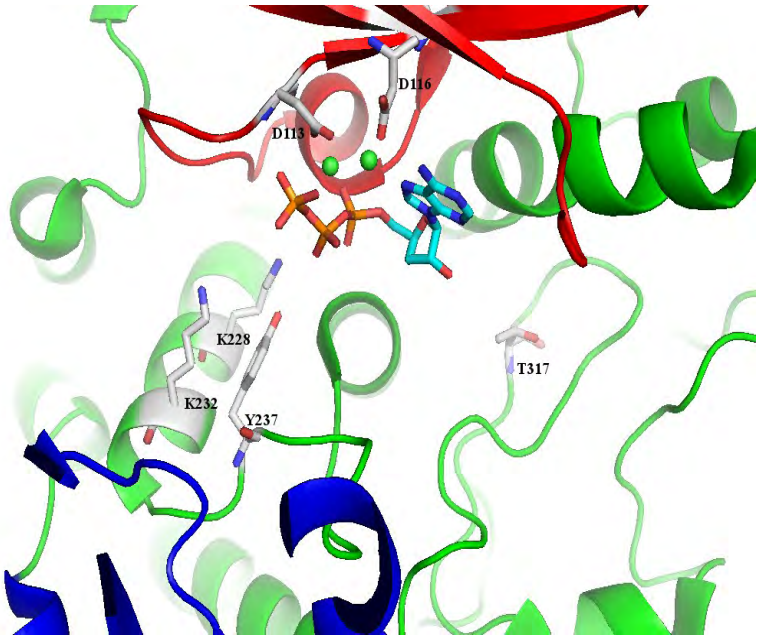


Figure 4: Poly(A)polymerase active site (PDB ID 1q79).
The Ntr domain is shown red, the central domain in green. The RBD is partially visible in dark blue. Residues D113 and D116 coordinate the Mg^{2+} (green) that stabilize the transition states during the catalytic reaction, while residues K228, K282 and Y237 recognize the incoming ATP. The picture was generated with PyMol.

Giardia lamblia

Giardia lamblia is a eukaryotic unicellular intestinal parasite that infects vertebrates (see Figure 6). It is one of the prime inducers of waterborne diarrhoea, also know as giardiasis, worldwide. Like most parasites, *Giardia* is extremely dependent on metabolites derived from the host, such as nucleotides and amino acids. Thus, it does not have the full set of enzymes typical for most free-living microorganisms (Adam, 1991)

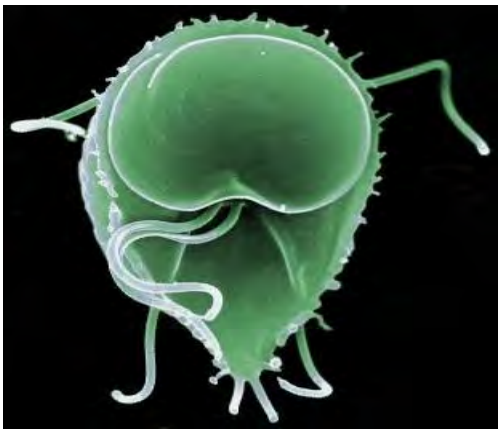


Figure 5: *Giardia lamblia* trophozoite.Scanning electron micrograph, false color. (Credit: Joel Mancuso, University of California, Berkeley; free source)

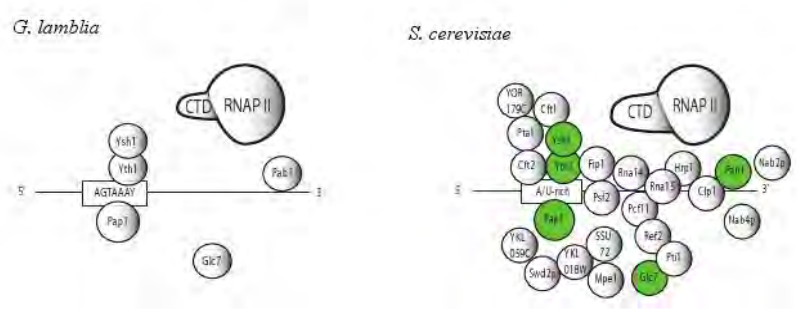


Figure 6: Comparison between *G .lamblia* and *S. cerivisiae* polyadenylation machinery. The figure was kindly provided by Staffan Svärd.

Recently, the genome of *G. lamblia* has been sequenced. The analysis of the data showed a simplified machinery for DNA replication, transcription and RNA processing. Despite being a eukaryote, *G. lamblia* exhibits signs of lateral gene transfer from bacteria, archaea and from a previously unidentified gene family (Morrison *et al.*, 2007).

The study of the less complex polyadenylation machinery (Figure 5) in *Giardia*, aims at the establishment of an *in vitro* system to reconstitute the whole polyadenylation reaction with recombinant proteins.

Aim

The aim of my study was to test the putative giardial poly(A) polymerase for activity. In order to do this, predictions about the biophysical and structural properties of the protein had to be acquired by computational analysis, which were used in the development of an expression and purification protocol. The obtained pure giardial poly(A) polymerase was to be used to develop an *in vitro* polyadenylation system with recombinant proteins.

Results

In silico *analysis of giardial gene with Entrez ID 5697771 confirm it is a putative poly(A) polymerase*

To analyse the domain organisation and the resemblance of the protein product of the giardial gene (Entrez ID 5697771) to poly(A) polymerases (PAP), I used the Pfam database (Finn *et al*., 2006). The main part of the peptide was homologous to the PAP central domain with an E-value[1] of $5.6x10^{-8}$. A nucleotidyl transferase (Ntr) domain was recognised near the N-terminus. The C-terminus was similar to the fungal ATP-synthase protein 8 (A67), which is part of the membrane associated domain of the ATP-synthase (Macreadie *et al.*, 1983). The predictions strongly suggested that the giardial protein is part of the Ntr family and may possesses ATP specificity. In contrast to canonical PAPs, gPAP did not have an RNA binding domain.

To investigate if gPAP had the conserved catalytic residues, I used Jpred: PDB Homologues search tool (Cuff & Barton, 2000). It searches for the closest sequence homologues in the structural data base (PDB (Berman *et al.,* 2000)). The three reliable results, with a Blast E-value of $2x10^{-23}$, were the three entries for bovine PAP (PDB ID 1q78, 1q79, 1f5a). An alignment of bovine PAP and gPAP (Figure 7) showed that the amino acids important for the catalytic reaction, D113, D115 and D167, in bovine PAP were conserved as D176, D178 and D230 in gPAP. The residues responsible for the coordination of the incoming ATP, K228, K232 and Y237, were conserved in gPAP as K293, R295 and Y304. It seemed that the functional group of K228 (*B. taurus* numbering) had been replaced by another positive amino acid R295 (*G. lamblia* numbering)(Figure 8).

```
gPAP: 86  KEQFQLVDNLLLSYLKNCKLYVSSEETRKKEAILRNLELIISDWVYETYKDAFLANMQFY 145
           KE   L+   L+  LK   ++   EE +++  IL  L  ++ +W+ E
1q79: 32  KETDCLLTQKLVETLKPFGVFEEEEELQRRILILGKLNNLVKEWIREI------------ 79

gPAP: 146 RKQPNMPRS--ERLV-RLYPFGSYYLGINEPSSDVDTVVIFPQYVRIADFFDKFPTIIGQ 202
           +  N+P+S  E +  +++ FGSY LG++   +D+D + + P++V  +DFF  F   +
1q79: 80  SESKNLPQSVIENVGGKIFTFGSYRLGVHTKGADIDALCVAPRHVDRSDFFTSFYDKLKL 139

gPAP: 203 MPQVSYFDCIRDAKVPLITLTYDSVDFDLSAAVMAVNRVTDAIPFLDPECTANMHEKSIL 262
             +V     + +A VP+I L +D ++ D+  A +A+  + + +   D     N+  + I
1q79: 140 QEEVKDLRAVEEAFVPVIKLCFDGIEIDILFARLALQTIPEDLDLRDDSLLKNLDIRCIR 199

gPAP: 263 SLNGYRTNIHVKSQFAGNNIFYSTFQNAVRALRLWCKRKNIYSNRCGFFGGINCIILVA 321
           SLNG R    +             F+  +RA++LW KR NIYSN  GF GG++  +LVA
1q79: 200 SLNGCRVTDEILHLVPN----IDNFRLTLRAIKLWAKRHNIYSNILGFLGGVSWAMLVA 254
```

Figure 7: Amino acids alignment of bovine and giardial poly(A)polymerases.
The conserved amino acids in the active site of bovine and giardial PAP are shown in red.

1 The E-value is the probability of finding the same or a better match by random chance. It is calculated by the following formula: $E = Kmn\, e^{-\lambda S}$, where S is the similarity score of the alignment, K and λ are scaling parameters that allow E values from difference searches to be compared, n is the size of the database and m is the length of the query sequence. An E-value of 1 assigned to an alignment means that it is just as likely to have occurred by chance alone (Karlin & Altschul, 1990).

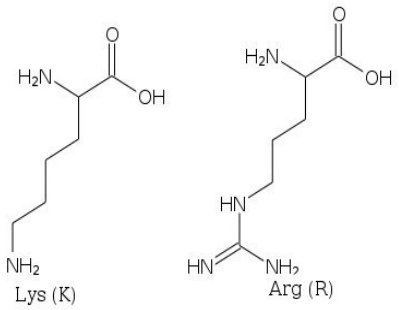


Figure 8: Positively charged amino acids lysine (on the left) and arginine (on the right).

Furthermore, to investigate the apparent lack of the RBD, a 3D structure prediction was performed using EsyPred3D tool (Labert e*t al.,* 2002). The derived structure was visualised in the PyMol molecular graphics program (DeLano, 2007) (Figure 9). A lot of the secondary structure elements were interpreted as loops, which indicated that these regions were predicted as thermodynamically unstable. Superimposition of the predicted structure of gPAP on the crystal structure of bovine PAP (1q79) (see Figure 10) suggested that the overall U-like shape of the central domain of bovine PAP was conserved. At the N-terminus, where the catalytic domain (Ntr domain) in bovine PAP is located, the structure was loop-rich and could not be defined more precisely. At the C-terminus there was no overlap between the two structures, because gPAP was shorter than 1q79. This verified the analysis from Pfam; lack of RBD.

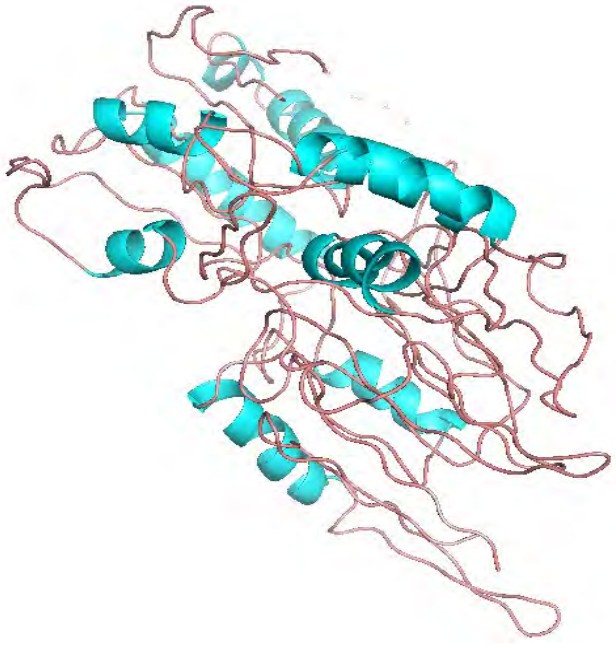


Figure 9: 3D structure prediction of giardial poly(A)polymerase.
The predicted helical structures are shown in light blue and the predicted loop regions are in light pink.

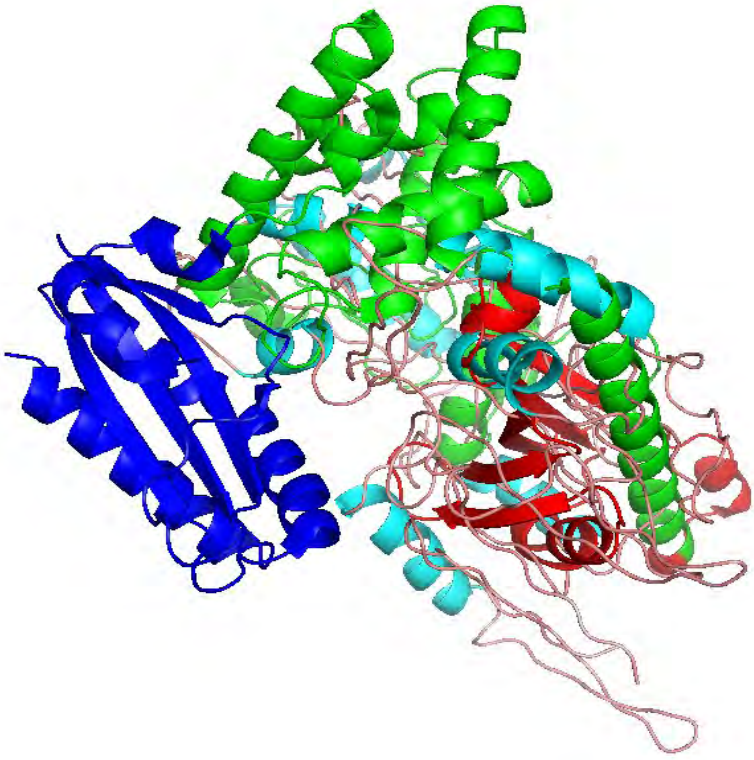


Figure 10: Superimposed giardial and bovine poly(A) polymerases.
Bovine poly(A) polymerase (PDB ID 1q79) in red, green and dark blue was superimposed on the predicted structure of giardial poly(A) polymerase. The colour code is maintained the same as in the other structure pictures: gPAP is in light blue and pink and it overlaps with the Ntr domain (red) and the central domain (green) of bovine PAP, but not with the RNA binding domain (dark blue).

Structural predictions for giardial PAP suggested a membrane-bound protein

In order to establish an efficient purification protocol, I needed more information about the biophysical parameters of gPAP. To predict them, I used ExPASy ProtParam tool (Lambert *et al.* 2002). The expected properties of gPAP were compared with those of bovine PAP. The results are summarised in Table 1. Both proteins were predicted to be fairly unstable with an *in vitro* half-life of 30 hours. This suggested that the purification technique should be straight forward and allow activity experiments within 30 hours after cell lysis. The pI of 8.08 indicated a basic protein, which should be considered in case of application of an ion-exchange chomatography.

According to the grand average hydropathicity index[3] (GRAVY) (for explanation of the term refer to Table 1), gPAP was a hydrophilic protein, but in comparison with bovine PAP less tolerant to water. In support of that, myrystyl motifs were recognised (by the same tool) and one of them was part of a prokaryotic lipoprotein motif (N-terminal location). Since the myrystyl tail is a hydrophobic sequence used for attaching proteins to the membrane, its presence pointed gPAP as a putative membrane-bound protein.

Table 1: Predicted parameters of giardial and bovine poly(A)polymerases

Protein	Number of amino acids	Molecular weight (Da)	Theoretical pI	Estimated half-life in vitro[1] (hours)	Aliphatic index[2]	Grand average hydropathicity (GRAVY)[3]	Instability index[4]
giardial PAP	674	77269	8.08	30	88.96	- 0.230	53.71[5]
bovine PAP	740	82441	6.96	30	89.30	- 0.343	57.12[5]

1. Prediction of the time it takes for the half of the amount of the protein in a cell to disappear after its synthesis in the cell. It relies on the "N-end rule" which relates the half-life of a protein to the identity of its N-terminal residue (Varshavsky, 1997).

2. Relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine).

3. The GRAVY value for a peptide or a protein is calculated as a sum hydropathy values of all the amino acods, devided by the number of residues in the sequence. Hydrophilic proteins have negative GRAVY (Kyte & Doolittle, 1982).

4. Provides an estimate of the stability of your protein in a test tube. It is calculated by the formula X(Ala) + a * X(Val) + b * (X(Ile) + X(Leu)), where X(Ala), X(Val), X(Ile), and X(Leu) are mole percent (100 X mole fraction)

of alanine, valine, isoleucine, and leucine. The coefficients a and b are the relative volume of valine side chain (a = 2.9)

and of Leu/Ile side chains (b = 3.9) to the side chain of alanine.

5. Predicted as unstable.

Giardial PAP forms inclusion bodies

The giardial poly(A) polymerase (gPAP) gene already had been cloned in the group of Staffan Svärd. The construct generated a fusion protein with an N-terminal glutathione S-transferase-tag (GST). I expressed the fusion protein in *E. coli* obtaining high yield of the gPAP-GST fusion protein (Figure 11, lane 3), but most of it was insoluble, located in inclusion bodies (Figure 11, lane 2). To purify PAP from the inclusion bodies, I treated the precipitate with 4 M urea and then refolded PAP by dialysis. This procedure solubilized around 20 % of the gPAP (data not shown). To test if the protein was properly refolded, I performed affinity purification with GST-affinity Sepharose. The solubilized protein exhibited no affinity to the matrix (data not shown), showing that the tag was folded incorrectly and therefore the whole construct most likely was misfolded.

In order to avoid the formation of inclusion bodies, I induced gPAP expression at higher cell density for a shorter period. Then a non-ionic detergent (Sarcosyl), which would bind to the hydrophobic patches and prevent aggregate formation, was added to the lysis buffer. The comparison of the amount of pelleted gPAP before addition of Sarcosyl (Figure 11, lane 2) and after its addition (Figure 12, lane 2) shows that gPAP had become more soluble in the presence of Sarkosyl.

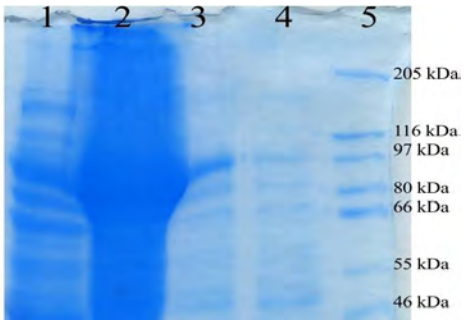


Figure 11: Giardial poly(A)polymerase forms inclusion bodies.
SDS-PAGE showing expression of gPAP. Lane 1, clear lysate; lane 2, insoluble gPAP located in the pellet, lane 3, induced culture; lane 4, culture before induction with 0.1 mM IPTG; lane 5, molecular marker. See Materials and methods section for details.

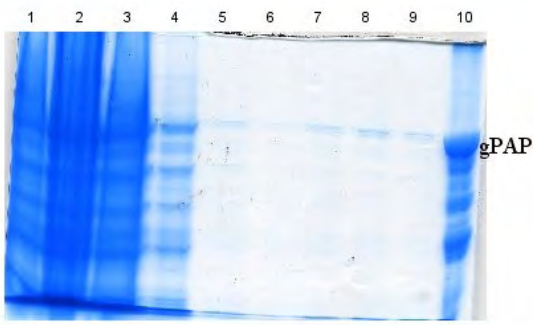


Figure 12: Improved expression and purification protocol of giardial poly (A) polymerase with glutathione-S-transferase affinity Sepharose.
SDS-PAGE
Lane 1, clear lysate; lane 2, pellet; lane 3, flow-through; lane 4, 1st wash 300 mM NaCl; lane 5, 2nd wash 100 mM NaCl; lane 6 - 3rd wash no salt; lanes 7 - 9 - elution with 5 mM GSH; lane 10 - remaining proteins on the beads. See Materials and methods section for details.

Giardial poly(A) polymerase binds unspecifically to glutathione-coupled Sepharose

The primary purification step of gPAP was affinity chromatography with Glutathione Sepharose 4B™. This is a Sepharose matrix that contains covalently linked reduced glutathione (GSH), which is the substrate of glutathione S-transferase. Bound proteins can be eluted with glutathione, or can be cleaved-off from the GST-tag and the untagged protein eluted from the matrix.

Initially, I tried to elute the whole fusion protein, but no matter what GSH concentration was used most of gPAP remained bound to the Sepharose beads. The same result was obtained when gPAP was cleaved from the tag with PreScission protease ™ (GE Healthcare). This suggested that gPAP was bound unspecifically to the resin. To overcome the unspecific binding of gPAP, I tested different washing and elution conditions with varying salt and non-ionic detergent concentrations. This resulted in around 10% eluted gPAP (Figure 11, lanes 4-9). Furthermore, I pre-incubated the Sepharose beads with bovine serum albumin (BSA) to occupy the unspecific binding sites on the matrix. This wide-spread technique (used for instance in ELISA) did not improve the elution quality, and gPAP was observed primarily bound to the Sepharose, but not in the eluate (Figure 13). Even after regeneration of the matrix, a clear protein band at 103 kDa was observed. This was the size of the fusion gPAP (Figure 13, lane 2). This led me to the conclusion that the Sepharose matrix was not appropriate for purification of gPAP.

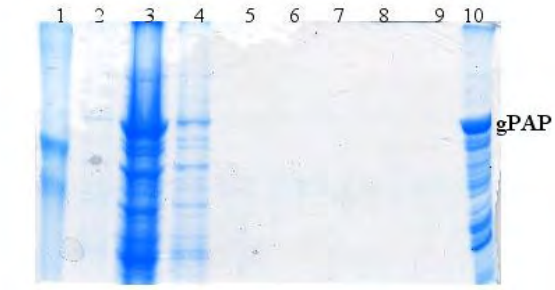


Figure 13: Giardial poly(A) polymerase binds unspecifically to glutathione-S-transferase affinity Sepharose beads blocked by BSA.
SDS-PAGE. Lane 1, BSA solution; lane 2, Sepharose matrix incubated with BSA; lane 3, flow-through; lanes 4-6, wash with 300 mM, 100 mM and without NaCl; lanes 7-9; elutions with 5 mM GSH; lane 10, proteins remaining on the matrix.

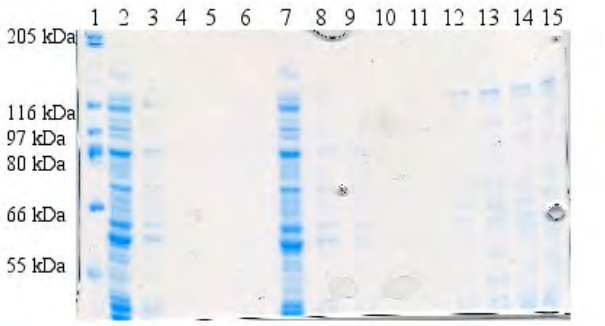


Figure 14: Purification of giardial poly(A) polymerase using magnetic glutathione-S-transferase affinity beads.
Lane 1, molecular marker. Lanes 2-6 - sample I 2.1 (the cells were enzymatically lysed, the lysis buffer contained PMSF and DTT, all procedures were performed at room temperature): lane 2, flow-through; lanes 3 - 5, wash; lane 6, elution. Lanes 7-11, sample I 2.2 (purification procedure included enzymatic cell lysis, lysis buffer contained PMSF and DTT and all procedures were performed at + 4°C): lane 7, flow-through; lanes 8-10, wash; lane 11, elution; lanes 12-15, proteins remaining on the beads after elution. See Table 3 for more details.

GST-tagged giardial poly(A) polymerase does not bind to glutathione-coated magnetic beads

In order to avoid the unspecific binding of gPAP to the matrix, I used a more inert matrix MagneGST™. Instead of modified agarose (Sepharose) the beads are made from a magnetic material, which allows a smaller bead size (10 μm) and thus increased binding capacity (10 mg fusion protein per 1 ml beads). In this assay I tested different lysis and elution conditions with varying temperature and buffer additives (see Materials and methods). Twelve different experiments gave the same result: GST-gPAP did not bind to this affinity matrix (Figure 14).

Polyadenylation activity could not be recovered with recombinant giardial poly(A) polymerase

To recover polyadenylation activity from the partially purified gPAP (Glutathione Sepharose 4B preparation), I used non-specific polyadenylation conditions. Previously those have been described by Wahle (Wahle, 1991) and modified in Anders Virtanen's lab by Christina Kyriakopoulou and Helena Nordvarg (Kyriakopoulou *et al.*, 2001).

The non-specific polyadenylation assay included in the reaction buffer an A_{15} substrate[2], [α-^{32}P]ATP, Mn^{2+} and different enzyme concentrations. For control reactions I used reaction mix without substrate. After the reaction was performed the RNA substrate was precipitated on CF/C filters with trichloracetic acid (TCA). The incorporation of radioactively labelled

2 RNA molecule consisting of 15 adenosine residues.

ATP was detected in a scintillator. However, no RNA polymerization activity was detected (data not shown).

To verify the results, the reaction was repeated. This time I used radioactively labeled A_{15} substrate and normal ATP. The reaction products were gel fractionated (Figure 15). The gel assay results confirmed those from the TCA precipitation. There were no heavier RNA substrates than the initial A_{15}. Thus, polyadenylation activity could not be recovered. One of the possible causes could have been incorrect folding of the protein, or incomplete purification of gPAP.

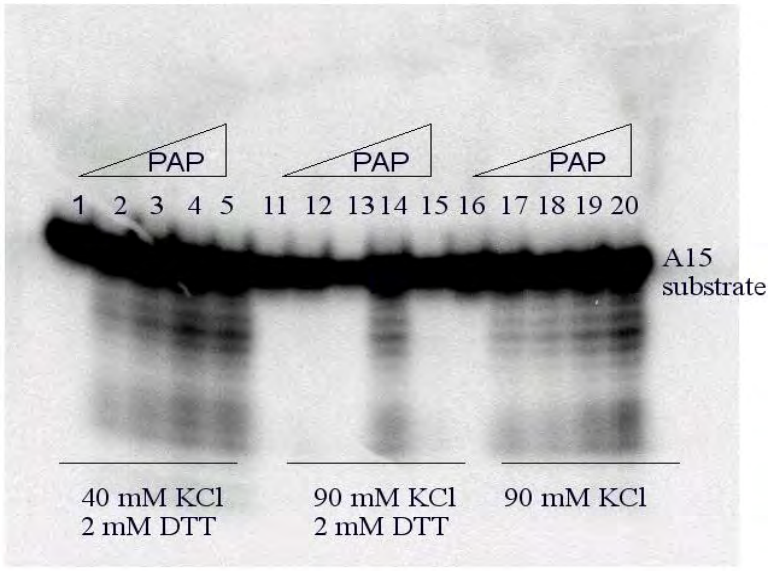


Figure 15: Products from activity assay for non-specific polyadenylation reaction, performed with giardial poly(A) polymerase, fractionated on 15 %RNA denaturating polyacrylamide gel. Lanes 1-5: Poly(A) reaction preformed in the presence of DTT, 40 mM KCl; lanes 11-15 - reaction performed in the presence of DTT, 90 mM KCl; lanes 16-20 - reaction performed without DTT, with 90 mM KCl. Lanes 1, 11 and 16 are control reaction containing only A_{15} substrate. (Note: The gel is over exposed thus, bands are much darker.) See Table 4 for details.

Discussion

The current work shed some light on the properties of the putative poly(A) polymerase from the microorganism *Giardia lamblia* (gPAP). The *in silico* analysis of gPAP showed that its domain architecture resembles that of a typical poly(A) polymerase (PAP). It has a PAP central domain and and N-terminal nucleotidyl transferase domain (Ntr). The PAP catalytic residues in the active site are conserved in gPAP. In contrast to bovine PAP, the giardial enzyme has a hydrophobic domain (fungal ATP synthase protein 8 domain) at its C-terminal site instead of an RNA binding domain (RBD). Furthermore, at the N-terminal of gPAP, a myrystyl motif was recognised as a part of a prokaryotic lipoprotein motif.

These findings lead to the assumption that gPAP possesses a poly(A) polymerisation capacity due to the conserved catalytic residues. Probably the catalytic mechanism is also divalent metal ion assisted as with other Ntrs. Since gPAP lacks an RBD, there is may be an assisting protein partner that facilitates the binding to the mRNA. If such a partner is established it may bind to the C-terminal hydrophobic domain of gPAP, as is the case of GLD-2 cytoplasmic poly(A) polymerase and its partner GLD-3 (Kwak *et al.*, 2004) . In addition, it seems that gPAP is N-terminally membrane-bound through the myrystyl motif.

One way to test the above hypothesis is to use viral MS2 coat protein, which binds to mRNA, instead of the assisting partner (Kwak *et al.*, 2004). Another possible reason for the absence of polyadenylation activity is the incorrect folding of gPAP. Since, the GST-tag did not bind to the MagneGST matrix, it must have been misfolded. Thus, gPAP might have been misfolded, too.

The poor purification of gPAP, may have been another result of the incorrect folding of the GST-tag. Thus, for future expression and purification procedures a smaller tag should be chosen, such a histidine- tag, or a thioredoxin-tag that would assist the proper folding (Kyriakopoulou *et al.*, 2001).

Materials and Methods

Strains and plasmids

Escerichia coli strains are listed in Table 2:

Table 2: *E. coli* strains:

Strains	Genotype	Source and references
BL21 (DE3) [1]	F-*ompT gal hsdSB* (rB^-mB^-) *gal dcm* (DE3)	Home-made $CaCl_2$ competent cells Studier, 1986
DH5α [2]	F^-, φ *80dlac*ZΔM15, Δ(*lac*ZYA-*arg*F)U169, *rec*A1, *end*A1, *hsd*R17(rk^-, mk^+), *pho*A, *sup*E44, *thi*-1, *gyr*A96, relA1λ^-	Home-made $CaCl_2$ competent cells Hanahad, D., 1983

1. Contains λDE3 lysogen that carries the gene for T7 RNA polymerase under the control of *lac*UV5 promotor. IPTG is required to induce expression of the T7 RNA polymerase. The strain is deficient in *lon* protease and *ompT* membrane protease, thus degradation of heterologous protein is reduced.
2. Host for blue/white screening utilizing the activity of β-galactosidase. This strain does not carry *lacl*^{q}, basically IPTG is not needed. The insert stability is ensured through mutation of *rec*A1, due to *end*A1 mutation the plasmid yield is greatly increased.

The poly(A) polymerase gene (Entrez GeneID 5697771) from *G. lamblia* cloned in pGEX-6P-3 overexpression vector was provided by Staffan Svärd.

Transformation

An aliqot of 50 µl competent cells was thawed on ice for 5 min and 1 µl plasmid (pGEX-6P-3-gPAP) was added. Cells with the plasmid were tap-mixed and incubated on ice for 2 min.This mixture was then heat-shocked for 40 sec at + 42 ºC and 2 min incubation on ice. An aliquot of 100 µl of SOC medium (Invitrogen©) was added and followed by incubation for 30 min at + 37 ºC. The transformed cells were plated on pre-warmed Luria agar (LA – Bactro-tryptone 10 g/L, yeast extract 5 g/L, NaCl 10 g/L, agar 15 g/L, pH 7.5) plates supplemented with the 50 µg/ml amp. After the cell mixture had diffused on the agar medium plates were inverted and incubated overnight at + 37 ºC.

Plasmid preparation

One colony from the transformed DH5α cells was used to inoculate 10 ml of Luria-Bertani medium (LB – 10 g/L Bactro-tryptone, 5 g/L yeast extract, 10 g/L NaCl, pH 7.5) supplemented with 100 µg/ml amp. The culture was grown for 6 hours at 37ºC with agitation and the plasmid was extracted using Plasmid Mini-prep kit™ (QIAgen©) as described in the kit manual.

Large scale expression of giardial poly(A) polymerase

One colony of transformed *E. coli* BL21 (DE3) cells was used to inoculate 20 ml LB containing 100 µg/ml amp. The culture was grown overnight at + 37 ºC without agitation, then harvested by centrifugation at 6000 g for 20 min. The pellet was resuspended in 5 ml fresh LB and used to inoculate 1 L LB containing 100 µg/ml amp in a 5 L flask. Expression was induced at $O.D._{600}$ 0.7 – 1.0 by addition of isopropyl β-D-1- thiogalactopyranoside (IPTG) to 0.1 mM final concentration for 2.5 hours. Finally, cells were harvested by centrifugation at 6000 g for 20 min at + 4 ºC, washed with 20 ml cold PBS (8 g/L NaCl, 0.2 g/L KCl, 1.44 g/L Na_2HPO_4, 0.24 g/L KH_2PO_4, pH 7.4) and stored at – 70 ºC.

Purification of GST-PAP using glutathione-Sepharose

The purification protocol was modified from GE Healthcare© instructions 52-2303-00 AG.

Slurry preparation: For 1 L culture of induced cells the bed volume (BV) was 0.5 ml. 0.7 ml 75 % glutathione-Sepharose was pipetted in a 15 ml falcon tube and centrifuged for 5 min at 500 g to remove the storage buffer. The slurry was washed three times with 5 ml 1x PBS and decanted at 500 g for 5 min. Finally, it was equilibrated with 0.5 ml cold lysis buffer. The resulting concentration of the slurry was 50 %. 1 ml of it was used for purification of gPAP from 1L .

Preparation of clear lysate and binding: Cell paste from 1 L induced cell culture (approximately 2.5 g) was thawed on ice and resuspended in 50 ml ST buffer (10 mM Tris-HCl pH 8.0 and 150 mM NaCl). Lysozyme to 100 µg/ml, dithiothreitol (DTT) to 5 mM and sodium lauryl sarcosinate (Sarkosyl) to 1.5 %final concentration were added and incubated for 15 min before the cells were sonicated. The sonication was performed with an MSE 100 Watt ultrasonic disintegrator (4 x 15 sec. with 1 min. pause between sonications, at amplitude 5). Cell debris was removed by centrifugation for 30 min at 43 000 g at + 4 ºC. The supernatant was transferred to a 50 ml falcon tube and supplemented with Triton X-100 to 3% (v/v) and phenylmethylsulphonyl fluoride (PMSF, serine protease inhibitor) to 0.1 mM final concentration. Finally, the affinity resin was added to the sample and incubated for 1 hour at + 4 ºC with rotation.

Elution and GST-tag cleavage: After binding, the slurry from the falcon tube was transferred into a BioRad© 10 ml chromatography column and washed consecutively with 5 ml 300 mM NaCl in 1x PBS, 5 ml 100 mM NaCl in 1x PBS and 5 ml 1x PBS. The protein was eluted with 3 x 0.5 ml elution buffer (50 mM Tris-HCl pH 8.0, 10 mM glutathione). All procedures were performed at + 4 ºC. To cleave off the GST-tag, the instructions provided by GE Healthcare© were applied (for product 07 0843-01 (PreScission protease™). Protein concentration was measured using BioRad© protein assay.

Refolding of giardial poly(A) polymerase from inclusion bodies

The pellet from the lysis step was resuspended in 20 ml 4 M urea at + 4 ºC and centrifuged at 10 000 g for 20 min. The resulting pellet was resuspended in 10 ml 8 M urea and centrifuged for 20 min at 10 000 g. The supernatant obtained from the 4 M urea was dialysed against 0.5 L dialysis buffer (1x PBS, 5 mM $MnCl_2$, 1% Triton X-100) at 4º C for 2 hours, and then against 1 L

dialysis buffer for 2 hours and overnight against 0.5 L dialysis buffer. The insoluble proteins were pelleted by centrifugation for 20 min at 10 000 g.

Purification of giardial poly(A) polymerase with glutathione-coupled magnetic beads (MagneGST™)

The protocol was designed for 10 ml induced cell culture according to Promega© instructions for use of products V8600, V8603, V8611 and V8612. For the assays 12 x 10 ml cultures were grown, as described in the "Large scale expression of giardial poly(A) polymerase" section. Those were used for 12 different purification conditions, which are shortly described in Table 3. The GST-affinity resin was prepared according to the manufacturer's instructions (0.5 ml resin). Six samples were enzymatically lysed in the presence of lysozyme and DNase I, the other six were mechanically disintegrated. The additives DTT, Triton X -100 and PMSF were added to the cell lysate right after cell disruption. Debris was removed by centrifugation in a benchtop microcentrifuge for 30 min at 5000 g. Temperature conditions (+ 4ºC and room temperature) were maintained throughout the whole purification procedure.

Table 3: Conditions used for the purifcation of giardial poly(A) polymerase with magnetic GST-affinity matrix.

		Samples											
		I						II					
Condition		1.1	1.2	2.1	2.2	3.1	3.2	1.1	1.2	2.1	2.2	3.1	3.2
Lysis conditions	Enzymatic[1] (lysozyme and DNase I)	+	+	+	+	+	+						
	Mechanic[2]							+	+	+	+	+	+
Lysis buffer components	1 mM PMSF	+	+	+	+	+	+	+	+	+	+	+	+
	5 mM DTT	+	+	+	+			+	+	+	+		
	0.5% Triton X-100	+	+					+	+				
Temperature[3]	+ 4ºC		+		+		+		+		+		+
	Room temperature	+		+		+		+		+		+	

1. Lysozyme 100 µg/ml, DNase I 100 µg/ml
2. Cell Disruptor ™, 2 Kba pressure
3. Temperature conditions refer for both lysis and purification steps.

5'-labelling and purification of RNA

For 20 µl reaction 0.5 µM 5'-dephosphorylated A_{15} RNA substrate (chemically synthesized from Sigma©), 10 µCi [γ-^{32}P]ATP, 2 µl 10x reaction buffer, 9 µl DNase and RNase free water and 1µl T4 polynucleotide kinase (10 U/µl) from Fermentas © were needed. RNA kinasing was performed for 30 min at + 37 ºC and stopped by addition of an equal volume RNA sample buffer (80% formamide, 0.1% v/v xylene cyanol, 0.1% v/v bromphenol blue, 50 mM Tris-HCl pH 8.0, 1 mM EDTA). Results were resolved 25 % polyacrylamide gel (See section "Polyacrylamide gel electrophoresis, staining and developing", Table 6) for 2h at 10W in 0.5x TBE buffer (5.4 g/L Tris base, 2.75 g/L boric acid, 0.46 g/L Na_4EDTA). Afterwards the gel was exposed to a photo-sensitive film and the labelled A_{15} substrate was cut out from the gel. RNA was eluted from the pieces in 300 µl water (DNase and RNase free supplemented with RNasin) by incubation on rotating platform overnight at + 4 ºC. The concentration of the product was measured by scintillation.

Non-specific polyadenylation reaction[3]

TCA precipitation

For 25 µl reaction reaction buffer contained 100 mM Tris-HCl pH 8.6, 10% (v/v)glycerol, 40 mM KCl[4], 0.1% (v/v) Nonident 40 (NP-40), 0.2 mg/ml methylated BSA, 0.5 mM DTT[5], 0.5 mM ATP mix (1.2 µCi of [α-^{32}P]ATP), 0.5 mM $MnCl_2$, 9 U Rnasin (Fermentas©), 10 µM A_{15}, with variation of the gPAP concentration from 0.168 µM to 1.34 µM (samples 1 -10 are described in Table 4).The reaction was carried out for 20 min at + 37º C. Upon completion each reaction was pipetted onto a GF/C filter. The filters were washed in a beaker with 5% tricloroacetic acid (TCA) and 1% tetrasodium pyrophosphate (NaPPi) for 5 min (5 ml per filter). This was followed by 3 x wash for 5 min with 5% TCA. Finally, the filters were washed with 99% ethanol for 3 min. All wash steps were performed on ice and the beaker was twirled for thorough washing. The filters were air-dried for 30 min and counted in a scintillator.

To calculate the percentage of incorporated [α-^{32}P]ATP, the following formula was used:

$$\frac{(cpm\ per\ sample)}{(total\ cpm)} x100$$

To calculate the concentration of incorporated [α-^{32}P]ATP:

$$\frac{((Initial\ concentration\ of [\alpha - 32P]\ ATP)\ x\ cpm\ per\ sample)}{(Intial\ concentration\ of [\alpha - 32P]\ ATP)}$$

3 The RNA substrate was not a mRNA but A_{15} (non-specific) and instead of Mg^{2+}, Mn^{2+} was used.

4 40 mM KCl was standard concentration for this reaction, in the gel fractionation assay 90 mM KCl was also used.

5 0.5 mM DTT was standard concentration for this reaction. Some reactions lacked DTT in order to test its effect on reaction efficiency.

Gel fractionation assay

Non-specific polyadenylation reaction was performed in the same way as the TCA precipitation assay, except that the A_{15} substrate was radioactively labelled and cold ATP was used (samples 11 – 20, Table 4). The reaction was stopped by addition of 20 µl RNA sample buffer (80% (v/v) formamide, 0.1% (v/v) xylenecyanol, 0.1% (v/v) bromphenol blue, 50 mM Tris- HCl pH 8.0, 1 mM EDTA).Then the samples were loaded on 15% RNA denaturation gel (see section "Polyacrylamide gel electrophoresis, staining and developing", Table 5) for 1.5 hours at 10 W in 0.5x TBE. Protective screen was used when handling the radioactive samples.

Table 4: Non-specific polyadenylation sample description.

	Sample																			
Component	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
gPAP [µM]	0	0.17	0.34	0.68	1.34	0	0.17	0.34	0.68	1.34	0	0.17	0.34	0.68	1.34	0	0.17	0.34	0.68	1.34
0.5 mM [α-^{32}P]ATP	+	+	+	+	+	+	+	+	+	+										
10 µM cold A_{15}						+	+	+	+	+										
0.5 mM cold ATP											+	+	+	+	+	+	+	+	+	+
10 µM 5'*A_{15} radioactive											+	+	+	+	+	+	+	+	+	+
40 mM KCl	+	+	+	+	+	+	+	+	+	+										
90 mM KCl											+	+	+	+	+	+	+	+	+	+
0.5 mM DTT	+	+	+	+	+	+	+	+	+	+										

Polyacrylamide gel electrophoresis, staining and developing

RNA PAGE

RNA gels were cast with 0.4 mm thick spacers. Before loading the samples the gel was pre-run for 30 min in 0.5x TBE at 10W. The electrophoresis was performed in 0.5x TBE at 10W for approximately 1.5 hours. Description of the 15 % polyacrylamide denaturating gel used in the giardial PAP activity assay and 25 % polyacrylamide gel used for RNA A_{15} substrate purification contents is presented in Tables 5 and 6.

Table 5: 15 % RNA denaturating polyacrylamide gel

Component	Amount
40 % acrylamide (29:1)	9.38 ml
5x TBE	5 ml
water	20.63 ml
Urea	10.5 g
10 % APS	0.25 ml
TEMED	0.03 ml
Final volume	25 ml

* crosslinker

Table 6: 25 % RNA native polyacrylamide gel

Component	Volume [ml]
40 % acrylamide (29:1)	15.6
5x TBE	5
water	4.4
10% APS	0.25
TEMED *	0.03
Final volume	25

* crosslinker

Sodium dodecyl sulfate polyacrylamide gel elctrophoresis

Each 7.5 % sodium dodecyl sulfate polyacrylamide gel consisted of stacking and separating gel described in Tables 7 and 8. The electrophoresis was performed in 1x SDS running buffer (25 mM Tris base pH 8.8, 250 mM glycine, 0.1% (w/v) SDS) at 130 V for approximately 80 min. Samples were prepared with 5x sample buffer (250 mM Tris-HCl pH 6.8, 10% SDS, 30% glycerol, 5% β-mercaptoethanol, 0.2% bromphenol blue).

Table 7: 7.5 % separating sodium dodecyl sulfate gel

Component	Volume [ml]
1.5 M Tris-HCl pH 8.8	1.25
30% acrylamide (37:5:1)	1.25
87% glycerol	0.33
water	2.15
20% SDS	0.05
10% APS	0.1
TEMED*	0.01
Final volume	5.14

*crosslinker added last

Table 8: 4 % stacking sodium dodecyl sulfate gel

Component	Volume [ml]
0.5 M Tris-HCl pH 6.8	0.25
30% acrylamide (29:1)	0.13
water	0.6
20% SDS	0.01
10% APS	0.05
TEMED*	0.01
Final volume	1.05

*crosslinker added last

Coomassie staining

The gel was placed in 100 ml Coomassie staining solution (3 parts 37% acetic acid (CH_3COOH), 7 parts isopropanol, 1.2 g/L Coomassie Brilliant Blue G250) and heated up in a microwave for approximately 1 min at 850 W and incubated with the stain for 20 min on a shaker. Destaining was performed by washing the gel for 5 min with dH_2O and 2x for 1 h with 100 ml destaining solution (5 parts 37% acetic acid (CH_3COOH), 12 parts isopropanol, 33 parts dH_2O, 50 parts dH_2O).

Silver staining

The gel was fixed for 10 min in 50 % methanol and 10 % acetic acid, then for10 min in 5 % methanol and 7% acetic acid and washed with dH2O for 10 min. Proteins were additionally fixed with 10% glutaraldehyde for 15 min, followed by 1 min wash with dH2O and 2x 10 min washes with dH2O. Fresh silver solution (1.9 ml 1M NaOH, 1.4 ml NH3, 21 ml dH2O, 4 ml 10% AgNO3 was added drop-wise with stirring without the silver precipitating. The solution was dilute to 100 ml with dH2O) was applied to the gel for 10 min. Then the gel was washed for 10 min with dH2O. Finally, 100 ml freshly prepared developing solution (100 µl 50 mg/ml citric acid, 50 µl formaldehyde, dH2O was added to 100 ml) was added. Developing was stopped by addition of 37% acetic acid. Gel was preserved in drying buffer 2 x 10 min (5 % methanol and 4 % glycerol).

Bradford assay

The Bradford reagent was prepared by diluting 1 part BioRad protein assay™ reagent with 3 parts ddH_2O. To 1 ml of reagent 20 µl protein sample was added. Absorbance was measured at 600 nm. For calculating the protein concentration bovine γ-globulin was used as standard.

Computational analysis

For the *in silico* analysis, the amino acid sequences were submitted to the search engine of

the data base and the results were acquired either online or via email. The different data bases are listed below, with a short introduction of their usage purpose and link to them.

ExPASY ProtParam is data base that allows computational prediction of different physical and chemical properties of a given protein - http://expasy.org/tools/protparam.html

Jpred: PDB Homologues is a server that predicts protein secondary structure. - http://www.compbio.dundee.ac.uk/~www-jpred/

EsyPred3D is an automated homology modelling program - http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/

Pfam is database which contains different protein families, each represented with multiple sequence alignments - http://pfam.sanger.ac.uk/

CATH database contains protein structure classification http://www.cathdb.info/latest/index.html

PyMol software (DeLano W L , 2007) was used to visualize structures from the protein data bank (PDB - http://www.rcsb.org/pdb/home/home.do).

The gene and protein sequence were obtained from NCBI Entrez Gene database: http://www.ncbi.nlm.nih.gov/sites/entrez, Gene ID 5697771

Acknowledgements

I would like to thank Anders Virtanen for the opportunity to work in his lab and have my own glimpse in the world of science. To Assad who welcomed me and was there for me the whole time. I will never forget Lei, Niklas, Per and Pontus, who introduced me to the lab and were helping when I was stumbling in bad results or computer problems.

I would also like acknowledge the synapse group – Ben, Christoph, Enzo,Greta and Nicole for the the 'fyka' and the discussions. My friend Konstantin, who listened when I was thinking aloud. And not to forget Vasili H. and Adrian S. who shared their professional experience with me.

I am very thankful to Karin Carlson for her advises when I was starting my project work and for her support throughout it.

Last by not least, it is my family and Miro which were always there when I was feeling down.

References

Adam, R.D., (1991). *The biology of Giardia spp*. Microbiol Rev. 55, 706-32.

Aravind, L. & Koonin, E.V., (1999). *DNA polymerase beta-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history*. Nucleic Acids Res. 27, 1609-18.

Balbo, P.B., Meinke, G. & Bohm, A., (2005). *Kinetic studies of yeast polyA polymerase indicate an induced fit mechanism for nucleotide specificity*. Biochemistry. 44, 7777-86.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., (2002) *The Protein Data Bank*. Nucleic Acids Research, 28 pp. 235-242.

Brown, J.W. & Reeve, J.N., (1986). *Polyadenylated RNA isolated from the archaebacterium Halobacterium halobium*. J Bacteriol. 166, 686-8.

Chekanova, J.A. & Belostotsky, D.A., (2003). *Evidence that poly(A) binding protein has an evolutionarily conserved function in facilitating mRNA biogenesis and export*. RNA. 9, 1476-90.

Cuff J. A., Barton G. J., (2000). *Application of multiple sequence alignment profiles to improve protein secondary structure prediction.* Proteins 2000;40:502-511.

DeLano, W.L., (2007). *The PyMOL Molecular Graphics System,* DeLano Scientific LLC, Palo Alto, CA, USA. http://www.pymol.org

Dzikiewicz, A. & Szweykowska-Kulińska, Z., (2006). *Nonsense-mediated mRNA decay (NMD)--on guard of mRNA quality*. Postepy Biochem. 52, 390-8.

Edmonds, M. & Abrams, R., (1960). *Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei*. J Biol Chem. 235, 1142-9.

Finn R.D., Jaina Mistry , Schuster-Böckler B. , Griffiths-Jones S., Hollich V., Lassmann T., Moxon S., Marshall M., Khanna A., Durbin R., Eddy S.R., Sonnhammer, E. L. L., and Bateman A., (2006). *Pfam: clans, web tools and services.* Nucl. Acids Res. 34: D247-D251.

Gallie, D.R., (1991). *The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency*. Genes Dev. 5, 2108-16.

Gallie, D.R., (1998). *A tale of two termini: a functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation*. Gene. 216, 1-11.

Jackson, R.J. & Standart, N., (1990). *Do the poly(A) tail and 3' untranslated region control mRNA translation?*. Cell. 62, 15-24.

Karlin, S. & Altschul, S.F., (1990) M*ethods for assessing the statistical significance of molecular sequence features by using general scoring schemes.* Proc. Natl. Acad. Sci. USA 87:2264-2268

Kyte, J. and Doolittle, R.F., (1982) *A simple method for displaying the hydropathic character of a protein.* J. Mol. Biol. 157, 105-132.

Kaufmann, I., Martin, G., Friedlein, A., Langen, H. & Keller, W., (2004). *Human Fip1 is*

a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J. 23, 616-26.

Kwak, J.E., Wang, L., Ballantyne, S., Kimble, J. & Wickens, M., (2004). *Mammalian GLD-2 homologs are poly(A) polymerases*. Proc Natl Acad Sci U S A. 101, 4407-12.

Kyriakopoulou, C.B., Nordvarg, H. & Virtanen, A., (2001). *A novel nuclear human poly(A) polymerase (PAP), PAP gamma*. J Biol Chem. 276, 33504-11.

Lambert C, Leonard N, De Bolle X, Depiereux E., (2002). *ESyPred3D: Prediction of proteins 3D structures.* Bioinformatics. 18(9):1250-1256

Macreadie, I.G., Novitski, C.E., Maxwell, R.J., John, U., Ooi, B.G., McMullen, G.L., (1983). *Biogenesis of mitochondria: the mitochondrial gene (aap1) coding for mitochondrial ATPase subunit 8 in Saccharomyces cerevisiae*. Nucleic Acids Res. 11, 4435-51.

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L., (2006). *Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease*. Nature. 444, 953-6.

Martin, G. & Keller, W., (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, 2593-603.

Martin, G. & Keller, W., (2007). *RNA-specific ribonucleotidyl transferases*. RNA. 13, 1834-49.

Martin, G., Keller, W. & Doublié, S., (2000). *Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP*. EMBO J. 19, 4193-203.

Martin, G., Möglich, A., Keller, W. & Doublié, S., (2004). *Biochemical and structural insights into substrate binding and catalytic mechanism of mammalian poly(A) polymerase*. J Mol Biol. 341, 911-25.

Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., (2007). *Genomic minimalism in the early diverging intestinal parasite Giardia lamblia*. Science. 317, 1921-6.

Murthy, K.G. & Manley, J.L., (1995). *The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation*. Genes Dev. 9, 2672-83.

O'Hara, E.B., Chekanova, J.A., Ingle, C.A., Kushner, Z.R., Peters, E. & Kushner, S.R., (1995). *Polyadenylylation helps regulate mRNA decay in Escherichia coli*. Proc Natl Acad Sci U S A. 92, 1807-11.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M., (1997). *CATH- A Hierarchic Classification of Protein Domain Structures.* Structure. Vol 5. No 8. p.1093-1108.

Sheets, M.D. & Wickens, M., (1989). *Two phases in the addition of a poly(A) tail*. Genes Dev. 3, 1401-12.

Studier FW, Moffatt BA., (1986) *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes*. J. Mol. Biol.189, 113-130.

Varshavsky, A., (1997) *The N-end rule pathway of protein degradation.* Genes Cells 2, 13-28.

Wahle, E., (1991). *Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors*. J Biol Chem. 266, 3131-9.

Wahle, E. & Keller, W., (1992). *The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors*. Annu Rev Biochem. 61, 419-40.

Wahle, E. & Keller, W., (1996). *The biochemistry of polyadenylation*. Trends Biochem Sci. 21, 247-50.

Yue, D., Maizels, N. & Weiner, A.M., (1996). *CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile Sulfolobus shibatae*. RNA. 2, 895-908.

Zarudnaya, M.I., Kolomiets, I.M. & Hovorun, D.M., (2002). *What nuclease cleaves pre-mRNA in the process of polyadenylation?*. IUBMB Life. 54, 27-31.