Analysis of the mammalian pre-mRNA 3' end processing cleavage complex and attempts of its reconstitution

Dissertation

zur Erlangung des akademischen Grades

Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Naturwissenschaftlichen Fakultät I
Biowissenschaften

der Martin-Luther-Universität Halle-Wittenberg

von Frau Anne Ulrika Knoth

geb. am 04.03.1979 in Halle (Saale)

Gutachter:

1. Prof. Dr. E. Wahle
2. Prof. Dr. W. Keller
3. Prof. Dr. S.-E. Behrens

Halle (Saale), den 09.06.2008

urn:nbn:de:gbv:3-000013658
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000013658]

Success is the ability to go from one failure to another
with no loss of enthusiasm.

(Winston Churchill)

Table of contents

Abbreviations

APS	Ammonium peroxodisulphate
ATP	Adenosine triphosphate
BoxB	RNA element, forming a specific hairpin loop, recognition element of the antiterminator protein N from phage λ
bp	Base pair
BSA	Bovine serum albumin
CASP	Metallo-β-lactamase (CPSF, Artemis, Snm1, Pso2)
CBC	Cap binding complex
CF I_m	Mammalian Cleavage factor I
CF II_m	Mammalian Cleavage factor II
Ci	Curie (1 Ci = 3.7 * 10^7 Becquerel)
CID	CTD interaction domain
CIP	Calf intestinal alkaline phosphatase
CP	Cleavage product
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
CTD	C-terminal domain of RNA polymerase II
C-terminal	Carboxy-terminal
DEAE	Diethylaminoethylcellulose, Cation exchanger
DEPC	Diethyl-pyrocarbonate
DMSO	Dimethylsulfoxide
DNA	Desoxyribonucleic acid
DSE	Downstream sequence element
DTT	1,4-Dithiothreitol
E. coli	*Escherischia coli*
EDTA	Ethylendiamintetraacetate sodium salt
EMSA	Electrophoretic mobility shift assay
FCS	Fetal calf serum
GAR	Glycine-arginine rich motif
GFP	Green fluorescence protein
GST	Glutathione S-transferase
GST-λN	Fusion protein consisting of GST and peptide N from phage λ
HAT	Half a TPR

HEK293	Human embryonic kidney 293 cells
HeLa	Human cell line obtained from cervical cancer cells
HEPES	4-(2-Hydroxyethyl)-1-piperacinethansulfon acid
HF	his_8-flag tag
hFip1	Human Fip1 protein
his_8	*Octa*-histidine tag
hnRNP	Hetero nuclear ribonucleoprotein
HRP	Horseradish peroxidase
IgG	Immune globulin G
IPP	Immune precipitation buffer
IPTG	Isopropyl-β-D-thiogalactopyranoside
kbp	Kilo base pair
K_D	Dissociation constant
kDa	Kilo Dalton
L3	Adenovirus 2, transcript from major late promoter
L3Δ	L3 with point mutation in AAUAAA to AAGAAA sequence
L3-α T	L3 RNA containing the aptamer for the binding to tobramycin at the 3' end of the RNA
LB	Luria-Bertani
$m^7G^{5'}ppp^{5'}Np$	Cap structure
MAPKKK	Mitogen-activated protein kinase kinase kinase
MEP50	Methylosome protein 50
MPF	M-phase factor (kinase)
mRNA	Messenger RNA
mRNP	mRNA·protein complex
MS	Mass spectrometry
NCS	Newborn calf serum
NHS	N-Hydroxysuccinimide
Ni-NTA	Nickel-nitrilotriacetic acid
NUDIX	Nucleoside diphosphate linked to some moiety X
nt	Nucleotide
N-terminal	Amino-terminal
NXT	Nuclear extract
NXT-A	Nuclear extract prepared in Halle (Wahle & Keller, 1994)

NXT-B	Nuclear extract prepared in Basel (Methods 5.10.2)
$OD_{600\ nm}$	Optic density measured at a wave length of 600 nm
PABPN1	Nuclear poly(A) binding protein 1
PAP	Poly(A) polymerase
PAPolA	Poly(A) polymerase A
PBS	Phosphate buffered saline
PBST	Phosphate buffered saline with Tween 20
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PhD	*Philosophiæ doctor*
PK	Protein kinase
PMSF	Phenylmethylsulphonylfluoride
poly(A)	Polyadenylate residues
pre-mRNA	Premature mRNA
PRMT5	Protein arginine methyltransferase 5
PVA	Polyvinyl alcohol
RBD	RNA binding domain
RNA	Ribonucleic acid
RNAP II	RNA polymerase II
RNase	Ribonuclease
RNP	Ribonucleoprotein particle
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RS-like	Like arginine-serine rich domain
RRM	RNA recognition motif
RT	Room temperature
s	Short
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS polyacrylamide gel electrophoresis
siRNA	Silencer RNA
SN	Supernatant
snRNA	Small nuclear RNA
snRNP	Small nuclear RNA protein complex

SR	Serine-arginine rich
SV40 late	SV40 virus transcript from late promoter
SV40 late Δ	SV40 late with point mutation in AAUAAA to AAGAAA sequence
TAP	Protein-tag consisting of two protein A domains and a calmodulin binding domain
TBE	Tris-borate-EDTA
TBS	Tris buffered saline
TE	Tris-EDTA
TEMED	N,N,N',N'-Tetramethylethylendiamine
TEV	Tobacco etch virus
TFIID	Transcription factor II D
TREX	Transcription export complex
TPR	Tetratricopeptide repeat domain
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
[U]	Units
UG-rich	RNA containing a Uracil-Guanine rich sequence
U-rich	RNA containing a Uracil rich sequence
USE	Upstream sequence element
UTP	Uracil triphosphate
UTR	Untranslated RNA
UV	Ultra violet
v/v	Volume part
w/v	Weight part
yRNA	Total yeast RNA
α T L3	L3 RNA containing the aptamer for the binding to tobramycin at the 5' end of the RNA
λN peptide affinity	Method developed by Czaplinski *et al*. (2005) using the GST-λN fusion protein and RNA containing the BoxB element
5' cap	$m^7G^{5'}ppp^{5'}Np$
3' dATP	3' Desoxy-ATP, cordycepin-triphosphate

1. Introduction

Organisms are made up of the sum of their genes and environmental influences. Early estimations of the number of human genes suggested 50.000 to 100.000 (Fields *et al*., 1994), just 45.000 (Green, 1999) or up to 140.000 genes in the human genome (Scott, 1999, Liang *et al*., 2000). As a result of the human genome sequencing project the estimated gene number dropped to 26.000 - 38.000 genes (Venter *et al*., 2001). Further analysis of databases that track protein coding genes revealed a number of around 20.500 genes (Pennisi, 2007).

One explanation for these dramatically decreased numbers, lies in the fact that transcripts of about 74 % of the human multi-exon genes are alternatively spliced (Johnson *et al*., 2003), alternatively processed or modified through RNA editing. Furthermore, some human proteins are involved in various complexes, catalysing different reactions.

Messenger RNA (mRNA) has been analysed for more than 40 years. The development of new methods to purify RNA and ribonucleoprotein particles (RNPs) was necessary to allow a better understanding of the mechanisms involved in RNA processing. This work aims to shed light on a small area of the maturation of mRNA in mammals, the 3' end processing reaction.

1.1 mRNA maturation

Transcription by RNA polymerase II (RNAP II) results in precursor mRNAs (pre-mRNA), which have to undergo several maturation events, and small nuclear RNAs. The carboxy-terminal domain (CTD) of the RNAP II acts as a binding platform for proteins. These proteins catalyse several steps of maturation of mRNA.

All mRNAs carry a cap structure at their 5' end. The three reactions leading to the cap structure (Shuman, 2001) occur before the transcript reaches a size of 20-50 nucleotides (nt) (Jove and Manley, 1984; Rasmussen and Lis, 1993). The RNAP II pauses to check the completion of the reactions (Mandal *et al*., 2004; Kim *et al*., 2004; Aguilera, 2005). The cap structure, bound by the cap binding complex (CBC), has an influence on transcript stability (Furuichi *et al*., 1977), on translation initiation (Muthukrishnan *et al*., 1975; Both *et al*., 1975), it is involved in splicing (Konarska *et al*., 1984; Lewis *et al*., 1996), in 3' end formation (Flaherty *et al*., 1997) and mRNA export (Izaurralde *et al*., 1992). The most important function is maybe, that the cap in combination with the poly(A) tail, marks the mRNA as fully intact and completely processed.

During maturation the pre-mRNA transcripts have to be spliced for the removal of introns, internal non-coding sequences. The phosphorylation of the CTD of RNAP II allows the

assembly of the spliceosome during transcription, which in turn enables the co-transcriptional splicing of many introns (Wetterberg *et al*., 1996).

In 1978 Ford and Hsu showed that mRNA maturation of the transcript of the simian virus 40 late promoter (SV40 late) involves 3' end cleavage of the primary transcript. This is the first step in 3' end formation after RNAP II has passed all possible poly(A) sites during transcription (Nevins and Darnell, 1978). A model is, that the recruitment of 3' end processing factors occurs at the promoter site, throughout the length of the gene and at the 3' end. CPSF (cleavage and polyadenylation specificity factor) is probably recruited to the RNAP II by the transcription factor II D (TFIID) when the transcription has started. Thus it may be associated with RNAP II during elongation (Dantonel *et al*., 1997). The cleavage reaction takes place after the assembly of the cleavage factors CF I_m and CF II_m, the cleavage stimulation factor (CstF) and the poly(A) polymerase (PAP) at their respective sequence elements. The 5' pre-mRNA fragment is then further polyadenylated, while the 3' end fragment is degraded. The cleavage reaction and its RNA sequence elements are further described in the following Chapter (1.2 and 1.3).

Following the cleavage the PAP starts to synthesise the poly(A) tail. After the polymerisation of 10 to 12 adenylate residues to the 3'-hydroxyl group of the 5' cleavage fragment, the nuclear poly(A) binding protein 1 (PABPN1) joins the complex and enhances further polyadenylation by PAP, through a direct interaction. In mammals, the poly(A) tail length is limited to a size of around 250 nt (Wahle, 1995). Until now histone mRNAs are the only known exception which lack poly(A) tails (Adesnik and Darnell, 1972; Greenberg and Perry, 1972).

A few mRNAs are modified post-transcriptionally by base conversions from adenine to inosine and cytosine to uracil (Wedekind *et al*., 2003).

Several RNA export factors are recruited directly to the RNA during splicing (Custódio *et al*., 2004). The mature mRNA forms a mRNA·protein complex (mRNP). The complete transcription and export complex (TREX) along with the THO complex plays a major role in the transport of mRNPs to the cytoplasm (Reed and Hurt, 2002; Reed and Cheng, 2005). The mRNP is transferred to the nuclear envelope and translocated through the nuclear pore complex into the cytoplasm for translation (Reed, 2003). The maturation steps of the pre-mRNA to the mRNA are schematically shown in Figure 1-1.

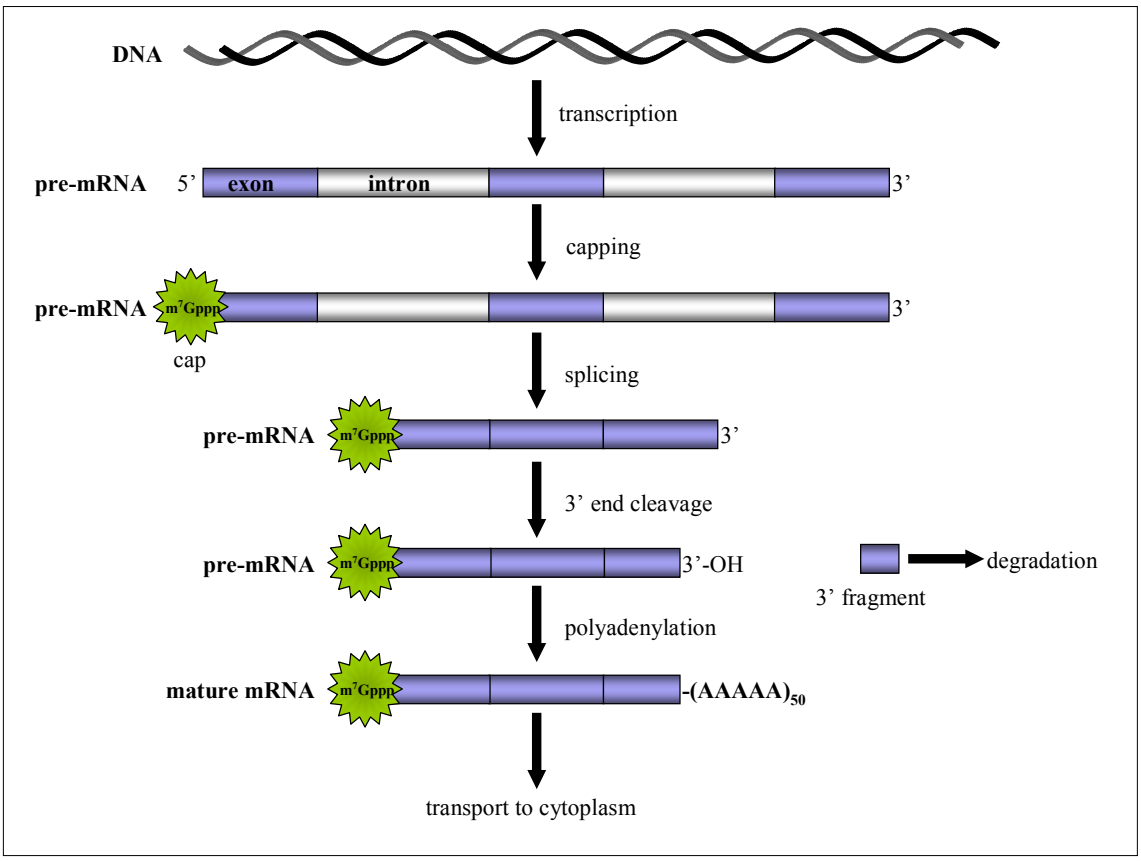


Figure 1-1 Schematic overview of mRNA maturation in the nucleus

1.2 Sequence elements in mRNA 3' end processing

In chemical terms, the 3' end formation is a simple reaction. A phosphodiester bond is hydrolysed in the pre-mRNA, and afterwards ATP is polymerised to the newly generated 3'-hydroxyl group. From a biochemical point of view, this reaction is much more complicated and lots of proteins are necessary. During 3' end processing the cleavage complex assembles onto the pre-mRNA and catalyses the endonucleolytic cleavage. Different sequence elements are necessary for the correct assembly of the proteins.

The highly conserved hexanucleotide AAUAAA was the first sequence element discovered in pre-mRNAs (Proudfoot and Brownlee, 1976). It is located 11-30 nt upstream of the poly(A) site (Proudfoot and Brownlee, 1976; Hagenbüchle *et al.*, 1979). The AAUAAA sequence is necessary for the binding of the tetrameric CPSF (Gilmartin and Nevins, 1989; Bardwell *et al.*, 1991; Keller *et al.*, 1991). CPSF-160K and -30K bind directly to the RNA (Murthy and Manley, 1995; Zhao *et al.*, 1999). CPSF is required for the cleavage and the polyadenylation reaction (Christofori and Keller, 1988; Gilmartin and Nevins, 1989; Takagaki *et al.*, 1989). Mutations in AAUAAA lead to a strong reduction or even to complete abolishment of cleavage (Fitzgerald and Shenk, 1981; Montell *et al.*, 1983; Higgs *et al.*, 1983; Gil and

Proudfoot, 1984; Wickens and Stephenson, 1984; Skolnik-David *et al*., 1987). This is due to reduced binding of CPSF. The hexanucleotid is highly conserved, but there are also other variants known, which are functional to a lower extent. The most common variant is AUUAAA (Chen and Shyu, 1995).

Another sequence element is the downstream element (DSE). It is weakly conserved and contains a short U-rich sequence and / or a GU-rich motif (Gil and Proudfoot, 1984; Hart *et al*., 1985a; McLauchlan *et al*., 1985; Conway and Wickens, 1985; McDevitt *et al*., 1986; Zarkower and Wickens, 1988). Salisbury and colleagues (2006) described that the DSE element consists of two parts. The UG-rich element is proximal, 5 to 10 nt, and the U-rich element is distal, 15-25 nt downstream of the cleavage position. Mutations in DSE cause a decrease in 3' end processing efficiency, but do not abolish the reaction (McDevitt *et al*., 1986). The DSE is bound by the cleavage stimulation factor via its 64K subunit (Weiss *et al*., 1991; MacDonald *et al*., 1994). The sequences around the AAUAAA and the DSE are not conserved, but the distance between them has effects on poly(A) site choice and efficiency in cleavage (Mason *et al*., 1986; McDevitt *et al*., 1986; Gil and Proudfoot, 1987; Chen *et al*., 2005). Cleavage occurs mostly after a CA dinucleotide (Fitzgerald and Shenk, 1981).

A binding site for the mammalian cleavage factor I (CF I_m) was recently discovered upstream of the poly(A) signal AAUAAA (Brown and Gilmartin, 2003). The UGUAN motif is bound by CF I_m and is present in different numbers in mRNAs. The L3 RNA, which contains the natural adenovirus 2 poly(A) site number 3, has two of these motifs. The mRNA of the poly(A) polymerase A gene contains four UGUAN motifs (Venkataraman *et al*., 2005). It was supposed that these sequence elements influence the poly(A) site selection (Venkataraman *et al*., 2005).

In the 3' untranslated region (UTR), another sequence element was found 13 - 48 nt upstream of the canonical poly(A) signal (Carswell and Alwine, 1989; DeZazzo and Imperiale, 1989) and was designated as upstream sequence element (USE). USEs are not essential for 3' end formation but play a role in poly(A) site choice (DeZazzo and Imperiale, 1989). Adenovirus 2 L1 and L3 transcripts contain a UUCUUUUU sequence (Prescott and Falck-Pederson, 1994) while SV40 late mRNA comprises three core USE elements with the consensus sequence AUUUGURA. They act in a distance-dependent manner from the AAUAAA signal and enhance the efficiency of 3' end processing additively (Schek *et al.*, 1992). Other USEs are not sequence homologues but act in the same manner and can be replaced by each other (Valsamakis *et al.*, 1991). A capped and spliced precursor mRNA with its sequence elements is schematically shown in Figure 1-2.

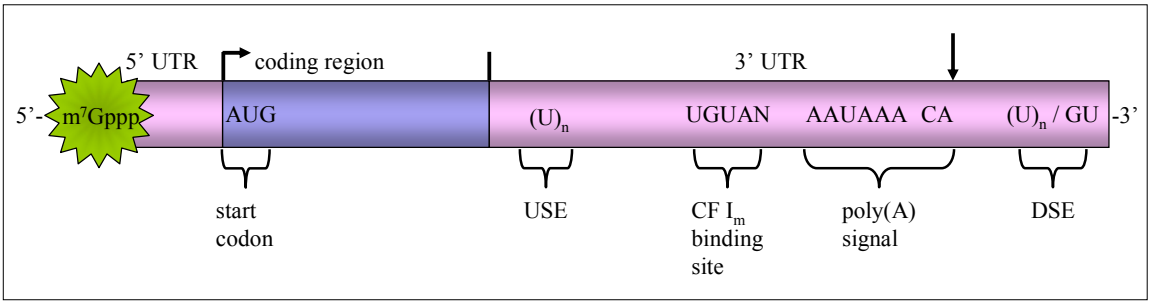


Figure 1-2 Sequence elements in a pre-mRNA
USE = upstream element, UTR = untranslated region, DSE = downstream element, arrow indicates cleavage site.

1.3 Proteins involved in mRNA 3' end processing

The proteins involved in the cleavage reaction are schematically shown in Figure 1-3. The well known and characterised factors CstF, CF I_m and the PAP assemble together with the less characterised factors CPSF and CF II_m onto the pre-mRNA, forming a complex active in cleavage.

During the polyadenylation reaction PABPN1 joins the complex after addition of about ten adenylate residues. It binds to the growing poly(A) tail and causes the length control of poly(A) tail of the mRNA. Most proteins involved in 3' end processing are significantly conserved from yeast to humans, which indicates the importance of the 3' end processing reaction.

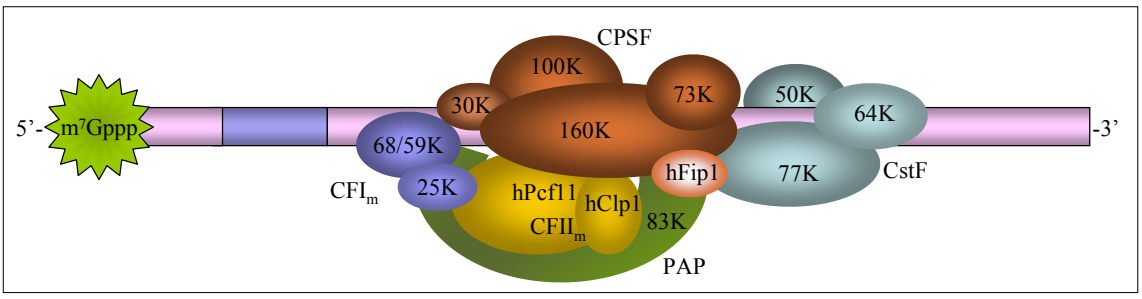


Figure 1-3 Overview about 3' end processing proteins and their binding positions onto the pre-mRNA
CPSF (Cleavage and polyadenylation specificity factor) (in brown) bound to AAUAAA, CstF (Cleavage stimulation factor) (in light blue) bound to downstream element, CF I_m (mammalian cleavage factor I) (dark blue) bound to UGUAN motif, CF II_m (mammalian Cleavage factor II) (khaki) and PAP (poly(A) polymerase) (green) bridges CPSF and CF I_m. RNA: cap structure in green, 3' and 5' UTR in violet, translated region in blue.

1.3.1 Cleavage stimulation factor, CstF

CstF binds to the GU / U rich downstream element (Beyer *et al*., 1997; Takagaki and Manley, 1997). The factor contains two polypeptides of 50K and 64K (Gilmartin and Nevins, 1991; Takagaki *et al*., 1990) which are bridged by a third subunit, CstF-77K (Takagaki and Manley, 1994). This factor is well characterised and can be reconstituted from purified subunits (Dettwiler, thesis, 2003).

CstF-50K contains seven WD-40 (β-transducin) repeats, which are implicated in binding to the phosphorylated C-terminal domain of RNAP II and to the BRCA1-associated protein BARD1 (Takagaki and Manley, 1992; McCracken *et al*., 1997; Kleiman and Manley, 1999; Fong and Bentley, 2001). Therefore, it has been suggested that CstF-50K plays an important role in linking the 3' end processing reaction to transcription (McCracken *et al*., 1997). CstF-50K dimerises and binds to CstF-77K (Takagaki and Manley, 1992). *In vitro* experiments revealed that CstF-50K is necessary for CstF activity (Takagaki and Manley, 1994).

CstF-64K binds to the downstream element via its N-terminal RNP type RNA recognition motif (RRM). Its C-terminal domain contains a long proline / glycine-rich region, which encloses 12 tandem copies of the MEARA / G amino acid (aa) motif. They form a long α-helical structure (Takagaki *et al*., 1992). In 2005 Deka and colleagues could show that during RNA binding the helix is unfolded. As the DSE is only weakly conserved, they suggested that an increased flexibility of the protein chain is necessary to bind multiple related RNA sequences. The C-terminal and N-terminal domains are connected by a so-called hinge region. CstF-64K interacts with several proteins. In 2000 Takagaki and Manley demonstrated its interaction with symplekin, a protein supposed to be involved in mammalian 3' end processing (Hofmann *et al*., 2002). Additionally, CstF-64K binds to hClp1, another protein of the pre-mRNA cleavage complex (de Vries *et al*., 2000). Paushkin and colleagues (2004) could show that CstF-64K was co-purified with Sen2 and Sen34 proteins, which are subunits of the human tRNA processing complex. Earlier publications indicated interactions with the transcriptional co-activator PC4 (positive factor 4) and transcription factor IIS (TFIIS) (Calvo and Manley, 2001; McCracken *et al*., 1997). The interaction of CstF-64K with PC4 could not be confirmed by Qu *et al.* (2007) using NMR technique. Another protein, hnRNP F, competes with CstF-64K for binding to the DSE element and thereby inhibits the cleavage reaction in mouse B cells (Veraldi *et al*., 2001).

CstF-64K exists in two forms. The first form is encoded on the X chromosome (Wallace *et al*., 1999). The other, the so-called τ variant, is encoded by a paralogous gene on chromosome 10 (Dass *et al*., 2001 and 2002). Both forms are highly related and share 74.9 % amino acid identity (Dass *et al*., 2002). Perhaps the τ variant has evolved through the inactivation of the X chromosome during meiosis (Handel *et al*., 1991 and 2004).

Different expression levels in mice and rats suggested that CstF-64K and CstF-64K τ can substitute for each other in some tissues and might have complementary functions in other tissues (Wallace *et al*., 2004). The CstF-64K τ protein contains two additional amino acid sequence inserts and contains only nine tandem repeats of MEARA / G (Dass *et al*., 2002). These differences lead to altered affinities for poly(U) and poly(GU). CstF-64K has a higher affinity for poly(U) and a lower affinity for poly(GU) than the τ variant (Monarez *et al*., 2007).

CstF-77K is highly conserved among eukaryotes (Mitchelson *et al*., 1993, Takagaki and Manley, 1994). It is comprised of an N-terminal HAT domain with twelve repeats, which might be involved in mediating protein-protein interactions (Preker and Keller, 1998) and a proline rich segment. CstF-77K interacts with PAP and CPSF-160K. This interaction is suggested to stabilise the CPSF·CstF·RNA complex (Murthy and Manley, 1995). An additional interaction with hFip1, a CPSF associated factor, was shown by Kaufmann and colleagues (2004). CstF-77K was also found to dimerise and bind to the CTD and TFIIS (Takagaki and Manley, 2000; McCracken *et al*., 1997).

1.3.2 Cleavage factor I_m, CF I_m

CF I_m is composed of a small subunit of 25 kDa and a large subunit of either 59, 68 or 72 kDa (Rüegsegger *et al*., 1996). It binds to the UGUAN motif of the pre-mRNA (Brown and Gilmartin, 2003). In absence of the AAUAAA motif, CF I_m can function as a primary determinant in poly(A) site recognition by recruitment of PAP and CPSF through the CF I_m·hFip1 interaction (Venkataraman *et al*., 2005). CF I_m is involved in alternative polyadenylation (Kubo *et al*., 2006). CF I_m activity was reconstituted *in vitro* with only the 25K and 68K subunits (Rüegsegger *et al*., 1998) or with 25K and 59K (Dettwiler *et al*., 2004).

CF I_m-25K has only one conserved domain containing the NUDIX motif. This motif is present in enzymes catalysing the hydrolysis of substrates, consisting of a nucleoside diphosphate linked to some other moiety X (Bessman *et al*., 1996). CF I_m-25K can bind RNA and interacts with all large subunits of CF I_m (Rüegsegger *et al*., 1996), PAP (Kim and Lee,

2001) and PABPN1 (Dettwiler *et al.*, 2004) as well as with U1snRNP-70K (Awasthi 2003) and AIP4 (E3 ubiquitin protein ligase) (Ingham *et al*., 2005). The analysis of the pre-mRNA of CF I_m-25K indicated that it has three different poly(A) sites in its 3' UTR. The largest pre-mRNA is ubiquitously and the two smaller pre-mRNAs are tissue specifically expressed (Kubo *et al*., 2006).

The large subunits of CF I_m, 59K and 68K, are encoded by paralogous genes, whereas the 72K subunit is a splicing variant of 68K which contains one additional exon (Dettwiler *et al*., unpublished). All three large subunits contain an N-terminal RNA recognition motif (RRM), a central proline-rich domain and a C-terminal RS-like domain (Rüegsegger *et al*., 1998). The RS domain is similar to that of the SR proteins which are involved in splicing (Graveley, 2000). All large subunits can bind the hClp1 protein of CF II_m, but they do not interact directly with each other (de Vries *et al*., 2000; Dettwiler, unpublished).

The direct interaction of the RS-like domain of CF I_m-59K with U2AF-65 (U2 snRNP auxiliary factor 65) links the 3' end processing to splicing. U2AF-65 recruits the CF I_m-59K / 25K dimer to the polyadenylation signal (Millevoi *et al*., 2006). Interestingly, the RS-like domain of CF I_m-68K can not interact with U2AF-65. However, it interacts with other members of the SR family of splicing factors like Srp20, 9G8 and hTra2β (Dettwiler *et al.*, 2004).

The RRM domain of the 68K subunit is involved in the protein-protein interaction with CF I_m-25K and not in RNA binding. Therefore the RS-like domain should be required for RNA binding (Dettwiler *et al*., 2004). Ingham and colleagues (2005) demonstrated that CF I_m-68K is bound *in vitro* by several WW-proteins like NEDD4-1, WWOX, CA150, FBP1 and FBP11. WW-domains of proteins mediate protein-protein interactions of proline rich motifs and phosphorylated serines, threonines and proline sites. The biological significance of these interactions is still not known.

1.3.3 Cleavage factor II_m, CF II_m

The activity of CF II_m was separated into two components during purification (de Vries *et al*., 2000). The first one, the essential complex (CF II_m A), contains hPcf11 and hClp1, CF I_m and several splicing and transcription factors. The second one, complex B, had a solely stimulatory function, and its composition is unknown (de Vries *et al*., 2000). Recent publications show that at least the hCpl1 protein is also present in complexes unrelated to pre-mRNA 3' end processing (Paushkin *et al*., 2004; Weitzer and Martinez, 2007).

hClp1 contains Walker A and B motifs, which are known to bind nucleotides (Walker *et al*., 1982). hClp1 is able to bind ATP and GTP (de Vries, unpublished results), and it is the only RNA kinase discovered in humans so far (Weitzer and Martinez, 2007). The free enzyme alone is able to phosphorylate synthetic siRNAs, so that they can be incorporated into the RNA-induced silencing complex (RISC). Furthermore, it phosphorylates ssRNA and dsDNA. Mutations in the Walker A motif lead to an inactivation of the kinase activity. Surprisingly, the Walker motifs are hClp1's only homology to other known kinases (Weitzer and Martinez, 2007). Paushkin and colleagues (2004) showed that hCpl1 is a component of the tRNA splicing endonuclease complex. Human Sen54 and Sen2, two subunits of the tRNA splicing endonuclease complex, interact directly with hClp1 (Paushkin *et al*., 2004). Experiments by Weitzer and Martinez (2007) revealed that kinase and endonuclease activities are present in a single complex and that the 5' phosphorylation of the 3' exon is necessary for tRNA splicing. They suggest, that the kinase activity of hCpl1 might affect mRNA 3' end processing by maintaining the 5' phosphate on the 3' cleavage fragment, which is necessary for the degradation by Xrn2. hCpl1 interacts directly with CPSF and CF I_m (de Vries *et al*., 2000).
Few data is available on hPcf11. Most results were obtained with yeast Pcf11p. It contains a CID (CTD interaction domain) at its N-terminus (Sadowski *et al*., 2003) and recognises the ser-2 phosphorylations of the RNAP II specifically (Licatalosi *et al*., 2002). The structure of the CTD-CID complex was published from Meinhart and Cramer (2004) showing that a β-turn of the CTD binds to a conserved groove in the CID domain of Pcf11. The sequence of hPcf11 possesses two zinc-finger motifs and 30 repeats of the consensus sequence LRFDG. Immunodepletion of hPcf11 disturbed the cleavage activity of the HeLa cells nuclear extract, whereas the polyadenylation activity was not affected (Kaufmann, unpublished results).

1.3.4 Poly(A) polymerase, PAP

PAP catalyses the addition of the poly(A) tail to the newly formed 3' hydroxyl group of the pre-mRNA during 3' end processing reaction. The enzyme belongs to the polymerase ß-type nucleotidyl-transferase super family (Holm and Sander, 1995; Martin and Keller, 1996). It is a template-independent RNA polymerase with low affinity for the RNA primer. PAP alone polyadenylates mRNA slowly in a distributive manner, adding one nucleotide or less per substrate binding event. The polyadenylation efficiency is highly increased by the addition of CPSF and PABPN1, which stabilise the RNA protein complex. Thereby the reaction becomes processive, which means that PAP adds several adenylate residues to the growing poly(A) tail before dissociation (Bienroth *et al*., 1993). *In vitro,* PAP polyadenylates RNAs unspecifically

in the presence of Mn^{2+}. However, in the presence of CPSF and Mg^{2+}, PAP shows a specific polyadenylation activity of pre-mRNAs with an AAUAAA poly(A) signal (Wahle, 1991b; Wittmann and Wahle, 1997).

The crystallographic structure of a PAP fragment (aa 20 to 498) showed a modular organisation with a compact tripartite domain structure. Its catalytic domain is N-terminally located, whereas the RRM is near the C-terminus (Martin and Keller, 1996; Martin *et al*., 1999, 2000 and 2004, Balbo and Bohm, 2007). PAP shares substantial structural homologies with other nucleotidyl transferases (Martin *et al*., 2000; Martin and Keller, 2007). Its C-terminus contains a ser / thr-rich region (SR). The activity of PAP can be down-regulated by phosphorylation at multiple sites of the SR region (Colgan *et al*., 1996 and 1998; Wahle and Rüegsegger, 1999). Three aspartates are essential for catalysis. This catalytic triad coordinates two of three active site metal ions. One of these metal ions gets in touch with the adenine ring of the ATP. Other conserved amino acids contact the nucleotide as well (Martin *et al*., 2000).

1.3.5 Cleavage and polyadenylation specificity factor, CPSF

CPSF is a multimeric protein complex which binds to the highly conserved AAUAAA sequence (Bardwell *et al*., 1991, Bienroth *et al*., 1991). It is necessary for the cleavage and the polyadenylation reaction. CPSF maybe is recruited to RNAP II by TFIID at the transcription initiation site and might be brought to the poly(A) signal by the elongating RNA polymerase II (Dantonel *et al*., 1997; Minvielle-Sebastia and Keller, 1999; reviewed by Proudfoot, 2004). Previously it was shown that CPSF interacts with U2 snRNP (Kyburz *et al*., 2006). This and the previously mentioned CF I_m-59K·U2AF-65 interaction are two examples which show the coupling of pre-mRNA 3' end processing and splicing. CPSF has also functions in the splicing of terminal introns *in vivo* (Li *et al*., 2001) and in cytoplasmic polyadenylation (Dickson *et al*., 1999). Four subunits are known for CPSF (30K, 73K, 100K and 160K) (Bienroth *et al*., 1991, Murthy and Manley, 1992, Jenny *et al*., 1994 and 1996, Barabino *et al*., 1997). An associated factor is Fip1 (Kaufmann *et al*., 2004).

CPSF binding to the AAUAAA signal is weak but can be enhanced by a cooperative interaction with CstF bound to the downstream signal sequence (Wilusz and Shenk, 1990; Weiss *et al*., 1991; Gilmartin and Nevins, 1991; MacDonald *et al*., 1994).

The 30K subunit contains five zinc fingers and a zinc knuckle motif, which are known to bind to nucleic acids. Barabino and colleagues (1997) showed that it binds preferentially to poly(U) sequences but it can also be cross-linked to AAUAAA-containing RNA (Jenny *et al*.,

1994). The Drosophila homologue Clp (clipper) showed endonucleolytic activity against RNA hairpins. This enzymatic activity was localised in the zinc finger motifs (Bai and Tolias, 1996). Due to this fact CPSF-30K has been proposed to be the nuclease (Zarudnaya *et al*., 2002). However, studies with recombinant CPSF-30K or its yeast homologue Yth1p could not confirm this idea (Ohnacker, Barabino and Keller, unpublished results).

Sequence alignments showed, that CPSF-73K and -100K belong to a metallo-β-lactamase / β-CASP subfamily (Callebaut *et al*., 2002). It was suggested that CPSF-73K is the endonuclease for 3' end processing (Ryan *et al*., 2004) and also for histone pre-mRNA processing (Dominski *et al*., 2005a), while CPSF-100K lacks some of the conserved amino acids in the active centre, which are necessary for predicted endonuclease activity. Ryan and colleagues (2004) showed that CPSF-73K can be UV cross-linked to the cleavage site. First experimental evidence, that CPSF-73K is the endonuclease, was provided by Mandel and colleagues (2006) by a crystal structure analysis of a fragment of hCPSF-73K (aa 1 - 460). In these experiments purified recombinant hCPSF-73K, expressed in *E. coli* showed an unspecific endonuclease activity. This activity was not present in the his396 mutant CPSF-73K, which is unable to bind zinc ions at the active centre. However, there was no evidence of any specific endonuclease activity.

CPSF-73K interacts with CPSF-100K (Calzado *et al*., 2004; Dominski *et al.,* 2005b). They share a sequence similarity of 49 % (Jenny *et al*., 1996). The function of CPSF-100K is unknown. It was predicted to be an inactive endonuclease, because it lacks the conserved amino acids which are necessary for the endonuclease activity. Therefore CPSF-100K is suggested to function as a regulator of the enzymatic activity of CPSF-73K (Aravind, 1999).

CPSF-160K contains a bipartite nuclear localisation signal and two RRMs (Jenny and Keller, 1995; Murthy and Manley, 1995). CPSF binds through its 160K subunit preferentially to RNAs containing the AAUAAA sequence (Moore *et al*., 1988). Nevertheless the binding of recombinant CPSF-160K to RNA is weak and is enhanced by the interaction with the other CPSF subunits (Murthy and Manley, 1995) as well as CstF (Wilusz *et al*., 1990). CPSF-160K interacts with CstF-77K and PAP (Murthy and Manley, 1995). Furthermore CPSF-160K interacts with TFIID, thereby forming a connection to transcription (Dantonel *et al*., 1997).

Fip1 was identified to be a subunit of the CPSF complex in 2004 by Kaufmann and colleagues. CPSF preparations from calf thymus mostly lack Fip1 (pers. communication Wahle). Therefore it seems only to be a CPSF associated factor. These fractions are active in polyadenylation. hFip1 stimulates the polyadenylation activity of PAP in an AAUAAA

independent manner (Kaufmann *et al*., 2004). hFip1 binds preferentially to U-rich RNA sequences. It was shown that hFip1 interacts directly with CPSF-30K, CstF-77K and PAP.

1.4 Reconstitution of cleavage activity

The reconstitution of the cleavage activity is an important step to obtain detailed information about the mechanisms of the pre-mRNA 3' end processing reaction, for example the function of each protein can be revealed using mutant subunits. Furthermore, the core of proteins, that are necessary for processing, can be elucidated. The first step for reconstitution was the isolation and purification of all known 3' end processing proteins so far. Rüegsegger and colleagues used CstF and CF I_m purified from HeLa cell nuclear extract, CPSF prepared from calf thymus (Bienroth *et al*., 1991), recombinant bovine PAP and a crude CF II_m preparation for their reconstitution assays. Cleavage activity was obtained using 2 nM CSPF, 2.4 nM CstF, 9.6 nM PAP and 10 nM CF I_m, whereas the amount of the partially purified CF II_m was not determined (Rüegsegger, thesis, 1997). Due to the fact that neither CF II_m nor CPSF preparations were completely pure, no detailed information about the proteins necessary for cleavage could be revealed. The purification of CF II_m by de Vries and colleagues (2000) allowed further reconstitution attempts with proteins obtained from HeLa cell NXT, which were however unsuccessful so far. Therefore Dettwiler and colleagues used the baculovirus system to express all the necessary factors in insect cells. These insect cells support post-translational modifications, which can influence the activity of the proteins (Dettwiler, thesis, 2003). Recombinantly expressed CF II_m containing hClp1 and hPcf11 was not active in a cleavage complex reconstituted from purified proteins, but when it is added to the NXT, depleted for hPcf11, there is activity observed. Even CPSF reconstituted from baculovirus co-expressed CPSF-30K / -100K, baculovirus co-expressed CPSF-73K / -160K and baculovirus expressed hFip1 was not active in cleavage assays using depleted NXT but showed polyadenylation stimulatory activity (Dettwiler, thesis, 2003).

1.5 Aim of this thesis

The pre-mRNA 3' end processing reaction has been studied for 20 years now. It is known that the unstable CPSF·RNA complex (Gilmartin and Nevins, 1989 and 1991; Weiss *et al*., 1991) is stabilised by CstF bound to the DSE (Åström *et al*., 1991, Gilmartin and Nevins, 1989 and 1991; Weiss *et al*., 1991, Wilusz *et al*., 1990). This complex acts as a platform for further binding of CF I_m, CF II_m and PAP (Christophory and Keller, 1988; Takagaki *et al*., 1988). But how the proteins catalyse the 3' end processing cleavage reaction remains unknown.

The failure of the reconstitution experiments lead to the conclusion, that there might be at least one missing factor. This assumption is supported by the comparison of the protein homologs of the 3' end processing machinery in yeast and mammals. Several yeast proteins like Nab4p, Nrd1p and Glc7p have no mammalian homologs and *vice versa* CF I_m-25K and CstF-50K (see Table 6.11 A and B, page 92).

One aim of this thesis was the purification of the complete and functional pre-mRNA 3' end processing cleavage complex assembled on an RNA substrate and the analysis of the bound proteins via mass spectrometry.

The publication of Paushkin and colleagues (2004) showed that the CF II_m subcomplex can be affinity purified from HEK293 cells stably expressing his_8-flag tagged hClp1. These preparations are active in cleavage assays using depleted NXT (Kyburz, thesis, 2006). The reconstitution of pre-mRNA 3' end processing cleavage reaction with factors, purified from human cell lines, was not tested so far. The second aim was therefore the affinity purification of cleavage factors stably expressed in human cells. CF I_m, CstF and CF II_m as well as CPSF (Wlotzka, diploma thesis, 2006) were affinity purified from HEK293 cells, analysed in their composition by mass spectrometry and tested for their activity in antibody-depleted nuclear extract. The total reconstitution of the cleavage reaction from these proteins, plus recombinant bovine PAP, would prove their activity and permit the possibility to address the function of each protein subunit.

2. Results

The pre-mRNA 3' processing complex in mammals consists of four multimeric protein factors, CPSF, CstF, CF I_m and CF II_m and a fifth factor, the poly(A) polymerase (PAP) (Wahle & Keller, 1992). Following the catalysis of the 3' end cleavage reaction, the complex is remodelled. *In vitro* the complex, consisting of CPSF and PAP, starts the polyadenylation of the 5' fragment. After the addition of about ten nucleotides PABPN1 joins the polyadenylation-specific complex, and a poly(A) tail of around 250 nucleotides is added to the pre-mRNA.

Bovine PAP, active in polyadenylation, can be purified as recombinant protein from *E. coli* in large quantities (Wahle, 1991b). Active CF I_m was reconstituted as a hetero-dimer of the subunits CF I_m-68K and CF I_m-25K or CF I_m-59K and CF I_m-25K respectively (Rüegsegger *et al*., 1998; Dettwiler *et al.*, 2004). Active CstF was purified from HeLa cell nuclear extracts (Takagaki *et al*., 1989 & 1990) and was shown to contain the three subunits CstF-50K, -64K and -77K (Gilmartin & Nevins, 1991; Takagaki *et al*., 1990). The composition of CF II_m was tentatively determined by purification from HeLa cells. After purification over seven columns the purified CF II_m consists mainly of hClp1 and hPcf11, but in addition all CF I_m subunits and various other proteins were found in the preparation as well (de Vries *et al.*, 2000). CPSF consists of four subunits: 30K, 73K, 100K and 160K (Bienroth *et al*., 1991, Murthy & Manley, 1992; Jenny *et al*., 1994). Fip1 was found to be an integral subunit of human CPSF and interacts with the PAP (Kaufmann *et al*., 2004), but Fip1 was not detected in CPSF preparations of *Bos taurus*. Therefore Fip1 seems to be an associated factor. Initial preparations of CPSF from calf thymus (I and II) and HeLa cells showed activity in cleavage and polyadenylation (Christofori & Keller, 1988; Gilmartin & Nevins, 1989) but repeated preparations were inactive in cleavage (personal communication Kyburz and own results). The analysis and reconstitution of the subunit composition of CPSF is the aim of the PhD work of Ringel and is not further discussed in here.

The aim of this thesis was the analysis of the protein composition of the pre-mRNA 3' end processing cleavage complex, called cleavage complex. For this purpose, it was tried to purify the complete and active cleavage complex. The first part of this Chapter describes different approaches for the purification of the cleavage complex assembled on immobilised mRNA.

2.1 Purification of the pre-mRNA 3' end processing cleavage complex via immobilised pre-mRNA

Different approaches have been developed to purify RNA·protein complexes (RNPs). They are based on the immobilisation of the RNA on a specific affinity matrix. For this reason a cell extract containing the proteins of interest has to be incubated with the RNA matrix. After RNP formation the complex is specifically eluted and further analysed. In this work the tobramycin affinity selection method and the λN peptide affinity method were applied to purify the cleavage complex.

2.1.1 Optimisations of the *in vitro* cleavage reaction

The affinity purification of RNA·protein complexes requires large amounts of assembled complexes. The amount of assembled, active cleavage complex can be roughly estimated through the quantity of 5' cleavage product obtained. Starting with cleavage reaction conditions described in Table 2-2, only 10 to 20 % of the input pre-mRNA was cleaved. For this reason optimisation of the reaction condition was necessary to obtain higher yields of cleavage product (CP) in cleavage assays. Optimisation of cleavage complexes were performed in electrophoretic mobility shift assays (EMSA).

The RNA substrates L3 and SV40 late were used for the cleavage reaction; their cleavage inactive variants L3Δ and SV40 late Δ were used as controls. These Δ variants have a point mutation in the highly conserved AAUAAA sequence, U is mutated to G. Furthermore, shorter variants of L3 and SV40 late lacking 100 nt respectively 80 nt of non-essential sequences at the 5' end were used. These substrate RNAs were designated as L3s and SV40s and showed no differences in the cleavage reaction, compared to the longer substrates.

The given reaction conditions (see Table 2-2) were varied, and the cleavage reaction was optimised with respect to substrates, competitors, salt concentrations, additives and incubation times as summarised in Table 2-1.

Magnesium ions coordinate the phosphate groups of ATP or 3' dATP, respectively. Increasing concentrations of $MgCl_2$ improved the activity of the cleavage reaction but activated RNases as well (Figure 2-1 A). For this reason, the concentration was kept at 1.5 mM $MgCl_2$.

In parallel, electrophoretic mobility shift assays were used to control the amount of specific complexes formed on RNA substrates. PVA inhibits the entry of complexes into the native gel (data not shown). Thus it was substituted by PEG 6000. This exchange leads to the separation of the complexes in the native gels.

The influence of PEG 6000 concentration in the cleavage assay is shown in Figure 2-1 B. The percentage of cleavage product increases with higher quantities of PEG 6000 in the reaction. Best results were obtained at a concentration around 3 %. As expected, the control RNA (L3Δ) was not cleaved.

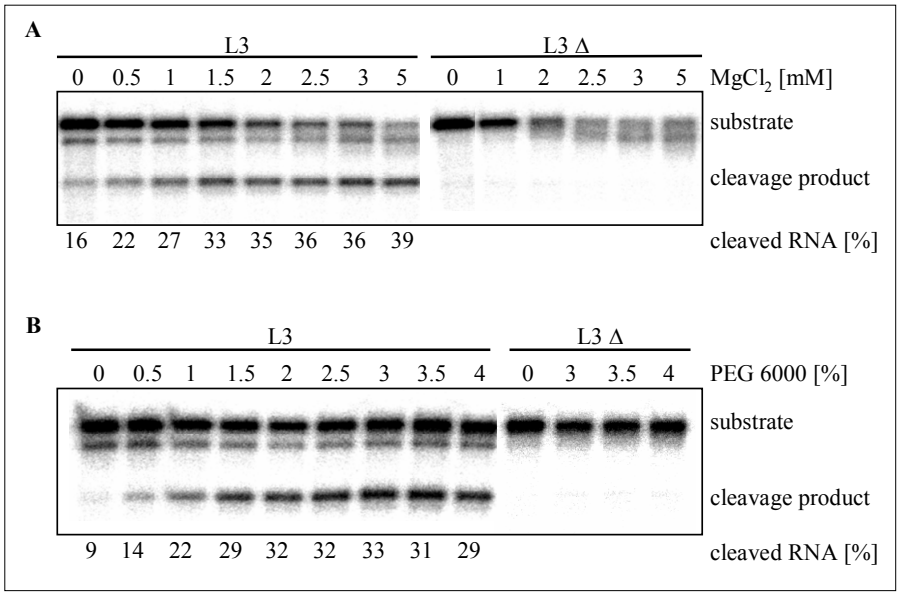


Figure 2-1 High yield of cleavage product revealed around 3 % of PEG 6000 per reaction, while increasing Mg^{2+} concentration leads to increased degradation
A) Standard protocol with different Mg^{2+} concentrations. Reaction performed for 2 hours at 30 °C. Cleavage product was calculated by measuring intensity of the band from the complete RNA per lane (100 %) compared to cleavage product with software ImageQuant (Methods 5.12). Purified L3 RNA contained a smaller degradation product, which appears for the L3Δ variant in unspecific degradation events.
B) Standard protocol with different concentrations of PEG 6000 instead of PVA. Reaction performed for 2 hours at 30 °C. Cleavage product was calculated by measuring intensity from the complete RNA per lane (100%) compared to cleavage product with software ImageQuant.

The use of different volumes of nuclear extract (NXT) changes the total KCl concentration in the reaction and may influence the activity. For this reason we tested different potassium chloride and potassium acetate concentrations. NXT-A, prepared as described in Wahle and Keller (1994) (see Chapter 5.10.2), showed the highest cleavage activity at 100 mM of potassium ions regardless of their anion (see Table 2-1). The activity decreased at higher concentrations of potassium acetate. We tested other salts like ammonium sulphate and ammonium acetate to change the ionic strength but found that they inhibited cleavage (data not shown).

Table 2-1 Cleavage assay optimisation in a 25 µL volume

component	concentrations tested	range for highest activity	comments
tRNA	0 – 5 µg/µL	0 – 0.5 µg/µL	no influence
total RNA (yeast)	0 – 5 µg/µL	0 µg/µL	degradation increased
substrate RNA	0.08 – 20 nM	0.4 – 1 nM	nearly 50 % of RNA substrate cleaved
NXT-A	1 – 15 µL	10 – 15 µL	nearly 40 % of RNA substrate cleaved
PEG 4000/6000	0 – 4 %	3 % PEG 6000	less RNA substrate cleaved with PEG 4000
PEG 6000	0 – 4 %	2 – 3.5 %	32 % RNA substrate cleaved
Mg^{2+}	0 – 5 mM	1.5 – 5 mM	unspecific degradation increased with higher [Mg^{2+}]
3’ dATP	0 – 2 mM	0.2 – 0.8 mM	reduced cleavage activity at higher concentrations
Potassium chloride	20 – 100 mM	100 mM	tested with 5 µL NXT-A
Potassium acetate	0 – 300 mM	100 mM	substituted for KCl, higher concentration inhibits reaction
Ammonium sulphate	0 – 100 mM	0 mM	inhibits reaction in 5 µL NXT-A
Ammonium acetate	0 – 300 mM	0 mM	inhibits reaction in 5 µL NXT-A
Zinc chloride	0 – 50 µM	0 µM	Ryan *et al*., 2004, no influence
time	30 – 120 min	120 min	increasing amounts of cleaved substrate RNA (*)
RNAP II CTD	50 – 600 ng	0 ng	no influence
phosp. RNAP II CTD	50 – 600 ng	0 ng	no influence

NXT-A preparation see Methods 5.10.2, * excess of substrate RNA permit the association of the cleavage complex after the dissociation following the endonucleolytic cut of another RNA molecule. CTD was unspecific phosphorylated using the MPF-complex (gift from Martin). The phosphorylation of the CTD was checked using α-ATP in a phosphorylation assay (15 min at 30 °C).

Dialysis of NXT leads to a partial loss of cleavage activity (Wahle & Keller, 1994). Ryan and co-workers suggested (2004) that CPSF-73K loses some of its Zn^{2+} ions during dialysis leading to an inactivated fraction of the protein. They showed that the addition of $ZnCl_2$ in a µM range is sufficient to restore the cleavage activity. According to their protocol, we tested different concentrations of $ZnCl_2$ in the cleavage assay, but were not able to confirm these results for dialysed NXT-A.

Ryan and colleagues (2002) showed that addition of the C-terminal domain of RNA-Polymerase II can increase the activity of the pre-mRNA 3' end processing complex. Tests using purified CTD and also CTD, which was unspecifically phosphorylated, showed in neither case an increase in the cleavage activity (data not shown).

After optimisation of single reaction parameters, combinations of those optimal parameters were used for cleavage assays. The observed increase in cleavage activity for these combinations was less than the additive effects of the optimisation of the single parameters. Therefore some concentrations were kept as before. Table 2-2 shows the conditions yielding the highest activity for the pre-mRNA 3' end processing reaction that were used for further experiments. 40 to 50 % of the input RNA was cleaved.

Table 2-2 Comparison cleavage assay conditions before and after optimisation

component (25 µL reaction volume)	conditions before optimisation	conditions after optimisation
NXT-A	12.5 µL	12.5 µL
RNA	4 nM	**2 nM**
DTT	2 mM	2 mM
$MgCl_2$	1.5 mM	1.5 mM
3' dATP	0.8 mM	0.8 mM
tRNA	0.1 µg/µL	0.1 µg/µL
PVA	2.6 %	-
PEG 6000	-	**3 %**
KCl	-	**50 mM**
creatine phosphate	20 mM	20 mM
RNAguard	10 U	10 U

Components which were changed are shown in bold letters.

2.1.2 Optimisation of cleavage complex formation using the EMSA technique

The EMSA method can determine if a protein or a protein mixture is capable of binding to a given RNA sequence. We performed electrophoretic mobility shift analysis (EMSA) to analyse the complex formation with RNA and to increase the specific binding of the proteins from the nuclear extract to the RNA, which was a requirement for the purification of RNA·protein complexes.

First the experiments were done according to the protocols of Humphrey and colleagues (1987). They detected a specific RNA·protein complex on ^{32}P-labelled L3 RNA after incubation with nuclear extract, which was not formed with the mutant RNA. Furthermore, this specific complex on L3 RNA was stable in competition assays using non-labelled competitor RNA with variants in the AAUAAA sequence, while it dissociated in the presence of non-labelled L3 RNA. They concluded, that the specific complex formation is dependent on an intact poly(A) signal.

The specific complex of substrate RNA was detected, but not of the RNA containing the U to G mutation in the AAUAAA sequence, as can be seen in Figure 2-2 (compare lanes 2 to 4 with 8 to 10). The specific complex was weakly populated compared to the heterogeneous complex containing unspecifically bound proteins. The complex assembled during 10 min and dissociated between 90 and 120 min after incubation, while no complex formation appeared for L3Δ. In order to increase the yield of specific complex, we tested different conditions and included competitors for unspecific protein binding like heparin. The results of these assays are summarised in Table 2-3.

Heparin is one of the most commonly used substances preventing unspecific protein binding to nucleic acids. We tested several concentrations and observed a better separation between the specific and the unspecific complex, but cleavage assays demonstrated that the lowest concentration of heparin inhibits the reaction completely. For further EMSA experiments the conditions were kept as before because no increase in the amount of specific protein complex was achieved by the variation of the conditions.

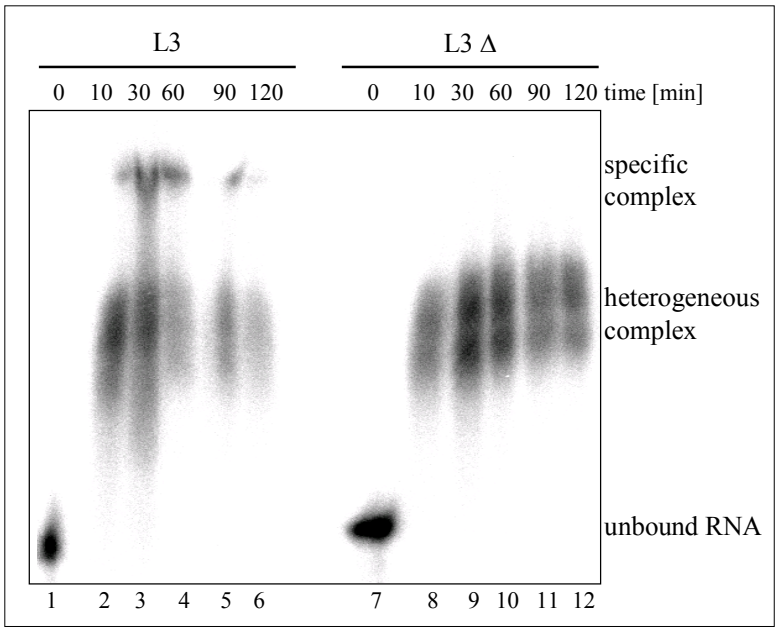


Figure 2-2 L3 is shifted in a specific complex
Conditions from Humphrey *et al*. (1987), with 0.8 mM 3' dATP, 20 mM creatine phosphate, 0.5 mM $MgCl_2$, 10 % glycerol, 50 mM KCl, 0.05 mM DTT, 1 % PVA, 0.1 mg / mL tRNA, 10 µL NXT-A. Reactions were performed at 30 °C for different time periods as indicated at the top and frozen in N_2. Prior to loading, 1 µL of Heparin [25 mg/mL] was added to 5 µL probe and incubated for 10 min on ice. The yield of cleavage complex was calculated in per cent by measuring intensity of blackness from the complete RNA per lane (100%) compared to cleavage product with software ImageQuant.

Table 2-3 EMSA optimisation

component	concentration	range for best complex formation	comments
Heparin	0 – 20 µg / µL	5 µg / µL	inhibits cleavage at lowest concentration, better separation of specific and heterogeneous complex
tRNA	0.02 – 0.4 µg/µL	0.02 – 0.08 µg/µL	no influence
yRNA (total)	0 – 0.4 µg/µL	-	no complex formation
Mg^{2+} (buffer, gel) (*)	-	-	no significant difference
Gel with/without agarose	-	-	no significant difference, better handling with agarose
gel run at room temperature	-	-	no significant difference

Reaction volume 25 µL, * $MgCl_2$ was added to the gel, respectively the running buffer to equalise the EDTA in the buffer. Agarose was added to the polyacrylamide solution for easier handling, it has no influence on the gel condition. Gels were run at 4 °C if not mentioned otherwise.

2.1.3 Complex purification using the tobramycin affinity selection method

Several approaches were developed to isolate functional ribonucleoprotein (RNP) complexes on a RNA affinity matrix. The tobramycin affinity selection method is one of these techniques. It was developed by Hartmuth and co-workers (2002) to obtain preparative amounts of purified pre-spliceosomes under native conditions and is based on the specific binding of an RNA aptamer to the aminoglycoside antibiotic tobramycin. We have chosen this method to purify the pre-mRNA 3' end processing cleavage complex on substrate RNAs like L3 and SV40 late from nuclear extracts of HeLa cells.

A purification scheme for the tobramycin affinity selection method is shown in Figure 2-3. Binding of the RNA to tobramycin is carried out by a special RNA sequence (the aptamer), which forms a hairpin loop. RNAs that include those aptamer sequences can be immobilised on tobramycin-Sepharose. After complex formation in nuclear extract, the complexes can be eluted with an excess of free tobramycin.

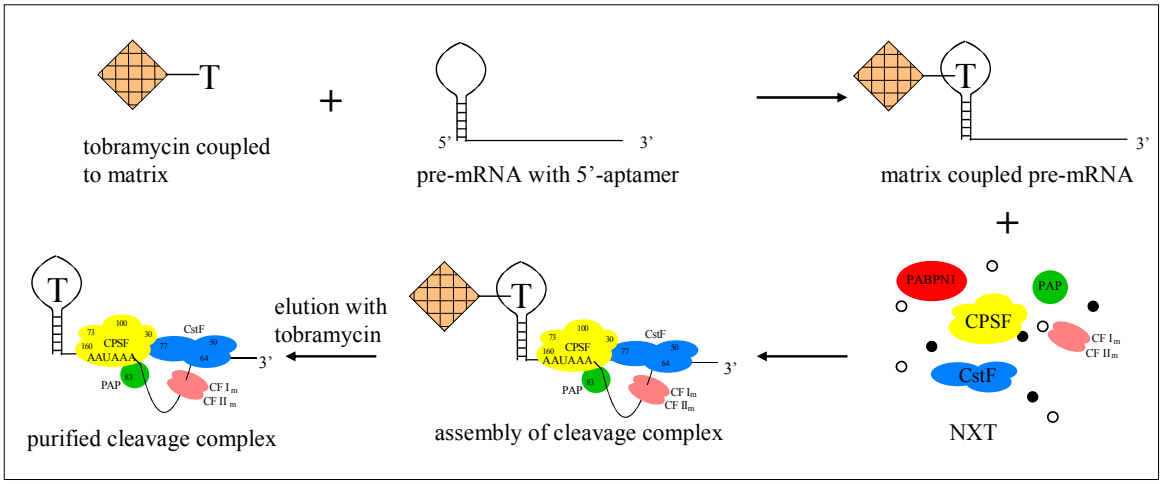


Figure 2-3 Scheme of tobramycin affinity purification
Modified from Hartmuth *et al*. (2002).

We prepared tobramycin coupled to Sepharose as described in Chapter 5.8.2, as well as different RNAs containing the aptamer at the 5' or 3' end (Chapter 5.8.1). All substrates are summarised in Table 6-2 in Chapter 6. The substrates were tested in standard cleavage and polyadenylation assays to verify that the aptamer does not influence the cleavage reaction. These experiments showed that the constructs containing the tobramycin aptamer are substrates for the cleavage reaction (data not shown).

Hartmuth observed that splicing reactions using immobilised RNAs are slowed down (pers. communication). For this reason kinetics of the cleavage reaction, using 5' immobilised RNAs, were done to optimise the ratio of formed cleavage complexes compared to the

cleaved substrates. Doing this time course, the expectation was to find cleaved RNA and untouched substrate RNAs in the elution fraction, whereas the 3' fragment of the RNA or its degradation products would be present in the supernatant of the complex binding reaction. Unfortunately, most RNA as well as the cleavage product was found in the supernatant of the binding reaction. Thus, we concluded that the RNA dissociates from the tobramycin matrix during incubation with nuclear extract. In consequence the RNP complex formation and the cleavage reaction may occur in the supernatant, whereas the assembly of a cleavage active complex on the tobramycin matrix was not achieved (data not shown).

Later tobramycin affinity experiments contained more input RNA. Although more than 50 % of the RNA dissociated from the matrix during NXT incubation, around 10 % of input RNA was still bound to the beads prior to elution. The analysis of these tobramycin affinity selection experiments by SDS-PAGE demonstrated that proteins from the nuclear extract were unspecifically bound in all samples, independently of the immobilised RNA, as seen in Figure 2-4 (compare lanes 3, 5, 7 and 8). In this experiment, an aliquot of tobramycin beads was incubated without any RNA for two hours in NXT-A to detect unspecific binding to the Sepharose matrix (lane 8). Another control, L3 (the RNA without aptamer) was not expected to bind the beads at all (lane 2 and 3). The protein profile is equal to that in lane 8 (no RNA),

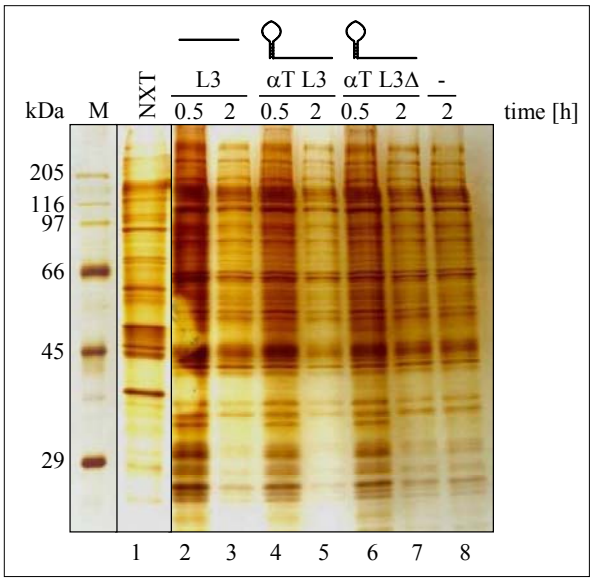


Figure 2-4 Specific complexes may not be formed on the tobramycin affinity matrix due to high protein background

Silver-stained SDS polyacrylamide gel showing the eluted fractions from tobramycin beads after 30 min or 2 hours incubation with NXT-A. Used RNAs and incubation time are indicated at the top. αT represents aptamer binding site for tobramycin at the 5' position. The secondary structures of the RNA substrates are indicated at the top. Lane 1 shows the proteins profile of the NXT ($^{1}/_{15}$ input). Protein marker weight is indicated at the left.

as expected. In lane 8, the same protein bands were present as in lanes 4 to 7, although here the specific RNA, respectively the L3Δ variant, was bound to the matrix. In conclusion, no specific protein binding to the immobilised RNA was achieved. Thus, we varied the reaction conditions to reduce unspecific protein binding to the tobramycin matrix.

2.1.3.1 Attempts to reduce unspecific protein binding to the matrix

First we added increasing amounts of BSA to the blocking buffer. We expected that this protein may block the Sepharose, thereby inhibiting the unspecific binding of other proteins. However, SDS-PAGE analysis showed that the amount of unspecifically bound protein was nearly the same.

Nuclear extracts contain membrane fragments, which were not completely removed during extract preparations. These membrane fragments may be the reason of unspecific protein binding and therefore interfere with the detection of the specific RNP formation. We tried to remove these fragments by ultracentrifugation of NXT-A for 3 hours at 250,000 xg. Fractions were taken from top to bottom and analysed by SDS-PAGE and cleavage assay. The pellet was dissolved in one additional volume of dialysis buffer D. Figure 2-5 A shows the distribution of proteins of each fraction. Fractions 1 to 7 look very much the same. Fraction 8 to 11 contained additional proteins up to sizes of around 300 kDa, and in the pellet an equal distribution of proteins of all sizes was found. All fractions were tested in standard cleavage assays and showed increasing activities from fraction 1 to fraction 11 (Figure 2-5 B). The resolubilised pellet was active in the cleavage assay as well.

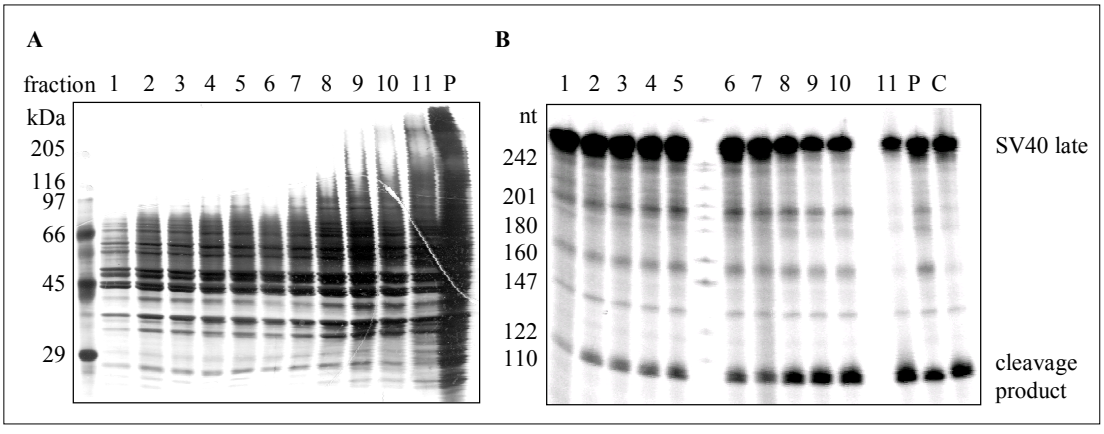


Figure 2-5 Ultracentrifuged NXT-A shows cleavage activity in all fractions
NXT-A was ultracentrifuged for 3 h at 250,000 xg. Fractions were taken from top (no. 1) to bottom (no. 11), the pellet was resuspended in one pellet volume buffer D (P). The samples were analysed by SDS-PAGE (A) and tested in a standard cleavage assay (B).
A) Protein standard weight is indicated in kilo dalton on the left.
B) Control reaction for cleavage activity was performed with NXT-A prior to ultracentrifugation (lane C). Nucleotide standard is indicated on the left (B).

Three fractions (1, 3 and 7) were tested in the tobramycin affinity selection method (see Figure 2-6). The eluted samples were analysed by silver-stained SDS-PAGE and Western blot analysis using a CPSF-100K antibody. In the silver-stained gel no difference in the protein pattern was detected between the fraction of SV40s - αT and SV40sΔ - αT (Figure 2-6 A), although SV40sΔ - αT seemed to bind slightly more protein. The amount of these proteins is lower than in the control extract without ultracentrifugation. The protein amount increased from fraction 1 to 7 independent of immobilised RNA. Western blot analysis, using an antibody against CPSF-100K, revealed equal amounts of CPSF-100K protein in each elution sample (Figure 2-6 B). We concluded that CPSF-100K was unspecifically bound to the Sepharose matrix.

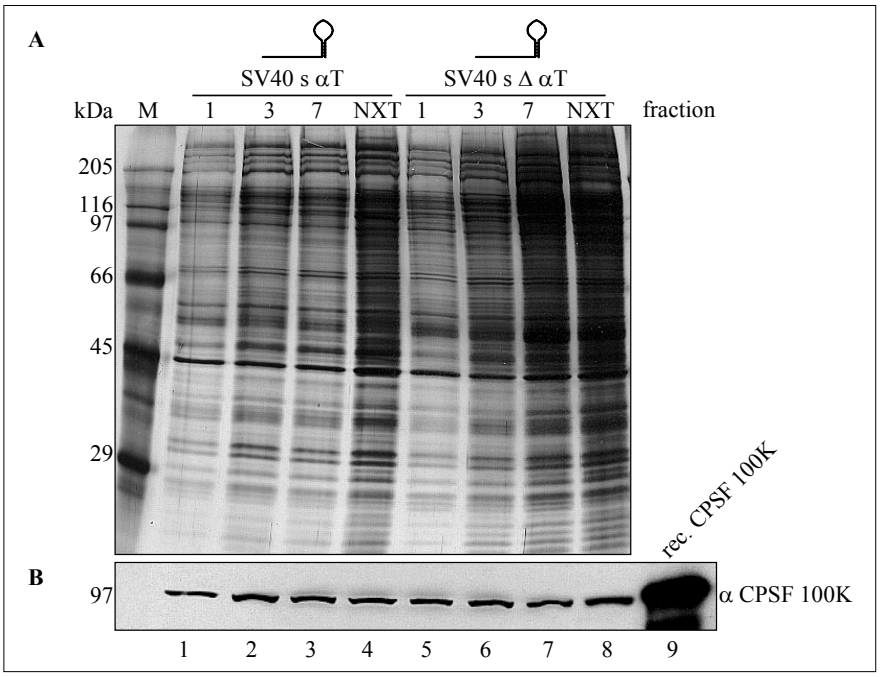


Figure 2-6 No specific complex is formed on the tobramycin affinity matrix using different fractions of ultracentrifuged nuclear extract (NXT-A)
A) Silver-stained SDS polyacrylamide gel of tobramycin elution fractions. RNA and fraction numbers are indicated at the top, NXT is the NXT-A prior to ultracentrifugation, which was used as a control. Protein marker (M) is indicated at the left.
B) Western blot analysis with CPSF-100K antibody of the elutions from tobramycin affinity method. Recombinant CPSF-100K was used as a positive control.

Only a modest decrease in unspecific protein binding to the tobramycin matrix was achieved using ultracentrifuged NXT. In addition, the ultracentrifugation of nuclear extract resulted in a decrease of cleavage activity.

Using the RNA substrates with the 3' aptamer instead of 5' aptamer did not lead to an improved signal either, as well as additional wash steps and the reduction of the amount of

beads. The reduction of used matrix resulted only in a slight decrease in background. No specific RNP complex formation was detectable in any case. The unsuccessful attempts to reduce the background and increase the specific complex formation are summarised in Table 2-4.

Table 2-4 Summary of varied conditions for tobramycin affinity selection method

condition	comments
addition of BSA to blocking buffer [0.5 – 5 mg/mL]	no difference in unspecific protein binding
additional washing with reaction buffer	unspecific proteins stayed bound to Sepharose
Sepharose depleted NXT [0.5 – 5 mg/mL]	less unspecific protein binding, no specific protein binding to RNA
ultracentrifugation of NXT (1h 100,000 xg)	less unspecific protein binding, no specific protein binding to RNA
5' or 3' immobilised substrate RNA	no differences, much unspecific protein binding, no specific protein binding to RNA
reduction of beads, increase of RNA amount	poor elution with tobramycin

2.1.4 Complex purification using the λN peptide affinity method

The λN peptide affinity method was developed by Czaplinski and colleagues (2005) (see Methods 5.9) to examine RNP formation on long RNAs. This approach is based on the λ phage N antiterminator peptide, which binds specifically to the BoxB sequence in λ phage RNA. The λN peptide is fused to GST and immobilised on a glutathione Sepharose matrix via GST·glutathione interaction. RNA containing the BoxB element is able to bind the λN peptide specifically and is immobilised on the Sepharose matrix. This method was tried for purification of the pre-mRNA 3' end processing cleavage complex using a magnetic glutathione matrix instead of glutathione Sepharose.

2.1.4.1 Characterisation of the GST-λN protein and the BoxB substrates

The λN peptide containing a N-terminal GST-tag and a C-terminal his_6-tag was expressed in BL21 pUBS. The purified protein was tested in electrophoretic mobility shift assay with L3 (negative control) and a RNA containing only two BoxB elements to confirm the specific binding of λN peptide to the BoxB element (Figure 2-7). Increasing amounts of GST-λN

protein were added to the RNAs. After 15 min incubation at room temperature the RNA·protein mixture was loaded on a native polyacrylamide gel. The $(BoxB)_2$-RNA is shown schematically on the left side of the Figure. Due to its secondary structures, three bands appeared in the lane without protein. These secondary structures were avoided, when the RNA was heated for one minute at 95 °C and afterwards chilled on ice prior to the protein incubation. Two of the secondary structures are shifted in the presence of the protein as indicated on the left side of the Figure. At the highest protein concentration most of the $(BoxB)_2$ RNA was shifted and bound to several λN protein molecules as seen by the appearing bands. The control RNA L3 was not bound, as expected. Two bands appeared. The lower band corresponded to the unfolded RNA, whereas the other band resulted from an unspecific secondary structure. The L3 RNA pattern was not changed, even at high amounts of GST-λN protein. We concluded that the expressed and purified GST-λN protein binds specifically to the BoxB RNA element.

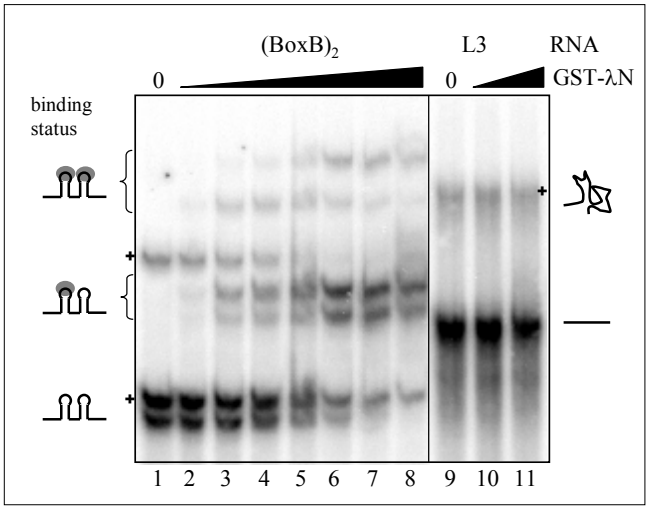


Figure 2-7 GST-λN protein specifically binds to BoxB containing RNA
EMSA using $(BoxB)_2$ RNA and L3 as substrates, increasing amounts of GST-λN protein were added to the reaction as indicated on the top. Both RNAs fold into secondary structures as indicated with +. These secondary structures could be avoided when the RNA was heated for one minute at 95 °C and afterwards chilled on ice. $(BoxB)_2$ RNA forms two hairpin loops which are bound by the GST-λN protein as indicated at the left side of the Figure. Grey ovals symbolise bound GST-λN molecules on the $(BoxB)_2$ RNA. $(BoxB)_2$ RNA has two main and one minor secondary structure which were all bound by the GST-λN protein. To control the specificity of the GST-λN binding, an RNA (L3) without BoxB element was incubated with GST-λN protein. This RNA has a minor secondary structure as well as indicated at the right site of the Figure. This RNA does not bind to the protein.

Several BoxB elements were inserted as tandem copies in plasmids coding for cleavage substrates (Table 6.4 in Chapter 6). The dissociation constant of the interaction of GST-λN protein with the L3 RNA containing two BoxB elements or none respectively, was determined by filter binding assays. The K_D for L3 was estimated to be about 1000-6000 nM,

while the K_D for $(BoxB)_2$-L3 was 3-17 nM. This indicated that the binding of the GST-λN protein is specific to the L3 substrate containing two BoxB motifs.

Cleavage assays demonstrated that the BoxB elements had no influence on the cleavage activity. $(BoxB)_2$-L3 was cleaved as efficiently as L3, and $(BoxB)_2$-L3Δ was no cleavage substrate as expected (data not shown).

2.1.4.2 Preliminary experiments of λN peptide affinity method

Paramagnetic beads, MagneGST™ glutathione particles (Promega) were used, because the experiments with the tobramycin affinity selection method demonstrated that Sepharose matrix bound many proteins from the nuclear extract unspecifically. First we tested RNA binding to the beads. RNA with and without BoxB elements were bound to the beads equally well. To prevent unspecific RNA binding, the addition of BSA and tRNA as competitors was tested, as summarised in Table 2-5. BSA and / or tRNA were incubated in buffer D with the beads for one or 12 hours to block the matrix. After three washes with buffer D, the radioactively labelled RNA was incubated with the beads in buffer D for 1 hour at 4 °C. The bound RNA was determined by scintillation counting. Whereas BSA reduced the unspecific binding significantly, the addition of tRNA did not reduce the unspecific binding further. Increased blocking time achieved no significant improvement.

Table 2-5 Preventing unspecific RNA binding to glutathione paramagnetic beads.

blocking reagent	blocking time [h]	unspecific RNA binding [%]
-	-	47
BSA	12	3
BSA + tRNA	12	5
BSA	1	2
BSA + tRNA	1	8

After blocking, beads were incubated with radioactively labelled RNA for one hour at 4 °C in buffer D. The radioactivity bound by the beads was measured by scintillation counting and given in % of input RNA.

In contrast to the protocol of Czaplinski and colleagues, who used SDS for elution, we used a specific elution with reduced glutathione to obtain native RNPs. Therefore we tested different protocols using only the GST-λN protein bound to magnetic matrix. Buffer 1 contained 10 mM reduced glutathione, 50 mM Tris pH 8.0 and 0.05 % Triton X-100. This elution buffer resulted in a good recovery of GST-λN protein, but eluted the blocking reagent BSA as well.

Buffer 2, as applied by Promega Inc. for specific elution, contained 50 mM reduced glutathione in 50 mM Tris pH 8.0. Neither GST-λN nor BSA was eluted by this elution buffer. Buffer 3 contained additional 1 % Triton X-100 and 100 mM NaCl in buffer 2 as suggested by Promega. Almost all GST-λN protein was eluted together with only faint amounts of BSA. As a control, the elution buffer was tested according to the protocol of Czaplinski. This elution buffer contained 0.1 % SDS, 2.5 mM Tris pH 6.8 and 5 mM DTT. The result was a recovery of GST-λN but as well as of BSA. Elution buffer 3 was chosen to use, since this allowed the specific elution of GST-λN protein and only small amounts of the blocking reagent BSA.

Next we examined whether the order of addition of components (RNA, the GST-λN protein and NXT) had any influence on complex assembly. First the GST-λN was bound to the beads and the RNP complex was assembled in a second tube by incubating RNA and NXT together. Only 22 % of the RNA was bound after incubation of RNA-NXT with the beads. A huge amount of unspecific RNA binding occurred in the control sample using RNA without BoxB as substrate. 55 % of the bound RNA was eluted. The silver-stained SDS polyacrylamide gel showed a high background of eluted proteins for these samples. In the second experiment we incubated NXT, RNA, GST-λN protein and magnetic beads together in one tube. In this case only 10 % of the input RNA was bound to the beads and the elution fractions displayed background as well. In the third attempt we added NXT to the pre-bound GST-λN-RNA beads. Here RNA binding was specific and sufficient for the following complex assembly experiment. The elution of the RNA was inefficient, only 15 % were eluted. Again we observed unspecific protein binding in the silver-stained SDS polyacrylamide gel. We decided to vary the conditions of the third experiment, to further improve the specific elution of RNA from the beads.

2.1.4.3 RNP formation with standard cleavage substrates during λN peptide affinity

To determine the incubation time that would allow maximum complex assembly and minimal unspecific protein binding, the cleavage reaction mix was incubated with the RNA bound to the magnetic matrix (RNA-affinity matrix). Samples were taken at different time points. The samples were analysed by silver-stained SDS polyacrylamide gel and western blot analysis; RNA binding and release was determined by scintillation counting. For this experiment we used $(BoxB)_2$-L3 RNA as substrate and a small RNA containing only two BoxB elements to detect proteins which were unspecifically bound to the BoxB elements. In Figure 2-8 A, a diagram of the distribution of the radioactively labelled RNA is shown, measured in the

different fractions during the experiment for three different time points. The comparison of the sample taken at 10 min with samples from longer incubation times showed a time dependent dissociation or degradation of the RNA. The specific elution resulted in a very low yield at all time points. Most of the RNA was still bound to the beads and was released only by incubation with SDS sample buffer. The analysis of the protein pattern in a silver-stained SDS polyacrylamide gel is shown in Figure 2-8 B. In the specific elution (E1) with reduced glutathione, the GST-λN protein was eluted. Unspecific protein binding increased with the time, whereas no specific complex formation was detectable at all. The RNA pattern is shown in Figure 2-8 C.

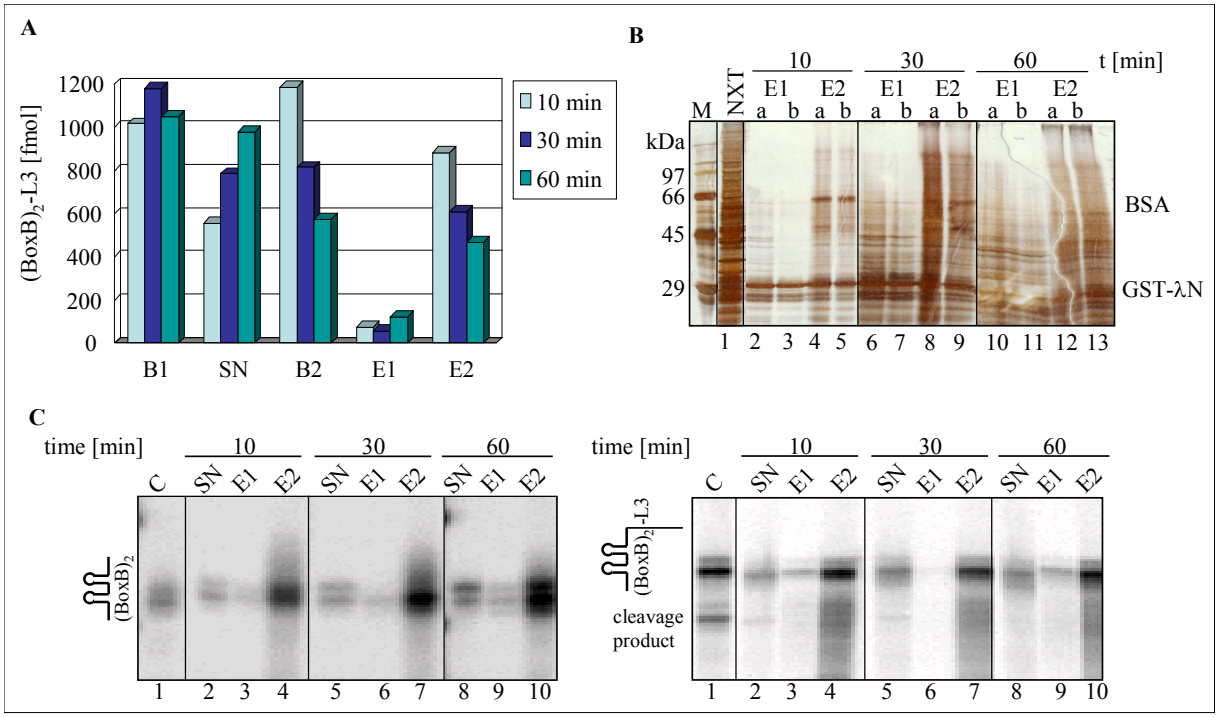


Figure 2-8 Longer incubation time lead to an increase in unspecific protein binding and unspecific RNA dissociation

λN peptide affinity experiment, 2000 fmol of labelled $(BoxB)_2$ or $(BoxB)_2$-L3 RNA were used for immobilisation. RNA was immobilised via binding to the λN protein separately for each time point. Immobilised RNA was determined by scintillation counting after RNA binding (B1). The matrix with bound RNA was incubated in the cleavage reaction mix containing NXT-A for 10, 30 and 60 min as indicated. The matrix was collected, the supernatant was removed and the RNA amount was determined (SN) as well as the RNA amount bound to the matrix (B2) after washing. The RNA·protein·complexes were specifically eluted with elution buffer 3 containing reduced glutathione (E1) and further unspecifically with SDS sample buffer (E2).

A) Diagram for the amount of radioactively labelled RNA in all steps during the λN peptide affinity for 3 time points.

B) Silver-stained SDS polyacrylamide gel shows the protein pattern in the elution fractions after 10, 30 and 60 min incubation on the RNA affinity matrix. $(BoxB)_2$ indicated by 'a' and $(BoxB)_2$-L3 all lanes with 'b'. Protein marker is indicated on the left site, time points on the top.

C) Autoradiography of the denaturing polyacrylamide gel shows that the cleavage reaction for the substrate $(BoxB)_2$-L3 RNA occurred only in the supernatant (SN) at 10 and 30 min time points. $^1/_3$ of the SN and elution fractions were treated with proteinase K digestion and the RNA was separated on the denaturing gel. A control standard cleavage reaction was performed for both RNAs (lanes C) to see the size of the cleavage product for $(BoxB)_2$-L3 RNA respectively that no reaction occurs with the $(BoxB)_2$ RNA. The RNA substrates and their theoretical secondary structure are indicated at the left site of the autoradiographies.

After RNA isolation from all samples, the RNAs were separated on a 10 % denaturing gel. As a control, standard cleavage assays were performed for $(BoxB)_2$ and $(BoxB)_2$-L3 RNA, to show the size of the cleavage product and that no cleavage occurs for $(BoxB)_2$ RNA. We observed a small amount of cleavage product in the flow through for the 10 and 30 min time points, but not in the elution lanes of the $(BoxB)_2$-L3 RNA as seen in lane E1 and E2. This result suggests that some of the RNA has been dissociated from the beads. The cleavage complex was formed only on the unbound RNA leading to cleavage of the RNA substrate in the solution.

To prevent further unspecific protein binding, the RNA-affinity matrix was additionally blocked with BSA in another experiment (Figure 2-9 A). For this experiment the same RNAs were used as before. As an additional control, $(Box)_2$-L3 RNA was incubated with GST protein.

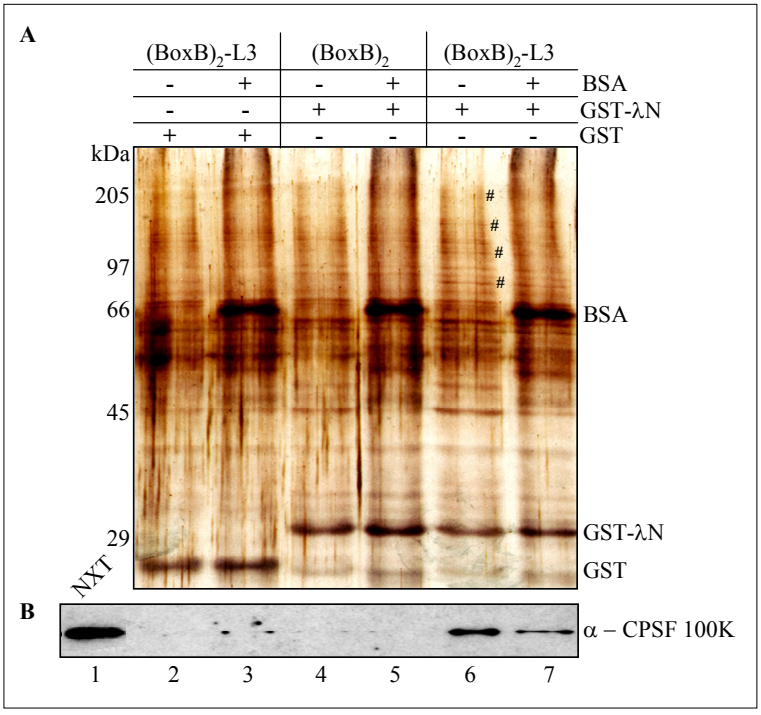


Figure 2-9 Different protein pattern in specific elution for $(BoxB)_2$-L3 RNA
λN peptide affinity experiment with an additional blocking step of the RNA-affinity matrix with BSA, $(BoxB)_2$ or $(BoxB)_2$-L3 RNA were used for immobilisation. Labelled RNA was immobilised via binding to the λN protein separately for each RNA substrate, afterwards the RNA-affinity matrix was splitted and one part of two was in addition blocked with 25 µg BSA for 1 h at 4 °C (lanes 3, 5 and 7). Subsequently, the reaction mixture containing NXT was applied for 20 min at 30 °C. The samples were washed three times in buffer D containing 0.1 % Triton X-100 and the RNA·protein·complexes were specifically eluted in elution buffer 3.
A) Silver-stained SDS polyacrylamide gel of λN peptide affinity specific elution fractions. $(BoxB)_2$-L3 (lane 2, 3, 6 and 7) and $(BoxB)_2$ (lane 4 and 5) were immobilised with GST-λN (lanes 4 to 7) or GST protein (lanes 2 and 3). # indicate protein bands, specific for $(BoxB)_2$-L3 RNA (lane 6), which were not present in the other samples.
B) Western blot analysis with CPSF-100K antibody of the specific elution fractions from λN peptide affinity. $^1/_{66}$ of NXT was used as a positive control for CPSF-100K (lane 1).

This RNA was not immobilised and all specifically eluted proteins were unspecifically bound to GST protein or the matrix. The samples were done in duplicate. One part of the two was in addition blocked by BSA after the RNA binding to prevent unspecific protein binding. Complex assembly was allowed for 20 min, which should be sufficient for complex formation. The RNA was specifically eluted by the elution buffer 3. Afterwards the beads were heated in SDS sample buffer for 5 min (at 95 °C) to remove all proteins still bound to the magnetic matrix.

We observed that in the specific elution fractions some additional bands appeared in the fraction containing the immobilised $(BoxB)_2$-L3 RNA. These protein bands were not present in the control lanes. Western blot analysis with CPSF-100K antibody showed a signal for both $(BoxB)_2$-L3 elution fractions, which did not occur in the other lanes (Figure 2-9 B).

$^1/_3$ of the elution fractions were separated in a SDS-PAGE next to 0.5 µL NXT ($^1/_{66}$ of NXT input). Only minimal amounts of CPSF were bound to the $(BoxB)_2$-L3 RNA. Due to the small amount of the eluted fraction, no further analysis of the additional protein bands was possible. In addition, this result could not be reproduced.

Additional variations of the incubation conditions are summarised in Table 2-6. Specific complex formation could not be observed again.

Table 2-6 Conditions tested with λN peptide affinity method

attempt	amount	comments
variation in binding order	-	parallel incubation of GST-λN protein and RNA resulted in highest specific RNA binding
incubation times	10'- 60'	dissociation of RNA from GST-λN protein increases with time
second blocking with BSA after RNA binding	5 µg BSA	less background, different protein bands, specific complex not reproducible
reduction of amount of matrix	0.1 and 1 µL packed beads	less background but unspecific protein binding to GST resp. GST-λN
elution at 4 °C		no difference in background
wash step with Triton X-100	0.1 % Triton X-100 in buffer D	no difference in background
NXT-B		no difference to NXT-A
10x more RNA	10 pmol substrate	no difference in background

2.1.4.4 λN peptide affinity with the new cleavage substrate PAPolA

Venkataraman and co-workers (2005) published that the UGUAN motif, which is present four times in the 3' UTR of the poly(A) polymerase A transcript (PAPolA), is bound by CF I_m. CF I_m interacts with PAP (Kim and Lee, 2001). The probably increased amount of bound CF I_m molecules to the PAPolA RNA and their interaction to PAP might stabilise the cleavage complex, formed on a RNA-affinity matrix. For that reason, the 3' UTR of PAPolA was cloned as described in Methods, Chapter 5.9.2.1. Afterwards the PAPolA transcript was used as a new substrate RNA for the λN peptide affinity method. The scheme of the construct with the known sequence motifs is shown in Figure 2-10.

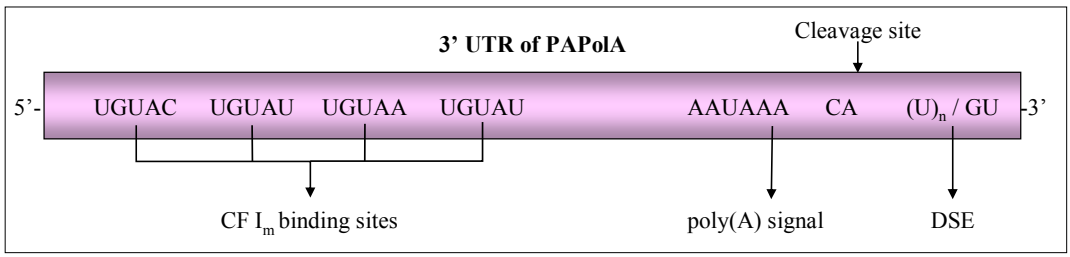


Figure 2-10 Structure of the cleavage substrate PAPolA
The 3' UTR of the poly(A) polymerase A contains four binding sites for CF I_m (Venkataraman *et al*., 2005). This is the UGUAN motif which is present twice in L3 RNA and only once in SV40 late.

To further stabilise the RNA binding to the GST-λN protein, an additional BoxB element was inserted into the plasmid coding for the cleavage substrate PAPolA. First it was confirmed that GST-λN binds to $(BoxB)_3$-PAPolA RNA in EMSA (Figure 2-11 A) (compare lane 1 and 2). Therefore $(BoxB)_3$-PAPolA was incubated with GST-λN protein. Lane 3 shows the assembly of the specific complex, which is not inhibited by additional binding of GST-λN protein (lane 4). The RNA was shifted mostly to the heterogeneous complex and a small amount to the specific complex. This complex assignment was described and analysed by Humphrey and co-workers (1987) (see Chapter 2.1.2). PAPolA RNA without BoxB element was not shifted with the GST-λN protein (compare lane 5 and 6). After incubation with NXT (lane 7) respectively additional GST-λN, the specific complex is formed on this RNA as well (lane 8), whereas the specific complex was not formed on the poly(A) signal mutant PAPolA RNA (data not shown). This result confirms that $(BoxB)_3$-PAPolA RNA was bound by GST-λN protein and the specific complex was formed.

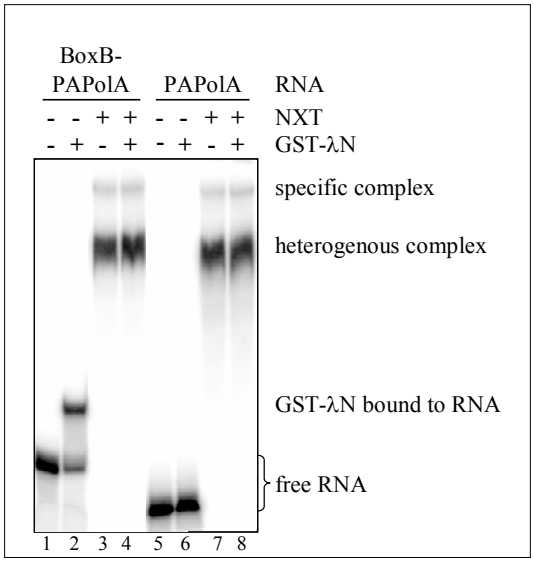


Figure 2-11 PAPolA RNA with three BoxB elements is specifically shifted by λN protein and forms the specific complex with NXT
EMSA using PAPolA and $(BoxB)_3$-PAPolA. Lanes 1 and 5 shows the free RNAs. These RNAs were incubated with GST-λN protein to observe the specific shift through the interaction of λN and BoxB (lane 2 and 6). Lanes 3 and 7 show the shifted RNAs after incubation with NXT and 4 and 8 with GST-λN and NXT. RNAs are indicated at the top, GST-λN and NXT addition are indicated by +, occurred bands are described at the right. Specific and heterogeneous complexes were defined in Humphrey *et al*., 1987. PAPolA Δ RNA showed no specific complex formation (data not shown).

These RNAs were cleaved in standard cleavage assays as expected (data not shown), whereas RNAs with the Δ mutation in their poly(A) signal were not cleaved. The addition of GST-λN protein in the cleavage assay had no influence on the cleavage activity (data not shown). These experiments indicate that the $(BoxB)_3$-PAPolA RNA can be used for λN peptide affinity purification.

The affinity purification experiments using the PAPolA RNAs showed only the elution of the GST-λN proteins and unspecific bound proteins. There was no difference between the $(BoxB)_3$-PAPolA compared to the control RNAs.
In another experiment, 10 times more RNA was used in an attempt to reduce the unspecific binding sites for proteins on the matrix. The $(BoxB)_3$-PAPolA was specifically eluted with reduced glutathione, but again contained only unspecifically bound proteins.
One reason for the failure of the RNA-affinity methods might be the instability of the complex. This was further analysed by shift assays.

2.1.5 Analysis of the pre-mRNA 3' end processing cleavage complex stability

Since it was not possible to purify the 3' end processing cleavage complex, the stability of the cleavage complex was suspected to be insufficient. In order to find out more about complex stability, the dissociation of the specific complex was analysed using the EMSA technique.

A radioactively labelled RNA substrate was used, NXT-A was added and the complex formation was allowed for 15 min at 30 °C. A 100 fold excess of non-labelled RNA substrate was added, which inhibited the re-association of all proteins dissociating from the specific complex to the labelled RNA substrate. Thus the intensity of the signal corresponding to the specific complex decreases in the autoradiography. Samples were taken over a period of two hours and frozen in liquid nitrogen to follow this process. The dissociation time course of the cleavage complex is shown in Figure 2-12. The protein·RNA complex was shifted into a specific and a heterogeneous complex compared to the unbound RNA (compare lanes 1 and 2). The band representing the specific complex disappeared. After 60 min the specific complex had completely dissociated. The heterogeneous complex was disassembled during that time also. The half-life of the specific complex was estimated to be 15 to 20 min from three different experiments.

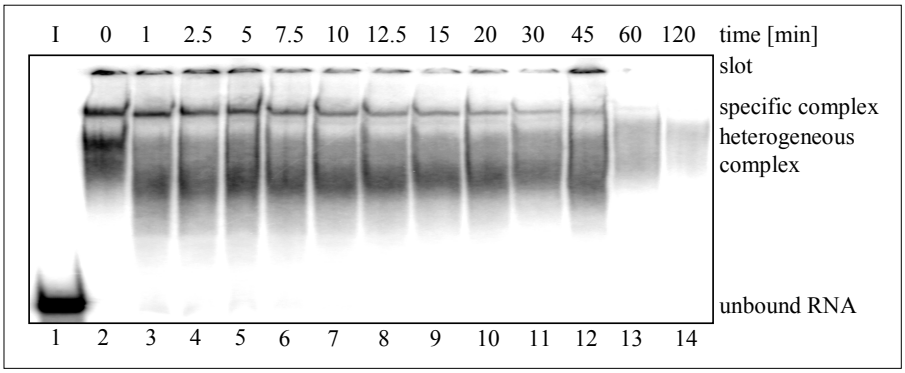


Figure 2-12 Analysis of the stability of the cleavage complex
Dissociation time course of cleavage complex formed on 100 fmol radioactively labelled RNA. Radioactively labelled RNA was pre-incubated with the cleavage reaction mixture containing NXT-A for 15 min at 30 °C. Then a 100 fold excess of non-labelled RNA was added and samples were taken at time points given and frozen in N_2. Lane 1 shows the radioactively labelled RNA after incubation in reaction mixture lacking NXT. The sample in lane 2 (time point 0) was taken prior to addition of non-labelled RNA and represents the maximal amount of specific complex which decreased during further incubation at 30 °C.

This result suggested that the specific complex is relatively unstable. Therefore a specific purification of the pre-mRNA 3' end cleavage complex may not be possible. Due to the negative results the work with the RNA affinity methods was abandoned. A new approach was chosen, to purify the proteins separately from human cell lines and to reconstitute the cleavage activity with these proteins.

2.2 Purification of tagged proteins for the reconstitution of the pre-mRNA 3' end processing cleavage complex

The reconstitution of the pre-mRNA 3' end processing reaction was achieved with recombinant and partially purified proteins (Rüegsegger *et al*., 1998). Recombinant PAP, recombinant CF I_m, CstF and CPSF purified from HeLa cells was used. CF II_m was only partially purified from HeLa cell nuclear extract (de Vries *et al*., 2000). It contains an essential protein complex CF II_m-A, consisting of hPcf11 and hClp1 and an unknown non-essential subcomplex, CF II_m-B, which enhances the cleavage activity. Mass spectrometry analysis of the final preparation, revealed more than 15 different proteins, including all CF I_m subunits (see Figure 2-17 B). Further reconstitution of the pre-mRNA 3' end processing reaction failed (Dettwiler, thesis, 2003).

CPSF was purified from HeLa cells or calf thymus in several steps (Bienroth *et al*., 1993). Several purifications of CPSF, purified by e.g. Kaufmann, Dettwiler, Kyburz and myself were active in polyadenylation but failed to reconstitute the cleavage activity using the same proteins as Rüegsegger respectively, failed to reconstitute CPSF depleted nuclear extract.

The development of methods to integrate DNA into the human cell genome allows the affinity purification of affinity-tagged proteins expressed in a human cell line. This approach was chosen to purify CF I_m, CstF, CF II_m and CPSF from HEK293 or HeLa cells with different affinity tags for reconstitution of the pre-mRNA 3' end cleavage reaction. Therefore, we cloned the open reading frames of all subunits of CF I_m and CstF in plasmids for expression of fusion proteins with an N-terminal TAP or his_8-flag-tag (HF) in human cell lines. These plasmids were stably integrated in the genome of HeLa or HEK293 cells. HEK293 cells stably expressing his_8-flag-tagged hClp1, a main subunit of CF II_m, were a kind gift of Trotta. Cells stably expressing CPSF HF-73K or 100K were obtained by Wlotzka during her diploma work.

2.2.1 Selection of cells stably expressing tagged proteins

2.2.1.1 Cloning of cDNAs encoding CF I_m and CstF subunits in plasmids for stable transfections

The DNA encoding the subunits of CF I_m and CstF was a gift from Dettwiler. PCR reactions were performed with 5' primers containing a Bam HI and 3' primers with Not I restrictions site. The PCR fragments were inserted in the Bam HI / Not I restricted pcDNA3.1 plasmid coding for a his_8-flag-tag at the N-terminus of the expressed protein (gift from Trotta). All subunits of CF I_m and CstF were also cloned into the pTRE2hyg TAP vector (gift from

Kühn). The expressed proteins contained an N-terminal TAP-tag. PCR reactions were performed with 5' primers with Nhe I and 3' primers with Not I restriction site. The purified plasmids from transformed *E. coli* were controlled by restriction and sequence analysis. Large amounts of endotoxin free plasmid DNA were purified using the Endo-free Maxi Kit® from Qiagen. All primers and constructed plasmids are summarised in Tables 6.7 to 6.9 in the Appendix.

2.2.1.2 Transfection of HEK293 and HeLa cells with CF I_m and CstF plasmids

Transfection tests showed that transfections using Lipofectamin 2000® resulted in high transformation efficiency of approximately 80-95 % for HEK293 cells, but not for HeLa cells. However, similar high efficiencies for HeLa cells could be achieved by using Lipofectin®.

Cells were seeded and transfected as described in the supplied protocols. Colonies formed from single transfected cells were isolated and grown adherently. Samples of these cells were analysed by western blot to control the expression of the tagged protein. For this we used antibodies directed against the flag or TAP-tag (protein A) and a protein specific antibody. Table 2-7 summarises the cell lines for CF I_m and CstF subunits with his_8-flag-tag or TAP-tag in HEK293 cells respectively HeLa cells.

An example of the western blot analysis is shown Figure 2-13 A. Cells expressing CF I_m HF-25K, HF-68K or HF-72K and CstF HF-50K were compared to non-transfected HEK293 cells. The detected bands showed the expected size for the proteins including 5 kDa for the HF-tag. Different expression levels of the proteins were due to the random integration of the plasmid into the genome of the cells and different numbers of integrated plasmids per cell.

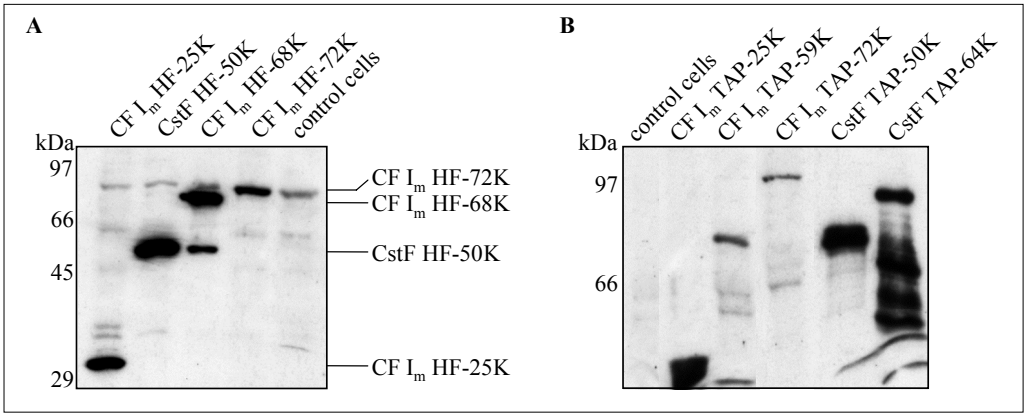


Figure 2-13 Cell lines stably expressing CF I_m and CstF subunits with his_8-flag-tag or TAP-tag
Western blot analysis, each lane contains total protein extracted from equal cell numbers from HEK293 cell lines stably expressing the HF-tagged protein as indicated on the top of the gel. Control cells are non-transfected HEK293 cells. The molecular weight marker is indicated at the left.
A) Antibody directed against the flag peptide
B) Antibody directed against protein A

Table 2-7 Stable cell lines expression CF I_m and CstF subunits with his_8-flag-tag or TAP-tag in HEK293 cells respectively HeLa cells

protein complex	subunit	cell line	tag	tested / positive clones
CF I_m	25K	HEK293	HF	21 / 19
		HEK293	TAP	29 / 23
		HeLa	TAP	9 / 4
	59K	HEK293	TAP	11 / 8
		HeLa	TAP	9 / 6
	68K	HEK293	HF	26 / 17
		HEK293	TAP	3 / 2
		HeLa	TAP	6 / 1
	72K	HEK293	HF	20 / 3
		HEK293	TAP	4 / 4
		HeLa	TAP	10 / 2
CstF	50K	HEK293	HF	25 / 11
		HEK293	TAP	13 / 1
		HeLa	TAP	6 / 2
	64K	HEK293	HF	20 / 6
		HEK293	TAP	9 / 5
		HeLa	TAP	7 / 3

HF abbreviation for his_8-flag-tag

2.2.2 Purification of proteins

2.2.2.1 Purification of TAP-tagged proteins was not successful

Cell lines were obtained for all CF I_m subunits and CstF-50K and -64K in HEK293 (Figure 2-13 B) and HeLa cells. Western blot analysis showed the proteins from lysed cells at expected sizes, subunit size plus 20 kDa for the TAP-tag.

HEK293 cells expressing CF I_m TAP-25K and CF I_m TAP-59K were grown adherently. Nuclear extracts were prepared from 7 mL packed cell pellet volume as described in Wahle & Keller (1994). After dialysis, the protein A purification was performed using IgG Sepharose. The proteins were bound for two hours at 4 °C. The matrix was washed five times and equilibrated in TEV buffer. The proteins were eluted with TEV protease at 16 °C for 1.5 h. The purifications were analysed by a silver-stained SDS polyacrylamide gel. Western blot analysis of selected fractions and a sample IgG beads treated with SDS sample buffer showed

that the TEV treatment did not lead to protein elution (data not shown). The proteins could be removed by the SDS sample buffer treatment only, which resulted in denatured proteins. Even after over night elution with TEV protease, no protein could be detected in the elution fractions. The TEV protease was active and successfully used in the purification of TAP-tagged proteins from yeast (Paquet, pers. communication).

Tests with a different batch of purified and active TEV protease (purified by Blank) revealed no elution of the proteins either. Ringel prepared TAP-tagged proteins from the same vector and obtained no elution either (pers. communication). It is possible that the TEV cleavage site is not accessible to the protease due to steric hindrances. Therefore we decided to focus on HF-tagged proteins for further experiments.

2.2.2.2 Purification of his_8-flag-tagged proteins

HEK293 cells stably expressing tagged CF I_m, CstF and CF II_m subunits were grown adherently. A packed cell volume of around 30-40 mL was used for purification as described in Paushkin and colleagues (2004) with minor changes (see Methods 5.11.2.3).

2.2.2.2.1 All CF I_m subunits co-purify with CF I_m-25K

CF I_m-25K binds to all large subunits of CF I_m (Rüegsegger *et al*., 1998; Dettwiler *et al*., 2004). Thus, the expectation was the co-purification of CF I_m-59K, CF I_m-68K and CF I_m-72K with CF I_m HF-25K. The first purification of CF I_m-25K was carried out following the protocol from Paushkin and colleagues (2004) using a flag and a Ni-NTA matrix.

The result of this purification is illustrated in Figure 2-14 A. The purified protein was analysed by mass spectrometry (Friedlein, Roche Inc., Basel): The HF-tagged CF I_m-25K subunit was identified, as well as the two large subunits of CF I_m-68K and -59K. A strong band appeared at a size of around 25K, which was identified as the light chain of IgG1 from mouse. This contamination occurred probably during the flag purification. Western blot analysis using antibodies direct against flag and CF I_m-25K, indicated that additionally endogenous CF I_m-25K was co-purified (Figure 2-14 B, lower panel): Two bands occurred in the western blot, one corresponding to the CF I_m HF-25K containing the 5 kDa his_8-flag-tag and the other corresponding to the endogenous 25K subunit. This confirmed the self-interaction of CF I_m-25K which was previously shown in GST pull-down assays (Dettwiler, thesis, 2003).

The antibody directed against CF I_m-68K recognises all three large subunits. They are visible in the last lane of the western blot (Figure 2-14 B, upper panel) as indicated on the right. The

activity of the protein was tested using antibody-depleted NXT-B (NXT-preparation following the protocol of Rüegsegger, see Methods 5.10.2) which lacks CF I_m. The depletion of the proteins is described in Methods (Chapter 5.11.2.5.1). To test the activity of the preparation, the activity of NXT-B was compared to the activity of the antibody-depleted nuclear extract reconstituted with the purified CF I_m, see Figure 2-14 C. Lane 1 demonstrated the activity of NXT-B prior to protein depletion. Controls with mock depletion (lanes 2) and a pre-serum depletion (lane 3) indicated that there was a decrease in the activity due to unspecific binding of proteins like CPSF to the Sepharose matrix, as shown before in Chapter 2.1.3.1.

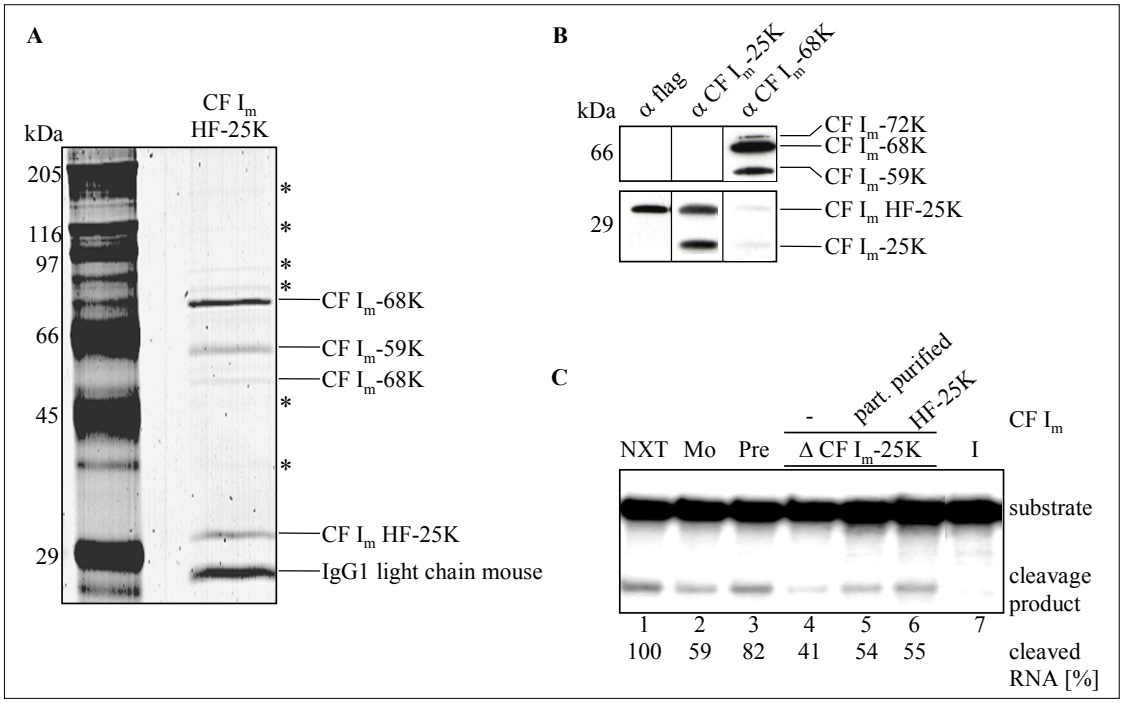


Figure 2-14 Endogenous CF I_m-25K and all large CF I_m subunits co-purifies with CF I_m HF-25K and show activity in reconstitution cleavage assay.
A) Polypeptide composition of CF I_m complex that was purified via the HF-tagged CF I_m-25K. The complex was purified via a flag-matrix followed by Ni-NTA purification. The protein bands were cut from the SDS polyacrylamide gel and identified by mass spectrometry, as indicated on the right. The molecular weight marker is indicated at the left. * indicates unidentified protein bands. Mass spectrometry analysis was done by Friedlein (Roche Inc.).
B) Western blot analysis of purified CF I_m-25K. The same lane was probed with antibodies directed against flag peptide and CF I_m subunits, as indicated at the top. The antibody directed against CF I_m-68K also recognises 59K and 72K. The protein standard in kDa is indicated at the left.
C) Activity test using depleted nuclear extract. NXT was depleted of CF I_m with an antibody directed against the 25K subunit of CF I_m. Mock depleted extract was incubated on protein A Sepharose without antibody. As an additional control, the nuclear extract was incubated with pre-serum of CF I_m-25K bound to the matrix. Lane 1 shows the maximal activity present in NXT, lane 7 the input RNA (I). In both controls (Mo = mock depleted NXT, lane 2 and Pre = pre-serum depleted NXT, lane 3) the activity is present, whereas it is partially reduced in NXT depleted with the CF I_m-25K antibody (lane 4). When partially purified CF I_m (from Dettwiler) is added to the depleted NXT (lane 5), the amount of cleavage product is slightly increased compared to the depleted NXT, showing that the loss of activity is due to specific removal of CF I_m from the NXT. Addition of approximately 50 ng affinity purified CF I_m HF-25K, containing the three large subunits, to the depleted NXT could also partially reconstitute the activity, showing that the preparation of CF I_m from HEK293 cells is active *in vitro*. CF I_m HF-25K on its own had no cleavage activity (data not shown).

The activity was reduced to 41 % in the CF I_m depleted NXT-B (lane 4) compared to NXT-B. Partially purified CF I_m was added as a positive control (lane 5) to show that the loss of cleavage activity in the depleted extract is due to the specific removal of CF I_m. The addition of approximately 50 ng of affinity purified CF I_m to the depleted NXT-B restored the cleavage activity to 55 % (lanes 5 and 6). In conclusion the affinity purified CF I_m is active in *in vitro* cleavage assays.

The second purification of tagged CF I_m-25K contained the proteins MEP50 and PRMT5, which are known contaminants of the flag purification (Chendrimada *et al*., 2005). This purification was active, too and was used for our total reconstitution assay as shown later in this thesis.

2.2.2.2.2 All CF I_m subunits co-purify with CF I_m-68K

CF I_m is thought to be a hetero-dimer, consisting of the 25K subunit and one of the three large subunits (Rüegsegger *et al*., 1996, Dettwiler *et al*., 2004). To confirm this data we purified CF I_m via the HF-tagged 68K subunit and expected to co-purify the 25K subunit and the 72K subunit, which is a splice variant of 68K.

However, mass spectrometry analysis identified not only the subunits HF-68K, 25K and 72K, but also the other large subunit 59K in the preparation (Figure 2-15 A). Together with the tagged version, the endogenous CF I_m-68K was co-purified as well. Three bands appeared in the western blot using the antibody directed against CF I_m-68K. The antibody recognised the CF I_m HF-68K together with CF I_m-72K, the endogenous CF I_m-68K and CF I_m-59K (Figure 2-15 B). We expected the 72K subunit to migrate, at the same position in the gel as the tagged 68K, containing the HF-tag, which leads to an increased size of around 5 kDa. The autoradiography of the activity test, using depleted nuclear extract for this CF I_m preparation is shown in Figure 2-15 C. The antibody-depletion was not complete, although some cleavage activity was still present in the activity test using CF I_m-68K-depleted NXT-B (lane 4). The amounts of cleavage products were calculated as relative amounts of the intensity of the cleavage product in the NXT-B (see Methods). The mock and the pre-serum depletions showed a loss of activity of more than 50 %, which were due to the eight hours depletion at 4 °C and the unspecific binding of CPSF to the Sepharose beads as shown before. The addition of approximately 60 ng of purified CF I_m allowed the reconstitution of the cleavage activity to 28 %, compared to the control depleted NXT-B alone, with only 7 % cleavage product.

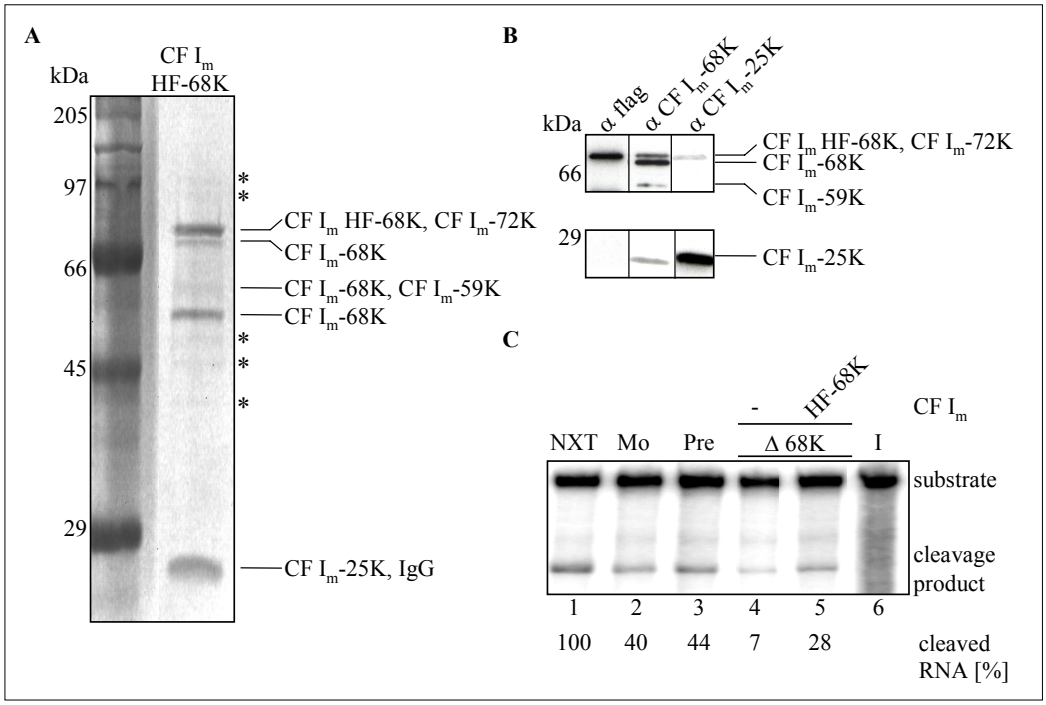


Figure 2-15 The endogenous CF I_m-68K together with CF I_m-25K and the large CF I_m subunits 59K and 72K co-purify with CF I_m HF-68K.
A) Polypeptide composition of CF I_m complex that was co-purified with his_8-flag-tagged CF I_m-68K (expression clone E2). The complex was purified over a flag matrix. The protein bands were cut from the SDS polyacrylamide gel and identified by mass spectrometry as indicated on the right. The molecular weight marker is indicated at the left. * indicate unidentified protein bands. The mass spectrometry analysis was performed by Jenö (Biozentrum Basel).
B) Western blot analysis of CF I_m HF-68K. The same lane was probed with antibodies directed against the flag peptide and CF I_m subunits as indicated on the top. The antibody directed against CF I_m-68K recognises also -59K and -72K. The migration of the molecular weight marker is indicated at the left.
C) Activity test using depleted nuclear extract. NXT was depleted of CF I_m with an antibody directed against the 68K subunit of CF I_m. Mock depleted extract was incubated on protein A Sepharose without antibody. As an additional control, the nuclear extract was incubated with pre-serum of CF I_m-25K bound to the matrix, the pre-serum of CF I_m-68K was not available. Lane 1 shows the maximal activity present in NXT, lane 6 the input RNA (I). In both controls (Mo = mock depleted NXT, lane 2 and Pre = pre-serum depleted NXT, lane 3) the activity is present, whereas it is reduced in NXT depleted with the CF I_m-68K antibody (lane 4). Addition of affinity purified CF I_m-68K, containing all subunits (CF I_m-25K, CF I_m-59K, CF I_m-68K and CF I_m-72K), to the depleted NXT can also reconstitute the activity, showing that the preparation of CF I_m from HEK293 cells is active *in vitro*. CF I_m HF-68K on its own had no cleavage activity (data not shown).

The difference in the cleavage activity between the controls with mock and pre-serum depleted NXT-B and the reconstitution (lane 5) may be due to the fact, that the concentration of other proteins e.g. CPSF or PAP was reduced during their interaction with the depleted CF I_m (personal communication Kyburz).

Since the random integration of the DNA, in the genome of the HEK293, might influence the CF I_m expression and composition, the CF I_m-68K purification was repeated using another clone. This CF I_m preparation was also active in the activity test using depleted NXT. This preparation contained more of the unspecifically bound proteins like MEP50 and PRMT5.

Two other proteins were identified by mass spectrometry as ser / thr kinase 38 and as prolyl 4-hydroxylase. Both were not identified in the first preparation. The sizes of the ser / thr kinase 38 (STK38) (around 55 kDa) and of the prolyl 4-hydoxylase (around 100 and 42 kDa) revealed that these proteins were probably the unidentified proteins in the first preparation, indicated by an asterisk (*). STK38 is a negative regulator of the MAPKKKs MEKK1 and MEKK2 (Enomoto *et al*., 2007), whereas prolyl 4-hydoxylase catalyses the formation of 4-hydroxyproline in collagen (Helaakoski *et al*., 1989). There are no hints in the literature that these proteins might have any functions in 3' end processing and they were not further investigated in this work.

2.2.2.2.3 CF I_m-68K is methylated by PRMT5

CF I_m was also purified from a cell line expressing the HF-tagged 72K subunit. The protein was expressed only weakly and degraded to a size of around 43 kDa during the purification. Again, the mass spectrometry data showed co-purification of all large subunits and the 25K subunit. In addition pICln was identified. Together with PRMT5 and MEP50, pICln forms a complex, the methylosome. This protein complex is known to catalyse the symmetric dimethylation of the Sm proteins D1 and D3 of the U snRNPs and several other proteins (Meister *et al*., 2001). The identification of the third methylosome subunit, pICln, was the reason for further analysis of the methylation of CF I_m.

We considered the possibility that the methylosome binds specifically to CF I_m and scanned for methylation motifs in the sequence of the CF I_m subunits. A GGRGRGR motif is present in the CF I_m-68K sequence. This motif might be a methylation site of the methylosome. In order to test this hypothesis, an *in vitro* methylation assay was performed. ^{14}C-labelled S-adenosyl methionine was added to all affinity purifications of CF I_m and CstF, and incubated for two hours at 30 °C. The proteins were separated by SDS-PAGE. The autoradiography of the SDS polyacrylamide gel, showed that bands occurred only in HF-tagged CF I_m preparations, which were affinity purified only via the flag tag (data not shown). These bands were not present in HF-tagged CstF preparations and in HF-tagged CF I_m preparations, additionally purified via Ni-NTA Sepharose. Reasons therefore are, that CstF probably is no substrate for methylosome / PRMT5 methylation and that MEP50 and PRMT5 are not further present in CF I_m preparations after the Ni-NTA purification.

A second *in vitro* methylation assay was performed using purified fragments of CF I_m-68K, recombinantly expressed in *E. coli* (purified by Dettwiler), containing the postulated methylation motif, and the HF-tagged CF I_m preparations containing PRMT5. The *E. coli*

CF I_m-68K fragments received a radioactive label, which was transferred from C^{14}-labelled S-adenosyl methionine to the protein. Therefore we conclude that the CF I_m was methylated *in vitro* (data not shown).

To confirm that CF I_m-68K is indeed methylated by PRMT5, Martin purified the CF I_m methylation activity from HeLa cell extract and could identify PRMT5 and MEP50 by mass spectrometry analysis. However, whether the observed methylation of CF I_m has any biological significance, is not known up to now.

2.2.2.2.4 CPSF subunits co-purify with CstF-64K

CstF consists of three subunits, CstF-50K, CstF-64K and CstF-77K (Gilmartin & Nevins, 1991; Takagaki *et al*., 1990). CstF-64K is known to bind to the downstream element of the pre-mRNA (Takagaki *et al*., 1992). CPSF-160K was previously shown to interact with CstF-77K by co-immunoprecipitation (Murthy *et al*., 1995). Pull-down assays using hFip1 revealed its interaction with CstF-77K (Kaufmann *et al*., 2004). Takagaki and colleagues demonstrated (2000) via co-immunoprecipitation, that symplekin interacts with CstF-64K. CstF was affinity purified from HEK293 cells stably expressing CstF HF-64K. It was purified via the flag matrix and analysed by mass spectrometry. We identified the three CstF subunits (Figure 2-16 A). This analysis revealed that additional proteins of the pre-mRNA 3' end processing complex were co-purified. The affinity purified tagged 64K subunit contained CPSF-160K, hFip and symplekin. Furthermore, the following proteins were observed: CPSF-100K, CF I_m-68K and CstF-64K τ, a variant of CstF-64K, which is encoded by a paralogous gene. MEP50, PRMT5 and IgG were contaminating proteins from the flag matrix.

Western blot analysis of the subunits of CstF is shown in Figure 2-16 B. The CstF HF-64K was detected by an antibody against the flag-tag and additionally by a specific CstF-64K antibody. The membrane was furthermore incubated with CstF-77K antibody before stripping, thus signals for both antibodies were visible. The CstF-50K antibody hardly recognised the 50K subunit, even in a dilution of 1:100.

Figure 2-16 C shows the cleavage activity test using NXT-B that was depleted for CstF-77K. Depletions with CstF-77K or -64K antibody were incomplete, resulting in a remaining cleavage activity of approximately 20 %, compared to the non-depleted NXT-B (lane 4 and lane 1). The mock and the pre-serum depletion showed a loss of activity to 40 or 44 % respectively. The addition of purified CstF to the depleted extract allowed the reconstitution of the activity to the value observed for mock depleted NXT-B. Another CstF preparation from HeLa cells, purified by Dettwiler, was used as a positive control.

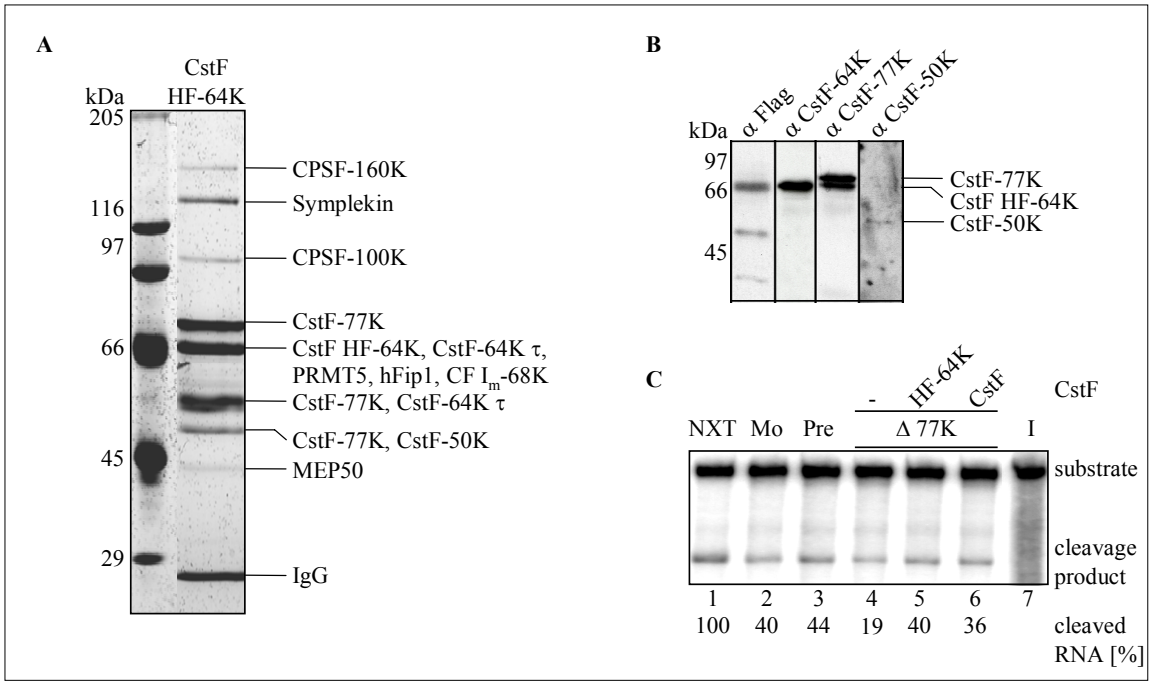


Figure 2-16 Co-elution of CPSF-100K, -160K, hFip1, CF I_m-68K and symplekin with CstF HF-64K.
A) Polypeptide composition of CstF which was purified via affinity tagged CstF-64K. The complex was purified only via the flag matrix. The protein bands were cut from the SDS polyacrylamide gel and identified by mass spectrometry as indicated on the right. The molecular weight marker is indicated at the left. * indicate unidentified protein bands. The mass spectrometry analysis was done by Jenö (Biozentrum Basel).
B) Western blot analysis of CstF HF-64K. The same lane was probed with antibodies directed against the flag peptide and CstF subunits as indicated on the top. The flag antibody recognised two smaller proteins. The protein marker weight is indicated at the left.
C) Activity test using antibody-depleted nuclear extract. NXT was depleted of CstF with an antibody directed against the 77K subunit. Mock-depleted extract was incubated on protein A Sepharose without antibody. As an additional control, the nuclear extract was incubated with pre-serum of CF I_m-25K bound to the matrix; pre-serum of CstF-77K was not available. Lane 1 shows the maximal activity present in NXT, lane 7 the input RNA (I). In both controls (Mo = mock-depleted NXT, lane 2 and Pre = pre-serum-depleted NXT, lane 3) the activity is present, whereas it is reduced in NXT depleted with the CstF-77K antibody (lane 4). Addition of affinity-purified CstF-64K, containing the three subunits of CstF, CstF-50K, CstF-64K and CstF-77K, to the depleted NXT can also reconstitute the activity, showing that the preparation of CstF from HEK293 cells is active *in vitro*. Purified CstF from HeLa cells is the positive control for the reconstitution (lane 6). Purified HF-tagged CstF on its own had no cleavage activity (data not shown).

Addition of this preparation to the depleted NXT-B, restored the cleavage activity to the same extent that was observed without depletion. CstF HF-64K on its own had no cleavage activity (data not shown).

Two preparations via the tagged CstF-50K subunit co-purified only the three CstF subunits 77K, 64K and HF-50K; of these CstF-64K was present only in small amounts. This could be due to the fact that CstF-50K and 64K do not interact directly. The cleavage activity of HF-tagged CstF-50K was tested using NXT-B depleted for CstF, but showed less reconstitution activity than CstF HF-64K (data not shown). Therefore the CstF HF-64K preparation was used for the total reconstitution assays of the pre-mRNA 3' end processing complex.

2.2.2.2.5 Affinity purification of CF II_m is very efficient compared to column preparations

First purifications of CF II_m were done by de Vries and colleagues (2000). The authors purified the complex from HeLa extracts using seven chromatography steps. The resulting purification is shown in Figure 2-17 B. The preparation contained the two known subunits hPcf11 and hClp1, and additional other proteins like all CF I_m subunits, transcriptions factors TFIIH p52 and p89, and splicing proteins U2AF-35 and -65.

In this work, the CF II_m complex was purified from HEK293 cells via the his_8-flag tagged hClp1 subunit, in a two step procedure. Firstly, CF II_m was purified on a flag matrix. The flag elution fractions were additionally applied to Ni-NTA Sepharose. Thereby, the known contaminants, PRMT5 and MEP50 (Chendrimada *et al*., 2005), were removed. The elutions from the first purification via the flag-tag and from the second purification via Ni-NTA matrix are shown in Figure 2-17 A. They contained fewer proteins compared to the earlier column preparation from de Vries *et al.* (2000). Mass spectrometry analysis of the proteins eluted from Ni-NTA matrix identified, apart from hPcf11 and hClp1, the tRNA intron endonuclease (tRIE) and proteins Sen2 and Sen54. The Sen proteins form together with tRIE, also known as Sen34, and hClp1, the tRNA splicing endonuclease complex. This complex catalyses the removal of introns from pre-tRNA (Paushkin *et al*., 2004).

Western blot analysis of the proteins eluted from the flag matrix and Ni-NTA matrix are shown in Figure 2-17 C. The hPcf11 antibody recognised proteins in a range of 140 to 180 kDa.

In Figure 2-17 D the activity test with depleted NXT-B is shown. An antibody against hPcf11 was used for the depletion. CF II_m on its own had no cleavage activity (lane 2). Lane 3 contained NXT-B as the positive control for the experiment. The amount of cleavage product from this lane was set at 100 %. The mock and the pre-serum depletion showed a loss of activity to 27 or 28 % respectively (lane 4 and 5). The depleted NXT-B showed no activity. The cleavage product appeared again after the addition of affinity purified CF II_m (lane 6). Purified CF II_m (de Vries) was used as a positive control for reconstitution of the hPcf11 depleted NXT-B (lane 7). We demonstrated that the affinity purified CF II_m reconstituted the cleavage activity of hPcf11-depleted nuclear extract.

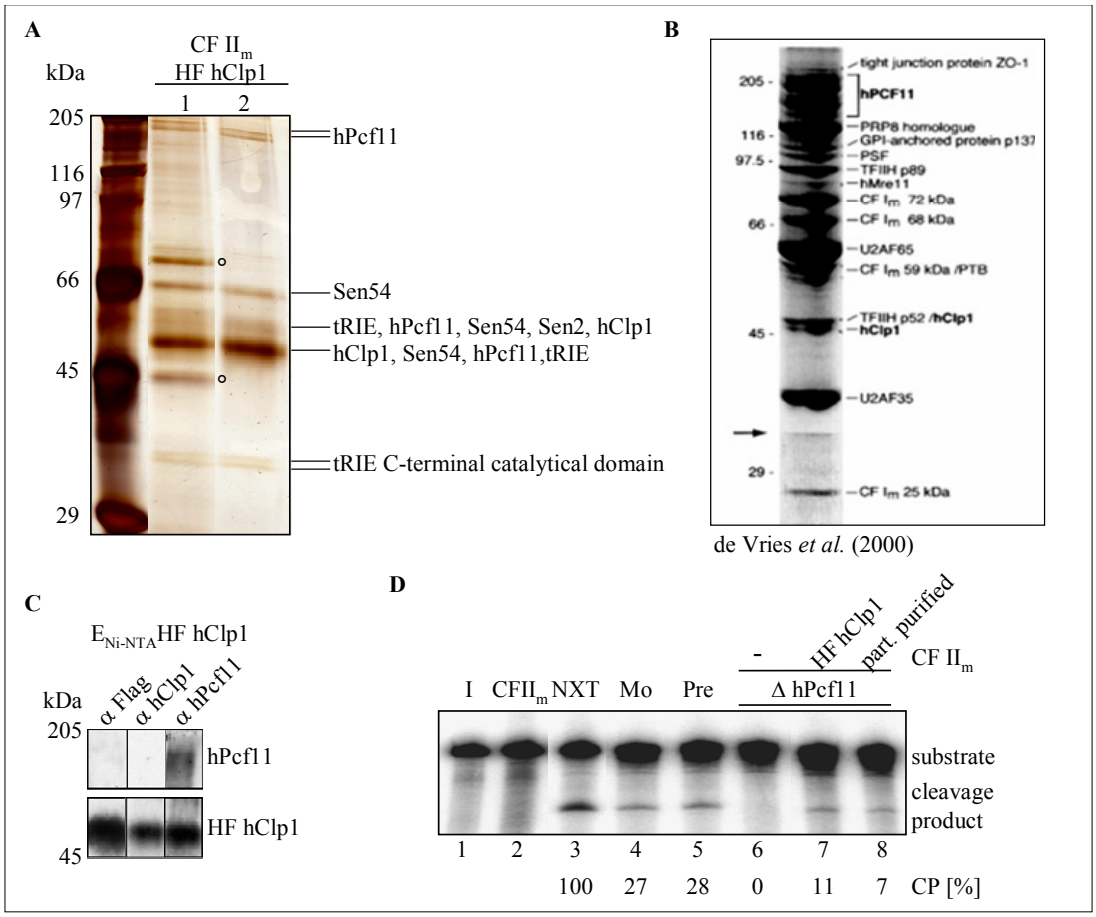


Figure 2-17 Purification of reconstitution active CF II_m with his_8-flag tagged hClp1 is efficient compared to column preparation

A) Polypeptide composition of CF II_m that was purified via his_8-flag tagged hClp1. The complex was first purified via flag matrix (1) and in addition via Ni-NTA Sepharose (2). The protein bands were cut from the SDS polyacrylamide gel and identified by mass spectrometry as indicated on the right. The molecular weight marker is indicated at the left. ° indicates the unspecific flag matrix bound proteins PRMT5 and MEP50 that are known to bind unspecifically to the flag matrix (Chendrimada *et al*., 2005). The mass spectrometry analysis was performed by Jenö (Biozentrum Basel). tRIE = tRNA Intron Endonuclease

B) Purification of CF II_m from de Vries after 7 columns. Figure taken from the publication of de Vries *et al*., 2000.

C) Western blot analysis of polypeptide composition of CF II_m from Ni-NTA elution as indicated at the top. The same lane was probed with antibodies directed against the flag peptide and CF II_m subunits as indicated on the top. The molecular weight marker is indicated at the left.

D) Activity test using depleted nuclear extract. NXT was depleted of CF II_m with an antibody directed against hPcf11. Mock-depleted extract was incubated on protein A Sepharose without antibody. As an additional control, the nuclear extract was incubated with pre-serum of CF II_m-hPcf11 bound to the matrix. Lane 1 shows the input RNA (I), lane 2 that affinity purified CF II_m had no cleavage activity itself. Lane 3 shows the maximal activity present in NXT. In the controls (Mo = mock-depleted NXT, lane 4 and Pre = pre-serum-depleted NXT, lane 5) the activity is present, whereas it is reduced in NXT depleted with the CF II_m hPcf11 antibody (lane 6). When purified CF II_m (de Vries, fraction 54-1) is added to the depleted NXT, cleavage activity can again be observed (lane 8), showing that the loss of activity is due to specific removal of CF II_m from the NXT. Addition of affinity purified CF II_m-hClp1 to the depleted NXT can also reconstitute the activity, showing that the preparation of CF II_m from HEK293 cells is active *in vitro*.

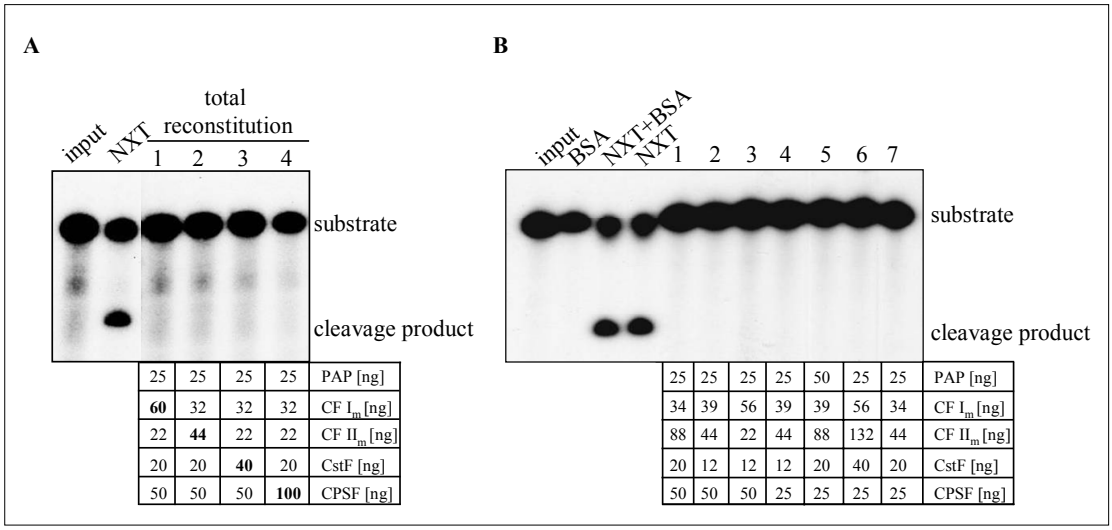
2.2.2.3 Total reconstitution of the pre-mRNA 3' end processing reaction with purified proteins failed

The term total reconstitution means the assembly of the functional pre-mRNA 3' end processing complex from well defined, purified protein components. This would reveal the complete protein composition of the complex and allow the further characterisation of the functions of the components by using variants of the subunits.

After the purification of active CF I_m, CF II_m and CstF, the missing components were PAP and CPSF. Whereas bovine PAP was purified from *E. coli*, CPSF was affinity purified in the same way as the other protein components expressed in HEK293 cells by Wlotzka. All purified proteins from HEK293 cells are summarised in Table 6.10 in the Appendix. CPSF showed only weak activity in cleavage assays using depleted NXT-B and only with the addition of CstF and PAP (Wlotzka, diploma thesis, 2006). These factors were probably depleted as well through their interaction with the depleted CPSF (pers. communication Kyburz). Nevertheless, it was tried to reconstitute the cleavage reaction with the affinity-purified protein complexes.

Different amounts of purified factors were used (see Figure 2-18 A). These amounts were determined in the activity assay using depleted nuclear extract and were sufficient for the reconstitution of the cleavage activity in depleted NXT-B. CPSF was purified via HF-tagged CstF-73K and tested for its cleavage activity in depleted NXT by Wlotzka (diploma thesis, 2006). In these assays CPSF hardly reconstituted the cleavage activity, despite addition of PAP and CstF, which are known to be co-depleted with the CPSF antibodies. For this reason higher amounts were used in the total reconstitution assays. Lane 2 shows NXT-B as a positive control. The expected cleavage product was observed. In four different experiments the amount of proteins was varied (see Figure 2-18 A) but no cleavage activity could be detected.

Another experiment is shown in Figure 2-18 B. Lanes 1 to 7 indicated the trials for the total reconstitution. Lane 7 contains additional BSA. BSA was added to stabilise the proteins. BSA had no cleavage activity as expected (lane 2) and did not show any influence on the cleavage activity of the NXT-B, comparing lane 3 and 4. As before, different amounts of the proteins were tested (see Table below), but no cleavage product was observed.



25	25	25	25	PAP [ng]
60	32	32	32	CF I_m [ng]
22	**44**	22	22	CF II_m [ng]
20	20	**40**	20	CstF [ng]
50	50	50	**100**	CPSF [ng]

25	25	25	25	50	25	25	PAP [ng]
34	39	56	39	39	56	34	CF I_m [ng]
88	44	22	44	88	132	44	CF II_m [ng]
20	12	12	12	20	40	20	CstF [ng]
50	50	50	25	25	25	25	CPSF [ng]

Figure 2-18 No total reconstitution with purified proteins.
Reconstitution assays with different amounts of purified proteins. The amount of protein, which was necessary to reconstitute the cleavage activity in the depleted NXT, was determined and used for the total reconstitutions assays. Approximately protein concentrations were determined by Bradford assay.
A) Total reconstitution using the determined protein amount or nearly the double amount (**bold**). NXT was used as control to show that the cleavage is possible in the used buffer conditions. Table at the bottom shows the proteins composition. Substrate and cleavage product are indicated at the right side.
B) Total reconstitution assay using additional BSA for protein stabilisation. BSA alone shows no cleavage activity and had no influence in the cleavage activity of the NXT. NXT lane serves as control to show that the condition works. Table at the bottom shows the proteins composition. Lane 7 contains additional 100 ng BSA. Reactions are indicated at the top, substrate and cleavage product at the right side.

All other attempts are summarised in Table 2-8. We used different cleavage substrates (L3, SV40 late and PAPolA), added other purified proteins like hFip1, CTD or unspecifically phosphorylated CTD but could never detect any activity. The amount of creatine phosphate was increased, because it was previously shown to function as a cofactor (Hirose & Manley, 1997) and to mimic a phospho-CTD (Ryan, 2007). In the control reaction, NXT-B showed a higher activity with the 2.5 fold higher creatine phosphate concentration. However, the addition of 2.5 times more creatine phosphate to the mixture of proteins, did not result in cleavage activity.

Rüegsegger described in her thesis (1997) that it might help to add the proteins sequentially to the reaction mix, as they bind to the RNA substrate. Several experiments were performed with sequential pre-incubation of the proteins, but a cleavage product could not be observed. Assuming that the proteins might need more time to assemble, they were incubated on ice over night before the two hours incubation at 30 °C. The analysis of the substrate RNA showed that no cleavage activity could be reconstituted (data not shown).

Table 2-8 Summary of the tested conditions for total reconstitution

condition
variation in the amounts of all proteins
variation of substrates, using L3, SV40 late and PAPolA
addition of CTD from mouse (expressed in *E. coli*)
addition of hFip1 (purified by Kaufmann)
increase in creatine phosphate up to 50 mM (2.5 fold)
sequential addition of proteins with additional pre-incubation at 30 °C
over night pre-incubation for proteins at 4 °C
addition of diluted NXT-B

One possible explanation for the inactivity of the protein mixture, could be that there is at least one factor missing, or one of them is inactive in the absence of extract components. The influence of small amounts of extract on the total reconstitution was examined (Figure 2-19). The activity of the cleavage reaction was only slightly increased (see lanes 10 to 12) compared to the diluted extract (lanes 4 to 6). The addition of 1 µL of NXT-B showed no further stimulation (comparing lane 7 and 13).

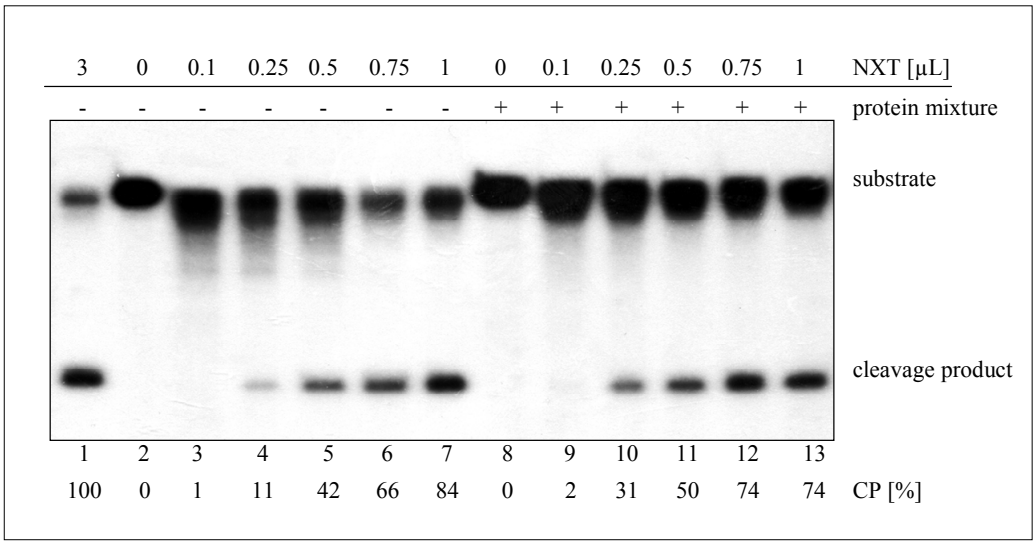


Figure 2-19 Weak stimulation of reconstitution of pre-mRNA 3' end processing activity with addition of NXT

Reconstitution cleavage assay was performed using a mixture of affinity purified proteins and PAP (lanes 8 to 13). Standard cleavage assay were performed with increasing amounts of NXT (lanes 2 to 7). The volume of used NXT is indicated at the top of the autoradiography. 3 µL of NXT were used as positive control. For better handling was the NXT diluted 1:10 in buffer D. The amount of cleavage product was determined for comparison of the experiments using NXT + protein mixture and solely NXT. This is indicated at the bottom.

The potential factor was attempted to purify by fractionating HeLa cell extract over a DEAE column. Several elution fractions, that were inactive alone, showed a significant cleavage activity, when mixed with the purified proteins in the total reconstitution assay. These fractions consisted of at least CPSF, CF I_m and CF II_m and small amounts of CstF, as identified in western blot analysis. They were pooled and applied to a Blue Sepharose column for further separation.

However, the dialysed elution fractions of the Blue Sepharose could not further reconstitute the cleavage activity in the activity assay (data not shown). Taken together, in our hands it was not possible to reconstitute the cleavage activity from purified proteins. It seems that at least one protein component is missing, inactive or present in too small amounts.

3. Discussion

The pre-mRNA 3' end processing reaction has been studied for more than 20 years. The well known factors CstF, CF I_m and PAP, assemble together with the less characterised factors CPSF and CF II_m onto the pre-mRNA, forming a cleavage complex. *In vitro* reconstitution of the 3' end processing complex with the known, purified proteins failed, so far. Comparison of the protein complexes catalysing the cleavage reaction, in yeast and humans, showed that there are still yeast proteins which have no identified human counterpart (see Table 6-11B, page 92). A possible reason for the failure so far of the reconstitution with mammalian proteins could be that there is still at least one missing component.

The aim of this thesis was the affinity purification and analysis of the mammalian pre-mRNA 3' end processing complex. This would allow the defining of all factors necessary for the pre-mRNA 3' end processing cleavage reaction. The functional reconstitution of the reaction with purified proteins would allow the testing of mutant subunits and reveal their function e.g. to prove the suggestion that CPSF-73K is the endonuclease responsible for the cleavage reaction.

3.1. Optimisation of the cleavage reaction conditions

The cleavage product obtained in an *in vitro* reaction would represent the assembly of a functional cleavage complex. Therefore, the cleavage reaction was optimised for the amount of cleavage product. The efficiency of the cleavage reaction strongly depends on the extract preparations, which were optimised for maximal activity. Published cleavage protocols varied in the buffer composition used during extract preparation (Rüegsegger *et al*., 1996; Manley, 1983). These extracts were prepared according to the Dignam protocol (Dignam *et al*., 1983) with several modifications. In the beginning of this work, the cleavage reaction protocol for the NXT was optimised according to the Wahle and Keller protocol (1994). The variation in the concentration of some components e.g. KCl, the addition of more extract and the reduction of the substrate RNA, allowed a slightly increased efficiency of the cleavage activity. Furthermore, the CTD of the RNAP II and different zinc ions concentrations were tested.

CTD acts as a landing platform for different protein factors during the transcription and allows efficient coordination of the mRNA maturation reactions. CPSF was shown to interact with the CTD during transcription initiation (Dantonel *et al*., 1997). Hirose and Manley (1998) suggested that the CTD is required for cleavage and showed the enhancement of the reaction by the addition of purified and phosphorylated CTD. In their experiment the purified

CTD was hyper-phosphorylated through incubation in nuclear extract. In our assays, no enhancement was observed after the addition of purified CTD or CTD, which was unspecifically phosphorylated by the MPF complex (cdc2/cyclin B). Reasons for the differences observed in the enhancement of the stimulation might be that it is not known if the purified, recombinant CTD was properly folded. Thus the CTD might not act as a binding platform for the 3' end processing factors. Nevertheless, we can not exclude, that the changed method of CTD phosphorylation may lead to a different amount and position of phosphorylated serines, which regulate the binding of 3' end processing factors.

Zink ions are bound in the β-lactamase domain of CPSF-73K. The loss of cleavage activity in nuclear extracts during dialysis was suggested to be due to the loss of bound Zn^{2+} by Ryan and colleagues (2002). They observed a slight stimulation of the cleavage reaction by the addition of $ZnCl_2$ to the cleavage reaction. However, we could not observe any stimulation by the addition of $ZnCl_2$. Conceivably, the loss of Zn^{2+} ions could be prevented, or at least reduced, by the addition of $ZnCl_2$ to the dialysis buffer.

Electrophoretic mobility shift assays were performed to estimate the yield of cleavage complexes assembled on RNA under different conditions. The formed RNA·protein complexes were so large, that around 80 % were not able to enter the native gel and were kept in the slots of the gel. Different changes in the reaction conditions were tested. The replacement of PVA by PEG 6000 allowed the entry of the cleavage complexes into the native gels and their separation, but did not lead to an increase in the amount of assembled cleavage complex.

Nevertheless, most RNA was bound in the heterogeneous complex (Figure 2-2). This complex consists of unspecifically bound proteins and migrates slower than the free RNA, but faster than the specific complex. Removing unnecessary sequences at the 5' end of the RNA did not affect the amount of heterogeneous complex. This means, that unspecific proteins compete for the binding sites with the 3' end processing factors, even in the presence of competitor RNA.

In conclusion, it was not possible to increase the amount of specific complex.

3.2 Poor complex formation on immobilised pre-mRNA

The tobramycin affinity selection method was used for attempts to purify the cleavage complex. The specific elution of the RNA·protein complex would allow an analysis of the proteins of the complex by mass spectrometry.

We showed that aptamer-containing RNA was still a substrate for the cleavage reaction and that the RNA was immobilised onto the matrix. However, it was not possible to purify the cleavage complex with this method. The cleavage product of the reaction was observed in the supernatant, but not on the matrix. This showed that no stable and / or active RNA·protein complex was formed on the matrix, whereas RNA, which dissociated from the matrix, was cleaved as a free RNA substrate.

The reasons might be, that the RNA dissociates from the matrix, through the unwinding of the aptamer structure, which is necessary for the binding to the tobramycin matrix. This aptamer unwinding might be catalysed by helicases, which are present in the extracts. Moreover the ratio between RNA and factors might be unbalanced, due to the large amount of RNA, which was used for the immobilisation and the concentration of the 3' end processing proteins in the extract. A further possibility might be that there is only a limited amount of active cleavage complex present in the extracts, which catalyses the cleavage reaction and dissociate afterwards. Additionally the huge amount of unspecifically bound proteins were bound and eluted from the Sepharose matrix (see Figure 2-4). These proteins might have covered any specific complex, eluted from the matrix. Furthermore the stability of the complex might be low compared to the protein complexes, e.g. spliceosome, which were purified with this method so far.

Since Sepharose matrix bound a lot of proteins unspecifically, we decided to use the λN peptide affinity method and a different matrix, MagneGST™, with a glutathione anchor. This polystyrol core matrix should be less susceptible to unspecific protein binding. The method of λN peptide affinity allows the binding of the substrate RNA to the matrix, via the recognition of the λN peptide by the BoxB RNA element. RNA containing BoxB elements was shown to be cleaved, the BoxB was recognised by the purified fusion protein GST-λN (see Figure 2-7) and the RNA was immobilised onto the matrix (see Figure 2-8).

However, also with this strategy, it was impossible to purify the cleavage complex. In only one occasion, a different protein pattern in the sample, compared to the specificity controls, was obtained in a silver-stained SDS gel, but this result was not reproduced. The same problems occurred for the λN peptide affinity method as for the tobramycin affinity selection

method used before; dissociation of RNA substrate from the matrix, the probable excess of RNA compared to 3' end processing proteins and the unspecific binding of proteins to the matrix. Even the preparation of a new RNA substrate, PAPolA, with two additional binding sites for CF I_m, did not lead to the purification of the complex.

What might the reasons be that the cleavage complex could be not purified assembled on a substrate RNA? The literature describes a complex consisting of CPSF, CstF and RNA. This result was obtained by electrophoretic mobility shift assays with partially purified CPSF (PF2) and CstF (CF1) without Mg^{2+} (Weiss *et al*., 1991). These two partially purified factors were not able to cleave a substrate RNA. Veraldi and colleagues (2000) described the purification of a polyadenylation complex on an immobilised RNA substrate with nuclear extract from mouse myeloma cells. These purified complexes were only tested in western blot analysis for CPSF and CstF subunits, but neither for the other subunits CF I_m, CF II_m and PAP nor cleavage activity. We suggest that the complexes purified in these experiments are likely to be the formerly described CPSF·CstF·RNA complexes. Veraldi and colleagues described that such gel filtration-purified complexes are stable for some hours, which fits with the previous data from Gilmartin and Nevins (1989), Prescott and Falck-Pederson (1992) and Weiss and colleagues (1991). Nevertheless, we did not detect any specific complex with the affinity-tag methods in western blot assays. Differences between the method of Veraldi and co-workers (2000) and the affinity methods used in this thesis include the region of the RNA used for the binding to the matrix. Veraldi used randomly biotinylated RNA substrates, whereas we included a secondary RNA structure at the 5' end of the substrates. Thereby the hairpin structure of the RNA would bring the 5' cap structure close to the matrix and might prevent binding of the cap binding complex (CBC). The CBC is known to enhance the cleavage activity (Hart *et al*., 1985, Gilmartin *et al*., 1988). Direct evidence for the influence of the CBC for 3' end processing was presented by Flaherty and colleagues (1997). They depleted the CBC from nuclear extracts and showed that the 3' end processing cleavage reaction was inhibited by approximately 80 %. Furthermore, the half-life time of the *in vitro* formed 3' end processing complexes was about 4 min in CBC-depleted extract, compared to 20 min in the mock-depleted extracts (Flaherty *et al*., 1997; Lewis and Izaurralde, 1997). From these experiments they suggested a "closed loop" of a functional pre-mRNA. Thereby the CBC stabilises and enhances the 3' end processing complex. The immobilisation of the pre-mRNA at the 5' end might not allow this pre-mRNA loop through steric hindrance. Thus the stabilising effect of the interaction of the CBC with the 3' end processing complex would be absent. 3' end immobilised RNA substrates were used, but even then no active or

immobilised cleavage complex was observed. The main problem of these RNA affinity methods may be the low abundance of 3' end processing factors compared to sum of proteins in the nucleus. We consider that the amounts of 3' end processing factors are lower than e.g. splicing factors, whereas no exact or estimated numbers for molecules per cell or nucleus are available for 3' end processing factors at the moment. Therefore we suggest to estimate these numbers, as it was done by Raue *et al*. (2007), who determined molecule numbers per cell for several yeast proteins. These numbers may allow a conclusion in the complex composition and may exemplify, if a known 3' end processing factor is present only in small amount and thereby limiting the complex assembly.

To exclude that the purification of complexes was not possible due to a weak complex stability, half life-time of these complexes was estimated using electrophoretic mobility shift assays. To estimate the stability of the complex, the conditions were tried to be identical as for the immobilisation assays. The protein·RNA complex was assembled on RNA for different time periods at 30 °C in tobramycin or GST-λN affinity assays as well as in the EMSA. Assembly conditions were based on the cleavage assay condition (see Table 2-2).
The result, a half-life time of 15 to 20 min, indicates that most of the complexes dissociated during the experiment. This fits with the stability data of Flaherty *et al.* (1997).
Cleavage assays showed that the cleavage product occurred after a lag-phase of approximately 15 min and increases during further incubation at 30 °C. Based on the results from EMSA and cleavage assays, it can be speculated that an inactive pre-cleavage complex is assembled during a period of approximately the 15 to 20 min probably consisting of the known proteins. During that time an unknown factor, e.g. a protein or a RNA, which might be present only in catalytic amounts, has to activate the complex. When activated, the complex catalyses the cleavage reaction and dissociates to allow the polyadenylation of the newly obtained 3' end. If the complex is not activated by the factor, the complex might dissociate as well, to newly assemble on another RNA substrate.

We would suggest to use a coupled transcription and 3' end processing system. Thereby a plasmid containing the sequence information for an RNA substrate with an affinity-tag would be *in vitro* transcribed. Together with the transcription machinery the 3' end processing complex would be formed with the elongating RNA polymerase II and could be purified via RNA affinity purification, if the are sufficient numbers of 3' end processing factors.

3.3 Purifications of the complex components

The second approach to analyse the complex composition was to affinity purify CstF, CF I_m and CF II_m expressed in human cell lines. These purified proteins were tested for the composition of the protein subunits and for unknown co-purifying proteins. For this purpose, the chosen proteins subunits were expressed with a HF-tag at the N-terminus of the protein, purified via the flag affinity matrix and analysed by mass spectrometry. Together with the purification of CPSF, prepared by Wlotzka, the cleavage reaction should be reconstituted *in vitro*.

3.3.1 CF I_m is probably a heterotetramer

CF I_m was thought to be a dimer and to be composed of the small subunit CF I_m-25K and a large subunit of CF I_m-59K, -68K or -72K (Rüegsegger *et al*., 1996). CF I_m activity was reconstituted *in vitro* using recombinant CF I_m-25K and -68K (Rüegsegger *et al*., 1998) or CF I_m-25K and -59K (Dettwiler *et al*., 2004). The self-interaction of CF I_m-25K was showed by GST pull-down assays (Dettwiler *et al*., 2004). An interaction between the large CF I_m subunits could not be observed in these pull-down assays (Dettwiler, thesis, 2003).

Firstly, CF I_m was purified via the tagged CF I_m-25K subunit. The expectation was to co-purify the large subunits of CF I_m and endogenous CF I_m-25K. This was indeed the case, and additionally confirmed the self-interaction of CF I_m-25K. The co-purification of the endogeneous form of CF I_m-25K is clearly seen in Figure 2-14B, where a single band, for the tagged CF I_m-25K with the flag specific antibody, and two bands, with the CF I_m-25K specific antibody, were observed in the western blot analysis. Nevertheless, co-purification of PAP, PABPN1, U1 snRNP-70K or AIP4, which were also shown to interact with CF I_m-25K, was not observed (see Introduction) (Kim and Lee, 2001; Dettwiler *et al*., 2004; Awasthi *et al*., 2003; Ingham *et al*., 2005). These interactions could not be confirmed in this thesis and might be only weak.

The preparations via the tagged form of CF I_m-68K contained the small CF I_m-25K subunit, the endogenous CF I_m-68K and small amounts of the other large subunits CF I_m-59K and -72K. In the case, that the large subunits do neither interact with each other nor with themselves, the co-purification of endogenous CF I_m-68K would be possible through the self-interaction of CF I_m-25K. Therefore we suggest that CF I_m is a heterotetramer.

The biological significance of this multimeric CF I_m might be the stronger interaction with the SR proteins by a tetrameric CF I_m containing two dimers of CF I_m-25K / -68K and the strengthening of the linkage to the spliceosome through the different interaction partners of

the large subunits (Figure 3-1). CF I_m-68K interacts with the SR proteins: SRp20, hTra2, 9G8 and SRm300, which belong to the splicing factors (Dettwiler *et al*., 2004). These interactions were not found for the CF I_m-59K (Dettwiler, thesis, 2003). Nevertheless CF I_m-25K / -59K was shown to interact with the U2AF-65 splicing factor which binds the pyrimidine tract of the last 3' splice site (Millevoi *et al*., 2006). These different interactions could link the 3' end processing with splicing (Figure 3-1).

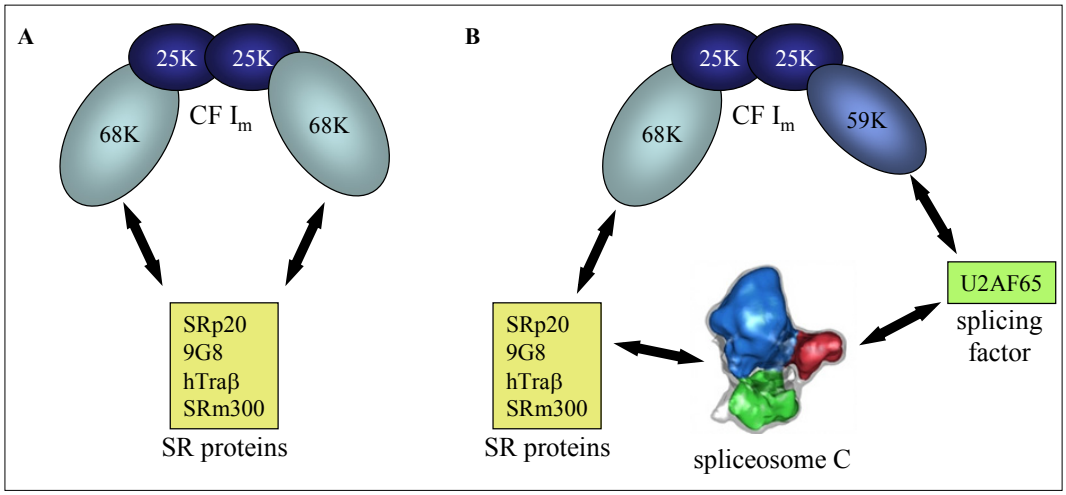


Figure 3-1 Interaction of the heterotetrameric model of CF I_m
Model based on self-interaction of CF I_m-25K and the co-purification of CF I_m-59K and endogenous CF I_m-68K via tagged CF I_m-68K.
A) Double dimer of CF I_m-25K / -68K. B) Heterotetramer of CF I_m. Double head arrows show the interaction between CF I_m-25K / -68K and the SR proteins (Dettwiler *et al*., 2004) and the CF I_m-25K / -59K with the U2 snRNP auxiliary factor 65 (U2AF-65) (Millevoi *et al*., 2006) and their interaction with the spliceosome. The 3D model of the spliceosome C complex is taken from Jurica *et al*. (2004) and is based on the structure obtained by EM.

3.3.2 Different methylation states of CF I_m-68K and CF I_m-59K

PRMT5 and MEP50 were co-purified in the affinity tagged preparations of CF I_m and CstF. These proteins are common contaminants of the flag purification (Chendrimada *et al*., 2005). PRMT5 and MEP50, together with pICln, belong to the methylosome complex which catalyses the symmetric dimethylation of arginines of e.g. the Sm proteins B/B', D1 and D3 (Brahms *et al*., 2000). Firstly, we assumed that these proteins were unspecifically co-purified with CstF as well as CF I_m. After the identification of the third subunit of the methylosome, the pICln protein, in CF I_m HF-72K and CF I_m HF-68K (second preparation) we further analysed the methylation of the CF I_m subunits, firstly in the literature and secondly in additional experiments.

CF I_m-68K was found to be immunoprecipitated with an antibody specifically recognising symmetric dimethylated arginines, while CF I_m-59K was not detected (Boisvert *et al*., 2003).

PRMT5 dimethylates glycine-arginine rich (GAR) motifs as well as single arginines of substrate proteins symmetrically (Richard *et al*., 2005; Pahlich *et al*., 2006). Sequence analysis revealed that CF I_m-68K contains a GRGRGR motif which is not present in CF I_m-59K (Figure 3-2).

```
CF Im-68K   1  MADGVDHINIYADVGEEFNQEAEYGGHDQIDLYDDVISPSANNGDAPEDRDYMDTLPPTV
CF Im-59K   1  MSEGVDLIDIYAD--EEFNQDPEFNNTDQIDLYDDVLTATSQPSDDRSSSTEPPPPVRQE

CF Im-68K  61 GDDVGKGAAPNVVYTYTG---KRIALYIGNLTWWTTDEDLTEAVHSLGVNDILEIKFFEN
CF Im-59K  59 PSPKPNNKTPAILYTYSGLRNRRAAVYVGSFSWWTTDQQLIQVIRSIGVYDVVELKFAEN
                                        RNP 2

CF Im-68K 118 RANGQSKGFALVGVGSEASSKKLMDLLPKRELHGQNPVVTPCNKQFLSQFEMQSRKTTQS
CF Im-59K 119 RANGQSKGYAEVVVASENSVHKLLELLPGKVLNGEKVDVRPATRQNLSQFEAQARKRECV
                    RNP1

CF Im-68K 178 GQMSGEGKAGPPGGSSRAAFPQGGRGRGRFPGAVPGGDRFPGPAGPGGPPPPFPAGQTPP
CF Im-59K 179 RVPRG---GIPPRAHSRDS-SDSADGR-----ATPSENLVPSSARVDKPPSVLPYFNRPP
                                                 proline rich domain

CF Im-68K 238 RPPLGPPGPPGPPGPPPPGQVLPPPLAGPPNRGDRPPPPVLFPGQPFGQPPLGPLPPGPP
CF Im-59K 230 SA-LPLMGLPPPPIPP------PPPLSSS-----------------FGVPP---------
                          proline rich domain

CF Im-68K 298 PPVPGYGPPPGPPPPQQGPPPPPGPFPPRPPGPLGPPLTLAPPPHLPGPPPGAPPPAPHV
CF Im-59K 257 -------PPPGIHYQHLMPPPP-----------------RLPPHLAVPPPGAIPPALHL
                                    proline rich domain

CF Im-68K 358 NPAFFPPPTNSGMPTSDSRGPPPTDPYGRPPPYDRGDYGPPGREMDTARTPLSEAEFEEI
CF Im-59K 292 NPAFFPPP-------NATVGPPPDTYMKASAPYNHHG----SRDSGPPPSTVSEAEFEDI
                          proline rich domain

CF Im-68K 418 MNRNRAISSSAISRAVSDASAGDYGSAIETLVTAISLIKQSKVSADDRCKVLISSLQDCL
CF Im-59K 341 MKRNRAISSSAISKAVSGASAGDYSDAIETLLTAIAVIKQSRVANDERCRVLISSLKDCL

CF Im-68K 478 HGIESKSYGSGS------RRERSRERDHSRSREKSRRHKS--RSRDRHDDYYRERSRERE
CF Im-59K 401 HGIEAKSYSVGASGSSSRKRHRSRERSPSRSRESSRRHRDLLHNEDRHDDYFQERNREHE
                                                 SR-like domain

CF Im-68K 530 RHRDRDRDR
CF Im-59K 461 RHRDRERDR
```

Figure 3-2 Sequence alignment of CF I_m-68K and CF I_m-59K
The subunits show 44.3 % identity, these amino acids are highlighted in gray. The RNA recognition motifs (aa 83 to 156 for CF I_m-68K and 84 to 158 for CF I_m-59K) consist of two RNPs which are indicated by solid boxes. The proline rich and SR-like domains are indicated by solid boxes, too. CF I_m-68K consists of a presumably methylation motif highlighted in red (aa 200-206). R400 (bold) of CF I_m-68K was only once found to be methylated by mass spectrometry analysis of affinity purified CF I_m.

Therefore affinity purified CF I_m proteins were analysed by mass spectrometry for methylated arginine residues. In only one occasion, it was found that arginine 400 of CF I_m-68K was methylated (see Figure 3-2) but this data was not convincing and needs to be verified. The peptide containing the GGRGRGR motif was not identified in the mass spectrometry.

Hence *in vitro* methylation assays were performed to examine whether bacterially expressed CF I_m-68K is specifically methylated by the methylosome complex, which was co-purified with the affinity purified CF I_m. This was indeed the case, whereas recombinant CF I_m-59K was not a substrate for methylation in these experiments. To confirm that PRMT5 is the methyl transferase of CF I_m-68K, the activity responsible for CF I_m-68K methylation was purified from HeLa cell extracts and determined to be PRMT5 in mass spectrometry analysis (experiments done by Martin).

These results indicate that in contrast to CF I_m-59K, CF I_m-68K is methylated at least by PRMT5. Arginine methylation is known to selectively modulate protein·protein interactions or maybe RNA·protein interactions, without changing the charge of the amino acids of the proteins (Bedford *et al*., 2000; Pahlich *et al*., 2005). The *in vivo* methylation of CF I_m-68K has to be confirmed as well as its regulatory function.

3.3.3 CstF interacts with two subunits of CPSF and with the τ variant of CstF-64K

CstF binds to the downstream element of the RNA and consists of three subunits. It is possible to reconstitute CstF from subunits separately expressed in baculovirus (Dettwiler, thesis, 2003). CstF was shown to interact *in vitro* with several protein subunits of the 3' end processing machinery. Most of these proteins interact with the largest subunit of CstF, the 77K subunit which bridges the RNA-bound CstF-64K with the CstF-50K subunit. Murthy and Manley (1995) showed that *in vitro* translated CstF-77K interacts with recombinant CPSF-160K. Kaufmann and colleagues used GST pull-down assays to show the interaction of CstF-77K with the recombinant N-terminal region of hFip1 (2004). CstF was further shown to interact with symplekin (Takagaki *et al*., 2000), hClp1 (de Vries *et al*., 2000), CTD, TFIIS (McCracken *et al*., 1997), Sen2 and Sen34 (Paushkin *et al*., 2004) and BARD1 (Kleiman *et al*., 1999).

CstF was purified via the tagged CstF-50K subunit. The mass spectrometry analysis of this affinity-purified protein identified the co-purifying protein subunits CstF-64K and -77K and τ variant of CstF-64K. None of the CstF interacting proteins e.g. hFip1, symplekin, TFIIS were co-purified with tagged CstF-50K. In contrast to this result, the preparation of tagged CstF-64K from HEK293 cells contained, in addition to the CstF subunits, CPSF-100K and

-160K, the associated factors hFip1 and symplekin. This confirms that the interaction between CstF and CPSF also occurs in extracts. Due to the fact that not all CPSF subunits were co-purified, we suppose that CPSF might exist in different complex species. This supposition is confirmed by data from Dickson *et al.* (1999), who showed that CPSF is involved in cytoplasmic polyadenylation in *Xenopus* oocytes.

The co-purification of CstF-64 τ indicates, that CstF might occur as a large protein complex. We suggest that it consists of four subunits: CstF-50K, -77K, -64K and -64K τ. An indication for oligomeric CstF was found in *Drosophila melanogaster*. Simonelig and colleagues used mutant Su(f) protein (suppressor of forked), the drosophila homolog of CstF-77K, to complement Su(f) mutant flies. This data leads to the model that several monomers of mutant Su(f) protein form a functional hybrid complex (Simonelig *et al*., 1996).

CstF-64K and CstF-64K τ show different affinities to poly(U) and poly(GU), which may result from differences in the domain downstream of the RBD (Monarez *et al.,* 2007) (see Figure 3-3). CstF, containing two CstF-64K variants, would allow the recognition of most DSE elements of mRNAs which are less conserved and contain a short U-rich sequence and/or a GU-rich motif. Nevertheless, this suggestion has to be confirmed *in vivo*.

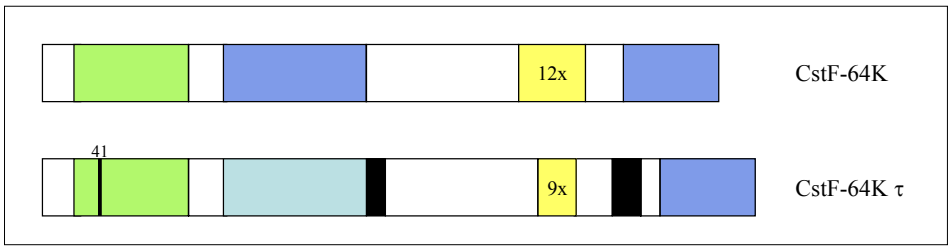


Figure 3-3 Comparison of human CstF-64K (577 aa) and human CstF-64K τ (616 aa)
Domains are highlighted in different colours RBD (green), CstF-77K interacting domain (blue), MEARA repeats (yellow). CstF-64K τ variant contains an aa exchange in position 41 (pro→ser) and two insertions (black), the CstF-77K interaction domain is divergent from CstF-64K and highlighted in light blue. Figure adopted from Dass *et al*., 2002.

3.3.4 hClp1 probably exists in two protein complexes

CF II_m was first purified by de Vries and colleagues (2000) and the preparation contained, next to hClp1 and hPcf11, a couple of other unrelated proteins like all CF I_m subunits, splicing and transcription factors. Paushkin and colleagues (2004) identified hClp1 as a co-eluting protein from HF-affinity purified tRNA splicing endonuclease complex. *Vice versa* they could co-elute the subunits of the tRNA splicing endonuclease complex with tagged hClp1. Only HF-hClp1 preparation contained additional hPcf11, the second known subunit of CF II_m.
CF II_m purified via tagged hClp1 (this work; Kyburz, thesis, 2006), as well as hPcf11 / hClp1 purified from baculovirus, were able to reconstitute the cleavage activity in hPcf11-depleted extract. Even bac-hPcf11 alone reconstituted the cleavage activity of hPcf11-depleted extract (Dettwiler, thesis 2003). This means, that hClp1 is still present in the depleted extract, probably in the tRNA splicing endonuclease complex. The hPcf11-depleted extract often showed a minor background activity, whereas depletion with the hClp1 antibody resulted only in a small reduction of the cleavage activity and was therefore not used for the activity assays. This data implies that, in a hPcf11-depleted extract, the hPcf11 protein is removed together with hClp1 as CF II_m complex, whereas additional hClp1 exists in the tRNA splicing endonuclease complex. Dissociation of the tRNA splicing endonuclease complex would allow the reconstitution of CF II_m with bac-hPcf11.

In the publication Paushkin *et al.* (2004) suggested a link between tRNA splicing and pre-mRNA 3' end processing. Kyburz confirmed that hClp1 is a subunit of the tRNA splicing endonuclease complex, but in contrast to Paushkin and co-authors, showed that purified HF-tagged Sen protein complexes were not able to reconstitute the cleavage activity in antibody-depleted NXT. Therefore, the suggestion of a link between tRNA splicing and 3' end processing was rejected (Kyburz, thesis, 2006).

Finally, we propose that hClp1 is part of two complexes, first of CF II_m and second of the tRNA splicing endonuclease complex. This model could explain the observed weak reduction of cleavage activity in hClp1-depleted extract. Thereby hClp1 is present in the tRNA splicing endonuclease complex and might be protected against antigen-antibody interaction. This model could be tested by the affinity purification of hPcf11 with an affinity-tag and the analysis of its co-eluting proteins. Here we would expect to co-purify only hClp1, but none of the tRNA splicing endonuclease complex subunits, if the model is true.

3.4 Reconstitution of the pre-mRNA 3' end cleavage reaction failed

The aim was to reconstitute the cleavage reaction with purified factors. Therefore almost all known protein factors necessary for 3' end processing (CF I_m, CF II_m, CstF and CPSF) were flag-purified from HEK293 cells. CF I_m, CF II_m and CstF were shown to be active by restoring the activity of depleted extract (this work). CPSF was kindly provided by Wlotzka, who detected a minimal cleavage activity after the addition of the protein together with CstF and PAP in CPSF-depleted extracts. Solely PAP was expressed in bacteria.

Reconstitution of cleavage was done using the same amount of purified factors which, was sufficient to restore the cleavage activity in antibody-depleted extract. In all experiments no cleavage product was observed. Additional experiments with increased amounts of purified proteins did not show any activity either. Further attempts using additional BSA, purified CTD or hFip1 failed, as well as the use of different RNA substrates, the increase of creatine phosphate concentration or pre-incubation of the proteins.

The CPSF preparation of Wlotzka showed only a very weak complementation of the cleavage activity in CPSF antibody-depleted extracts. Therefore a calf thymus CPSF preparation was used. However, this did not result in a cleavage product in the reconstitution assay.

Reconstitution assays by Rüegsegger contained partially purified CF II_m and a cleavage active CPSF preparation from calf thymus. Purified and cleavage active CF II_m is now available with affinity tagged hClp1. However, CPSF preparations from calf thymus done by Dettwiler, Kaufmann, Kyburz (pers. communication Kyburz) and myself could not restore the cleavage activity in CPSF-depleted extracts but all CPSF preparations showed polyadenylation activity. The baculovirus expressed and reconstituted CPSF was inactive in cleavage, too (Dettwiler, thesis, 2003).

One possibility is that the used concentrations of the purified tagged 3' end processing factors in the total reconstitution assay are to low and / or contained partially inactive protein. The preparation of larger amount of tagged protein from HEK293 cells would allow further tries of reconstituting the cleavage reaction.

Furthermore, it is possible that an unknown factor is missing. This factor could be limited in extracts and might function as a catalyst for the reaction. This factor could be purified by fractionation of HeLa cell nuclear extract. These fractions have to be tested for their ability to reconstitute the cleavage reaction with the other purified factors.

The final reconstitution of completely purified protein factors from the cleavage reaction will lead to a new possibility of characterising the function of each subunit.

4. Summary

Mammalian mRNAs undergo several maturation steps in the nucleus. Next to capping and splicing, 3' end processing is necessary for the maturation of RNA polymerase II transcripts. The reaction at the 3' end leading to the poly(A) tail is a two step reaction. The pre-mRNA is cleaved by a 3' end cleavage complex containing the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulation factor (CstF), the cleavage factors I (CF I_m) and II (CF II_m) and the poly(A) polymerase (PAP). After the endonucleolytic cut, the pre-mRNA is polyadenylated by a complex containing PAP, CPSF and, later on, the nuclear poly(A) binding protein (PABPN1). While the polyadenylation reaction can be reconstituted *in vitro* by purified and recombinant proteins, the reconstitution of the 3' end processing cleavage reaction has not been possible so far. The cleavage reaction was reconstituted with partially purified factors from calf thymus and recombinant proteins (Rüegsegger *et al.*, 1996 and thesis), but these assays did not result in a knowledge of all protein subunits necessary for the reaction. Further reconstitution assay with recombinant proteins expressed in baculovirus failed (Dettwiler, thesis, 2003).

The aim of this PhD work was the purification of a complete and functional pre-mRNA 3' end processing cleavage complex. The complex should be assembled on an immobilised RNA and purified via RNA affinity methods. The mass spectrometry analysis of the bound proteins should identify all proteins necessary for the pre-mRNA 3' end processing reaction.

First we used the tobramycin affinity selection method. This method allows the immobilisation of pre-mRNA, through the interaction of a RNA aptamer sequence with the aminoglycosid antibiotic tobramycin, on a Sepharose matrix. Several attempts were made to adapt the method for the cleavage complex. Nevertheless, it was not possible to purify a pre-mRNA cleavage complex. Strong unspecific interactions of CPSF to the Sepharose matrix were observed. Another RNA affinity chromatography approach was used, based on the interaction of the BoxB RNA element from Phage λ and the antiterminator protein N. The use of a paramagnetic matrix reduced the unspecific binding of proteins, but this approach did not result in an active cleavage complex either. Further attempts, e.g. another RNA substrate, which contains more CF I_m binding sites, did not show the desired results. Additional analysis done by electrophoretic mobility shift assay, revealed a low cleavage complex stability, which might be the reason for the failure of these methods.

Thus we changed the tools for the affinity purification of the cleavage factors from human cells, for the reconstitution of the cleavage reaction. Therefore HEK293 cells were transfected

with plasmids coding for subunits of CF I_m and CstF with an N-terminal double affinity tag, which contained an octa-histidine stretch and the flag peptide.

Purified tagged CF I_m was able to reconstitute the cleavage activity of CF I_m depleted NXT. Purification via different subunits revealed that CF I_m probably is a multimeric protein complex, consisting at least two of each of the two subunits. We assume, that two copies of CF I_m-25K self-interact and each CF I_m-25K interacts with a large subunit of CF I_m. This larger CF I_m complex might support the coupling of the 3' end processing reaction to splicing, through stronger interactions respective to different interaction partners.

CstF purified via the 64K subunit complements the cleavage activity in CstF depleted NXTs. Mass spectrometry analysis of this protein preparation revealed the co-purification of some CPSF subunits. This demonstrates the tight interaction of CPSF and CstF, but also that the CPSF complex might exist in different complexes.

CF II_m was purified via HF-hClp1 and tested in the same way. The cleavage-active CF II_m contained three proteins of the tRNA splicing endonuclease complex in addition to hClp1 and hPcf11. These proteins co-elute with the HF-hClp1 protein. hClp1 is present in two complexes, the tRNA splicing endonuclease complex (Paushkin *et al*., 2004) and CF II_m.

CPSF was purified via tagged CPSF-73K and contained all known subunits, and the associated factors hFip1 and symplekin, but showed only weak activity in CPSF-depleted NXT.

These proteins were used together with recombinant poly(A) polymerase in an attempt to reconstitute the cleavage complex. Although a set of different experimental conditions were tested, cleavage activity was not observed. The reasons for this could be, the weak activity of the CPSF used, the low concentration of the proteins in the assay, or that there is still an unknown protein factor, protein or RNA, which associates with the known complex components during the cleavage reaction. This factor could be limited in extracts and might function as a catalyst for the reaction.

In future it is necessary to purify larger amounts of 3' end processing factors from HEK293 cells and to identify the unknown factor. This might allow reconstitution of the cleavage complex with the necessary concentrations of the proteins. The estimation of the protein molecule number per cells for the 3' end processing factors might help to find the ratio of the factors necessary for cleavage. The reconstitution of the cleavage reaction with defined protein complexes will provide a new approach to analyse the functions of each protein subunits, e.g. to finally prove that the CPSF-73K is the endonuclease of the cleavage reaction.

5. Material and Methods

5.1 Chemicals

Chemicals were obtained from GE Healthcare, Biozym, Bio-Rad, Fluka, Merck, Roth, Serva and Sigma in a high purity (p.A.).

Table 5.1 Chemicals

name	supplier
acrylamide bisacrylamide 19:1	Accugel, National Diagnostics
Agarose	Invitrogen
Ammonium peroxodisulphat (APS)	Merck
Boric acid	Roth
Bromphenol blue	Merck
Chloroform	Merck
Diethyl-pyrocarbonate (DEPC)	Sigma
Dithiothreitol (DTT)	Gerbu
Ethanol p.a.	Roth, Merck
Ethylenediamine tetra-acetic acid (EDTA)	Merck
Glycerol	Roth
Hydrochloric acid	Merck
Magnesium acetate	Merck
Magnesium chloride	Merck
N,N,N',N'-Tetra-methyl-ethylen-diamine (TEMED)	Merck
Nonidet P40	Fluka
Phenol, TE saturated	Roth
Polyvinyl alcohol (PVA)	Sigma
Potassium chloride	Roth
RNasin	Promega
Sodium acetate	Merck
Sodium-dodecylsulphate (SDS)	Merck
Tris(hydroxymethyl)aminomethane (Tris)	Roth
tRNA	Boehringer
Urea	Roth, Invitrogen
Xylene cyanol blue	Merck

5.1.1 Antibiotics

Table 5.2 Antibiotics

antibiotic	supplier
Ampicillin	Roth
Carbenicillin	Roth
Chloramphenicol	Merck
Kanamycin	Merck
Tetracyclin	Merck

5.2 Enzymes and proteins

Table 5.3 Enzymes used in this work

enzyme / protein	supplier
Bovine serum albumin (BSA)	New England Biolabs
BSA Fraction V	Merck
Calf intestinal alkaline phosphatase (CIP)	Roche Diagnostics
CF I_m (crude preparation)	gift from Dettwiler
CF IIm (54-1)	gift from de Vries
CPSF IV	own preparation
CstF (from HeLa cells)	gift from Dettwiler
GST	gift from Kühn
hFip1	gift from Kaufmann
Methylated BSA	gift from Körner
MPF	gift from Martin
RNAguard	Roche Diagnostics
RNasin	Promega
RNase A	Roth
PAP	gift from Martin; resp. own preparation
Proteinase K	Merck
Pwo DNA polymerase	Roche Diagnostics, Peqlab
SP6 RNA polymerase	Roche Diagnostics
T4-DNA-Ligase	New England Biolabs
T4-Polynukleotidkinase	New England Biolabs
TEV protease	Invitrogen
TEV protease	gift from Blank

All used restriction enzymes were obtained from New England Biolabs or MBI Fermentas. Additional enzymes and proteins are listed below in the Table 5.3.

5.3 Kits

Kits were used according to the protocols supplied by the manufacturers.

Table 5.4 Summary of used kits with the corresponding companies

name	company
ABI Prism BigDye Terminator Cycle Reaction Kit® 1.1	Applied Biosystems
Colloidal blue Staining Kit®	GIBCO-Invitrogen
Endofree Maxi Kit®	Qiagen
Gene Elution Kit®	Sigma Aldrich
Jetsorb®	Genomed
Qiagen Plasmid Midi Kit®	Qiagen
Rapid DNA Ligation Kit®	Roche
SP6 Megascript®	Ambion

5.4 Buffers and solutions

The standard buffers were prepared as described in Sambrook and co-workers (2001).

FBK100: 100 mM KCl
50 mM Tris-HCl pH 7.9
10 % glycerol
0.2 mg/mL methylated BSA
0.01 % Nonidet P40

Gel extraction buffer: 500 mM Ammonium acetate
10 mM Magnesium acetate
0.1 mM EDTA
0.1 % SDS

Buffer D: 20 mM HEPES KOH pH 7.9
20 % glycerol
100 mM KCl
0.2 mM EDTA
0.5 mM DTT

Ponceau S solution	0.25 g Ponceau S
	1 % acetic acid
Stripping buffer:	100 mM β-Mercaptoethanol
	2 % SDS
	62.5 mM Tris-HCl pH 6.7
PBS:	137 mM NaCl
	2.7 mM KCl
	20 mM Na_2HPO_4
	2 mM KH_2PO_4
PBST:	PBS
	0.05 % Tween 20
Urea PAGE S1	20 % acrylamide / bisacrylamide 19:1
	8.3 M Urea
	1 x TBE
Urea PAGE S2	8.3 M Urea
	1 x TBE
acrylamide solution	30 % acrylamide / bisacrylamide 80:1

5.5 General methods

General molecular biology techniques like cloning, transformation and DNA preparation were performed according to Sambrook and co-workers (2001). Kits (Methods 5.3) were used, if possible. Primers used are shown in Tables 7.1, 7.3, 7.5, 7.7 and 7.8. Sequences of obtained plasmids were confirmed by cycle sequencing on an ABI PRISM® 310 Genetic analyser (PE Applied Biosystems) and were analysed with the software Four peaks® and SeqMan®.

5.6 Protein methods

5.6.1 Bradford assay

Protein concentrations were determined according to Bradford (1976). BSA was used as a standard. 20 µL of a known BSA concentration or protein dilution were added to 750 µL of diluted Bradford reagent and were incubated for 15 min at room temperature prior to absorption measuring. The observed absorptions of 8 to 10 different known BSA concentrations were measured at 595 nm and plotted in a calibration curve. The absorption of unknown proteins was measured at least three times and the protein concentration was calculated from the calibration curve.

5.6.2 Protein concentration

If necessary proteins were concentrated in Microcon Y-10 spin columns (1.5 mL volume, Amicon, Millipore) at 4 °C according to the manufacturer's protocol.

5.6.3 Staining of proteins in SDS-polyacrylamide gels

Proteins separated on SDS-PAGE were stained with Coomassie brilliant blue G250 or, for mass spectrometry analysis, with Colloidal blue staining kit® or with silver nitrate according to Nesterenko *et al*. (1994).

5.6.4 Filter binding assay

The filter binding assay is an approach to determine the dissociation constant (K_D) of an RNA protein interaction, based on the retention of a protein on nitrocellulose filters by non-covalent interactions. Radioactively labelled RNA, which is bound to the protein remains on the nitrocellulose filter, while unbound RNA is not retarded. The amount of bound RNA can be quantified by scintillation counting.

The RNA was incubated with increasing amounts of protein in FBK100 in a total volume of 50 µL for 10 min at room temperature. Filters were soaked in wash buffer (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1mM EDTA) containing 0.1 µg/mL tRNA to prevent unspecific RNA binding to the nitrocellulose filters. The complete RNA protein reaction mixture was applied to the filter, washed with 5 mL ice-cold wash buffer and the bound radioactivity was measured in a scintillation counter. The amount of RNA retained and the protein concentration were plotted in a double-reciprocal function (1/[RNA] vs. 1/[protein]]. The x-intercept of the linear graph represents $-1/K_D$. The K_D value was calculated from three independent measurements (see Wahle *et al*., 1993).

5.7 RNA methods

5.7.1 *In vitro* transcription of radioactively and non-labelled RNA

Reaction mix (50 µL): 1x SP6 transcription buffer
1 µg linearised DNA
10 mM m^7GpppG cap
1 mM DTT
1 mM ATP, CTP
0.1 mM UTP, GTP
20 µCi α P^{32}-UTP
40 U RNAguard
40 U SP6 RNA polymerase

Transcription reactions were performed using SP6 RNA polymerase according to the protocol supplied by Roche. DNA templates were linearised with Dra I, or in case of templates containing an additional inserted sequence element, with Pvu II. The RNAs were capped by addition of m^7GpppG cap structure (New England Bioscience) and radioactively labelled by incorporation of [α-^{32}P] UTP (GE Healthcare). Transcription products were separated on a 6 % denaturing urea gel. Full length transcripts were excised from the gel and incubated in 400 µL gel extraction buffer over night at room temperature. The eluted RNA was phenol / chloroform extracted, ethanol precipitated, dissolved in DEPC treated H_2O and stored at -20 °C. Non-labelled and capped transcripts were prepared with SP6 Megascript® Kit according to the manufacturer's protocol and purified with a pre-packed G-50 Sephadex spin column (Roche), phenol / chloroform extracted, ethanol precipitated, dissolved in DEPC-treated H_2O and stored at -20 °C.

The radioactively and non-labelled RNA were mixed to obtain large amounts of RNA which were necessary for RNP complex purification, and to be able to control the RNA binding to the matrix in these purifications.

5.7.2 Denaturing Urea PAGE

Denaturing urea PAGE was used to separate RNA molecules according to their length. The gels were prepared by mixing Urea PAGE S1 and Urea PAGE S2 solutions and the polymerisation was started by addition of TEMED and 10 % APS, e.g. for a 20 mL gel 300 µL APS and 30 µL TEMED were added. Before loading urea polyacrylamide gels, a pre-run was performed at 20 W for 30 min in 1x TBE. Samples were resuspended in formamide buffer, denatured for 5 min at 95 °C and loaded. The gel electrophoresis was usually

performed until the xylene cyanol blue dye reached the middle of the gel. The urea gel was exposed over night to Phosphorimager screens or to X-ray films.

5.7.3 Standard cleavage reaction

The optimised conditions for the cleavage reaction obtained in this thesis (see Results 2.1., Table 2-2) are termed 'standard' for this work. They include a 25 µL cleavage activity test with 50-100 fmol RNA using the following reaction mixture. For cleavage complex purification larger volumes of this standard mixture were used.

Standard Cleavage reaction mix:	12.5 µL NXT-A (preparation, Methods 5.10.2)
(25 µL volume)	2 mM DTT
	1.5 mM $MgCl_2$
	0.8 mM 3' dATP
	0.1 µg/µL tRNA
	3.5 % PEG 6000
	50 mM KCl
	20 mM creatine phosphate
	5-10 U RNasin

The reaction were incubated for 2 hours at 30 °C and stopped by proteinase K treatment (Wahle, 1991a) followed by ethanol precipitation. The RNA was analysed by denaturing urea PAGE (5.7.2).

5.7.4 Electrophoretic mobility shift assay (EMSA)

2x Gels (20 x 20 cm size)	0.5 g agarose (in 50 mL deionised water, melted and equilibrated at 55 °C for 30 min)
	3 % acrylamide / bisacrylamide 80:1
	0.5 x TBE
	adjust to 100 mL deionised water

This method can determine if a protein or a protein mixture is capable of binding to a given RNA sequence. A native polyacrylamide gel was pre-run at 300 V for 30 min at 4 °C. Protein and RNA were incubated for 15 min at room temperature in an appropriate reaction buffer and loaded onto the gel without a dye. One lane contained DNA dye containing Bromphenol blue. The gel was run until the Bromphenol blue reached the end of the gel. The gel was dried and exposed over night to a Phosphorimager screens or to X-ray films.

5.8 Tobramycin affinity selection method

The tobramycin affinity selection method is a strategy for the purification of native RNP complexes. It was developed by Hartmuth and colleagues (2002) to obtain preparative amounts of purified pre-spliceosomes under native conditions and is based on the specific binding of an RNA aptamer to the aminoglycoside antibiotic tobramycin. RNA containing the tobramycin recognition aptamer J6f1 was immobilised on the tobramycin-derivatised Sepharose (Hartmuth *et al*., 2004).

5.8.1 Preparation of RNA for tobramycin affinity selection method

The DNA sequence coding for the aptamer J6f1 was inserted into plasmids containing the sequence of cleavage substrates L3 or SV40 late or shorter variants lacking 100 or 80 nt at the 5' end under control of a SP6 promotor. If the cleavage complex is active during the preparation, we expected to purify only a partial complex, missing the 3' cleavage fragment and proteins which dissociate during the complex remodeling for the polyadenylation reaction. Therefore the aptamer was inserted at the 5' and 3' end of the substrate, respectively, to compare the composition of the purified complexes containing the 5'- or the 3'-cleavage product. All oligonucleotides and plasmids are summarised in Tables 6.1 - 6.2.

5.8.2 Preparation of tobramycin coupled NHS-Sepharose

Blocking buffer A:	0.2 M ethanolamine HCl pH 8.0 (freshly prepared)	Coupling buffer:	0.2 M $NaHCO_3$ 0.5 M NaCl pH 8.3 NaOH

4x Binding buffer:
80 mM Tris-HCl pH 8.1
4 mM $CaCl_2$
4 mM $MgCl_2$
0.8 mM DTT

Blocking buffer B:	1x binding buffer 300 mM KCl 0.1 mg/mL tRNA 0.5 mg/mL BSA 0.01 % Nonidet P40	RNA binding buffer:	1x binding buffer 145 mM KCl 0.1 mg/mL tRNA

Wash buffer W145:	1x binding buffer	Wash buffer W75:	1x binding buffer
	0.2 mM DTT		0.2 mM DTT
	145 mM KCl		75 mM KCl
	0.1 % Nonidet P40		0.1% Nonidet P40

Elution E145T: 1x binding buffer
0.2 mM DTT
5 mM Tobramycin
145 mM KCl
2 mM $MgCl_2$

Tobramycin was dissolved in coupling buffer and stored at -20 °C as a 100 mM solution. 2 mL of NHS-Sepharose fast flow (packed volume) were washed four times with 9 mL of 1 mM HCl, and the matrix was collected by centrifugation for 10 seconds at 2000 xg (4 °C). The matrix was incubated with 5 mM Tobramycin in 1 mL coupling buffer, rotating over night at 4 °C. The resulting tobramycin Sepharose was collected by centrifugation for 5 min at 2000 xg at 4 °C and the supernatant was completely removed. The matrix was blocked by incubation with 8 mL blocking buffer A by rotation for two hours at room temperature. The beads were collected by centrifugation, washed three times with 9 mL PBS and twice with PBS containing 0.02 % NaN_3.

To confirm the specificity of RNA binding the beads were incubated with aptamer RNA and control RNA without an aptamer sequence. 15 µL of tobramycin Sepharose (packed volume) per reaction were incubated over night with blocking buffer B at 4 °C with rotation, the beads were collected by centrifugation and incubated with 40 pmol radioactively labelled RNA in 400 µL binding buffer for 1.5 h at 4 °C with rotation. A 5 µL sample was taken from the mixture to determine the total input of radioactivity (A_0), and the beads were washed three times with 1 mL W145 buffer. The wash supernatants of one sample were collected in one tube (W). The washed beads (B1) were also measured in a scintillation counter prior to elution with 50 µL E145 buffer per aliquot for 15 min at 4 °C with head over tail rotation. After centrifugation the RNA amounts of beads and supernatants were determined by scintillation counting. The matrix was shown to bind at least 75 % of the aptamer and less than 5 % of the unspecific RNA.

5.8.3 Tobramycin affinity selection experiment

Cleavage reaction mix (500 µL): 20 x standard cleavage reaction mix

The RNA was coupled to the beads (15 µL packed volume) as described above. The matrix was washed and incubated with 500 µL cleavage reaction mix for 2 hours at 30 °C. The beads were washed three times with 750 µL of W75 buffer and eluted with 250 µL E145T buffer for 15 min at 4 °C. 5 µL samples were taken at each step and the beads were measured in a scintillation counter after every step to verify the RNA binding and elution. The RNA protein complex formation was analysed by preparing 10 % SDS polyacrylamide gels with samples from supernatants, washes and elutions for silver staining and western blotting (5.11.2.1). Different conditions in the RNP formation step such as altered incubation times, composition of cleavage reaction mix with different NXT-A and a downscale of the beads were tested, as mentioned in the results.

5.9 λN peptide affinity method

As a further alternative, the λN peptide affinity approach was used, which was developed by Czaplinski and co-workers (2005) for the examination of RNP formation on long RNAs. This approach is based on the λ phage N anti-terminator protein, which specifically binds to the BoxB sequence in λ phage RNA. The 21 amino acid peptide of the protein λN was fused to GST and the fusion protein was immobilised by coupling of the GST tag to the glutathione Sepharose. Two tandem copies of the BoxB sequence were inserted in the plasmids coding for the RNA of interest. The RNA is immobilised on the matrix by λN-BoxB interaction.

5.9.1 Expression of λN-GST protein

The plasmid pGLambdaH2 coding for the λN peptide with an N-terminal GST and a C-terminal $(his)_6$-tag was a kind gift of Czaplinski. The purification was done according to the protocol given in supplementary data in the publication of Czaplinski and co-workers (2005). After the Ni-NTA purification the protein was dialysed (2 x 3 h, 1 L buffer D). The concentration of the protein was determined by Bradford assay (Chapter 5.6.1). The yield of GST-λN protein resulting from 800 mL *E. coli* culture were 24 mg protein with a concentration of 1.7 mg/mL and 54.5 µM, respectively.

5.9.2 Preparation of RNA substrates containing the BoxB sequence

The BoxB element was inserted several times as tandem repeats into plasmids containing the cleavage substrates L3. The oligonucleotides used as well as the plasmids are summarised in the Appendix (Tables 6.3, 6.4, 6.5 and 6.6).

5.9.2.1 PAPolA – a new cleavage substrate

The work of Venkataraman and colleagues (2005) showed that the 3' UTR of the poly(A) polymerase A mRNA contains four UGUAN motifs which are binding sites for CF I_m. We expected this substrate to stabilise the cleavage complex for the purification, due to more binding sites for CF I_m compared to L3 (2 UGUAN motifs) and SV40 late (one UGUAN motif). The new substrate was analysed in cleavage activity assays and EMSA prior to λN peptide affinity (see results).

5.9.3 λN peptide affinity experiment

Starting from the protocol given by Czaplinski and co-workers (2005) we varied several parameters: Different blocking reagents and their incubation time on the matrix; the order of binding the fusion protein, the RNA and the proteins from the NXT-A on the matrix; the length of incubation times of NXT-A; different elution buffers and the new substrate PAPolA as described in the results.

The following fundamental protocol was used: The magnetic glutathione beads (4 µL slurry per reaction) were washed 3 x with buffer D and pre-incubated with 100 µg BSA in buffer D for one hour at 4 °C with rotation and washed again three times with buffer D. The beads were incubated with 1 pmol RNA and 8 µg GST-λN protein for one hour at 4 °C. A small sample was removed to determine the total radioactivity of the input RNA. The RNA affinity matrix was washed twice with buffer D and the bound RNA was determined. To assemble the mRNA processing complex, the RNA affinity matrix was incubated with 250 µL standard cleavage reaction mix for 15 min at 30 °C on a shaker. The supernatant was removed and the beads were washed three times with buffer D, followed by elution in buffer D containing 50 mM reduced glutathione, 0.1 % triton X-100 and 100 mM NaCl. The beads were further incubated in SDS sample buffer to determine the elution efficiency.

5.10 Cell culture

5.10.1 Media, solutions and cell culture dishes

The components were obtained from Invitrogen with exception of Hygromycin and cell culture flasks from Falcon.

HeLa suspension Medium:	1 x MEM Joklik
	1 x Penicillin, Streptomycin (5000 U/ mL; 5000µg/mL)
	10 % NCS (newborn calf serum)
HeLa Medium complete:	MEM + 4500 mg/L glucose + GlutaMAX™ I + pyruvate
	10 % fetal calf serum (FCS)
	1 x Penicillin Streptomycin (5000 U/ mL; 5000µg/mL)
HEK Medium complete:	D-MEM + 4500 mg/L glucose + GlutaMAX™ I + pyruvate
	10 % FCS
	1 x Penicillin Streptomycin (5000 U/ mL; 5000µg/mL)
Selective Medium:	complete Medium containing Hygromycin (HeLa: 200 mg/L, HEK293 150 mg/L)
Freezing Medium:	90 % FCS
	10 % DMSO

OptiMEM (serum free Medium)
Lipofectin® (1 mg/mL)
Lipofectamin 2000® (1 mg/mL)
Hygromycin (50 mg/mL) in PBS (GIBCO Invitrogen, A.G. Scientific and Calbiochem)
Trypsin EDTA (1x)

5.10.2 HeLa cell suspension culture and nuclear extract preparation

1 x MEM Joklik medium was prepared according to the suppliers instructions. HeLa cells were thawed, washed in complete medium, resuspended in complete medium and transferred to a small spinner flask. The cells were counted and further diluted with medium to a concentration of 2 x 10^5 cells per mL. The cells were incubated at 37 °C and slowly stirred (200 rpm). Every two to three days the cells were counted and diluted to around 2 x 10^5 cells per mL. 6 L of cells suspension were harvested and NXT was prepared as described in Wahle & Keller (1994). NXT resulting from this protocol is termed NXT-A.

NXT-B was prepared with the same protocol with few changes: all buffers contained proteinase inhibitors (PMSF, Leupeptin, Pepstatin). Tris-HCl pH 7.9 was used instead of HEPES, buffer C contained 20 mM KCl instead of NaCl, and the high salt buffer C contained 800 mM $(NH_4)_2SO_4$. Dialysis buffer D contained additionally 0.02 % Nonidet P40 and 20 mM $(NH_4)_2SO_4$ instead of 100 mM KCl, and an ultracentrifugation step was performed (60 min in Centrikon T-2000 Ultracentrifuge with rotor TFT 65.38) in the end prior to dialysis. NXT resulting from this protocol is termed NXT-B.

5.10.3 Cultivation and transfection of adherent HeLa cells

HeLa cells were grown in 75 cm^2 tissue culture flasks. Frozen cells were thawed fast, washed with PBS and seeded in complete medium. The cells were grown for two to three days. For further cultivation the cells were diluted 1:10 or 1:20 and seeded into new flasks. Transfections were performed with Lipofectin® as described in the manufacturer's protocol. The concentration of Hygromycin necessary for efficient selection was determined to be 200 µg/mL medium.

A plasmid coding for GFP (gift from P. Guye, Biozentrum Basel) was used to test the transfection efficiency by observing the green fluorescence of the transfected cells. The transfected cells were grown in 150 mm plates in selection medium until isolated foci were observed and could be marked with a pen. The colonies were picked with a pipette tip and cultivated separately. Samples of 2.5 x 10^5 cells were used for western blot analysis (5.11.2.1) to control the expression of the transfected gene. Clones, stably expressing the protein subunits, were slowly frozen in several aliquots and stored in liquid nitrogen for later use.

For protein purifications, appropriate cells were seeded into 150 x 25 mm plates with 30 mL selective medium. Approximately 1 to 1.5 mL of packed cell volume was harvested from ten plates with confluently grown cells. Harvested cells were collected by centrifugation at 500 xg for 10 min at room temperature and frozen in liquid nitrogen.

5.10.4 Cultivation and transfection of adherent HEK293 cells

HEK293 adherent cells were grown and harvested as described for adherent HeLa cells. For transfections 3 x 10^5 cells were seeded into a six-well plate per well and were incubated overnight at 37 °C with 5 % CO_2 in an humidified incubator. Transfections were performed at 70 - 80 % confluence using 4 µg of plasmid DNA and 10 µL of Lipofectamin 2000® according to the Invitrogen protocol. Medium was changed after 12-14 hours. One day later, the cells were seeded in selective medium in 150 mm plates. The concentration of

Hygromycin necessary for efficient selection of HEK293 cells was determined to be 150 µg/mL medium. Stable cell lines were selected as described above.

5.10.5 Cell lysis for western blot analysis

RIPA buffer:	50 mM Tris-HCl pH 7.9	freshly added: 0.2 mM $NaVO_3$
	150 mM NaCl	0.5 mM DTT
	10 mM NaF	0.5 mM PMSF
	0.4 mM EDTA	2 µg/mL Pepstatin
	10 % glycerol	2 µg/mL Leupeptin
	1 % Nonidet P40	
	0.1 % SDS	
	0.5 % Sodium deoxycholat	

Lysates were analysed by western blot analysis to verify expression of tagged proteins. 2.5 x 10^5 cells per sample were resuspended in 60 µL RIPA buffer and incubated at 4 °C for 10 min in a shaker followed by centrifugation for 5 min at 3000 xg at 4 °C. The supernatant was transferred to a fresh tube containing 60 µL SDS sample buffer. After incubation for 5 min at 95 °C the cell lysate was analysed by SDS-PAGE followed by western blot analysis (Chapter 5.11.2.1).

5.11 Purification of proteins containing different affinity tags from human cells

5.11.1 Cloning of plasmids expressing TAP-tagged and his_8-flag-tagged proteins

One part of this thesis was the reconstitution of the 3' end cleavage complex with purified proteins from human cells. We therefore inserted the cDNA encoding the subunits of the proteins CF I_m, CstF and CF II_m in plasmids for the expression of the proteins in human cells. We used the plasmid pTRE2hyg-TAP (Kühn, see Appendix) expressing an N-terminal TAP tagged protein and pcDNA3.1 which expressed the protein with a N-terminal his_8-flag tag. The plasmid pcDNA3.1 was a kind gift of Trotta. The primers used and the obtained plasmids are listed in Tables 6.7, 6.8 and 6.9 in the Appendix.

5.11.2 Purification and analysis of the expressed proteins

5.11.2.1 Western blot analysis

Milk powder (non fat, dry milk) (Rapilait, Migros)

ECL Western blotting detection reagents and analysis system (GE Healthcare Bioscience)

Table 5.5 Primary antibodies

name	directed against	clone	from	company / preparer
his	his_6 peptide	monoclonal	mouse	Clontech
F3165	flag peptide	monoclonal	mouse	Sigma
P-3775	protein A	polyclonal	rabbit	Sigma
CBP	Calmodulin binding protein	monoclonal	mouse	Open Biosystems
CF I 30K	CF I_m-25K	polyclonal	rabbit	Rüegsegger / Basel
CF I 70K	CF I_m-68K	polyclonal	rabbit	Rüegsegger / Basel
# 237	CPSF-30K	polyclonal	rabbit	Barabino / Basel
# 009	CPSF-73K	polyclonal	rabbit	Jenny / Basel
66	CPSF-100K	polyclonal	rabbit	Kühn / Halle
# 0680	CPSF-160K	polyclonal	rabbit	Jenny / Basel
# 97	hFip1	polyclonal	rabbit	Kaufmann / Basel
CstF 50K	CstF-50K	polyclonal	rabbit	Dettwiler / Basel CstF2
	CstF-64K	polyclonal	mouse	Abnova
CstF3	CstF-77K	monoclonal	mouse	Abnova
# 171	CF II_m hClp1	polyclonal	rabbit	de Vries / Basel
# 868	CF II_m hPcf11	polyclonal	rabbit	Kaufmann / Basel
0658	PAP	polyclonal	rabbit	Martin / Basel
# 57478	PABPN1	polyclonal	rabbit	Wahle / Halle

Table 5.6 Secondary antibodies containing a HRP conjugate

name	directed against	clone	prepared in	company
α-rabbit	rabbit IgG	polyclonal	swine	DakoCytomation
α-mouse	mouse IgG	polyclonal	rabbit	DakoCytomation

The proteins were separated by SDS PAGE and transferred onto a nitrocellulose membrane by semi-dry electroblotting at 1750 V for 75 min according to Sambrook *et al*. (2001). The efficiency of the blot was determined by staining the membrane with Ponceau S solution. After destaining with PBST (Chapter 5.4), the membrane was blocked with 5 % milk in PBST for 1 hour at room temperature, followed by five wash steps, and the membrane was incubated for 1 hour with the primary antibody in PBST. After three washes for 10 minutes each, the membrane was incubated with the secondary antibody for one hour and washed again prior to detection with horseradish peroxidase substrate. If necessary the membrane was stripped for incubation with other antibodies according to the protocol from ECL Western blotting detection reagent. The membrane was incubated by rotation in 25 mL of stripping buffer for 30 min at 50 °C. Afterwards the membrane was blocked again with milk in PBST prior to incubation with further antibodies.

5.11.2.2 Purification of TAP-tagged proteins from nuclear extract

IPP 150:
10 mM Tris-HCl pH 8.0
150 mM NaCl
10 % glycerol
0.1 % Nonidet P40
Protease inhibitors

TEV cleavage buffer:
10 mM Tris-HCl pH 8.0
150 mM NaCl
10 % glycerol
0.1 % Nonidet P40
0.5 mM EDTA
1 mM DTT
Protease inhibitors

IgG Sepharose™ 6 fast flow (GE Healthcare) was washed four times with IPP150 then incubated with NXT diluted in IPP150 for three hours at 4 °C. The beads were collected by centrifugation, washed three times with IPP150 and twice with TEV cleavage buffer. The proteins were eluted in 1 mL TEV cleavage buffer containing 20 µL TEV protease for 1.5 hours at 16 °C with occasional mixing.

5.11.2.3 Purification of flag-tagged proteins from total cell extract

Anti-Flag® M2 Affinity Gel (Sigma, A2220)
Flag peptide (F3290) (5 mg/mL) in TBS (Sigma)

TBS: 50 mM Tris-HCl pH 7.4
150 mM NaCl

buffer B: 250 mM NaCl
30 mM Tris-HCl pH 7.0
1 mM EDTA
5 % glycerol
0.1 % Triton X-100

buffer N: 200 mM NaCl
40 mM Tris-HCl pH 7.0
2 mM $MgCl_2$
5 % glycerol
0.05 % Triton X-100

buffer W: 400 mM NaCl
30 mM Tris-HCl pH 7.0
1 mM EDTA
5 % glycerol
0.05 % Triton X-100

protease inhibitors
(added to all buffers before use):
1 µg/mL Leupeptin
1 µg/mL Pepstatin
1 mM PMSF

For protein purification 25 µL of Flag-Agarose matrix (packed volume) was used per 1 mL of packed cells. For purifications approximately 30 mL packed cells were used. All steps were carried out at 4 ºC. The matrix was washed according to the protocol supplied by Sigma and equilibrated in buffer B. The cells were resuspended in half of their volume of buffer B, disrupted by sonification (seven times for 10 s, 50% duty, intensity 6) with a Branson Sonifier 250 and centrifuged for one hour at 18.000 xg. It was not possible to filter the centrifuged total cell extract as described in Paushkin and colleagues (2004). We therefore increased the centrifugation force and time to remove larger particles. Further purification steps were done according to the protocol from Paushkin and colleagues (2004). The second purification step, the affinity purification via Ni-NTA Sepharose, was not performed for later CF I_m and CstF protein preparations. Glycerol was added to the elution fractions to a final concentration of 20 %, and the proteins were stored at -80 ºC. The proteins were concentrated (5.6.2) and dialysed against buffer D. Samples of all steps were analysed by SDS polyacrylamide gels electrophoresis for Western blotting (5.11.2.1) and silver staining.

5.11.2.4 Mass spectrometry

The sample preparation for mass spectrometry was done according to the protocols from Jenö (Biozentrum Basel). The proteins were carboxymethylated with iodoacetamide prior to loading onto an SDS-polyacrylamide gel. The Colloidal blue stained protein bands were cut out. The proteins were digested with trypsin or chymotrypsin and identified via MS / MS by Jenö at Biozentrum Basel.

5.11.2.5 Activity tests

5.11.2.5.1 Depletion of nuclear extract

IPP 150:	150 mM KCl	IPP 500:	500 mM KCl
	20 mM Tris-HCl pH 7.9		20 mM Tris-HCl pH 7.9
	0.01 % Nonidet P40		0.01 % Nonidet P40

This method allows specific removal of a protein from nuclear extracts (NXT). A protein-specific antibody was coupled to a matrix and incubated with the protein mixture in the NXT. Protein A Sepharose CL-4B was prepared according to the protocol delivered by GE Healthcare and stored in 100 mM Tris-HCl pH 7.9 with 0.1 % NaN_3. The following control reactions were done in parallel using NXT-B: Pre-incubated NXT without protein A Sepharose to show the loss of reaction activity during the preparation; mock depleted NXT to demonstrate that proteins unspecifically bound to the Sepharose have no influence in the reaction, and a sample with the pre-serum of the antibody to show that the loss of activity obtained by the protein specific antibody is due to the removal of the specific protein. 50 µL packed protein A Sepharose per reaction were washed and equilibrated in IPP 500 over night with rotation at 4 °C. The beads were collected and 50 µL polyclonal antibody was added in an end volume of 500 µL IPP 500 and incubated with rotation over night at 4 °C. The beads were washed four times with 1 mL IPP 500 and further incubated with 50 µL of NXT for four hours rotating at 4 °C. After centrifugation at 500 xg for 5 min the supernatant was tested in cleavage assays for activity. Antibodies used for depletion are summarised in Table 5.7.

Table 5.7 The antibodies used for NXT depletion

antibody	depletion efficiency
polyclonal CF I_m-25K (CF I_m 30K)	60 %
polyclonal CF I_m-68K (CF I_m 70K)	93 %
polyclonal CstF-50K	no depletion
polyclonal CstF-64K (CstF 2)	70 %
monoclonal CstF-77K (CstF3)	80 %
polyclonal CPSF-30K (# 237)	95 %
polyclonal CPSF-73K (# 009)	no depletion
polyclonal CPSF-160K (# 0680)	no depletion
polyclonal hFip (# 97)	95 %
polyclonal hClp1 (# 171)	no depletion
polyclonal hPcf11 (# 868)	95 %

For the depletion of CstF-77K from NXT we coupled the monoclonal antibody to protein G Sepharose CL-4B, all other antibodies were coupled to protein A Sepharose CL-4B. Depletion efficiency: cleavage product obtained from NXT prior to depletion was set to 100 %. The remaining cleavage product occurring in depleted NXT was determined and subtracted from 100%.

5.11.2.5.2 Activity test using depleted nuclear extract

This approach was used to test the activity of the protein which is missing in the depleted NXT. All other necessary components should be still present in the extract.

Reaction mix (25 µL): contain DTT, $MgCl_2$, 3' dATP, tRNA, KCl and creatine phosphate as described in 5.7.3.
2.6 % PVA
0.01 µg/µL creatine kinase
15 fmol RNA substrate
3-5 µL depleted NXT
10-200 ng purified protein
adjust to 25 µL with DEPC treated H_2O

Depleted NXT was pre-incubated with the purified protein for 15 min on ice prior to the addition of reaction mix and the incubation for two hours at 30 °C. To control the reconstitution, cleavage assays were performed using the pre-incubated NXT, mock depleted NXT, pre-serum depleted NXT and the CF I_m-25K depleted NXT instead of NXT. After two hours incubation the reactions were stopped by proteinase K treatment (Wahle, 1991a), and the RNA was precipitated and analysed by denaturing PAGE containing 8.3M urea (5.7.2).

5.11.2.5.3 Total reconstitution assay

The total reconstitution assay is a cleavage assay using only purified proteins to test whether the activity can be reconstituted.

Reaction mix (25 µL): cleavage reaction mix (5.11.2.5.1) without NXT
CF I_m, CF II_m, CstF, CPSF, PAP in various amounts

The proteins were incubated on ice for 15 min. The reaction mix was added and the samples were incubated for two hours at 30 °C. One sample contained NXT instead of the proteins and served as a positive control for the reaction. The reaction was stopped by proteinase K treatment (Wahle, 1991a) and the RNA was precipitated and analysed by denaturing Urea PAGE (5.7.2).

5.12 Computational analysis of RNA bands

To determine the activity of NXT or specific complex formation in EMSA, the obtained cleavage products or specific complexes, visible on the autoradiographies, had to be transformed into numbers. The films or phosphorimager screens were scanned and the digital autoradiographies were analysed by the software ImageQuant.

Each sample lane was scanned. To compare the samples, it was necessary uniform the total radioactively labelled RNA (input) per lane, by calculation of an input-factor. Afterwards the numbers for the cleavage products were determined (CP) and multiplied with the input-factor. The CP of control reaction was set to 100 % and CP of the samples was calculated in terms of percent.

6. Appendix

Table 6.1 Primers for cloning the tobramycin aptamer into pSP64 plasmid

name	sequence
HindIII α T PstI	AGC TTg gct tag tat agc gag gtt tag cta cac tcg tgc tga gcc CTG CA
PstI α T HindIII	Ggg ctc agc acg agt gta gct aaa cct cgc tat act aag ccA
EcoRI α T PvuII	AAT TCg gct tag tat agc gag gtt tag cta cac tcg tgc tga gcc CA
PvuII α T EcoRI	CTG ggc tca gca cga gtg tag cta aa ctc gct ata cta agc cgG
Mutagen SV40Δ 3'-5'	aat gaa tgc aat tgt tgt tgt taa ctt gtt tct tgc agc
L3 s PstI	TTA TAA CTG CAG tct ttt tgt cac ttg aaa aac a
SV40 late s PstI	TTA TAA CTG CAG tgc ttt att tgt gaa att tgt g

Capital letters show 5' overhangs, underlined sequences correspond to recognition sites for the restrictions enzymes as indicated in the primer name. The first four primers were aligned pair wise and ligated into the opened plasmids, α T in the name corresponds to the tobramycin aptamer J6f1, grey underlaid is the mutated poly(A) signal, s = short version, lacking 80 or 100 nt at the 5' end

Table 6.2 Plasmids coding for cleavage substrates with the tobramycin aptamer

plasmid coding for RNA with 5' aptamer	RNA substrate	plasmid coding for RNA with 3' aptamer
pSP64 α T - L3	L3	pSP64 L3 - α T
pSP64 α T - L3 Δ	L3, Δ correspond to point mutation U to G in PAS	pSP64 L3 Δ - α T
pSP64 α T - L3 s	L3, s abb. for short, correspond to deletion of 100 nt at the 5' end)	pSP64 L3 s - α T
pSP64 α T - L3 s Δ	L3, short (see above)	pSP64 L3 s Δ - α T
pSP64 α T - L3 pre	L3 pre, polyadenylation substrate	n.c.
pSP64 α T - L3 pre Δ	L3 pre, Δ (see above)	n.c.
pSP64 α T - SV40 late	SV40 late	pSP64 SV40 late - α T
pSP64 α T - SV40 late Δ	SV40 late, Δ (see above)	pSP64 SV40 late Δ - α T
pSP64 α T - SV40 late s	SV40 late, short (80 nt deletion at the 5' end)	pSP64 SV40 late s - α T
pSP64 α T - SV40 late s Δ	SV40 late, short, Δ (see above)	pSP64 SV40 late s Δ - α T

pSP64 is the vector used for SP6 transcription, α T indicates the tobramycin aptamer and its position in front (5') or in the end (3') of the RNA substrate, PAS is poly(A) signal, n.c. not cloned.

Table 6.3 Primers for cloning the BoxB sequence into pSP64 plasmid

name	sequence
HindIII BoxB 5'-3'	AGC TTg ggc cct gaa gaa ggg ccc G
HindIII BoxB 3'-5'	AGC TCg ggc cct tct tca ggg ccc A
EcoRI BoxB 5'-3'	AAT TCg ggc cct gaa gaa ggg ccc G
EcoRI BoxB 3'-5'	AAT TCg ggc cct tct tca ggg ccc A
HindIII bb PCR	ACA CAC AAG CTT ggg ccc tga aga agg
SpeI bb PCR	ACA CAC ACT AGT ggg ccc ttc ttc agg

Capital letters show overhangs, underlined sequences correspond to recognition site for the indicated restriction enzyme, Restriction enzymes in the name were used for cloning, number of 'b' corresponds to the numbers of BoxB elements

Table 6.4 Plasmids coding for cleavage substrates with BoxB element

name	insert
pSP64 bb L3	two BoxB hairpins 5' from L3
pSP64 bb L3 Δ	two BoxB hairpins 5' from L3 Δ
pSP64 bbb L3	three BoxB hairpins 5' from L3
pSP64 bbbb L3	three BoxB hairpins 5' from L3
pSP64 bbbb L3 Δ	four BoxB hairpins 5' from L3 Δ

pSP64 is the vector used for SP6 transcription, number of 'b' corresponds to the numbers of BoxB elements, Δ corresponds to point mutation U to G in poly(A) signal

Table 6.5 Primers for PAPolA cloning

name	sequence
PAPolA PstI fill up	ACA CAC CTG CAG act gga gtt tgc ttt gtt tta tag tat ctg tac tcc ttg tat ttt tca aga gct att ttg taa aca gat gat gta ttt ctc cat tga aaa cac aat aa
PAPolA PvuII fill up	ACA CAC CAG CTG taa act tat gat tac tgg aaa aaa gta gta gtc aaa aga aat cag gca act gtg aga ttg tgc tgt ttt ttt ttt att gtg ttt tca atg gag aaa tac atc a
Mutagenese PAPolA	TTA ATG CAG CTG taa act tat gat tac tgg aaa aaa gta gta gtc aaa aga aat cag gca act gtg aga ttg tgc tgt ttt ttt ttt ctt gtg tt

Capital letters show 5' overhangs, underlined sequences correspond to recognition site for the indicated restriction enzyme, Restriction enzymes in the name are used for cloning, grey underlaid is the mutated poly(A) signal

Table 6.6 Plasmids coding for cleavage substrate PAPolA

name	insert
pSP64 PAPolA	3' UTR of poly(A) polymerase A
pSP64 PAPolA Δ	3' UTR of poly(A) polymerase A with point mutation U to G
pSP64 bb PAPolA	two BoxB elements 5' from PAPolA
pSP64 bb PAPolA Δ	two BoxB elements 5' from PAPolA Δ
pSP64 bbb PAPolA Δ	three BoxB elements 5' from PAPolA Δ

pSP64 is the vector used for SP6 transcription, number of 'b' corresponds to the numbers of BoxB elements, Δ correspond to point mutation U to G in poly(A) signal

Table 6.7 Primers for the construction of protein expression plasmids

name	sequence
NheI CF I_m-25K	CTC TCT GCT AGC tct gtg gta ccg ccc aat cgc t
Bam HI CF I_m-25K	ACA CAC GGA TCC gtc tgt ggt acc gcc caa tcg ct
CF I_m-25K NotI	CTC TCT GCG GCC GCt cag ttg taa ata aaa ttg aac ctg
NheI CF I_m-59K	CTC TCT GCT AGC tca gaa gga gtg gac ttg att g
Bam HI CF I_m-59K	GTG CAA GGA TCC gtc aga agg agt gga ctt gat tg
CF I_m-59K NotI	CTC TCT GCG GCC GCt cag tgg tgc cgg tcc cgt tct cta tc
NheI CF I_m-68K	CTC TCT GCT AGC gcg gac ggc gtg gac cac ata g
Bam HI CF I_m-68K	GTG CAA GGA TCC ggc gga cgg cgt gga cca cat ag
CF I_m-68K NotI	CTC TCT GCG GCC GCc taa cga tga cga tat tcg cgc tct
NheI CstF-50K	CTC TCT GCT AGC gta cag aac caa agt ggg ctt
Bam HI CstF-50K	ACA CAC GGA TCC gta cag aac caa agt ggg ctt ga
CstF-50K NotI	CTC TCT GCG GCC GCt cag tca gtg gtc gat ctc c
NheI CstF-64K	CTC TCT GCT AGC ggc ggg ttt gac tgt gag aga
Bam HI CstF-64K	ACA CAC GGA TCC ggc ggg ttt gac tgt gag aga cc
CstF-64K NotI	CTC TCT GCG GCC GCt caa ggt gct cca gtg gat t
NheI CstF-77K	CTC TCT GCT AGC gtc agg aga cgg agc cac gga
Bam HI CstF-77K	TTT TTT GGA TCC gtc agg aga cgg agc cac gga g
CstF-77K NotI	CTC TCT GCG GCC GCc tac cga atc cgc ttc tgc t

Capital letters show overhangs, underlined sequences correspond to recognition site for the indicated restriction enzyme, position of restriction enzyme in front corresponds to an upstream primer, in the end to a downstream primer

Table 6.8 Sequencing primers for DNA coding for CF I_m and CstF subunits.

name	sequence
TAP 950 – 970	CTC AGC AGC CAA CCG CTT TAA
his-flag forward	GAA ATA GAT ATT AAG AAA ACA
CF I_m-59K 1400 – 1420	GGT GGT GGT AGC CTC TGA AA
CF I_m-59K 1750 – 1770	CAC CTC TCT CCT CAA GCT TT
CF I_m-68K 1400 – 1420	TGG TGT TGG ATC TGA AGC AT
CF I_m-68K 1800 – 1820	CTA GCT GGG CCT CCT AAT CG
CF I_m-72K 1670 – 1690	GCC AGC AGG ACC AGG AGG GC
CstF-50K 350 – 370	GTA GAG ATG GAC AGT TAA TA
CstF-50K 700 – 720	TCG GAA CTC AGC ATC CTA CT
CstF-64K 1480 – 1500	AGG AGG CAC GGA ACA TGT TAC
CstF-64K 1880 – 1900	AGT GTC CAT GGA ACG GGG GCA
CstF-77K 381 – 400	CTT TGC ACT GGA TAA AAT TG
CstF-77K 781-800	AGC AAC CCT CTT CGT ACA GA
CstF-77K 1181-1200	GAA GAA TGA TAT TTA AAA AA
CstF-77K 1581-1600	ATA CAA GTT CAT GGA TTT AT
CstF-77K 1981-2000	GGT GGG GCC CCA GAG CTA GC

Names correspond to protein subunits and numbers to nucleotides in original pBluescript® II KS vector

Table 6.9 Plasmids coding for protein expression in human cells lines

name	insert	affinity tag
pcDNA3.1 HF CF I_m-25K	cDNA of CF I_m-25K	HF
pcDNA3.1 HF CF I_m-59K	cDNA of CF I_m-59K	HF
pcDNA3.1 HF CF I_m-68K	cDNA of CF I_m-68K	HF
pcDNA3.1 HF CF I_m-72K	cDNA of CF I_m-72K	HF
pcDNA3.1 HF CstF-50K	cDNA of CstF-50K	HF
pcDNA3.1 HF CstF-64K	cDNA of CstF-64K	HF
pcDNA3.1 HF hPcf11	cDNA of hPcf11	HF
pTRE2hyg TAP CF I_m-25K	cDNA of CF I_m-25K	TAP
pTRE2hyg TAP CF I_m-59K	cDNA of CF I_m-59K	TAP
pTRE2hyg TAP CF I_m-68K	cDNA of CF I_m-68K	TAP
pTRE2hyg TAP CF I_m-72K	cDNA of CF I_m-72K	TAP
pTRE2hyg TAP CstF-50K	cDNA of CstF-50K	TAP
pTRE2hyg TAP CstF-64K	cDNA of CstF-64K	TAP
pTRE2hyg TAP CstF-77K	cDNA of CstF-77K	TAP

pcDNA3.1 and pTRE2hyg plasmids were used for expression of indicated proteins in human cell lines HEK293 and HeLa, HF abbr. for his_8-flag

Table 6.10 Summary of purified proteins

protein	pur. subunit	purification	activity	comment
CF I_m	25K	flag + Ni-NTA	+	all CF I_m subunits
	25K	flag	+	all CF I_m subunits
	68K D12	flag	+	all CF I_m subunits
	68K E2	flag	+	68K and 25K, contains other subunits in low amounts
	72K	flag	+	all CF I_m subunits, HF-72K C-terminally degraded
CstF	50K	flag + Ni-NTA	+/-	all CstF subunits
	50K	flag	+	all CstF subunits
	64K	flag	+	all CstF subunits
CF II_m	hClp1	flag + Ni-NTA	+/-	hPcf11 co-purified
	hClp1	flag + Ni-NTA	+	hPcf11 co-purified
CPSF	100K	flag + Ni-NTA	-	all CPSF subunits
(Wlotzka)	100K	flag	-	all CPSF subunits
	73K	flag	+/-	all CPSF subunits

Purified subunit indicates the affinity purified subunit, flag indicates the affinity purification over a flag matrix, Ni-NTA indicates the affinity purification over Ni-NTA Sepharose, activities were tested with depleted NXT see 5.12.2.5.3., CPSF-100 K was purified from Wlotzka during her diploma work.

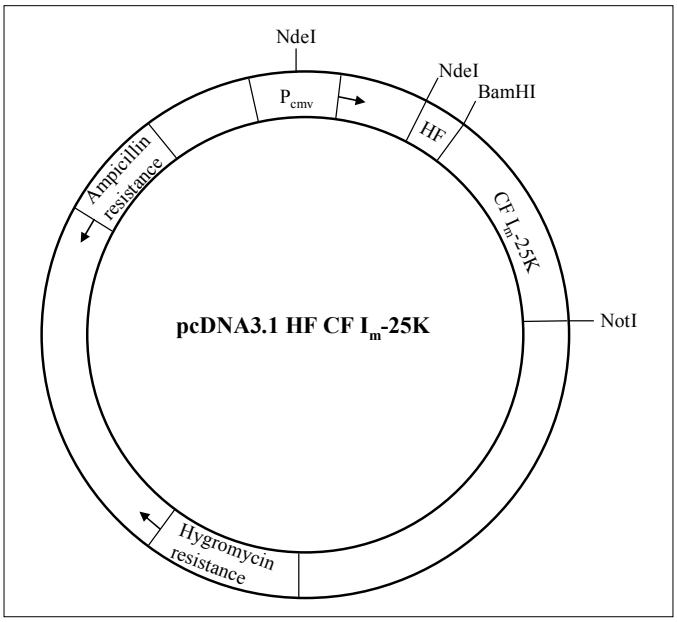


Figure 6.1 **Plasmid Map of pcDNA3.1 CF I_m HF-25K**
The plasmid pcDNA3.1 hSen 34 was a gift from Trotta. The cDNA coding for CF I_m-25K was amplified and ligated in frame into the BamHI / Not I opened plasmid.

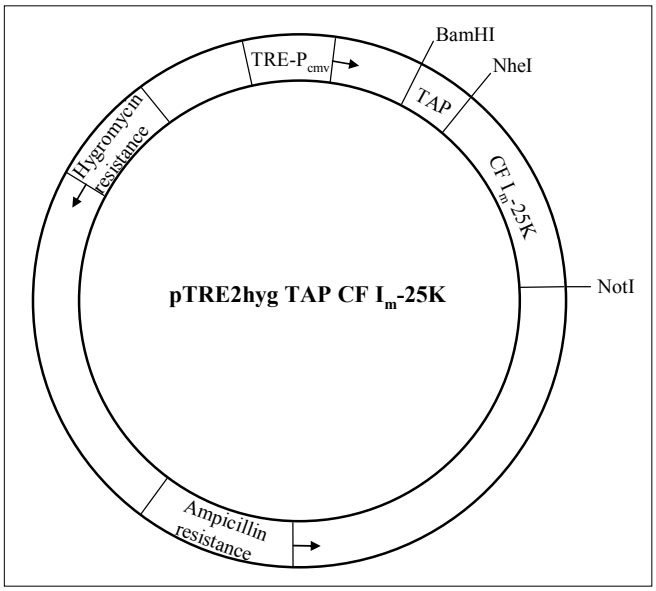


Figure 6.2 **Plasmid map of pTRE2hyg TAP CF I_m-25K**
The TAP tag was amplified from pBS1761 using PCR primers and ligated into the BamHI / NheI pTRE2hyg plasmid gift from Kühn. The cDNA of CF I_m-25K was amplified by PCR and ligated in frame into the NheI / Not I opened pTRE2hyg TAP plasmid.

Table 6.11 A Comparison of human cleavage complex proteins with yeast homologs and *vice versa* (B)

Factor	Human	Size [K]	Yeast	RefSeq	Factor	Size [K]	Identity [%]	Motif
CPSF	CPSF-30K	30	Yth1p	NP_015432	CPF	24.5	48	Yth1
	CSPF-73K	77.5	Ysh1p	Q06224	CPF	87.7	40	Ysh1, Lactamase_B
	CPSF-100K	88.4	Ysh1p	Q06224	CPF	87.7	26	Ysh1, Lactamase_B
	CPSF-160K	161	Cft1 / Yhh1p	Q06632.	CPF	153.4	25	SFT1, CPSF_A
	hFip1	66.5	Fip1p	NP_012626	CPF	35.7	48	Fip1
	Symplekin	141	Pta1p	Q01329	CPF	88.5	27	
CstF	CstF-50K	48.4	**Pwp2p**	NP_009984		104	21	2x WD40
	CstF-64K	61	Rna15p	P25299	CF IA	32.8	47	RRM
	CstF-77K	82.9	Rna14p	CAA89771	CF IA	80	27	RNA14, Suf
CF I_m	CF I_m-25K	26.2	**tRNA guanosine-2'-O-methyltransferase TRM3**	Q07527		165	29	CspR
	CF I_m-59K	52	Rna15p	P25299	CF IA	32.8	35	RRM
			polyadenylate-binding protein	AAA34838	CPF	64.4	31	4x RRM, Poly(A)
			Hrp1p / Nab4p	AAB18142	CF IB	59.6	24	2x RRM
			Nsr1p	NP_011675		44.5	24	2x RRM
	CF I_m-68K	59.2	Hrp1p / Nab4p	Q99383	CF IB	60	29	2x RRM
			polyadenylate-binding protein	AAA34838	CPF	64.4	31	4x RRM, Poly(A)
	CF I_m-72K	63.5	polyadenylate-binding protein	AAA34838	CPF	64.4	31	4x RRM, Poly(A)
			Hrp1p / Nab4p	AAB18142	CF IB	59.6	24	2x RRM
CF II_m	hClp1	47.6	Clp1p	Q08685	CF IA	50.2	25	Clp1
			Protein GRC3	Q07845		72.2	28	Clp1, GTPase
	hPcf11	174	Pcf11p	NP_010514	CF IA	71.9	39	CID
PAP	PAP	82.8	Pap1p	NP_012927	CPF	64.5	46	PAP-central, PAP-RNA bind, NTP-transf_2

Table 6.11 B Comparison of yeast cleavage complex proteins with human homologs

Factor	Yeast	Size [K]	Human	RefSeq	Factor	Size [K]	Identity [%]	Motif
Cleavage factor I A, CF I A	Rna14p	76	CstF-77K	Q12996	CstF	82.9	27	9x HAT
			PRP39	Q86UA1		107	20	7x HAT
	Rna15p	38	CstF-64K	P33240	CstF	61	47	RRM
			CstF-64K	AAH28239.1		64.4	46	RRM
			CstF-64K hypothetic	CAB66681.1		64.4	46	RRM
	Pcf11p	72	hPcf11	O94913	CF II_m	184	30	CID
			KIAA0824	BAA74847.1		183	30	CID
	Clp1p	50	hClp1	Q92989	CF II_m	47.6	24	GTPase
			ATP/GTP binding protein	NP_006822.1		47.6	24	GTPase
Cleavage factor I B, CF IB	Nab4p	73	**hnRNP D**	BAA09524.1		24.6	41	RRM
			hnRNP A2/B1	P22626		37.4	38	RRM
	Nrd1p	64	**RNA-binding motif protein 16**	NP_055707.3		140.5	22	RRM, CID
			Splicing factor, RS-rich 15	NP_065757.1		125.9	21	
Cleavage and Polyadenylation factor, CPF	Yhh1p	150	CPSF-160K	EAW82106.1	CPSF	160.9	25	SFT1, CPSF_A
			CPSF A	AAC50293.1	CPSF	160.9	25	SFT1, CPSF_A
	Ydh1p	105	CPSF-100K	Q9P2I0	CPSF	88.4	28	Lactamase_B, Ysh1
			KIAA1367 protein	BAA92605.1		65.6	22	Ysh1
			CPSF-73K-like	CAI23179.1		58.2	25	Ysh1
	Ysh1p	100	CPSF-73K	AAH11654.1	CPSF	77.5	40	Lactamase_B, Ysh1
	PTA1p	85	SYMPK variant	BAE06092.1		118.8	27	
			Symplekin	AAC50667.1		126.5	27	
			KIDINS 220 protein	AAI30611.1		185.5	21	3x ANK
	PAP1p	64	PAP	P51003	PAP	82.8	46	NTP_transf_2, PAP central, PAP_RNA bind

	Ref2p	60	**Unnamed protein**	BAB15631		127	29	DOP1
			Hypothetical protein	CAB82395		142	29	DOP1
	Mpe1p	58	**Retinoblastom binding protein 6**	AAH63524		102	30	RING, Zn-finger
	Fip1p	55	hFip1	AAQ88277	CPSF	66.5	48	Fip1
			Fip1L1	AAH24016		63	39	Fip1
	PSF2p	53	**WDR33**	NP_060853		146	38	WD40
			WDR51a	AAI10878.1		39.8	26	WD40
	Pti1p	47	CstF-64K	P33240	CstF	61	28	RRM
			CstF-64K τ	AAH28239.1	CstF	64.4	24	RRM
	Swd2p	37	**WD40 protein**	AAQ88631		35	33	WD40
			Transmembran protein 113	EAW65202.1		35	33	WD40
	Glc7p	36	**Protein Phosphatase 1 γ**	NP_002701.1		37	85	PP2Ac
			Ser/Thr specific protein Phosphatase	CAA52169.1		37	85	PP2Ac
	Yth1p	26	CPSF-30K	NP_006684.1	CPSF	30	48	Zn-finger, clipper
	Ssu72p	23	**PNAS-120**	AAaK07538.1		22.6	45	Ssu72
	Syc1p	21	**Ssu72 RNA polymerase II CTD phosphatase homolog**	NP_054907.1		22.5	45	Ssu72

Homolog proteins which are not known to belong to the cleavage complex in mammals or yeast are designated in bold.

References

Adesnik, M., Darnell, J.E. (1972) Biogenesis and characterization of histone messenger RNA in HeLa cells, J. Mol. B. **67** (3), 397-406

Aguilera, A. (2005) Cotranscriptional mRNP assembly: from the DNA to the nuclear pore, Curr. Opin. Cell. Biol. **17** (3), 242-250

Aravind, L. (1999) An evolutionary classification of the metallo-beta-lactamase fold proteins, In Silico Biol. **1** (2), 69-91

Åström, A., Åström, J., Virtanen, A. (1991) A simple procedure for isolation of eukaryotic mRNA polyadenylation factors, Eur. J. Biochem. **202** (3), 765-773

Awasthi, S., Alwine, J.C. (2003) Association of polyadenylation cleavage factor I with U1 snRNP, RNA, **9** (11), 1400-1409

Bai, C., Tolias, P.P. (1996) Cleavage of RNA hairpins mediated by a developmentally regulated CCCH zinc finger protein, Mol. Cell. Biol. **16** (12), 6661-6667

Balbo, P.D., Bohm, A. (2007) Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis, Structure **15** (9), 1117-1131

Barabino, S.M., Hübner, W., Jenny, A., Minvielle-Sebastia, L., Keller, W. (1997) The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins, Genes Dev. **11** (13), 1703-1716

Bardwell, V.J., Wickens, M., Bienroth, S., Keller, W., Sproat, B.S., Lamond, A.I. (1991) Site-directed ribose methylation identifies 2'-OH groups in polyadenylation substrates critical of AAUAAA recognition and poly(A) addition, Cell **65** (1), 125-133

Bedford, M.T., Frankel, A., Yaffe, M.B., Clarke, S., Leder, P., Richard, S. (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains, J. Biol. Chem. **275** (21), 16030-16036

Bessman, M.J., Frick, D.N., O'Handley, S.F. (1996) The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes, J. Biol. Chem. **271** (41), 25059-25062

Beyer, K., Dandekar, T., Keller, W. (1997) RNA ligands selected by cleavage stimulation factor contain distinct sequence motifs that function as downstream elements in 3' end processing of pre-mRNA, J. Biol. Chem. **272** (42) 26769-26779

Bienroth, S., Wahle, E., Suter-Crazzolara, C., Keller, W. (1991) Purification of the cleavage and polyadenylation factor involved in the 3' processing of mRNA precursors, J. Biol. Chem. **266** (29) 19768-19776

Bienroth, S., Keller, W., Wahle, E. (1993) Assembly of a processive mRNA polyadenylation complex, EMBO J. **12** (2), 585-594

Boisvert, F.-M., Côté, J., Boulanger, M.-C., Richard, S. (2003) A proteomic analysis of arginine-methylated protein complexes, Mol. Cell. Proteomics **2** (12), 1319-1330

Both, G.W., Furuichi, Y., Muthukrishnan, S., Shatkin, A.J. (1975) Ribosome binding to reovirus mRNA in protein synthesis requires 5' terminal 7-methylguanosine, Cell **6** (2), 185-195

Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. **72**, 248-254

Brahms, H., Raymackers, J., Union, A., de Keyser, F., Meheus, L., Lührmann, R. (2000) The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies, J. Biol. Chem. **275** (22), 17122-17129

Brown, K.M., Gilmartin, G.M. (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im, Mol. Cell **12** (6), 1467-1476

Callebaut, I., Moshous, D., Mornon, J.P., de Villartay, J.P. (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family, Nucleic Acids Res. **30** (16), 3592-3601

Calvo, O., Manley, J.L. (2001) Evolutionarily conserved interaction between CstF-64 and PC4 links transcription, polyadenylation, and termination, Mol. Cell **7** (5), 1013-1023

Calzado, M.A., Sancho, R., Muñoz, E. (2004) Human immunodeficiency virus type 1 Tat increases the expression of cleavage and polyadenylation specificity factor 73-kilodalton subunit modulating cellular and viral expression, J. Virol. **78** (13), 6846-6854

Carswell, S., Alwine, J.C. (1989) Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences, Mol. Cell. Biol. **9** (10), 4248-4258

Chen, C.-Y. A., Shyu, A.-B. (1995) AU-rich elements: characterization and importance in mRNA degradation, Trends Biochem. Sci. **20** (11), 465-470

Chen, F., MacDonald, C.C., Wilusz, J. (1995) Cleavage site determinants in the mammalian polyadenylation signal, Nucleic Acids Res. 23 (14), 2614-2620.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., Shiekhattar, R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing, Nature **436** (7051), 740-744

Christofori, G., Keller, W. (1988) 3' Cleavage and polyadenylation of mRNA precursors in vitro requires a poly(A) polymerase, a cleavage factor, and a snRNP, Cell **54** (6), 875-889

Colgan, D.F., Murthy, K.G.K., Prives, C., Manley, J.L. (1996) Cell-cycle related regulation of poly(A) polymerase by phosphorylation, Nature **384** (6606), 282-285

Colgan, D.F., Murthy, K.G.K., Zhao, W., Prives, C., Manley, J.L. (1998) Inhibition of poly(A) polymerase requires p34cdc/cyclin B phosphorylation of multiple consensus and non-consensus sites, EMBO J. **17** (4), 1053-1062

Conway, L., Wickens, M. (1985) A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3' termini in frog oocytes, Proc. Natl. Acad. Sci. USA, **82** (12), 3949-3953

Custódio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B.J., Carmo-Fonseca, M. (2004) In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei, RNA **10** (4), 622-633

Czaplinski, K., Köcher, T., Schelder, M., Segref, A., Wilm, M., Mattaj, I.W. (2005) Identification of 40LoVe, a Xenopus hnRNP D family protein involved in localizing a TGF-beta-related mRNA during oogenesis, Dev. Cell **8** (4), 505-515

Dantonel, J.-C., Murthy, K.G.K., Manley, J.L., Tora, L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA, Nature **389** (6649), 399-402

Dass, R., Attaya, E.N., Wallace, A.M., McDonald, C.C. (2001) Overexpression of the CstF-64 and CPSF-160 polyadenylation protein messenger RNAs in mouse male germ cells, Biol. Reprod. **64** (6), 1722-1729

Dass, R., McDaniel, L., Schultz, R.A., Attaya, E., McDonald, C.C. (2002) The gene CSTF2T, encoding the human variant CstF-64 polyadenylation protein tauCstF-64, lacks introns and may be associated with male sterility, Genomics **80** (5), 509-514

Deka, P., Rajan, P.K., Perez-Canadillas, J.M., Varani, G. (2005) Protein and RNA dynamics plays key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human CstF-64 protein, J. Mol. Biol. **347** (4), 719-733

Dettwiler, S. (2003) Characterization of factors involved in cleavage and polyadenylation of mammalian messenger RNA precursors, Thesis, University of Basel.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., Barabino, S.M. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor I mediate RNA binding, protein-protein interactions and subcellular localisation, J. Biol. Chem. **279** (34) 35788-35797

de Vries, H., Rüegsegger, U., Hübner, W., Friedlein, A., Langen, H., Keller, W. (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors, EMBO J. **19** (21), 5895-5904

DeZazzo, J.D., Imperiale, M.J. (1989) Sequences upstream of AAUAAA influences poly(A) site selection in a complex transcription unit, Mol. Cell. Biol. **9** (11), 4951-4961

Dickson, K.S., Bilger, A., Ballantyne, S., Wickens, M.P. (1999) The cleavage and polyadenylation specificity factor in Xenopus laevis oocytes is a cytoplasmic factor involved in regulated polyadenylation, Mol. Cell. Biol. **19** (8), 5707-5717

Dignam, J.D., Lebovitz, R.M., Roeder, R.G. (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei, Nucleic Acids Res. **11** (5) 1475-1489

Dominski, Z., Yang, X-c., Purdy, M., Wagner, E.J., Marzluff, W.F. (2005a) A CPSF-73 homologue is required for cell cycle progression but not for growth and interacts with a protein having features of CPSF-100, Mol. Cell. Biol. **25** (4), 1489-1500

Dominski, Z., Yang, X-c., Marzluff, W.F. (2005b) The polyadenylation factor CPSF-73 is involved in histone-pre-mRNA processing, Cell **123** (1), 37-48

Enomoto, A., Kido, N., Ito, M., Morita, A., Matsumoto, Y., Takamatsu, N., Hosoi, Y., Miyagawa, K. (2007) Negative regulation of MEKK1/2 signaling by Serine-Threonine kinase 38 (STK38), Oncogene, Advanced online publication, 01.10.2007

Fields, C., Adams, M.D., White, O., Venter, J.C. (1994) How many genes in the human genome? Nat. Genet. **7** (3), 345-346

Fitzgerald, M., Shenk, T. (1981) The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs, Cell **24** (1), 251-260

Flaherty, S.M., Fortes, P., Izaurralde, E., Mattaj, I.W., Gilmartin, G.M. (1997) Participation of the nuclear cap binding complex in pre-mRNA 3' end processing, Proc. Natl. Acad. Sci. USA **94** (22), 11893-11898

Fong, N., Bentley, D.L. (2001) Capping, splicing and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD, Genes Dev. **15** (14), 1783-1795

Ford, J.P., Hsu, M.-T. (1978) Transcription pattern of in vivo-labeled late simian virus 40 RNA : equimolar transcription beyond the mRNA 3' terminus, J. Virol. **28** (3), 795-801

Furuichi, Y., LaFiandra, A., Shatkins, A.J. (1977) 5'-Terminal structure and mRNA stability, Nature **266** (5599), 235-239

Gil, A., Proudfoot, N.J. (1984) A sequence downstream of AAUAAA is required for rabbit beta-globin mRNA 3'-end formation, Nature **312** (5993), 473-474

Gil, A., Proudfoot, N.J. (1987) Position-dependent sequence elements downstream of AAUAAA are required for efficient rabbit beta-globin mRNA 3' end formation, Cell **49** (3), 399-406

Gilmartin, G.M., McDevitt, M.A., Nevins, J.R. (1988) Multiple factors are required for specific RNA cleavage at a poly(A) addition site, Genes Dev. **2** (5), 578-587

Gilmartin, G.M., Nevins, J.R. (1989) An ordered pathway of assembly of components required for polyadenylation site recognition and processing, Genes Dev. **3** (12B), 2180-2189

Gilmartin, G.M., Nevins, J.R. (1991) Molecular analyses of two poly(A) site-processing factors that determine the recognition and efficiency of cleavage of pre-mRNA, Mol. Cell. Biol. **11** (5) 2432-2438

Graveley, B.R. (2000) Sorting out the complexity of SR proteins functions, RNA **6** (9), 1197-1211

Green, P. (1999) Interpreting the genome, Presentation at Bridging the Gap between Sequences and Function, Cold Spring Harbour Laboratory, NY, September 1999

Greenberg, J.R., Perry, P.R. (1972). Relative occurrence of polyadenylic acid sequences in messenger and heterogeneous nuclear RNA of L cells as determined by poly(U)-hydroxylapatite chromatography, J. Mol. Biol. **72** (1), 91-98

Hagenbüchle, O., Krikeles, M.S., Sprague, K.U. (1979) The nucleotide sequence adjacent to poly(A) in silk fibroin messenger RNA, J. Biol. Chem. **254** (15), 7157-7162

Handel, M.A., Hunt, P.A., Kot, M.C. (1991) Role of sex chromosomes in the control of male germ-cell differentiation, Ann. N.Y. Acad. Sci. **637**, 64-73

Handel, M.A. (2004) The XY body: a specialized meiotic chromatin domain, Exp. Cell. Res. **296** (1), 57-63

Hart, R.P., McDevitt, M.A., Ali, H., Nevins, J.R. (1985 a) Definition of essential sequences and functional equivalence of elements downstream of the adenovirus E2A and the early simian virus 40 polyadenylation sites, Mol. Cell. Biol. **5** (11), 2975-2983

Hart, R.P., McDevitt, M.A., Ali, H., Nevins, J.R. (1985) (Part 2) Poly(A) site cleavage in a HeLa nuclear extract is dependent on downstream sequences, Cell 43 (3), 677-683

Hartmuth, K., Urlaub, H., Vornlocher, H.-P., Will, C.L., Gentzel, M., Wilm, M., Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method, Proc. Natl. Acad. Sci. USA **99** (26), 16719-16724

Hartmuth, K., Vornlocher, H.-P., Lührmann, R. (2004) Tobramycin affinity tag purification of spliceosomes, Methods Mol. Biol. **257**, 47-64

Helaakoski, T., Vuori, K., Myllylä, R., Kivirikko, K.I., Pihlajaniemi, T. (1989) Molecular cloning of the alpha-subunit of human prolyl 4-hydroxylase: the complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts, Proc. Natl. Acad. Sci. USA **86** (12), 4392-6

Higgs, D.R., Goodbourn, S.E., Lamb, J., Clegg, J.B., Weatherall, D.J., Poudfoot N.J. (1983) Alpha-thalassaemia caused by a polyadenylation signal mutation, Nature **306** (5941), 398-400

Hirose, Y., Manley, J.L. (1997) Creatine phosphate, not ATP, is required for 3' end cleavage of mammalian pre-mRNA in vitro, J. Biol. Chem. **272** (47), 29636-29642

Hirose, Y., Manley, J.L. (1998) RNA polymerase II is an essential mRNA polyadenylation factor, Nature **395** (6697), 93-96

Hofmann, I., Schnölzer, M., Kaufmann, I., Franke, W.W. (2002) Symplekin, a constitutive protein of karyo- and cytoplasmic particles involved in mRNA biogenesis in Xenopus laevis oocytes, Mol. Biol. Cell **13** (5), 1665-1676

Holm, L., Sander, C. (1995) DNA polymerase beta belongs to an ancient nucleotidyltransferase superfamily, Trends Biochem. Sci. **20** (9), 345-347

Humphrey, T., Christofori, G., Lucijanic, V., Keller, W. (1987) Cleavage and polyadenylation of messenger RNA precursors in vitro occurs within large and specific 3' processing complexes, EMBO J. **6** (13), 4159-4168

Ingham, R.J., Colwill, K., Howard, C., Dettwiler, S., Lim, C.S.H., Yu, J., Hersi, K., Raadijmakers, J., Gish, G., Mbamalu, G., Taylor, L., Yeung, B., Vassilovski, G., Amin, M., Chen, F., Matskova, L., Winberg, G., Ernberg, I., Linding, R., O'Donell, P., Starostine, A., Keller, W., Metalnikow, P., Stark, C., Pawson, T. (2005) WW domains provide a platform for the assembly of multiprotein networks, Mol. Cell. Biol. **25** (16), 7092-7106

Izaurralde, E., Stepinski, J., Darzynkiewicz, E., Mattaj, I.W. (1992) A cap binding protein that may mediate nuclear export of RNA polymerase II-transcribed RNAs, J. Cell. Biol. **118** (6), 1287-1295

Jenny, A., Hauri, H.-P., Keller, W. (1994) Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit, Mol. Cell. Biol. **14** (12), 8183-8190

Jenny, A., Keller, W. (1995) Cloning of cDNAs encoding the 160 kDa subunit of the bovine cleavage and polyadenylation specificity factor, Nucleic Acids Res. **23** (14) 2629-2635

Jenny, A., Minvielle-Sebastia, L., Precker, P.J., Keller, W. (1996) Sequences similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I, Science **274** (5292), 1514-1517

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., Shoemaker, D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays, Science **302** (5653), 2141-4

Jove, R., Manley, J.L. (1984) In vitro transcription from the adenovirus 2 major late promoter utilizing templates truncated at promoter-proximal sites, J. Biol. Chem. **259** (13) 8513-8521

Jurica, M.S., Sousa D., Moore, M.J., Grigorieff, N. (2004) Three-dimensional structure of C complex spliceosomes by electron microscopy, Nat. Struct. Mol. Biol. **11** (3), 1147-1149

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** (3), 616-626

Keller, W., Bienroth, S., Lang, K.M., Christofori, G. (1991) Cleavage and polyadenylation factor CPF specifically interacts with the pre-mRNA 3' processing signal AAUAAA, EMBO J. **10** (13), 4241-4249

Kim, H., Lee, Y. (2001) Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I, Biochem. Biophys. Res. Commun. **289** (2), 513-518

Kim, H.-J., Jeong, S.-H., Heo, J.-H., Jeong, S.-J., Kim, S.-T., Youn, H.-D., Han, J.-W., Lee, H.-W., Cho, E.-J. (2004) mRNA capping enzyme activity is coupled to an early transcription elongation, Mol. Cell. Biol. **24** (14) 6184-6193

Kleiman, F.E., Manley, J.L. (1999) Functional interaction of BRCA1-associated BARD1 with polyadenylation factor CstF -50, Science **285** (5433), 1576-1579

Konarska, M.M., Padgett, R.A., Sharp, P.A. (1984) Recognition of the cap structure in splicing in vitro of mRNA precursors, Cell **38** (3), 731-736

Kubo, T., Wada, T., Yamaguchi, Y., Shimizu, A., Handa, H. (2006) Knock-down of 25 kDa subunit of cleavage factor Im in HeLa cells alters alternative polyadenylation within 3'-UTRs, Nucleic Acids Res. **34** (21), 6264-6271

Kyburz, A., Friedlein, A., Langen, H., Keller, W. (2006) Direct interactions between subunits of CPSF and the U2 snRNP contribute to the coupling of pre-mRNA 3' end processing and splicing, Mol. Cell **23** (2), 195-205

Kyburz, A. (2006) Characterization of factors involved in the coupling of 3' end processing and splicing and in the 3' end formation of mRNA precursors, Thesis, University of Basel

Lewis, J.D., Izaurralde, E., Jarmolowski, A., McGuigan, C., Mattaj, I.W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site, Genes Dev. **10** (13), 1683-1698

Lewis, J.D., Izaurralde, E. (1997) The role of the cap structure in RNA processing and nuclear export, Eur. J. Biochem. **247** (2), 461-469

Li, Y., Chen, Z.-Y., Wang, W., Baker, C.C., Krug, R.M. (2001) The 3'-end-processing factor CPSF is required for the splicing of single-intron pre-mRNA in vivo, RNA **7** (6), 920-931

Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L., Quackenbusch, J. (2000) Gene index analysis of the human genome estimates approximately 120,000 genes, Nat. Genet. **25** (2), 239-240

Licatalosi, D.D., Geiger, G., Minet, M., Schroeder, S., Cilli, K., McNeil, J.B., Bentley, D.L. (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II, Mol. Cell **9** (5), 1101-1111

MacDonald, C.C., Wilusz, J., Shenk, T. (1994) The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location, Mol. Cell. Biol. **14** (10), 6647-6654

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J., Reinberg, D. (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II, Proc. Natl. Acad. Sci. USA **101** (20), 7572-7577

Mandel, C.R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J.L, Tong, L. (2006) Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease, Nature **444** (7121), 953-956

Manley, J.L. (1983) Accurate and specific polyadenylation of mRNA precursors in a soluble whole-cell lysate, Cell **33** (2), 595-605

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** (10), 2593-2603

Martin, G., Jenö, P., Keller, W. (1999) Mapping of ATP binding regions in poly(A) polymerases by photoaffinity labeling and by mutational analysis identifies a domain conserved in many nucleotidyltransferases, Protein Sci. **8** (11), 2380-2391

Martin, G., Keller, W., Doublié, S. (2000) Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP, EMBO J. **19** (16), 4193-4203

Martin, G., Keller, W. (2004) Sequence motifs that distinguish ATP(CTP):tRNA nucleotidyl transferases from eubacterial poly(A) polymerases, RNA **10** (6), 899-906

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

Mason, P.J., Elkington, J.A., Lloyd, M.M., Jones, M.B., Williams, J.G. (1986) Mutations downstream of the polyadenylation site of a Xenopus beta-globin mRNA affect the position but not the efficiency of 3' processing, Cell **46** (2), 263-270

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription, Nature **385** (6614), 357-361

McDevitt, M.A., Hart, R.P., Wong, W.W., Nevins, J.R. (1986) Sequences capable of restoring poly(A) site function define two distinct downstream elements, EMBO J. **5** (11), 2907-2913

McLauchlan, J., Gaffney, D., Whitton, J.L., Clements, J.B. (1985) The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3' termini, Nucleic Acids Res. **13** (4), 1347-1368

Meinhart, A., Cramer, P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors, Nature **430** (6996), 223-226

Meister, G., Bühler, D., Pillai, R., Lottspeich, F., Fischer, U. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs, Nat. Cell. Biol. **3** (11), 945-949

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., Vagner, S. (2006) An interaction between U2AF 65 and CF I_m links the splicing and 3' end processing machineries, EMBO J. **25** (20), 4854-4864

Minvielle-Sebastia, I., Keller, W. (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription, Curr. Opin. Cell. Biol. **11** (3), 352-257

Mitchelson, A., Simonelig, M., Williams, C., O'Hare, K. (1993) Homology with Saccharomyces cerevisiae RNA14 suggests that phenotypic suppression in Drosophila melanogaster by suppressor of forked occurs at the level of RNA stability, Genes Dev. **7** (2), 241-249

Monarez, R.R., MacDonald, C.C., Dass, B. (2007) Polyadenylation proteins CstF-64 and tauCstF-64 exhibit differential binding affinities for RNA polymers, Biochem. J. **401** (3), 651-658

Montell, C., Fischer, E.F., Caruthers, M.H., Berk, A.J. (1983) Inhibition of RNA cleavage but not polyadenylation by a point mutant in mRNA 3' consensus sequence AAUAAA, Nature **305** (5935), 600-605

Moore, C.L., Skolnik-David, H., Sharp, P.A. (1988) Sedimentation analysis of polyadenylation-specific complexes, Mol. Cell. Biol. **8** (1), 226-233

Murthy, K.G.K., Manley, J.L. (1992) Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus, J. Biol. Chem. **267** (21), 14804-14811

Murthy, K.G.K., Manley, J.L. (1995) The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation, Genes Dev. **9** (21), 2672-2683

Muthukrishnan, S., Both, G.W., Furuichi, Y., Shatkin, A.J. (1975) 5'-Terminal 7-methylguanosine in eukaryotic mRNA is required for translation, Nature **255** (5503), 33-37

Nesterenko, M.V., Tilley, M., Upton, S.J. (1994) A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels, J. Biochem. Biophys. Methods **28** (3), 239-242

Nevins, J.R., Darnell Jr., J.E. (1978) Steps in the processing of Ad2 mRNA: Poly(A)+ nuclear sequences are conserved and poly(A) addition precedes splicing, Cell **15**, 1477-1493

Pahlich, S., Zakaryan, R.P., Gehring, H. (2006) Protein arginine methylation: Cellular functions and methods of analysis, Biochim. Biophys. Acta **1764** (12), 1890-1903

Paushkin, S.V., Patel, M., Furia, B.S., Peltz, S.W., Trotta, C.R. (2004) Identification of a human endonuclease complex reveals a link between tRNA splicing and pre-mRNA 3' end formation, Cell **117** (3), 311-321

Pennisi, E. (2007) Working the (gene count) numbers: finally, a firm answer? Science **316** (5828), 1113

Preker, P.J., Keller, W. (1998) The HAT helix, a repetitive motif implicated in RNA processing, Trends Biochem. Sci. **23** (1), 15-16

Prescott, J., Falck-Pedersen, E. (1992) Varied poly(A) site efficiency in the adenovirus major late transcription unit, J. Biol. Chem. **267** (12), 8175-8181

Prescott, J., Falck-Pedersen, E. (1994) Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites, Mol. Cell. Biol. **14** (7), 4682-4693

Proudfoot, N. (2004) New perspectives on connecting mRNA 3' end formation to transcription, Curr. Opin. Cell. Biol. **16** (3), 272-278

Proudfoot, N.J., Brownlee, G.G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA, Nature **263** (5574), 211-214

Qu, X., Perez-Canadillas, J.-M., Agrawal, S., De Baecke, J., Cheng, H., Varani, G., Moore, C. (2007) The C-terminal domains of vertebrate CstF-64 and its yeast orthologue Rna15 form a new structure critical for mRNA 3' end processing, J. Biol. Chem. **282** (3), 2101-2115

Rasmussen, E.B., Lis, J.T. (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes, Proc. Natl. Acad. Sci. USA **90** (17), 7923-7927

Raue, U., Oellerer, S., Rospert, S. (2007) Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence, J. Biol. Chem. **282** (11), 7809-78

Reed, R. (2003) Coupling transcription, splicing and mRNA export, Curr. Opin. Cell. Biol. **15** (3), 326-331

Reed, R., Cheng, H. (2005) TREX, SR proteins and export of mRNA, Curr. Opin. Cell. Biol. **17** (3), 269-273

Reed, R., Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing, Cell **108** (4), 523-531

Richard, S., Morel, M., Cléroux, P. (2005) Arginine methylation regulates IL-2 gene expression: a role for protein arginine methyltransferase 5 (PRMT5), Biochem. J. **388** (Pt 1), 379-386

Rüegsegger, U., Beyer, K., Keller, W. (1996) Purification and characterization of human cleavage factor Im involved in the 3' end processing of mRNA precursors, J. Biol. Chem. **271** (11), 6107-6113

Rüegsegger, U. (1997) Characterization of factors involved in 3' end processing of mammalian messenger RNA precursors, Thesis, University of Basel.

Rüegsegger, U., Blank, D., Keller, W. (1998) Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits, Mol. Cell **1** (2), 243-253

Ryan, K., Murthy, K.G.K., Kaneko, S., Manley, J.L. (2002) Requirements of the RNA polymerase II C-terminal domain for recognition pre-mRNA 3' cleavage, Mol. Cell. Biol. **22** (6), 1684-1692

Ryan, K., Calvo, O., Manley, J.L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease, RNA **10** (4), 565-573

Ryan, K. (2007) Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation, RNA Biol. **4** (1), 26-33

Sadowski, M., Dichtl, B., Hübner, W., Keller, W. (2003) Independent functions of the yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination, EMBO J. **22** (9), 2167-2177

Salisbury, J., Hutchison, K.W., Graber, J.H. (2006) A multispecies comparison of the metazoan 3'- processing downstream elements and the CstF-64 RNA recognition motif, BMC Genomics **7**: 55

Sambrook, J., Fritsch, E.F., Mantiatis, T. (2001) Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory Press, New York

Schek, N., Cooke, C., Alwine, J.C. (1992) Definition of the upstream efficiency element of the simian virus 40 late polyadenylation signal by using in vitro analyses, Mol. Cell. Biol. **12** (12), 5386-5393

Simonelig, M., Elliot, K., Mitchelson, A., O'Hare, K. (1996) Interallelic complementation at the *suppressor of forked* locus of drosophila reveals complementation between Suppressor of forked proteins mutated in different regions, Genet. **142**, 1225-1235

Skolnik-David, H., Moore, C.L., Sharp, P.A. (1987) Electrophoretic separation of polyadenylation-specific complexes, Genes Dev. **1** (7), 672-682

Scott, R. (1999) The future in understanding the molecular basis of life, Presentation at Eleventh International Genome Sequencing and Analysis Conference, Miami, 1999.

Shuman, S. (2001) Structure, mechanism, and evolution of the mRNA capping apparatus, Prog. Nucleic Acid Res. Mol. Biol. **66**, 1-40

Takagaki, Y., Ryner, L.C., Manley, J.L. (1988) Separation and characterization of a poly(A) polymerase and a cleavage/specificity factor required for pre-mRNA polyadenylation, Cell **52** (5), 731-742

Takagaki, Y., Ryner, L.C., Manley, J.L. (1989) Four factors are required for 3'-end cleavage of pre-mRNAs, Genes Dev. **3** (11), 1711-1724

Takagaki, Y., Manley, J.L., MacDonald, C.C., Wilusz, J., Shenk, T. (1990) A multisubunit factor, CstF, is required for polyadenylation of mammalian pre-mRNAs, Genes Dev. **4** (12A), 2112-2120

Takagaki, Y., Manley, J.L. (1992) A human polyadenylation factor is a G protein beta-subunit homologue, J. Biol. Chem. **267** (33), 23471-23474

Takagaki, Y., Manley, J.L. (1994) A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein, Nature **372** (6505), 471-474

Takagaki, Y., Manley, J.L. (1997) RNA recognition by the human polyadenylation factor CstF, Mol. Cell. Biol. **17** (7), 3907-3914

Takagaki, Y., Manley, J.L. (2000) Complex protein interaction within the human polyadenylation machinery identify a novel component, Mol. Cell. Biol. **20** (5), 1515-1525

Valsamakis, A., Zeichner, S., Carswell, S., Alwine, J.C. (1991) The human immunodeficiency virus type 1 polyadenylylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation, Proc. Natl. Acad. Sci. USA **88** (6), 2108-2112

Venkataraman, K., Brown, K.M., Gilmartin, G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition, Genes Dev. **19** (11), 1315-1327

Venter, J.G., Adams, M.D., *et al*. (2001) The sequence of the human genome, Science **291** (5507), 1304-1351

Veraldi, K.L., Edwalds-Gilbert, G., MacDonald, C.C., Wallace, A.M., Milcarek, C. (2000) Isolation and characterization of polyadenylation complexes assembled in vitro, RNA **6** (5), 768-777

Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Gangster, L.-H., Wilusz, J., Milcarek, C. (2001) hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells, Mol. Cell. Biol. **21** (4), 1228-1238

Wahle, E. (1991a) A novel poly(A) binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation, Cell **66** (4), 759-768

Wahle, E. (1991b) Purification and characterization of a mammalian polyadenylate polymerase involved in the 3' end processing of messenger RNA precursors, J. Biol. Chem. **266** (5), 3131-3139

Wahle, E. (1995) Poly(A) tail length control is caused by termination of processive synthesis, J. Biol. Chem. **270** (6), 2800-2808

Wahle, E., Keller, W. (1992) The biochemistry of 3' end cleavage and polyadenylation of messenger RNA precursors, Annu. Rev. Biochem. **61**, 419-440

Wahle, E., Keller, W. (1994) RNA processing, Oxford University press, Vol. II, 5-6

Wahle, E., Lustig, A., Jenö, P., Maurer, P. (1993) Mammalian poly(A) binding protein II. Physical properties and binding to polynucleotides, J. Biol. Chem. **268** (4), 2937-2945

Wahle, E., Rüegsegger, U. (1999) 3'-End processing of pre-mRNA in eukaryotes, FEMS Microbiol. Rev. **23** (3), 277-295

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, EMBO J. **1** (8), 945-951

Wallace, A.M., Dass, B., Ravnik, S.E., Tonk, V., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., MacDonald, C.C. (1999) Two distinct forms of the 64,000 Mr protein of the cleavage stimulation factor are expressed in mouse male germ cells, Proc. Natl. Acad. Sci. USA **96** (12), 6763-6768

Wallace, A.M., Denison, T.L., Attaya, E.N., MacDonald, C.C. (2004) Developmental distribution of the polyadenylation protein CstF-64 and the variant tauCstF-64 in mouse and rat testis, Biol. Reprod. **70** (4), 1080-1087

Wedekind, J.E., Dance, G.S., Sowden, M.P., Smith, H.C. (2003) Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business, Trends Genet. **19** (4), 207-216

Weiss, E.A., Gilmartin, G.M., Nevins, J.R. (1991) Poly(A) site efficiency reflects the stability of complex formation involving the downstream element, EMBO J. **10** (1), 215-219

Weitzer, S., Martinez, J. (2007) The human RNA kinase hClp1 is active on 3' transfer RNA exons and short interfering RNAs, Nature **447** (7141), 222-227

Wetterberg, I., Baurén, G., Wieslander, L. (1996) The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns, RNA 2 (7), 641-651

Wickens, M., Stephenson, P. (1984) Role of the conserved AAUAAA sequence: four AAUAAA point mutants prevent mRNA 3' end formation, Science **226** (4678), 1045-1051

Wilusz, J., Shenk, T. (1990) An uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal, Mol. Cell. Biol. **10** (12), 6397-6407

Wittmann, T., Wahle, E. (1997) Purification and characterization of full-length mammalian poly(A) polymerase, Biochim. Biophys. Acta **1350** (3), 1244-1248

Wlotzka, W. (2006) A new approach to purify and characterise cleavage and polyadenylation specificity factor involved in 3' end processing of messenger RNA precursors, Diploma thesis, University of Hannover.

Zarkower, D., Wickens, M. (1988) A functional redundant downstream element in SV40 late pre-mRNA is required for mRNA 3'-end formation and for assembly of a precleavage complex in vitro, J. Biol. Chem. **263** (12), 5780-5788

Zarudnaya, M.I., Kolomiets', I.M., Hovorun, D.M. (2002) What nuclease cleaves pre-mRNA in the process of polyadenylation? IUBMB Life **54** (1), 27-31

Zhao, J., Hyman, L., Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis, Microbiol. Mol. Biol. Rev. **63** (2), 405-445

Acknowledgement

I would like to thank …

my supervisor **Elmar Wahle** for giving me the opportunity to do my Ph.D. in his lab. I am grateful for his continuous support, discussion and comments for this work also during my time in Basel.

Walter Keller, who allowed me to continue my thesis at the Biozentrum; his constant interest in the progress of this work and helpful comments.

Uwe Kühn, who taught me a lot of my scientific tools during my diploma thesis and in the beginning of this thesis. This was an excellent basis. Thanks for his helpful suggestions, discussions and the correction of this work.

Paul Jenö, Suzette Moes and **Arno Friedlein** for doing the mass spectrometry analysis.

Christiane Rammelt for helpful and interesting discussions and suggestion, correcting this work, being a friend and bringing a feeling of home to Basel.

Andrea Kyburz, who introduced my to the his_8-flag tag topic, her interest and discussions till the birth of Lea.

Klaus Hartmuth and **Irene Ochsner** for introducing me to the tobramycin-affinity selection method.

Verena Widmer, who helped me a lot in cell culture and prepared NXT-B.

all present and previous members of the Keller and the Wahle Lab for discussion and support.

all members of the lab 403 especially **Gudrun** and **Andrea** as well as all the previously diploma students of Uwe: **Henning, Katharina, Sabine, Anke, Miriam** and **Lisette**.

Marty Lynch, spending some early morning hours on the telephone to discuss the English comma rules.

my friends who accompanied me: **Nils, Nadine, Franzi, Kristina** and **Knut** (thanks for the correction of this work) and spent time with me apart from the Labs.

My largest thanks go to **my parents**, **my sister** and **her family**, and my partner **Markus** for support, motivation and their love.

Curriculum vitae

Name: Anne Ulrika Knoth
Date of birth: 04.03.1979
Place of birth: Halle/Saale

Education and Academic Training:

1985-1987	Polytechnische Oberschule „Adam Kuckhoff“, Halle
1987-1991	Oberschule mit erweitertem Fremdsprachenunterricht „Dr. Kurt Fischer“ / „Johann-Gottfried Herder“, Halle
1991-1997	Gymnasium „Johann-Gottfried Herder“, Halle
1997	A-level (Abitur)
1997-2002	Studies in biochemistry at the Martin-Luther-University Halle-Wittenberg
2000	2-month practical training in Laboratory of Dr. M. Mostiack Vieira in Laboratorio Clinico Santa Rita LTDA, Caçador, Brazil
2001-2002	Diploma thesis in the laboratory of Prof. E. Wahle (“Untersuchungen zur Längenkontrolle der Polyadenylierung”)
2002	Diploma
2002-2005	Ph.D. student at the Institute of Biochemistry / Biotechnology of the Martin-Luther-University Halle-Wittenberg (“Analysis of the mammalian pre-mRNA 3’ end processing cleavage complex and attempts of its reconstitution”)
2005-2006	18-month guest Ph.D. student in the laboratory of Prof. W. Keller at the Biozentrum of the University of Basel /Switzerland
2006-2008	Ph.D. student at the Institute of Biochemistry/Biotechnology of the Martin-Luther-University Halle-Wittenberg (“Analysis of the mammalian pre-mRNA 3’ end processing cleavage complex and attempts of its reconstitution”)

Publication:

Kerwitz, Y., Kuhn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E. and Wahle, E. (2003) Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. EMBO J, 22, 3705–3714.

Kühn, U., Gündel, M., Knoth, A., Rüdel, S., Kerwitz, Y., Wahle, E. (2008) The mechanism of poly(A) tail length control. Manuscript in preparation.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen Hilfsmittel als die angegebenen benutzt habe. Die anderen Arbeiten entnommenen Angaben, wurden dem Wortlaut oder Sinn entsprechend wiedergegeben und sind durch Quellenangaben gekennzeichnet.

_____

Halle, Januar 2008